On the generic geometry of parametrized polynomial systems in biology and statistics

Oskar Henriksson

This thesis has been submitted to the PhD School of
The Faculty of Science, University of Copenhagen

Oskar Henriksson
Department of Mathematical Sciences, University of Copenhagen
Universitetsparken 5, 2100 København Ø, Denmark
oskar.henriksson@math.ku.dk

This thesis has been submitted to the PhD School of the Faculty of Science,
University of Copenhagen, Denmark, on the 31st of December 2024.

Academic advisor:	Elisenda Feliu University of Copenhagen, Denmark
Assessment committee:	Paul Breiding University of Osnabrück, Germany
	Fatemeh Mohammadi KU Leuven, Belgium
	Fabien Pazuki (chair) University of Copenhagen, Denmark

ISBN 978-87-7125-238-5

Acknowledgements

First and foremost, I want to express my deepest gratitude to my advisor, Elisenda Feliu, for introducing me to the beautiful and exciting world of applied algebraic geometry, and for all your encouragement, guidance and unwavering support.

I would also like to thank all past and present members of the applied algebra and geometry group in in Copenhagen who I have been fortunate to get to know: AmirHosein Sadeghimanesh, Angélica Torres, Beatriz Pascual-Escudero, Laura Brustenga i Moncusí, Máté Telek, Nidhi Kaihnsa, Joan Ferrer Rodríguez, and Carles Checa. Thanks for generously sharing your knowledge, and for filling my time in Copenhagen with interesting conversations, laughter, and lots of *hygge*.

Next, I want to direct a big thank you to all my amazing collaborators: Carlos Améndola, Balázs Boros, Diego Cifuentes, Gheorghe Craciun, Maize Curiel, Jan Draisma, Elisenda Feliu, Paul Helminck, Jiaxin Jin, Annachiara Korchmaros, Kaie Kubjas, Beatriz Pascual-Escudero, Kristian Ranestad, Yue Ren, Jose Rodriguez, Diego Rojas La Luz, Benjamin Schröter, Lisa Seccia, Máté Telek, Polly Yu, and Teresa Yu. Thank you for making every project we have worked on such a fun and rewarding experience.

A special thanks to Jose Rodriguez, as well as Gheorghe Craciun, Diego Rojas La Luz and Thomas Yahl, for being wonderful hosts during my research visit in Madison, and for contributing to making this a very productive and memorable time of my PhD.

I am also deeply appreciative of Nina Weisse, Morten Haupt Petersson, and the rest of the administrative team at the Department of Mathematical Sciences in Copenhagen, for providing excellent administrative assistance throughout my PhD studies.

My mathematical journey began long before Copenhagen, and I want to thank my school teachers, especially Gullevi Salomonsson, Tobias Roos, Per Wilhelmsson, Stefan Rosén, and Gulli Jansson, for going above and beyond in supporting my interest in science. I'm also deeply grateful to my academic mentors during my BSc and MSc studies, who all had a formative impact on my scientific development: Kenneth Wärnmark, Sigmundur Gudmundsson, Arne Meurman, Nathalie Wahl, and Andrea Bianchi.

Finally, to my beloved family: There is no way I could have done this without your endless love and support. Thank you for always being there to cheer me on.

Oskar Henriksson
Copenhagen, December 2024

Summary

This thesis is based on seven papers about polynomial systems of equations arising in biology and statistics which are *parametrized*, in the sense that their coefficients depend on parameters. Using techniques from algebraic geometry, the papers explore various properties of the solution sets of these systems that hold for generic parameter values.

Paper A concerns the problem of numerically approximating the solutions of systems with generically finite solution sets. We give an algorithm for constructing homotopies and start systems from tropical intersection data, that allows numerically solving such systems with homotopy continuation methods in a way that leads to a generically optimal number of paths. We also develop techniques for computing the necessary tropical intersection data and generic root counts for several classes of systems, including steady state systems of chemical reaction networks.

Papers B, C and D are centered around vertically parametrized systems, which arise in both chemical reaction network theory and optimization. In Paper B, we draw up a general framework for studying generic consistency and nondegeneracy of such systems over both the complex and real numbers. This framework is then applied in Paper C to study reaction-network-theoretic properties such as absolute concentration robustness and nondegenerate multistationarity, and in Paper D to investigate parametric toricity.

Papers E and F focus on the method of moments for statistical parameter estimation. In Paper E, we study the determinantal structure and singularities of the moment varieties of the exponential, chi-squared, gamma, and inverse Gaussian distribution. We also determine the number of moments needed to obtain generic unique parameter identifiability for mixtures of the exponential and chi-squared distribution. In Paper F, we build on the results from Paper E, combined with the theory of secant defectivity of surfaces, to determine the number of moments needed to obtain generic finite identifiability for mixtures of the inverse Gaussian and gamma distribution.

Finally, in Paper G, we investigate the problem of identifiability in the field of 3D genome reconstruction for diploid organisms. We prove generic finite identifiability from unphased Hi-C data, and also show that the algebraic complexity of the problem significantly decreases in the presence of even a small amount of phased data that distinguishes maternal and paternal genomic loci. Based on this result, we devise a new reconstruction approach, based on homotopy continuation and optimization.

Dansk resumé

Denne afhandling tager udgangspunkt i syv artikler, der omhandler polynomielle ligningssystemer, der opstår i biologi og statistik, og som er *parametriserede* i den forstand, at deres koefficienter afhænger af parametre. Ved hjælp af teknikker fra algebraisk geometri udforskes i artiklerne forskellige egenskaber ved løsningsmængderne til disse systemer for generiske parameterværdier.

Artikel A omhandler problemet med numerisk tilnærmelse af løsningerne til systemer med generisk endeligt antal løsninger. Vi giver en algoritme til at konstruere homotopier og startsystemer fra tropiske skæringspunkter, hvilket muliggør numerisk løsning af sådanne systemer med homotopifortsættelse med et generisk optimalt antal baner. Vi udvikler også teknikker til at beregne de nødvendige tropiske skæringsdata og det generiske antal rødder for flere klasser af systemer.

Artiklerne B, C og D fokuserer på vertikalt parametriserede systemer, som opstår inden for teorien om kemiske reaktionsnetværk og optimering. I Artikel B opstiller vi en generel ramme til at undersøge om et sådant system generisk er løsbart og ikke-degenereret over både de komplekse og de reelle tal. Denne ramme anvendes i Artikel C til at studere reaktionsnetværksteoretiske egenskaber såsom absolut koncentrationsrobusthed og ikke-degenereret multistationaritet, og i Artikel D til at undersøge parametrisk toricitet.

Artiklerne E og F behandler momentmetoden til statistisk estimation af parametrer. I Artikel E undersøger vi den determinantiske struktur og singulariteterne i momentvarieteterne ved eksponential-, chi-i-anden-, gamma- og invers-gauss-fordelingen. Vi bestemmer også antallet af momenter, der er nødvendigt for at opnå generisk unik parameteridentifikation for mixturer af eksponential- og chi-i-anden-fordelingen. I Artikel F bygger vi på resultaterne fra Artikel E, kombineret med teorien om fladers sekantdefektivitet, til at bestemme antallet af momenter, der er nødvendige for at opnå generisk endelig identificerbarhed for mixturer af invers-gauss- og gammafordelingen.

I Artikel G undersøger vi spørgsmålet om identificerbarhed inden for 3D-rekonstruktion av genom for diploide organismer. Vi beviser, at vi har generisk endelig identificerbarhed ud fra ufasede Hi-C-data, og at problemets algebraiske kompleksitet falder betydeligt hvis en lille mængde fasede data er til stede, der skelner mellem genomiske loci fra moderen og faderen. Motiveret af dette resultat foreslår vi en ny rekonstruktionsmetode baseret på homotopifortsættelse og optimering.

Contents

Papers

1 Introduction

This thesis lies in the field of *applied algebraic geometry*, where one studies the geometry of algebraic varieties—solution sets of polynomial systems—with a view towards applications. More specifically, the thesis focuses on polynomial systems that involve *parameters*, which are typically assumed to be unknown or varying. Understanding all possible geometries that the solution set can attain as the parameter values vary is typically very challenging, but for many properties there is a *generic* behavior that is common for all parameters except those in a subset of measure zero in parameter space. The overarching goal of the thesis is to understand various such generic properties for systems arising in three applications—chemical reaction network theory, statistical parameter estimation, and 3D genome reconstruction—as well as the broader problem of effectively solving generic instances of parametrized systems.

Below, we give an overview of each of these areas, with focus on the background and the main contributions of the thesis.

1.1 Numerical algebraic geometry

The problem of numerically approximating solutions to systems of equations is as old as numerical analysis itself, with iterative methods based on Newton's method being a staple in many areas of applied mathematics [128]. While having wide applicability to a wide range of systems, many of these methods only find single solutions, without providing any systematic way to find them all. One of the central goals of numerical algebraic geometry is to find approximations of *all* complex solutions of a polynomial system, through a technique called *homotopy continuation*, which was developed in this context in a series of works in the second half of the 20th century, including [45, 59, 143].

One of the challenges of these techniques, is that they require *a priori* knowledge of upper bounds on the number of complex solutions of the system. This is typically accomplished by viewing the target system at hand as an instance of a parametrized system, and using the *generic root count* of the parametrized system as an upper bound. The original works on homotopy continuation mostly focused on the total degree bound provided by Bézout's theorem, but a breakthrough took place in 1995 when Huber and

Sturmfels in [80] gave an algorithm for constructing homotopies based on the mixed volume bound of Bernstein [17], Khovanskii [90] and Kouchnirenko [92], which takes into account the Newton polytopes of the system at hand, not just the total degree.

Since then, many other methods have been developed for computing sharper bounds and associated homotopies, by taking into account more refined structural features that systems appearing applications can exhibit, using tools such as decomposable projections [6], tropical geometry [100], and Khovanskii bases [29].

The main contribution of Paper A is a general algorithm for computing start systems and homotopies for parametrized systems with generically finitely many solutions from tropical intersection data. These homotopies are generically optimal, in the sense that they give rise to the generic root count many paths to trace. With this, we generalize the polyhedral homotopies of Huber–Sturmfels [80], and the more recent tropical techniques of Leykin–Yu [100]. We also develop techniques for making the tropical intersection data computable for specific classes of parametrized systems (including the vertically parametrized systems of Chapter 4), and implement these in a Julia package based on the computer algebra system `OSCAR` [121], and `HomotopyContinuation.jl` [27].

1.2 The structure of steady state equations

The field of reaction network theory aims to qualitatively understand the dynamics of networks of interaction processes in fields like chemistry, cellular biology and ecology, usually without explicit knowledge of parameters appearing in the models. In its current form, the field originates from work by Horn and Jackson [78] and Feinberg [56] in the 1970s.

The use of algebraic-geometric techniques in the field is much more recent, and goes back to work by Gatermann [62], and the seminal paper on toric dynamical systems by Craciun, Dickenstein, Shiu, and Sturmfels [41]. These techniques have led to fruitful progress on many topics, especially regarding the set of steady states, including concepts such as multistationarity [37, 39, 51], and in particular bistability [83, 138, 140], as well as global stability [7, 28, 64], boundary steady states and persistence [48, 63, 69, 110, 135], absolute concentration robustness [61, 114, 133], and model selection [71, 74, 107, 108].

Yet, many fundamental geometric questions about the geometry of the set of steady states remain open. In Papers B, C and D, we address several such questions, by focusing on a particular *vertically parametrized* structure that the systems that describe the steady states display under the assumption of power-law kinetics.

Paper B lays the foundation for the other two papers. Here, we prove that the incidence varieties of vertically parametrized systems are smooth and irreducible, and give a rank condition that characterizes when a vertically parametrized system is generically consistent, and have a zero locus that is generically smooth of expected

dimension. We also give versions of these results that hold over the real and positive numbers.

In Paper C, we apply the results from Paper B to answer several geometric questions and clarify example-based observations that have appeared in the literature over the years. As an immediate consequence of the results from Paper B, we prove a generic finiteness result that answers a question from [22]. We also give an ideal-theoretic characterization of generic absolute concentration robustness, which strengthens sufficient conditions for fixed parameter values developed in [61, 115] and a linear algebra condition that characterizes the weaker local version of this property introduced in [123]. We furthermore prove part of a conjecture about nondegenerate multistationarity by Joshi and Shiu [85].

In Paper D, we use the framework from Papers B and C to study the problem of detecting when the set of positive steady states equals a scaling of a given positive toric variety for generic (or all) parameter values. This is an old problem that goes back to the concept of deficiency theory and complex balancing in the works by Horn, Jackson and Feinberg [56, 57, 78], as well as the work on toric dynamical systems [41], and is motivated by the fact that it simplifies the analysis of multistationarity [116, 118]. Recently, the problem has been approached with quantifier elimination techniques [126], and through various sufficient conditions based on binomiality [38, 116, 127, 131]. In Paper D, we give a *necessary* linear algebra condition for parametric toricity, and a range of new sufficient conditions. We also introduce and study the weaker concept of *local toricity*, where each connected component of the set of positive steady states is a scaling of a given positive toric variety.

1.3 Parameter identifiability from moments

The *method of moments* for statistical parameter estimation was first proposed in Pearson's 1894 paper [125], where he analyzed data (on the size of two crab populations) that was assumed to come from a mixture of two Gaussian distributions. The idea is that the moments of many classical distributions are polynomial in the distribution parameters, meaning that said parameters can be estimated from sample moments by solving a system of polynomial equations.

In the 19th century, solving large multivariate polynomial systems was a challenging task. In Pearson's case of a mixture of two Gaussians, he was able to reduce the problem to solving a univariate degree-9 polynomial, and wrote himself in [125]:

> "*The analytical difficulties ... are so considerable, that it may be questioned whether the general theory could ever be applied in practice ...*"

His colleague Carl Charlier wrote (see, e.g., [113, page 3]):

> "*Mr. Pearson has indeed possessed the energy to perform this heroic task in some instances ... but I fear that he will have few successors.*"

Nevertheless, with the advent of numerical analysis, the method of moments has become a standard technique in mathematical statistics, especially when combined with various optimization techniques [73]. Over the last two decades, the method of moments has also attracted newfound theoretical interest from an algebraic point of view, when it comes to the problem of *identifiability*. In 2004, Lazard proved the first unique identifiability result for Guassian mixture distributions [97], and the first broader treatment of moments from the point of view of algebraic geometry appeared in Belkin and Sinha's 2010 paper [13]. Another geometric perspective was introduced in [4], when Améndola, Faugère and Sturmfels introduced the concept of moment varieties, which has later been followed up by several works with an algebraic viewpoint concerning identifiability of various types of multivariate Gaussian mixtures [1, 4, 5, 18, 19, 104], as well as estimation techniques based on numerical algebraic geometry [6, 104].

Moment identifiability for other distributions is a much less explored topic. In [91], the authors study uniform distributions on polytopes, and mixtures distributions have been treated in the Dirac and Pareto case [68], as well as for product distributions [2]. In Papers E and F, we take some new toward extending the theory of mixture distributions beyond the Gaussian case. In Paper E, we determine the number of moments needed to have generic unique identifiability for mixtures of the exponential and chi-squared distribution, and investigate experimentally the number of moments required to have generic finite identifiability for mixtures of the gamma and inverse Gaussian distribution. We also undertake a detailed study of the moment varieties for these distributions, prove that they all admit determinantal realizations, and describe the Hilbert series and singularities for the inverse Gaussian and gamma distribution. Using the information about the singular loci, we prove in Paper F our conjectured condition for generic finite identifiability from Paper E, using the theory of secant defective surfaces and intersection theory.

1.4 Identifiability in 3D genome reconstruction

The field of *3D genome reconstruction* is a relatively new area in genomics, stemming from the observation that not only the DNA sequence itself, but also the *spatial organization* of the chromosomes plays an important role in how genes are expressed. In its current form, the field emerged in the early 2000's, when an experimental setup was proposed in [46] for estimating the relative pairwise distances between points on a chromosome. This later turned into a technique called *high-throughput chromosome conformation capture* (Hi-C) [102].

In the case of *haploid* cells (where there is a single copy of each chromosome), it is well-understood that Hi-C data can be used to successfully infer the 3D structure of a chromosome, but for *diploid* cells (where there are two copies of each chromosome), the situation is complicated by the fact that most Hi-C data will be unphased (i.e., completely agnostic about which genes come from the maternal or paternal copy of a chromosome). In practice, most methods rely on combining unphased Hi-C data with additional data, and Segal asks rhetorically "Can 3D diploid genome reconstruction from unphased Hi-C data be salvaged?" in the title of the review paper [132].

A fundamental question that lends itself to algebraic techniques is whether we at all have identifiability from unphased Hi-C data. The first algebraic work in this direction came in 2022, when Belyaeva, Kubjas, Sun, and Uhler in [14] proved non-identifiability under a simple model for how Hi-C data depends on the configuration, and studied various types of additional data that is sufficient to obtain identifiability.

In Paper G, we study another and more common model for how Hi-C data depends on the structure, and prove generic finite identifiability, but with a very large identifiability degree. We furthermore investigate how adding some *phased* data (that distinguishes maternal and paternal genetic loci) influences the identifiability properties, inspired by the simulation-based work [32]. We show that even a small amount of phased data gives rise to a stronger *local* identifiability property, with a significantly lower identifiability degree. Based on this result, we also propose a new reconstruction approach, where first, the configuration of the distinguishable loci is estimated with standard semi-definite programming methods, before the configuration of the indistinguishable loci is estimated in a two-stage procedure: first individually with numerical algebraic geometry, and then collectively through local optimization.

Structure of the thesis

The thesis is based on seven papers, which are gathered at the end, following an initial synopsis consisting of six chapters (the first of which is this introduction) and a bibliography. The rest of the synopsis is structured as follows.

Chapter 2 gives a background on parametrized polynomial systems, which forms the foundation for the rest of the thesis. Chapter 3 focuses on enumerative and numerical algebraic geometry, with emphasis on the tropical homotopies constructed in Paper A. Chapter 4 is devoted to vertically parametrized systems, including the general theory of Paper B, the applications to reaction network theory explored in Paper C, and the results on toricity from Paper D. In Chapter 5, we focus on the method of moments, including the determinantal structure of the moment varieties of several classical distributions indentified in Paper E, and the identifiability results obtained in Paper F. Finally, Chapter 6 is dedicated to 3D genome reconstruction, concentrating on the identifiability results and reconstruction techniques developed in Paper G.

Notation and conventions

For $n \in \mathbb{Z}_{>0}$, we write $[n] := \{1, 2, \ldots, n\}$. For a set S, the cardinality is denoted $|S|$. The componentwise Hadamard product is denoted by $\star\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}^n$. For a vector $t \in (\mathbb{C}^*)^d$ and a matrix $A = (a_{ij}) \in \mathbb{Z}^{d\times n}$, we write t^A for the vector in $(\mathbb{C}^*)^n$ with $(t^A)_j = \prod_{i=1}^d t_i^{a_{ij}}$.

The ideal generated by the polynomials in a tuple $F = (f_1, \ldots, f_r) \in K[x_1^{\pm}, \ldots, x_n^{\pm}]^r$ is denoted by $\langle F\rangle$. The corresponding zero locus in the torus $(K^*)^n$ is denoted by $\mathbb{V}_{K^*}(F)$. In the case when $K = \mathbb{R}$, the set of zeros in the positive orthant $\mathbb{R}^n_{>0}$ is denoted by $\mathbb{V}_{>0}(S)$. The $r \times n$ Jacobian matrix of F is denoted by $J_F = (\partial f_i/\partial x_j)$.

2

Preliminaries on parametrized polynomial systems

In this chapter, we introduce some general terminology and foundational facts about parametrized polynomial systems that will be used throughout the thesis. Particular emphasis is placed on the dimension of the algebraic varieties cut out by generic instances of a parametrized system. We end the chapter by introducing some of the key examples of parametrized systems that will appear in subsequent chapters. The notation and terminology follow in large parts that of Paper B.

2.1 Basic terminology

By a *parametrized polynomial system*, we will mean a tuple of (Laurent) polynomials

$$F = (f_1, \dots, f_r) \in \mathbb{C}[p_1, \dots, p_k, x_1^{\pm}, \dots, x_n^{\pm}]^r$$

with variables $x = (x_1, \dots, x_n)$ and coefficients that are polynomials in parameters $p = (p_1, \dots, p_k)$. We assume that the variables can take values in the *state space* $(\mathbb{C}^*)^n$, and that the parameters take values in the *parameter space* $\mathbb{C}^k$. We will be interested in properties that hold for *generic* choices of parameters, in the sense that they hold in a nonempty Zariski open subset of $\mathbb{C}^k$. We extend this concept of genericity to subsets $S \subseteq \mathbb{C}^k$ by saying that a property holds generically in S if holds in a nonempty open subset of S with respect to the subspace topology inherited from $\mathbb{C}^k$.

We will be interested in how the geometry of the set of zeros $\mathbb{V}_{\mathbb{C}^*}(F_p) \subseteq (\mathbb{C}^*)^n$ for the specialization $F_p = F(p, \cdot) \in \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}]^r$ varies as we specialize the system at different points $p \in \mathbb{C}^k$. Geometrically, this corresponds to describing the fibers of

$$\mathcal{I} := \{(p, x) \in \mathbb{C}^k \times (\mathbb{C}^*)^n : F_p(x) = 0\} \xrightarrow{\pi} \mathbb{C}^k, \quad (p, x) \mapsto p\,,$$

where $\mathcal{I}$ is the *incidence variety* of the system. We denote by $\mathcal{Z} = \pi(\mathcal{I})$ the set of parameters for which the system is compatible. If the Zariski closure $\overline{\mathcal{Z}}$ (which coincides with the Euclidean closure by constructibility of $\mathcal{Z}$) is all of $\mathbb{C}^k$, we say that the system is *generically consistent*. This is equivalent to $\pi \colon \mathcal{I} \to \mathbb{C}^k$ being a dominant morphism.

We also introduce the terminology of nondegeneracy. A zero $x^* \in (\mathbb{C}^*)^n$ of F is said to be *nondegenerate* if the Jacobian $J_F = (\partial F_i/\partial x_j)$ has rank r at x^*. We point out two properties of nondegenerate zeros that we will be important in subsequent chapters.

Lemma 2.1 (Proposition B.2.14, Theorem B.2.15)**.** *Let $F \in \mathbb{C}[p_1, \dots, p_k, x_1^{\pm}, \dots, x_n^{\pm}]^r$, and suppose F_p has a nondegenerate zero for some $p \in \mathbb{C}^k$. Then the following holds:*

(i) $\overline{\mathcal{Z}} = \mathbb{C}^k$.

(ii) If $r = n$, then all zeros of F_p are nondegenerate for generic $p \in \mathbb{C}^k$.

Remark 2.2. The theory in this chapter is formulated for Laurent polynomials, with the complex torus $(\mathbb{C}^*)^n$ as the state space, but all the content of this chapter also holds for polynomials with nonnegative exponents and $\mathbb{C}^n$ as the state space. Note, however, that in Chapters 3 and 4, working in $(\mathbb{C}^*)^n$ will play an important role; in the former case because tropical geometry is defined in the torus, and in the latter case because working in the torus is crucial for our parametrization of the incidence variety.

2.2 Generic dimension

Throughout the thesis, the dimension of $\mathbb{V}_{\mathbb{C}^*}(F_p) \subseteq (\mathbb{C}^*)^n$ for generic $p \in \mathbb{C}^k$ will be of central importance. In principle, this can be determined through a Gröbner basis computation over the field $\mathbb{C}(p_1, \dots, p_k)$ (cf. [40, Chapter 9]), but this is rarely computationally feasible. In practice, it is more fruitful to study the generic dimension through the dimension of $\overline{\mathcal{Z}}$, with the help of the following consequence of the classical theorem of dimension of fibers [117, Theorem 3.13, Corollary 3.15].

Proposition 2.3. *Suppose $\mathcal{I}$ is irreducible. Then for every $p \in \mathbb{C}^k$, it holds that*

$$\dim(Y) \geq \dim(\mathcal{I}) - \dim(\overline{\mathcal{Z}}) \quad \text{for every irreducible component } Y \text{ of } \mathbb{V}_{\mathbb{C}^*}(F_p),$$

with equality for generic $p \in \mathcal{Z}$. In particular, F is generically consistent if and only if $\dim \mathbb{V}_{\mathbb{C}^*}(F_p) = \dim(\mathcal{I}) - k$ for some (and hence generic) $p \in \mathbb{C}^k$.

One way to verify generic consistency is to investigate the Jacobian J_π of π. We make use of the following characterization of generic consistency in Papers E and G.

Proposition 2.4 ([117, Proposition 3.6])**.** *Let $\mathcal{I}$ be irreducible. Then F is generically consistent if and only if $\operatorname{rk}(J_\pi(p, x)) = k$ for some $(p, x) \in \mathcal{I}$.*

2.3 Real and positive solutions

In many applications, we are working with systems with real coefficients, i.e.,

$$F = (f_1, \dots, f_r) \in \mathbb{R}[p_1, \dots, p_k, x_1^{\pm}, \dots, x_n^{\pm}]^r,$$

and we are interested in the real solutions $\mathbb{V}_{\mathbb{R}^*}(F_p) \subseteq (\mathbb{R}^*)$ or even the positive ones $\mathbb{V}_{>0}(F_p) \subseteq \mathbb{R}^n_{>0}$ for $p \in \mathbb{R}^k$. This type of question belongs to the realm of *real and semialgebraic geometry*, which in many ways is qualitatively different from algebraic geometry over the complex numbers; see [20] for a classical overview of the field.

Under mild conditions, the geometry of the complex solutions is representative of the real solutions, in the sense that the latter form a Zariski dense subset. More precisely, we have the following, which forms a foundation for several of the results in Paper B.

Proposition 2.5 ([20, Proposition 3.3.16], [123, Theorem 6.5])**.** *Let $F \in \mathbb{R}[x_1^{\pm}, \dots, x_n^{\pm}]^r$ and $U \subseteq (\mathbb{R}^*)^n$ be a Euclidean open subset (e.g., $U = \mathbb{R}^n_{>0}$). If $\mathbb{V}_{\mathbb{C}^*}(F) \subseteq (\mathbb{R}^*)^n$ is irreducible, and the intersection with U contains a nonsingular real point, then the intersection $\mathbb{V}_{\mathbb{R}^*}(F) \cap U$ is Zariski dense in $\mathbb{V}_{\mathbb{C}^*}(F)$.*

2.4 Classes of parametrized systems

We end the chapter by introducing a couple of special cases of parametrized polynomial systems that will appear throughout the thesis.

Freely parametrized systems

The arguably most well-studied examples of parametric systems are those with fixed finite support sets $\mathcal{A}_1, \dots, \mathcal{A}_r \subseteq \mathbb{Z}^n$ and freely varying coefficients, which we call *freely parametrized systems* in Paper B. More precisely, the freely parametrized system associated with the supports $(\mathcal{A}_1, \dots, \mathcal{A}_r)$ is given by

$$F = \Big(\sum_{\alpha \in \mathcal{A}_i} a_{i,\alpha} x^\alpha \Big)_{i \in [r]} \in \mathbb{C}\Big[(a_{i,\alpha} : i \in [r], \alpha \in \mathcal{A}_i), x_1^{\pm}, \dots, x_n^{\pm}\Big],$$

where the coefficients $a_{i,\alpha}$ are seen as parameters.

A common theme in the theory of freely parametrized systems is that the generic geometry of the solution sets is determined by the *Newton polytopes* of the polynomials. For instance, the celebrated BKK theorem (named after Bernstein, Kouchnirenko and Khovanskii's pioneering work in the 1970's) tells us that the generic number of zeros of a square freely parametrized system is the mixed volume of the Newton polytopes. More recent work also studies the monodromy group [54], and characterizes generic irreducibility of the zero locus [89, 148] in terms of the Newton polytopes.

Systems that arise in applications often have additional structure, which gives rise to dependencies among the coefficients (or even some coefficients being fixed), such that the generic behavior might differ from the generic behavior predicted by the Newton polytopes alone. Hence, it is desirable to generalize the "BKK toolkit" (to borrow a

term from [55]) of results on the generic geometry of freely parametrized systems to non-freely parametrized cases, in a way that accounts for the additional structure.

In Paper A, we do this for the problem of numerically solving generic instances of non-freely parametrized systems, by generalizing the polyhedral homotopies of Huber and Sturmfels [80]. In Papers B and D, we study the properties of generic consistency and generic toricity in the context of vertically parametrized systems (introduced at the end of this section).

Systems with parametric constant terms

An extreme case of parametrized systems with very few freely varying coefficients are those where each polynomial has an independently varying parametric constant term, but all other coefficients are fixed. More precisely, they are of the form

$$F = (g_1 - b_1, \dots, g_r - b_r) \in \mathbb{C}[b_1, \dots, b_r, x_1^{\pm}, \dots, x_n^{\pm}]^r,$$

where $g_1, \dots, g_r \in \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}]$. In this case, the incidence variety is isomorphic to the state space $(\mathbb{C}^*)^n$, and the solution sets correspond to fibers of the map

$$(\mathbb{C}^*)^n \to \mathbb{C}^r, \quad x \mapsto (g_1(x), \dots, g_r(x)).$$

This construction can also be extended to allow for rational functions $g_i = p_i/q_i$ with $p_i, q_i \in \mathbb{C}[x_1, \dots, x_n]$, which give rise to a rational map

$$\mathbb{C}^n \dashrightarrow \mathbb{C}^r, \quad x \mapsto (g_1(x), \dots, g_r(x))$$

defined on the distinguished open subset $\{x \in \mathbb{C}^n : q_1(x) \cdots q_r(x) \neq 0\}$. Systems of this form appear in Papers E and F, where the g_i are moments of a statistical distribution depending on unknown distribution parameters, and in Paper G, where the g_i are contact counts depending on unknown coordinates of DNA segments along a chromosome.

Horizontally parametrized systems

A generalization of freely parametrized systems that has attracted particular attention the past decade is what we call *horizontally parametrized systems*, where each polynomial is a general linear combination of some fixed sets $\mathcal{S}_1, \dots, \mathcal{S}_r \subseteq \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}]$ of support polynomials (that are not necessarily monomial), i.e., systems of the form

$$F = \Big(\sum_{g \in \mathcal{S}_i} a_{i,g} g : i \in [r] \Big) \in \mathbb{C}\Big[(a_{i,g} : i \in [r], g \in \mathcal{S}_i), x_1^{\pm}, \dots, x_n^{\pm}\Big]^r.$$

The terminology "horizontal" comes from [76], and refers to the fact that each parameter appears in a single row of the coefficient matrix with respect to the set of

monomials appearing in the system. Note that a horizontally parametrized family is freely parametrized system precisely when all the sets $\mathcal{S}_1, \dots, \mathcal{S}_r$ consist of monomials.

In their seminal works [87, 88], Kaveh and Khovanskii give a framework for studying such systems through the theory of Newton–Okounkov bodies, which has recently been explored in, e.g., [23, 119] in the study of coupled oscillators.

Vertically parametrized systems

Another generalization of freely parametrized systems, that will play a particularly central role in this thesis, are those that we call *vertically parametrized systems*, where we allow linear dependencies among coefficients appearing in front of the same monomial in different polynomials. Concretely, such systems can be written as

$$F = C(a \star x^M) \in \mathbb{C}[a_1, \dots, a_m, x_1^{\pm}, \dots, x_n^{\pm}]^s,$$

for a matrix $M \in \mathbb{Z}^{n\times m}$ whose columns are the exponent vectors of the monomials appearing in the system, a vector of parameters $a = (a_1, \dots, a_m)$ that scale the monomials, and a matrix $C \in \mathbb{C}^{s\times m}$ whose rows encode linear combinations of the scaled monomials. Analogously to the horizontally parametrized case, the term "vertical" was introduced in [76], and refers to the fact that each parameter appears in a single column of the coefficient matrix with respect to the set of monomials appearing in the system.

Remark 2.6. A freely parametrized system with support sets $(\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_s)$ in $\mathbb{Z}^n$ can be expressed as a vertically parametrized system by setting

$$C = \begin{bmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_s \end{bmatrix} \quad \text{with} \quad C_i = [1 \ \cdots \ 1] \in \mathbb{C}^{1\times|\mathcal{A}_i|}, \quad \text{and} \quad M = [M_1 \ \cdots \ M_s], \tag{2.1}$$

where the columns $M_i \in \mathbb{Z}^{n\times|\mathcal{A}_i|}$ are taken to be the elements of $\mathcal{A}_i$.

Remark 2.7. An first observation to make is that the generic behavior of vertically parametrized systems cannot be explained only in terms of the Newton polytopes, since acting on C from the left by $\mathrm{GL}_s(\mathbb{C})$ can give cancellations that change the supports. Instead, we will see in Chapters 3 and 4 that the behavior is dictated by an interplay between the exponent matrix M and the (column matroid of the) coefficient matrix C.

We furthermore define *augmented vertically parametrized systems* to be a concatenation of a vertically parametrized system and a linear system with parametric constant terms of the form

$$F = \Big(C(a \star x^M),\ Lx - b\Big) \in \mathbb{C}[a_1, \dots, a_m, b_1, \dots, b_\ell, x_1^{\pm}, \dots, x_n^{\pm}]^{s+\ell},$$

for $C \in \mathbb{C}^{s\times m}$, $M \in \mathbb{Z}^{n\times m}$ scaling parameters $a = (a_1, \dots, a_m)$, and an additional coefficient matrix $L \in \mathbb{C}^{\ell\times n}$ and parametric constant terms $b = (b_1, \dots, b_\ell)$ for $\ell \geq 0$.

Augmented vertically parametrized systems naturally appear as the steady state systems for chemical reaction networks, as explained in Section 4.1. They also appear in optimization as the critical points systems for freely parametrized polynomials, as discussed in the introduction of Paper B.

3 Counting and approximating roots with tropical geometry

This chapter is an overview of the results of Paper A. We begin this chapter by introducing the key ideas of homotopy continuation, and the role that generic root counts play in this context, and then go on to explain the techniques developed in Paper A for constructing homotopies based on tropical intersection data.

3.1 Homotopy continuation

The basic idea of homotopy continuation is to turn the problem of solving a polynomial system into a problem of solving a system of ordinary differential equations, which is a standard problem in numerical analysis. The general strategy for doing this for a given square *target system* $F \in \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}]^n$ can be said to consist of the following three main steps, which are schematically illustrated in Figure 3.1:

1. Construct a *start system* $G \in \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}]^n$ with at least as many isolated zeros as F, which are easy to approximate.
2. Construct a *homotopy* $H\colon [0,1] \times (\mathbb{C}^*)^n \to (\mathbb{C}^*)^n$ with $H(0,\cdot) = G$ and $H(1,\cdot) = F$.
3. For each $\zeta \in \mathbb{V}_{\mathbb{C}^*}(G)$, numerically solve the initial problem

$$\frac{\partial H}{\partial x}\frac{dx}{dt} + \frac{\partial H}{\partial t} = 0, \quad x(0) = \zeta\,.$$

   If the corresponding trajectory $x(t)$ converges to a point in $(\mathbb{C}^*)^n$ as $t \to 1$, the limit is a zero of F. By making appropriate choices of G and H, we can expect to find all isolated points of $\mathbb{V}_{\mathbb{C}^*}(F)$ in this way.

The ideas behind this go back to works such as [45, 59] in the second half of the 20th century, and have since been implemented in software such as `PHCpack` [143], `Bertini` [12], and `HomotopyContinuation.jl` [27].

There are many interesting theoretical and practical aspects of all three of the points mentioned above, which are treated in the field of numerical algebraic geometry. For a

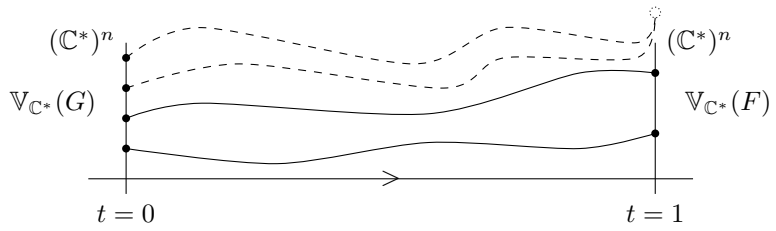


Figure 3.1: Schematic illustration of homotopy continuation. In this example, the start system has 4 zeros, whereas the target system only has 2, and we obtain two diverging superfluous paths (dashed).

comprehensive overview, we refer to the textbook [136] and the review [11]. In what follows, our focus will lie on construction of start systems and associated homotopies that lead to as few as possible superfluous paths to trace. A central challenge with this is that it relies on having an *a priori* upper bound on $|\mathbb{V}_{\mathbb{C}^*}(F)|$, and the problem is thus intimately related to enumerative geometry.

3.2 Generic root counts

Given a target system $F^* \in \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}]^n$, a common approach to obtain upper bounds on $|\mathbb{V}_{\mathbb{C}^*}(F^*)|$ is to view it as an instance $F^* = F_p$ of a parametrized system $F \in [p_1, \dots, p_k, x_1^{\pm}, \dots, x_n^{\pm}]^n$. By the following classical result, F has a generic root count, which is an upper bound on the number of isolated zeros of F^*.

Theorem 3.1 (Parameter continuation theorem)**.** *Let $F \in [p_1, \dots, p_k, x_1^{\pm}, \dots, x_n^{\pm}]^n$. Then there exists a* generic root count *$N \in \mathbb{Z}_{\geq 0}$ and a* discriminant variety *$\Delta \subsetneq \mathbb{C}^k$ such that F_p has at most N nondegenerate zeros for all $p \in \mathbb{C}^k$, with equality for $p \in \mathbb{C}^k \setminus \Delta$.*

See [24] for a recent proof based on Gröbner bases in the affine setting. For parametrized system with generically finitely many and generically nondegenerate zeros (e.g., square vertically parametrized systems by Theorem 4.4), this coincides with the definition of generic root count that we use in Paper A (see Definition A.2.4).

Given an embedding of F^* into a parametrized system F with $F^* = F_p$, the next step is to identify a $q \in \mathbb{C}^k \setminus \Delta$ such that F_q is easy to solve, and then construct a sufficiently generic path $\gamma\colon [0,1] \to \mathbb{C}^k$ from q to p that avoids Δ (which is possible, since Δ has at least real codimension 2 in $\mathbb{C}^k$). We can then choose $H(t,x) = F_{\gamma(t)}(x)$. See [136, Section 7.1] for details.

Example 3.2. If all exponents of F^* are nonnegative and the ith entry has total degree d_i, we can let F be the freely parametrized system where the ith entry includes all monomials of total degree at most d_i. Then the generic root count of F is the product $d_1 \cdots d_n$ by Bézout's theorem, and we can pick $G = (x_1^{d_1} - 1, \dots, x_n^{d_n} - 1)$ as the start system, and take H to be the straight-line homotopy $H(t,x) = (1-t)\,\gamma\, G + t\, F^*$ for a random $\gamma \in \mathbb{C}$ (see the discussion in [136, Section 8.4.1]).

The default approach in many software packages for homotopy continuation is to embed F^* into the corresponding freely parametrized system. By the renowned BKK theorem of Bernstein [17], Khovanskii [90] and Kouchnirenko [92], the generic root count of a freely parametrized system is the mixed volume of the Newton polytopes. In the 1995 paper [80], Huber and Sturmfels proposed an algorithm that for a given target system constructs a collection of homotopies and associated binomial start systems that together give rise to the mixed volume number of paths, and thus can be expected to find all zeros of the target system.

Parametrized systems that arise in practice often have dependencies between the coefficients that make them correspond to a subset of the discriminant variety of the associated freely parametrized system, meaning that the mixed volume bound might not be sharp. Examples of scenarios where this happens include reaction network theory [69, 119], the theory of coupled oscillators [25], and rigidity theory [30]. On the other hand, many systems have coefficients generic enough to fall outside the BKK discriminant. This can be detected with "Bernstein's second theorem" [17, Theorem B], which characterizes when the mixed volume bound is attained in terms of facial subsystems, which has successfully been employed for various Lagrangian systems in optimization [26, 103].

For systems for which the mixed volume bound is not sharp, other combinatorial data can be used to obtain sharper root bounds and associated homotopies. A tool that has attracted much interest recently is Khovanskii bases and the related Newton–Okounkov bodies, which are based on work by Kaveh and Khovanskii [88], and recently have been employed for, e.g., coupled oscillators [25] and reaction networks [119], and which can be used to construct homotopies as shown in [29].

Another type of combinatorial data that carries intersection-theoretic information comes from *tropical geometry*, which has been explored as a way to obtain sharper root counts in, e.g., [76, 77, 100]. Some first steps towards constructing homotopies form such data are taken by Leykin and Yu in [100]. In Paper A, we extend this method to a general framework for constructing homotopies based on tropical intersection data. This will be the topic for the remainder of this chapter.

3.3 Preliminaries from tropical geometry

Tropical geometry is a central topic in contemporary combinatorial algebraic geometry, and allows studying properties of a classical variety through the combinatorial structure of its tropicalization, which is a polyhedral complex. Here, we give a brief introduction to some of the most central constructions that we will need in the rest of this chapter. For a more comprehensive overview of the field, we refer to the textbook [106].

Let $K = \mathbb{C}\{\{t\}\} = \bigcup_{n>0} \mathbb{C}((t^{1/N}))$ be the field of complex Puiseux series, with the valuation map

$$\nu\colon \mathbb{C}\{\{t\}\}^* \to \mathbb{R}, \qquad \sum_{i=i_0}^{\infty} c_i t^{i/N} \mapsto \min\{i/N : c_{i/N} \neq 0\}.$$

The *tropicalization* $\operatorname{trop}(f)$ of a polynomial $f = \sum_{\alpha\in S} c_\alpha x^\alpha \in K[x_1^{\pm}, \dots, x_n^{\pm}]$ with support $S \subseteq \mathbb{Z}^n$ and nonzero Puiseux series coefficients is a function given by

$$\operatorname{trop}(f)(x) = \min_{\alpha\in S} \Big(\nu(c_\alpha) + \alpha \cdot x\Big),$$

where $\cdot$ denotes the standard inner product. The *initial form* of f with respect to a weight $w \in \mathbb{R}^n$ is given by

$$\operatorname{in}_w(f) := \sum_{\substack{\alpha\in S \text{ such that} \\ \nu(c_\alpha)+\alpha\cdot w = \operatorname{trop}(f)(w)}} t^{-\nu(c_\alpha)} c_\alpha\, x^\alpha \in \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}].$$

For an ideal $I \subseteq K[x_1^{\pm}, \dots, x_n^{\pm}]$, the *initial ideal* with respect to a weight $w \in \mathbb{R}^n$ is defined as

$$\operatorname{in}_w(I) = \langle \operatorname{in}_w(f) : f \in I\rangle \subseteq \mathbb{C}[x_1^{\pm}, \dots, x_n^{\pm}].$$

As a set, the *tropicalization* of an ideal I is given by

$$\operatorname{Trop}(I) = \{w \in \mathbb{R}^n : \operatorname{in}_w(I) \text{ contains no monomials}\}.$$

More precisely, we can define $\operatorname{Trop}(I)$ as a subcomplex of the Gröbner complex of the homogenization of I, which turns $\operatorname{Trop}(I)$ into a weighted polyhedral complex [106, Chapters 2–3]. Yet another way to think of $\operatorname{Trop}(I)$ is given by the fundamental theorem of tropical geometry [106, Theorem 3.2.3], which says that it is equal to $\overline{\nu(\mathbb{V}_{K^*}(I))}$, where ν is applied componentwise, and the line denotes the Euclidean closure in $\mathbb{R}^n$.

The *stable intersection* of two tropicalizations Σ_1 and Σ_2 in $\mathbb{R}^n$ is defined as

$$\Sigma_1 \wedge \Sigma_2 = \{\sigma_1 \cap \sigma_2 : \sigma_1 \in \Sigma_1, \sigma_2 \in \Sigma_2, \dim(\sigma_1 + \sigma_2) = n\},$$

where $\Sigma_1 \wedge \Sigma_2$ inherits multiplicities from Σ_1 and Σ_2 as detailed in Definition A.2.12.

A key result for computing tropicalizations is the following version of the transverse intersection lemma of Bogart, Jensen, Speyer, Sturmfels and Thomas [21].

Theorem 3.3 (Theorem A.2.14)**.** *Let $I, J \subseteq K[x_1^{\pm}, \dots, x_n^{\pm}]$ be complete intersections, such that $\operatorname{Trop}(I)$ and $\operatorname{Trop}(J)$ intersect transversally. Then the following holds:*

(i) $\operatorname{Trop}(I + J) = \operatorname{Trop}(I) \wedge \operatorname{Trop}(J)$.

(ii) $\operatorname{in}_w(I + J) = \operatorname{in}_w(I) + \operatorname{in}_w(J)$ *for all* $w \in \operatorname{Trop}(I + J)$.

3.4 Homotopies via tropical geometry

Let $F = (f_1, \dots, f_n) \in \mathbb{C}[p_1, \dots, p_k, x_1^{\pm}, \dots, x_n^{\pm}]^n$ be a parametrized system with generically zero-dimensional, nondegenerate zero locus, and let $p \in \mathbb{C}^k$ be a generic choice of parameters. The main result of Paper A is Algorithm A.3.1 for obtaining the root count of F_p and an associated collection of homotopies for approximating $\mathbb{V}_{\mathbb{C}^*}(F_p)$.

The idea of the algorithm is as follows: Form a "perturbation" $q = t^u \star p \in \mathbb{C}\{\{t\}\}^k$ of the parameter values for some generic $u \in \mathbb{Q}^k$. By the Newton–Puiseux theorem, as $t \to 0$, the zeros of F_q can be approximated by Puiseux series of the form

$$x(t) = \begin{pmatrix} c_1 t^{w_1} + \text{higher-order terms} \\ \vdots \\ c_n t^{w_n} + \text{higher-order terms} \end{pmatrix} \text{ for some } c \in (\mathbb{C}^*)^n \text{ and } w \in \mathbb{Q}^n.$$

For such a Puiseux series to correspond to an actual zero, the terms with minimal exponents in $F_q(x(t))$ need to cancel as $t \to 0$, which happens precisely when

$$w \in \operatorname{Trop}(\langle F_q \rangle) \quad \text{and} \quad c \in \mathbb{V}_{\mathbb{C}^*}(\operatorname{in}_w(\langle F_q \rangle)) \,. \tag{3.1}$$

In fact, it turns out that for each such choice of w and c, we get a path that can be used to approximate a zero of F_p. More precisely, we get a homotopy

$$H^{(w)}(t, x) = \left(t^{-\operatorname{trop}(f_{i,q})(w)} f_{i,q}(t^w \star x) : i \in [n] \right)$$

that (up to certain "early game" issues discussed in Remark A.3.3) can be used trace the zeros of $\operatorname{in}_w(\langle F_q \rangle)$ from $t = 0$ to $t = 1$.

Remark 3.4. Computing the tropical intersection data of (3.1) is very challenging in general, since the known general algorithms rely on heavy Gröbner computations [65]. To make use of Algorithm A.3.1 for concrete systems, our strategy is therefore to decompose the system (possibly after re-embedding it, to include more polynomials and variables) into systems that are easy to tropicalize (e.g., binomial, linear or reciprocal linear varieties), and whose tropicalizations generically intersect transversally, so that Theorem 3.3 can be used.

Example 3.5. We illustrate the strategy described in the previous remark for a vertically parametrized system $C(a \star x^M)$. This system can be re-embedded and decomposed into a binomial part $y - a \star x^M$ and a linear part Cy, where $y = (y_1, \dots, y_m)$ are new variables. The tropicalization $\operatorname{Trop}(\langle Cy \rangle)$ is a tropical linear space, given by the Bergman fan of the matroid associated to $\operatorname{row}(C)$. The tropicalization $\operatorname{Trop}(\langle y - a \star x^M \rangle)$ is the kernel of $[-M^\top \mid \mathrm{id}_m]$ shifted by $(0, \nu(a))$. For $u \in \mathbb{Q}^m$ sufficiently general, $\operatorname{Trop}(\langle y - t^u \star a \star x^M \rangle)$ and $\operatorname{Trop}(\langle Cy \rangle)$ intersect transversally, so Theorem 3.3 is applicable.

In Paper A, we discuss various practical considerations for applying Algorithm A.3.1 to vertically and horizontally parametrized systems. We also illustrate through a series of case studies how the method can be used for systems appearing in reaction network theory, the theory of coupled oscillators, and for systems from rigidity theory.

3.5 Open problems

An interesting direction for future work is to better understand and improve the computational complexity of the tropical computations needed for (3.1).

The vertically parametrized scenario of Example 3.5 is a reasonable first objective, since it is the most straight-forward case. Our experiments indicate that with our current implementation, computing the data (3.1) is feasible for vertically parametrized systems up to approximately 15 variables, with both the tropicalization of the linear space and the stable intersection acting as bottle necks.

In practice, it is often the case that only a small part of the tropical linear space contributes to the stable intersection. It is therefore of interest to develop new methods for computing the relevant parts of stable intersections without explicitly computing the full spaces being intersected. A particularly interesting directions to explore in future work is *tropical homotopy continuation* [44, 82], as well as methods for locally computing *only the relevant parts* of the tropical linear space.

Another intersecting direction is to systematically exploit the fact that vertically parametrized systems that appear in the chemical application often have a toric structure, as discussed in Section 4.4. Since a toric variety can be tropicalized directly, this allows circumventing the re-embedding and to calculate the stable intersection in an ambient space of lower dimension, which has the potential to speed up computations.

4

Vertically parametrized systems and reaction networks

This chapter focuses on the concept of (augmented) *vertically parametrized systems*. We begin in Section 4.1 by recalling how they arise in reaction network theory, and then proceed to explain the results on consistency and nondegeneracy from Paper B in Section 4.2, and discuss the implications of these results for reaction network theory developed in Paper C in Section 4.3. Finally, we discuss results on toricity of vertically parametrized systems developed in Paper D in Section 4.4.

Recall from Section 2.4 that an *augmented vertically parametrized system* is one that can be written as

$$F = \Big(C(a \star x^M),\ Lx - b \Big) \in \mathbb{C}[a_1, \dots, a_m, b_1, \dots, b_\ell, x_1^{\pm}, \dots, x_n^{\pm}]^{s+\ell},$$

for a coefficient matrix $C \in \mathbb{C}^{s \times m}$, a matrix $M \in \mathbb{Z}^{n \times m}$ whose columns encode m monomials that are scaled by the parameters $a = (a_1, \dots, a_m)$, as well as a coefficient matrix $L \in \mathbb{C}^{\ell \times n}$ encoding $\ell \geq 0$ affine forms with parametric constant terms $b = (b_1, \dots, b_\ell)$. If $\ell = 0$, we simply call the system *vertically parametrized*.

4.1 Connection to reaction network theory

Here, we provide a quick overview of the basic terminology of reaction network theory, and the connection to vertically parametrized systems; for a more comprehensive introduction to the field, we refer the reader to [50, 52, 58].

By a *reaction network*, we mean a collection of some number m of *reactions* of form

$$\alpha_{1j} X_1 + \cdots + \alpha_{nj} X_n \longrightarrow \beta_{1j} X_1 + \cdots + \beta_{nj} X_n, \quad j \in [m] \tag{4.1}$$

that describe interactions between some number n of *species* $X_1, \dots, X_n$, which represent molecules, cells, or other interacting entities. An important goal of reaction network theory is to study how the concentration x_i of each species X_i varies over time, under various models for the dynamics.

A common choice of dynamical model is *power law kinetics*, where one assumes that the jth reaction takes place at a rate proportional to a monomial $x_1^{M_{1j}} \cdots x_n^{M_{nj}}$, with a proportionality constant $a_j > 0$ called *rate constant* that depends on the environment. Under this assumption, the concentrations $x = (x_1, \ldots, x_n)$ evolve according to the autonomous system

$$\dot{x}(t) = C(a \star x(t)^M), \tag{4.2}$$

where $C = (\beta_{ij} - \alpha_{ij}) \in \mathbb{R}^{n \times m}$ is the *stoichiometric matrix* and $M = (M_{ij}) \in \mathbb{Z}^{n \times m}$ is the *kinetic matrix* of the network. A particularly well-studied case is *mass-action kinetics*, which corresponds to $M = (\alpha_{ij})$. This model for the dynamics of a network goes back to work by Guldberg and Waage in the 1870's [70], and has since then become one of the standard choices for modeling the dynamics of reaction networks [144].

Reaction networks with power law kinetics can display very rich dynamics, including Hopf bifurcations, oscillations and chaotic behavior; see [72] for a recent discussion on the dynamics networks can exhibit. Here, we focus only on the *positive steady states*, which are the positive zeros $x \in \mathbb{R}^n_{>0}$ of the vertically parametrized system $C(a \star x^M)$.

In practice, the rows of C often have linear dependencies, in which case the trajectory of (4.2) for a given initial value $x(0) = x_0 \in \mathbb{R}^n_{>0}$ is confined to the affine linear space $x_0 + \mathrm{im}(C)$. This is often encoded in the following way: Let $L \in \mathbb{R}^{\ell \times n}$ be a matrix whose rows form a basis for $\ker(C^\top)$. The entries of Lx are called *conserved quantities* of the network, and the entries of Lx_0 are called the associated *total amounts* for the initial state x_0. For a choice $b \in L(\mathbb{R}^n_{>0})$ of total amounts, the affine linear space where trajectories are confined is given by $\mathbb{V}(Lx - b)$, and the positive steady states in this space are the positive zeros of the augmented vertically parametrized system

$$\big(C(a \star x^M),\ Lx - b\big) \tag{4.3}$$

where the variables are the concentrations x, and the parameters are the rate constants a and the total amounts b; see Figure 4.1 for an illustration. By removing linearly dependent rows of C to obtain a matrix of full rank, (4.3) becomes a square system.

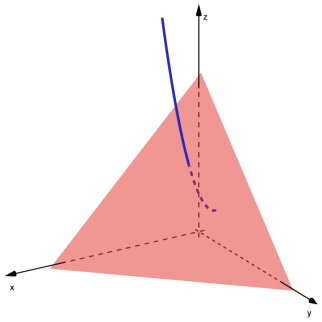


Figure 4.1: Schematic illustration of $\mathbb{V}_{>0}(C(a \star x^M))$ (blue) and $\mathbb{V}_{>0}(Lx - b)$ (red) for a particular choice of parameters $(a, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell$.

Example 4.1. The following network from [133] appears in Papers C and D, and models the IDHKP-IDH system in bacterial metabolism:

$$X_1 + X_2 \underset{a_2}{\overset{a_1}{\rightleftharpoons}} X_3 \xrightarrow{a_3} X_1 + X_4$$
$$X_3 + X_4 \underset{a_5}{\overset{a_4}{\rightleftharpoons}} X_5 \xrightarrow{a_6} X_2 + X_3$$

Here, X_2 is the enzyme isocitrate dehydrogenase, which is turned into a phosphorylated form X_4 by an enzyme X_1, and converted back to its original form by an enzyme X_3. The steady states are the zeros of the augmented vertically parametrized system

$$F = \begin{pmatrix} -a_1x_1x_2 + a_2x_3 + a_3x_3 \\ -a_1x_1x_2 + a_2x_3 + a_6x_5 \\ a_4x_3x_4 - a_5x_5 - a_6x_5 \\ -2x_1 + x_2 - x_3 + x_4 - b_1 \\ x_1 + x_3 + x_5 - b_2 \end{pmatrix} \in \mathbb{R}[a_1, \dots, a_6, b_1, b_2, x_1^{\pm}, \dots x_5^{\pm}]^3, \tag{4.4}$$

defined by the following matrices (row 3 and 4 of C are omitted by linear dependence):

$$C = \begin{bmatrix} -1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad L = \begin{bmatrix} -2 & 1 & -1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

We end the section with a few words about realizability questions. First of all, we note that a vertically parametrized system $C(a \star x^M)$ with $s = n$ comes from a network with mass-action kinetics if and only if both M and $M + C$ lies in $\mathbb{Z}_{\geq 0}^{n \times m}$; to construct the network, we just take $\alpha_{ij} = M_{ij}$ and $\beta_{ij} = (M + C)_{ij}$ in (4.1).

Furthermore, a specific system $Dx^E \in \mathbb{R}[x_1, \dots, x_n]^n$ with $D = (d_{ij}) \in \mathbb{R}^{n \times m}$ and $E = (e_{ij}) \in \mathbb{Z}_{\geq 0}^{n \times m}$ is the steady system from a network with mass-action kinetics for some choice of rate constants if and only if $e_{ij} > 0$ whenever $d_{ij} < 0$. This is the content of the famous realizability theorem of Hárs and Tóth [81]. The proof is constructive and constructs a network with reactions

$$\sum_{k=1}^{n} e_{kj} X_k \xrightarrow{|d_{ij}|} \sum_{k=1}^{n} (e_{kj} + \delta_{ik} \operatorname{sign}(d_{ij})) X_k \quad \text{for each } (i, j) \in [n] \times [m] \text{ with } d_{ij} \neq 0,$$

where $|d_{ij}|$ is the rate constant for the reaction, and δ_{ik} denotes the Kronecker delta.

An immediate consequence of this is the following folklore universality theorem, which says that the positive part of any affine variety is the set of positive steady states of some mass-action network (see [50, Section 2] for a discussion of the square case).

Theorem 4.2. *For any $n, s \in \mathbb{Z}_{>0}$ and any $G = (g_1, \dots, g_s) \in \mathbb{R}[x_1, \dots, x_n]^s$, one can find a steady state system $C(a \star x^M)$ and a choice of positive rate constants a such that*

$$\mathbb{V}_{>0}(G) = \mathbb{V}_{>0}(C(a \star x^M)).$$

Proof. In the square case $s = n$, we apply the Hárs–Tóth condition to $x \star G$, noting that $\mathbb{V}_{>0}(x \star G) = \mathbb{V}_{>0}(G)$. If $s < n$, we add $n - s$ zeros to the tuple G and apply the square case. If $s > n$, we apply the square case to $(g_1, \dots, g_{n-1}, g_n^2 + g_{n+1}^2 + \cdots + g_s^2)$, noting that $\mathbb{V}_{>0}(g_1, \dots, g_{n-1}, g_n^2 + g_{n+1}^2 + \cdots + g_s^2) = \mathbb{V}_{>0}(G)$. □

An important take-away of Theorem 4.2 is that it is very challenging to understand the geometry of the set of positive steady states for *all* networks and *all* choices of rate constants. However, as we will see in the following sections, if we focus on properties of networks that hold for *generic* parameter values, many geometric properties becomes remarkably well-behaved. In particular, the behavior of the positive real solution sets often qualitatively agrees with the behavior over the nonzero complex numbers, which makes the it possible to capture it with classical algebraic geometry.

4.2 Consistency, nondegeneracy and dimension

The starting point of the analysis in Paper B is the fact that the incidence variety

$$\mathcal{I} = \{(a, b, x) \in \mathbb{C}^m \times \mathbb{C}^\ell \times (\mathbb{C}^*)^n : C(a \star x^M) = Lx - b = 0\}$$

admits a parametrization

$$\phi\colon \ker(C) \times (\mathbb{C}^*)^n \to \mathcal{I}, \quad (w, h) \mapsto (w \star h^M, Lh^{-1}, h^{-1}), \tag{4.5}$$

which also restricts to a parametrization of the positive incidence variety

$$(\ker(C) \cap \mathbb{R}^m_{>0}) \times \mathbb{R}^n_{>0} \to \mathcal{I} \cap (\mathbb{R}^m_{>0} \times \mathbb{R}^\ell \times \mathbb{R}^n_{>0}),\ (w, h) \mapsto (w \star h^M, Lh^{-1}, h^{-1}). \tag{4.6}$$

The parametrization (4.6) has previously appeared in the form of convex parameters [35], and plays a central role in, e.g., the study of regions of multistationarity [37]. By compositing with the projection $\pi\colon \mathcal{I} \to \mathbb{C}^k$, this also gives a parametrization of the set

$$\mathcal{Z} = \Big\{(k, b) \in \mathbb{C}^m \times \mathbb{C}^\ell : \mathbb{V}_{\mathbb{C}^*}\Big(C(a \star x^M), Lx - b\Big) \neq \varnothing\Big\}$$

and the positive analog

$$\mathcal{Z}_{>0} = \Big\{(k, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell : \mathbb{V}_{>0}\Big(C(a \star x^M), Lx - b\Big) \neq \varnothing\Big\}.$$

Using the parametrization ϕ, we prove that $\mathcal{I}$ has the following important properties.

Theorem 4.3 (Theorem B.3.1)**.** *For an augmented vertically parametrized system, the incidence variety $\mathcal{I}$ is a nonsingular irreducible variety of dimension $m + n - \operatorname{rk}(C)$.*

Through a combination of Sard's lemma and Proposition 2.3, we obtain a characterization of generic consistency and generic nonsingularity of the zeros of the system, in terms of the rank condition

$$\operatorname{rk}\begin{bmatrix} C\operatorname{diag}(w)M^\top \operatorname{diag}(h) \\ L \end{bmatrix} = \operatorname{rk}(C) + \ell \quad \text{for some } (w,h) \in \ker(C) \times (\mathbb{C}^*)^n. \tag{4.7}$$

Theorem 4.4 (Theorem B.3.7)**.** *Let $F = (\,C(a \star x^M),\, Lx - b\,)$ be an augmented vertically parametrized system. Then one of the following scenarios holds:*

(i) ***Generic consistency:*** *If (4.7) is satisfied, then $\overline{\mathcal{Z}} = \mathbb{C}^m \times \mathbb{C}^\ell$, and for generic $(a,b) \in \mathcal{Z}$, all zeros of $F_{a,b}$ are nondegenerate, and $\mathbb{V}_{\mathbb{C}^*}(F_{a,b})$ is of pure dimension $n - (\operatorname{rk}(C) + \ell)$.*

(ii) ***Generic inconsistency:*** *If (4.7) is not satisfied, then $\overline{\mathcal{Z}} \subsetneq \mathbb{C}^m \times \mathbb{C}^\ell$, and for all $(a,b) \in \mathcal{Z}$, all zeros of $F_{a,b}$ are degenerate, and all irreducible components of $\mathbb{V}_{\mathbb{C}^*}(F_{a,b})$ have dimension strictly greater than $n - (\operatorname{rk}(C) + \ell)$.*

Furthermore, the ideal $\langle F_{a,b}\rangle \subseteq \mathbb{C}[x_1^\pm, \dots, x_n^\pm]$ is radical for generic $(a,b) \in \mathbb{C}^m \times \mathbb{C}^\ell$.

This generalizes (and provides a new, elementary proof of) the previously known condition for a freely parametrized system with support sets $(\mathcal{A}_1, \dots, \mathcal{A}_s)$ to be generically consistent (see [148, Lemma 1] and [89, Theorem 11]):

$$\text{There exists a linearly independent tuple } (u_1, \dots, u_s) \in \textstyle\prod_{i=1}^s \operatorname{Lin}(\mathcal{A}_i)\,, \tag{4.8}$$

where $\operatorname{Lin}(\mathcal{A}_i) \subseteq \mathbb{R}^n$ is the direction of the affine hull of $\mathcal{A}_i$. In the language of [15, 137], this corresponds to $(\mathcal{A}_1, \dots, \mathcal{A}_s)$ being an *essential family*.

Motivated by the applications to reaction network theory, we also provide a *positive real version* of Theorem 4.4. In the reaction-network-theoretic setting, this shows that for generic rate constants, there are finitely many positive steady states reachable from a generic initial condition, which answers an open question from [22, Section 5].

Theorem 4.5 (Theorem B.3.7)**.** *Let $F = (\,C(a\star x^M),\, Lx-b\,)$ be an augmented vertically parametrized system with $C \in \mathbb{R}^{s\times m}$ and $L \in \mathbb{R}^{\ell\times n}$. Suppose $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Then one of the following two scenarios hold:*

(i) ***Consistency in a Euclidean open set:*** *If condition (4.7) is satisfied, then $\mathcal{Z}_{>0}$ has nonempty Euclidean interior, and for generic $(a,b) \in \mathcal{Z}_{>0}$, all zeros are nondegenerate with $\mathbb{V}_{>0}(F_{a,b})$ being a smooth manifold of dimension $n-(\operatorname{rk}(C)+\ell)$.*

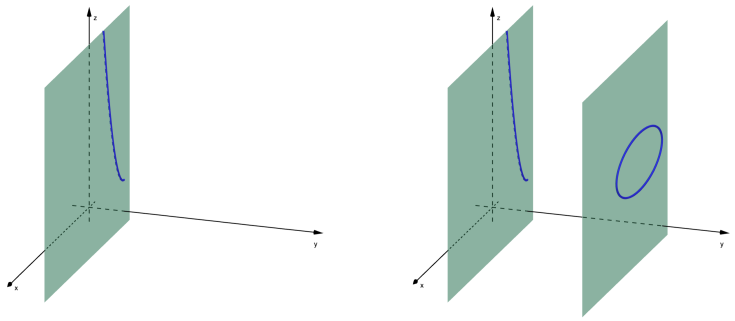
(ii) ***Generic inconsistency:*** *If condition (4.7) is not satisfied, then $\mathcal{Z}_{>0}$ is nonempty but has empty Euclidean interior, and for all $(a,b) \in \mathcal{Z}_{>0}$, all zeros of $F_{a,b}$ are degenerate.*

Example 4.6. For the IDHKP-IDH network in Example 4.1, we have that the vector $w = (2, 1, 1, 2, 1, 1)$ lies in $\ker(C) \cap \mathbb{R}^6_{>0}$ and satisfies the rank condition (4.7) with $h = (1, \dots, 1)$. Hence, there exists a nonempty Euclidean open subset of pairs $(a, b) \in \mathbb{R}^6_{>0} \times \mathbb{R}^2$ for which $\mathbb{V}_{>0}(C(a \star x^M), Lx - b)$ is nonempty and consists of finitely many points. Likewise, there is a nonempty Euclidean open subset of rate constants $a \in \mathbb{R}^6_{>0}$ for which $\mathbb{V}_{>0}(C(a \star x^M))$ is a smooth surface.

4.3 ACR and multistationarity

The perspective developed in Paper B turns out to be a powerful framework for understanding several chemically relevant geometric properties of the set of positive steady states, which we explore in Paper C.

We begin by discussing the notion of *absolute concentration robustness* (ACR), which was introduced by Shinar and Feinberg in their seminal paper [133]. In geometric terms, a network has ACR in the species X_i for a choice of rate constants $a \in \mathbb{R}^m_{>0}$ if the positive steady states are contained in a coordinate hyperplane $\mathbb{V}(x_i - c)$ for some $c \in \mathbb{R}_{>0}$. A weaker notion is that of *local ACR* in a species X_i, which means that the positive steady states are contained in finitely many hyperplanes of the form $\mathbb{V}(x_i - c)$ for $c \in \mathbb{R}_{>0}$. This was introduced in [123] as a necessary condition for ACR. See Figure 4.2 for an illustration.



(a) **(b)**

Figure 4.2: Schematic illustration of the set of positive steady states (blue) for a network with species X, Y, and Z that displays (a): ACR in Y; (b): local ACR (but not ACR) in Y.

Several previous works have focused on finding conditions that guarantee ACR or local ACR for *specific* or *all* values of the rate constants, which has turned out to be a very challenging task, see, e.g., [61] for a recent overview. In Paper B, we show that the problem is more well-behaved if we instead focus on *generic* local ACR and *generic* ACR, by giving a full characterization of both properties.

Theorem 4.7 (Theorem C.4.4, Corollary C.4.7)**.** *Consider a network with stoichiometric matrix $C \in \mathbb{R}^{n\times m}$ and kinetic matrix $M \in \mathbb{Z}^{n\times m}$ for which $\ker(C) \cap \mathbb{R}^m_{>0}$ is nonempty and $C(a\star x^M)$ is generically consistent. The network exhibits local ACR in X_i for generic rate constants if and only if it satisfies one of the following equivalent conditions:*

- *It holds that $\operatorname{rk}(C\operatorname{diag}(w)M_{\setminus i}^\top) < \operatorname{rk}(C)$ for all $w \in \ker(C)$, where $M_{\setminus i}$ is M without the ith row.*
- *There is a nonzero polynomial $g \in \left(\left\langle C(a\star x^M)\right\rangle : (x_1\cdots x_n)^\infty\right)\cap\mathbb{R}(a_1,\ldots,a_m)[x_i]$.*

The polynomial g is unique up to scaling, and if it has a single positive root for generic $a \in \mathbb{R}^m_{>0}$, then the network displays generic ACR in X_i.

The last part can be seen as a strengthening of the sufficient conditions ACR for for *specific* choices of rate constants found in [61, Proposition 3.8].

Example 4.8. For our running example with the IDHKP-IDH network from Example 4.1, we find the polynomial

$$a_4a_6x_4 - (a_3a_5 + a_3a_6) \in \left(\left\langle C(a\star x^M)\right\rangle : (x_1\cdots x_5)^\infty\right) \cap \mathbb{R}(a_1,\ldots,a_6)[x_4]$$

which clearly has a unique zero for all $a \in \mathbb{R}^6_{>0}$. We conclude that we have generic ACR in X_4. Indeed, this is one of the main features of this network that is studied in [133].

Next, we turn to the concept of multistationarity. A network with steady state system $F = (C(a\star x^M), Lx - b)$ is said to have the *capacity for multistationarity* if there is some choice of parameters $(a, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell$ such that $\mathbb{V}_{>0}(F_{a,b})$ has multiple points. Deciding whether a network admits multistationarity has been a central problem in the field of reaction network theory since its genesis in the 1970's, and has recently attracted attention in the biochemical context, since it is believed to play a role in cellular decision making [96]. Many sufficient and necessary conditions have been developed in recent years, and we refer to [39] for an overview of the state of the art.

One approach to detecting multistationarity is to study small networks that appear as submotives in larger ones. If we have *nondegenerate multistationarity* of a small network, in the sense that the multiple points of $\mathbb{V}_{>0}(F_{a,b})$ are nondegenerate zeros of $F_{a,b}$, then there are conditions that allow lifting the multistationarity to larger networks; see, e.g., [9, 31, 42, 84]. Motivated by this, a natural question is whether networks with the capacity for multistationarity also has the capacity for *nondegenerate* multistationarity. The *Nondegeneracy Conjecture* from 2017 of Joshi and Shiu says that the answer is yes under mild assumptions.

Conjecture 4.9 ([85, Conjecture 2.3])**.** Let $F = (C(a\star x^M), Lx - b)$ be a steady state system for a reaction network. Suppose that $\mathbb{V}_{>0}(F_{a,b})$ is finite for all $(a, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell$, and that the system admits p positive zeros for some parameters. Then there is a choice of parameters such that the system has p *nondegenerate* positive zeros.

Over the last years, the conjecture has been shown to be true for special classes of networks [83, 134]. In Paper C, we make some progress on the general case for $p = 2$, by assuming nodegeneracy, exploiting the parametrization of $\mathcal{Z}_{>0}$, and applying topological facts about maps between spaces of the same dimension (e.g., invariance of domain).

Theorem 4.10 (Theorem C.5.2)**.** *The nondegeneracy conjecture is true in the case $p = 2$ for networks that admit at least one nondegenerate positive steady state.*

In fact, it turns out that one can relax the global finiteness assumption of Conjecture 4.9 to more local assumptions; see Section 5 of Paper C for the full statement, and a discussion about how much the assumptions can be relaxed.

4.4 Parametric toricity

An important feature that steady state systems from biochemically relevant networks often display, on top of their vertically parametrized structure, is that there are *affine dependencies* between the monomial exponents. In databases such as ODEbase [105], it is therefore common to find networks where each nonempty set of positive steady states is the positive part of a scaling of a given toric variety.

This parametric notion of toricity has played a central role in the field already since the works of Horn and Jackson [78], although it was not put in the language of toric geometry until the work on *toric dynamical systems* by Craciun, Dickenstein, Shiu, and Sturmfels [41]. In Paper D, we explore this phenomenon from a new perspective, focusing on the vertically parametrized structure of the steady state system.

Definition 4.11. We say that a vertically parametrized system $F = C(a \star x^M)$ with real coefficients and $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$ displays *parametric $\mathcal{T}_A$-toricity* for an exponent matrix $A \in \mathbb{Z}^{d \times n}$ if, for each choice of parameters $a \in \mathbb{R}^m_{>0}$, the set of positive zeros $\mathbb{V}_{>0}(F_a)$ is either empty or equal to a coset $z \star \mathcal{T}_A$ of the positive toric variety $\mathcal{T}_A = \{t^A : t \in \mathbb{R}^d_{>0}\}$ (viewed as a multiplicative group) for some point $z \in \mathbb{R}^n_{>0}$ depending on a.

This notion is motivated by the fact that a network with parametric $\mathcal{T}_A$-toricity has the capacity for multistationarity if and only if it satisfies the rank condition

$$\operatorname{rk} \begin{bmatrix} B^\top \operatorname{diag}(h) \\ L \end{bmatrix} < n \quad \text{for some } h \in \mathbb{R}^n_{>0}, \tag{4.9}$$

where B is a Gale dual matrix to A, in the sense that the columns form a basis for $\ker(A)$ [116, 118, 131]. Note in particular that (4.9) does not rely on having an explicit formula for how the scaling vector z depends on a in the definition of parametric toricity.

The general problem of detecting when a variety is toric is classical, and has recently been approached with symbolic [66] and Lie-theoretic [86, 109] tools, but our question,

about whether a whole *parametrized family* of semialgebraic sets is toric is much less explored. In principle, it can be answered with quantifier elimination techniques [126], but for practical purposes, such computations are rarely feasible. Instead, several *sufficient* conditions have been developed in the context of reaction network theory, either based on binomiality [38, 116, 127, 131], or the graph-theoretic concept of complex-balanced steady states [28, 41, 57, 78].

In Paper D, our starting point is instead a *necessary* condition for parametric $\mathcal{T}_A$-toricity, namely, that $\mathbb{V}_{>0}(F_a)$ is *invariant* under componentwise multiplication by $\mathcal{T}_A$ for each $a \in \mathbb{R}^m_{>0}$. This is equivalent to each $\mathbb{V}_{>0}(F_a)$ being a union of (possibly zero or infinitely many) cosets of $\mathcal{T}_A$, and is fully characterized by a linear algebra condition.

Theorem 4.12. *Let $F = C(a \star x^M)$ be a vertically parametrized system with $C \in \mathbb{R}^{s\times m}$, $M \in \mathbb{Z}^{n\times m}$ and $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Let $\mathscr{C}$ be a basis for $\ker(C)$ of circuit vectors. Then $\mathbb{V}_{>0}(F_a)$ is $\mathcal{T}_A$-invariant for all $a \in \mathbb{R}^m_{>0}$ for a given $A \in \mathbb{Z}^{d\times n}$ if and only if*

$$A(M_i - M_j) = 0 \text{ for all } i, j \in \operatorname{supp}(v) \text{ and } v \in \mathscr{C}. \tag{4.10}$$

Furthermore, for a given $a \in \mathbb{R}^m_{>0}$ and under the assumption of $\mathcal{T}_A$-invariance, the following is a bijection onto the set of cosets $\mathbb{V}_{>0}(F_a)/\mathcal{T}_A$ for any $b \in A(\mathbb{R}^n_{>0})$:

$$\mathbb{V}_{>0}(F_a, Ax - b) \to \mathbb{V}_{>0}(F_a)/\mathcal{T}_A, \quad z \mapsto z \star \mathcal{T}_A.$$

The condition (4.10) also characterizes $\mathcal{T}_A$-invariance for the complex varieties $\mathbb{V}_{\mathbb{C}^*}(F_a)$, making this is an example how the complex behavior of a vertically parametrized system often agrees with the positive real behavior, as alluded to at the end of Section 4.1. It also generalizes the situation for freely parametrized systems, where $\mathcal{T}_A$-invariance precisely corresponds to *quasihomogeneity* of the polynomials.

The second part of the theorem is a consequence of Birch's theorem, and the domain of the bijection is the set of solutions to an augmented vertically parametrized system. In Paper D, we give conditions for when it contains a unique point, based on techniques previously developed for proving monostationarity in reaction network theory.

With these conditions in place, we are able to both rule out and prove toricity for biologically relevant networks in the database ODEbase [105] with up to a hundred reactions, where previous sufficient criteria either gave inconclusive results or were computationally infeasible.

Example 4.13. For the IDHKP-IDH network from Example 4.1, it is not hard to see that we for all $a \in \mathbb{R}^6_{>0}$ have the monomial parametrization

$$\mathbb{R}^2_{>0} \to \mathbb{V}_{>0}(C(a \star x^M)), \quad (t_1, t_2) \mapsto \left(t_1, t_2, \tfrac{a_1}{a_2+a_3}\, t_1 t_2, \tfrac{a_3(a_5+a_6)}{a_4 a_6}, \tfrac{a_1 a_3}{a_6(a_2+a_3)}\, t_1 t_2\right).$$

However, one can also apply our conditions to find $\mathcal{T}_A$-invariance for the matrix

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix},$$

and we prove in Example D.6.4 that the coset-counting system $(C(a \star x^M), Ax - b)$ has a unique positive zero for all $(a, b) \in \mathbb{R}^6_{>0} \times A(\mathbb{R}^5_{>0})$. Hence, we conclude that we have parametric toricty. Note that since the 4th column of A is zero, this also proves that we have ACR in X_4 for all parameter values. Using the Gale dual matrix

$$B = \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

for C, we find that

$$\det \begin{bmatrix} B^\top \operatorname{diag}(h) \\ L \end{bmatrix} = -(h_1h_3h_4 + h_1h_4h_5 + 2\,h_2h_3h_4 + h_2h_4h_5 + h_3h_4h_5)$$

which does not vanish for any $h \in \mathbb{R}^5_{>0}$, so the rank condition (4.9) is not satisfied, and we conclude that the network does not have the capacity for multistationarity.

4.5 Open problems

An interesting future research direction is to continue the program of generalizing results for freely parametrized systems to the vertically parametrized setting.

A natural candidate for a next result to generalize is characterization of *generic irreducibility* of Khovanskii [89] and Yu [148]. This is an interesting property from the point of view of reaction network theory, since it simplifies the analysis of many properties. For instance, generic irreducibility combined with generic *local* ACR (for which we have a simple rank condition from Theorem 4.7) implies generic ACR. Similarly, generic irreducibility combined with the toric invariance of Theorem 4.12 gives toricity for generic parameter values.

Another interesting concept to generalize is *extremal genericity* of Bender and Spaenlehauer [15]. In [15, Proposition 1.12], it is shown that only the coefficients corresponding to the vertices of the Newton polytopes need to be chosen generically for a freely parametrized system to attain the generic properties given by Theorem 4.4 (whereas the remaining coefficients can be fixed). It is natural to believe that the circuits of the column matroid of C play a role, together with the polyhedral structure of the convex hull of M, in determining which parameters a need to be chosen generically to obtain the generic behavior in the vertically parametrized setting.

Yet another interesting phenomenon to study in the vertically parametrized setting is the property of *generically lacking irreducible components in the coordinate hyperplanes*. When a system with nonnegative exponents has this property, it means that studying its generic behavior in $(\mathbb{C}^*)^n$ captures its generic behavior in $\mathbb{C}^n$. This is especially

interesting in the case of root counting, since our current tropical methods only work in the torus. In the reaction-network theoretic setting, it is related to questions of boundary steady states and permanence. Several results exist in the (square) freely-parametrized setting [101, 129]. One of the simplest sufficient conditions is the existence of an independent constant parameter in each polynomial of the system, which also holds in the vertical setting.

Theorem 4.14 (Theorem B.3.19)**.** *Let* $F = (C(a \star x^M), Lx - b)$ *be an augmented vertical system with nonnegative exponents and* $C \in \mathbb{C}^{s\times m}$ *with* $\operatorname{rk}(C) = s$*. Suppose that for some index set* $I \subseteq [m]$*, the submatrix* M_I *is the zero matrix, and the submatrix* C_I *is diagonal (after reordering of the columns) and of full rank. Then the variety* $\mathbb{V}_{\mathbb{C}}(F_{a,b})$ *in* $\mathbb{C}^n$ *lacks components contained in coordinate hyperplanes for generic* $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$*.*

In the reaction-network-theoretic scenario, this is satisfied for networks where each species has an *inflow* reaction $0 \to X_i$, which is common in certain engineering applications (see, e.g., [58, Section 4.2.1] on continuous-flow stirred-tank reactors). However, it would be desirable to have more widely applicable conditions.

5 The method of moments

This chapter provides an overview of Papers E and F. The emphasis is on our results concerning identifiability, which is the topic of Section 5.2, as well as the determinantal structure of the moment varieties of several classical distributions, discussed in Section 5.3.

5.1 The moment system

The *method of moments* is a classical statistical method, going back to Pearson's 1894 paper [125]. In its simplest form, the aim is to solve the following parameter estimation problem: Given a random variable X (assumed to be univariate, for simplicity) with probability density p_θ depending on some unknown parameters $\theta = (\theta_1, \dots, \theta_n)$, estimate θ from a sample $x_1, \dots, x_N$. For many distributions, the moments $m_r = \mathbb{E}[X^r]$ for $r \in \mathbb{Z}_{\geq 0}$ are rational functions in the distribution parameters θ. We can therefore estimate θ by computing an appropriate number d of sample moments $\widehat{m}_r = \frac{1}{N}\sum_{i=1}^N x_i^r$ for $r \in [d]$, and then solving the *moment system*, given by

$$F = \left(m_1(\theta) - \widehat{m}_1,\ \ m_2(\theta) - \widehat{m}_2,\ \ \dots,\ \ m_d(\theta) - \widehat{m}_d\right). \tag{5.1}$$

We will view this as a system with *parametric constant terms*, in the sense of Chapter 2, with the distribution parameters θ as variables, and the sample moments as parameters (note the slight conflict in terminology here). Provided that the sample size N is large enough, it follows from the law of large numbers that we can expect arbitrarily good approximations of θ among the zeros of (5.1), and in many cases, one can show that the method of moments gives rise to a consistent estimator [146, Theorem 9.6].

The fact that the method of moments is a polynomial estimation approach gives it many desirable properties from a computational perspective [13]. For example, provided that d is large enough for (5.1) to have generically finitely many zeros, there is a well-defined generic root count (the *identifiability degree*), which gives a bound on the number of (complex) zeros of (5.1). This is an advantage over other estimation methods such as maximum likelihood estimation, which can sometimes lack upper bounds on the number of critical points [3].

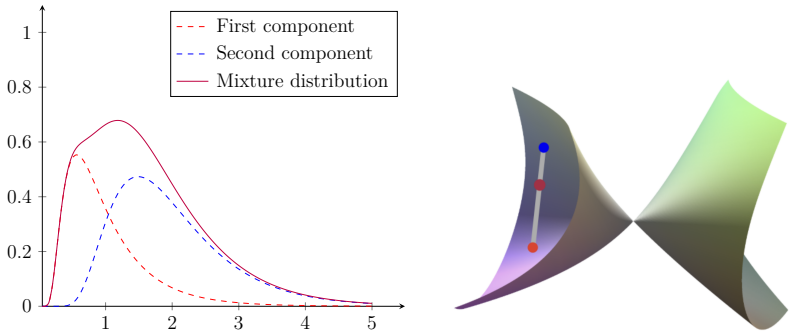


Figure 5.1: Schematic illustration of a point on $\mathrm{Sec}_2(\mathcal{M}_3)$ for the inverse Gaussian distribution, and a corresponding two-component mixture distribution. The figure is based on figures from Paper E.

5.2 Moment identifiability

A fundamental information-theoretic question one can ask about the method of moments is how many $d > 0$ sample moments are needed to have *generic finite identifiability*, in the sense that the system (5.1) has finitely many complex zeros for generic values of the sample moments $\widehat{m}_1, \ldots, \widehat{m}_d$. Geometrically, this corresponds to asking when the map

$$\mathbb{C}^n \dashrightarrow \mathbb{C}^d, \quad \theta \mapsto (m_1(\theta), \ldots, m_d(\theta))$$

has generically finite fibers. Since the map is rational, the Euclidean closure of the image is an algebraic variety $\mathcal{M}_d \subseteq \mathbb{C}^d$, referred to as the dth *moment variety*, and by the theorem of dimension of fibers (cf. Proposition 2.3), we have generic finite identifiability precisely when $\dim(\mathcal{M}_d) = n$. A central focus of the algebraic-geometric study of the method of moments is therefore to determine the dimension of moment varieties.

Of particular interest are *mixture distributions*. Suppose we have a given parametric distribution with dth moment variety $\mathcal{M}_d$ of sufficient dimension for generic finite identifiability. Then the moment variety of the k-mixture of this distribution is the *kth secant variety* $\mathrm{Sec}_k(\mathcal{M}_d)$ (the Euclidean closure of the union of all k-secants through points in $\mathcal{M}_d$); see Figure 5.1 for a sketch, and Section F.2.1 for a detailed explanation.

In this setting, finite identifiability corresponds to proving that $\mathcal{M}_d$ is *k-nondefective*, in the sense that

$$\dim(\mathrm{Sec}_k(\mathcal{M}_d)) = k \dim(\mathcal{M}_d) + k - 1.$$

Generic unique identifiability (up to the label-swapping symmetry on the mixture components) corresponds to proving that a generic point of $\mathrm{Sec}_k(\mathcal{M}_d)$ lies on a unique k-secant through $\mathcal{M}_d$. Secant varieties, nondefectivity and the relation to identifiability problems have been central topics in algebraic geometry for a long time, and goes back at least to works by the Italian school of algebraic geometry in the late 19th and early 20th century (see [16] for a historical overview). In more recent years, classical nondefectivity results have found applications in areas such as rigidity theory [43], sums of squares decompositions [122], and tensor decomposition [16].

In the context of mixture identifiability, most work has focused on Gaussians [1, 4, 5, 18, 19, 104], but also product distributions [2], and Dirac and Pareto distributions [68]. In Papers E and F, we tackle mixtures of several other classical distributions.

The simplest cases are 1-parameter distributions, for which the moment varieties are curves. It is well known that nonplanar curves are always nondefective (see, e.g., the discussion in [150, Section 1]), which gives generic finite identifiability for many distributions. In Paper E, we go further, and show that for the *exponential* and *chi-squared* distribution, the moment variety $\mathcal{M}_d$ is linearly isomorphic to the standard rational normal curve, which has the following strong consequence.

Theorem 5.1 (Proposition E.5.1)**.** *For k-mixtures of the exponential and chi-squared distribution, we have generic unique identifiability from the first $2k-1$ moments.*

In Paper F, we treat mixtures of the *inverse Gaussian* and *gamma* distribution, which are 2-parameter distributions for which $\mathcal{M}_d$ is a surface for $d \geq 2$. This is precisely analogous to results proven for the Gaussian case in [5, 104]. Several of the key tools used in those proofs are the same as the ones we use. This includes, in particular, a classification of defective surfaces due to Terracini [33, 139], where the relevant cases can be ruled out by intersection-theoretic calculations, and a recent result on generic uniqueness of k-secants passing through points on secant varieties from [111].

Theorem 5.2 (Theorems F.1.1 and F.1.2)**.** *For k-mixtures of the inverse Gaussian distribution and k-mixtures of the gamma distribution, we have*

(i) generic finite identifiability from the first $3k-1$ moments;

(ii) generic unique identifiability from the first $3k+2$ moments.

The more complicated structure of our underlying moment varieties compared to the Gaussian case led us to develop a different and more general approach for the necessary intersection-theoretic calculations, relying only on a *parametrization* of the underlying moment variety, instead of a *determinantal realization* with linear entries (which is needed in the approach of [5], but which we do not have for the inverse Gaussian case).

It is currently an open question whether we have generic unique identifiability from less than $3k+2$ moments. In Paper E, we conjecture that $3k$ moments suffices, based on numerical experiments.

5.3 Determinantal structure of moment varieties

In order to apply the techniques we used in Paper F, we first needed a thorough understanding of the moment surfaces for the inverse Gaussian and gamma distribution, especially regarding the singular loci of their projective closures $\bar{\mathcal{M}}_d \subseteq \mathbb{P}^d$. Motivated

by this, we undertook in Paper E a careful analysis of these projective varieties and their homogeneous vanishing ideals, with the following as our main result. In particular, part (ii) proves [67, Conjecture 3.2.5].

Theorem 5.3 (Sections E.3 and E.4)**.** *Let $d \geq 3$.*

(i) For the inverse Gaussian distribution, the singular locus of $\bar{\mathcal{M}}_d$ is given by the line $\mathbb{V}(x_0, x_1, \ldots, x_{d-2})$ and the point $\mathbb{V}(x_1, x_2, \ldots, x_d)$ in $\mathbb{P}^d$. The vanishing ideal of $\bar{\mathcal{M}}_d$ is generated by the maximal minors of the $(3 \times d)$-matrix

$$\begin{pmatrix} x_0^2 & x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-2} \\ 0 & x_1 & 3x_2 & 5x_3 & 7x_4 & \cdots & (2d-3)x_{d-1} \\ x_1^2 & x_2 & x_3 & x_4 & x_5 & \cdots & x_d \end{pmatrix}.$$

The Hilbert series of the coordinate ring is $\frac{1}{(1-t)^3}\left(1 + (d-2)t + \binom{d-1}{2}t^2 + \binom{d-1}{2}t^3\right)$.

(ii) For the gamma distribution, the singular locus of $\bar{\mathcal{M}}_d$ is given by the two points $\mathbb{V}(x_0, x_1, \ldots, x_{d-1})$ and $\mathbb{V}(x_1, x_2, \ldots, x_d)$ in $\mathbb{P}^d$. The vanishing ideal of $\bar{\mathcal{M}}_d$ is given by the maximal minors of the $(3 \times d)$-matrix

$$\begin{pmatrix} 0 & x_1 & 2x_2 & 3x_3 & \cdots & (d-1)x_{d-1} \\ x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-1} \\ x_1 & x_2 & x_3 & x_4 & \cdots & x_d \end{pmatrix}.$$

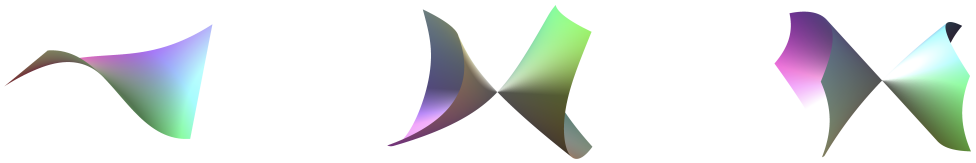
The Hilbert series of the coordinate ring is $\frac{1}{(1-t)^3}\left(1 + (d-2)t + \binom{d-1}{2}t^2\right)$.

In both cases, the minors described above constitute a Gröbner basis with respect to the reverse lexicographic ordering, and the coordinate rings are Cohen–Macaulay.

These results have a similar flavor as those previously proven for the Gaussian distribution in [4], for which $\bar{\mathcal{M}}_d$ is cut out by the maximal minors of

$$\begin{pmatrix} 0 & x_0 & 2x_1 & 3x_2 & \cdots & (d-1)x_{d-2} \\ x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-1} \\ x_1 & x_2 & x_3 & x_4 & \cdots & x_d \end{pmatrix}.$$

See also [67, Chapter 3] where the ideal-structure of several affine moment varieties is analyzed. Similarly to the previous section, the more complicated structure of the determinantal realizations for our distributions (which, contrary to the Gaussian case, are not Hankel matrices) led us to use more powerful machinery from commutative algebra for the proofs, including the Thom–Porteous–Giambelli formula for the degree of determinantal ideals for the inverse Gaussian case (see Section E.3), and polarization and Stanley–Reisner theory for the gamma distribution (see Section E.4).



(a) Gaussian **(b)** Inverse Gaussian **(c)** Gamma

Figure 5.2: Illustration of the moment variety $\mathcal{M}_3$ for three distributions. Figure from Paper E.

5.4 Open problems

While the distributions discussed in previous sections are interesting in their own right, we also see these results as a stepping stone towards a more general theory. In particular, it would be interesting to identify unifying statistical criteria for 1-parameter exponential families that ensure that the moment variety is linearly isomorphic to the standard rational normal curve, as well as criteria for 2-parameter exponential families that ensure properties (1) and (2) of Theorem 5.2.

Another direction for future work is to prove identifiability results for mixtures of *3-parameter distributions*, where the underlying moment variety is a *threefold*, rather than a surface (e.g., the scaled non-central chi-squared distribution, or variations of the generalized beta distribution). In this setting, there is also a classification of defectivity [34], with many of the defective cases having similar flavor as those in Terracini's classification of defective surfaces.

Finally, the method of moments can also be analyzed from the point of view of optimization. In practical applications (see, e.g., [73]), it is common not to directly solve the system (5.1), but instead make the system overdetermined, and solve the optimization problem

$$\min_{\theta \in \mathbb{C}^n} \sum_{i=1}^{d} (m_i(\theta) - \widehat{m}_i)^2. \tag{5.2}$$

One way to quantify the complexity of this problem is to considerthe geometric optimization problem

$$\min_{m \in \mathcal{M}_d} \sum_{i=1}^{d} (m_i - \widehat{m}_i)^2,$$

for which the number of complex critical points for generic sample moments $\widehat{m}$ is the *Euclidean distance degree* (ED degree) of the variety $\mathcal{M}_d$.

The ED degree of a variety was first introduced by Draisma, Horobeţ, Ottaviani, Sturmfels, and Thomas in [53], and can, in the language of Chapter 3, be seen as a generic root count of a Lagrangian system, with $\widehat{m}$ as parameters.

Determining the ED degree for various classes of varieties is still an active area of research in applied algebraic geometry, and has recently been studied for, e.g., toric

varieties [75], phylogenetic varieties [60], multi-view varieties [36], and neurovarieties [94]. The ED degree has, however, not been explored for moment varieties. In Paper E, we take the first steps towards studying this problem computationally.

Proposition 5.4 (Proposition E.5.6)**.** *For the inverse Gaussian, gamma, and Gaussian distribution, the ED degree of $\mathcal{M}_3$ is 12, 10, and 7, respectively.*

Finding the ED degrees of $\mathcal{M}_d$ for $d > 3$ for these distributions remains an open problem. Part of the challenge in applying the existing theory of ED degrees is that moment varieties typically have singularities. An interesting direction for future work is to apply the machinery of Chern–Mather classes, which has been used to compute ED degrees for similar singular varieties in, e.g., [94, 112], combined with the fact that these moment varieties have highly structured determinantal realizations.

6

3D genome reconstruction

In this chapter, we give an overview of Paper G. We begin by explaining the geometric setup of 3D genome reconstruction in Section 6.1, and then discuss the problem of identifiability in Section 6.2. We end with a discussion about how to practically carry out a reconstruction in Section 6.3, as well as some open problems in Section 6.4.

6.1 Problem formulation

The goal of *3D genome reconstruction* is to recover the 3D configuration of chromosomes from (indirect) information about the pairwise distances between various points along them. This is motivated by a large body of works in genomics that indicate that the structure plays an important role in a wide range of cellular processes, including gene regulation [8, 141] and the DNA damage repair system [98, 145].

The field encompasses several different techniques, including both *single-cell methods* (see [10] for an overview, and [49] for a recent algebraic work in this direction), and *bulk methods* that collect data from large amounts of cell [120]. The latter can be divided into *ensemble methods*, which aim to find several structures that together explain data [79, 124, 130], and *consensus methods*, which aim to find a single main structure that explains the data [99, 142, 149]. We will focus on consensus methods in what follows.

The simplest and most well-studied setting for 3D genome reconstruction is *haploid* cells, where each chromosome comes in a single copy, as opposed to *diploid* cells, that carry both a maternal and paternal copy of each chromosome. In the haploid scenario, a chromosome can be modeled as a sequence of some number n of points

$$(x_1, x_2, \ldots, x_n) \in (\mathbb{R}^3)^n$$

where each point x_i corresponds to a DNA segment of some length determined by the resolution of the experimental data (for instance, in Section 6.3, we show results for a dataset with $n = 343$, and a resolution of 500 000 base pairs per segment).

The goal is to recover the sequence $(x_i)_{i=1}^n$ up to scaling and rigid transformations (viewed as an action of $\mathrm{O}(3) \ltimes \mathbb{R}^3$), based on indirect data about the pairwise distances

of the points. Typically, this data comes from a technique called *high-throughput chromosome conformation capture* (Hi-C for short), which measures how often various points of a chromosome appear in close proximity to each other; see [95] for details. More specifically, we get a contact count $c_{ij} > 0$ for each $\{i, j\} \in \binom{[n]}{2}$, which is assumed to depend on $\|x_i - x_j\|$ through the power-law formula

$$c_{ij} = \beta \|x_i - x_j\|^{\alpha}, \tag{6.1}$$

where $\alpha < 0$ is called the conversion factor (and depends on the experimental setup), and $\beta > 0$ is a scaling factor. See [102] for a discussion of this model. For simplicity, we will set $\alpha = -2$ in most of this chapter, and since we only aim to reconstruct chromosomes structures up to scaling, we can without loss of generality fix $\beta = 1$.

For a given α, what we just described is a variation of the classical Euclidean distance problem of finding a point configuration given pairwise distances; it can also be seen as a rigidity theory problem for a complete graph, as in [30]. It is well known that we have unique identifiability in this setting (see, e.g., [93, Proposition 3.2]). For noisy data, there are furthermore established methods in the realm of 3D genome reconstruction based on semidefinite programming (e.g., the software ChromSDE [149]) or maximum likelihood estimation (e.g., the software PASTIS [142]) for determining the structure.

In the diploid case, we instead want to determine the 3D structure of a *chromosome pair*, which can be modeled as a sequence of homologous pairs of points

$$\Big((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\Big) \in (\mathbb{R}^3 \times \mathbb{R}^3)^n \,,$$

where x_i corresponds to a maternal segment, and y_i to a homologous paternal segment. In the simplest experimental setups, the segments corresponding to the points x_i and y_i cannot be distinguished, which means that each contact count c_{ij} will be an aggregated count of four interactions, depending on $\|x_i - x_j\|$, $\|x_i - y_j\|$, $\|y_i - x_j\|$ and $\|y_i - y_j\|$, as illustrated by Figure 6.1.

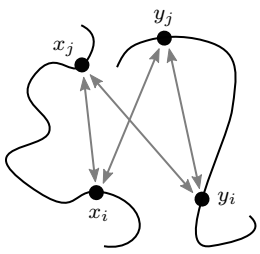


Figure 6.1: Four interactions contributing to c_{ij}. Adapted from Paper G.

6.2 Identifiability in the diploid setting

In the diploid setting, it is not immediately clear whether the structure can be identified from the data. Since the number of constraints coming from the contacts is $\binom{n}{2}$, which grows faster than the number of unknown coordinates $6n-6$ (modulo the action of $\mathrm{SO}(3) \ltimes \mathbb{R}^3$), one could naively expect to have at least generic finite identifiability for sufficiently large n. However, in [14], it is shown that under the model

$$c_{ij} = \frac{\beta}{\|x_i - x_j\|^2 + \|x_i - y_j\|^2 + \|y_i - x_j\|^2 + \|y_i - y_j\|^2}, \tag{6.2}$$

we do not have finite identifiability for any n.

In Paper G, we investigate another model that generalizes (6.1) to the diploid setting, where we assume that the four contacts aggregate additively, similarly to the models used in [32, 147], so that

$$c_{ij} = \beta\Big(\|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha + \|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha\Big) \tag{6.3}$$

for an exponent $\alpha < 0$ and a scaling factor $\beta > 0$. In this setting, we prove generic finite identifiability for $\alpha = -2$ and $n = 12$ by applying Proposition 2.4 to the map $(\mathbb{C}^3 \times \mathbb{C}^3)^n \dashrightarrow \mathbb{C}^{\binom{[n]}{2}}$ given by

$$\underbrace{(x_i, y_i)_{i=1}^n}_{\text{Configuration}} \mapsto \underbrace{\Big(\tfrac{1}{Q(x_i-x_j)} + \tfrac{1}{Q(x_i-y_j)} + \tfrac{1}{Q(y_i-x_j)} + \tfrac{1}{Q(y_i-y_j)}\Big)_{1\le i<j\le n}}_{\text{Contact counts}},$$

where we let $Q\colon \mathbb{C}^3 \to \mathbb{C}$ be defined by $Q(z) = z_1^2 + z_2^2 + z_3^2$. For $n = 12$, we find a point where the Jacobian has full rank 66, and show that for generic contact counts, all irreducible components of the 6-dimensional fiber are orbits of the group of complex rigid transformations $\mathrm{SO}(3, \mathbb{C}) \ltimes \mathbb{C}^3$. From this, we obtain the following.

Theorem 6.1 (Theorem G.2)**.** *For generic* $(c_{ij}) \in \mathbb{R}^{\binom{[12]}{2}}$*, the system*

$$\|x_i - x_j\|^2 + \|x_i - y_j\|^2 + \|y_i - x_j\|^\alpha + \|y_i - y_j\|^2 = c_{ij} \quad \text{for } \{i,j\} \in \tbinom{[12]}{2} \tag{6.4}$$

has finitely many solutions $(x_i, y_i)_{i=1}^{12} \in (\mathbb{R}^3 \times \mathbb{R}^3)^{12}$ *up to the action of* $\mathrm{O}(3) \ltimes \mathbb{R}^3$*.*

In practice, it is unlikely to be computationally feasible to carry out reconstructions by solving (6.4), as certified numerical experiments with monodromy (see Remark 3 in Paper G) show that the identifiability degree is at least 1000 for $n = 12$, thus indicating a very high algebraic complexity in finding the structure from the aggregated counts.

This fits with the general sentiment that diploid 3D genome reconstruction from aggregated counts alone is very hard, and that additional data is needed to find reliable

reconstructions; see [132] for a recent overview of the state of the art. For instance, in [14], the authors prove identifiability using information about higher-order contacts between more than two genomic loci, and develop a reconstruction method based around this. In [32], they instead assume that some of the data is *phased* in the sense that it distinguishes maternal and paternal copies of the same gene, and experimentally show that this improves the quality of the reconstruction in simulated data sets. In Paper G, we study an idealized version of this scenario, and rigorously prove a stronger generic finite identifiability result than Theorem 6.1.

Our setup is that we assume that there is a partition $[n] = D \sqcup I$ of the pairs into distinguishable pairs D and indistinguishable ones I, in the sense that x_i and y_i can be distinguished experimentally if $i \in D$, but not if $i \in I$. This gives rise to three types of contacts, illustrated in Figure 6.2:

- If $i, j \in D$, then c_{ij} splits into four observable *unambiguous* contact counts

$$c^U_{XX}(i,j) = \beta\|x_i - x_j\|^\alpha, \quad c^U_{XY}(i,j) = \beta\|x_i - y_j\|^\alpha,$$
$$c^U_{YX}(i,j) = \beta\|y_i - x_j\|^\alpha, \quad c^U_{YY}(i,j) = \beta\|y_i - y_j\|^\alpha.$$

- If $i \in D$ and $j \in I$, then c_{ij} splits into two *partially ambiguous* contact counts

$$c^P_X(i,j) = \beta\Big(\|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha\Big), \qquad c^P_Y(i,j) = \beta\Big(\|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha\Big).$$

- If both $i, j \in I$, then c_{ij} is said to be an *ambiguous* contact count, which we from now on will denote

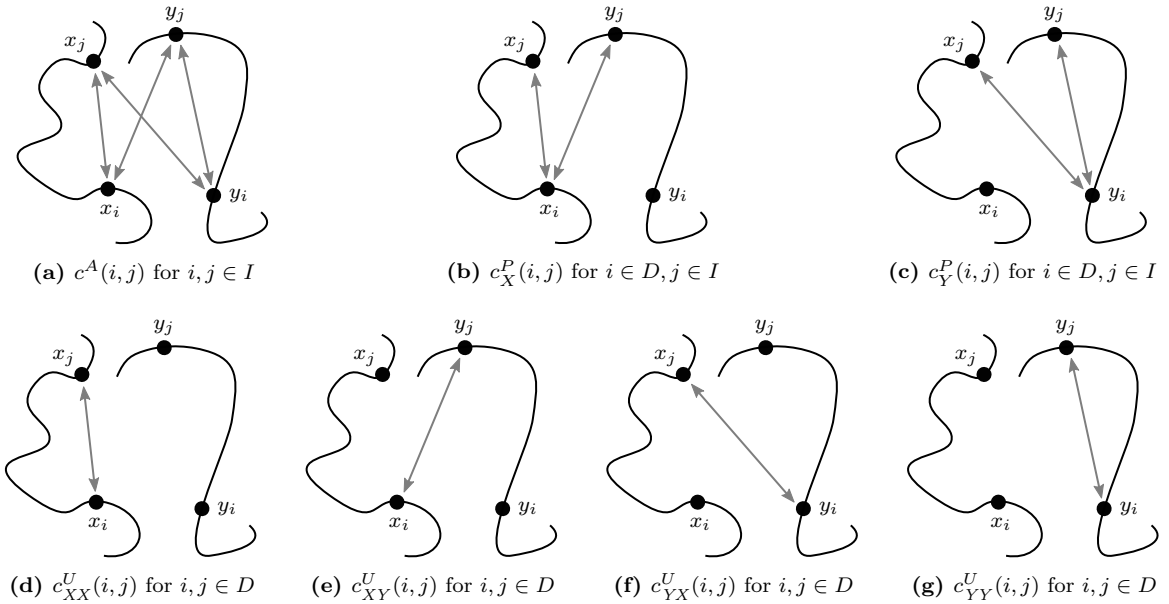
$$c^A(i,j) = \beta\Big(\|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha + \|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha\Big).$$

In Paper G, we show that $|D| \geq 3$ is enough to ensure generic finite identifiability. The starting point of the analysis is to first note that the pairs $(x_i, y_i)_{i \in D}$ are uniquely identifiable up to the $\mathrm{SO}(3) \ltimes \mathbb{R}^3$ action, since they corresponds to the haploid scenario (6.1) with $2|D|$ points. We therefore take $(x_i, y_i)_{i \in D}$ for given, and use them to determine $(x_i, y_i)_{i \in I}$.

It turns out that if $|D| \geq 3$, this can be done with generic finite identifiability for each pair (x_i, y_i) with $i \in I$ *considered individually*. In the $\alpha = -2$ case, this can be shown through an application of Proposition 2.4 to

$$(\mathbb{C}^3)^6 \times \mathbb{C}^3 \times \mathbb{C}^3 \dashrightarrow (\mathbb{C}^3 \times \mathbb{C})^6, \quad \Big((z_i)_{i=1}^6, x, y\Big) \mapsto \left(z_i, \tfrac{1}{Q(z_i - x)} + \tfrac{1}{Q(z_i - x)}\right)_{i=1}^6,$$

which shows that this is a dominant rational map, and therefore has generically finite fibers. Concretely, this means that for a given indistinguishable pair (x, y), and six known points $z_1, \ldots, z_6 \in \mathbb{R}^3$ coming from sufficiently generic distinguishable pairs, as well as sufficiently generic corresponding partially ambiguous contacts, $c_1, \ldots, c_6 \geq 0$, we will have at most finitely many compatible values of (x, y).



(a) $c^A(i,j)$ for $i,j \in I$ **(b)** $c^P_X(i,j)$ for $i \in D, j \in I$ **(c)** $c^P_Y(i,j)$ for $i \in D, j \in I$

(d) $c^U_{XX}(i,j)$ for $i,j \in D$ **(e)** $c^U_{XY}(i,j)$ for $i,j \in D$ **(f)** $c^U_{YX}(i,j)$ for $i,j \in D$ **(g)** $c^U_{YY}(i,j)$ for $i,j \in D$

Figure 6.2: Ambiguous, partially ambiguous and unambiguous contact counts. Figure from Paper G.

For more general rational exponents $\alpha < 0$ with $\frac{\alpha}{2} = \frac{a}{b}$ for $a, b \in \mathbb{Z}$ with $b > 0$ and $\gcd(a,b) = 1$, the proof can be adapted by considering the variety

$$X = \Big\{ \Big((z_i, r_i, s_i)_{i=1}^6, x, y \Big) \in (\mathbb{C}^3 \times \mathbb{C} \times \mathbb{C})^6 \times \mathbb{C}^3 \times \mathbb{C}^3 : \\ Q(x - z_i)^a = r_i^b \neq 0,\ Q(y - z_i)^a = s_i^b \neq 0 \text{ for all } i \in [6] \Big\},$$

and the projection

$$X \to (\mathbb{C}^3 \times \mathbb{C})^6, \quad \Big((z_i, r_i, s_i)_{i=1}^6, x, y \Big) \mapsto (z_i, r_i + s_i)_{i=1}^6.$$

By a careful analysis of the Jacobian of the projection, we conclude that each component of X maps dominantly to the domain, and generic finite identifiability follows.

Theorem 6.2 (Corollary G.1)**.** *Let $\alpha \in \mathbb{Q} \cap (-\infty, 0)$. Then the system*

$$\|x - z_i\|^\alpha + \|y - z_i\|^\alpha = c_i \quad \textit{for } i \in [6] \tag{6.5}$$

has finitely many solutions in $\mathbb{R}^3 \times \mathbb{R}^3$ for generic $z_1, \ldots, z_6 \in \mathbb{R}^3$ and $c_1, \ldots, c_6 \in \mathbb{R}_{>0}$.

Numerical experiments indicate that the identifiability degree for $\alpha = -2$ is 80, or 40 if we count up to the symmetry $(x, y) \mapsto (y, x)$. This makes it feasible to solve (6.5) numerically a large number of times, which plays a key role in the next section.

6.3 A new reconstruction method

Inspired by the "local" identifiability result Theorem 6.2, we suggest a new reconstruction approach for partially phased data in Paper G, where the main novelty lies in an initial pair-by-pair estimation for (x_i, y_i) with $i \in I$, which reduces the likelihood of obtaining non-global local minima in subsequent optimization steps.

Concretely, our method consists of the following main steps:

1. Estimation of $(x_i, y_i)_{i\in D}$ with a standard method for haploid 3D genome reconstruction (e.g., ChromSDE [149]).
2. Preliminary estimation of each (x_i, y_i) for $i \in I$ individually, based on repeatedly solving (6.5) for $z_1, \dots, z_6 \in \cup_{j\in D}\{x_j, y_j\}$ and corresponding partially ambiguous contacts $c_1, \dots, c_6$.
3. Refinement of $(x_i, y_i)_{i\in I}$ by applying standard local optimization techniques to the optimization problem

$$\min_{\{x_i,y_i\}_{i\in I}} \sum_{i\in D, j\in I} \left(\left(c_X^P(i,j) - \frac{1}{\|x_i - x_j\|^2} - \frac{1}{\|x_i - y_j\|^2} \right)^2 + \left(c_Y^P(i,j) - \frac{1}{\|y_i - x_j\|^2} - \frac{1}{\|y_i - y_j\|^2} \right)^2 \right). \tag{6.6}$$

4. A final clustering step to disambiguate between the estimations (x_i, y_i) and (y_i, x_i) for each $i \in I$ (see Section 4.4 of Paper G for details).

In Section 5.1 of Paper G, we show through experiments on simulated data sets that this approach compares favorably to standard methods for diploid 3D genome reconstruction under our modeling assumptions (see Figure 6.3 for some examples of reconstructions based on our method). In Section 5.2, we apply our technique to a data set for mouse X chromosomes previously studied in [32, 47], and recover previously known structural features (see Figure 6.4 for the reconstruction). The contact counts used as input for the estimation, as well as the computed contact counts after the reconstruction are displayed in Figure 6.5 in the form of matrices

$$c^U = \begin{bmatrix} c^U_{XX} & c^U_{XY} \\ c^U_{YX} & c^U_{YY} \end{bmatrix} \in \mathbb{R}^{2|D|\times 2|D|}, \quad c^P = \begin{bmatrix} c^P_X \\ c^P_Y \end{bmatrix} \in \mathbb{R}^{2|D|\times|I|}, \quad c^A \in \mathbb{R}^{|I|\times|I|},$$

where $c^U_{XX}, c^U_{XY}, c^U_{YX}, c^U_{YY}, c^A$ are interpreted as symmetric matrices.

6.4 Open problems

The results in Paper G opens up several geometric problems. First of all, it would be interesting if one can prove that the identifiability degree of (6.5) is 80. This problem has similar flavor as the rigidity-theoretic counting problems explored in [30] (see also

Section 6.4 of Paper A). Hence, this poses an interesting challenge for current machinery in enumerative geometry and rigidity theory. It would also be interesting to better understand how the number of *real* solution vary in parameter space, since this has a big influence on the computational complexity of the reconstruction approach.

Secondly, Step 2 in the reconstruction approach of Section 6.3 can be replaced by an optimization step as follows. Suppose $|D| \geq 4$, pick seven points $z_1, \dots, z_7 \in \cup_{i \in D}\{x_i, y_i\}$ with partially ambiguous contacts $c_1, \dots, c_7$ with respect to (x, y), and solve

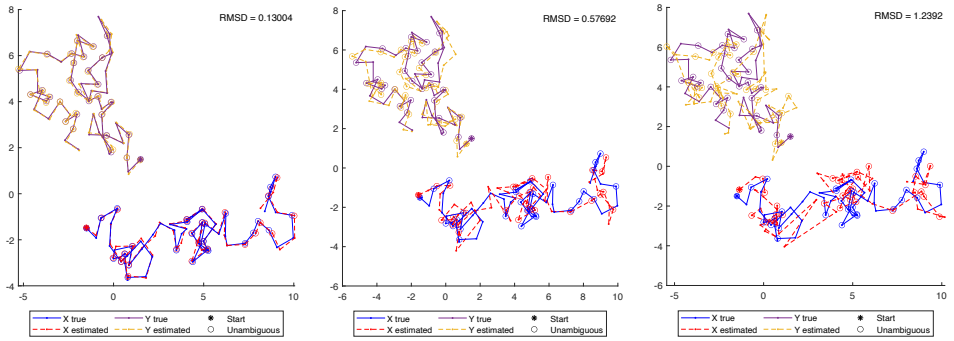
$$\min_{(x,y)\in\mathbb{R}^3\times\mathbb{R}^3} \sum_{i=1}^{7} \left(c_i - \tfrac{1}{\|x-z_i\|^2} - \tfrac{1}{\|y-z_i\|^2}\right)^2.$$

It would be interesting to understand the complexity of this optimization problem, compared to the identifiability degree discussed in the previous paragraph. In particular, a starting point can be to study the ED degree of the variety parametrized by

$$\mathbb{C}^3 \times \mathbb{C}^3 \dashrightarrow \mathbb{C}^7, \quad (x, y) \mapsto \left(\tfrac{1}{Q(z_i-x)} + \tfrac{1}{Q(z_i-y)}\right)_{i=1}^{7}$$

for generic $z_1, \dots, z_7 \in \mathbb{C}^3$.

Finally, we note that throughout Paper G, we use objective functions based on residual sums of squares, which is also the case in the software package ChromSDE [149]. Other common tools such as PASTIS [32, 142] and ASHIC [147] instead use statistical likelihood functions. It would be interesting to systematically study the properties of these different objective functions, and in particular the number of critical points that they give rise to.



(a) $\varepsilon = 0.10$ **(b)** $\varepsilon = 0.50$ **(c)** $\varepsilon = 0.90$

Figure 6.3: Reconstruction of simulated data sets for varying noise levels. Figure from Paper G.

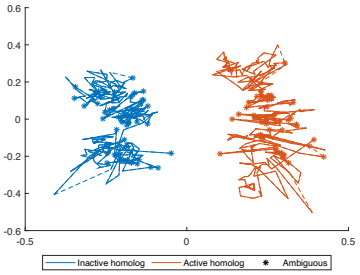
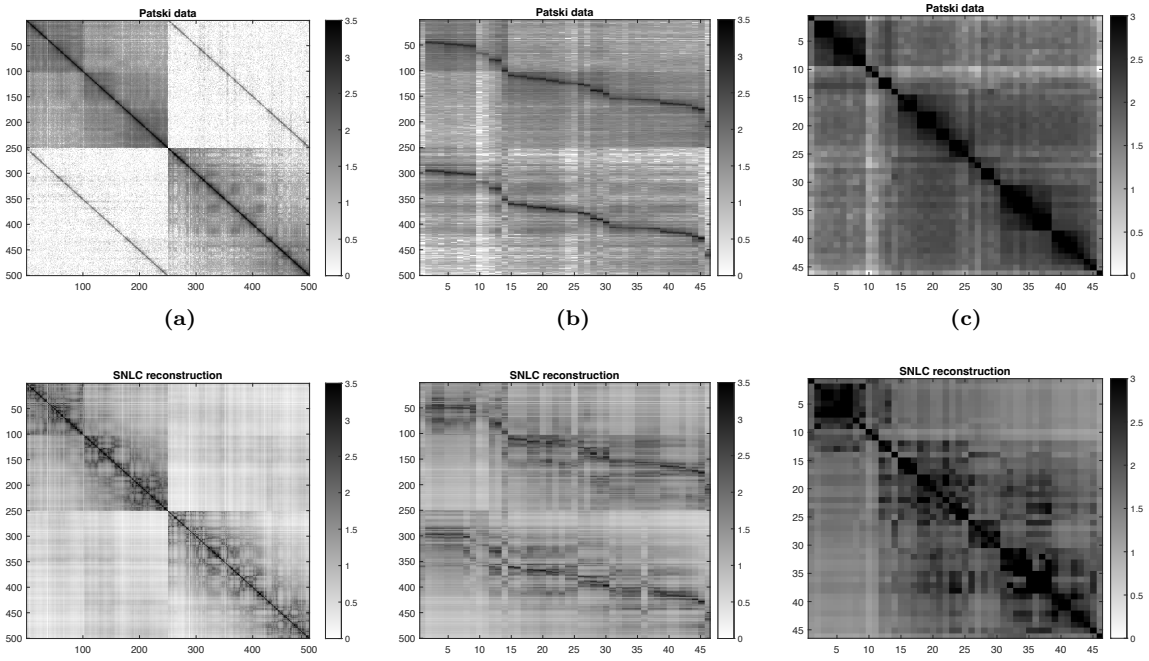


Figure 6.4: Reconstruction of mouse X chromosomes of the Patski cell line. Figure from Paper G.



(a) **(b)** **(c)**

(d) **(e)** **(f)**

Figure 6.5: Contact count matrices obtained from the original Patski dataset (after preprocessing), and from our reconstruction: (a,d): c^U; (b,e): c^P; (c,f): c^A. Figure from Paper G.

Bibliography

[1] D. Agostini, C. Améndola, and K. Ranestad. Moment identifiability of homoscedastic Gaussian mixtures. *Found. Comput. Math.*, 21(3):695–724, 2021.

[2] Y. Alexandr, J. Kileel, and B. Sturmfels. Moment varieties for mixtures of products. In *Proceedings of the International Symposium on Symbolic & Algebraic Computation (ISSAC 2023)*, pages 53–60. ACM, New York, 2023.

[3] C. Améndola, M. Drton, and B. Sturmfels. Maximum likelihood estimates for Gaussian mixtures are transcendental. In *Mathematical aspects of computer and information sciences*, volume 9582 of *Lecture Notes in Comput. Sci.*, pages 579–590. Springer, 2016.

[4] C. Améndola, J.-C. Faugère, and B. Sturmfels. Moment varieties of Gaussian mixtures. *J. Algebr. Stat.*, 7(1):14–28, 2016.

[5] C. Améndola, K. Ranestad, and B. Sturmfels. Algebraic identifiability of Gaussian mixtures. *Int. Math. Res. Not. IMRN*, (21):6556–6580, 2018.

[6] C. Améndola, J. Lindberg, and J. I. Rodriguez. Solving parameterized polynomial systems with decomposable projections, 2016. Preprint: `arXiv:1612.08807`.

[7] D. F. Anderson. A proof of the global attractor conjecture in the single linkage class case. *SIAM J. Appl. Math.*, 71(4):1487–1508, 2011.

[8] F. Ay, E. M. Bunnik, N. Varoquaux, S. M. Bol, J. Prudhomme, J.-P. Vert, W. S. Noble, and K. G. Le Roch. Three-dimensional modeling of the p. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.*, 24(6):974–988, 2014.

[9] M. Banaji and C. Pantea. The inheritance of nondegenerate multistationarity in chemical reaction networks. *SIAM J. Appl. Math.*, 78:1105–1130, 2018.

[10] K. Banecki, S. Korsak, and D. Plewczynski. Advancements and future directions in single-cell Hi-C based 3D chromatin modeling. *Comput. Struct. Biotechnol. J.*, 2024.

[11] D. J. Bates, P. Breiding, T. Chen, J. D. Hauenstein, A. Leykin, and F. Sottile.

Numerical nonlinear algebra, 2023. Preprint: `arXiv:2302.08585v2`.

[12] D. J. Bates, A. J. Sommese, J. D. Hauenstein, and C. W. Wampler. *Numerically solving polynomial systems with Bertini*. SIAM, 2013.

[13] M. Belkin and K. Sinha. Polynomial learning of distribution families. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 103–112. IEEE, 2010.

[14] A. Belyaeva, K. Kubjas, L. J. Sun, and C. Uhler. Identifying 3D genome organization in diploid organisms via Euclidean distance geometry. *SIAM J. Math. Data Sci.*, 4(1):204–228, 2022.

[15] M. Bender and P. J. Spaenlehauer. Dimension results for extremal-generic polynomial systems over complete toric varieties. *J. Algebra*, 646:0021–8693, 2024.

[16] A. Bernardi, E. Carlini, M. V. Catalisano, A. Gimigliano, and A. Oneto. The hitchhiker guide to: Secant varieties and tensor decomposition. *Mathematics*, 6(12):314, 2018.

[17] D. N. Bernstein. The number of roots of a system of equations. *Funct. Anal. Appl.*, 9:183–185, 1976.

[18] A. T. Blomenhofer. Gaussian mixture identifiability from degree 6 moments, 2023. Preprint: `arXiv:2307.03850`.

[19] A. T. Blomenhofer, A. Casarotti, M. Michałek, and A. Oneto. Identifiability for mixtures of centered Gaussians and sums of powers of quadratics. *Bull. Lond. Math. Soc.*, 55(5):2407–2424, 2023.

[20] J. Bochnak, M. Coste, and M. Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, 1998.

[21] T. Bogart, A. N. Jensen, D. Speyer, B. Sturmfels, and R. R. Thomas. Computing tropical varieties. *J. Symb. Comput.*, 42(1-2):54–73, 2007.

[22] B. Boros, G. Craciun, and P. Y. Yu. Weakly reversible mass-action systems with infinitely many positive steady states. *SIAM J. Appl. Math*, 80(4):1936–1946, 2020.

[23] V. Borovik, P. Breiding, J. del Pino, M. Michałek, and O. Zilberberg. Khovanskii bases for semimixed systems of polynomial equations – a case of approximating stationary nonlinear Newtonian dynamics, 2023. Preprint: `arXiv:2306.07897v1`.

[24] V. Borovik and P. Breiding. A short proof for the parameter continuation theorem. *J. Symb. Comput.*, 127:8, 2025. Id/No 102373.

[25] P. Breiding, M. Michałek, L. Monin, and S. Telen. The algebraic degree of coupled oscillators, 2022. Preprint: `arXiv:2208.08179v1`.

[26] P. Breiding, F. Sottile, and J. Woodcock. Euclidean distance degree and mixed volume.

Found. Comput. Math., 22(6):1743–1765, 2022.

[27] P. Breiding and S. Timme. HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.

[28] L. Brustenga i Moncusí, G. Craciun, and M.-S. Sorea. Disguised toric dynamical systems. *J. Pure Appl. Algebra*, 226(8):107035, 2022.

[29] M. Burr, F. Sottile, and E. Walker. Numerical homotopies from Khovanskii bases. *Math. Comput.*, 92(343):2333–2353, 2023.

[30] J. Capco, M. Gallet, G. Grasegger, C. Koutschan, N. Lubbes, and J. Schicho. The number of realizations of a laman graph. *SIAM J. Appl. Algebra Geom.*, 2(1):94–125, 2018.

[31] D. Cappelletti, E. Feliu, and C. Wiuf. Addition of flow reactions preserving multistationarity and bistability. *Math. Biosc.*, 320(108295), 2020.

[32] A. G. Cauer, G. Yardimci, J.-P. Vert, N. Varoquaux, and W. S. Noble. Inferring diploid 3D chromatin structures from Hi-C data. In *19th International Workshop on Algorithms in Bioinformatics (WABI 2019)*, 2019.

[33] L. Chiantini and C. Ciliberto. Weakly defective varieties. *Trans. Amer. Math. Soc.*, 354(1):151–178, 2002.

[34] L. Chiantini and C. Ciliberto. On the classification of defective threefolds. In *Projective varieties with unexpected properties*, pages 131–176. Walter de Gruyter, Berlin, 2005.

[35] B. L. Clarke. *Stability of Complex Reaction Networks*, pages 1–215. John Wiley & Sons, Ltd, 1980.

[36] E. Connelly, T. Duff, and J. Loucks-Tavitas. Algebra and geometry of camera resectioning. *Math. Comput.*, 2024.

[37] C. Conradi, E. Feliu, M. Mincheva, and C. Wiuf. Identifying parameter regions for multistationarity. *PLoS Comput. Biol.*, 13(10):e1005751, 2017.

[38] C. Conradi and T. Kahle. Detecting binomiality. *Adv. Appl. Math.*, 71:52–67, 2015.

[39] C. Conradi and C. Pantea. Multistationarity in biochemical networks: results, analysis, and examples. In *Algebraic and combinatorial computational biology*, pages 279–317. Elsevier, 2019.

[40] D. A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2015.

[41] G. Craciun, A. Dickenstein, A. Shiu, and B. Sturmfels. Toric dynamical systems. *J.*

Symbolic Comput., 44:1551–1565, 2009.

[42] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: extensions to entrapped species models. *Syst. Biol. (Stevenage)*, 153:179–186, 2006.

[43] J. Cruickshank, F. Mohammadi, A. Nixon, and S. Tanigawa. Identifiability of points and rigidity of hypergraphs under algebraic constraints, 2023. Preprint: `arXiv:2305.18990`.

[44] O. Daisey and Y. Ren. A framework for generalized tropical homotopy continuation. In K. Buzzard, A. Dickenstein, B. Eick, A. Leykin, and Y. Ren, editors, *Mathematical Software – ICMS 2024*, pages 331–339, Cham, 2024. Springer Nature Switzerland.

[45] D. F. Davidenko. On a new method of numerical solution of systems of nonlinear equations. *Doklady Akad. Nauk SSSR (N.S.)*, 88:601–602, 1953.

[46] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing chromosome conformation. *science*, 295(5558):1306–1311, 2002.

[47] X. Deng, W. Ma, V. Ramani, A. Hill, F. Yang, F. Ay, J. B. Berletch, C. A. Blau, J. Shendure, Z. Duan, et al. Bipartite structure of the inactive mouse X chromosome. *Genome Biol.*, 16(1):1–21, 2015.

[48] A. Deshpande and M. Gopalkrishnan. Autocatalysis in reaction networks. *Bull. Math. Biol.*, 76(10):2570–2595, 2014.

[49] S. Dewar, G. Grasegger, K. Kubjas, F. Mohammadi, and A. Nixon. Single-cell 3D genome reconstruction in the haploid setting using rigidity theory, 2024. Preprint: `arXiv:2407.10700`.

[50] A. Dickenstein. Biochemical reaction networks: an invitation for algebraic geometers. In *Mathematical Congress of the Americas*, volume 656 of *Contemp. Math.*, pages 65–83. Amer. Math. Soc., Providence, RI, 2016.

[51] A. Dickenstein, M. Pérez Millán, A. Shiu, and X. Tang. Multistationarity in structured reaction networks. *Bull. Math. Biol.*, 81(5):1527–1581, 2019.

[52] A. Dickenstein. Where algebra and geometry meet systems biology. *SIAM News*, 57(7):1, 3, 2024.

[53] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.*, 16(1):99–149, 2016.

[54] A. Esterov. Galois theory for general systems of polynomial equations. *Compos. Math.*, 155(2):229–245, 2019.

[55] A. Esterov. Engineered complete intersections: slightly degenerate Bernstein–Kouchnirenko–Khovanskii, 2024. Preprint: `arXiv:2401.12099`.

[56] M. Feinberg. Complex balancing in general kinetic systems. *Arch. Rational. Mech. Anal.*, 49(3):187, 1972.

[57] M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rational. Mech. Anal.*, 132(4):311, 1995.

[58] M. Feinberg. *Foundations of chemical reaction network theory*, volume 202 of *Applied Mathematical Sciences*. Springer, Cham, 2019.

[59] C. B. García and W. I. Zangwill. Finding all solutions to polynomial systems and other systems of equations. *Math. Programming*, 16(2):159–176, 1979.

[60] L. D. García Puente, M. Garrote-López, and E. Shehu. Computing algebraic degrees of phylogenetic varieties. *Algebr. Stat.*, 14(2):215–231, 2023.

[61] L. D. García Puente, E. Gross, H. A. Harrington, M. Johnston, N. Meshkat, M. Pérez Millán, and A. Shiu. Absolute concentration robustness: Algebra and geometry. *J. Symb. Comput.*, 128:102398, 2025.

[62] K. Gatermann. Counting stable solutions of sparse polynomial systems in chemistry. *Contemp. Math.*, 286:53–70, 2001.

[63] M. Gopalkrishnan. Catalysis in reaction networks. *Bull. Math. Biol.*, 73(12):2962–2982, 2011.

[64] M. Gopalkrishnan, E. Miller, and A. Shiu. A geometric approach to the global attractor conjecture. *SIAM J. Appl. Dyn. Syst.*, 13(2):758–797, 2014.

[65] P. Görlach, Y. Ren, and L. Zhang. Computing zero-dimensional tropical varieties via projections. *Comput. Complexity*, 31(1):5, 2022.

[66] D. Grigoriev, A. Iosif, H. Rahkooy, T. Sturm, and A. Weber. Efficiently and effectively recognizing toricity of steady state varieties. *Math. Comput. Sci.*, 15(2):199–232, 2020.

[67] A. Grosdos Koutsoumpelias. *Algebraic Methods for the Estimation of Statistical Distributions*. PhD thesis, Osnabrück University, 2020.

[68] A. Grosdos Koutsoumpelias and M. Wageringel. Moment ideals of local Dirac mixtures. *SIAM J. Appl. Algebra Geom.*, 4(1):1–27, 2020.

[69] E. Gross, H. A. Harrington, Z. Rosen, and B. Sturmfels. Algebraic systems biology: a case study for the Wnt pathway. *Bull. Math. Biol.*, 78(1):21–51, 2016.

[70] C. M. Guldberg and P. Waage. Über die chemische Affinität. *J. Prakt. Chem.*, 19:69–114, 1879.

[71] J. Gunawardena. Distributivity and processivity in multisite phosphorylation can be distinguished through steady-state invariants. *Biophys. J.*, 93(11):3828–3834,

2007.

[72] V. Gáspár and J. Tóth. Reaction extent or advancement of reaction: a definition for complex chemical reactions. *Chaos*, 33(4):22, 2023. Id/No 043141.

[73] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[74] H. A. Harrington, K. L. Ho, T. Thorne, and M. P. Stumpf. Parameter-free model discrimination criterion based on steady-state coplanarity. *Proc. Natl. Acad. Sci. U.S.A.*, 109(39):15746–15751, 2012.

[75] M. Helmer and B. Sturmfels. Nearest points on toric varieties. *Math. Scand.*, 122(2):213–238, 2018.

[76] P. A. Helminck and Y. Ren. Generic root counts and flatness in tropical geometry, 2022. Preprint: `arXiv:2206.07838v2`.

[77] I. Holt and Y. Ren. Generic root counts of tropically transverse systems – an invitation to tropical geometry in OSCAR, 2023.

[78] F. Horn and R. Jackson. General mass action kinetics. *Arch. Rational Mech. Anal.*, 47:81–116, 1972.

[79] M. Hu, K. Deng, Z. Qin, J. Dixon, S. Selvaraj, J. Fang, B. Ren, and J. S. Liu. Bayesian inference of spatial organizations of chromosomes. *PLoS Comput. Biol.*, 9(1):e1002893, 2013.

[80] B. Huber and B. Sturmfels. A polyhedral method for solving sparse polynomial systems. *Math. Comput.*, 64(212):1541–1555, 1995.

[81] V. Hárs and J. Tóth. On the inverse problem of reaction kinetics. In *Colloquia Mathematica Societatis János Bolyai 30., Qualitative Theory of Differential Equations, Szeged (Hungary)*, pages 363–379, 1979.

[82] A. N. Jensen. Tropical homotopy continuation, 2016. Preprint: `arXiv:1601.02818v1`.

[83] Y. Jiao, X. Tang, and X. Zeng. Multistability of small zero-one reaction networks, 2024. Preprint: `arXiv:2406.11586`.

[84] B. Joshi and A. Shiu. Atoms of multistationarity in chemical reaction networks. *J. Math. Chem.*, 51(1):153–178, 2013.

[85] B. Joshi and A. Shiu. Which small reaction networks are multistationary? *SIAM J. Appl. Dyn. Syst.*, 16(2):802–833, 2017.

[86] T. Kahle and J. Vill. Efficiently deciding if an ideal is toric after a linear coordinate change, 2024. Preprint: `arXiv:2408.14323`.

[87] K. Kaveh and A. G. Khovanskii. Mixed volume and an extension of intersection

theory of divisors. *Mosc. Math. J.*, 10(2):343–375, 479, 2010.

[88] K. Kaveh and A. G. Khovanskii. Newton-Okounkov bodies, semigroups of integral points, graded algebras and intersection theory. *Ann. of Math. (2)*, 176(2):925–978, 2012.

[89] A. G. Khovanskii. Newton polyhedra and irreducible components of complete intersections. *Izv. Math.*, 80(1):263–284, 2016.

[90] A. G. Khovanskii. Newton polyhedra and the genus of complete intersections. *Funct. Anal. Appl.*, 12(1):38–46, 1978.

[91] K. Kohn, B. Shapiro, and B. Sturmfels. Moment varieties of measures on polytopes. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 21:739–770, 2020.

[92] A. G. Kouchnirenko. Polyèdres de Newton et nombres de Milnor. *Invent. Math.*, 32(1):1–31, 1976.

[93] N. Krislock. *Semidefinite Facial Reduction for Low-Rank Euclidean Distance Matrix Completion*. PhD thesis, University of Waterloo, 2010.

[94] K. Kubjas, J. Li, and M. Wiesmann. Geometry of polynomial neural networks. *Algebr. Stat.*, 15(2):295–328, 2024.

[95] D. L. Lafontaine, L. Yang, J. Dekker, and J. H. Gibcus. Hi-C 3.0: Improved protocol for genome-wide chromosome conformation capture. *Curr. Protoc.*, 1(7):e198, 2021.

[96] M. Laurent and N. Kellershohn. Multistability: a major means of differentiation and evolution in biological systems. *Trends Biochem. Sciences*, 24(11):418–422, 1999.

[97] D. Lazard. Injectivity of real rational mappings: the case of a mixture of two gaussian laws. *Math. Comput. Simul.*, 67(1-2):67–84, 2004.

[98] C.-S. Lee, R. W. Wang, H.-H. Chang, D. Capurso, M. R. Segal, and J. E. Haber. Chromosome position determines the success of double-strand break repair. *Proc. Natl. Acad. Sci. U.S.A.*, 113(2):E146–E154, 2016.

[99] A. Lesne, J. Riposo, P. Roger, A. Cournac, and J. Mozziconacci. 3D genome reconstruction from chromosomal contacts. *Nat. methods*, 11(11):1141–1143, 2014.

[100] A. Leykin and J. Yu. Beyond polyhedral homotopies. *J. Symbolic Comput.*, 91:173–180, 2019. MEGA 2017, Effective Methods in Algebraic Geometry, Nice (France), June 12-16, 2017.

[101] T. Li and X. Wang. The BKK root count in $\mathbb{C}^n$. *Math. Comput.*, 65(216):1477–1484, 1996.

[102] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy,

A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009.

[103] J. Lindberg, L. Monin, and K. Rose. The algebraic degree of sparse polynomial optimization, 2023. Preprint: `arXiv:2308.07765v2`.

[104] J. Lindberg, C. Améndola, and J. I. Rodriguez. Estimating Gaussian mixtures using sparse polynomial moment systems, 2021. Preprint: `arXiv:2106.15675v3`.

[105] C. Lüders, T. Sturm, and O. Radulescu. Odebase: a repository of ode systems for systems biology. *Bioinf. Adv.*, 2(1), 2022.

[106] D. Maclagan and B. Sturmfels. *Introduction to tropical geometry*, volume 161 of *Grad. Stud. Math.* Providence, RI: American Mathematical Society (AMS), 2015.

[107] A. L. MacLean, Z. Rosen, H. M. Byrne, and H. A. Harrington. Parameter-free methods distinguish Wnt pathway models and guide design of experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 112(9):2652–2657, 2015.

[108] A. K. Manrai and J. Gunawardena. The geometry of multisite phosphorylation. *Biophys. J.*, 95(12):5533–5543, 2008.

[109] A. Maraj and A. Pal. Symmetry Lie algebras of varieties with applications to algebraic statistics, 2023.

[110] M. Marcondes de Freitas, E. Feliu, and C. Wiuf. Intermediates, catalysts, persistence, and boundary steady states. *J. Math. Biol.*, 74(4):887–932, 2017.

[111] A. Massarenti and M. Mella. Bronowski's conjecture and the identifiability of projective varieties, 2022. Preprint: `arXiv:2210.13524v3`.

[112] L. Maxim, J. I. Rodriguez, and B. Wang. Applications of singularity theory in applied algebraic geometry and algebraic statistics, 2023. Preprint: `arXiv:2305.19842`.

[113] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake. Finite mixture models. *Annu. Rev. Stat. Appl.*, 6(1):355–378, 2019.

[114] N. Meshkat, A. Shiu, and A. Torres. Absolute concentration robustness in networks with low-dimensional stoichiometric subspace. *Vietnam J. Math.*, 50:623–651, 2022.

[115] M. P. Millán. *Métodos algebraicos para el estudio de redes bioquímicas*. PhD thesis, Universidad de Buenos Aires, 2011. Available at `https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis_n5103_PerezMillan.pdf`.

[116] M. P. Millán, A. Dickenstein, A. Shiu, and C. Conradi. Chemical reaction systems with toric steady states. *Bull. Math. Biol.*, 74:1027–1065, 2012.

[117] D. Mumford. *Algebraic Geometry I, Complex Projective Varieties*. Classics in Mathematics. Springer Verlag, 1976.

[118] S. Müller, E. Feliu, G. Regensburger, C. Conradi, A. Shiu, and A. Dickenstein. Sign conditions for injectivity of generalized polynomial maps with applications to chemical reaction networks and real algebraic geometry. *Found. Comput. Math.*, 16(1):69—-97, 2015.

[119] N. K. Obatake and E. Walker. Newton-Okounkov bodies of chemical reaction systems. *Adv. in Appl. Math.*, 155:Paper No. 102672, 27, 2024.

[120] O. Oluwadare, M. Highsmith, and J. Cheng. An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. *Biol. Proced. Online*, 21(1):1–20, 2019.

[121] The OSCAR Team. *OSCAR: Open Source Computer Algebra Research system*. Available at https://www.oscar-system.org.

[122] G. Ottaviani and E. Teixeira Turatti. Generalized identifiability of sums of squares. *J. Algebra*, 661:641–656, 2025.

[123] B. Pascual-Escudero and E. Feliu. Local and global robustness at steady state. *Math. Methods Appl. Sci.*, 45(1):359–382, 2022.

[124] J. Paulsen, M. Sekelja, A. R. Oldenburg, A. Barateau, N. Briand, E. Delbarre, A. Shah, A. L. Sørensen, C. Vigouroux, B. Buendia, et al. Chrom3D: Three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol.*, 18(1):1–15, 2017.

[125] K. Pearson. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond., Ser. A, Contain. Pap. Math. Phys. Character*, 187:253–318, 1896.

[126] H. Rahkooy and T. Sturm. Parametric toricity of steady state varieties of reaction networks. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing*, pages 314–333, Cham, 2021. Springer International Publishing.

[127] H. Rahkooy and T. Sturm. Testing binomiality of chemical reaction networks using comprehensive Gröbner systems. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing*, pages 334–352, Cham, 2021. Springer International Publishing.

[128] W. C. Rheinboldt. *Methods for solving systems of nonlinear equations*, volume 70 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1998.

[129] J. M. Rojas and X. Wang. Counting affine roots of polynomial systems via pointed

Newton polytopes. *J. Complexity*, 12(2):116–133, 1996.

[130] M. Rousseau, J. Fraser, M. A. Ferraiuolo, J. Dostie, and M. Blanchette. Three-dimensional modeling of chromatin structure from interaction frequency data using markov chain monte carlo sampling. *Bioinform.*, 12(1):414, 2011.

[131] A. Sadeghimanesh and E. Feliu. The multistationarity structure of networks with intermediates and a binomial core network. *Bull. Math. Biol.*, 81:2428–2462, 2019.

[132] M. R. Segal. Can 3D diploid genome reconstruction from unphased Hi-C data be salvaged? *NAR Genom. Bioinform.*, 4(2):lqac038, 2022.

[133] G. Shinar and M. Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327:1389–1391, 2010.

[134] A. Shiu and T. de Wolff. Nondegenerate multistationarity in small reaction networks. *Discrete Contin. Dyn. Syst. Ser. B*, 24(6):2683–2700, 2019.

[135] A. Shiu and B. Sturmfels. Siphons in chemical reaction networks. *Bull. Math. Biol.*, 72(6):1448–1463, 2010.

[136] A. J. Sommese and C. W. I. Wampler. *The numerical solution of systems of polynomials. Arising in engineering and science.* River Edge, NJ: World Scientific, 2005.

[137] B. Sturmfels. On the Newton polytope of the resultant. *J. Algebraic Combin.*, 3(2):207–236, 1994.

[138] X. Tang and Z. Zhang. Multistability of reaction networks with one-dimensional stoichiometric subspaces. *SIAM J. Appl. Dyn. Syst.*, 21(2):1426–1454, 2022.

[139] A. Terracini. Su due problemi concernenti la determinazione di alcune classi di superficie, considerati da G. Scorza e da J. Palatini. *Atti Soc. Natur. e Matem. Modena*, 5(6), 1921.

[140] A. Torres and E. Feliu. Symbolic proof of bistability in reaction networks. *SIAM J. Appl. Dyn. Syst.*, 20(1):1–37, 2021.

[141] C. Uhler and G. Shivashankar. Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nat. Rev. Mol. Cell Biol.*, 18(12):717–727, 2017.

[142] N. Varoquaux, F. Ay, W. S. Noble, and J.-P. Vert. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics*, 30(12):i26–i33, 2014.

[143] J. Verschelde. Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. Math. Softw.*, 25(2):251–276, 1999.

[144] E. O. Voit, H. A. Martens, and S. W. Omholt. 150 years of the mass action law.

PLoS Comput. Biol., 11(1):e1004012, 2015.

[145] H. Wang, X. Xu, C. M. Nguyen, Y. Liu, Y. Gao, X. Lin, T. Daley, N. H. Kipniss, M. La Russa, and L. S. Qi. CRISPR-mediated programmable 3D genome positioning and nuclear organization. *Cell*, 175(5):1405–1417, 2018.

[146] L. Wasserman. *All of statistics: a concise course in statistical inference*. Springer, 2004.

[147] T. Ye and W. Ma. ASHIC: Hierarchical Bayesian modeling of diploid chromatin contacts and structures. *Nucleic Acids Res.*, 48(21):e123–e123, 2020.

[148] J. Yu. Do most polynomials generate a prime ideal? *J. Algebra*, 459:468–474, 2016.

[149] Z. Zhang, G. Li, K.-C. Toh, and W.-K. Sung. Inference of spatial organizations of chromosomes using semi-definite embedding approach and Hi-C data. In *Annual international conference on research in computational molecular biology*, pages 317–332. Springer, 2013.

[150] B. Ådlandsvik. Joins and higher secant varieties. *Math. Scand.*, 61(2):213–222, 1987.

Papers

A

A tropical method for solving parametrized polynomial systems

Paul Alexander Helminck
Mathematical Institute
Tohoku University

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Yue Ren
Department of Mathematical Sciences
Durham University

Publication details

Preprint: https://doi.org/10.48550/arXiv.2409.13288 (2024)

A TROPICAL METHOD FOR SOLVING PARAMETRIZED POLYNOMIAL SYSTEMS

PAUL ALEXANDER HELMINCK, OSKAR HENRIKSSON, AND YUE REN

Abstract. We give a framework for constructing generically optimal homotopies for parametrized polynomial systems from tropical data. Here, generically optimal means that the number of paths tracked is equal to the generic number of solutions. We focus on two types of parametrized systems – vertically parametrized and horizontally parametrized systems – and discuss techniques for computing the tropical data efficiently. We end the paper with several case studies, where we analyze systems arising from chemical reaction networks, coupled oscillators, and rigid graphs.

1. Introduction

Solving systems of polynomial equations is a fundamental task throughout applied mathematics; for instance, polynomials govern the motion of robots [SW05], the phase of coupled oscillators [CMMN19], and the concentrations of species in biological systems [Dic16]. A staple for solving polynomial systems numerically over the complex numbers is *homotopy continuation* [BBC+23], which traces the solutions of an easy-to-solve start system to the desired solutions of the target system along a path in the space of polynomial systems, commonly called a *homotopy*.

Homotopy continuation is known for being able to compute a single solution to a polynomial system in average polynomial time, thereby answering Smale's 17th problem in the positive [BP11; BC11; Lai17]. However, constructing homotopies for computing all solutions to a polynomial system remains a major challenge.

Ideally, a homotopy should be both fast to construct and *optimal* in the sense that the number of paths equals the number of solutions of the target system. However, constructing optimal homotopies requires a priori knowledge of the number of solutions of the target system, which is difficult to obtain efficiently. This means that in practice, there is a tradeoff between the speed of construction and minimizing the number of superfluous paths.

For instance, computing the upper bound on the number of solutions given by Bézout's theorem is fast, as it only requires multiplying degrees. The (normalized) mixed volume bound given by Bernstein's theorem [Ber76], on the other hand, will often be lower than the Bézout bound for sparse systems, but is substantially harder to compute. The actual number of solutions can, in principle, be computed through a Gröbner basis computation, but in this case, homotopy continuation becomes

irrelevant since there are more effective solvers available if given a Gröbner basis, such as eigenvalue solvers (see [CLO05, Chapter 2] for an introduction, and [Cox20, Section 2.1] for an overview of recent progress).

The mixed volume is known to be a sharp bound provided that the coefficients of the system are generic, but for many parametrized systems that arise in applications, relations among the coefficients lead to fewer than mixed volume many solutions (see, e.g., [GHRS16] and [BBP+23]). This has prompted a wealth of works on finding upper bounds on the number of solutions under weaker genericity assumptions, involving techniques such as tropical geometry [HR22; HR23], Khovanskii bases and Newton–Okounkov bodies [KK12; OW24; BBP+23], and toric geometry and facial subsystems [BKK+21; BSW22; LMR23].

Given a bound on the number of solutions, it is a problem in its own right to construct a homotopy (or a collection of homotopies) that trace precisely that many paths. For Bernstein's mixed volume bound, this problem was solved by Huber and Sturmfels in their seminal paper [HS95] in the form of a construction called *polyhedral homotopies* (see also [VVC94; Li99]). Since their introduction, polyhedral homotopies have become a popular default strategy in several systems such as HomotopyContinuation.jl [BT18], PHCpack [Ver99], and Hom4PS [CLL14]. Recent work on constructing better homotopies include [LY19] (using tropical geometry), [BSW23] (using Khovanskii bases), and [DTWY24] (using toric geometry).

In this paper, we propose a generalization of polyhedral homotopies for constructing homotopies that realize the tropical root bounds from [HR22; HR23], building on and extending ideas from Leykin and Yu [LY19].

Contents of the paper and links to the existing literature. Section 2 goes through the necessary theoretical background on parametrized polynomial systems and tropical geometry.

The main goal of Section 3 is to describe a natural generalization of polyhedral homotopies. Algorithm 3.1 describes how to construct generically optimal start systems and homotopies for solving a parametrized (Laurent) polynomial system

$$\mathcal{F} = \{f_1, \dots, f_n\} \subseteq \mathbb{C}[a_1, \dots, a_m][x_1^{\pm}, \dots, x_n^{\pm}]$$

for a generic choice $P \in (\mathbb{C}^*)^m$ of parameters, using the following tropical data:

(1) The zero-dimensional tropicalization $\operatorname{Trop}(\langle \mathcal{F}_Q \rangle)$, where $\mathcal{F}_Q$ is the system specialized at a perturbation $Q := (t^{v_1}P_1, \dots, t^{v_m}P_m) \in \mathbb{C}\{\{t\}\}^m$ of the parameters, for generic exponents $v \in \mathbb{Q}^m$.

(2) The zeros of the initial ideals $V(\operatorname{in}_w(\langle \mathcal{F}_Q \rangle)) \subseteq (\mathbb{C}^*)^n$ for all $w \in \operatorname{Trop}(\langle \mathcal{F}_Q \rangle)$.

Algorithm 3.1 builds on the same connection between tropical geometry and polynomial system solving that polyhedral homotopy rests on: Consider $\mathcal{F}_Q$ as a one-parameter family of polynomial systems with parameter t and variables x, whose specialization at $t = 1$ equals a given system to be solved. By the Newton–Puiseux theorem, solutions around $t = 0$ are parametrized by Puiseux series, and consequently, their convergence or divergence at $t = 0$ is governed by their coordinatewise valuations. In [HS95], the coordinatewise valuations are computed using mixed cells, whereas in our work, they will be computed from tropical stable intersections.

The objective of Sections 4 and 5 is to explain how Algorithm 3.1 can be used to obtain optimal homotopies efficiently for two types of parametrized polynomial systems:

Section 4 considers *vertically parametrized* polynomial systems and describes how to obtain their tropical data efficiently. Vertical systems arise for example from chemical reaction networks, see [Dic16] for a general introduction and [FHP23] for a writeup closer to the language of this article. Similar systems also arise from Lagrangian systems in polynomial optimization such as the maximum likelihood estimations for log-linear models, where tropical techniques have been applied in [BDH24]. We employ the idea in [HR22, Section 6.1], explain how the tropical data above can be computed from the intersection of a tropical linear space and a tropical binomial variety, and discuss the computational challenges involved in doing so.

Section 5 considers *horizontally parametrized* polynomial systems as in the works of Kaveh and Khovanskii [KK12]. Obtaining the required tropical data for horizontal systems is a highly non-trivial task, but we discuss two techniques that cover many systems that arise in practice:

(1) Identifying a *tropically transverse base* for the polynomial support (as in [HR22, Example 6.12] and [HR23]). This is the topic of Section 5.2.
(2) Embedding the family into a larger family, by introducing parameters into the polynomial support. This might make the resulting root bound larger, while still being an improvement compared to the Bernstein bound. This is the topic of Section 5.3.

Both these techniques result in new, Bernstein generic systems, in such a way that an adapted version of polyhedral homotopies can be used.

The purpose of Section 6 is to demonstrate that our techniques can be applied to several examples from the existing literature:

(1) In Section 6.1, we examine steady state equations of the WNT pathway [GHRS16] by relaxing it to a vertically parametrized system.
(2) In Section 6.2, we regard the equations for Duffing oscillators [BBP+23] as horizontally parametrized systems with transverse base.

(3) In Section 6.3, we consider the Kuramoto equations with phase delays [CKL22] as relaxed horizontally parametrized systems.

(4) In Section 6.4, we study the realizations of a Laman graph [CGG+18].

In Section 7, we summarize our results and outline future research direction.

A Julia implementation of our algorithm based on OSCAR [OSCAR] and HomotopyContinuation.jl [BT18], as well as code for the examples appearing in the paper, can be found in the repository

https://github.com/oskarhenriksson/TropicalHomotopies.jl .

Acknowledgments. The authors would like to thank Elisenda Feliu, Máté L. Telek, and Benjamin Schröter for their involvement in the early phases of this project, as well as for helpful comments and discussions. The authors also thank Paul Breiding and Sascha Timme for help with HomotopyContinuation.jl.

Paul Helminck was supported by the UKRI Future Leaders Fellowship "Tropical Geometry and its applications" (MR/S034463/2), and is supported by a JSPS Postdoctoral Fellowship (Grant No. 23769) and KAKENHI 23KF0187. Oskar Henriksson is partially supported by the Novo Nordisk project (NNF20OC0065582), as well as the European Union under the Grant Agreement no. 101044561, POSALG. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or European Research Council (ERC). Neither the European Union nor ERC can be held responsible for them. Yue Ren is supported by the UKRI Future Leaders Fellowship "Tropical Geometry and its applications" (MR/S034463/2 & MR/Y003888/1).

2. Background

In this section we briefly recall some basic concepts of parametrized polynomial systems and tropical geometry that are of immediate interest to us.

Convention 2.1. For the remainder of the article, we fix:

(1) An algebraically closed field K of characteristic 0 with a possibly trivial valuation $\mathrm{val}\colon K^* \to \mathbb{R}$, and residue field $\mathfrak{K}$. By [MS15, Lemma 2.1.15] there exists a splitting $\mathrm{val}(K^*) \to K^*$, which we assume to be fixed. For example:

   (a) For the field of complex numbers $K = \mathbb{C}$ and the trivial valuation $\mathrm{val}\colon \mathbb{C}^* \to \mathbb{R}$, a possible splitting is the map $\mathrm{val}(\mathbb{C}^*) = \{0\} \to \mathbb{C}^*$, $0 \mapsto 1$,

   (b) For the field of complex Puiseux series $K = \mathbb{C}\{\{t\}\}$ and the usual valuation $\mathrm{val}\colon \mathbb{C}\{\{t\}\}^* \to \mathbb{R}$, a possible splitting is $\mathrm{val}(K^*) = \mathbb{Q} \to K^*, \lambda \mapsto t^\lambda$.

   Following the notation of [MS15], we denote the image of $\lambda \in \mathrm{val}(K^*)$ under the splitting as t^λ.

(2) An m-dimensional affine space K^m with coordinate ring $K[a] \coloneqq K[a_1, \ldots, a_m]$ and field of fractions $K(a) \coloneqq K(a_1, \ldots, a_m)$. We refer to the a as *parameters*, K^m as the *parameter space* and points $P \in K^m$ as *choices of parameters*.

(3) An n-dimensional torus $(K^*)^n$ with coordinate ring $K[x^{\pm}] \coloneqq K[x_1^{\pm}, \ldots, x_n^{\pm}]$. We refer to the x as *variables*.

(4) We write $K^m \times (K^*)^n$ for the product variety with coordinate ring $K[a][x^{\pm}] \coloneqq K[a_1, \ldots, a_m][x_1^{\pm}, \ldots, x_n^{\pm}]$. We refer to elements $f \in K[a][x^{\pm}]$ as *parametrized (Laurent) polynomials*, ideals $\mathcal{I} \subseteq K[a][x^{\pm}]$ as *parametrized (Laurent) polynomial ideals*, and finite sets $\{f_1, \ldots, f_k\} \subseteq K[a][x^{\pm}]$ as *parametrized (Laurent) polynomial systems*.

2.1. **Parametrized polynomial systems.** In this section we recall some basic concepts of parametrized polynomial systems over algebraically closed fields.

Definition 2.2. Let $f \in K[a][x^{\pm}]$ be a parametrized polynomial, say $f = \sum_{\alpha \in \mathbb{Z}^n} c_\alpha x^\alpha$ with $c_\alpha \in K[a]$ and $x^\alpha \coloneqq x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n$. For any choice of parameters $P \in K^m$ we define the *specialization* of f at P to be

$$f_P \coloneqq \sum_{\alpha \in \mathbb{Z}^n} c_\alpha(P) x^\alpha \in K[x^{\pm}].$$

Similarly, for a parametrized polynomial ideal $\mathcal{I} \subseteq K[a][x^{\pm}]$ and $P \in K^m$, we define the *specialization* of $\mathcal{I}$ at P to be

$$\mathcal{I}_P \coloneqq \langle f_P : f \in I \rangle \subseteq K[x^{\pm}].$$

Let $B_P = K[x^{\pm}]/\mathcal{I}_P$ be its coordinate ring. The *root count* of $\mathcal{I}$ at P is the vector space dimension $\ell_{\mathcal{I},P} \coloneqq \dim_K(B_P) \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$.

Remark 2.3. The integer $\ell_{\mathcal{I},P}$ is the number of points in the variety $V(\mathcal{I}_P) \subseteq (K^*)^n$ counted with a suitable multiplicity [CLO05, Corollary 2.5]. In particular, if $\mathcal{I}_P$ is zero-dimensional and radical, we have $\ell_{\mathcal{I},P} = |V(\mathcal{I}_P)|$ [CLO05, Corollary 2.6].

Definition 2.4. The *generic specialization* of a parametrized polynomial ideal $\mathcal{I} \subseteq K[a][x^{\pm}]$ is the ideal $\mathcal{I}_{K(a)} \subseteq K(a)[x^{\pm}]$ generated by $\mathcal{I}$ under the inclusion $K[a][x^{\pm}] \subseteq K(a)[x^{\pm}]$. The quotient ring $B_{K(a)} \coloneqq K(a)[x^{\pm}]/\mathcal{I}_{K(a)}$ is a vector space over the field $K(a)$, and the *generic root count* of $\mathcal{I}$ is its vector space dimension

$$\ell_{\mathcal{I},K(a)} \coloneqq \dim_{K(a)}(B_{K(a)}) \in \mathbb{Z}_{\geq 0} \cup \{\infty\}.$$

The *generic dimension* of $\mathcal{I}$ is the Krull dimension of $B_{K(a)}$. If the generic dimension is zero (equivalently, if $\ell_I < \infty$), we say that $\mathcal{I}$ is *generically zero-dimensional*. We say that $\mathcal{I}$ is *generically a complete intersection* if $B_{K(a)} \cong K(a)[z_1, ..., z_r]/\langle f_1, ..., f_k \rangle$ for polynomials $f_1, \ldots, f_k \in K(a)[z_1, \ldots, z_r]$ and $\dim(B_{K(a)}) = r - k$.

The *generic root count* and *generic dimension* of a parametrized polynomial system $\mathcal{F} \subseteq K[a][x^{\pm}]$ are the generic root count and generic dimension of the

parametrized ideal it generates. We say that $\mathcal{F}$ is *generically zero-dimensional* if the parametrized ideal it generates is generically zero-dimensional.

Remark 2.5.

(1) The properties in Definition 2.4 are generic in the sense that they reflect the behavior over a Zariski-dense open subset of K^m. For example, if $\mathcal{I}$ is generically zero-dimensional, then there is a Zariski-dense open subset $U \subseteq K^m$ such that $\ell_{\mathcal{I},P} = \ell_{\mathcal{I},K(a)}$ for all $P \in U$ [HR22, Remark 2.4].

(2) The generic root count $\ell_{\mathcal{I},K(a)}$ is invariant under field extensions. In particular, if $\mathcal{I} \subseteq \mathbb{C}[a][x^{\pm}]$ is a parametrized polynomial ideal over the complex numbers, and $\tilde{\mathcal{I}} = \langle \mathcal{I} \rangle \subseteq \mathbb{C}\{\{t\}\}[a][x^{\pm}]$ is the parametrized polynomial ideal over the complex Puiseux series that is generated by the elements of $\mathcal{I}$, then their generic root counts coincide, $\ell_{\mathcal{I},\mathbb{C}(a)} = \ell_{\tilde{\mathcal{I}},\mathbb{C}\{\{t\}\}(a)}$.

Consequently, from Section 3 onward, we may consider all parametrized polynomial systems over $\mathbb{C}$ as parametrized polynomial systems over $\mathbb{C}\{\{t\}\}$ without changing their generic root count.

We will close this subsection with the known result that embedding parametrized polynomial systems can only raise the generic root count. This is relevant for Section 5, where we embed a difficult family of polynomial system into an easier larger family of polynomial systems. Moreover, it implies that the generic root count is an upper bound in the sense that $\ell_{\mathcal{I},P} \leq \ell_{\mathcal{I},K(a)}$ for any $P \in K^m$ where $\ell_{\mathcal{I},P}$ is finite, provided $\mathcal{I}$ is generically zero-dimensional and a complete intersection.

Definition 2.6. Let $K[a][x^{\pm}] \coloneqq K[a_1,\dots,a_m][x_1^{\pm},\dots,x_n^{\pm}]$ and $K[b][x^{\pm}] \coloneqq K[b_1,\dots,b_l][x_1^{\pm},\dots,x_n^{\pm}]$ be two parametrized polynomial rings with the same variables x but different parameters a and b. Let $\mathcal{I}_1 \subseteq K[a][x^{\pm}]$ and $\mathcal{I}_2 \subseteq K[b][x^{\pm}]$ be two parametrized ideals. We say $\mathcal{I}_1$ is *embedded* in $\mathcal{I}_2$, if there is a ring homomorphism $K[b] \to K[a]$ such that $\mathcal{I}_1$ is the ideal generated by the image of $\mathcal{I}_2$ under the induced ring homomorphism $K[b][x^{\pm}] \to K[a][x^{\pm}]$. Similarly, we call a system $\mathcal{F}_1 \subseteq K[a][x^{\pm}]$ *embedded* in a system $\mathcal{F}_2 \subseteq K[b][x^{\pm}]$, if $\mathcal{F}_1$ is the image of $\mathcal{F}_2$ under the induced ring homomorphism.

Proposition 2.7. *Let $\mathcal{I}_1 \subseteq K[a][x^{\pm}]$ and $\mathcal{I}_2 \subseteq K[b][x^{\pm}]$ be two generic complete intersections. If $\mathcal{I}_1$ is embedded in $\mathcal{I}_2$, then $\ell_{\mathcal{I}_1} \leq \ell_{\mathcal{I}_2}$.*

Proof. Follows from [HR22, Lemma 5.2]. □

Example 2.8. Consider the polynomial system $F \coloneqq \{f_1, f_2\} \subseteq \mathbb{C}[x_1^{\pm}, x_2^{\pm}]$ given by

$$f_1 = x^2 + y^2 + x + y + 1 \quad \text{and} \quad f_2 = 3x^2 + 3y^2 + 5x + 7y + 11.$$

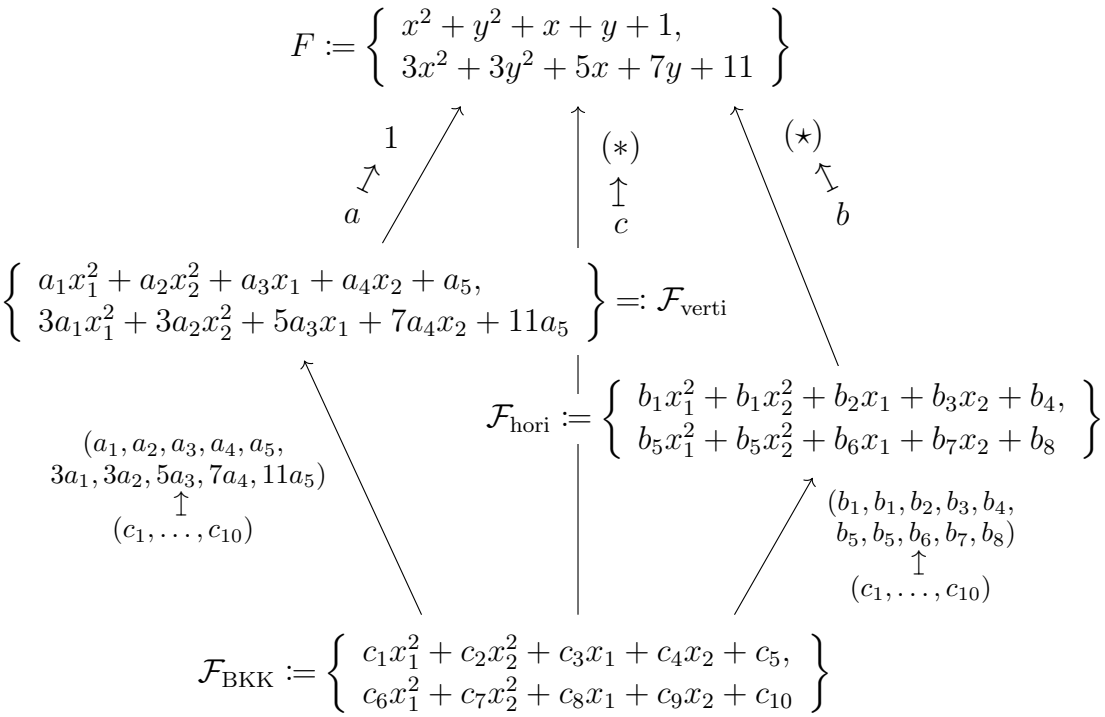


FIGURE 1. Three embeddings of the system F, where $(*)$ and $(\star)$ are suitable choices of parameters in $\mathbb{C}^{10}$ and $\mathbb{C}^8$, respectively.

The polynomials define two ellipses in the complex torus $(\mathbb{C}^*)^2$ intersecting in two points. Figure 1 shows three different parametrized polynomial systems it can be embedded to. By Bernstein's Theorem, the generic root counts of $\mathcal{F}_{\text{BKK}}$ is the mixed volume of the Newton polytopes, which is 4. One can show that the generic root counts of both $\mathcal{F}_{\text{verti}}$ and $\mathcal{F}_{\text{hori}}$ are 2. The systems $\mathcal{F}_{\text{verti}}$ and $\mathcal{F}_{\text{hori}}$ are examples of *vertically* and *horizontally parametrized system*, which are discussed in Section 4 and Section 5, respectively.

2.2. **Tropical geometry.** For tropical geometry, we will follow the notation of [MS15] as closely as possible with one key difference: we tropicalize ideals instead of varieties as we are not only interested in solutions of polynomial systems, but also their multiplicity. The resulting tropical varieties will be balanced polyhedral complexes instead of supports thereof. Our definition of tropical varieties will therefore rely on some of the results in [MS15, Sections 3.3 and 3.4].

Definition 2.9. Let $\mathfrak{K}$ be the residue field of K ($\mathfrak{K} = K$ if the valuation is trivial). The *initial form* of a polynomial $f \in K[x^{\pm}]$, say $f = \sum_{\alpha \in S} c_\alpha x^\alpha$ with support $S \subseteq \mathbb{Z}^n$ and coefficients $c_\alpha \in K^*$, with respect to a weight vector $w \in \mathbb{R}^n$ is given by

$$\text{in}_w(f) := \sum_{\substack{\alpha \in S \text{ with} \\ \text{val}(c_\alpha)+w\cdot\alpha \text{ minimal}}} \overline{t^{-\text{val}(c_\alpha)} c_\alpha}\, x^\alpha \in \mathfrak{K}[x^{\pm}].$$

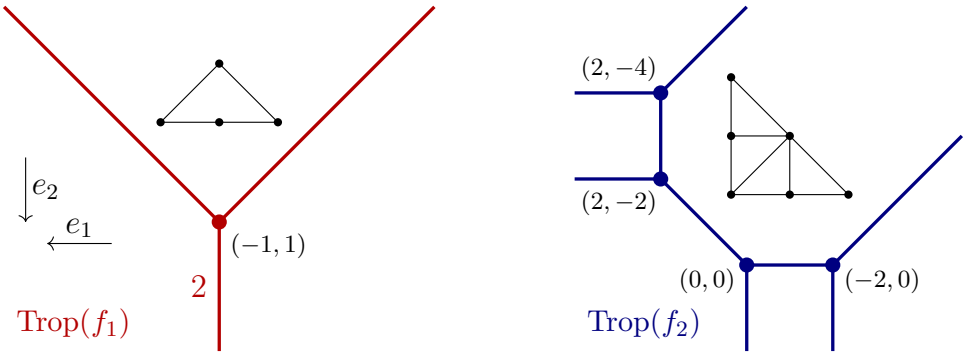


FIGURE 2. Tropical hypersurfaces and Newton polytopes of the two polynomials from Example 2.11.

The *initial ideal* of an ideal $I \subseteq K[x^{\pm}]$ with respect to $w \in \mathbb{R}^n$ is given by

$$\mathrm{in}_w(I) \coloneqq \langle\, \mathrm{in}_w(f) \mid f \in I \,\rangle \subseteq \mathfrak{K}[x^{\pm}].$$

Defining tropical varieties as a subcomplex of the Gröbner complex of the homogenized ideal is quite technical. We will omit the complete definition, refer the reader to [MS15, Chapters 2 and 3] for details, and rather explain them in some easy cases that are of importance to us.

Definition 2.10. Let $I \subseteq K[x^{\pm}] = K[x_1^{\pm}, \dots, x_n^{\pm}]$ be a Laurent polynomial ideal. The *tropical variety* or *tropicalization* of I is defined to be:

$$\mathrm{Trop}(I) \coloneqq \Big\{ w \in \mathbb{R}^n \mid \mathrm{in}_w(I) \neq \mathfrak{K}[x^{\pm}] \Big\}.$$

By the Fundamental Theorem [MS15, Theorem 3.2.3], if the valuation val is nontrivial, we also have

$$\mathrm{Trop}(I) \coloneqq \mathrm{cl}\Big(\Big\{ \mathrm{val}(z) \in \mathbb{R}^n \mid z \in V(I) \Big\} \Big)$$

where $\mathrm{val}(\cdot)$ denotes coordinatewise valuation, and $\mathrm{cl}(\cdot)$ denotes Euclidean closure.

By the Structure Theorem [MS15, Theorem 3.3.5], we can use the Gröbner complex to make $\mathrm{Trop}(I)$ into a weighted polyhedral complex that, if I is prime, is balanced and connected in codimension one.

Example 2.11. If $I = \langle f \rangle \in K[x^{\pm}]$ is principal, then $\mathrm{Trop}(f) \coloneqq \mathrm{Trop}(I)$ is referred to as a *tropical hypersurface*, and it is dual to the regular subdivision of the Newton polytope induced by the valuation of its coefficients. The multiplicities are dual to the lattice length of the edges of the Newton polytope.

Consider the two polynomials, whose tropical hypersurfaces and corresponding subdivisions of the Newton polytopes are illustrated in Figure 2:

$$f_1 = t^2x_1^2 + x_1x_2 + 1 \text{ and } f_2 = t^2x_1^2 + x_1x_2 + t^6x_2^2 + x_1 + t^2x_2 + 1 \in \mathbb{C}\{\{t\}\}[x_1^{\pm}, x_2^{\pm}].$$

The idea behind the duality is that for any $(z_1, z_2) \in (\mathbb{C}\{\{t\}\}^*)^2$ with $f_i(z_1, z_2) = 0$ the lowest t-degree terms in $f_i(z_1, z_2)$ must cancel. Hence the monomials of the initial form $\mathrm{in}_w(f_i)$, $w := \mathrm{val}((z_1, z_2)) \in \mathbb{R}^2$ must form a positive-dimensional cell in the Newton subdivision. For instance:

$\mathrm{Trop}(f_1)$ has a ray $(-1, 1) + \mathbb{R}_{\geq 0} \cdot (0, 1)$, that contains weight vectors $w \in \mathbb{R}^n$ such that $\mathrm{in}_w(f_1) = x_1^2 + 1$ and the coordinatewise valuations of solutions of the form $(t^{-1} \cdot (c + z_1'), t \cdot z_2') \in V(f_1)$, where $c \in \mathbb{C}^*$ with $c^2 = 1$, and $z_i \in \mathbb{C}\{\{t\}\}$ with $\mathrm{val}(z_i') > 0$. When substituting the solution into f_1, we see that the monomials of the initial form contribute to the terms of lowest t-degree:

$$f_1\big(t^{-1} \cdot (c + z_1'), t \cdot z_2'\big) = \underbrace{t^2 \cdot \big(t^{-1} \cdot (c + z_1')\big)^2}_{\mathrm{val}=0} + \underbrace{\big(t^{-1} \cdot (c + z_1')\big) \cdot \big(t \cdot z_2'\big)}_{\mathrm{val}>0} + \underbrace{1}_{\mathrm{val}=0}.$$

Next, we introduce stable intersections of balanced polyhedral complexes using both the definition in [MS15, Definition 3.6.5] and the equivalent formulation in [MS15, Proposition 3.6.12].

Definition 2.12. Let Σ_1, Σ_2 be two weighted balanced polyhedral complexes in $\mathbb{R}^n$. Their *stable intersection* is defined to be the polyhedral complex

$$\Sigma_1 \wedge \Sigma_2 := \{\sigma_1 \cap \sigma_2 \mid \sigma_1 \in \Sigma_1, \sigma_2 \in \Sigma_2, \dim(\sigma_1 + \sigma_2) = n\}$$

with the multiplicities for the top-dimensional polyhedra given by

$$\mathrm{mult}_{\Sigma_1 \wedge \Sigma_2}(\sigma_1 \cap \sigma_2) := \sum_{\tau_1, \tau_2} \mathrm{mult}_{\Sigma_1}(\tau_1)\, \mathrm{mult}_{\Sigma_2}(\tau_2)[N : N_{\tau_1} + N_{\tau_2}].$$

Here, the sum is taken over all maximal $\tau_1 \in \Sigma_1$ and $\tau_2 \in \Sigma_2$ containing $\sigma_1 \cap \sigma_2$ with $\tau_1 \cap (\tau_2 + \varepsilon \cdot v) \neq \emptyset$ for some generic $v \in \mathbb{R}^n$ and $\varepsilon > 0$ sufficiently small. Moreover, N denotes the standard lattice $\mathbb{Z}^n$, and N_{τ_i} denotes the sublattice generated by the linear span of the τ_i translated to the origin. Alternatively, it can be defined as:

$$\Sigma_1 \wedge \Sigma_2 := \lim_{\varepsilon \to 0} \Sigma_1 \cap (\Sigma_2 + \varepsilon \cdot v).$$

Example 2.13. Figure 3 illustrates the stable intersection of $\mathrm{Trop}(f_1)$ and $\mathrm{Trop}(f_2)$ from Example 2.11. It consists of four points, each of multiplicity 1.

The following is a generalization of the Transverse Intersection Theorem [BJSST07, Lemma 15] from tropical varieties as *supports of polyhedral complexes* to tropical varieties as *balanced polyhedral complexes*.

Theorem 2.14. *Let $I, J \subseteq K[x^{\pm}]$ be complete intersections. Suppose that* $\mathrm{Trop}(I)$ *and* $\mathrm{Trop}(J)$ *intersect transversally. Then*

$$\mathrm{Trop}(I + J) = \mathrm{Trop}(I) \wedge \mathrm{Trop}(J).$$

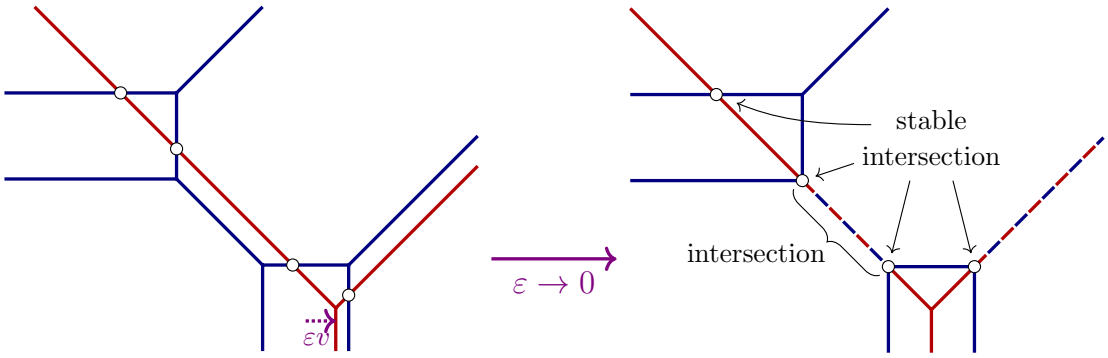


FIGURE 3. The stable intersection of the two tropical plane curves of Example 2.11.

Moreover, for all $w \in \mathrm{Trop}(I+J)$ *we have*

$$\mathrm{in}_w(I+J) = \mathrm{in}_w(I) + \mathrm{in}_w(J).$$

Proof. Set-theoretically the first statement is [MS15, Theorem 3.4.12]. The fact that the multiplicities match follows from [OP13, Corollary 5.1.3], which requires I and J to be Cohen–Macaulay. The latter is implied from I and J being complete intersections. The second statement on the initial ideals is proven in the proof of [MS15, Theorem 3.4.12], see in particular [MS15, Equation 3.4.3]. □

We end the section with a small lemma that we need in the next section.

Lemma 2.15. *Let* $P \in (\mathbb{C}^*)^m$*, and let* $X \subsetneq \mathbb{C}\{\{t\}\}^m$ *be proper Zariski-closed subset. Then* $t^v \cdot P \notin X$ *for generic* $v \in \mathbb{Q}^m$*.*

Proof. As $X \subseteq \mathbb{C}\{\{t\}\}^m$ is proper and Zariski closed, $X \cap (\mathbb{C}\{\{t\}\}^*)^m \subseteq (\mathbb{C}\{\{t\}\}^*)^m$ is also proper and Zariski closed. Hence $\mathrm{Trop}(I(X))$ has positive codimension in $\mathbb{R}^m$, and the complement of its support is dense. This immediately gives the desired statement. □

3. Tropical homotopies

In this section, we explain how to construct generically optimal homotopies for a parametrized polynomial system using data about its tropicalization. Algorithm 3.1 naturally generalizes polyhedral homotopies, see [BBC+23, Algorithm 3.1], and is a variation of an idea found in the works of Leykin and Yu [LY19].

Algorithm 3.1 (Homotopies from tropical data)**.**

Input: $(\mathcal{F}, P)$, where

(1) $\mathcal{F} = \{f_1, \dots, f_n\} \subseteq \mathbb{C}[a][x^{\pm}]$ is a square, parametrized polynomial system that is generically zero-dimensional and generically radical,

(2) $P \in (\mathbb{C}^*)^m$.

From hereon, we will regard $\mathcal{F}$ as a parametrized polynomial system over $\mathbb{C}\{\{t\}\}$, and set $\mathcal{I} \coloneqq \langle \mathcal{F} \rangle \subseteq \mathbb{C}\{\{t\}\}[a][x^{\pm}]$.

Output: A finite set $\{(H_i, V_i) \mid i = 1, \dots, r\}$, where

(1) $H_i \subseteq \mathbb{C}\{\{t\}\}[x^{\pm}]$, a homotopy with $\mathcal{I}_P = \langle H_i|_{t=1} \rangle$,
(2) $V_i \subseteq V(H_i|_{t=0}) \subseteq (\mathbb{C}^*)^n$, starting solutions,

so that

(1) every point in $V(\mathcal{I}_P)$ is connected to a point in V_i via H_i for some $i \in [r]$,
(2) $\sum_{i=1}^{r} |V_i| = \ell_{\mathcal{I}, \mathbb{C}\{\{t\}\}(a)}$.[1]

1: Pick a generic choice of parameter valuations $v \in \mathbb{Q}^m$ and set $Q \coloneqq t^v \cdot P \in \mathbb{C}\{\{t\}\}^m$.
2: Compute the tropical data required for homotopy construction:
(1) $\mathrm{Trop}(\mathcal{I}_Q) \subseteq \mathbb{R}^n$,
(2) $V^{(w)} \subseteq (\mathbb{C}^*)^n$, a sufficiently precise approximation of $V(\mathrm{in}_w(\mathcal{I}_Q)) \subseteq (\mathbb{C}^*)^n$ for each $w \in \mathrm{Trop}(\mathcal{I}_Q)$.
3: Construct the homotopies

$$H^{(w)} \coloneqq \left\{ t^{-\operatorname{trop}(f)(w)} \cdot f(t^w \cdot x) \mid f \in \mathcal{F}_Q \right\} \subseteq \mathbb{C}\{\{t\}\}[x^{\pm}] \text{ for all } w \in \mathrm{Trop}(\mathcal{I}_Q).$$

4: **return** $\{(H^{(w)}, V^{(w)}) \mid w \in \mathrm{Trop}(\mathcal{I}_Q)\}$

Proof of correctness. Without loss of generality, we may assume that $v \in \mathbb{Z}^m$, so that $\mathcal{F}_Q \subseteq \mathbb{C}[t^{\pm}][x^{\pm}]$. By Lemma 2.15, $\langle \mathcal{F}_Q \rangle$ is zero-dimensional and radical. Hence, $\mathcal{F}_Q$ may be used for homotopy continuation with target system $\mathcal{F}_Q|_{t=1} = \mathcal{F}_P$ and starting system $\mathcal{F}_Q|_{t=\varepsilon}$ for $\varepsilon > 0$ sufficiently small by [SW05, Theorem 7.1.1]. The solutions of $\mathcal{F}_Q$ may diverge however at $t = 0$. By the Newton–Puiseux theorem, each homotopy path is parametrized by a Puiseux series around $t = 0$, and the changes of coordinates in Line 3 ensures that the homotopy paths whose Puiseux series have coordinatewise valuation w do not diverge at $t = 0$, without affecting the solutions at $t = 1$. □

Example 3.2 (Polyhedral homotopies)**.** Consider $F = \{f_1, f_2\}$ from [BBC+23, Example 11]:

$$f_1 \coloneqq 5 - 3x_1^2 - 3x_2^2 + x_1^2x_2^2, \quad f_2 \coloneqq 1 + 2x_1x_2 - 5x_1x_2^2 - 3x_1^2x_2 \in \mathbb{C}[x_1^{\pm}, x_2^{\pm}].$$

For the input of Algorithm 3.1, consider the parametrized system $\mathcal{F} = \{f_1, f_2\}$ with

$$\begin{aligned} f_1 &\coloneqq a_{0,0} + a_{2,0}x_1^2 + a_{0,2}x_2^2 + a_{2,2}x_1^2x_2^2, \\ f_2 &\coloneqq b_{0,0} + b_{1,1}x_1x_2 + b_{1,2}x_1x_2^2 + b_{2,1}x_1^2x_2 \in \mathbb{C}[a, b][x^{\pm}] \end{aligned} \tag{3.1}$$

and parameters

$$P \coloneqq (\underset{a_{0,0}}{5}, -3, -3, \underset{a_{0,2}}{1}, \underset{a_{2,2}}{1}, \underset{b_{0,0}}{2}, \underset{b_{1,1}}{-5}, \underset{b_{1,2}}{-3}, \underset{b_{2,1}}{}) \in \mathbb{C}^8,$$

so that $\mathcal{F}_P = F$.

[1] $|V_i|$ is counted with multiplicity as V_i may not be smooth, see Remark 3.3.

In Step 1, consider the choice of parameter valuations

$$v := (\underset{a_{0,0}}{0}, \underset{a_{2,0}}{0}, \underset{a_{0,2}}{0}, \underset{a_{2,2}}{0}, \underset{b_{0,0}}{0}, \underset{b_{1,1}}{2}, \underset{b_{1,2}}{3}, \underset{b_{2,1}}{3}) \in \mathbb{Q}^8,$$

so that $f_{1,Q}, f_{2,Q}$ coincide with h_1, h_2 in [BBC+23, Example 11]:

$$\begin{aligned} f_{1,Q} &= 5 - 3x_1^2 - 3x_2^2 + x_1^2x_2^2, \\ f_{2,Q} &= 1 + 2t^2x_1x_2 - 5t^3x_1x_2^2 - 3t^3x_1^2x_2 \in \mathbb{C}\{\{t\}\}[x_1, x_2], \end{aligned}$$

For the tropical data in Step 2, note that by Theorem 2.14 and without the need for any Gröbner basis computations, we obtain the tropicalization $\mathrm{Trop}(\mathcal{I}_Q) = \{(0, -\frac{3}{2}), (-\frac{3}{2}, 0)\}$, see Figure 4 (a), as well as the initial ideals

$$\begin{aligned} \mathrm{in}_w(\langle f_{1,Q}, f_{2,Q}\rangle) &= \mathrm{in}_w(\langle f_{1,Q}\rangle) + \mathrm{in}_w(\langle f_{2,Q}\rangle) = \langle \mathrm{in}_w(f_{1,Q}), \mathrm{in}_w(f_{2,Q})\rangle \\ &= \begin{cases} \langle 3x_2^2 + x_1^2x_2^2, 1 + 5x_1x_2^2\rangle & \text{for } w = (0, -\frac{3}{2}) \\ \langle 3x_1^2 + x_1^2x_2^2, 1 - 3x_1^2x_2\rangle & \text{for } w = (-\frac{3}{2}, 0) \end{cases} \end{aligned}$$

which yields the following $4 + 4$ starting solutions in $(\mathbb{C}^*)^2$

$$\begin{aligned} V^{(0,-\frac{3}{2})} &= \left\{ \left(z_1, \pm\sqrt{-\tfrac{1}{5z_1}} \right) \;\middle|\; z_1 = \pm\sqrt{-3} \right\} \quad \text{and} \\ V^{(-\frac{3}{2},0)} &= \left\{ \left(\pm\sqrt{\tfrac{1}{3z_2}}, z_2 \right) \;\middle|\; z_2 = \pm\sqrt{-3} \right\}. \end{aligned}$$

As for homotopies, for $w = (0, -\frac{3}{2})$ in Step 3, we get

$$\begin{aligned} \mathrm{trop}(f_{1,Q})\left(0, -\tfrac{3}{2}\right) &= \min\left(0, 2\cdot 0, 2\cdot(-\tfrac{3}{2}), 2\cdot 0 + 2\cdot(-\tfrac{3}{2})\right)(w) = -3 \\ \mathrm{trop}(f_{2,Q})\left(0, -\tfrac{3}{2}\right) &= \min\left(0, 2 + 0 + (-\tfrac{3}{2}), 5 + 0 + 2\cdot(-\tfrac{3}{2}), 3 + 2\cdot 0 + (-\tfrac{3}{2})\right) = 0 \end{aligned}$$

and hence

$$\begin{aligned} t^3 f_{1,Q}(x_1, t^{-3/2}x_2) &= 5t^3 - 3t^3x_1^2 - 3x_2^2 + x_1^2x_2^2, \\ t^0 f_{2,Q}(x_1, t^{-3/2}x_2) &= 1 + 2t^{1/2}x_1x_2 - 5x_1x_2^2 - 3t^{3/2}x_1^2x_2 \in \mathbb{C}\{\{t\}\}[x_1, x_2]. \end{aligned}$$

This homotopy is the same as [BBC+23, Example 11] for $\alpha = (0, -3)$ after substituting t by s^2, making all exponents integer. The same holds for $w = (-\frac{3}{2}, 0)$, which will reconstruct the homotopies of [BBC+23, Example 11] for $\gamma = (-3, 0)$.

More generally, any polyhedral homotopy can be obtained from Algorithm 3.1 with a parametrized input system where every coefficient has its own parameter as in Equation (3.1). Obtaining the $\mathrm{Trop}(\mathcal{I}_Q)$ for tropical data in Step 2 is usually done by mixed cells, see Figure 4 (b), and obtaining the starting solutions $V^{(w)}$ is easy as the initial ideals $\mathrm{in}_w(\mathcal{I}_Q)$ will be binomial.

To borrow a term from chess, we refer to the numerical challenges of initiating the tracing of the homotopies $(H^{(w)}, V^{(w)})$ produced by Algorithm 3.1 around $t = 0$ as the *early game* of path tracking. Depending on the parametrized polynomial

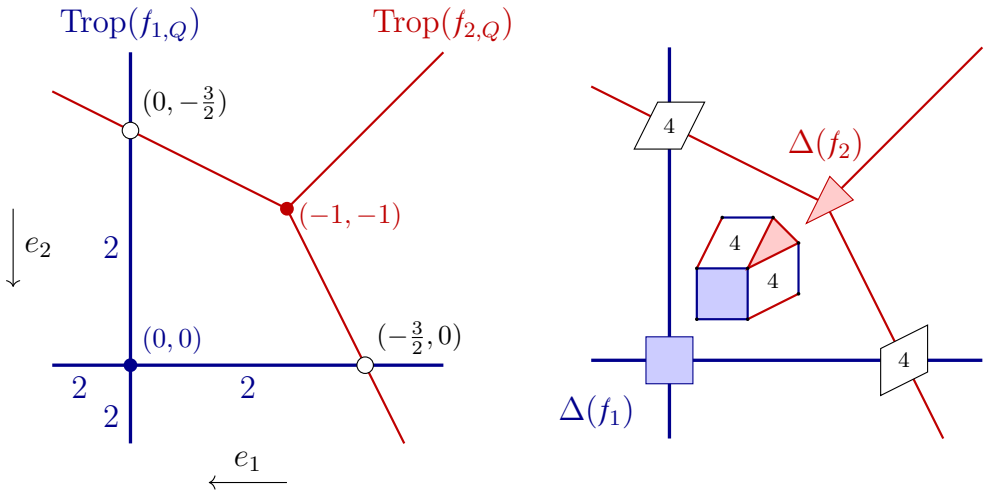


FIGURE 4. The tropical intersection from Example 3.2 and its mixed cells. Here, $\Delta(\cdot)$ denotes the Newton polytope.

system $\mathcal{F}$ at hand, the following are two of the main early game issues one needs to address. (However, as we will see in later sections, none of these issues arise for the parametrized systems considered in Sections 4, 5.2 and 5.3.)

Remark 3.3 (Early game technicalities)**.** All potential issues with the starting system stem from the following two technicalities:

(i) the initial ideal $\mathrm{in}_w(\langle \mathcal{F}_Q \rangle) \subseteq \mathbb{C}[x^{\pm}]$ need not be radical,

(ii) at $t = 0$, the homotopy $H^{(w)}(0, x) \subseteq \mathbb{C}[x^{\pm}]$ need not generate the initial ideal.

Both (i) and (ii) lead to the problem that, at $t = 0$ and for $z \in V^{(w)}$, the Jacobian $J(H^{(w)}|_{t=0})(z) \in \mathbb{C}^{n\times n}$ is not invertible, which means using the usual predictor-corrector methods described in [BBC+23, Section 2.4] is not possible.

This is a well-known issue that can be worked around by approximating the evaluation of the Puiseux series solution at small $t = \varepsilon > 0$ by $z{\cdot}\varepsilon^w := (z_1{\cdot}\varepsilon^{w_1}, \ldots, z_n{\cdot}\varepsilon^{w_n})$. If the initial ideal $\mathrm{in}_w(\langle \mathcal{F}_Q \rangle)$ is not radical, this may further require higher order terms, see [Stu02, Section 3.3] and [LY19, Remark 6].

If $\mathrm{in}_w(\langle \mathcal{F}_Q \rangle)$ is radical, we can alternatively construct the homotopies $H^{(w)}$ from the tropical Gröbner bases used to compute $V^{(w)}$; see Example 5.4. If the Gröbner basis is not square, we can construct square homotopies by taking a generic linear combination of the Gröbner basis elements.

The technicalities outlined in Remark 3.3 motivate the following definition:

Definition 3.4. Let $\mathcal{I} \subseteq \mathbb{C}[a][x^{\pm}]$ be a parametrized polynomial ideal.

- We say that $\mathcal{I}$ has *binomial*, *complete intersection*, or *radical initials* at a choice of parameters $Q \in \mathbb{C}\{\{t\}\}^m$, if for all $w \in \mathrm{Trop}(\mathcal{I}_Q)$ the initial ideal $\mathrm{in}_w(\mathcal{I}_Q)$ is binomial, a complete intersection, or radical, respectively.
- We say $\mathcal{I}$ has one of the aforementioned properties *sufficiently often*, if there is a Zariski dense set $U \subseteq \mathbb{C}\{\{t\}\}^m$ such that the property holds for all $Q \in U$.
- We say $\mathcal{I}$ has one of the aforementioned properties *generically*, if there is a Zariski open and dense set $U \subseteq \mathbb{C}\{\{t\}\}^m$ such that the properties above hold for all $Q \in U$.

We conclude the section by showing why binomial initial ideals are highly desirable. The result is used in Section 4.

Theorem 3.5. *Suppose that $\mathcal{I} \subseteq \mathbb{C}\{\{t\}\}[a][x^{\pm}]$ is generically zero-dimensional. If $\mathcal{I}$ has binomial initials at $Q \in \mathbb{C}\{\{t\}\}^m$, then $\mathcal{I}$ has radical initials at Q. In particular, if $\mathcal{I}$ has binomial initials sufficiently often, it is generically radical.*

Proof. The first part follows from [ES96, Corollary 2.2], which states that binomial ideals are radical over fields of characteristic 0.

Let $Q \in \mathbb{C}\{\{t\}\}^m$ be a choice of parameters such that $\mathcal{I}_Q$ is zero-dimensional and has binomial initial ideals. By [HK12, Proposition 3.9], this means that $\mathcal{I}_Q$ is schön, and we can follow the argument before [HK12, Proposition 3.9] to show that $\mathcal{I}_Q$ is radical. As the set of parameters $Q \in \mathbb{C}\{\{t\}\}^m$ for which $\mathcal{I}_Q$ is radical is Zariski-open, this implies that $\mathcal{I}$ is generically radical. □

Example 3.6. Many naturally occurring types of parametrized polynomial ideals $\mathcal{I} \subseteq \mathbb{C}[a][x^{\pm}]$ have binomial initials sufficiently often.

If $\mathcal{I}$ is linear in the variables x, then its initial ideals $\mathrm{in}_w(\mathcal{I}_Q)$ are binomial provided $w \in \mathrm{Trop}(\mathcal{I}_Q)$ lies in the relative interior of a maximal polyhedron. This is a consequence of [BLMM17, Lemma 1], and the fact that all multiplicities on a tropical linear space are 1.

If $\mathcal{I}$ is generated by parametrized polynomials $\mathcal{F} = \{f_1, \dots, f_n\}$ where each coefficient is its own parameter, such as $\mathcal{F}_{\mathrm{BKK}}$ in Example 2.8, then it is easy to find $Q \in \mathbb{C}\{\{t\}\}^m$ such that $\mathcal{I}_Q$ has binomial initials: Simply find $Q \in \mathbb{C}\{\{t\}\}^m$ such that

(1) the valuations of the coefficients of the $f_{i,Q}$ induce maximal subdivisions on the Newton polytopes,
(2) the tropicalizations $\mathrm{Trop}(f_{i,Q})$ intersect transversally.

Finding such parameters (or rather their valuations) is the first step in constructing polyhedral homotopies, see [BBC+23, Algorithm 3.1 Line 4]. The fact that the initials are binomial is a consequence of Theorem 2.14.

4. Vertically parametrized polynomial systems

In this section, we discuss vertically parametrized polynomial systems, which are inspired from the steady state equations of chemical reaction networks and certain Lagrangian systems in polynomial optimization. See [FHP23] for a discussion on the generic geometry of vertically parametrized systems, and [HR22, Section 6.1] for a discussion on their generic root counts. The goal of this section is to show how Algorithm 3.1 can be carried out efficiently for these systems, and how they satisfy all desirable properties for it.

Definition 4.1. Given a multiset $S = \{\alpha_1, \dots, \alpha_m\} \subseteq \mathbb{Z}^n$ of exponent vectors, a *vertically parametrized* system with exponents S is a parametrized system

$$\mathcal{F} \coloneqq \{f_1, \dots, f_n\} \subseteq \mathbb{C}[a][x^{\pm}] \coloneqq \mathbb{C}[a_1, \dots, a_m][x_1^{\pm}, \dots, x_n^{\pm}],$$

of the form

$$f_i \coloneqq \sum_{j=1}^{m} c_{i,j} a_j x^{\alpha_j}$$

for some coefficients $c_{i,j} \in \mathbb{C}$ (note that we allow $c_{i,j} = 0$).

For the remainder of Section 4, we fix a vertically parametrized system $\mathcal{F}$ with some exponent vectors S, and let $\mathcal{I} \coloneqq \langle \mathcal{F} \rangle \subseteq \mathbb{C}\{\{t\}\}[a][x^{\pm}]$ denote the ideal generated by $\mathcal{F}$ over $\mathbb{C}\{\{t\}\}$.

4.1. Tropical data for homotopy construction.

In this subsection, we will show that vertically parametrized systems are especially suited for Algorithm 3.1.

Definition 4.2. The *modification* of a vertically parametrized system $\mathcal{F}$ is given by

$$\hat{\mathcal{F}} \coloneqq \{\hat{f}_i, \hat{g}_j \mid i \in [n], j \in [m]\} \subseteq \mathbb{C}[a][x^{\pm}, y^{\pm}] \coloneqq \mathbb{C}[a][x_i^{\pm}, y_j^{\pm} \mid i \in [n], j \in [m]].$$

where

$$\hat{f}_i \coloneqq \sum_{j=1}^{m} c_{i,j} a_j y_j \quad \text{and} \quad \hat{g}_j \coloneqq y_j - x^{\alpha_j}.$$

In what follows, we let $\hat{\mathcal{I}}_{\text{lin}} \coloneqq \langle \hat{f}_i \mid i \in [n] \rangle$ and $\hat{\mathcal{I}}_{\text{bin}} \coloneqq \langle \hat{g}_j \mid j \in [m] \rangle$ denote the ideals generated by the $\hat{f}_i$ and $\hat{g}_j$ respectively in $\mathbb{C}\{\{t\}\}[a][x^{\pm}, y^{\pm}]$, and set $\hat{\mathcal{I}} \coloneqq \hat{\mathcal{I}}_{\text{lin}} + \hat{\mathcal{I}}_{\text{bin}}$.

This modification gives a way to compute $\operatorname{Trop}(\mathcal{I}_Q)$ as a stable intersection, which was also explained through the notion of toric equivariance in [HR22, Section 6.1].

Lemma 4.3. *For a generic choice of parameters $Q \in \mathbb{C}\{\{t\}\}^m$, we have an isomorphism of (zero-dimensional) weighted polyhedral complexes*

$$\begin{array}{ccc} \mathbb{R}^n & & \mathbb{R}^n \times \mathbb{R}^m \\ \cup| & & \cup| \\ \operatorname{Trop}(\mathcal{I}_Q) & \xleftrightarrow{\cong} & \operatorname{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q}) \wedge \bigwedge_{j=1}^{m} \operatorname{Trop}(\hat{g}_{j,Q}) \\ (w_i)_{i\in[n]} & \xleftarrow{\pi} & \big((w_i)_{i\in[n]}, (w_j)_{j\in[m]}\big) \end{array} \tag{4.1}$$

Proof. As $\operatorname{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q})$ and the $\operatorname{Trop}(\hat{g}_{j,Q})$ intersect transversally for generic v, we have $\operatorname{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q}) \wedge \bigwedge_{j=1}^m \operatorname{Trop}(\hat{g}_{j,Q}) = \operatorname{Trop}(\hat{\mathcal{I}}_Q)$ by Theorem 2.14. Moreover, it is straightforward to show that $\hat{\mathcal{I}}_Q \cap \mathbb{C}\{\{t\}\}[x^{\pm}] = \mathcal{I}_Q$, which means that the projection in Equation (4.1) is an isomorphism. □

Step 2 of Algorithm 3.1 requires $V(\operatorname{in}_w(\mathcal{I}_Q))$. For vertically parametrized systems, we can obtain generators of $\operatorname{in}_w(\mathcal{I}_Q)$ through a simple linear Gröbner basis computation.

Lemma 4.4. *Suppose we have*

(1) $Q \coloneqq t^v \cdot P \in \mathbb{C}\{\{t\}\}^m$ for some $P \in (\mathbb{C}^*)^m$ and some $v \in \mathbb{Q}^m$,

(2) $w \in \operatorname{Trop}(\mathcal{I}_Q)$,

(3) $\hat{w} \in \operatorname{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q}) \wedge \bigwedge_{j=1}^m \operatorname{Trop}(\hat{g}_{j,Q})$ with $\pi(\hat{w}) = w$ as in Lemma 4.3, and

(4) $\hat{G} \subseteq \hat{\mathcal{I}}_{\mathrm{lin},Q}$ a tropical Gröbner basis with respect to $\hat{w}$.

Then $\operatorname{in}_w(\mathcal{I}_Q) = \langle \operatorname{in}_w(\hat{g}|_{y_j=x^{\alpha_j}}) \mid \hat{g} \in \hat{G}\rangle$, where $(\cdot)|_{y_j=x^{\alpha_j}}$ denotes substituting all y_j by x^{α_j}. In particular, $\{\hat{g}|_{y_j=x^{\alpha_j}}\} \subseteq \mathcal{I}_Q$ is a Gröbner basis with respect to w.

Proof. Note that

$$\langle \operatorname{in}_{\hat{w}}(\hat{g})|_{y_j=x^{\alpha_j}} \mid \hat{g} \in \hat{G}\rangle = \langle \operatorname{in}_{\hat{w}}(\hat{g}) \mid \hat{g} \in \hat{G}\rangle|_{y_j=x^{\alpha_j}} = \operatorname{in}_{\hat{w}}(\hat{\mathcal{I}}_{\mathrm{lin},Q})|_{y_j=x^{\alpha_j}}.$$

Hence, it suffices to show that $\operatorname{in}_w(\mathcal{I}_Q) = \operatorname{in}_{\hat{w}}(\hat{\mathcal{I}}_{\mathrm{lin},Q})|_{y_j=x^{\alpha_j}}$.

For the "$\subseteq$" inclusion, it suffices to consider elements of the form $\operatorname{in}_w(g) \in \operatorname{in}_w(\mathcal{I}_Q)$ for some $g = \sum_{i=1}^r q_i f_{i,Q} \in \mathcal{I}_Q$ with $q_i \in \mathbb{C}\{\{t\}\}[x^{\pm}]$. Let $\hat{g} \coloneqq \sum_{i=1}^r q_i \hat{f}_{i,Q} \in \mathcal{I}_{\mathrm{lin},Q}$, so that $g = \hat{g}|_{y_j=x^{\alpha_j}}$. Due to (2) and (3), we then have

$$\operatorname{in}_w(g) = \operatorname{in}_w(\hat{g}|_{y_j=x^{\alpha_j}}) = \operatorname{in}_{\hat{w}}(\hat{g})|_{y_j=x^{\alpha_j}} \in \operatorname{in}_{\hat{w}}(\hat{\mathcal{I}}_{\mathrm{lin},Q})|_{y_j=x^{\alpha_j}}.$$

For the "$\supseteq$" inclusion, consider $\hat{g} = \sum_{i=1}^r c_i \hat{f}_{i,Q} \in \hat{\mathcal{I}}_{\mathrm{lin},Q}$ with $c_i \in \mathbb{C}\{\{t\}\}$. As $\hat{\mathcal{I}}_{\mathrm{lin},Q}$ is linear, it suffices to consider a linear generator $\hat{g}$. Let $g \coloneqq \sum_{i=1}^r c_i f_{i,Q}$, so that $g = \hat{g}|_{y_j=x^{\alpha_j}}$. Due to (2) and (3), we again have

$$\operatorname{in}_{\hat{w}}(\hat{g})|_{y_j=x^{\alpha_j}} = \operatorname{in}_{\hat{w}}(\hat{g}|_{y_j=x^{\alpha_j}}) = \operatorname{in}_w(g) \in \operatorname{in}_w(\mathcal{I}_Q).$$

□

Example 4.5. Let $\mathcal{F}_{\text{verti}} = \{f_1, f_2\} \subseteq \mathbb{C}[a][x^{\pm}]$ be the vertically parametrized system from Example 2.8, given by

$$\begin{aligned} f_1 &= a_1x_1^2 + a_2x_2^2 + a_3x_1 + a_4x_2 + a_5 \quad \text{and} \\ f_2 &= 3a_1x_1^2 + 3a_2x_2^2 + 5a_3x_1 + 7a_4x_2 + 11a_5, \end{aligned}$$

and consider its modification

$$\begin{aligned} \hat{f}_1 &= a_1y_1 + a_2y_2 + a_3y_3 + a_4y_4 + a_5y_5, \quad \hat{g}_1 = y_1 - x_1^2, \quad \hat{g}_3 = y_3 - x_1. \\ \hat{f}_2 &= 3a_1y_1 + 3a_2y_2 + 5a_3y_3 + 7a_4y_4 + 11a_5y_5 \quad \hat{g}_2 = y_2 - x_2^2, \quad \hat{g}_4 = y_4 - x_2, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \hat{g}_5 = y_5 - 1. \end{aligned}$$

(Note that the introduction of y_3, y_4, y_5 is not strictly necessary, as their monomials are linear or constant. They can be omitted as a computational optimization.)

For $Q := (t, 1, 1, t, 1) \in \mathbb{C}\{\{t\}\}^m$, we obtain $\text{Trop}(\hat{\mathcal{I}}_{\text{lin},Q})$ and $\text{Trop}(\hat{\mathcal{I}}_{\text{bin},Q})$ that intersect transversally in a single point $\hat{w} = 0 \in \mathbb{R}^7$ of multiplicity 2. By Lemma 4.3, this means $\text{Trop}(I_Q) = \{(0,0)\}$, and for $w = (0,0)$, Algorithm 3.1 Line 3 gives the homotopy

$$\begin{aligned} H^{(w)} &= \Big(t^0 \cdot \big(t^{1-0}x_1^2 + t^{0-0}x_2^2 + t^{0-0}x_1 + t^{1-0}x_2 + t^0\big), \\ &\qquad t^0 \cdot \big(3t^{1-0}x_1^2 + 3t^{0-0}x_2^2 + 5t^{0-0}x_1 + 7t^{1-0}x_2 + 11t^0\big)\Big) \\ &= \Big(tx_1^2 + x_2^2 + x_1 + tx_2 + 1,\ \ 3tx_1^2 + 3x_2^2 + 5x_1 + 7tx_2 + 11\Big). \end{aligned}$$

Note that $H^{(w)} = (f_{1,Q}, f_{2,Q})$ due to $w = (0,0)$. At $t = 0$, we have the equations $x_2^2 + x_1 + 1 = 0 = 3x_2^2 + 5x_1 + 11$, which are not binomial. In order to obtain binomial equations, we can compute a (tropical) Gröbner basis of $\mathcal{I}_Q$ with respect to w using Lemma 4.4, which is:

$$g_1 := x_1 + tx_2 + 4 \quad \text{and} \quad g_2 := 4tx_1^2 + 4x_2^2 + 3x_1 + 2tx_2$$

Using these instead of the $f_{i,Q}$ in Algorithm 3.1 Line 3 gives us a homotopy $H^{(w)} = (g_1, g_2)$ that is binomial at $t = 0$.

We conclude this subsection, by noting that vertically parametrized polynomial systems exhibit all desirable properties for Algorithm 3.1 that are discussed at the end of Section 3.

Theorem 4.6. *Let $\mathcal{F}$ be a generically zero-dimensional vertically parametrized system. Then $\mathcal{F}$ generically has binomial, complete intersection, and radical initials.*

Proof. Consider the generating sets $\{\text{in}_{\hat{w}}(\hat{g}|_{y_j = x^{\alpha_j}}) \mid \hat{g} \in \hat{G}\} \subseteq \text{in}_w(\mathcal{I}_Q)$ of Lemma 4.4. The complete intersection property follows from the fact that the cardinality of a (minimal) Gröbner basis $\hat{G}$ of a linear ideal $\hat{\mathcal{I}}_{\text{lin},Q}$ always equals their codimension, which is n. To show binomiality, observe that for $Q = t^v \cdot P$ with $v \in \mathbb{R}^n$

generic we may assume that $\hat{w}$ lies in the relative interior of a maximal polyhedron of $\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q})$. Hence the $\mathrm{in}_{\hat{w}}(\hat{\mathcal{I}}_{\mathrm{lin},Q})$ is binomial by [BLMM17, Lemma 1], and therefore its Gröbner basis $\hat{G}$ can be chosen to be binomial. Radicality then follows from Theorem 3.5. □

See [FHP23] for an alternative proof of generic radicality, as well as explicit conditions for when a vertically parametrized system is generically zero-dimensional.

Remark 4.7. In some cases, the n-codimensional tropical linear space $\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q})$ is a *tropical complete intersection*, namely

$$\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q}) = \mathrm{Trop}(\hat{h}_1) \wedge \cdots \wedge \mathrm{Trop}(\hat{h}_n) \text{ for linear } \hat{h}_1, \ldots, \hat{h}_n \in \hat{\mathcal{I}}_{\mathrm{lin},Q}. \tag{4.2}$$

By [MS15, Theorem 3.6.1], Condition (4.2) holds if and only if the column matroid of the coefficient matrix $(c_{i,j}Q_j)_{i\in[n],j\in[m]} \in \mathbb{C}\{\{t\}\}^{n\times m}$ is *transversal*; see [Bon10, Section 2.2] for a definition. The latter can be tested in POLYMAKE [GJ00], and, if the matroid is transversal, the transversal presentation that is computed during the test can be used to construct the $\hat{h}_i$.

If Condition (4.2) holds, then one can show that the generic root count of the original system $\mathcal{F}$ is the mixed volume of the Newton polytopes of the $\hat{h}_i|_{y_j=x^{\alpha_j}} \in \mathbb{C}\{\{t\}\}[x^{\pm}]$:

$$\begin{aligned}\ell_{\mathcal{I},\mathbb{C}(a)} &\overset{\text{Lemma 4.3}}{=} \ell_{\hat{\mathcal{I}},\mathbb{C}(a)} \overset{\text{[MS15, Theorem 4.6.8]}}{=} \mathrm{MV}(\Delta(\hat{h}_i), \Delta(\hat{g}_j) \mid i \in [n], j \in [m]) \\ &= \mathrm{MV}(\Delta(\hat{h}_i|_{y_j=x^{\alpha_j}}) \mid i \in [n]),\end{aligned}$$

where $\Delta(\cdot)$ denotes the Newton polytope, $\mathrm{MV}(\cdot)$ denotes the mixed volume, and the final equality follows from the fact that the mixed volume is invariant under toric reembeddings. Instead of using Algorithm 3.1, one can now simply construct polyhedral homotopies for the $\hat{h}_i|_{y_j=x^{\alpha_j}} \in \mathbb{C}\{\{t\}\}[x^{\pm}]$. Note that the mixed volume of the $\hat{h}_i|_{y_j=x^{\alpha_j}}$ need not be the mixed volume of $\mathcal{F}_Q$.

4.2. **Remarks on computational ingredients.** We close this section with a few remarks on some computations that are required for obtaining the necessary tropical data for Algorithm 3.1 via Lemma 4.3 and Lemma 4.4.

Remark 4.8 (Tropical intersections)**.** Lemma 4.3 requires the computation of

$$\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q}) \wedge \bigwedge_{j=1}^{m} \mathrm{Trop}(\hat{g}_{j,Q}). \tag{4.3}$$

Constructing $\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q})$ and intersecting it with $\bigwedge_{j=1}^{m} \mathrm{Trop}(\hat{g}_{j,Q})$ are both known to be difficult tasks individually. However, computing the intersection in Equation (4.3) can be done faster than the sum of its constituencies. In [Jen16], Jensen develops a tropical homotopy continuation approach for computing transverse intersection of n hypersurfaces in $\mathbb{R}^n$. The resulting number is the mixed volume of the polynomials of

the hypersurfaces [MS15, Theorem 4.6.8]. Both Gfan [Gfan] and HomotopyContinuation.jl [BT18], and by proxy also Singular [DGPS24] and Macaulay2 [M2], rely on it for mixed volume computations.

Similar to homotopy continuation for polynomial systems, the basic idea of tropical homotopy continuation is to deform an easy starting intersection to a desired target intersection. What makes this process possible in tropical geometry is the duality between tropical hypersurfaces and regular subdivisions of Newton polytopes. As tropical linear spaces are dual to matroid subdivisions of matroid polytopes, this approach can in principle be generalized to intersections of hypersurfaces and tropical linear spaces as required in Lemma 4.3. The first steps were done in [DR24].

Remark 4.9 (Tropical Gröbner bases)**.** Lemma 4.4 requires the computation of a set $G \subseteq I$ such that $\mathrm{in}_w(I) = \langle \mathrm{in}_w(g) \mid g \in G \rangle$ for an ideal $I \subseteq \mathbb{C}\{\{t\}\}[x^{\pm}]$ and $w \in \mathrm{Trop}(I)$. Such G are also known as tropical Gröbner bases [MS15, Section 2.4]. Similar to classical Gröbner bases, they can be computed using Buchberger-like algorithms [CM19; MR20] or using `F4-/F5-`like algorithms [Vac18].

Most importantly for our purposes, if I is generated by linear polynomials, we can compute G using a single Gaussian elimination on the Macaulay matrix of its generators [Vac18, Algorithm 3.2.2]. Hence, we will generally consider tropical Gröbner bases of linear ideals a computational non-issue.

Remark 4.10 (Binomial systems)**.** Homotopy continuation generally requires starting solutions for the path tracking. Naturally, these starting solutions should come from systems that are easy to solve. The starting systems produced by Lemma 4.4 are binomial, similar to polyhedral homotopies. Binomial systems are easily solvable as, modulo a change of coordinates that can be computed using a Smith normal form on the exponent matrix (see, e.g., [CL14]), any binomial system is of the form

$$x_1^{d_1} - c_1 = 0, \qquad \ldots, \qquad x_n^{d_n} - c_n = 0.$$

5. Horizontally parametrized polynomial systems

In this section, we discuss horizontally parametrized polynomial systems, which were prominently studied by Kaveh and Khovanskii [KK12] and many others using the theory of Newton–Okounkov bodies. We show that the ideas from Section 4 are insufficient for addressing general systems of such type, and focus on two related types of parametrized polynomial systems instead: One is a particular class of horizontally parametrized systems, the other is a relaxation of horizontally parametrized systems, i.e., a larger parametrized family of polynomial systems as in Proposition 2.7.

Definition 5.1. A *horizontally parametrized* system with polynomial support $R = \{q_1, \dots, q_m\} \subseteq \mathbb{C}[x^{\pm}]$ is a parametrized system

$$\mathcal{F} \coloneqq \{f_1, \dots, f_n\} \subseteq \mathbb{C}[a][x^{\pm}] \coloneqq \mathbb{C}[a_{i,j} \mid i \in [n], j \in [m]][x_1^{\pm}, \dots, x_n^{\pm}],$$

of the form

$$f_i \coloneqq \sum_{j=1}^{m} c_{i,j} a_{i,j} q_j \tag{5.1}$$

for some coefficients $c_{i,j} \in \mathbb{C}$ (note that we allow $c_{i,j} = 0$).

For the remainder of Section 5, we fix a horizontally parametrized system $\mathcal{F} \subseteq \mathbb{C}[a][x^{\pm}]$ with polynomial support $R = \{q_1, \dots, q_m\} \subseteq \mathbb{C}[x^{\pm}]$. We furthermore let $\mathcal{I} \coloneqq \langle \mathcal{F} \rangle \subseteq \mathbb{C}\{\{t\}\}[a][x^{\pm}]$ be the ideal generated by $\mathcal{F}$.

It is commonplace to only consider solutions in a Zariski open space that depends on the polynomial support R [KK12, Definition 4.5]. In Lemma 5.3 below, this is done by saturation. Note that such systems indeed satisfy the requirements of Algorithm 3.1, see [KK12, Theorem 4.9] or [HR22, Proposition 6.9].

5.1. **The challenge of horizontally parametrized systems.** In [LY19], Leykin and Yu suggest considering the following modification.

Definition 5.2. The *modification* of a horizontally parametrized system $\mathcal{F}$ is

$$\hat{\mathcal{F}} \coloneqq \{\hat{f}_i, \hat{g}_j \mid i \in [n], j \in [m]\} \subseteq \mathbb{C}[a][x^{\pm}, y^{\pm}] \coloneqq \mathbb{C}[a][x_i^{\pm}, y_j^{\pm} \mid i \in [n], j \in [m]]$$

where

$$\hat{f}_i \coloneqq \sum_{j=1}^{m} c_{i,j} a_{i,j} y_j \quad \text{and} \quad \hat{g}_j \coloneqq y_j - q_j.$$

We will use the notation $\hat{\mathcal{I}}_{\text{lin}} \coloneqq \langle \hat{f}_i \mid i \in [n] \rangle$ and $\hat{\mathcal{I}}_{\text{nlin}} \coloneqq \langle \hat{g}_j \mid j \in [m] \rangle$ for the ideals generated by the $\hat{f}_i$ and $\hat{g}_j$ respectively in $\mathbb{C}\{\{t\}\}[a][x^{\pm}, y^{\pm}]$, and set $\hat{\mathcal{I}} \coloneqq \hat{\mathcal{I}}_{\text{lin}} + \hat{\mathcal{I}}_{\text{nlin}}$.

As in Lemma 4.3, the tropicalization of the horizontal modification also decomposes and relates to the tropicalization of the original ideal. The following lemma is an extension of [LY19, Lemma 4].

Lemma 5.3. *For a generic choice of parameters $Q \in \mathbb{C}\{\{t\}\}^m$, we have an isomorphism of (zero-dimensional) weighted polyhedral complexes*

$$\begin{array}{ccc} \mathbb{R}^n & & \mathbb{R}^n \times \mathbb{R}^R \\ \cup\mathrm{I} & & \cup\mathrm{I} \\ \mathrm{Trop}(\mathcal{I}_Q : (\prod_{j=1}^m q_j)^\infty) & \xleftrightarrow{\cong} & \Big(\bigwedge_{i \in [n]} \mathrm{Trop}(\hat{f}_{i,Q}) \Big) \wedge \mathrm{Trop}(\hat{\mathcal{I}}_{\text{nlin},Q}) \\ (w_i)_{i \in [n]} & \longmapsto & \big((w_i)_{i \in [n]}, (w_q)_{q \in R} \big). \end{array} \tag{5.2}$$

Proof. As the $\mathrm{Trop}(\hat{f}_{i,Q})$ and $\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{nlin},Q})$ intersect transversally for generic v, we have $(\bigwedge_{i\in[n]} \mathrm{Trop}(\hat{f}_{i,Q})) \wedge \mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{nlin},Q}) = \mathrm{Trop}(\hat{\mathcal{I}}_Q)$ by Theorem 2.14. To show that the projection in Equation (5.2) is an isomorphism, we need to show that there is an isomorphism

$$\mathbb{C}\{\{t\}\}[x^{\pm}] \Big/ \mathcal{I}_Q : (\textstyle\prod_{j=1}^m q_j)^\infty \xrightarrow{\sim} \mathbb{C}\{\{t\}\}[x^{\pm}, y^{\pm}] \Big/ \hat{\mathcal{I}}_Q, \quad \overline{x} \longmapsto \overline{x}.$$

To show that the map is well-defined, let $h \in \mathbb{C}\{\{t\}\}[x^{\pm}]$ with $h \cdot q^\beta \in \mathcal{I}_Q$ for some $\beta \in \mathbb{Z}^m_{\geq 0}$. As $\mathcal{I}_Q \subseteq \hat{\mathcal{I}}_Q$ as sets, we have $h \cdot q^\beta \in \hat{\mathcal{I}}_Q$, which implies $h \cdot y^\beta \in \hat{\mathcal{I}}_Q$ and consequently $h \in \hat{\mathcal{I}}_Q$.

The map is clearly surjective, as $\overline{x}$ maps to $\overline{x}$, and $\overline{q}_j$ maps to $\overline{q}_j = \overline{y}$.

To show that the map is injective, let $h \in \hat{\mathcal{I}}_Q \cap \mathbb{C}\{\{t\}\}[x^{\pm}]$, say

$$h = \Big(\sum_{i=1}^n h_{1,i} \cdot \hat{f}_{i,Q}\Big) + \Big(\sum_{j=1}^m h_{2,j} \cdot \hat{g}_{j,Q}\Big) \text{ for some } h_{1,i}, h_{2,j} \in \mathbb{C}\{\{t\}\}[x^{\pm}, y^{\pm}]$$

Substituting y_j by q_j on both sides (leaving the left side unchanged) then yields

$$h = \Big(\sum_{i=1}^n h'_{1,i} \cdot f_{i,Q}\Big) \qquad \text{for some } h'_{1,i} \in \mathbb{C}\{\{t\}\}[x^{\pm}]$$

showing that $h \in \mathcal{I}_Q$. □

Example 5.4. Consider $\mathcal{F}_{\mathrm{hori}} = \{f_1, f_2\} \subseteq \mathbb{C}[b][x^{\pm}]$ from Example 2.8 given by

$$f_1 \coloneqq b_1x_1^2 + b_1x_2^2 + b_2x_1 + b_3x_2 + b_4 \quad \text{and} \quad f_2 \coloneqq b_5x_1^2 + b_5x_2^2 + b_6x_1 + b_7x_2 + b_8,$$

which was to be solved for $P = (1,1,1,1,3,5,7,11)$, and its modification

$$\hat{f}_1 \coloneqq b_1y_1 + b_2y_2 + b_3y_3 + b_4y_4, \qquad \hat{g}_1 \coloneqq y_1 - (x_1^2 + x_2^2), \qquad \hat{g}_3 \coloneqq y_3 - x_2,$$
$$\hat{f}_2 \coloneqq b_5y_1 + b_6y_2 + b_7y_3 + b_8y_4, \qquad \hat{g}_2 \coloneqq y_2 - x_1, \qquad \hat{g}_4 \coloneqq y_4 - 1.$$

Note that all $\mathrm{Trop}(\hat{g}_{j,Q})$ are intersecting transversally, which means

$$\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{nlin},Q}) = \bigwedge_{j=1}^4 \mathrm{Trop}(\hat{g}_{j,Q}),$$

making it easy to compute. Moreover, the introduction of y_2, y_3, y_4 is not strictly necessary as q_2, q_3, q_4 are variables or constants. They can be omitted for the sake of optimization.

For $Q \coloneqq (t^3, 1, t^2, t^3, 3t^2, 5t^2, 7t^3, 11t^2)$, we obtain $\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{lin},Q})$ and $\mathrm{Trop}(\hat{\mathcal{I}}_{\mathrm{nlin},Q})$ that intersect transversally in a single point $\hat{w} = (2,0,0,2,0,0)$ of multiplicity 2, as illustrated in Figure 5. By Lemma 5.3, this means $\mathrm{Trop}(\mathcal{I}_Q) = \{(2,0)\}$, and for

$w = (2, 0)$ Algorithm 3.1 Line 3 gives the homotopy

$$\begin{aligned} H^{(w)} &= \Big(t^{-2} \cdot \big(\; t^{3+4}x_1^2 + \; t^{3+0}x_2^2 + \; t^{0+2}x_1 + \; t^{2+0}x_2 + \quad t^3\big), \\ &\qquad t^{-2} \cdot \big(3t^{2+4}x_1^2 + 3t^{2+0}x_2^2 + 5t^{2+2}x_1 + 7t^{3+0}x_2 + 11t^2\big)\Big) \\ &= \Big(t^5x_1^2 + tx_2^2 + x_1 + x_2 + t, \;\; 3t^4x_1^2 + 3x_2^2 + 5t^2x_1 + 7tx_2 + 11\Big). \end{aligned}$$

At $t = 0$, we have the binomial equations $x_1 + x_2 = 0 = 3x_2^2 + 11$, which has two solutions, and at $t = 1$ we obtain the target system $\mathcal{F}_{\text{hori}}|_{b=1}$.

In contrast, for $Q := (1, t, t^2, 1, 1, t, t^2, 1)$, we obtain $\text{Trop}(\hat{\mathcal{I}}_{\text{lin},Q})$ and $\text{Trop}(\hat{\mathcal{I}}_{\text{nlin},Q})$ that intersect transversally in a single point $w = -(1, 1, 0, 1, 1, 0)$ of multiplicity 2. By Lemma 5.3, this means $\text{Trop}(\mathcal{I}_Q) = \{-(1, 1)\}$. And for $w = -(1, 1)$, Line 3 in Algorithm 3.1 gives the homotopy

$$\begin{aligned} H^{(w)} &= \Big(t^2 \cdot \big(\; t^{0-2}x_1^2 + \; t^{0-2}x_2^2 + \; t^{1-1}x_1 + \; t^{2-1}x_2 + 1\big), \\ &\qquad t^2 \cdot \big(3t^{0-2}x_1^2 + 3t^{0-2}x_2^2 + 5t^{1-1}x_1 + 7t^{2-1}x_2 + 11\big)\Big) \\ &= \Big(x_1^2 + x_2^2 + t^2x_1 + t^3x_2 + t^2, \;\; 3x_1^2 + 3x_2^2 + 5t^2x_1 + 7t^3x_2 + 11t^2\Big). \end{aligned}$$

At $t = 0$, we have the binomial equations $x_1^2 + x_2^2 = 0 = 3x_1^2 + 3x_2^2$, which is problematic as they cut out a one-dimensional solution set. However the system has only two solutions for $t > 0$ sufficiently small, and one can show that those solutions converge to the two solutions of $\text{in}_w(\mathcal{I}_Q) = \langle -2x_1 - 8, 3x_1^2 + 3x_2^2\rangle$ as t goes to 0. See Remark 3.3 for more details on homotopy continuation under such circumstances.

Alternatively, consider the following two polynomials that form a (tropical) Gröbner basis of $\mathcal{I}_Q$ with respect to w:

$$g_1 := -2tx_1 - 4t^2x_2 - 8 \quad \text{and} \quad g_2 := 3x_1^2 + 5tx_1 + 3x_2^2 + 7t^2x_2 + 11.$$

Using them instead of the $f_{i,Q}$ in Algorithm 3.1 Line 3 gives us

$$\begin{aligned} H^{(w)} &= \Big(t^0 \cdot \big(-2t^{1-1}x_1 - 4t^{2-1}x_2 - 8\big), \\ &\qquad t^2 \cdot \big(3t^{0-2}x_1^2 + 3t^{0-2}x_2^2 + 5t^{1-1}x_1 + 7t^{2-1}x_2 + 11\big)\Big) \\ &= \Big(-2x_1 - 4tx_2 - 8, \;\; 3x_1^2 + 3x_2^2 + 5t^2x_1 + 7t^3x_2 + 11t^2\Big). \end{aligned}$$

At $t = 0$, we obtain a binomial generating set of $\text{in}_w(\mathcal{I}_Q)$ used above.

Unlike Lemma 4.3 however, no fast approach is known for computing the intersection in Lemma 5.3. Unlike the binomial ideal $\hat{\mathcal{I}}_{\text{bin},Q}$ of Lemma 4.3, the nonlinear ideal $\hat{\mathcal{I}}_{\text{nlin},Q}$ in Lemma 5.3 has no easily exploitable structure in general. Computing $\text{Trop}(\hat{\mathcal{I}}_{\text{nlin},Q})$ using current algorithms would require several Gröbner basis computations [BJS+07; MR20].

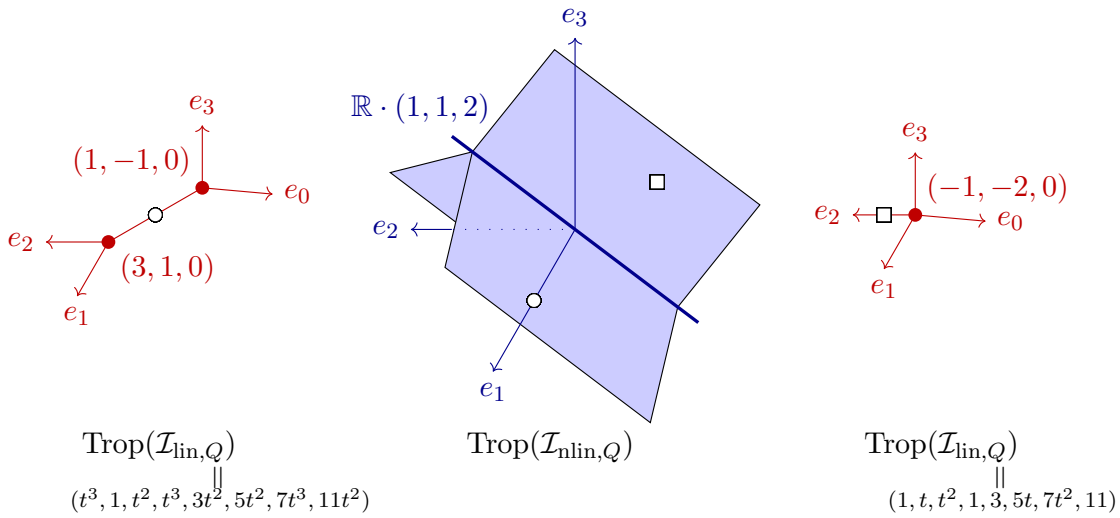


FIGURE 5. The transverse intersections of Example 5.4 illustrated in $\mathbb{R}^3 \cong \{e_{y_2} = e_{x_1}, e_{y_3} = e_{x_2}, e_{y_4} = 0\} \subseteq \mathbb{R}^6$ with $e_1 := e_{x_1}$, $e_2 := e_{x_2}$, $e_3 := e_{y_1}$, and $e_0 := -e_1 - e_2 - e_3$.

Example 5.5. Consider the following horizontally parametrized system that has generic root count 3, polynomial support $R = \{(1+x_1+x_2)^3, (1+x_1+x_2)^2, x_1, 1\}$, and suppose we want to solve it for $P = (1,1,1,1,2,3,5,7) \in \mathbb{C}^8$:

$$\begin{aligned} f_1 &= a_1(1+x_1+x_2)^3 + a_2(1+x_1+x_2)^2 + a_3x_1 + a_4 \\ f_2 &= a_5(1+x_1+x_2)^3 + a_6(1+x_1+x_2)^2 + a_7x_1 + a_8. \end{aligned} \tag{5.3}$$

Its modification is given by

$$\begin{aligned} \hat{f}_1 &= a_1y_1 + a_2y_2 + a_3y_3 + a_4y_4, \quad & \hat{g}_1 &= y_1 - (1+x_1+x_2)^3, \quad & \hat{g}_3 &= y_3 - x_1, \\ \hat{f}_2 &= a_5y_1 + a_6y_2 + a_7y_3 + a_8y_4, \quad & \hat{g}_2 &= y_2 - (1+x_1+x_2)^2, \quad & \hat{g}_4 &= y_4 - 1. \end{aligned}$$

We will continue this example in the upcoming sections.

5.2. **Supports with tropically transverse base.** One way to deal with the difficulty of horizontally parametrized systems is to impose extra conditions on the polynomial support. In [HR23], the authors focus on the following special case of horizontally parametrized systems.

Definition 5.6 ([HR23, Definition 9])**.** We say the polynomial support $R = \{q_1, \dots, q_m\} \subseteq \mathbb{C}[x^{\pm}]$ has a *tropically transverse base*, if there are sets $S = \{b_1, \dots, b_l\} \subseteq \mathbb{C}[x^{\pm}]$ and $B = \{\beta_1, \dots, \beta_l\} \subseteq \mathbb{Z}^m$ such that

(1) $q_j = b^{\beta_j} = \prod_{k=1}^{l} b_k^{\beta_{j,k}}$ for $j \in [m]$,
(2) $\mathrm{Trop}(b_1), \dots, \mathrm{Trop}(b_l)$ intersect transversally.

Definition 5.7 ([HR23, Definition 10])**.** Let $\mathcal{F}$ have a polynomial support with tropically transverse base S, as in Definition 5.6. The *(two-stage) modification* of $\mathcal{F}$ is defined to be

$$\begin{aligned}\hat{\mathcal{F}} &:= \{\hat{f}_i, \hat{g}_j, \hat{h}_k \mid i \in [n], j \in [m], k \in [l]\} \\ &\subseteq \mathbb{C}[a][x^{\pm}, y^{\pm}, z^{\pm}] := \mathbb{C}[a][x_i^{\pm}, y_j^{\pm}, z_k^{\pm} \mid i \in [n], j \in [m], k \in [l]]\end{aligned}$$

where

$$\hat{f}_i := \sum_{j=1}^{m} c_{i,j} a_j z_j, \quad \hat{g}_j := z_j - \prod_{k=1}^{l} y_k^{\beta_{j,k}}, \quad \text{and} \quad \hat{h}_k := y_k - b_k. \tag{5.4}$$

For the remainder of this subsection, we let $\hat{\mathcal{I}} := \langle \hat{\mathcal{F}} \rangle \subseteq \mathbb{C}\{\{t\}\}[a][x^{\pm}, y^{\pm}, z^{\pm}]$ be the parametrized ideal generated by $\hat{\mathcal{F}}$.

Theorem 5.8 ([HR23, Theorem 2])**.** *With the notation above,*

$$\ell_{\hat{\mathcal{I}}, \mathbb{C}(a)} = \mathrm{MV}\left(\Delta(f_i), \Delta(g_j), \Delta(h_k) \mid i \in [n], j \in [m], k \in [l]\right).$$

As a corollary of Theorem 5.8, polyhedral homotopies are optimal for System (5.4). Furthermore, the polyhedral homotopies for the modified System (5.4) can be reformulated to homotopies for the original System (5.1) by back-substituting the modification variables, as illustrated by the next example.

Example 5.9. The polynomial support of System (5.3) in Example 5.5 has the tropically transverse base

$$S = \{1 + x_1 + x_2, x_1\}, \quad B = \{(3,0), (2,0), (0,1), (0,0)\}.$$

Its two-stage modification is thus given by

$$\begin{aligned} \hat{f}_1 &= a_1 z_1 + a_2 z_2 + a_3 z_3 + a_4 z_4, & \hat{g}_1 &= z_1 - y_1^3, & \hat{h}_1 &= y_1 - (1 + x_1 + x_2), \\ \hat{f}_2 &= a_5 z_1 + a_6 z_2 + a_7 z_3 + a_8 z_4, & \hat{g}_2 &= z_2 - y_1^2, & \hat{h}_2 &= y_2 - x_1. \\ & & \hat{g}_3 &= z_3 - y_2, & & \\ & & \hat{g}_4 &= z_4 - 1. & & \end{aligned} \tag{5.5}$$

By Theorem 5.8, the generic root count of the modified system is its mixed volume, and polyhedral homotopies are optimal for solving it.

Recall from Example 3.2 that polyhedral homotopies require choosing valuations for all coefficients. For System (5.5), it actually suffices to simply choose $Q = (t^4, t^2, 1, 1, 2t^{11}, 3t^7, 5, 7t)$, which yields six tropical hypersurfaces that intersect

transversally in two points:

$$\Big(\bigwedge_{i=1}^{2}\operatorname{Trop}(\hat{f}_{i,Q})\Big)\wedge\Big(\bigwedge_{j=1}^{4}\operatorname{Trop}(\hat{g}_{j,Q})\Big)\wedge\Big(\bigwedge_{k=1}^{2}\operatorname{Trop}(\hat{h}_{k,Q})\Big)$$
$$=\Big\{(1,-1,-1,1,-3,-2,1,0),(1,-2,-2,1,-6,-4,1,0)\Big\}.$$

For $w=(1,-1,-1,1,-3,-2,1,0)$, the resulting homotopies from Algorithm 3.1 Line 3 are:

$$\hat{h}_1^{(w)}=tz_1+z_2+tz_3+z_4,\qquad \hat{h}_3^{(w)}=z_1-y_1^3,\quad \hat{h}_7^{(w)}=y_1-(t+t^2x_1+x_2),$$
$$\hat{h}_2^{(w)}=2t^7z_1+3t^4z_2+5z_3+7z_4,\quad \hat{h}_4^{(w)}=z_2-y_1^2,\quad \hat{h}_8^{(w)}=y_2-x_1.$$
$$\hat{h}_5^{(w)}=z_3-y_2,$$
$$\hat{h}_6^{(w)}=z_4-1.$$

Note that we can undo the modification by substituting the y and the z to obtain a homotopy involving only the x:

$$h_1^{(w)}=t(t+t^2x_1+x_2)^3+(t+t^2x_1+x_2)^2+tx_1+1,$$
$$h_2^{(w)}=2t^7(t+t^2x_1+x_2)^3+3t^4(t+t^2x_1+x_2)^2+5x_1+7,$$

which for $t=0$ yields a binomial system with 2 solutions in $(\mathbb{C}^*)^2$ (the latter may be seen easier by specializing the $\hat{h}_i^{(w)}$ at $t=0$).

Similarly, for $w=(1,-2,-2,1,-6,-4,1,0)$, we obtain a homotopy that for $t=0$ yields a binomial system with 1 solution in $(\mathbb{C}^*)^2$. The total number of starting solutions therefore equals the generic root count of System (5.3), which is 3.

5.3. **Support relaxation.** Another way to deal with the difficulty of horizontally parametrized systems, besides imposing extra conditions as in Section 5.2, is by embedding them into a larger and easier parametrized family in the sense of Proposition 2.7. For instance, the following relaxation ensures that the modification is Bernstein generic.

Definition 5.10. Suppose that the polynomial support $R=\{q_1,\dots,q_m\}$ has the form $q_j=\sum_{k=1}^{l}q_{j,k}x^{\alpha_k}$ with $q_{j,k}\in\mathbb{C}$ for some finite set of exponent vectors $S:=\{\alpha_1,\dots,\alpha_l\}\subseteq\mathbb{Z}^n$ and some coefficients $q_{j,k}\in\mathbb{C}$ (note that we may have $q_{j,k}=0$). The *relaxation* of $\mathcal{F}$ is the parametrized system

$$\mathcal{F}^\sharp:=\{f_1^\sharp,\dots,f_n^\sharp\}\subseteq\mathbb{C}[a,b][x^\pm]:=\mathbb{C}[a_{i,j},b_{j,k}\mid i\in[n],j\in[m],k\in[l]][x_1^\pm,\dots,x_n^\pm],$$

where

$$f_i^\sharp:=\sum_{j=1}^{m}c_{i,j}a_{i,j}\underbrace{\sum_{k=1}^{l}q_{j,k}b_{j,k}x^{\alpha_k}}_{=:q_j^\sharp}.$$

We use $\mathcal{I}^\sharp \subseteq \mathbb{C}\{\{t\}\}[a,b][x^{\pm}]$ to denote the ideal generated by $\mathcal{F}^\sharp$.

Definition 5.11. The *modification* of $\mathcal{F}^\sharp$ is the system

$$\hat{\mathcal{F}}^\sharp \coloneqq \{\hat{f}_i^\sharp, \hat{g}_j^\sharp \mid i \in [n], j \in [m]\} \subseteq \mathbb{C}[a,b][x^{\pm}, y^{\pm}] \coloneqq \mathbb{C}[a,b][x_i^{\pm}, y_j^{\pm} \mid i \in [n], j \in [m]].$$

where

$$\hat{f}_i^\sharp \coloneqq \sum_{j=1}^{m} c_{i,j} a_{i,j} y_j \quad \text{and} \quad \hat{g}_j^\sharp \coloneqq y_j - q_j^\sharp.$$

We use $\hat{\mathcal{I}}_{\text{lin}}^\sharp$ and $\hat{\mathcal{I}}_{\text{nlin}}^\sharp$ to denote the ideals in $\mathbb{C}\{\{t\}\}[a,b][x^{\pm}, y^{\pm}]$ generated by the $\hat{f}_i^\sharp$ and $\hat{g}_j^\sharp$, respectively, and set $\hat{\mathcal{I}}^\sharp = \hat{\mathcal{I}}_{\text{lin}}^\sharp + \hat{\mathcal{I}}_{\text{nlin}}^\sharp$.

Corollary 5.12. *We have* $\ell_{\hat{\mathcal{I}}^\sharp, \mathbb{C}(a)} = \text{MV}(\hat{f}_i^\sharp, \hat{g}_j^\sharp \mid i \in [n], j \in [m])$.

Proof. This follows directly from Theorem 2.14. □

Example 5.13. The relaxed horizontal modification for the horizontal system form Example 5.5 is given by

$$\begin{aligned}
\hat{f}_1 &= a_1 y_1 + a_2 y_2 + a_3 y_3 + a_4 y_4, \\
\hat{f}_2 &= a_5 y_1 + a_6 y_2 + a_7 y_3 + a_8 y_4, \\
\hat{g}_1 &= y_1 - (b_{1,1} x_1^3 + 3b_{1,2} x_2 x_1^2 + 3b_{1,3} x_1^2 + 3b_{1,4} x_2^2 x_1 + 6b_{1,5} x_2 x_1 + 3b_{1,6} x_1 \\
&\qquad + b_{1,7} x_2^3 + 3b_{1,8} x_2^2 + 3b_{1,9} x_2 + b_{1,10}), \\
\hat{g}_2 &= y_2 - (b_{2,1} x_1^2 + 2b_{2,2} x_2 x_1 + 2b_{2,4} x_1 + b_{2,5} x_2^2 + 2b_{2,6} x_2 + b_{2,7}), \\
\hat{g}_3 &= y_3 - b_{3,1} x_1, \\
\hat{g}_4 &= y_4 - b_{4,1}.
\end{aligned} \tag{5.6}$$

By Corollary 5.12, the generic root count of the relaxed modification is equal to its mixed volume, which is 6, exceeding the generic root count of the original System (5.3) but staying below the original mixed volume of 9. Similar as in Example 5.9, we can construct polyhedral homotopies for System (5.6) and undo the modification by substituting the y to obtain a homotopy purely in the x, tracing only 6 paths instead of the 9 paths of polyhedral homotopies.

We conclude this section with a short proof that the generic root count of relaxed system never exceeds the mixed volume of the original system, i.e., that the homotopies that one obtains from the polyhedral homotopies of the modified system will never involve more paths than the polyhedral homotopies of the original system.

Proposition 5.14. *Let $\mathcal{I}$ be the horizontally parametrized ideal from Definition 5.1, let $\hat{\mathcal{I}}$ be its modification from Definition 5.2, let $\mathcal{I}^\sharp$ be its relaxation from Definition 5.10, and let $\hat{\mathcal{I}}^\sharp$ be its relaxed modification from Definition 5.11. Then the*

generic root counts satisfy the following chain of inequalities:

$$\ell_{\mathcal{I}:(\prod_{i=1}^m q_i)^\infty,\mathbb{C}(a)} = \ell_{\hat{\mathcal{I}},\mathbb{C}(a)} \leq \ell_{\hat{\mathcal{I}}^\sharp,\mathbb{C}(a,b)} \leq \ell_{\mathcal{I}^\sharp,\mathbb{C}(a,b)} \leq \mathrm{MV}(\mathcal{F}).$$

Proof. The first equation follows from Lemma 5.3, whereas the first and third inequality follow from Proposition 2.7. For the second inequality, notice that for all $Q \in (\mathbb{C}^*)^m$ there is an isomorphism

$$K[x^{\pm}] \Big/ \mathcal{I}_Q^\sharp \colon \big(\textstyle\prod_{j=1}^m q_{j,Q}\big)^\infty \longrightarrow K[x^{\pm}, y^{\pm}]/\hat{\mathcal{I}}_Q^\sharp, \qquad \overline{x} \longmapsto \overline{x},$$

which can be proven with the same arguments as in the proof of Lemma 5.3, and that by definition root counts can only decrease with saturation. In other words, for all $Q \in (\mathbb{C}^*)^m$ the root count of $\hat{\mathcal{I}}_Q^\sharp$ is smaller or equal to that of $\mathcal{I}_Q^\sharp$. □

Remark 5.15. Using [HR22, Proposition 5.16], one can also show that if $\mathcal{I}$ is generically zero-dimensional, then $\hat{\mathcal{I}}^\sharp$ has the same generic root count as $\mathcal{I}^\sharp$.

6. Case Studies

In this section, we revisit several examples of parametrized systems in existing literature, and show how our techniques can be used in their context. Notebooks with computations for the case studies can be found in the repository

https://github.com/oskarhenriksson/TropicalHomotopies.jl .

6.1. **Steady states of chemical reaction networks.** Consider the chemical reaction network of the WNT pathway as studied in [GHRS16]:

$$\begin{aligned}
&\mathrm{X}_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{X}_2 && \mathrm{X}_3 + \mathrm{X}_6 \underset{k_{15}}{\overset{k_{14}}{\rightleftharpoons}} \mathrm{X}_{15} \xrightarrow{k_{16}} \mathrm{X}_3 + \mathrm{X}_7\\
&\mathrm{X}_2 + \mathrm{X}_4 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathrm{X}_{14} \xrightarrow{k_5} \mathrm{X}_2 + \mathrm{X}_5 && \mathrm{X}_7 + \mathrm{X}_9 \underset{k_{18}}{\overset{k_{17}}{\rightleftharpoons}} \mathrm{X}_{17} \xrightarrow{k_{19}} \mathrm{X}_6 + \mathrm{X}_9\\
&\mathrm{X}_5 + \mathrm{X}_8 \underset{k_7}{\overset{k_6}{\rightleftharpoons}} \mathrm{X}_{16} \xrightarrow{k_8} \mathrm{X}_4 + \mathrm{X}_8 && \mathrm{X}_6 + \mathrm{X}_{11} \underset{k_{21}}{\overset{k_{20}}{\rightleftharpoons}} \mathrm{X}_{19} \xrightarrow{k_{22}} \mathrm{X}_6 + \emptyset\\
&\mathrm{X}_4 + \mathrm{X}_{10} \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} \mathrm{X}_{18} \xrightarrow{k_{11}} \mathrm{X}_4 + \emptyset && \mathrm{X}_{11} \xrightarrow{k_{23}} \emptyset\\
&\emptyset \xrightarrow{k_{12}} \mathrm{X}_{10} && \mathrm{X}_{11} + \mathrm{X}_{12} \underset{k_{25}}{\overset{k_{24}}{\rightleftharpoons}} \mathrm{X}_{13}\\
&\mathrm{X}_{10} \xrightarrow{k_{13}} \emptyset && \mathrm{X}_2 \underset{k_{27}}{\overset{k_{26}}{\rightleftharpoons}} \mathrm{X}_3\\
&\mathrm{X}_5 \underset{k_{29}}{\overset{k_{28}}{\rightleftharpoons}} \mathrm{X}_7 && \mathrm{X}_{10} \underset{k_{31}}{\overset{k_{30}}{\rightleftharpoons}} \mathrm{X}_{11}.
\end{aligned}$$

The steady states of chemical reaction networks under mass action kinetics can be described by systems that are close to being vertically parametrized (in [FHP23] these systems are referred to as *augmented* vertically parametrized systems, see also [RT24] for a discussion from the point of view of positive tropicalizations).

For the WNT pathway, we obtain a square parametrized polynomial system with 19 variables x and 36 parameters k, c:

$$\begin{aligned}
f_1 &= k_1x_1 - (k_2 + k_{26})x_2 + k_{27}x_3 - k_3x_2x_4 + (k_4 + k_5)x_{14}\\
f_2 &= k_{26}x_2 - k_{27}x_3 - k_{14}x_3x_6 + (k_{15} + k_{16})x_{15}\\
f_3 &= -k_{28}x_5 + k_{29}x_7 - k_6x_5x_8 + k_5x_{14} + k_7x_{16}\\
f_4 &= -k_{14}x_3x_6 - k_{20}x_6x_{11} + k_{15}x_{15} + k_{19}x_{17} + (k_{21} + k_{22})x_{19}\\
f_5 &= k_{28}x_5 - k_{29}x_7 - k_{17}x_7x_9 + k_{16}x_{15} + k_{18}x_{17}\\
f_6 &= k_{12} - (k_{13} + k_{30})x_{10} - k_9x_4x_{10} + k_{31}x_{11} + k_{10}x_{18}\\
f_7 &= -k_{23}x_{11} + k_{30}x_{10} - k_{31}x_{11} - k_{20}x_6x_{11} - k_{24}x_{11}x_{12} + k_{25}x_{13} + k_{21}x_{19}\\
f_8 &= k_{24}x_{11}x_{12} - k_{25}x_{13}\\
f_9 &= k_3x_2x_4 - (k_4 + k_5)x_{14}\\
f_{10} &= k_{14}x_3x_6 - (k_{15} + k_{16})x_{15}\\
f_{11} &= -k_6x_5x_8 + (k_7 + k_8)x_{16}\\
f_{12} &= k_{17}x_7x_9 - (k_{18} - k_{19})x_{17}\\
f_{13} &= k_9x_4x_{10} - (k_{10} + k_{11})x_{18}\\
f_{14} &= k_{20}x_6x_{11} - (k_{21} + k_{22})x_{19}\\
f_{15} &= x_1 + x_2 + x_3 + x_{14} + x_{15} - c_1\\
f_{16} &= x_4 + x_5 + x_6 + x_7 + x_{14} + x_{15} + x_{16} + x_{17} + x_{18} + x_{19} - c_2\\
f_{17} &= x_8 + x_{16} - c_3\\
f_{18} &= x_9 + x_{17} - c_4\\
f_{19} &= x_{12} + x_{13} - c_5.
\end{aligned} \tag{6.1}$$

The mixed volume of System (6.1) is 56, but the generic root count is only 9 [GHRS16, Theorem 1.1]. We can embed System (6.1) into a vertically parametrized family by introducing new parameters for all monomials of the so-called *conservation laws* $f_{15}, \dots, f_{19}$ that lack parametric coefficients, and Algorithm 3.1 will construct homotopies with exactly 9 paths (see the GitHub repository for examples of such homotopies). The general fact that this type of embedding does not increase the generic root count for any steady state system is the subject of an upcoming preprint.

The main bottleneck is the computation of the tropical intersection data for Algorithm 3.1 via Lemma 4.3. It consists of:

(1) the computation of $\mathrm{Trop}(\hat{\mathcal{I}}_{Q,\mathrm{lin}})$, which consists of 78 983 maximal polyhedra,
(2) the intersection $\mathrm{Trop}(\hat{\mathcal{I}}_{Q,\mathrm{lin}}) \wedge \mathrm{Trop}(\hat{\mathcal{I}}_{Q,\mathrm{lin}})$, which consists of 9 points counted with multiplicity.

The bottlenecks (1) and (2) required 43 seconds and 21 seconds, respectively, on a MacBook Air with an Apple M2 chip and 16 GB of RAM. We note that the matroid associated with $\mathrm{Trop}(\hat{\mathcal{I}}_{Q,\mathrm{lin}})$ is non-transversal, so Remark 4.7 does not apply.

6.2. **Periodic solutions to Duffing oscillators.** In this section, we consider equations arising from coupled driven non-linear oscillators as discussed by Borovik, Breiding, del Pino, Michałek, and Zilberberg in their work on semimixed systems of polynomial equations [BBP+23]. To be precise, we consider the following parametrized system in [BBP+23, Section 6.2] with variables u_i, v_i and parameters ω_i, α, γ, λ:

$$\begin{aligned}
\left(\frac{\omega_0^2-\omega_1^2}{2}-\frac{\lambda\omega_0^2}{4}\right)u_1+\frac{\gamma\omega_1}{2}v_1+\frac{3}{8}\alpha u_1(u_1^2+v_1^2)+\frac{3}{4}\alpha u_1(u_2^2+v_2^2)=0,\\
\left(\frac{\omega_0^2-\omega_1^2}{2}+\frac{\lambda\omega_0^2}{4}\right)v_1-\frac{\gamma\omega_1}{2}u_1+\frac{3}{8}\alpha v_1(u_1^2+v_1^2)+\frac{3}{4}\alpha v_1(u_2^2+v_2^2)=0,\\
\frac{(\omega_0^2-\omega_2^2)}{2}u_2+\frac{\gamma\omega_2}{2}v_2+\frac{3}{8}\alpha u_2(u_2^2+v_2^2)+\frac{3}{4}\alpha u_2(u_1^2+v_1^2)=0,\\
\frac{(\omega_0^2-\omega_2^2)}{2}v_2-\frac{\gamma\omega_2}{2}u_2+\frac{3}{8}\alpha v_2(u_2^2+v_2^2)+\frac{3}{4}\alpha v_2(u_1^2+v_1^2)=0.
\end{aligned}$$

In [BBP+23, Section 6.2], the authors relax the system to the following parametrized polynomial system with variables u_i, v_i and parameters $a_{i,j}$:

$$\begin{aligned}
f_1 &:= a_{1,0}+a_{1,1}u_1+a_{1,2}v_1+a_{1,3}u_1(u_1^2+v_1^2)+a_{1,4}u_1(u_2^2+v_2^2),\\
f_2 &:= a_{2,0}+a_{2,1}u_1+a_{2,2}v_1+a_{2,3}v_1(u_1^2+v_1^2)+a_{2,4}v_1(u_2^2+v_2^2),\\
f_3 &:= a_{3,0}+a_{3,1}u_2+a_{3,2}v_2+a_{3,3}u_2(u_1^2+v_1^2)+a_{3,4}u_2(u_2^2+v_2^2),\\
f_4 &:= a_{4,0}+a_{4,1}u_2+a_{4,2}v_2+a_{3,3}v_2(u_1^2+v_1^2)+a_{4,4}v_2(u_2^2+v_2^2),
\end{aligned} \tag{6.2}$$

and use their theory to deduce an upper bound of 25 solutions for $\mathbb{C}^4$.

One way to apply our techniques to System (6.2) is to note that it is a horizontal system with tropically transverse base. By omitting some of the new variables y, z, its two-stage modification can be simplified to:

$$\begin{aligned}
&\hat{f}_1 := a_{1,0}+a_{1,1}u_1+a_{1,2}v_1+a_{1,3}z_{1,1}+a_{1,4}z_{1,2} && \hat{g}_1 := z_{1,1}-u_1y_1 && \hat{h}_1 := z_{1,2}-u_1y_2\\
&\hat{f}_2 := a_{2,0}+a_{2,1}u_1+a_{2,2}v_1+a_{2,3}z_{2,1}+a_{2,4}z_{2,2} && \hat{g}_2 := z_{2,1}-v_1y_1 && \hat{h}_2 := z_{2,2}-v_1y_2\\
&\hat{f}_3 := a_{3,0}+a_{3,1}u_2+a_{3,2}v_2+a_{3,3}z_{3,1}+a_{3,4}z_{3,2} && \hat{g}_3 := z_{3,1}-u_2y_1 && \hat{h}_3 := z_{3,2}-u_2y_2\\
&\hat{f}_4 := a_{4,0}+a_{4,1}u_2+a_{4,2}v_2+a_{4,3}z_{4,1}+a_{4,4}z_{4,2} && \hat{g}_4 := z_{4,1}-v_2y_1 && \hat{h}_4 := z_{4,2}-v_2y_2\\
&\hat{k}_1 := y_1-(u_1^2+v_1^2) && \hat{k}_2 := y_2-(u_2^2+v_2^2)
\end{aligned} \tag{6.3}$$

By Theorem 5.8, the generic root count is the mixed volume of the system above, which reconfirms the upper bound of 25 in [BBP+23]. Explicit homotopies constructed from this modification can be found in the GitHub repository.

We note that while the techniques in [BBP+23] require the polynomial support to have a constant term (see [BBP+23, Theorem 3.8]), our techniques do not. Hence,

if one is only interested in the solutions inside the torus $(\mathbb{C}^*)^4$, one can set $a_{i,0} = 0$ in Systems (6.2) and (6.3) to obtain a smaller upper bound of 16.

6.3. **Equilibria of the Kuramoto model.** In this section, we consider the Kuramoto equations with phase delays as discussed by Chen, Korchevskaia, and Lindberg in [CKL22, Section 2.4]. For a graph G with vertices $[n]$, these form a parametrized system with variables x and parameters $\overline{w}$, a, C:

$$f_{G,i} = \overline{w}_i - \sum_{j \in N_G(i)} a_{ij}\Big(\frac{x_i C_{ij}}{x_j} - \frac{x_j}{x_i C_{ij}}\Big) \quad \text{for } i \in [n], \tag{6.4}$$

where $N_G(i)$ denotes the set of vertices adjacent to the vertex i.

In [CKL22, Corollary 3.9], the authors show that the number of solutions equals the (normalized) volume of the adjacency polytope of G. This result is beyond a simple application of our techniques. However, we can easily express the generic root count as a mixed volume, by viewing System (6.4) as a horizontal system with relaxed polynomial support as discussed in Section 5.3. Consider its modification:

$$\begin{aligned} \hat{f}_{G,i} &= \overline{w}_i - \sum_{j \in N_G(i)} a_{ij} y_{ij} && \text{for } i \in [n], \\ \hat{g}_{ij} &= y_{ij} - \Big(\frac{x_i C_{ij}}{x_j} - \frac{x_j}{x_i C_{ij}}\Big) && \text{for } \{ij\} \in E(G), \end{aligned} \tag{6.5}$$

and observe that all resulting tropical hypersurfaces intersect transversally. Hence, the generic root count of System (6.4) is the mixed volume of System (6.5).

6.4. **Realizations of Laman graphs.** In [CGG+18], the authors study the problem of finding realizations of minimally 2-rigid so-called *Laman graphs*, given generic assignments of the edge lengths. Such realizations correspond to solutions of a parametrized system consisting of equations that are either linear or linear in the reciprocals of the variables, which can readily be treated with our techniques.

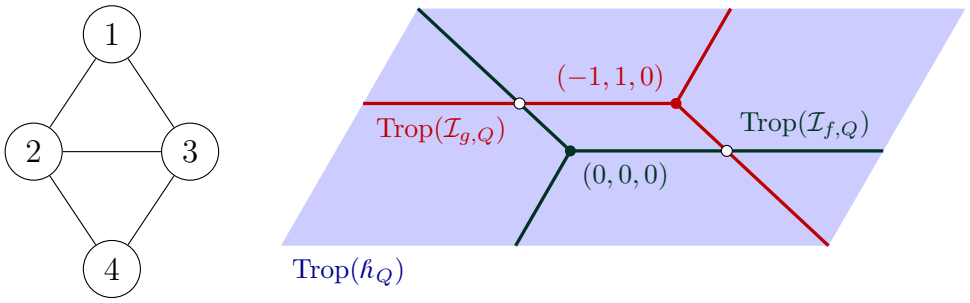
As an example, consider the Laman graph G in Figure 6(a), for which the realizations correspond to the solutions of the system

$$\begin{aligned} f_1 &\coloneqq x_{12} + x_{23} - x_{13}, \quad g_1 \coloneqq \lambda_{12} x_{12}^{-1} + \lambda_{23} x_{23}^{-1} - \lambda_{13} x_{13}^{-1}, \quad \hbar \coloneqq x_{23} - 1, \\ f_2 &\coloneqq x_{23} + x_{34} - x_{24}, \quad g_2 \coloneqq \lambda_{23} x_{23}^{-1} + \lambda_{34} x_{34}^{-1} - \lambda_{24} x_{24}^{-1}. \end{aligned} \tag{6.6}$$

with variables x_{ij} encoding the edge directions, and parameters λ_{ij} encoding the edge lengths. Note that this is a slightly simplified version of the system in [CGG+18, Example 2.25], where we have substituted y_{ij} by $\lambda_{ij} x_{ij}^{-1}$, and x_{ji} by $-x_{ij}$ for $i < j$.

Consider the parametrized ideals $\mathcal{I} \coloneqq \langle f_1, f_2, g_1, g_2, h_Q \rangle$, $\mathcal{I}_f \coloneqq \langle f_1, f_2 \rangle$, and $\mathcal{I}_g \coloneqq \langle g_1, g_2 \rangle$, and let $P = (1,1,1,1,1)$ be the target parameters. To compute the tropical data required for constructing the homotopies, note that for $Q = t^v \cdot P$ for generic $v \in \mathbb{Z}^{E(G)}$, it holds that

$$\operatorname{Trop}(\mathcal{I}_Q) = \operatorname{Trop}(\mathcal{I}_{f,Q}) \wedge \operatorname{Trop}(\mathcal{I}_{g,Q}) \wedge \operatorname{Trop}(\hbar_Q),$$



(a) (b)

FIGURE 6. A Laman graph and its resulting tropical intersection

since the intersection is transversal, as sketched in Figure 6(b). For $v_{ij} = i + j$, we obtain the following four intersection points, each of multiplicity 1:

$$(0,1,0,1,0),\ (0,1,0,0,2),\ (-2,-2,0,1,0),\ (-2,-2,0,0,2) \in \mathbb{R}^{\{12,13,23,24,34\}}.$$

This shows that the generic root count is equal to 4, and the four intersection points can be used to compute the solutions of System (6.6) for the parameters P through Algorithm 3.1. For example, the intersection point $w = e_{13}+e_{24}$ yields the homotopy

$$H^{(w)} = \Big(x_{12}+x_{23}-tx_{13}, x_{23}+x_{34}-tx_{24}, t^2x_{23}^{-1}+x_{12}^{-1}-x_{13}^{-1}, t^2x_{34}^{-1}+x_{23}^{-1}-x_{24}^{-1}, -1+x_{23}\Big),$$

and the binomial start system

$$\Big(x_{12}+x_{23}, x_{23}+x_{34}, x_{12}^{-1}-x_{13}^{-1}, x_{23}^{-1}-x_{24}^{-1}, -1+x_{23}\Big),$$

which has a unique solution in the torus. Homotopies for the other intersection points can be found in our GitHub repository.

7. Conclusion and outlook

In this work, we have explained how to construct generically optimal homotopies for parametrized polynomial systems from their tropical data, and listed conditions on the systems that result in desirable properties of the resulting homotopies (Section 3). We then discussed how said tropical data can be obtained for two classes of parametrized polynomial systems: vertically parametrized systems (Section 4) and horizontally parametrized systems (Section 5).

The data of vertically parametrized systems can be obtained from the intersection of a tropical linear space and a tropical binomial variety. In contrast, the data of horizontally parametrized systems is more difficult to obtain, and we have proposed two relaxations. Finally, we have highlighted several examples from the literature where our techniques can be applied (Section 6).

Two main challenges for the efficient application of our techniques remain:

(1) develop and implement tropical homotopy continuation to speed up the computation of the required tropical data (see timings in Section 6.1; ongoing work, see [DR24] for a preliminary paper),

(2) implement a numerically robust path tracker for tropical homotopies.

Another future research direction is to identify more classes of parametrized systems whose generic root count is below their mixed volume and whose tropical data is efficiently computable. Such classes are important targets for relaxation, independent of whether they arise directly in practise.

References

[BBC+23] D. J. Bates, P. Breiding, T. Chen, J. D. Hauenstein, A. Leykin, and F. Sottile. *Numerical Nonlinear Algebra*. 2023. eprint: arXiv:2302.08585v2.

[BBP+23] V. Borovik, P. Breiding, J. del Pino, M. Michałek, and O. Zilberberg. *Khovanskii bases for semimixed systems of polynomial equations – a case of approximating stationary nonlinear Newtonian dynamics*. 2023. eprint: arXiv:2306.07897v1.

[BC11] P. Bürgisser and F. Cucker. "On a problem posed by Steve Smale". In: *Ann. Math. (2)* 174.3 (2011), pp. 1785–1836. DOI: 10.4007/annals.2011.174.3.8.

[BDH24] E. Boniface, K. Devriendt, and S. Hoşten. *Tropical toric maximum likelihood estimation*. 2024. eprint: arXiv:2404.10567v1.

[Ber76] D. N. Bernstein. "The number of roots of a system of equations". In: *Funct. Anal. Appl.* 9 (1976), pp. 183–185. DOI: 10.1007/BF01075595.

[BJS+07] T. Bogart, A. N. Jensen, D. Speyer, B. Sturmfels, and R. R. Thomas. "Computing Tropical Varieties". In: *Journal of Symbolic Computation* 42.1-2 (2007), pp. 54–73. DOI: 10.1016/j.jsc.2006.02.004.

[BJSST07] T. Bogart, A. N. Jensen, D. Speyer, B. Sturmfels, and R. R. Thomas. "Computing tropical varieties". In: *J. Symb. Comput.* 42.1-2 (2007), pp. 54–73. DOI: 10.1016/j.jsc.2006.02.004.

[BKK+21] C. Borger, T. Kahle, A. Kretschmer, S. Sager, and J. Schulze. *Liftings of polynomial systems decreasing the mixed volume*. 2021. eprint: arXiv:2105.10714v1.

[BLMM17] L. Bossinger, S. Lamboglia, K. Mincheva, and F. Mohammadi. "Computing Toric Degenerations of Flag Varieties". In: *Combinatorial Algebraic Geometry: Selected Papers From the 2016 Apprenticeship Program*. Ed. by G. G. Smith and B. Sturmfels. Springer New York, 2017, pp. 247–281. DOI: 10.1007/978-1-4939-7486-3_12.

[Bon10] J. E. Bonin. *An Introduction to Transversal Matroids*. 2010. URL: https://bpb-us-e1.wpmucdn.com/blogs.gwu.edu/dist/3/152/files/2016/04/TransversalNotes-2ltphpb.pdf.

[BP11] C. Beltrán and L. M. Pardo. "Fast linear homotopy to find approximate zeros of polynomial systems". In: *Found. Comput. Math.* 11.1 (2011), pp. 95–129. DOI: 10.1007/s10208-010-9078-9.

[BSW22] P. Breiding, F. Sottile, and J. Woodcock. "Euclidean distance degree and mixed volume". In: *Found. Comput. Math.* 22.6 (2022), pp. 1743–1765. DOI: s10208-021-09534-8.

[BSW23] M. Burr, F. Sottile, and E. Walker. "Numerical homotopies from Khovanskii bases". In: *Mathematics of Computation* 92.343 (2023), pp. 2333–2353.

[BT18] P. Breiding and S. Timme. "HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia". In: *International Congress on Mathematical Software*. Springer. 2018, pp. 458–465. DOI: 10.1007/978-3-319-96418-8_54.

[CGG+18] J. Capco, M. Gallet, G. Grasegger, C. Koutschan, N. Lubbes, and J. Schicho. "The Number of Realizations of a Laman Graph". In: *SIAM Journal on Applied Algebra and Geometry* 2.1 (2018), pp. 94–125. DOI: 10.1137/17M1118312.

[CKL22] T. Chen, E. Korchevskaia, and J. Lindberg. *On the Typical and Atypical Solutions to the Kuramoto Equations*. 2022. eprint: arxiv:2210.00784v2.

[CL14] T. Chen and T.-Y. Li. "Solutions to systems of binomial equations". In: *Ann. Math. Sil.* 28 (2014), pp. 7–34. URL: www.sbc.org.pl/Content/129017/zip/.

[CLL14] T. Chen, T.-L. Lee, and T.-Y. Li. "Hom4PS-3: a parallel numerical solver for systems of polynomial equations based on polyhedral homotopy continuation methods". In: *Mathematical software – ICMS 2014. 4th international congress, Seoul, South Korea, August 5–9, 2014. Proceedings*. Berlin: Springer, 2014, pp. 183–190. DOI: 10.1007/978-3-662-44199-2_30.

[CLO05] D. A. Cox, J. Little, and D. O'Shea. *Using algebraic geometry*. 2nd ed. Vol. 185. Grad. Texts Math. New York, NY: Springer, 2005. DOI: 10.1007/b138611.

[CM19] A. J. Chan and D. Maclagan. "Gröbner bases over fields with valuations". In: *Math. Comp.* 88 (2019), pp. 467–483. DOI: 10.1090/mcom/3321.

[CMMN19] T. Chen, J. Mareček, D. Mehta, and M. Niemerg. "Three Formulations of the Kuramoto Model as a System of Polynomial Equations". In: *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 2019, pp. 810–815. DOI: 10.1109/ALLERTON.2019.8919934.

[Cox20] D. A. Cox. *Applications of polynomial systems. With contributions by Carlos D'Andrea, Alicia Dickenstein, Jonathan Hauenstein, Hal Schenck, and Jessica Sidman*. Vol. 134. CBMS Reg. Conf. Ser. Math. Providence, RI: American Mathematical Society (AMS), published for the Conference Board of the Mathematical Sciences (CBMS), 2020. DOI: 10.1090/cbms/134.

[DGPS24] W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann. SINGULAR *4-4-0 — A computer algebra system for polynomial computations*. Available at http://www.singular.uni-kl.de. 2024.

[Dic16] A. Dickenstein. "Biochemical reaction networks: an invitation for algebraic geometers". In: *Mathematical congress of the Americas. First mathematical congress of the Americas, Guanajuato, México, August 5–9, 2013*. Providence, RI: American Mathematical Society (AMS), 2016, pp. 65–83. DOI: 10.1090/conm/656/13076.

[DR24] O. Daisey and Y. Ren. "A Framework for Generalized Tropical Homotopy Continuation". In: *Mathematical Software – ICMS 2024*. Ed. by K. Buzzard, A. Dickenstein, B. Eick, A. Leykin, and Y. Ren. Cham: Springer Nature Switzerland, 2024, pp. 331–339. DOI: 10.1007/978-3-031-64529-7_32.

[DTWY24] T. Duff, S. Telen, E. Walker, and T. Yahl. "Polyhedral homotopies in Cox coordinates". In: *Journal of Algebra and Its Applications* 23.04 (2024), p. 2450073. DOI: 10.1142/S0219498824500737.

[ES96] D. Eisenbud and B. Sturmfels. "Binomial ideals". In: *Duke Math. J.* 84.1 (1996), pp. 1–45. DOI: 10.1215/S0012-7094-96-08401-X.

[FHP23] E. Feliu, O. Henriksson, and B. Pascual-Escudero. *Generic consistency and nondegeneracy of vertically parametrized systems*. 2023. eprint: arXiv:2304.02302v4.

[Gfan] A. N. Jensen. *Gfan, a software system for Gröbner fans and tropical varieties*. Available at http://home.imf.au.dk/jensen/software/gfan/gfan.html.

[GHRS16] E. Gross, H. A. Harrington, Z. Rosen, and B. Sturmfels. "Algebraic systems biology: a case study for the Wnt pathway". In: *Bull. Math. Biol.* 78.1 (2016), pp. 21–51. DOI: 10.1007/s11538-015-0125-1.

[GJ00] E. Gawrilow and M. Joswig. "polymake: a framework for analyzing convex polytopes". In: *Polytopes - combinatorics and computation. DMV-seminar Oberwolfach, Germany, November 1997*. Basel: Birkhäuser, 2000, pp. 43–73. DOI: 10.1007/978-3-0348-8438-9_2.

[HK12] D. Helm and E. Katz. "Monodromy filtrations and the topology of tropical varieties". In: *Can. J. Math.* 64.4 (2012), pp. 845–868. DOI: 10.4153/CJM-2011-067-9.

[HR22] P. A. Helminck and Y. Ren. *Generic root counts and flatness in tropical geometry*. 2022. eprint: arXiv:2206.07838v2.

[HR23] I. Holt and Y. Ren. *Generic root counts of tropically transverse systems – An invitation to tropical geometry in OSCAR*. 2023. eprint: arXiv:2311.18018v1.

[HS95] B. Huber and B. Sturmfels. "A polyhedral method for solving sparse polynomial systems". In: *Math. Comput.* 64.212 (1995), pp. 1541–1555. DOI: 10.2307/2153370.

[Jen16] A. N. Jensen. *Tropical Homotopy Continuation*. 2016. eprint: arXiv:1601.02818v1.

[KK12] K. Kaveh and A. G. Khovanskii. "Newton-Okounkov bodies, semigroups of integral points, graded algebras and intersection theory". In: *Ann. Math. (2)* 176.2 (2012), pp. 925–978. DOI: 10.4007/annals.2012.176.2.5.

[Lai17] P. Lairez. "A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time". In: *Found. Comput. Math.* 17.5 (2017), pp. 1265–1292. DOI: 10.1007/s10208-016-9319-7.

[Li99] T.-Y. Li. "Solving polynomial systems by polyhedral homotopies". In: *Taiwanese J. Math.* 3.3 (1999), pp. 251–279. DOI: 10.11650/twjm/1500407124.

[LMR23] J. Lindberg, L. Monin, and K. Rose. *The algebraic degree of sparse polynomial optimization*. 2023. eprint: arXiv:2308.07765v2.

[LY19] A. Leykin and J. Yu. "Beyond polyhedral homotopies". In: *Journal of Symbolic Computation* 91 (2019). MEGA 2017, Effective Methods in Algebraic Geometry, Nice (France), June 12-16, 2017., pp. 173–180. DOI: 10.1016/j.jsc.2018.06.019.

[M2] D. R. Grayson and M. E. Stillman. *Macaulay2, a software system for research in algebraic geometry*. Available at http://www2.macaulay2.com.

[MR20] T. Markwig and Y. Ren. "Computing tropical varieties over fields with valuation". In: *Found. Comput. Math.* 20.4 (2020), pp. 783–800. DOI: 10.1007/s10208-019-09430-2.

[MS15] D. Maclagan and B. Sturmfels. *Introduction to tropical geometry*. Vol. 161. Grad. Stud. Math. Providence, RI: American Mathematical Society (AMS), 2015.

[OP13] B. Osserman and S. Payne. "Lifting tropical intersections". In: *Doc. Math.* 18 (2013), pp. 121–175. DOI: 10.4171/dm/394.

[OSCAR] The OSCAR Team. *OSCAR – Open Source Computer Algebra Research system*. Available at https://www.oscar-system.org.

[OW24] N. K. Obatake and E. Walker. "Newton-Okounkov bodies of chemical reaction systems". In: *Adv. in Appl. Math.* 155 (2024), Paper No. 102672, 27. DOI: 10.1016/j.aam.2024.102672. URL: https://doi.org/10.1016/j.aam.2024.102672.

[RT24] K. Rose and M. L. Telek. *Computing positive tropical varieties and lower bounds on the number of positive roots*. 2024. eprint: arXiv:2408.15719v1.

[Stu02] B. Sturmfels. *Solving systems of polynomial equations*. 97. American Mathematical Soc., 2002.

[SW05] A. J. Sommese and C. W. I. Wampler. *The numerical solution of systems of polynomials. Arising in engineering and science*. River Edge, NJ: World Scientific, 2005. DOI: 10.1142/5763.

[Vac18] T. Vaccon. "Matrix-F5 algorithms and tropical Gröbner bases computation". In: *J. Symb. Comput.* 89 (2018), pp. 227–254. DOI: 10.1016/j.jsc.2017.11.014.

[Ver99] J. Verschelde. "Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation". In: *ACM Trans. Math. Softw.* 25.2 (1999), pp. 251–276. DOI: 10.1145/317275.317286.

[VVC94] J. Verschelde, P. Verlinden, and R. Cools. "Homotopies exploiting Newton polytopes for solving sparse polynomial systems". In: *SIAM Journal on Numerical Analysis* 31.3 (1994), pp. 915–930. DOI: 10.1137/0731049.

Mathematical Institute, Tohoku University, Japan.
Email address: paul.helminck.a6@tohoku.ac.jp
URL: https://paulhelminck.wordpress.com

Department of Mathematical Sciences, University of Copenhagen, Denmark.
Email address: oskar.henriksson@math.ku.dk
URL: https://oskarhenriksson.se

Department of Mathematical Sciences, Durham University, United Kingdom.
Email address: yue.ren2@durham.ac.uk
URL: https://yueren.de

B

Generic consistency and nondegeneracy of vertically parametrized systems

Elisenda Feliu
Department of Mathematical Sciences
University of Copenhagen

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Beatriz Pascual-Escudero
Department of Mathematics and Informatics
Universidad Politécnica de Madrid

Publication details

Preprint: https://doi.org/10.48550/arXiv.2304.02302 (2024)

GENERIC CONSISTENCY AND NONDEGENERACY OF VERTICALLY PARAMETRIZED SYSTEMS

ELISENDA FELIU, OSKAR HENRIKSSON, AND BEATRIZ PASCUAL-ESCUDERO

Abstract. We determine the generic consistency, dimension and nondegeneracy of the zero locus over $\mathbb{C}^*$, $\mathbb{R}^*$ and $\mathbb{R}_{>0}$ of vertically parametrized systems: parametric polynomial systems consisting of linear combinations of monomials scaled by free parameters. These systems generalize sparse systems with fixed monomial support and freely varying parametric coefficients. As our main result, we establish the equivalence among three key properties: the existence of nondegenerate zeros, the zero set having generically the expected dimension, and the system being generically consistent. Importantly, we prove that checking whether a vertically parametrized system has these properties amounts to an easily computed matrix rank condition.

1. Introduction

Polynomial systems that arise in applications often have fixed support, with coefficients that depend on unknown parameters, and a common theme in applied algebraic geometry is to ask how the geometry of the solution set varies with the value of the parameters. A particularly well-studied setting is when all coefficients vary independently from each other. We will refer to such systems as ***freely parametrized systems***. A simple example of this is

$$\begin{pmatrix} a_1x_1^2x_2^2 + a_2x_1^2x_3 + a_3x_3x_4^2 \\ a_4x_1^2x_3 + a_5x_3x_4^2 + a_6x_1x_3x_4 \end{pmatrix} \tag{1.1}$$

with parameters $a = (a_1, \dots, a_6)$ and variables $x = (x_1, x_2, x_3, x_4)$. Many results are known about how the geometry of zero sets of freely parametrized systems for generic choices of parameters depends on the geometry of the associated Newton polytopes. For instance, Bernstein's theorem describes the generic number of roots in $(\mathbb{C}^*)^n$ in the square case [Ber75], and more recent work characterizes generic nonemptiness and generic irreducibility of the zero locus of (possibly non-square) freely parametrized systems [Yu16, Kho16].

In many applications, the systems have additional structure apart from a fixed support, which gives rise to algebraic dependencies between the coefficients (in particular, some coefficients might be fixed). This happens for instance in enumerative geometry [EH16], optimization [LMR23], statistics [HS14], dynamics [BMMT22], reaction network theory [OW24], and for extremal-generic systems [BS24].

In this work, we will study a particular generalization of freely parametrized systems that we refer to as ***vertically parametrized systems*** (a terminology that first appeared in [HR22]), where we allow linear dependencies among the coefficients of terms with the same monomial. More specifically, a vertically parametrized system is one that can be written as

$$F = C(a \star x^M) \in \mathbb{C}[a_1, \dots, a_m, x_1^{\pm}, \dots, x_n^{\pm}]^s$$

where $M \in \mathbb{Z}^{n\times m}$ is a matrix encoding monomials, the parameters $a = (a_1, \dots, a_m)$ scale the monomials via component-wise multiplication $\star$, and $C \in \mathbb{R}^{s\times m}$ is a matrix of full row rank, whose rows encode linear combinations of the scaled monomials. A simple example of a vertically parametrized system with the same support as (1.1) is

$$\begin{pmatrix} a_1x_1^2x_2^2 + 3a_2x_1^2x_3 + a_3x_3x_4^2 \\ a_2x_1^2x_3 + 2a_3x_3x_4^2 + a_4x_1x_3x_4 \end{pmatrix}. \tag{1.2}$$

More generally, we will be interested in *linear sections with fixed direction* of the zeros of a vertically parametrized system, in the sense that we also include $\ell \geq 0$ affine forms $Lx - b$ for a fixed matrix $L \in \mathbb{C}^{\ell\times n}$, and parameters $b = (b_1, \dots, b_\ell)$, to obtain an ***augmented vertically***

parametrized system of the form

$$F = \Big(C(a \star x^M),\, Lx - b \Big) \in \mathbb{C}[a_1, \ldots, a_m, b_1, \ldots, b_\ell, x_1^{\pm}, \ldots, x_n^{\pm}]^{s+\ell}.$$

For instance, an example of such a linear section of (1.2) is given by

$$\begin{pmatrix} a_1 x_1^2 x_2^2 + 3a_2 x_1^2 x_3 + a_3 x_3 x_4^2 \\ a_2 x_1^2 x_3 + 2a_3 x_3 x_4^2 + a_4 x_1 x_3 x_4 \\ x_2 + 2x_3 - b_1 \\ x_2 + x_3 - b_2 \end{pmatrix}. \tag{1.3}$$

One of the first main results of this work is a characterization of when an augmented vertically parametrized system is *generically consistent*, in the sense that the variety $\mathbb{V}_{\mathbb{C}^*}(F_{a,b}) \subseteq (\mathbb{C}^*)^n$ is nonempty for generic choices of parameters $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$. The key condition is the following:

$$\operatorname{rk}\Big(\begin{bmatrix} C \operatorname{diag}(w) M^\top \operatorname{diag}(h) \\ L \end{bmatrix} \Big) = s + \ell \quad \text{for some } (w, h) \in \ker(C) \times (\mathbb{C}^*)^n. \tag{1.4}$$

This condition ensures that $F_{a,b}$ has a nondegenerate zero (i.e., a zero where the Jacobian of $F_{a,b}$ does not have full rank) for some $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$. Using this, we will show in Example 3.12 that the systems (1.1) and (1.2) are generically consistent, whereas (1.3) is generically inconsistent.

Theorem A (Theorem 3.7)**.** *Let $F = (\, C(a \star x^M),\, Lx - b\,)$ be an augmented vertically parametrized system with $C \in \mathbb{C}^{s \times m}$ of full row rank, $M \in \mathbb{Z}^{n \times m}$ and $L \in \mathbb{C}^{\ell \times n}$. Then there are precisely two possibilities:*

- *(i)* ***Generic consistency:*** *If (1.4) holds, then for generic $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$, it holds that $\mathbb{V}_{\mathbb{C}^*}(F_{a,b})$ is nonempty of pure dimension $n - (s + \ell)$, and all zeros are nondegenerate.*
- *(ii)* ***Generic inconsistency:*** *If (1.4) does not hold, then $\mathbb{V}_{\mathbb{C}^*}(F_{a,b})$ is empty for generic $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$. Whenever $\mathbb{V}_{\mathbb{C}^*}(F_{a,b})$ is nonempty, its dimension is strictly greater than $n - (s + \ell)$ and all zeros are degenerate.*

Furthermore, the ideal $\langle F_{a,b} \rangle \subseteq \mathbb{C}[x_1^{\pm}, \ldots, x_n^{\pm}]$ is radical for generic $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$.

For a freely parametrized system with supports $\mathcal{S}_1, \ldots, \mathcal{S}_s \subseteq \mathbb{Z}^n$, we see in Corollary 3.14 that (1.4) specializes to the following condition, which has previously been proven by other means in [Yu16, Lemma 1] and [Kho16, Theorem 11]:

$$\text{There exists a linearly independent tuple } (u_1, \ldots, u_s) \in \textstyle\prod_{i=1}^s \operatorname{Lin}(\mathcal{S}_i)\,. \tag{1.5}$$

Here, $\operatorname{Lin}(\mathcal{S}_i) \subseteq \mathbb{R}^n$ denotes the direction of the affine hull of $\mathcal{S}_i$. In the square case $s = n$, this, together with Bernstein's theorem, also recovers the usual characterization of when the mixed volume is nonzero (see, e.g. [Sch13, Theorem 5.1.7]).

In many applications, it is natural to require the parameters or the variables to be (positive) real numbers. If a vertically parametrized system has been defined over the real numbers, in the sense that C and L have real entries, we have the following real version of Theorem A (a more general result that encompasses both these versions is given in Theorem 3.7).

Theorem B (Theorem 3.7)**.** *Let $F = (\, C(a \star x^M),\, Lx - b\,)$ be an augmented vertically parametrized system, with $C \in \mathbb{R}^{s \times m}$ of full row rank, $M \in \mathbb{Z}^{n \times m}$ and $L \in \mathbb{R}^{\ell \times n}$. Suppose that $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Then there are two cases:*

- *(i)* ***Consistency in a Euclidean open set:*** *If condition (1.4) holds, then $\mathbb{V}_{\mathbb{C}^*}(F_{a,b}) \cap \mathbb{R}^n_{>0} \neq \varnothing$ for (a, b) in a nonempty Euclidean open subset of $\mathbb{R}^m_{>0} \times \mathbb{R}^\ell$. For generic such parameter values, all zeros are nondegenerate and it holds that $\dim(\mathbb{V}_{\mathbb{C}^*}(F_{a,b}) \cap \mathbb{R}^n_{>0}) = n - (s + \ell)$ as a semialgebraic set.*
- *(ii)* ***Generic inconsistency:*** *If condition (1.4) does not hold, then the set of $(a, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell$ for which $\mathbb{V}_{\mathbb{C}^*}(F_{a,b}) \cap \mathbb{R}^n_{>0} \neq \varnothing$ is nonempty but has empty Euclidean interior, and all zeros are degenerate.*

By specializing Theorem B to real freely parametrized systems with prescribed signs for the coefficients, we find in Corollary 3.17 that the zero locus contains positive points (i.e., in $\mathbb{R}^n_{>0}$) for coefficients in a nonempty Euclidean open set exactly when (1.5) holds and each polynomial has at least one positive and one negative coefficient.

There are two key features of augmented vertically parametrized systems that we use in the proof of Theorem 3.7. The first of them is the well-behaved geometry of the incidence variety

$$\mathcal{I} := \Big\{(a, b, x) \in \mathbb{C}^m \times \mathbb{C}^\ell \times (\mathbb{C}^*)^n : C(a \star x^M) = Lx - b = 0\Big\},$$

which allows us to use classical results from algebraic geometry about the dimension of fibers to relate generic consistency of the system with the generic dimension of the solution set.

Theorem C (Theorem 3.1)**.** *For an augmented vertically parametrized system, the incidence variety $\mathcal{I}$ admits a rational parametrization, and is a nonsingular irreducible variety of dimension $m + n - s$.*

This generalizes the well-known fact that the incidence variety is irreducible for freely parametrized systems (see, e.g., [PS93, Proof of Proposition 2.3]). In the square case (when $s + \ell = n$), another important consequence of Theorem C is that the Galois action of the monodromy group is transitive, which ensures that augmented vertically parametrized systems can be solved numerically with monodromy methods [DHJ+18].

As a second key ingredient, we show in Proposition 3.3 that degenerate zeros correspond to the critical points of the parametrization of $\mathcal{I}$. This allows us to invoke Sard's lemma and relate the dimension of the complex and real varieties, and thereby derive Theorem B from Theorem A.

Although Theorem A and Theorem C are stated for zeros in the complex torus $(\mathbb{C}^*)^n$, we show in Theorem 3.19 that if $M \in \mathbb{Z}^{n\times m}_{\geq 0}$ and each polynomial contains an independent constant term, then both results also hold for the full zero locus in $\mathbb{C}^n$. In fact, we prove that an augmented vertically parametrized system with independent constant terms generically does not have any irreducible component contained in a coordinate hyperplane. This generalizes [LW96, Lemma 2.1] from the freely parametrized and square case to the vertically parametrized setting.

The conclusions of this work can be contrasted with the following non-vertically parametrized systems, where neither case (i) nor (ii) of Theorem A apply:

(i) The following system is *horizontally parametrized* in the language of [HR22] (in the sense that we allow linear dependencies among coefficients appearing within the same equation):

$$\begin{pmatrix} a_1x_1^2 - a_1x_1 - a_3x_1 + a_3 \\ a_2x_1x_2 - a_2x_2 - a_4x_1 + a_4 \end{pmatrix} = \begin{pmatrix} (x_1 - 1)(a_1x_1 - a_3) \\ (x_1 - 1)(a_2x_2 - a_4) \end{pmatrix}. \tag{1.6}$$

The system is generically consistent but the zero sets have dimension one as they all contain the line $\{x_1 = 1\}$ (and the nondegenerate point $(a_3/a_1, a_4/a_2)$ whenever $a_1a_2 \neq 0$).

(ii) In the system

$$\begin{pmatrix} a_1x_1^2 + a_3x_2 + a_4x_2x_3 \\ 2a_1x_1x_2 + a_2x_1^2 + a_3x_2 \\ a_1x_2^2 + a_2x_1^2 - a_4x_2x_3 \end{pmatrix}, \tag{1.7}$$

the parameter a_1 accompanies different monomials in different equations, namely x_1^2, x_1x_2, x_2^2. The system is generically consistent with zero-dimensional zero set, but all zeros are degenerate.

(iii) Similarly to (i), the second polynomial factors in the system

$$\begin{pmatrix} a_1x_1 - a_2x_2 \\ a_1^2x_1^2 - a_2^2x_2^2 \end{pmatrix}, \tag{1.8}$$

from which it follows that the set of zeros is generically one-dimensional, and that all zeros are degenerate. This system is not linear in the parameters.

The incidence varieties of systems (1.6) and (1.7) are reducible, while it is irreducible for (1.8), but of dimension $3 > m + n - s$.

We end the introduction by briefly describing two applications where vertically parametrized systems appear naturally: optimization and dynamical systems.

Critical points in optimization: A freely parametrized single Laurent polynomial can be written as $f = a^\top x^M$ for a full row rank matrix $M \in \mathbb{Z}^{n\times m}$ and parameters $a = (a_1, \dots, a_m)$. Its critical points in $(\mathbb{C}^*)^n$ are the zeros of the square vertically parametrized system

$$x \star \nabla f = M(a \star x^M) \in \mathbb{C}[a_1, \dots, a_m, x_1^{\pm}, \dots, x_n^{\pm}]. \tag{1.9}$$

It follows from Theorem A that f_a has critical points in $(\mathbb{C}^*)^n$ for generic $a \in \mathbb{C}^m$ if and only if

$$\operatorname{rk}(M \operatorname{diag}(w) M^\top) = n \quad \text{for some } w \in \ker(M). \tag{1.10}$$

If this is the case, the number of critical points is generically finite (c.f. Example 3.13).

Steady states of reaction networks: A common model for the dynamics of reaction networks is (generalized) *mass-action kinetics*, where the evolution of some quantities x vary according to an autonomous system of ordinary differential equations of the form

$$\dot{x}(t) = C(a \star x(t)^M),$$

where $C \in \mathbb{Z}^{n\times m}$ is called the stoichiometric matrix, $a \in \mathbb{R}^m_{>0}$ the vector of reaction rate constants, and $M \in \mathbb{Z}^{n\times m}$ the kinetic matrix. The steady states are then given by the zeros of an augmented vertically parametrized system

$$\Big(C(a \star x^M),\ Lx - b\Big)$$

where the rows of L encode conserved quantities (first integrals), and b encodes total amounts. For an introduction to the theory of reaction networks and mass-action kinetics, we refer to [Dic16, Fei19], and to [MR12] for the notion of generalized mass-action kinetics.

Example 1.1. The following network from [HVMM11] models the interactions between the HIV virus and the T-cells and macrophages of an infected individual:

$$\begin{array}{llll}
\mathsf{T} + \mathsf{V} \longrightarrow 2\,\mathsf{T} + \mathsf{V} & \mathsf{M} + \mathsf{V} \longrightarrow \mathsf{M_i} & 0 \rightleftharpoons \mathsf{T} & \mathsf{T_i} \longrightarrow 0 \\
\mathsf{T} + \mathsf{V} \longrightarrow \mathsf{T_i} & \mathsf{T_i} \longrightarrow \mathsf{V} + \mathsf{T_i} & 0 \longrightarrow \mathsf{M} & \mathsf{M} \longrightarrow 0 \\
\mathsf{M} + \mathsf{V} \longrightarrow 2\,\mathsf{M} + \mathsf{V} & \mathsf{M_i} \longrightarrow \mathsf{V} + \mathsf{M_i} & \mathsf{V} \longrightarrow 0 & \mathsf{M_i} \longrightarrow 0\,.
\end{array}$$

Modeled with mass-action kinetics, we get a stoichiometric and kinetic matrix (the system lacks conserved quantities) that satisfy the conditions in Theorem B(i). From this, we conclude that there is a nonempty Euclidean open subset of parameter space where the system has a positive steady state. The existence of positive steady states is one of the key features of the model discussed in [HVMM11]. If the processes $0 \to \mathsf{T}$ and $0 \to \mathsf{M}$ (which correspond to regeneration of T-cells and macrophages) are removed, Theorem B(ii) applies, and hence the set of parameter values for which there are positive steady states is nonempty but has empty Euclidean interior.

In a follow-up paper [FHPE24], we use the results from this paper to study the generic geometry of steady state varieties, and strengthen several previous statements about reaction networks, in particular concerning absolute concentration robustness from [PEF22, GPGH$^+$25] and nondegenerate multistationarity from [CFMW17].

The paper is organized as follows. The study of augmented vertically parametrized systems is the content of Section 3. Before that, we devote Section 2 to a more general discussion of parametric polynomial systems with irreducible incidence varieties and where we allow for restricted domains of parameters and variables. We focus on the connection between nondegeneracy of zeros, generic consistency, the generic dimension of the varieties, and the radicality of the ideals, with the main results being gathered in Theorem 2.17.

Notation and conventions. We let $\star\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}^n$ denote the Hadamard product, given by $(x \star y)_i = x_i y_i$. For a field K, we let $K^* = K \setminus \{0\}$ be the group of units. For a matrix $A = (a_{ij}) \in \mathbb{Z}^{n\times m}$ and a vector $x \in (K^*)^n$, we let $x^A \in K^m$ be defined by $(x^A)_j = x_1^{a_{1j}} \cdots x_n^{a_{nj}}$ for $j = 1, \ldots, m$. The zero matrix of size $n \times m$ is denoted $0_{n\times m}$, and Id_n is the identity matrix of size n. When saying that a property holds *generically* in a family indexed by some parameters in a set $\mathcal{P} \subseteq \mathbb{C}^k$, we mean that it holds in a nonempty open subset of $\mathcal{P}$, with respect to the subspace topology induced by the Zariski topology on $\mathbb{C}^k$. For a set $S \subseteq \mathbb{C}^n$, we let $\overline{S}$ denote the Zariski closure of S in $\mathbb{C}^n$.

Acknowledgements. EF and OH have been funded by the Novo Nordisk Foundation project with grant reference number NNF20OC0065582. BP has been funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie IF grant agreement No 794627 and the Spanish Ministry of Economy project with reference number PGC2018-095392-B-I00. This work has also been funded by the European Union under the Grant Agreement number 101044561, POSALG. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or European Research Council (ERC). Neither the European Union nor ERC can be held responsible for them.

The authors thank Ignacio González Mantecón and Anne Shiu for helpful feedback on earlier versions of the manuscript, and Bernd Sturmfels for pointing us to [Yu16].

2. Preliminaries of parametric polynomial systems

In this section, we study the generic properties of the zero set of parametric systems, under the assumption that the incidence variety is irreducible of known dimension. This sets the background theory of vertically parametrized systems that we will develop in Section 3.

We focus mainly on zeros in $(\mathbb{C}^*)^n$, $(\mathbb{R}^*)^n$ and $\mathbb{R}^n_{>0}$, and we therefore allow Laurent polynomials with negative exponents. However, we will show later that the theory in this section also extends to $\mathbb{C}^n$ if one restricts to polynomials with nonnegative exponents (see Remark 2.19 and Section 3.7).

2.1. Framework.

We consider a parametric (Laurent) polynomial system of the form

$$F = (f_1, \ldots, f_r) \in \mathbb{C}[p_1, \ldots, p_k, x_1^{\pm}, \ldots, x_n^{\pm}]^r,$$

for some integers $k, n, r > 0$ with $r \leq n$, where we view $p = (p_1, \ldots, p_k)$ as parameters in $\mathbb{C}^k$ and $x = (x_1, \ldots, x_n)$ as variables in $(\mathbb{C}^*)^n$. Throughout this section, when writing $\mathbb{C}[p, x^{\pm}]$, we implicitly assume that p has k entries and x has n entries. We consider the ***incidence variety***

$$\mathcal{I} := \{(p, x) \in \mathbb{C}^k \times (\mathbb{C}^*)^n : F(p, x) = 0\}$$

and the projection map to parameter space

$$\pi\colon \mathcal{I} \to \mathbb{C}^k, \quad (p, x) \mapsto p,$$

where the system F is implicit in the notation. For each choice of parameters $p \in \mathbb{C}^k$, we get a specialized system

$$F_p := F(p, \cdot) \in \mathbb{C}[x^{\pm}]^r.$$

We identify the very affine variety

$$\mathbb{V}_{\mathbb{C}^*}(F_p) = \{x \in (\mathbb{C}^*)^n : F_p(x) = 0\} \subseteq (\mathbb{C}^*)^n$$

with the fiber $\pi^{-1}(p)$ of the projection map. We also form the set of parameter values for which the system is consistent:

$$\mathcal{Z} := \{p \in \mathbb{C}^k : \mathbb{V}_{\mathbb{C}^*}(F_p) \neq \varnothing\} = \pi(\mathcal{I}) \subseteq \mathbb{C}^k. \tag{2.1}$$

We say that F is ***generically consistent*** if $\mathcal{Z}$ is Zariski dense in $\mathbb{C}^k$. Since $\mathcal{Z}$ is constructible (see, e.g., [CLO15, Theorem 3.2.3]), this is equivalent to $\mathcal{Z}$ containing a Zariski open subset of $\mathbb{C}^k$. Hence, if F is generically consistent, a property holds generically in $\mathcal{Z}$ if and only if it holds generically in $\mathbb{C}^k$.

We will restrict the parameter space to a subset $\mathcal{P} \subseteq \mathbb{C}^k$ (with $\mathcal{P} = \mathbb{R}^k_{>0}$ as the main example), and we therefore extend the notation in (2.1) to

$$\mathcal{Z}_\mathcal{P} := \mathcal{Z} \cap \mathcal{P} = \{p \in \mathcal{P} : \mathbb{V}_{\mathbb{C}^*}(F_p) \neq \varnothing\} = \pi(\mathcal{I} \cap (\mathcal{P} \times (\mathbb{C}^*)^n)) \subseteq \mathbb{C}^k. \tag{2.2}$$

Under mild assumptions, the choice of $\mathcal{P}$ does not affect generic consistency of the system.

Lemma 2.1. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$ and $\mathcal{P} \subseteq \mathbb{C}^k$. Then the following holds:*

(i) If $\overline{\mathcal{Z}_\mathcal{P}} = \mathbb{C}^k$, then $\overline{\mathcal{Z}} = \mathbb{C}^k$.

(ii) If $\overline{\mathcal{P}} = \mathbb{C}^k$, then the converse of (i) holds.

Proof. Part (i) is immediate, since $\mathcal{Z}_\mathcal{P} \subseteq \mathcal{Z}$. For part (ii), note that if $\overline{\mathcal{Z}} = \mathbb{C}^k$, then $\mathcal{Z}$ contains a nonempty Zariski open subset U of $\mathbb{C}^k$. The intersection $U \cap \mathcal{P}$ is Zariski dense in $\mathbb{C}^k$ since it is the intersection of a nonempty Zariski open and a Zariski dense set, and $\overline{\mathcal{Z}_\mathcal{P}} = \mathbb{C}^k$ follows. □

With this notation in place, we proceed now to study the generic dimension of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ for $p \in \mathcal{P}$ in relation to whether the system is generically consistent. We then move on to study zeros in subsets $\mathcal{X} \subseteq (\mathbb{C}^*)^n$. We also relate the study of nonemptiness and dimension of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ to the concept of *nondegeneracy*. The connections among these properties are summarized in Theorem 2.17 towards the end of this section.

2.2. **Generic consistency and dimension.** An immediate first observation is that for each $p \in \mathcal{P}$, the principal ideal theorem [Eis95, Theorem 10.2] gives that all irreducible components of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ have dimension at least $n - r$. In particular,

$$\dim(\mathbb{V}_{\mathbb{C}^*}(F_p)) \geq n - r \qquad \text{for all } p \in \mathcal{Z}. \tag{2.3}$$

If (2.3) holds with equality for a given $p \in \mathcal{P}$, then all irreducible components have dimension $n - r$, and we say that $\mathbb{V}_{\mathbb{C}^*}(F_p)$ has ***pure dimension*** $n - r$. In what follows, the bound (2.3) will be referred to as the ***expected dimension*** of $\mathbb{V}_{\mathbb{C}^*}(F_p)$. Likewise, the expected dimension of $\mathcal{I}$ is $k + n - r$.

Remark 2.2. As neither of the expected dimensions $\dim(\mathcal{I}) = k+n-r$ and $\dim(\mathbb{V}_{\mathbb{C}^*}(F_p)) = n-r$ can be attained if the coefficient matrix of F (seen as a polynomial system in $\mathbb{C}[p, x^{\pm}]^r$) fails to have full rank, a natural preprocessing step when studying the dimension of the zero set is to remove linear dependencies in the entries of F.

Theorem 2.3. *Let $\mathcal{P} \subseteq \mathbb{C}^k$ and $F \in \mathbb{C}[p, x^{\pm}]^r$ be such that the incidence variety $\mathcal{I}$ is irreducible. Then:*

(i) For all $p \in \mathcal{Z}$, it holds that

$$\dim(Y) \geq \dim(\mathcal{I}) - \dim(\overline{\mathcal{Z}}) \quad \text{for all irreducible components } Y \subseteq \mathbb{V}_{\mathbb{C}^*}(F_p)$$

with equality for generic $p \in \overline{\mathcal{Z}}$.

(ii) If $\overline{\mathcal{Z}_\mathcal{P}} = \mathbb{C}^k$, then $\mathbb{V}_{\mathbb{C}^*}(F_p)$ has pure dimension $\dim(\mathcal{I}) - k$ for generic $p \in \mathcal{Z}_\mathcal{P}$.

(iii) If $\overline{\mathcal{Z}_\mathcal{P}} \subsetneq \mathbb{C}^k$ and $\overline{\mathcal{P}} = \mathbb{C}^k$, then all irreducible components of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ have dimension greater than $\dim(\mathcal{I}) - k$ for all $p \in \mathcal{Z}$.

Proof. Part (i) follows by applying the classical theorem of dimension of fibers (see, e.g., [Mum76, Theorem 3.13, Corollary 3.15]) to the canonical projection map

$$\pi \colon \mathcal{I} \to \overline{\mathcal{Z}}, \quad (p, x) \mapsto p.$$

Part (ii) follows by noting that (i) and Lemma 2.1(i) together give that there is a nonempty Zariski open subset $U \subseteq \mathbb{C}^k$ such that $\mathbb{V}_{\mathbb{C}^*}(F_p)$ has pure dimension $\dim(\mathcal{I}) - k$ for all $p \in U$. In particular, this holds for $p \in \mathcal{Z}_\mathcal{P} \cap U$, which is nonempty since $\overline{\mathcal{Z}_\mathcal{P}} = \mathbb{C}^k$. Finally, part (iii) follows directly from (i) and Lemma 2.1(ii). □

Combining part (ii) and (iii) of Theorem 2.3 when $\overline{\mathcal{P}} = \mathbb{C}^k$, we obtain that $\dim(\mathbb{V}_{\mathbb{C}^*}(F_p)) = \dim(\mathcal{I}) - k$ for some $p \in \mathcal{Z}_{\mathcal{P}}$ if and only if $\mathcal{Z}_{\mathcal{P}}$ is Zariski dense in $\mathbb{C}^k$.

Example 2.4. For the system in (1.8) with $n = k = r = 2$, we have $\mathcal{Z} = (\mathbb{C}^*)^2$ and $\mathcal{I}$ is irreducible of dimension 3. By Theorem 2.3(ii), it holds that $\dim(\mathbb{V}_{\mathbb{C}^*}(F_p)) = 3 - 2 = 1$ for generic $p \in (\mathbb{C}^*)^2$, which we already noticed in (1.8). In this case, the generic dimension differs from the expected dimension $n - r = 0$ from (2.3).

We end by connecting consistency and dimension with the notion of flatness. Recall that a morphism of varieties $f\colon X \to Y$ is said to be flat at a point $x \in X$ if the induced homomorphism of local rings $\mathscr{O}_{Y,f(x)} \to \mathscr{O}_{X,x}$ is a flat ring homomorphism.

Proposition 2.5. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$ be a parametric system such that $\mathcal{I}$ is irreducible. Then the following holds:*

(i) If the projection $\pi\colon \mathcal{I} \to \mathbb{C}^k$ is flat at generic points $(p, x) \in \mathcal{I}$, then $\overline{\mathcal{Z}} = \mathbb{C}^k$.

(ii) If $\mathcal{I}$ is nonsingular, then the converse of (i) holds.

Proof. It follows from [Sta24, 00ON] that if $\pi\colon \mathcal{I} \to \mathbb{C}^k$ is flat at a point $(p, x) \in \mathcal{I}$, then the local dimension of $\pi^{-1}(p)$ at (p, x) is $\dim(\mathcal{I}) - k$, and [Sta24, 00R4] gives that the converse is true if $\mathcal{I}$ is nonsingular (and hence in particular Cohen–Macaulay). The desired result now follows from Theorem 2.3. □

2.3. **Restricting the ambient space.** We consider now the zero set obtained by restricting the ambient space to a subset $\mathcal{X} \subseteq (\mathbb{C}^*)^n$. The main example we have in mind is the positive orthant $\mathcal{X} = \mathbb{R}^n_{>0}$. For $K \in \{\mathbb{R}, \mathbb{C}\}$, $F \in K[x^{\pm}]^r$, and $\mathcal{X} \subseteq (K^*)^n$, we let the variety $\mathbb{V}^{\mathcal{X}}_{K^*}(F)$ be defined as the union of the irreducible components of $\mathbb{V}_{K^*}(F)$ that intersect $\mathcal{X}$. Clearly, the expected dimension $n - r$ from (2.3) is a lower bound on the dimension of $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F)$ as long as $\mathbb{V}_{\mathbb{C}^*}(F) \cap \mathcal{X} \neq \varnothing$.

Remark 2.6. As $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F) \subseteq (\mathbb{C}^*)^n$ is a Zariski closed set, we have

$$\overline{\mathbb{V}_{\mathbb{C}^*}(F) \cap \mathcal{X}} \subseteq \mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F), \tag{2.4}$$

but equality might not hold (consider $F = (x_1 - 1)^2 + (x_2 - 1)^2$ and $\mathcal{X} = \mathbb{R}^2_{>0}$). In general, equality in (2.4) holds if $\mathcal{X}$ is a Euclidean open subset of $(\mathbb{C}^*)^n$, or if $\mathcal{X}$ is a Euclidean open subset of $(\mathbb{R}^*)^n$ and additionally each irreducible component of $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F)$ contains a nonsingular real point (see [PEF22, Theorem 6.5] and [BCR98, Proposition 3.3.16]). Furthermore, if $F \in \mathbb{R}[x^{\pm}]^r$, then $n - r$ is a lower bound of the dimension of $\mathbb{V}_{\mathbb{R}^*}(F)$ if it is nonempty and each of its irreducible components contains a nonsingular point. Similarly, $n - r$ is a lower bound on the dimension of $\mathbb{V}_{\mathbb{R}^*}(F) \cap \mathbb{R}^n_{>0}$ as a semialgebraic set if it is nonempty and each irreducible component of $\mathbb{V}_{\mathbb{R}^*}(F)$ that intersects $\mathbb{R}^n_{>0}$ contains a nonsingular point.

For $\mathcal{P} \subseteq \mathbb{C}^k$ and $\mathcal{X} \subseteq (\mathbb{C}^*)^n$, we generalize definition (2.2) to restrict to zeros in $\mathcal{X}$:

$$\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) := \{p \in \mathcal{P} : \mathbb{V}_{\mathbb{C}^*}(F_p) \cap \mathcal{X} \neq \varnothing\} = \pi(\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})) \subseteq \mathbb{C}^k,$$

where recall that π is the projection to parameter space. With this notation, $\mathcal{Z}_{\mathcal{P}}((\mathbb{C}^*)^n) = \mathcal{Z}_{\mathcal{P}}$, and we have the following inclusions:

$$\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) \subseteq \mathcal{Z}_{\mathcal{P}} \subseteq \mathcal{Z}, \qquad \mathcal{Z}_{\mathcal{P}}(\mathcal{X}) \subseteq \mathcal{Z}_{\mathbb{C}^k}(\mathcal{X}) \subseteq \mathcal{Z}. \tag{2.5}$$

If $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is Zariski dense in $\mathbb{C}^k$, then so is $\mathcal{Z}_{\mathcal{P}}$. If $\mathcal{I}$ is irreducible and attains its expected dimension $k + n - r$, then Theorem 2.3(ii) readily gives that $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F_p)$ has pure dimension $n - r$ for generic $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$. The following example illustrates that Theorem 2.3(iii) might not extend to subsets $\mathcal{X}$ unless extra requirements are imposed.

Example 2.7. Let $\mathcal{P} = \mathbb{R}^3$, $\mathcal{X} = \mathbb{R}^2_{>0}$, and $f = (p_1x_1 - p_2x_2)^2 + p_3^2x_1$. The incidence variety $\mathcal{I}$ is irreducible and attains its expected dimension 4, and $\mathbb{V}_{\mathbb{C}^*}(f_p)$ has dimension 1 for generic p. The set

$$\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) = \{p \in \mathbb{R}^3 \colon p_3 = 0, p_1p_2 > 0\}$$

is not Zariski dense in $\mathbb{C}^3$, but still, the dimension of $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(f_p)$ takes the expected value of 1 for all $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

In what follows, the sets $\mathcal{P}$ and $\mathcal{X}$ will be required to satisfy that $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$, which ensures the following property.

Lemma 2.8. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$, $\mathcal{P} \subseteq \mathbb{C}^k$, and $\mathcal{X} \subseteq (\mathbb{C}^*)^n$ be such that $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$. Then*

$$\overline{\mathcal{Z}_{\mathcal{P}}(\mathcal{X})} = \overline{\mathcal{Z}_{\mathcal{P}}} = \overline{\mathcal{Z}_{\mathbb{C}^k}(\mathcal{X})} = \overline{\mathcal{Z}}.$$

In particular $\mathcal{Z}$ is Zariski dense in $\mathbb{C}^k$ if and only if $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is.

Proof. As $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$, we have

$$\overline{\mathcal{Z}_{\mathcal{P}}(\mathcal{X})} = \overline{\pi(\mathcal{I} \cap (\mathcal{P} \times \mathcal{X}))} = \overline{\pi(\overline{\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})})} = \overline{\pi(\mathcal{I})} = \overline{\mathcal{Z}},$$

which together with the inclusions in (2.5) give the desired statement. □

Remark 2.9. Assume that $\mathcal{I}$ is nonsingular and irreducible. If $\mathcal{P}$ and $\mathcal{X}$ are Euclidean open subsets of $\mathbb{C}^k$ and $(\mathbb{C}^*)^n$, respectively, then $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$ if this intersection is nonempty. The same conclusion holds if F has real coefficients and $\mathcal{P}$ and $\mathcal{X}$ are Euclidean open subsets of $\mathbb{R}^k$ and $\mathbb{R}^n$, respectively. This follows from the nonsingularity of $\mathcal{I}$, together with the fact that for a complex irreducible variety defined by polynomials with real coefficients that has at least one real nonsingular point, the real points form a Zariski dense subset of the complex variety [BCR98, Proposition 3.3.16]. Observe that in Example 2.7, the intersection $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$, which is not dense in $\mathcal{I}$, consists only of singular points.

2.4. **Nondegeneracy, Euclidean interior, and (real) dimension.** This subsection recalls the concept of *nondegeneracy*, as a means of studying the dimension of a variety, both theoretically and computationally. Additionally, we relate Zariski denseness of $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ in $\mathbb{C}^k$ to $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ having nonempty Euclidean interior in $\mathcal{P}$.

Definition 2.10. Given a polynomial system $F = (f_1, \ldots, f_r) \in \mathbb{C}[x^{\pm}]^r$, a zero $x^* \in \mathbb{V}_{\mathbb{C}^*}(F)$ is called ***nondegenerate*** if the Jacobian matrix $J_F(x^*) := \left(\frac{\partial f_i}{\partial x_j}(x^*)\right)_{ij}$ has rank r. Otherwise, x^* is called ***degenerate***.

Unlike the weaker notion of *nonsingularity*, nondegeneracy depends on the particular tuple $F \in \mathbb{C}[x^{\pm}]^r$, and not just the variety $\mathbb{V}_{\mathbb{C}^*}(F)$. The relationship between nondegenerate zeros and nonsingular points, and the connection to dimension, is summarized in the following proposition, which gathers several well-known results.

Proposition 2.11.

(i) Let $F \in \mathbb{C}[x^{\pm}]^r$ and $x^* \in (\mathbb{C}^*)^n$ be a nondegenerate zero of F. Then x^* is a nonsingular point of $\mathbb{V}_{\mathbb{C}^*}(F)$ and belongs to a unique irreducible component, which has dimension $n - r$.

(ii) Let $F \in \mathbb{R}[x^{\pm}]^r$ and $x^* \in (\mathbb{R}^*)^n$ be a nondegenerate zero of F. Then there is a unique irreducible component of $\mathbb{V}_{\mathbb{R}^*}(F)$ containing x^*. This irreducible component is Zariski dense in the irreducible component of $\mathbb{V}_{\mathbb{C}^*}(F)$ containing x^* and has dimension $n - r$.

(iii) Let $F \in \mathbb{C}[p, x^{\pm}]^r$. Then the set $\mathcal{U}$ of points $(p, x) \in \mathcal{I}$ for which x is a nondegenerate zero of F_p is a Zariski open subset of $\mathcal{I}$.

Proof. Statement (i) is [CLO15, Theorem 9.6.9]. For (ii), the uniqueness follows from [CLO15, Theorem 9.6.8(iv)], whereas the density and dimensionality claims follow from Remark 2.6 (see also [BCR98, Theorem 3.3.10]). For statement (iii), consider the Jacobian $J_F = (\partial f_i/\partial x_j) \in \mathbb{C}[p, x^{\pm}]^{r\times n}$, and note that the complement $\mathcal{I} \setminus \mathcal{U} \subseteq \mathcal{I}$ is cut out by the r-minors of J_F, and therefore forms a Zariski closed subset of $\mathcal{I}$. □

Theorem 2.3 allows us to determine the generic dimension of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ from the Zariski denseness of $\mathcal{Z}_{\mathcal{P}}$, while the implicit function theorem (as we will use below in Proposition 2.14) gives that nondegeneracy implies that $\mathcal{Z}_{\mathcal{P}}$ has nonempty Euclidean interior. The key for making the connection between consistency and the Euclidean interior is the following topological property.

Definition 2.12. A nonempty subset $A \subseteq \mathbb{C}^k$ is said to be ***locally Zariski dense*** (in $\mathbb{C}^k$) if $U \cap A$ is Zariski dense in $\mathbb{C}^k$ for any Euclidean open subset $U \subseteq \mathbb{C}^k$ with $U \cap A \neq \varnothing$.

Simple examples of locally Zariski dense sets in $\mathbb{C}^k$ include all nonempty Euclidean open subsets of $\mathbb{C}^k$, $\mathbb{R}^k$, and $\mathbb{R}^k_{>0}$. In general, any nonempty subset $S \subseteq \mathbb{R}^k$ for which, with the Euclidean topology, its closure agrees with the closure of its interior, is locally Zariski dense. Any locally Zariski dense set is in particular Zariski dense, but the converse is not true (for instance, $\mathbb{Z}^k$ is Zariski dense in $\mathbb{C}^k$, but not locally Zariski dense).

Lemma 2.13. *Let $\mathcal{P} \subseteq \mathbb{C}^k$ be locally Zariski dense. If a subset $S \subseteq \mathcal{P}$ has nonempty Euclidean interior in $\mathcal{P}$, then S is Zariski dense in $\mathbb{C}^k$.*

Proof. By hypothesis, there exists an open Euclidean ball $B \subseteq \mathbb{C}^k$ such that $\varnothing \neq B \cap \mathcal{P} \subseteq S$. The Zariski closures satisfy $\mathbb{C}^k = \overline{B \cap \mathcal{P}} \subseteq \overline{S}$ as $\mathcal{P}$ is locally Zariski dense. Hence $\mathbb{C}^k = \overline{S}$. □

Proposition 2.14. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$ and $\mathcal{P} \subseteq \mathbb{C}^k$ be locally Zariski dense. Assume that F_{p^*} has a nondegenerate zero x^* in $(\mathbb{C}^*)^n$ for some $p^* \in \mathcal{P}$. Then the following statements hold:*

(i) $\mathcal{Z}_{\mathcal{P}}$ has nonempty Euclidean interior in $\mathcal{P}$ and is Zariski dense in $\mathbb{C}^k$.

(ii) If in addition the incidence variety $\mathcal{I}$ is irreducible, then $\dim(\mathcal{I}) = k+n-r$, and $\mathbb{V}_{\mathbb{C}^*}(F_p)$ has pure dimension $n-r$ for generic $p \in \mathcal{Z}_{\mathcal{P}}$.

Proof. For (i), by assumption we have $\operatorname{rk}(J_{F_{p^*}}(x^*)) = r$. Let $A \in \mathbb{C}^{(n-r)\times n}$ be a matrix whose rows extend the rows of $J_{F_{p^*}}(x^*)$ to a basis of $\mathbb{C}^n$. Then x^* is a nondegenerate zero of the square system $\widetilde{F}_{p^*}$, where

$$\widetilde{F} := \begin{pmatrix} F \\ Ax - Ax^* \end{pmatrix} \in \mathbb{C}[p, x^{\pm}]^n.$$

The complex implicit function theorem [Huy05, Proposition 1.1.11] now gives that there exists an open Euclidean neighborhood B of p^* contained in $\widetilde{\mathcal{Z}}$ (the set $\mathcal{Z}$ for $\widetilde{F}$) and hence in $\mathcal{Z}$. Intersecting B with $\mathcal{P}$, we obtain the first part of (i). By Lemma 2.13, $\mathcal{Z}_{\mathcal{P}}$ is Zariski dense in $\mathbb{C}^k$.

For (ii), as the pair (p^*, x^*) is a nondegenerate zero of F and $\mathcal{I}$ is irreducible, Proposition 2.11(i) gives that $\mathcal{I}$ has the expected dimension. This fact, (i) and Theorem 2.3(ii) give now the second part of (ii). □

In certain scenarios within the setting of Theorem 2.3, all zeros of F_p will be nondegenerate for generic $p \in \mathcal{P}$. A simple such condition is that the system is square.

Theorem 2.15. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$ with $n = r$, and let $\mathcal{P} \subseteq \mathbb{C}^k$ be locally Zariski dense. Assume that the incidence variety $\mathcal{I}$ is irreducible and that F_{p^*} has a nondegenerate zero for some $p^* \in \mathcal{P}$. Then all zeros of F_p are nondegenerate for all p in a nonempty Zariski open subset of $\mathcal{Z}_{\mathcal{P}}$.*

Proof. Consider the proper Zariski closed subset set $D \subsetneq \mathcal{I}$ consisting of the points (p, x) for which x is a degenerate zero of F_p (c.f. Proposition 2.11(iii)). As $\dim(\mathcal{I}) = k$ by Proposition 2.14(ii), we have $\dim(\overline{\pi(D)}) \leq \dim(D) < k$. For all p in the nonempty Zariski open set $U := \mathbb{C}^k \setminus \overline{\pi(D)}$, all zeros of F_p are nondegenerate. The result now follows from $U \cap \mathcal{Z}_{\mathcal{P}} \neq \varnothing$, as $\mathcal{Z}_{\mathcal{P}}$ is Zariski dense by Proposition 2.14(i). □

We conclude this subsection by noting that nondegeneracy allows us to also assert radicality of the ideals generated by the systems. The following proposition follows from standard commutative algebra arguments. We give a proof in Appendix A for completeness.

Proposition 2.16. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$, and suppose that all zeros of F_p in $(\mathbb{C}^*)^n$ are nondegenerate for generic $p \in \mathbb{C}^k$. Then the following holds:*

(i) The ideals $\langle F_p \rangle \subseteq \mathbb{C}[x^{\pm}]$ and $\langle F_p \rangle \cap \mathbb{C}[x] \subseteq \mathbb{C}[x]$ are radical for generic $p \in \mathbb{C}^k$.

(ii) The ideals $\langle F \rangle \subseteq \mathbb{C}(p)[x^{\pm}]$ and $\langle F \rangle \cap \mathbb{C}(p)[x] \subseteq \mathbb{C}(p)[x]$ are radical.

2.5. **Main implications on dimension, Zariski denseness and nondegeneracy.** In the previous subsections, we have studied the generic dimension of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ in relation to the existence of nondegenerate zeros and topological properties of $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$. For $F \in \mathbb{C}[p, x^{\pm}]^r$, $\mathcal{P} \subseteq \mathbb{C}^k$, and $\mathcal{X} \subseteq (\mathbb{C}^*)^n$, we consider from now on the following statements:

(deg1) F_p has a nondegenerate zero in $(\mathbb{C}^*)^n$ for some $p \in \mathbb{C}^k$.

(degX1) F_p has a nondegenerate zero in $\mathcal{X}$ for some $p \in \mathcal{P}$.

(degXG) There is a nonempty Zariski open subset $\mathcal{U} \subseteq \mathcal{I}$ such that for all $(p, x) \in \mathcal{U} \cap (\mathcal{P} \times \mathcal{X})$, x is a nondegenerate zero of F_p.

(degAllG) For generic $p \in \mathcal{Z}$, all zeros of F_p in $(\mathbb{C}^*)^n$ are nondegenerate.

(setE) $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ has nonempty Euclidean interior in $\mathcal{P}$.

(setZ) $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is Zariski dense in $\mathbb{C}^k$.

(flatG) The projection $\pi \colon \mathcal{I} \to \mathbb{C}^k$ is flat at generic $(p, x) \in \mathcal{I}$.

(dim1) $\mathbb{V}_{\mathbb{C}^*}(F_p)$ has pure dimension $n - r$ for at least one $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

(dimG) $\mathbb{V}_{\mathbb{C}^*}(F_p)$ has pure dimension $n - r$ for generic $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

(dimX) $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F_p)$ has pure dimension $n - r$ for generic $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

(rad) F_p generates a radical ideal in $\mathbb{C}[x^{\pm}]$ for generic $p \in \mathbb{C}^k$, and F generates a radical ideal in $\mathbb{C}(p)[x^{\pm}]$.

(reg) $\mathbb{V}_{\mathbb{C}^*}(F_p)$ is a nonsingular complex algebraic variety for generic $p \in \mathbb{C}^k$.

(real) If $F \in \mathbb{R}[p, x^{\pm}]^r$, $\mathcal{P} \subseteq \mathbb{R}^k$ and $\mathcal{X} \subseteq (\mathbb{R}^*)^n$, then $\mathbb{V}^{\mathcal{X}}_{\mathbb{R}^*}(F_p)$ and $\mathbb{V}_{\mathbb{R}^*}(F_p)$ have pure dimension $n - r$ for generic $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

Recall that $n - r$ is the expected dimension of $\mathbb{V}_{\mathbb{C}^*}(F_p)$, while $k + n - r$ is the expected dimension of $\mathcal{I}$. The following main theorem gathers the conclusions of our results so far.

Theorem 2.17. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$, $\mathcal{P} \subseteq \mathbb{C}^k$ locally Zariski dense and $\mathcal{X} \subseteq (\mathbb{C}^*)^n$. Assume that the incidence variety $\mathcal{I}$ is irreducible of dimension $k + n - r$ and that $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$. The following implications hold:*

(i) (deg1), (degX1) *and* (degXG) *are equivalent, and* (degAllG) *implies any of these. If in addition $r = n$, then these four statements are all equivalent.*

(ii) (setZ), (dimG) *and* (dim1) *are equivalent.*

(iii) (setE) *implies* (setZ).

(iv) (deg1) *implies* (setZ).

(v) (dimG) *implies* (dimX).

(vi) (degAllG) *implies* (rad), (reg) *and* (real).

(vii) (flatG) *implies* (setZ), *and the converse is true if $\mathcal{I}$ is nonsingular.*

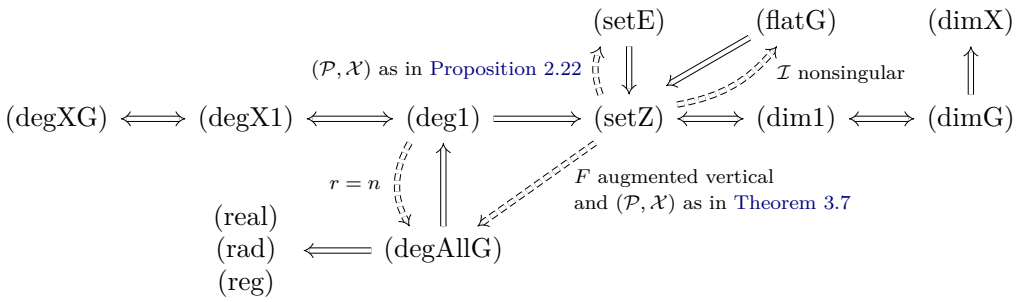


FIGURE 1. Graphical illustration of the implications in Theorem 2.17. The dashed arrows indicate implications that hold under additional assumptions.

Proof. Note throughout that by hypothesis, $\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) \neq \varnothing$. Statement **(iii)** is a consequence of Lemma 2.13 and statement **(iv)** follows from Proposition 2.14(i) and Lemma 2.8.

We show next **(i)**. We have that (degXG)⇒(degX1), as $\mathcal{U}$ is a nonempty Zariski open set and $\mathcal{I}\cap(\mathcal{P}\times\mathcal{X})$ is Zariski dense in $\mathcal{I}$, guaranteeing that the intersection $\mathcal{U}\cap(\mathcal{P}\times\mathcal{X})$ is nonempty. The implication (degX1)⇒(deg1) is immediate, (deg1)⇒(degXG) follows from Proposition 2.11(iii), and (degAllG) trivially implies (deg1). Theorem 2.15 gives (deg1)⇒(degAllG) when $r = n$ as by **(iv)**, $\mathcal{Z}$ is Zariski dense in $\mathbb{C}^n$ and hence contains a nonempty Zariski open set of $\mathbb{C}^k$.

For **(ii)**, for the implication (setZ)⇒(dimG), Lemma 2.8 gives that $\mathcal{Z}_{\mathcal{P}}$ is Zariski dense whenever $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is, and hence (dimG) follows from Theorem 2.3(ii). The implication (dimG)⇒(dim1) is immediate. Finally, if (dim1) holds, then Theorem 2.3 gives that $\mathcal{Z}_{\mathcal{P}}$ is Zariski dense, and by Lemma 2.8 so is $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$, giving (setZ).

For **(v)**, the implication holds as $\mathbb{V}^{\mathcal{X}}_{\mathbb{C}^*}(F_p)$ is the nonempty union of irreducible components of $\mathbb{V}_{\mathbb{C}^*}(F_p)$ if $p \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

For **(vi)**, the implications from (degAllG) to (real) and (reg) follow from Proposition 2.11. The implication (degAllG)⇒(rad) follows from Proposition 2.16.

Finally, **(vii)** is the content of Proposition 2.5. □

The condition that $\mathcal{I}\cap(\mathcal{P}\times\mathcal{X})$ is Zariski dense in $\mathcal{I}$ only imposes very mild conditions on $\mathcal{X}$, as the following small example illustrates.

Example 2.18. Let $f = x - p$, $\mathcal{P} = \mathbb{C}$ and $\mathcal{X} = \mathbb{Q}^*$. Then $\mathcal{I} = \{(p,p) : p \in \mathbb{C}^*\}$. The intersection $\mathcal{I}\cap(\mathcal{P}\times\mathcal{X}) = \{(p,p) : p \in \mathbb{Q}^*\}$ is dense in $\mathcal{I}$ and $\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) = \mathbb{Q}^*$ is dense in $\mathbb{C}$. As (setZ) holds, Theorem 2.17 applies and the generic dimension is 0. Note that (setE) does not hold, showing that the implication (setZ)⇒(setE) is not true for general F and $\mathcal{X}$.

Remark 2.19. We have chosen to focus on the complex torus because removal of zero coordinates will be necessary in the systems studied in the next section. However, for systems with nonnegative exponents $F \in \mathbb{C}[p,x]^r$, if the ***affine incidence variety***

$$\mathcal{I}_{\mathbb{C}} := \{(p,x) \in \mathbb{C}^k \times \mathbb{C}^n : F_p(x) = 0\}$$

is irreducible, Theorem 2.3 holds also in the affine setting over $\mathbb{C}$ (rather than $\mathbb{C}^*$) after replacing $\mathbb{V}_{\mathbb{C}^*}(F_p)$ by $\mathbb{V}_{\mathbb{C}}(F_p)$. Additionally, Lemma 2.8, Propositions 2.11, 2.14 and 2.16, as well as Theorem 2.15 extend easily to $\mathbb{C}$ by allowing $\mathcal{X} \subseteq \mathbb{C}^n$. We conclude that Theorem 2.17 holds in the affine setting over $\mathbb{C}$ as long as $\mathcal{I}_{\mathbb{C}}$ is irreducible and $\mathcal{I}_{\mathbb{C}}\cap(\mathcal{P}\times\mathcal{X})$ is Zariski dense in $\mathcal{I}_{\mathbb{C}}$.

In Section 3, we will see that for augmented vertically parametrized systems, the nine first statements (deg1)–(dimG) from the beginning of the subsection are equivalent. Thus, all the desired properties of the polynomial system will rely on the existence of a nondegenerate zero in the complex torus. The key part of the argument will be to show that $\mathcal{Z}$ has nonempty Euclidean interior if and only if F has a nondegenerate zero. This phenomenon need not happen for general families, as example (1.8) shows.

2.6. **Zariski denseness and nonempty Euclidean interior.** For the stronger result discussed at the end of the previous subsection to hold, we will need to restrict to sets $\mathcal{P}$ and $\mathcal{X}$ where the converse of Lemma 2.13 holds, that is (setZ) implies (setE) (c.f. Example 2.18). In this final subsection, we identify pairs $(\mathcal{P}, \mathcal{X})$ for which this holds.

Lemma 2.20.

(i) A constructible set $S \subseteq \mathbb{C}^k$ is Zariski dense in $\mathbb{C}^k$ if and only if it has nonempty Euclidean interior in $\mathbb{C}^k$.

(ii) A semialgebraic set $S \subseteq \mathbb{R}^k$ is Zariski dense in $\mathbb{C}^k$ if and only if it has nonempty Euclidean interior in $\mathbb{R}^k$.

Proof. For part (i), as S is constructible, $S = \bigcup_{i=1}^t Z_i \cap U_i$, for some irreducible Zariski closed sets $Z_i \subseteq \mathbb{C}^k$ and nonempty Zariski open subsets $U_i \subseteq \mathbb{C}^k$. The statement now follows from the fact that the Zariski closure satisfies $\overline{S} = \bigcup_{i=1}^t Z_i$. In case (ii), [BCR98, Proposition 2.8.2] gives that S is Zariski dense in $\mathbb{C}^k$ if and only if its semialgebraic dimension is k, which in turn is equivalent to S having nonempty Euclidean interior in $\mathbb{R}^k$. □

In order to apply Lemma 2.20, we note that $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is constructible or semialgebraic, if both $\mathcal{P}$ and $\mathcal{X}$ are defined by algebraic data in the sense of the following definition.

Definition 2.21. A pair of sets $(\mathcal{P}, \mathcal{X})$ where $\mathcal{P} \subseteq \mathbb{C}^k$ and $\mathcal{X} \subseteq \mathbb{C}^n$ is said to be ***algebraically defined*** if it satisfies one of the following conditions:

(i) $\mathcal{P}$ is a Zariski open subset of $\mathbb{C}^k$ and $\mathcal{X}$ is constructible.

(ii) $\mathcal{P} \subseteq \mathbb{R}^k$ and $\mathcal{X} \subseteq \mathbb{R}^n$ are both semialgebraic, and $\mathcal{P}$ is locally Zariski dense.

(iii) $\mathcal{P} \subseteq \mathbb{R}^k$ is semialgebraic and locally Zariski dense, and $\mathcal{X} \subseteq \mathbb{C}^n$ is constructible.

Examples of algebraically defined pairs include $(\mathbb{R}^k, \mathbb{R}^n_{>0})$, $(\mathbb{R}^k, (\mathbb{R}^*)^n)$, and $(\mathbb{C}^k, (\mathbb{C}^*)^n)$.

Proposition 2.22. *Let $F \in \mathbb{C}[p, x^{\pm}]^r$ and $(\mathcal{P}, \mathcal{X}) \subseteq \mathbb{C}^k \times (\mathbb{C}^*)^n$ be an algebraically defined pair. Then:*

(i) $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is either a semialgebraic subset of $\mathbb{R}^k$ or a constructible subset of $\mathbb{C}^k$.

(ii) (setZ) and (setE) in Theorem 2.17 are equivalent.

Proof. For (i), in case Definition 2.21(i), the set $\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) \subseteq \mathbb{C}^k$ is constructible by Chevalley's theorem. For the case Definition 2.21(ii), the Tarski–Seidenberg Theorem [BCR98, Theorem 2.2.1] gives that $\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) \subseteq \mathbb{R}^k$ is semialgebraic. Finally, in case Definition 2.21(iii), Chevalley's theorem gives that $\mathcal{Z}_{\mathbb{C}^k}(\mathcal{X}) \subseteq \mathbb{C}^k$ is constructible. This in turn implies that

$$\mathcal{Z}_{\mathcal{P}}(\mathcal{X}) = \mathcal{Z}_{\mathbb{C}^k}(\mathcal{X}) \cap \mathcal{P} = (\mathcal{Z}_{\mathbb{C}^k}(\mathcal{X}) \cap \mathbb{R}^k) \cap \mathcal{P}$$

is an intersection of semialgebraic sets and therefore semialgebraic (note that the real points of a constructible set in $\mathbb{C}^k$ form a semialgebraic set in $\mathbb{R}^k$).

For (ii), the reverse implication is Lemma 2.13 as $\mathcal{P}$ is locally Zariski dense. For the forward implication, by Lemma 2.20, if $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is Zariski dense in $\mathbb{C}^k$, then it has nonempty Euclidean interior in $\mathbb{R}^k$ in cases (ii) and (iii) of Definition 2.21 and in $\mathbb{C}^k$ in case (i). This in turn implies that $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ has nonempty Euclidean interior in $\mathcal{P}$, concluding the proof. □

Remark 2.23. The condition of $\mathcal{P}$ being locally Zariski dense cannot be replaced by $\mathcal{P}$ being Zariski dense in Lemma 2.13 and in Definition 2.21(ii),(iii). The fact that $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is semialgebraic, gives that $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ being Zariski dense in $\mathbb{C}^k$ is equivalent to having nonempty Euclidean interior in $\mathbb{R}^k$. The latter might not be equivalent to having nonempty Euclidean interior in $\mathcal{P}$. For example, this fails for the semialgebraic set $\mathcal{P} = \mathbb{R}^2_{>0} \sqcup \{p_1 + p_2 = 0\}$ if $\mathcal{Z}_{\mathcal{P}}(\mathcal{X})$ is contained in the line with equation $p_1 + p_2 = 0$. This phenomenon cannot happen if, with the Euclidean topology, the closure of $\mathcal{P}$ equals the closure of the interior of $\mathcal{P}$ in $\mathbb{R}^k$, as $\mathcal{P}$ is then locally Zariski dense. The latter will typically be the case in applications.

3. Vertically parametrized systems

We now turn our attention to vertically parametrized systems. We study their incidence varieties and the set of parameter values for which all zeros are degenerate, and use this to prove our main result Theorem 3.7. We end the section by commenting on how our results specialize for freely parametrized systems, as well as on conditions that ensure that our results extend from $\mathbb{C}^*$ to $\mathbb{C}$.

3.1. **Structure of the systems.** By a ***vertically parametrized system*** (or ***vertical system*** for short), we mean a parametric system $F \in \mathbb{C}[a, x^{\pm}]^s$ for $s \leq n$, with parameters $a = (a_1, \dots, a_m)$ and variables $x = (x_1, \dots, x_n)$, with the following properties:

- The coefficients of the monomials in x are homogeneous linear forms in a.
- Each a_i appears in at most one column of the coefficient matrix of F regarded in $\mathbb{C}(a)[x^{\pm}]^s$.
- The coefficient matrix of F regarded in $\mathbb{C}[a, x^{\pm}]^s$ has full rank.

This is equivalent to the existence of matrices $C \in \mathbb{C}^{s\times m}$ and $M \in \mathbb{Z}^{n\times m}$ such that

$$F = C(a \star x^M), \qquad \operatorname{rk}(C) = s \leq n. \tag{3.1}$$

The rows of C produce s linear combinations of m monomials encoded by the columns of M, where the i-th monomial is scaled by the parameter a_i. Note that M might have repeated columns; this corresponds to the same monomial appearing several times in the system, accompanied with different parameters.

More generally, we will consider ***augmented vertically parametrized systems*** of the form

$$\Big(C(a \star x^M),\ Lx - b\Big) \in \mathbb{C}[a, b, x^{\pm}]^{s+\ell}, \qquad s + \ell \leq n,$$

where we also include ℓ affine equations, encoded by a coefficient matrix $L \in \mathbb{C}^{\ell\times n}$ and parametric constant terms $b = (b_1, \dots, b_\ell)$. Geometrically, this corresponds to intersecting the variety given by $C(a \star x^M)$ by a parallel translate of $\ker(L)$. We observe that as C has full row rank, the coefficient matrix of the system F, regarded in $\mathbb{C}[a, b, x^{\pm}]^s$, has full rank, independently of the rank of L (c.f. Remark 2.2). When $\ell = 0$, an augmented vertical system is simply a vertical system.

Similarly to what we have done in the previous section, we will restrict the parameter values to some sets $\mathcal{A} \subseteq \mathbb{C}^m$ and $\mathcal{B} \subseteq \mathbb{C}^\ell$, and the variable values to some set $\mathcal{X} \subseteq (\mathbb{C}^*)^n$. In the notation of the previous section, the parameter space, number of parameters, and number of polynomials in the system are, respectively,

$$\mathcal{P} = \mathcal{A} \times \mathcal{B} \subseteq \mathbb{C}^m \times \mathbb{C}^\ell, \qquad k = m + \ell \qquad \text{and} \qquad r = s + \ell. \tag{3.2}$$

3.2. **The incidence variety.** We now show that the incidence varieties for augmented vertical systems are irreducible and nonsingular everywhere, and derive some basic facts about their geometry. In particular, it follows that Theorem 2.3 is applicable.

To this end, with $h^{-1} = (h_1^{-1}, \dots, h_n^{-1})$, we consider the map

$$\phi\colon\ \ker(C) \times (\mathbb{C}^*)^n \to \mathbb{C}^m \times \mathbb{C}^\ell \times (\mathbb{C}^*)^n, \quad (w, h) \mapsto (w \star h^M, Lh^{-1}, h^{-1}). \tag{3.3}$$

Theorem 3.1. *Let $F = (C(a \star x^M), Lx - b)$ be an augmented vertical system with $C \in \mathbb{C}^{s\times m}$ of full rank s, $L \in \mathbb{C}^{\ell\times n}$ with $s + \ell \leq n$ and $M \in \mathbb{Z}^{n\times m}$. Then the following statements hold:*

(i) The map ϕ is injective and $\mathcal{I} = \operatorname{im}(\phi)$.

(ii) $\mathcal{I}$ is irreducible of dimension $m + n - s$ and has no singular points.

Proof. To prove (i), injectivity is straightforward, and for surjectivity onto $\mathcal{I}$, we first note that if $(w, h) \in \ker(C) \times (\mathbb{C}^*)^n$, then

$$F(w \star h^M, Lh^{-1}, h^{-1}) = (C((w \star h^M) \star (h^{-1})^M), Lh^{-1} - Lh^{-1}) = (C(w), 0) = (0, 0).$$

Hence, $\operatorname{im}(\phi) \subseteq \mathcal{I}$. To show the reverse inclusion, given $(a, b, x) \in \mathcal{I}$, we have that $a \star x^M \in \ker(C)$. By letting $w = a \star x^M$ and $h = x^{-1}$, we obtain $\phi(w, h) = (a, b, x)$.

For part (ii), the irreducibility and dimension claims follow from the isomorphism of varieties $\mathcal{I} \cong \ker(C) \times (\mathbb{C}^*)^n$ from (i). To prove nonsingularity, we observe that each $(a, b, x) \in \mathcal{I}$ is a nondegenerate zero of F. Indeed, the Jacobian of F has the form

$$J_F(a, b, x) = \begin{bmatrix} C \operatorname{diag}(x^M) & 0_{s\times\ell} & * \\ 0_{\ell\times m} & -\mathrm{Id}_\ell & L \end{bmatrix}.$$

As both $C \operatorname{diag}(x^M)$ and Id_ℓ have maximal row rank for $x \in (\mathbb{C}^*)^n$, so has $J_F(a, b, x)$. □

3.3. **Nondegenerate zeros and generic consistency.** It follows from Proposition 2.14 that if an augmented vertical system $F = (C(a \star x^M), Lx - b)$ has a nondegenerate zero, then $\mathcal{Z}$ has nonempty Euclidean interior in the parameter space $\mathbb{C}^m \times \mathbb{C}^\ell$. We will see next that the converse also holds, that is, if $\mathcal{Z}$ has nonempty Euclidean interior, then necessarily the system F has a nondegenerate zero for some $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$. We start by deriving a parametrization of $\mathcal{Z}$ from (3.3), and proceed to study how degenerate zeros behave with respect to this parametrization. This, together with Sard's lemma, gives the main result Theorem 3.5 of this subsection.

Let $G \in \mathbb{C}^{m\times(m-s)}$ be Gale dual to C, in the sense that the columns of G form a basis for $\ker(C)$. The parametrization ϕ in (3.3) gives rise to a parametrization of $\mathcal{Z}$ via the Laurent polynomial map

$$\Phi := \pi \circ \phi \circ (G \times \mathrm{id}) \colon \mathbb{C}^{m-s} \times (\mathbb{C}^*)^n \to \mathbb{C}^m \times \mathbb{C}^\ell, \qquad (u, h) \mapsto ((Gu) \star h^M, Lh^{-1}),$$

where id is the identity on $(\mathbb{C}^*)^n$, and $\pi \colon \mathbb{C}^m \times \mathbb{C}^\ell \times (\mathbb{C}^*)^n \to \mathbb{C}^m \times \mathbb{C}^\ell$ the canonical projection.

Consider now the set of parameters that give rise to degenerate zeros:

$$\mathcal{D} := \{(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell : F_{a,b} \text{ has a degenerate zero in } (\mathbb{C}^*)^n\} \subseteq \mathcal{Z}. \tag{3.4}$$

We introduce the following matrix for $(w, h) \in \mathbb{C}^{m-s} \times (\mathbb{C}^*)^n$

$$Q(w, h) := \begin{bmatrix} C \operatorname{diag}(w) M^\top \operatorname{diag}(h) \\ L \end{bmatrix} \in \mathbb{C}^{(s+\ell)\times n}, \tag{3.5}$$

and the set

$$\Delta := \{(u, h) \in \mathbb{C}^{m-s} \times (\mathbb{C}^*)^n : \operatorname{rk}(Q(Gu, h)) < s + \ell\}.$$

Proposition 3.2. *Let $F = (C(a \star x^M), Lx - b)$ be an augmented vertical system and consider the notation above. Then the following statements hold:*

(i) Let $w \in \ker(C)$, $x, h \in (\mathbb{C}^*)^n$, $a \in \mathbb{C}^m$ and $b \in \mathbb{C}^\ell$ satisfy $(a, b, x) = \phi(w, h)$. Then

$$J_{F_{a,b}}(x) = Q(w, h).$$

(ii) $F_{a,b}$ has a nondegenerate zero in $(\mathbb{C}^*)^n$ for some $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$ if and only if $\operatorname{rk}(Q(w, h)) = s + \ell$ for some $w \in \ker(C)$ and $h \in (\mathbb{C}^*)^n$.

(iii) $\mathcal{Z} = \operatorname{im}(\Phi)$ and $\mathcal{D} = \Phi(\Delta)$.

Proof. An easy computation shows that the Jacobian of $F_{a,b}$ is given by

$$J_{F_{a,b}}(x) = \begin{bmatrix} C \operatorname{diag}(a \star x^M) M^\top \operatorname{diag}(x^{-1}) \\ L \end{bmatrix},$$

from which (i) follows. Statement (iii) follows from part (i) and Theorem 3.1(i). The description of $\mathcal{Z}$ in (ii) follows from Theorem 3.1(i), and that of $\mathcal{D}$ follows from this and part (i). □

Proposition 3.3. *With the notation above, Δ agrees with the set of critical points of Φ (equivalently, $J_\Phi(u,h)$ does not have full rank if and only if $(u,h) \in \Delta$).*

Proof. We start by noting that $(u,h) \in \mathbb{C}^{m-s} \times (\mathbb{C}^*)^n$ is a critical point of Φ if and only if $\dim(\ker(J_\Phi(u,h))) > n-(s+\ell)$, and that $(u,h) \in \Delta$ if and only if $\dim(\ker(Q(Gu,h))) > n-(s+\ell)$. Hence, all we need is to show that $\dim(\ker(J_\Phi(u,h))) = \dim(\ker(Q(Gu,h)))$.

A simple computation gives that

$$J_\Phi(u,h) = \begin{bmatrix} \operatorname{diag}(h^M) G & \operatorname{diag}(Gu \star h^M) M^\top \operatorname{diag}(h^{-1}) \\ 0 & -L \operatorname{diag}(h^{-2}) \end{bmatrix},$$

and hence its rank agrees with that of

$$P(u,h) = \begin{bmatrix} G & \operatorname{diag}(Gu) M^\top \operatorname{diag}(h) \\ 0 & L \end{bmatrix} \in \mathbb{C}^{(m+\ell)\times(m+n-s)}.$$

As $CG = 0$ we further have that

$$\begin{bmatrix} C & 0 \\ 0 & \mathrm{Id}_\ell \end{bmatrix} P(u,h) = \begin{bmatrix} 0_{(s+\ell)\times(m-s)} & Q(Gu,h) \end{bmatrix} \tag{3.6}$$

and

$$\ker\Big(\begin{bmatrix} C & 0 \\ 0 & \mathrm{Id}_\ell \end{bmatrix} \Big) \cap \operatorname{im}(P(u,h)) = \operatorname{im}(G) \times \{0\}.$$

From this and (3.6) we obtain

$$\begin{aligned} \dim(\ker(P(u,h)) &= \dim\Big(\ker\Big(\begin{bmatrix} C & 0 \\ 0 & \mathrm{Id}_\ell \end{bmatrix} P(u,h)\Big)\Big) - \dim\Big(\ker\Big(\begin{bmatrix} C & 0 \\ 0 & \mathrm{Id}_\ell \end{bmatrix}\Big) \cap \operatorname{im}(P(u,h))\Big) \\ &= \dim(\ker(Q(Gu,h)) + (m-s) - (m-s) = \dim(\ker(Q(Gu,h)). \end{aligned}$$ □

The next statement is given for zeros in $(\mathbb{C}^*)^n$ and for the full parameter space $\mathbb{C}^m \times \mathbb{C}^\ell$. We will later see in Theorem 3.7 that the statement also holds for more general sets $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{X}$.

Proposition 3.4. *The set $\mathcal{D}$ in (3.4) is contained in a hypersurface of $\mathbb{C}^m \times \mathbb{C}^\ell$.*

Proof. As $\mathcal{D} = \Phi(\Delta)$ by Proposition 3.2(iii), it follows from Proposition 3.3 that $\mathcal{D}$ is the set of critical values of Φ. Now Sard's Lemma (see, e.g., [SEDS17, Proposition B.2] and [Mum76, Proposition 3.7]), adapted to very affine varieties, gives that $\dim(\overline{\mathcal{D}}) < m + \ell$. □

We summarize our conclusions on degenerate zeros of augmented vertical systems as follows.

Theorem 3.5. *Let $F = (C(a \star x^M), Lx - b)$ be an augmented vertical system with $C \in \mathbb{C}^{s\times m}$ of full rank s, $L \in \mathbb{C}^{\ell\times n}$ with $s + \ell \le n$ and $M \in \mathbb{Z}^{n\times m}$. The following statements are equivalent:*

(i) For all $(a,b) \in \mathcal{Z}$, all zeros of $F_{a,b}$ in $(\mathbb{C}^*)^n$ are degenerate.

(ii) For every $(a,b) \in \mathcal{Z}$, it holds that $F_{a,b}$ has a degenerate zero in $(\mathbb{C}^*)^n$, i.e., $\mathcal{D} = \mathcal{Z}$.

(iii) $\mathcal{Z}$ has empty Euclidean interior in $\mathbb{C}^m \times \mathbb{C}^\ell$.

Additionally, if $\mathcal{Z}$ has nonempty Euclidean interior in $\mathbb{C}^m \times \mathbb{C}^\ell$, then all zeros of $F_{a,b}$ are nondegenerate for generic choices of $(a,b) \in \mathbb{C}^m \times \mathbb{C}^\ell$.

Proof. The implication (i)⇒(ii) holds trivially, while (ii)⇒(iii) is a consequence of Proposition 3.4. The implication (iii)⇒(i) follows by Proposition 2.14. For the last statement, it is enough to consider the Zariski open subset $\mathcal{U} = \mathcal{Z} \setminus \overline{\mathcal{D}}$ of $\mathcal{Z}$, which is nonempty by Proposition 3.4. □

Remark 3.6. For $\mathbb{G} = \mathbb{R}^*$ and $\mathbb{G} = \mathbb{R}_{>0}$, it is straightforward to verify that restricting the map $\phi\colon \ker(C) \times (\mathbb{C}^*)^n \to \mathcal{I}$ to $(\ker(C) \cap \mathbb{G}^m) \times \mathbb{G}^n$ provides a parametrization of $\mathcal{I} \cap ((\mathbb{G}^m \times \mathbb{R}^\ell) \times \mathbb{G}^n)$, which after projection gives a parametrization of $\mathcal{Z}_{\mathbb{G}^m \times \mathbb{R}^\ell}(\mathbb{G}^n)$.

3.4. **The main theorem on generic dimension.** We now apply the results of the previous subsections to complete the equivalences in Theorem 2.17, and thereby unify Theorem A and Theorem B from the introduction. Recall from (3.2) the relation between the notation in Theorem 2.17 and the current section, and recall the matrix $Q(w,h)$ defined in (3.5).

Theorem 3.7. *Let $F = (C(a \star x^M), Lx - b)$ be an augmented vertical system with $C \in \mathbb{C}^{s\times m}$ of full rank s, $L \in \mathbb{C}^{\ell\times n}$ with $s + \ell \le n$ and $M \in \mathbb{Z}^{n\times m}$. Let $\mathcal{P} = \mathcal{A} \times \mathcal{B} \subseteq \mathbb{C}^m \times \mathbb{C}^\ell$ and suppose that $(\mathcal{P}, \mathcal{X})$ is an algebraically defined pair of sets, for which $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$. Then the following holds:*

(i) The statements (deg1)*,* (degX1)*,* (degXG)*,* (degAllG)*,* (setE)*,* (setZ)*,* (flatG)*,* (dim1) *and* (dimG) *are all equivalent to the condition*

$$\operatorname{rk}(Q(w,h)) = s + \ell \quad \text{for some } (w,h) \in \ker(C) \times (\mathbb{C}^*)^n.$$

(ii) Any of the statements mentioned above implies (dimX)*,* (reg) *and* (real)*.*

(iii) The statement (rad) *holds, independently of the other statements.*

Proof. The equivalence between (deg1) and the matrix rank condition is Proposition 3.2(ii). As $\mathcal{I}$ is irreducible and nonsingular of dimension $m + n - s$, the assumptions for Theorem 2.17 are satisfied, and there is an equivalence between (setZ) and (flatG). Therefore, for (i) and (ii), it suffices to show that (setZ) implies (degAllG) and (setE). The equivalence between (setZ) and (setE) is Proposition 2.22(ii). By Lemma 2.8 and Proposition 2.22, $\mathcal{Z}$ has nonempty Euclidean interior in $\mathbb{C}^m \times \mathbb{C}^\ell$ if (setZ) holds. By Theorem 3.5, this implies (degAllG). Finally, to prove (iii), we note that if (degAllG) holds, so does (rad) by Theorem 2.17. If, on the other hand, (degAllG) does not hold, then (i) gives that (setZ) does not hold either, and hence it follows by Theorem 2.3 that $\mathbb{V}_{\mathbb{C}^*}(F_{a,b}) = \varnothing$ for generic $(a,b) \in \mathbb{C}^m \times \mathbb{C}^\ell$. Thus, generically, $\langle F_{a,b}\rangle = \mathbb{C}[x^{\pm}]$, which is radical. □

Remark 3.8.

(i) As indicated in Theorem 2.17(i), the implication (deg1)⇒(degAllG) already followed from Theorem 2.15 for square augmented vertical systems. The results of this section have shown that this also holds in the non-square case.

(ii) For any F, $\mathcal{P}$ and $\mathcal{X}$ satisfying the assumptions of Theorem 3.7, any statement in Theorem 3.7(i) is equivalent to (deg1), and hence to any of the same statements for $\mathcal{P} = \mathbb{C}^m \times \mathbb{C}^\ell$ and $\mathcal{X} = (\mathbb{C}^*)^n$ instead.

(iii) If (degAllG) holds under the assumptions of Theorem 3.7, then we also have that all zeros of $F_{a,b}$ are nondegenerate for generic $(a,b) \in \mathcal{Z}_{\mathcal{P}}(\mathcal{X})$.

(iv) When $\ell = 0$, the rank condition in Theorem 3.7(i) simplifies to

$$\operatorname{rk}(C \operatorname{diag}(w) M^\top) = s \quad \text{for some } w \in \ker(C).$$

(v) If a vertical system $C(a \star x^M)$ is generically consistent, then so is the augmented vertical system $(C(a \star x^M), Lx - b)$ for generic $L \in \mathbb{C}^{\ell\times n}$ for $\ell \le n - s$.

Theorem 3.7 tells us that for augmented vertical systems, only one of two extreme scenarios occurs, provided the parameter sets $\mathcal{A}$ and $\mathcal{B}$ and the domain $\mathcal{X}$ of the variables satisfy some mild conditions: Either the system is generically consistent, has generically the expected dimension $n - (s + \ell)$, and generically all zeros are nondegenerate, or the system is generically inconsistent, has never the expected dimension over $\mathbb{C}$, and all zeros are degenerate.

In Section 3.5 we will discuss, and show several examples of, how to computationally check which of these two scenarios we are in, for a given augmented vertical system.

Remark 3.9. The conclusions of Theorem 3.7 also hold for parametric systems $F_{\cdot,b}$ given by the restriction of an augmented vertical system F to a fixed value of $b \in L((\mathbb{C}^*)^n)$. In this case, the incidence variety is irreducible and nonsingular of dimension $n+m-(s+\ell)$, with parametrization

$$\ker(C) \times U_b \to \mathbb{C}^m \times (\mathbb{C}^*)^n, \qquad (w,h) \mapsto (w \star h^M, h^{-1}),$$

where $U_b := \{h \in (\mathbb{C}^*)^n : Lh^{-1} = b\}$. Using this, the proof of Proposition 3.3 (and subsequently Theorem 3.7) can be adapted to this system.

Remark 3.10 (Non-vertical systems)**.** One of the restrictions we imposed on vertical systems is that each parameter always multiplies the same monomial. Relaxing this restriction might make Theorem 3.7 fail. In particular, this more general class of systems might not give rise to irreducible incidence varieties, and hence the implications in Theorem 2.17 do not necessarily hold. For example, (setZ) and (deg1) might hold, but not (dimG). Additionally, the implication (setZ)⇒(deg1), which holds for vertical systems by Theorem 3.7, might not hold. This is illustrated by the examples we saw in (1.6) and (1.7).

3.5. **Computational considerations.** Consider an augmented vertical system with the following scenarios regarding the sets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{X}$, which are common in applications:

(a) $\mathcal{B} = \mathbb{R}^\ell$ and
$(\mathcal{A}, \mathcal{X}) \in \{(\mathbb{R}^m_{>0}, \mathbb{R}^n_{>0}), ((\mathbb{R}^*)^m, \mathbb{R}^n_{>0}), ((\mathbb{R}^*)^m, (\mathbb{R}^*)^n), (\mathbb{R}^m, \mathbb{R}^n_{>0}), (\mathbb{R}^m, (\mathbb{R}^*)^n)\}$.

(b) $\mathcal{A} \in \{(\mathbb{C}^*)^m, \mathbb{C}^m\}$, $\mathcal{B} = \mathbb{C}^\ell$, and $\mathcal{X} = (\mathbb{C}^*)^n$.

For all of these, $\mathcal{P} = \mathcal{A} \times \mathcal{B}$ is locally Zariski dense and the pair $(\mathcal{P}, \mathcal{X})$ is algebraically defined. Assuming F has real coefficients and $\mathcal{X} \subseteq (\mathbb{R}^*)^n$, it follows from Remark 2.9 that $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}$ whenever $\mathcal{I} \cap (\mathcal{P} \times \mathcal{X}) \neq \varnothing$. This, in turn, is characterized as follows.

Proposition 3.11. *Let $\mathcal{A}, \mathcal{B}, \mathcal{X}$ be as in (a) or (b) above, and F be an augmented vertical system (with real coefficients in case (a)). Then $\mathcal{I} \cap (\mathcal{A} \times \mathcal{B} \times \mathcal{X}) \neq \varnothing$ if and only if* $\ker(C) \cap \mathcal{A} \neq \varnothing$.

Proof. The conditions on $\mathcal{A}$ and $\mathcal{X}$ imply that for all $a \in \mathcal{A}$ and $x \in \mathcal{X}$, it holds $a \star x^M \in \mathcal{A}$. Hence, if $\ker(C) \cap \mathcal{A} = \varnothing$, then $\mathbb{V}_{\mathbb{C}^*}(F_{a,b}) = \varnothing$ for all $(a,b) \in \mathcal{A} \times \mathcal{B}$. On the other hand, if $\ker(C) \cap \mathcal{A} \neq \varnothing$, then $(1,\ldots,1) \in \mathbb{V}_{\mathbb{C}^*}(F_{a,b}) \cap \mathcal{X}$ for all $a \in \ker(C) \cap \mathcal{A}$, with $b := L(1,\ldots,1)^\top \in \mathcal{B}$. □

In the special case $\mathcal{A} = \mathbb{R}^m_{>0}$, checking $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$ corresponds to showing the existence of an interior point of the polyhedral cone $\ker(C) \cap \mathbb{R}^m_{\geq 0}$. If C has rational entries, this is a straightforward computation using linear programming.

The rank condition from Theorem 3.7 can be checked in the following way: Let $G \in \mathbb{C}^{m \times (m-s)}$ be a Gale dual to C, whose columns form a basis for $\ker(C)$. We want to check whether there exists some $(u,h) \in \mathbb{C}^{m-s} \times (\mathbb{C}^*)^n$ such that

$$\operatorname{rk} \begin{bmatrix} C \operatorname{diag}(Gu) M^\top \operatorname{diag}(h) \\ L \end{bmatrix} = s + \ell. \tag{3.7}$$

Equality (3.7) holds for all (u,h) in a Zariski open subset of $\mathbb{C}^{m-s} \times (\mathbb{C}^*)^n$, so if this set is nonempty, then (3.7) holds for a randomly chosen (u,h) with probability 1, given an appropriate probability measure on $\mathbb{C}^{m-s} \times (\mathbb{C}^*)^n$. Hence, we pick a random pair (u,h) and compute the rank in (3.7) with exact arithmetic. If the rank is $s+\ell$, then we are in the generically consistent scenario. If not, we can suspect that we are in the generically inconsistent scenario, and to conclusively prove this, we view the matrix in (3.7) as a symbolic matrix with indeterminates (u,h) and verify that all $(s+\ell)$-minors are the zero polynomial.

In the special case when $\ell = 0$, it follows from Remark 3.8(iv) that it suffices to check whether there is some $u \in \mathbb{C}^{m-s}$ such that

$$\operatorname{rk}(C \operatorname{diag}(Gu) M^\top) = s. \tag{3.8}$$

Example 3.12. Consider the vertical system (1.2) from the introduction. Using the matrices

$$C = \begin{bmatrix} 1 & 3 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{bmatrix}, \quad M = \begin{bmatrix} 2 & 2 & 0 & 1 \\ 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 5 & 3 \\ -2 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we obtain

$$C \operatorname{diag}(Gu) M^\top = \begin{bmatrix} -2u_1 & 10u_1 + 6u_2 & -5u_1 - 3u_2 & 2u_1 \\ -4u_1 - u_2 & 0 & 0 & 4u_1 + u_2 \end{bmatrix},$$

and we see that (3.8) holds for $u = (1, 1)$. Hence, the system (1.2) is generically consistent. Consider now the augmented vertical system (1.3), with C and M as above, and

$$L = \begin{bmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

In this case, the matrix in condition (3.7) is square, and its determinant is the zero polynomial, when (u, h) are viewed as variables. We conclude that system (1.3) is generically inconsistent. (We do, however, obtain a generically consistent system for any choice of $L \in \mathbb{C}^{2\times 4}$ with full rank, such that column 1 or 4, as well as column 2 or 3, have at least one nonzero entry.)

Example 3.13. Going back to the application to critical points in optimization mentioned in the introduction, we consider bivariate polynomials of the form $f = a_1x_1 + a_2x_1x_2 + a_3x_2^2$. The matrix M in the square vertical system in (1.9) and a Gale dual matrix are

$$M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}, \qquad G = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

This gives that

$$M \operatorname{diag}(Gu) M^\top = \begin{bmatrix} 0 & -2u \\ -2u & 2u \end{bmatrix},$$

which has rank 2 for all $u \neq 0$. As the rank condition in (1.10) holds, f_a has a finite, positive number of critical points in $(\mathbb{C}^*)^2$ for generic $a \in \mathbb{C}^3$. If we now impose the coefficients to be real such that $a_1, a_3 > 0$ and $a_2 < 0$, the polynomial f can be written instead as $f = a_1x_1 - a_2x_1x_2 + a_3x_2^2$ with $a_i > 0$. The coefficient matrix of the vertical system encoding the critical points becomes

$$C = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 2 \end{bmatrix},$$

which satisfies $\ker(C) \cap \mathbb{R}^3_{>0} \neq \varnothing$. From this, Theorem 3.7 with $\mathcal{A} = \mathbb{R}^3_{>0}$ and $\mathcal{X} = \mathbb{R}^2_{>0}$ tells us that f_a has positive critical points for a in a subset of $\mathbb{R}^3_{>0}$ with nonempty Euclidean interior.

3.6. **Freely parametrized systems.** We now turn our attention to the special case of freely parametrized systems, and record some corollaries of our main results.

Given finite sets $\mathcal{S}_1, \ldots, \mathcal{S}_s \subseteq \mathbb{Z}^n$, we consider the corresponding freely parametrized family

$$\mathcal{F}_{\text{free}}(\mathcal{S}_1, \ldots, \mathcal{S}_s) := \{F = (f_1, \ldots, f_s) \in \mathbb{C}[x^{\pm}]^s : \operatorname{supp}(f_i) \subseteq \mathcal{S}_i \text{ for all } i = 1, \ldots, s\}$$

and say that a property holds generically in $\mathcal{F}_{\text{free}}(\mathcal{S}_1, \ldots, \mathcal{S}_s)$ if it does so under the isomorphism $\mathcal{F}_{\text{free}}(\mathcal{S}_1, \ldots, \mathcal{S}_s) \cong \prod_{i=1}^s \mathbb{C}^{\mathcal{S}_i}$ that identifies each polynomial with its coefficients.

The set $\mathcal{F}_{\text{free}}(\mathcal{S}_1, \ldots, \mathcal{S}_s)$ can be encoded as a vertical system of the form (3.1) in the following way. Let m_i be the cardinality of $\mathcal{S}_i$, set $m = m_1 + \cdots + m_s$, and define $C \in \mathbb{C}^{s\times m}$ to be the block diagonal matrix

$$C = \begin{bmatrix} C_1 & 0 & \ldots & 0 \\ 0 & C_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & C_s \end{bmatrix}, \qquad \text{with} \quad C_i = [1 \ \cdots \ 1] \in \mathbb{C}^{1\times m_i}. \tag{3.9}$$

Similarly, let $M = [M_1 \ \cdots \ M_s] \in \mathbb{Z}^{n\times m}$ be the block matrix where the columns of $M_i \in \mathbb{Z}^{n\times m_i}$ are the elements of $\mathcal{S}_i$ in some fixed order. For the vertical system $F = C(a \star x^M)$, it then holds that $\mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s) = \{F_a : a \in \mathbb{C}^m\}$. We refer to this F as the ***vertical system associated with*** $\mathcal{S}_1, \dots, \mathcal{S}_s$.

Let $V_1, \dots, V_s \subseteq \mathbb{R}^n$ be sets of vectors. Following the terminology in [Per69, Yu16], we define an ***independent transversal*** of $(V_1, \dots, V_s)$ to be a linearly independent tuple $(v_1, \dots, v_s) \in \prod_{i=1}^s V_i$. It is a well-known linear algebra fact (see, e.g., [Per69, Theorem 1] and [Kho16, Theorem 4]) that the existence of such an independent transversal is equivalent to

$$\dim\Big(\sum_{j\in J} \operatorname{span}_{\mathbb{R}}(V_i)\Big) \le |J| \quad \text{for all } J \subseteq [s]. \tag{3.10}$$

Consider now support sets $\mathcal{S}_1, \dots, \mathcal{S}_s \subseteq \mathbb{Z}^n$, and let

$$\operatorname{Lin}(\mathcal{S}_i) := \operatorname{span}_{\mathbb{R}}\{u - v : u, v \in \mathcal{S}_i\} \subseteq \mathbb{R}^n$$

denote the direction of the affine hull of $\mathcal{S}_i$. Then $(\mathcal{S}_1, \dots, \mathcal{S}_s)$ is called an ***essential family*** in [Stu94, BS24] if $(\operatorname{Lin}(\mathcal{S}_1), \dots, \operatorname{Lin}(\mathcal{S}_s))$ satisfies (3.10). It is shown in [Yu16, Lemma 1] and [Kho16, Theorem 11] that this completely characterizes when $\mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$ is generically consistent. We recover this fact as a corollary of Theorem 3.7.

Corollary 3.14. *Let $\mathcal{S}_1, \dots, \mathcal{S}_s \subseteq \mathbb{Z}^n$, and let $\mathcal{G}_i \subseteq \mathbb{C}^n$ be a generating set of $\operatorname{Lin}(\mathcal{S}_i)$ for $i = 1, \dots, s$. The following statements are equivalent:*

(i) The dimension of $\mathbb{V}_{\mathbb{C}^*}(F)$ is $n - s$ for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$.

(ii) The tuple $(\mathcal{G}_1, \dots, \mathcal{G}_s)$ admits an independent transversal.

(iii) The tuple $(\operatorname{Lin}(\mathcal{S}_1), \dots, \operatorname{Lin}(\mathcal{S}_s))$ admits an independent transversal.

Furthermore:

- *If the equivalent statements hold, then $\mathbb{V}_{\mathbb{C}^*}(F)$ is pure-dimensional and all zeros of F are nondegenerate for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$.*
- *If the equivalent statements do not hold, then $\mathbb{V}_{\mathbb{C}^*}(F)$ is empty for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$, with dimension strictly larger than $n - s$ if nonempty.*
- *The ideal $\langle F \rangle$ is radical for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$.*

Proof. Let $F = C(a \star x^M)$ be the vertical system associated with $\mathcal{S}_1, \dots, \mathcal{S}_s$. By Theorem 3.7, for $\mathcal{P} = \mathbb{C}^m$ and $\mathcal{X} = (\mathbb{C}^*)^n$, statement (i) is equivalent to $\operatorname{rk}(C \operatorname{diag}(Gu) M^\top) = s$ for some $u \in \mathbb{C}^{m-s}$, with G a matrix whose columns form a basis of $\ker(C)$. By the form of C, we can choose G to be

$$G = \begin{bmatrix} G_1 & 0 & \dots & 0 \\ 0 & G_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & G_s \end{bmatrix}, \qquad \text{with} \quad G_i = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{bmatrix} \in \mathbb{C}^{m_i \times (m_i - 1)}.$$

We index $u \in \mathbb{C}^{m-s}$ in block form as $u = (u_1, \dots, u_s)$ with $u_i = (u_{i1}, \dots, u_{i,m_i-1}) \in \mathbb{C}^{m_i - 1}$. Then, the product $C \operatorname{diag}(Gu)$ is the block diagonal matrix of the same shape as C but with blocks given by

$$C_i \operatorname{diag}(G_i u_i) = [u_{i1} \quad \dots \quad u_{i,m_i-1} \quad -(u_{i1} + \cdots + u_{i,m_i-1})].$$

The i-th row of $C \operatorname{diag}(Gu) M^\top$ is then the row vector

$$\big(u_{i1}(M_{i1} - M_{i,m_i}) + \cdots + u_{i,m_i-1}(M_{i,m_i-1} - M_{i,m_i})\big)^\top.$$

The set of vectors obtained by varying all the entries of u_i is precisely $\operatorname{Lin}(\mathcal{S}_i)$. Hence, we have shown that $\operatorname{rk}(C \operatorname{diag}(Gu) M^\top) = s$ for some $u \in \mathbb{C}^{m-s}$ (i.e. (i) holds) if and only if (iii) holds.

Clearly, (ii) implies (iii). For the reverse implication, assume that (ii) does not hold. Then, for all choices of $\omega_i \in \mathcal{G}_i$, $i = 1, \dots, s$, the matrix with columns $\omega_1, \dots, \omega_s$ has rank smaller than s; that is, all s-minors are zero. For any $(v_1, \dots, v_s) \in \prod_{i=1}^{s} \operatorname{Lin}(\mathcal{S}_i)$, we have that v_i is a linear combination of the elements of $\mathcal{G}_i$. Then, the multilinear expansion of the determinant gives that each s-minor of the matrix with columns $v_1, \dots, v_s$ is a linear combination of s-minors of matrices with i-th column in $\mathcal{G}_i$, $i = 1, \dots, s$. Hence it is zero, showing that (iii) does not hold. This concludes the proof of (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii).

The bullet points are a direct consequence of Theorem 3.7. □

Example 3.15. Consider the freely parametrized system

$$\widetilde{F} = \begin{pmatrix} a_1x_1x_3 + a_2x_2x_3 + a_3x_3 + a_4 \\ a_5x_1^2 + a_6x_2x_3 + a_7x_3 + a_8 \\ a_9x_1^2 + a_{10}x_2x_3 + a_{11}x_3 + a_{12} \end{pmatrix}$$

with supports

$$\mathcal{S}_1 = \{(1,0,1), (0,1,1), (0,0,1), (0,0,0)\}, \quad \mathcal{S}_2 = \mathcal{S}_3 = \{(2,0,0), (0,1,1), (0,0,1), (0,0,0)\}.$$

This system is generically consistent since there exists a linearly independent tuple

$$\Big((1,0,1), (2,0,0), (0,1,1)\Big) \in \operatorname{Lin}(\mathcal{S}_1) \times \operatorname{Lin}(\mathcal{S}_2) \times \operatorname{Lin}(\mathcal{S}_3).$$

Remark 3.16. If a vertical system F is generically consistent, so is any other vertical system $\widetilde{F}$ with the same supports and less dependencies between the coefficients (systems (1.1) and (1.2) in the introduction provide an example of this). However, restricting a vertical system by keeping the supports and adding dependencies among the coefficients does not preserve generic consistency. For example, consider the vertical system

$$F = \begin{pmatrix} a_1x_1x_3 + a_2x_2x_3 + a_3x_3 + a_4 \\ a_5x_1^2 + a_2x_2x_3 + a_3x_3 + a_4 \\ a_6x_1^2 + a_2x_2x_3 + a_3x_3 + a_4 \end{pmatrix}, \tag{3.11}$$

which can be seen as a specialization of the generically consistent system from Example 3.15. It is clear from inspection that the zero locus is empty unless $a_5 = a_6$, and in this case, the dimension of the zero locus is 1. The system is thus generically inconsistent.

The specialization of Theorem 3.7 for vertical systems to the case $\mathcal{A} = \mathbb{R}^m_{>0}$ and $\mathcal{X} = \mathbb{R}^n_{>0}$ allows us to study freely parametrized systems with real coefficients with prescribed sign. Specifically, we consider maximal dimensional orthants $\mathcal{O}_i$ of $(\mathbb{R}^*)^{\mathcal{S}_i}$, and consider the subset $\mathcal{F}_{\mathcal{O}_1,\dots,\mathcal{O}_s}(\mathcal{S}_1, \dots, \mathcal{S}_s)$ of $\mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$ consisting of the polynomials in the image of $\prod_{i=1}^{s} \mathcal{O}_i$ under the isomorphism $\prod_{i=1}^{s} \mathbb{C}^{\mathcal{S}_i} \cong \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$.

Corollary 3.17. *Let $\mathcal{S}_1, \dots, \mathcal{S}_s \subseteq \mathbb{Z}^n$. For $i = 1, \dots, s$, let $\mathcal{G}_i \subseteq \mathbb{C}^n$ be a generating set of $\operatorname{Lin}(\mathcal{S}_i)$ and $\mathcal{O}_i$ be a maximal dimensional orthant of $(\mathbb{R}^*)^{\mathcal{S}_i}$. The following statements are equivalent:*

(i) The semialgebraic set $\mathbb{V}_{\mathbb{R}^*}(F) \cap \mathbb{R}^n_{>0}$ has dimension $n - s$ for F in a subset of $\mathcal{F}_{\mathcal{O}_1,\dots,\mathcal{O}_s}(\mathcal{S}_1, \dots, \mathcal{S}_s)$ with nonempty Euclidean interior.

(ii) The tuple $(\mathcal{G}_1, \dots, \mathcal{G}_s)$ admits an independent transversal and $\mathcal{O}_i \notin \{\mathbb{R}^{\mathcal{S}_i}_{>0}, \mathbb{R}^{\mathcal{S}_i}_{<0}\}$ for all $i = 1, \dots, s$.

Proof. Let $F = C(a \star x^M)$ be the vertical system associated with $\mathcal{S}_1, \dots, \mathcal{S}_s$, and let $\widetilde{C}$ be obtained by multiplying by -1 each column of C_i in (3.9) corresponding to a negative coordinate of $\mathcal{O}_i$. Then $\mathcal{F}_{\mathcal{O}_1,\dots,\mathcal{O}_s}(\mathcal{S}_1, \dots, \mathcal{S}_s)$ is identified with the vertical system $\widetilde{C}(a \star x^M)$ with $\mathcal{A} = \mathbb{R}^m_{>0}$. The result now follows from Corollary 3.14, Theorem 3.7 and Proposition 3.11, after realizing that $\ker(\widetilde{C}) \cap \mathbb{R}^n_{>0} \neq \varnothing$ if and only if $\mathcal{O}_i \neq \mathbb{R}^{\mathcal{S}_i}_{>0}, \mathbb{R}^{\mathcal{S}_i}_{<0}$ for all $i = 1, \dots, s$. □

In the freely parametrized setting, it is known from [BS24, Proposition 1.12] that one can obtain the generic properties predicted by Corollary 3.14, by only requiring the coefficients corresponding to the vertices of the Newton polytopes to be generic (while the other coefficients can be arbitrarily fixed). It is an interesting problem for future research to determine which parameters can be fixed in a vertical system while still retaining the generic properties given by Theorem 3.7.

3.7. **Extension to $\mathbb{C}$.** A natural question is to what extent Theorem 3.7 also holds for zeros in the whole affine space $\mathbb{C}^n$ rather than just in the torus $(\mathbb{C}^*)^n$. As explained in Remark 2.19, we can extend Theorem 2.17 to zeros in $\mathbb{C}^n$, but our proofs of the irreducibility of $\mathcal{I}$ for augmented vertical systems and of Theorem 3.5 rely heavily on the fact that we can invert the entries of $x \in (\mathbb{C}^*)^n$. The following example illustrates that the affine incidence variety

$$\mathcal{I}_{\mathbb{C}} = \{(a, b, x) \in \mathbb{C}^m \times \mathbb{C}^\ell \times \mathbb{C}^n : F_{a,b}(x) = 0\}$$

might be reducible for an augmented vertical system.

Example 3.18. For the following variation of Example (3.11)

$$\begin{pmatrix} a_1x_1x_3 + a_2x_2x_3 + a_3x_3 + a_4 \\ a_5x_1^2 + a_2x_2x_3 + a_3x_3 + a_4 \\ a_6x_1^3 + a_2x_2x_3 + a_3x_3 + a_4 \end{pmatrix},$$

the affine incidence variety has two irreducible components:

$$\mathcal{I}_{\mathbb{C}} = \mathbb{V}_{\mathbb{C}}(x_1, a_2x_2x_3 + a_3x_3 + a_4) \cup \mathbb{V}_{\mathbb{C}}(a_1x_3 - a_6x_1^2, a_1x_3 - a_5x_1, a_1x_1x_3 + a_2x_2x_3 + a_3x_3 + a_4).$$

The system is generically consistent over $\mathbb{C}^*$ (and thus also over $\mathbb{C}$), and hence the zero locus over $\mathbb{C}^*$ is generically 0-dimensional. However, the zero locus over $\mathbb{C}$ is generically 1-dimensional, due to the curve in the $\{x_1 = 0\}$ coordinate hyperplane, which contradicts the expected dimension of $n - s = 0$.

A simple criterion that guarantees that the affine incidence variety is irreducible is that each polynomial appearing in F has a constant term involving a unique parameter. More precisely, we have the following result, where (i) is an affine analog of Theorem 3.1, (ii) is an affine analog of Theorem 3.7, and (iii) can be seen as an extension of [LW96, Lemma 2.1] to augmented vertical systems.

Theorem 3.19. *Let $F = (C(a \star x^M), Lx - b) \in \mathbb{C}[a, b, x]^{s+\ell}$ be an augmented vertical system with $C \in \mathbb{C}^{s\times m}$ of full rank s, $L \in \mathbb{C}^{\ell\times n}$ with $s + \ell \leq n$ and $M \in \mathbb{Z}_{\geq 0}^{n\times m}$. Suppose that for some indices $i_1 < \cdots < i_s$, the submatrix of C given by the columns with these indices is diagonal of rank s, and that the corresponding submatrix of M is the zero matrix. Then the following holds:*

- *(i) The affine incidence variety $\mathcal{I}_{\mathbb{C}}$ is nonsingular and irreducible of dimension $m + n - s$.*
- *(ii) Suppose that $(\mathcal{A} \times \mathcal{B}, \mathcal{X}) \subseteq \mathbb{C}^m \times \mathbb{C}^\ell \times \mathbb{C}^n$ is an algebraically defined pair, and assume that $\mathcal{I}_{\mathbb{C}} \cap (\mathcal{A} \times \mathcal{B} \times \mathcal{X})$ is Zariski dense in $\mathcal{I}_{\mathbb{C}}$. Then the affine analogs of* (deg1), (degX1), (degXG), (degAllG), (setE), (setZ), (flatG), (dim1), (dimG) *are equivalent, any of these statements imply* (dimX), (real) *and* (reg), *and* (rad) *holds independently of the other statements.*
- *(iii) For generic $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$, the variety $\mathbb{V}_{\mathbb{C}}(F_{a,b})$ has no irreducible component contained in a coordinate hyperplane of $\mathbb{C}^n$.*

Proof. To prove (i), we parametrize the affine incidence variety as follows. Without loss of generality, we assume that the first s columns of C form the diagonal matrix. Then $a_1, \ldots, a_s$ are terms of the entries $1, \ldots, s$ of $C(a \star x^M)$ respectively. Hence $C(a \star x^M) = 0$ is equivalent to $(a_1, \ldots, a_s) = \psi(a', x)$ where $a' = (a_{s+1}, \ldots, a_m)$ and ψ is a polynomial function. This gives the parametrization $(\psi(a', x), a', Lx, x)$ of $\mathcal{I}_{\mathbb{C}}$ in terms of $m + n - s$ free parameters. To show nonsingularity, we proceed analogously to the proof of Theorem 3.1(ii), after noting that the first s columns of $C \operatorname{diag}(x^M)$ are independent of $x \in \mathbb{C}^n$ and form a diagonal matrix of full rank s.

For statement (ii), the equivalences in Theorem 2.17 and the equivalence between (setZ) and (setE) hold by part (i), Remark 2.19 and the extension of Proposition 2.22(ii) to $\mathbb{C}$. The only obstruction lies in proving that (setZ) implies (degAllG); with that in place, the rest of the claims in (ii) follow analogously to the proof of Theorem 3.7.

The proof of (setZ) $\Rightarrow$(degAllG) is based on the following construction. For every subset $I \subseteq \{1, \dots, n\}$, let $F_I \in \mathbb{C}[a, b, x]^{s+\ell}$ be the system obtained by letting $x_i = 0$ for $i \notin I$ in F. As each entry of F_I has a constant term, the coefficient matrix of the vertically parametrized part of F_I has maximal rank. Hence F_I is again an augmented vertical system.

Let $\mathcal{X}^* := \mathcal{X} \cap (\mathbb{C}^*)^n$ and let $\Theta_I \subseteq \mathbb{C}^n$ be the set defined by $x_i = 0$ if $i \notin I$. For $(a, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$, there are surjective maps $\psi_I \colon \mathbb{V}_{\mathbb{C}^*}(F_{I,a,b}) \to \mathbb{V}_{\mathbb{C}}(F_{a,b}) \cap \Theta_I$ obtained by sending $x^* \in \mathbb{V}_{\mathbb{C}^*}(F_{I,a,b})$ to $x' \in \Theta_I$ defined as $x'_i = x^*_i$ if $i \in I$ and zero otherwise. We obtain in particular that

$$\mathbb{V}_{\mathbb{C}}(F_{a,b}) = \bigcup_{I \subseteq \{1,\dots,n\}} \psi_I(\mathbb{V}_{\mathbb{C}^*}(F_{I,a,b})) \quad \text{and} \quad \mathcal{Z}_{F,\mathcal{P}}(\mathcal{X}) = \bigcup_{I \subseteq \{1,\dots,n\}} \mathcal{Z}_{F_I,\mathcal{P}}(\mathcal{X}^*). \tag{3.12}$$

(Here and below, subindices F, F_I are added to differentiate among the augmented vertical systems.) Additionally, the columns of $J_{F_{a,b}}(\psi_I(x^*))$ and of $J_{F_{I,a,b}}(x^*)$ indexed by I agree, and $J_{F_{I,a,b}}(x^*)$ is zero outside these columns. Hence, if x^* is a nondegenerate zero of $F_{I,a,b}$, then so is $\psi_I(x^*)$ as a zero of $F_{a,b}$.

Set $\mathcal{P} = \mathcal{A} \times \mathcal{B}$. If (setZ) holds for F, then $\mathcal{Z}_{F_I,\mathcal{P}}(\mathcal{X}^*)$ is Zariski dense for at least one F_I by the second equality in (3.12). In this case, so is $\mathcal{Z}_{F_I} = \mathcal{Z}_{F_I,\mathbb{C}^m \times \mathbb{C}^\ell}((\mathbb{C}^*)^n)$ and hence, using the implication (setZ) $\Rightarrow$(degAllG) from Theorem 3.7 for F_I, there exists a nonempty Zariski open subset $U_I \subseteq \mathbb{C}^m \times \mathbb{C}^\ell$ such that all zeros of $F_{I,a,b}$ in $(\mathbb{C}^*)^n$ are nondegenerate for $(a, b) \in U_I$. Consider the nonempty Zariski open set U obtained by intersecting U_I over all subsets I for which $\mathcal{Z}_{F_I,\mathcal{P}}(\mathcal{X}^*)$ is Zariski dense, and removing the Zariski closed sets $\overline{\mathcal{Z}_{F_I,\mathcal{P}}(\mathcal{X}^*)}$ for all other subsets I. Then, for all $(a, b) \in U$ and for all I, all zeros of $F_{I,a,b}$ in $(\mathbb{C}^*)^n$ are nondegenerate, and hence so are all zeros of $F_{a,b}$ in $\mathbb{C}^n$, which gives (degAllG).

Finally, to prove (iii), we begin by noting that the statement is trivial if F is generically inconsistent. Hence, we assume F is generically consistent. If we set $x_i = 0$, we obtain an augmented vertically parametrized system $F_{|x_i=0}$ with independent constant terms and $n - 1$ variables. Suppose this system is generically consistent. Then, for generic (a, b), all irreducible components of $\mathbb{V}_{\mathbb{C}}(F_{a,b}) \cap \{x_i = 0\} \cong \mathbb{V}_{\mathbb{C}}((F_{|x_i=0})_{a,b})$ have dimension $n - 1 - (s + \ell)$ by part (ii). But by (ii) again, all components of $\mathbb{V}_{\mathbb{C}}(F_{a,b})$ have dimension $n - (s + \ell)$. Hence, none of them is fully contained in $\{x_i = 0\}$ for generic (a, b). □

In the freely parametrized setting, we recover [LSEDSV21, Proposition 2.1].

Corollary 3.20. *For $\mathcal{S}_1, \dots, \mathcal{S}_s \subseteq \mathbb{Z}^n_{\geq 0}$ with $0 \in \mathcal{S}_i$ for $i = 1, \dots, s$, the following are equivalent:*

(i) $\mathbb{V}_{\mathbb{C}}(F)$ has pure dimension $n - s$ for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$.

(ii) $\mathbb{V}_{\mathbb{C}^*}(F)$ has pure dimension $n - s$ for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$.

(iii) The tuple $(\mathcal{S}_1, \dots, \mathcal{S}_s)$ admits an independent transversal.

Proof. Note that $\mathcal{S}_i$ is a generating set of $\text{Lin}(\mathcal{S}_i)$ as $0 \in \mathcal{S}_i$. Hence (ii) and (iii) are equivalent by Corollary 3.14.

Let H be the vertical system associated with $\mathcal{S}_1, \dots, \mathcal{S}_s$. To show that (i) implies (ii), assume (dimG) holds for H over $\mathbb{C}$. If m is the sum of the cardinalities of $\mathcal{S}_1, \dots, \mathcal{S}_s$, taking Zariski closures in affine space, we have

$$\overline{\mathcal{Z}((\mathbb{C}^*)^n)} = \overline{\pi(\mathcal{I})} = \overline{\pi(\overline{\mathcal{I}})} = \overline{\pi(\mathcal{I}_{\mathbb{C}})} = \overline{\mathcal{Z}(\mathbb{C}^n)} = \mathbb{C}^m,$$

where in the third equality we use that $\mathcal{I}_{\mathbb{C}}$ is irreducible by Theorem 3.19(i) and that $\mathcal{I} \neq \varnothing$, and in the last equality, we use the implication (dimG)$\Rightarrow$(setZ) over $\mathbb{C}$ of Theorem 3.19(ii). So, by Theorem 3.7, $\mathbb{V}_{\mathbb{C}^*}(F)$ has pure dimension $n - s$ for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$ and hence (ii) holds.

Conversely, if (ii) holds, then (deg1) holds for H over $\mathbb{C}^*$ by Corollary 3.14. Hence, (deg1) holds for H over $\mathbb{C}$ as well, and the implication (deg1)⇒(dimG) over $\mathbb{C}$ of Theorem 3.19(ii) gives that $\mathbb{V}_{\mathbb{C}}(F)$ has pure dimension $n-s$ for generic $F \in \mathcal{F}_{\text{free}}(\mathcal{S}_1, \dots, \mathcal{S}_s)$, that is, (i) holds. □

Example 3.21. We modify the generically consistent system from Example 3.18 to

$$\begin{pmatrix} a_1x_1x_3 + a_2x_2x_3 + a_3x_3 + a_4 \\ a_5x_1^2 + a_2x_2x_3 + a_3x_3 + a_6 \\ a_7x_1^3 + a_2x_2x_3 + a_3x_3 + a_8 \end{pmatrix},$$

where each polynomial now contains an independent constant term (a_4, a_6 and a_8, respectively). By Theorem 3.19(i), the affine incidence variety in $\mathbb{C}^8 \times \mathbb{C}^3$ is irreducible of dimension 8. The system still falls in the generically consistent scenario, and it therefore follows from Theorem 3.19(ii) that the zero locus over $\mathbb{C}$ generically has the expected dimension 0.

An interesting problem for future research is to explore conditions other than those in Theorem 3.19 that ensure that the zero locus of an augmented vertical system generically has no component contained in the coordinate hyperplanes, similar to what has been done in the square freely parametrized setting (see, e.g., [LW96, RW96]).

Appendix A. Nondegeneracy and radicality

In this appendix we prove Proposition 2.16. The proof is based on a series of lemmas, the first of which is a well-known commutative algebra fact adapted to the Laurent polynomial setting.

Lemma A.1.

(i) If all zeros of $F \in \mathbb{C}[x]^r$ in $\mathbb{C}^n$ are nondegenerate, then $\langle F \rangle \subseteq \mathbb{C}[x]$ is radical.

(ii) If all zeros of $F \in \mathbb{C}[x^{\pm}]^r$ in $(\mathbb{C}^*)^n$ are nondegenerate, then $\langle F \rangle \subseteq \mathbb{C}[x^{\pm}]$ is radical.

Proof. Part (i) follows by [Eis95, Corollary 16.20] (which, in turn, is a consequence of the principal ideal theorem [Eis95, Theorem 10.2] and the Jacobian criterion [Eis95, Theorem 16.19]), combined with the fact that being reduced is a local property of rings.

For (ii), first of all note that multiplying the polynomials $F = (f_1, \dots, f_r)$ by monomials neither changes the ideal $\langle F \rangle \subseteq \mathbb{C}[x^{\pm}]$ nor the nondegeneracy of their zero locus over $\mathbb{C}^*$. Hence, without loss of generality, we can assume $F \in \mathbb{C}[x]^r$. In order to show that $\langle F \rangle \subseteq \mathbb{C}[x^{\pm}]$ is radical, it is now enough to show that

$$I := \langle f_1, \dots, f_r, x_1y_1 - 1, \dots, x_ny_n - 1 \rangle \subseteq \mathbb{C}[x_1, \dots, x_n, y_1, \dots, y_n] =: \mathbb{C}[x, y]$$

is radical. The Jacobian matrix of the generators defining I has the block form

$$J := \begin{bmatrix} J_F(x) & 0 \\ \operatorname{diag}(y) & \operatorname{diag}(x) \end{bmatrix} \in \mathbb{C}^{(r+n)\times 2n},$$

and has rank $r+n$ for all $(x, y) \in \mathbb{V}_{\mathbb{C}}(I)$, as for any such point, $x \in (\mathbb{C}^*)^n$ is a zero of F, and hence nondegenerate. Hence, it follows by (i) that I is radical. □

The following lemma allows us to switch back and forth between radicality in the Laurent polynomial ring and the usual polynomial ring.

Lemma A.2. *Let K be a field and let $F \in K[x^{\pm}]^r$. Then $\langle F \rangle \subseteq K[x^{\pm}]$ is radical if and only if $\langle F \rangle \cap K[x] \subseteq K[x]$ is radical.*

Proof. The "only if" direction holds as contractions of radical ideals are radical. For the "if" direction, assume $\langle F \rangle \cap K[x]$ is radical and that $f^u \in \langle F \rangle$ for some $f \in K[x^{\pm}]$ and some integer $u > 0$. Let $N > 0$ be an integer such that $(x_1 \cdots x_n)^{Nu} f^u \in \langle F \rangle \cap K[x]$. Then, by assumption, $(x_1 \cdots x_n)^N f \in \langle F \rangle \cap K[x]$, from which it follows that $f \in \langle F \rangle$. □

For a parametric system $F \in \mathbb{C}[p, x^{\pm}]^r$ with $p = (p_1, \dots, p_k)$, we next relate radicality for generic parameter values p to radicality over the field $\mathbb{C}(p)$ of rational functions in the parameters.

Lemma A.3. *Let $F = (f_1, \dots, f_r) \in \mathbb{C}[p, x^{\pm}]^r$. If $\langle F_p \rangle \subseteq \mathbb{C}[x^{\pm}]$ is radical for generic $p \in \mathbb{C}^k$, then $\langle F \rangle \subseteq \mathbb{C}(p)[x^{\pm}]$ is a radical ideal.*

Proof. The proof will rely on Gröbner bases, so we will reformulate the lemma from a statement about Laurent polynomial rings to a statement about usual polynomial rings. We will use a subscript R in the ideal notation $\langle \cdot \rangle_R$ to indicate that the ideal is generated in the ring R whenever this is not clear from the context. Since monomials in x are units in $\mathbb{C}[p, x^{\pm}]$, we can without loss of generality assume that $F \in \mathbb{C}[p, x]^r$. Then contraction corresponds to saturation:

$$\langle F \rangle_{\mathbb{C}(p)[x^{\pm}]} \cap \mathbb{C}(p)[x] = \langle F \rangle_{\mathbb{C}(p)[x]} : (x_1 \cdots x_n)^{\infty},$$
$$\langle F_p \rangle_{\mathbb{C}[x^{\pm}]} \cap \mathbb{C}[x] = \langle F \rangle_{\mathbb{C}[x]} : (x_1 \cdots x_n)^{\infty} \text{ for } p \in \mathbb{C}^k.$$

By Lemma A.2, the lemma at hand says that if $I_p := \langle F_p \rangle_{\mathbb{C}[x]} : (x_1 \cdots x_n)^{\infty}$ is radical for generic $p \in \mathbb{C}^k$, then $I := \langle F \rangle_{\mathbb{C}(p)[x]} : (x_1 \cdots x_n)^{\infty}$ is radical.

We can construct a generating set of I that generically specializes to a generating set of I_p in the following way. Let $\widetilde{G}$ be a Gröbner basis of $\langle f_1, \dots, f_r, 1 - x_1 \cdots x_n\, y \rangle$ in the ring $\mathbb{C}(p)[x_1, \dots, x_n, y]$ with respect to the lexicographic ordering $y > x_1 > \cdots > x_n$. Then $G := \widetilde{G} \cap \mathbb{C}(p)[x]$ is a Gröbner basis for I by [CLO15, Theorem 4.4.14]. Now $\widetilde{G}$ specializes to a Gröbner basis $\widetilde{G}_p$ of $\langle f_{1,p}, \dots, f_{r,p}, 1 - x_1 \cdots x_n\, y \rangle$ for generic p by [CLO15, Theorem 6.3.1], so that $\widetilde{G}_p \cap \mathbb{C}[x]$ is a Gröbner basis for I_p. Also, $\widetilde{G}_p \cap \mathbb{C}[x] = G_p$ for generic p. Hence, there exists a nonempty Zariski open set $U \subseteq \mathbb{C}^k$ such that G specializes to a Gröbner basis G_p of I_p for all $p \in U$.

We now prove the contraposition of the desired result. If I is not radical, then there exists $h \in \mathbb{C}(p)[x]$ and $N \geq 0$ such that the normal form of h^N with respect to G is zero, while the normal form of h is nonzero. Let Z be the proper Zariski closed subset of $\mathbb{C}^k$ where the denominators of the quotients and remainders of the division of h^N and h by G vanish, and where the normal form of h vanishes. Then, for all p in the nonempty Zariski open subset $U \setminus Z$, the normal form of h_p by G_p is nonzero and that of h_p^N is zero. Hence, $h_p \notin I_p$ but $h_p \in \operatorname{rad}(I_p)$. This gives a contradiction, showing the statement. □

Proof of Proposition 2.16. If all zeros of F_p are nondegenerate for generic $p \in \mathbb{C}^k$, then, by Lemma A.1, we have that $\langle F_p \rangle \subseteq \mathbb{C}[x^{\pm}]$ is radical for generic $p \in \mathbb{C}^k$. From this, Lemma A.2, gives that $\langle F_p \rangle \cap \mathbb{C}[x] \subseteq \mathbb{C}[x]$ also is generically radical in $\mathbb{C}[x]$, and by Lemma A.3, the ideal $\langle F \rangle \subseteq \mathbb{C}(p)[x^{\pm}]$ is radical. Using again Lemma A.2 with the field $\mathbb{C}(p)$, we obtain that $\langle F \rangle \cap \mathbb{C}(p)[x] \subseteq \mathbb{C}(p)[x]$ is radical. □

References

[BCR98] J. Bochnak, M. Coste, and M. Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, 1998.

[Ber75] D. N. Bernshtein. The number of roots of a system of equations. *Funct. Anal. Appl.*, 9:183–185, 1975.

[BMMT22] P. Breiding, M. Michałek, L. Monin, and S. Telen. The algebraic degree of coupled oscillators, 2022. Preprint: `arXiv:2208.08179v1`.

[BS24] M. Bender and P. J. Spaenlehauer. Dimension results for extremal-generic polynomial systems over complete toric varieties. *J. Algebra*, 646:0021–8693, 2024.

[CFMW17] C. Conradi, E. Feliu, M. Mincheva, and C. Wiuf. Identifying parameter regions for multistationarity. *PLOS Comput. Biol.*, 13(10):e1005751, 2017.

[CLO15] D. A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer International Publishing, 2015.

[DHJ+18] T. Duff, C. Hill, A. Jensen, K. Lee, A. Leykin, and J. Sommars. Solving polynomial systems via homotopy continuation and monodromy. *IMA J. Numer. Anal.*, 39(3):1421–1446, 2018.

[Dic16] A. Dickenstein. Biochemical reaction networks: an invitation for algebraic geometers. In *Mathematical Congress of the Americas*, volume 656 of *Contemp. Math.*, pages 65–83. Amer. Math. Soc., Providence, RI, 2016.

[EH16] D. Eisenbud and J. Harris. *3264 and All That: A Second Course in Algebraic Geometry*. Cambridge University Press, 2016.

[Eis95] D. Eisenbud. *Commutative Algebra: With a View Toward Algebraic Geometry*. Graduate Texts in Mathematics. Springer, 1995.

[Fei19] M. Feinberg. *Foundations of chemical reaction network theory*, volume 202 of *Applied Mathematical Sciences*. Springer, Cham, 2019.

[FHPE24] E. Feliu, O. Henriksson, and B. Pascual-Escudero. The generic geometry of steady state varieties, 2024. Preprint: `arXiv:2412.17798`.

[GPGH+25] L. D. García Puente, E. Gross, H. A. Harrington, M. Johnston, N. Meshkat, M. Pérez Millán, and A. Shiu. Absolute concentration robustness: Algebra and geometry. *J. Symb. Comput.*, 128:102398, 2025.

[HR22] P. A. Helminck and Y. Ren. Generic root counts and flatness in tropical geometry, 2022. Preprint: `arXiv:2206.07838v2`.

[HS14] J. Huh and B. Sturmfels. Likelihood geometry. In *Combinatorial algebraic geometry*, volume 2108 of *Lecture Notes in Math.*, pages 63–117. Springer, Cham, 2014.

[Huy05] D. Huybrechts. *Complex Geometry: An Introduction*. Universitext. Springer-Verlag, Berlin, 2005.

[HVMM11] E. A. Hernandez-Vargas, D. Mehta, and R. H. Middleton. Towards modeling HIV long term behavior. *IFAC Proceedings Volumes*, 44(1):581–586, 2011. 18th IFAC World Congress.

[Kho16] A. G. Khovanskii. Newton polyhedra and irreducible components of complete intersections. *Izv. Math.*, 80(1):263–284, 2016.

[LMR23] J. Lindberg, L. Monin, and K. Rose. The algebraic degree of sparse polynomial optimization, 2023. Preprint: `arXiv:2308.07765v2`.

[LSEDSV21] G. Labahn, M. Safey El Din, É. Schost, and T. Vu. Homotopy techniques for solving sparse column support determinantal polynomial systems. *J. Complex.*, 66:101557, 2021.

[LW96] T. Li and X. Wang. The BKK root count in $\mathbb{C}^n$. *Math. Comput.*, 65(216):1477–1484, 1996.

[MR12] S. Müller and G. Regensburger. Generalized mass action systems: Complex balancing equilibria and sign vectors of the stoichiometric and kinetic-order subspaces. *SIAM J. Appl. Math.*, 72(6):1926–1947, 2012.

[Mum76] D. Mumford. *Algebraic Geometry I, Complex Projective Varieties*. Classics in Mathematics. Springer Verlag, 1976.

[OW24] N. K. Obatake and E. Walker. Newton-Okounkov bodies of chemical reaction systems. *Adv. in Appl. Math.*, 155:Paper No. 102672, 27, 2024.

[PEF22] B. Pascual-Escudero and E. Feliu. Local and global robustness at steady state. *Math. Methods Appl. Sci.*, 45(1):359–382, 2022.

[Per69] H. Perfect. A generalization of Rado's theorem on independent transversals. *Proc. Cambridge Philos. Soc.*, 66:513–515, 1969.

[PS93] P. Pedersen and B. Sturmfels. Product formulas for resultants and Chow forms. *Math. Z.*, 214(3):377–396, 1993.

[RW96] J. M. Rojas and X. Wang. Counting affine roots of polynomial systems via pointed Newton polytopes. *J. Complexity*, 12(2):116–133, 1996.

[Sch13] R. Schneider. *Convex bodies: the Brunn–Minkowski theory*, volume 151. Cambridge University Press, 2013.

[SEDS17] M. Safey El Din and É. Schost. A nearly optimal algorithm for deciding connectivity queries in smooth and bounded real algebraic sets. *J. ACM*, 63(6), 2017.

[Sta24] The Stacks Project. Available at `https://stacks.math.columbia.edu`, 2024.

[Stu94] B. Sturmfels. On the Newton polytope of the resultant. *J. Algebraic Combin.*, 3(2):207–236, 1994.

[Yu16] J. Yu. Do most polynomials generate a prime ideal? *J. Algebra*, 459:468–474, 2016.

Authors' addresses:

Elisenda Feliu, University of Copenhagen `efeliu@math.ku.dk`
Oskar Henriksson, University of Copenhagen `oskar.henriksson@math.ku.dk`
Beatriz Pascual-Escudero, Universidad Politécnica de Madrid `beatriz.pascual@upm.es`

C

The generic geometry of steady state varieties

Elisenda Feliu
Department of Mathematical Sciences
University of Copenhagen

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Beatriz Pascual-Escudero
Department of Mathematics and Informatics
Universidad Politécnica de Madrid

Publication details

Preprint: https://doi.org/10.48550/arXiv.2412.17798 (2024)

THE GENERIC GEOMETRY OF STEADY STATE VARIETIES

ELISENDA FELIU, OSKAR HENRIKSSON, AND BEATRIZ PASCUAL-ESCUDERO

Abstract. We answer several fundamental geometric questions about reaction networks with power-law kinetics, on topics such as generic finiteness of steady states, robustness, and nondegenerate multistationarity. In particular, we give an ideal-theoretic characterization of generic absolute concentration robustness, as well as conditions under which a network that admits multiple steady states also has the capacity for nondegenerate multistationarity. The key tools underlying our results come from the theory of vertically parametrized systems, and include a linear algebra condition that characterizes when the steady state system has positive nondegenerate zeros.

1. Introduction

A fundamental object of interest in the study of chemical reaction networks is the set of *positive steady states*, which under the assumption of mass-action kinetics (or more generally power-law kinetics) are the zeros of a polynomial system. More precisely, for a network with n species with concentrations $x = (x_1, \dots, x_n)$, participating in m reactions with reaction rate constants $\kappa = (\kappa_1, \dots, \kappa_m)$, the positive steady states are the positive zeros of the system

$$f_\kappa(x) := N(\kappa \circ x^M),$$

where $N \in \mathbb{R}^{n\times m}$ is the stoichiometric matrix, $M \in \mathbb{Z}^{n\times m}$ is the kinetic matrix, x^M is the vector of monomials with exponent vectors given by the columns of M, and $\circ$ denotes componentwise multiplication.

The ordinary differential equations that model a mass-action system often have linear first integrals that are independent of the choice of reaction rate constants. These correspond to conservation laws of the form $Lx = b$ where the rows of $L \in \mathbb{R}^{d\times n}$ are chosen as a basis for the left kernel of N and $b = (b_1, \dots, b_d)$ are total amounts. The steady states compatible with a choice of total amounts are given as the zeros of the polynomial system

$$F_{\kappa,b}(x) := \begin{pmatrix} N(\kappa \circ x^M) \\ Lx - b \end{pmatrix}.$$

Many questions about the steady states of a reaction network are related to understanding the geometry of the positive zero sets $\mathbb{V}_{>0}(f_\kappa) \subseteq \mathbb{R}^n_{>0}$ and $\mathbb{V}_{>0}(F_{\kappa,b}) \subseteq \mathbb{R}^n_{>0}$ for varying values of the parameters $\kappa \in \mathbb{R}^m_{>0}$ and $b \in \mathbb{R}^d$. Even though the study of reaction networks in the current mathematical formalism goes back at least to Feinberg, Horn and Jackson in the 1970's [Fei72, HJ72], many fundamental properties about the underlying polynomial systems are still not fully understood, including properties such as dimension, finiteness, and singularities. Understanding the relations among these concepts is often necessary to lift results about small networks to larger networks where they appear as submotives (see, e.g., [BP18, CF06, JS13]), and a prerequisite for using machinery from algebraic geometry to study the steady states (see, e.g., [PM11, Section 6.5], [PEF22]).

Understanding the geometry of $\mathbb{V}_{>0}(f_\kappa)$ and $\mathbb{V}_{>0}(F_{\kappa,b})$ for all possible networks and parameter values is a very challenging task. Indeed, it follows from the classical *Hungarian lemma* [HT79] that the positive part of *any* algebraic variety can appear as $\mathbb{V}_{>0}(f_\kappa)$ for some network and some choice of $\kappa \in \mathbb{R}^m_{>0}$, which means that the set of steady states can have very complicated geometry. However, one of the main messages of this paper is that the problem becomes much more well-behaved if we change it to instead understand $\mathbb{V}_{>0}(f_\kappa)$ and $\mathbb{V}_{>0}(F_{\kappa,b})$ *up to perturbations* of the parameter values (i.e., in *open regions* of parameter space). In particular, we show that for many properties, the behavior up to perturbation agrees qualitatively with the behavior of the complex varieties $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ and $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$ in $(\mathbb{C}^*)^n$ for generic $(\kappa, b) \in \mathbb{C}^{m+d}$, and hence lends itself to be studied with tools from algebraic geometry.

We first tackle the question of *finiteness*. By a simple equation count, it is reasonable to expect that $\mathbb{V}_{>0}(f_\kappa)$ has codimension $\mathrm{rk}(N)$, and that $\mathbb{V}_{>0}(F_{\kappa,b})$ should be finite. Nevertheless, it is easy to construct examples of networks where these sets have higher-than-expected dimension for some parameter values. In fact, this can be done even for very well-behaved families of networks, such as those that are endotactic [KD24] or weakly reversible [BCY20] (see Example 3.12). However, it has been an open question whether this type of pathology can arise for open regions of parameter space (see [BCY20, Section 5]).

As our first main result, we answer this question in the negative. In fact, we show that there are two possible scenarios for $\mathbb{V}_{>0}(F_{\kappa,b})$, depending on whether or not the network is *nondegenerate*, in the sense that $\ker(N) \cap \mathbb{R}^m_{>0} \neq \varnothing$ and it satisfies the rank condition

$$\mathrm{rk} \begin{bmatrix} N \operatorname{diag}(w) M^\top \operatorname{diag}(h) \\ L \end{bmatrix} = n \quad \text{for some } (w,h) \in \ker(N) \times (\mathbb{R}^*)^n. \tag{1.1}$$

None of the scenarios allow infinitely many steady states in an open region of parameter space.

Theorem A (Theorem 3.4)**.** *Suppose that we have a network with* $\ker(N) \cap \mathbb{R}^m_{>0} \neq \varnothing$*. If the network is* ***nondegenerate****, then:*

- *The set* $\mathcal{Z}_{\mathrm{cc}}$ *of* $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$ *for which* $\mathbb{V}_{>0}(F_{\kappa,b})$ *is nonempty has nonempty interior.*
- $\mathbb{V}_{>0}(F_{\kappa,b})$ *is finite for generic* $(\kappa, b) \in \mathcal{Z}_{\mathrm{cc}}$*.*
- *The Jacobian* $J_{F_{\kappa,b}}(x)$ *is nonsingular on* $\mathbb{V}_{>0}(F_{\kappa,b})$ *for generic* $(\kappa, b) \in \mathcal{Z}_{\mathrm{cc}}$*.*

If the network is ***degenerate****, then*

- *The set* $\mathcal{Z}_{\mathrm{cc}}$ *of* $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$ *for which* $\mathbb{V}_{>0}(F_{\kappa,b})$ *is nonempty but has empty interior.*
- *The Jacobian* $J_{F_{\kappa,b}}(x)$ *is singular on all of* $\mathbb{V}_{>0}(F_{\kappa,b})$ *for all* $(\kappa, b) \in \mathcal{Z}_{\mathrm{cc}}$*.*

Analogous statements relating emptiness and dimension are given for $\mathbb{V}_{>0}(f_\kappa)$ in Theorem 3.1. Both of these results are an application of the theory of *vertically parametrized systems* developed in [FHPE24]. Condition (1.1) is easy to check computationally for specific networks, which we demonstrate in Section 3.2, where we check it for all networks in the database ODEbase [LSR22]. In Section 3.3, we also prove that it is satisfied by large families of networks, including weakly reversible networks, injective networks, conservative networks lacking boundary steady states, and the networks of the deficiency one theorem.

Next, we treat the problem of ***absolute concentration robustness*** (ACR). The term ACR was introduced in [SF10] and has been extensively studied, e.g., [ASBL99, CGK20, GPGH$^+$25, KPMD$^+$12, MST22, PEF22]. We say that a network has ACR in X_i for a given $\kappa \in \mathbb{R}^m_{>0}$ if $\mathbb{V}_{>0}(f_\kappa)$ is nonempty and contained in a parallel translate of the i-th coordinate plane. The weaker notion of ***local ACR*** was introduced in [PEF22] and means that $\mathbb{V}_{>0}(f_\kappa)$ is contained in a finite union of translates of a coordinate hyperplane.

Understanding when ACR or local ACR arise for all $\kappa \in \mathbb{R}^m_{>0}$ is a very challenging algebraic-geometric problem, as has recently been explored in detail in [GPGH$^+$25, PEF22], but the *generic* counterpart of these problems, where we only require that a property holds for almost all $\kappa \in \mathbb{R}^m_{>0}$, turns out to be much more well-behaved. Using the theory of vertically parametrized systems, we prove that the rank condition from [PEF22, Section 5] precisely characterizes generic local ACR. In particular, it is a necessary condition for ACR to hold in a Euclidean open subset. We also strengthen the sufficient ideal-theoretic condition from [GPGH$^+$25, Proposition 3.8] to a complete characterization of generic ACR.

Theorem B (Theorem 4.4, Corollary 4.7)**.** *For a nondegenerate network, the following are equivalent:*

(i) The network has generic local ACR for X_i*.*

(ii) $\mathrm{rk}(N \operatorname{diag}(w) M_{\setminus i}^\top) < \mathrm{rk}(N)$ *for all* $w \in \ker(N)$*, where* $M_{\setminus i}$ *is* M *without the* i*-th row.*

(iii) There exists a nonconstant polynomial $g \in (\langle f_\kappa \rangle : (x_1 \cdots x_n)^\infty) \cap \mathbb{R}(\kappa_1, \ldots, \kappa_m)[x_i]$*.*

In particular:

- *If g has a single positive root for generic $\kappa \in \mathbb{R}^m_{>0}$, the network has generic ACR for X_i.*
- *If the network does not have generic local ACR, then, for generic $\kappa \in \mathbb{R}^m_{>0}$, the network does not have (local) ACR for X_i.*

Finally, we also study the property of ***multistationarity***, which refers to $\mathbb{V}_{>0}(F_{\kappa,b})$ having at least two elements for some $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$. When, in addition, the steady states are stable, this property is believed to underlie cellular decision processes, in that it offers robust switch-like behavior as a response to changes in the parameters via a phenomenon known as *hysteresis* [LK99]. The study of multistationarity can be found at the roots of reaction network theory, with celebrated results on complex balancing [Fei72, Fei87, HJ72] and numerous algorithms and criteria to decide upon its existence or lack thereof, e.g., [CF05, CFRS07, JEKF18, PMDSC12]; see [JS15] for an overview.

Several results for inferring multistationarity from reduced models require, however, a stronger condition, namely that there is a choice of parameters for which $\mathbb{V}_{>0}(F_{\kappa,b})$ has at least two steady states where the Jacobian is nonsingular, as they rely on the implicit function theorem or homotopy continuation, see, e.g., [BP18, CF06, CFW20, FW13, JS13, JTZ24]. This property is referred to as ***nondegenerate multistationarity***.

In [JS17, Conjecture 2.3], the authors conjecture that if $\mathbb{V}_{>0}(F_{\kappa,b})$ is finite for all $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$, and has cardinality p for some choice of parameters, then there is also a choice of parameters such that the network has p *nondegenerate* positive steady states. This is known as the *Nondegeneracy Conjecture*. It has been proven for small networks (with at most 2 species and 2 reactions, which can be reversible) in [JS17, SdW19] and for $\mathrm{rk}(N) = 1$ in [JTZ24], but the general case remains open. Here, we prove the conjecture in the $p = 2$ case for nondegenerate networks (see Theorem 5.2 for the full statement, which allows for milder assumptions on the network).

Theorem C (Theorem 5.2)**.** *A nondegenerate network for which $\mathbb{V}_{>0}(F_{\kappa,b})$ is finite for all $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$, and which has at least two positive steady states for some choice of parameters, admits at least two nondegenerate positive steady states.*

Organization of the paper. In Section 2, we fix the notation for the rest of the paper, and recall some basic terminology of chemical reaction network theory, as well as some results regarding the Jacobian of f_κ and $F_{\kappa,b}$. In Section 3, we discuss the connection between the properties of nondegeneracy, dimension and consistency, leading up in particular to Theorem A. We discuss also computational aspects, as well as how nondegeneracy relates to other properties of reaction networks such as weak reversibility and the dimension of the kinetic subspace. In Section 4, we address ACR, state Theorem B and give several examples to clarify the relation to previous work. After this, we devote Section 5 to nondegenerate multistationarity. Finally, we give the proofs relying on technical aspects of vertically parametrized systems in Section 6.

Acknowledgements. EF and OH have been funded by the Novo Nordisk Foundation project with grant reference number NNF20OC0065582. BP has been funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie IF grant agreement No 794627 and the Spanish Ministry of Economy project with reference number PID2022-138916NB-I00. This work has also been funded by the European Union under the Grant Agreement no. 101044561, POSALG. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or European Research Council (ERC). Neither the European Union nor ERC can be held responsible for them.

The authors thank Anne Shiu for helpful feedback on earlier versions of the manuscript.

2. Reaction networks and steady states

In this section we give a quick overview of reaction networks, introduce the varieties of positive steady states, and fix notation that will be used in the rest of the paper.

A ***reaction network*** is simply a collection of *reactions*

$$\alpha_{1j}X_1 + \cdots + \alpha_{nj}X_n \xrightarrow{\kappa_j} \beta_{1j}X_1 + \cdots + \beta_{nj}X_n, \quad j = 1, \ldots, m,$$

that model interactions among *species* $X_1, \ldots, X_n$. Each of these interactions transforms a *complex* (a $\mathbb{Z}_{\geq 0}$-linear combination of the species) called the *reactant*, into another complex called the *product*. The reactions are labelled by positive real numbers $\kappa_j > 0$ called *reaction rate constants*, which play a role in the dynamics of the system represented by the network.

This way, a reaction network can be considered as a digraph, with reactions as directed edges, labeled by the reaction rate constants, and complexes as nodes.

Example 2.1. A simple representation of enzymatic transfer of calcium ions between the endoplasmic reticulum and the cytosol gives rise to the following network with $n = 4$ species and $m = 6$ reactions:

$$0 \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} X_1 \qquad X_1 + X_2 \xrightarrow{\kappa_3} 2X_1 \qquad X_1 + X_3 \underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} X_4 \xrightarrow{\kappa_6} X_2 + X_3.$$

Here, X_1 stands for cytosolic calcium, X_2 for calcium in the endoplasmic reticulum, and X_3 is an enzyme catalyzing the transfer via the formation of an intermediate X_4 [GES05].

Under the assumption of *power-law kinetics* with ***kinetic matrix*** $M = (\nu_{ij}) \in \mathbb{R}^{n\times m}$, the evolution of the concentration $x = (x_1, \ldots, x_n)$ of the species $X_1, \ldots, X_n$ over time is given by the autonomous system

$$\frac{dx}{dt} = N(\kappa \circ x^M), \qquad x \in \mathbb{R}^n_{>0}, \tag{2.1}$$

where $N = (\beta_{ij} - \alpha_{ij}) \in \mathbb{Z}^{n\times m}$ is the ***stoichiometric matrix***, x^M denotes the vector whose j-th entry is the product $x_1^{\nu_{1j}} \cdots x_n^{\nu_{nj}}$ (i.e., the monomial in x with exponents given by the j-th column of M), and $\kappa \circ x^M$ is the entry-wise product of the two vectors $\kappa = (\kappa_1, \ldots, \kappa_m)$ and x^M.

For the specific case in which the kinetics is ***mass action*** [GW64], $M = (\alpha_{ij}) \in \mathbb{Z}^{n\times m}_{\geq 0}$ is the *reactant matrix*. In the examples, we consider mass-action kinetics unless stated otherwise.

For the rest of our work, we will assume that $M \in \mathbb{Z}^{n\times m}$, that is M has integer entries, when referring to power-law kinetics. The results extend to systems where $M \in \mathbb{Q}^{n\times m}$ using the approach in [PEF22, Section 4.2]. As we allow negative exponents, we mostly work in the Laurent polynomial ring $\mathbb{R}[x^{\pm}] = \mathbb{R}[x_1^{\pm}, \ldots, x_n^{\pm}]$ (or over $\mathbb{C}$).

Observe that (2.1) is well defined also over $\mathbb{R}^n_{\geq 0}$ if M has nonnegative entries. Under mild additional assumptions, namely that all negative monomials of the expression for dx_i/dt are multiples of x_i, both $\mathbb{R}^n_{\geq 0}$ and $\mathbb{R}^n_{>0}$ are forward invariant [Vol72]. In particular, this is the case under mass-action kinetics.

A (positive) ***steady state*** of the ODE system in (2.1) is a tuple $x = (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}$ such that $N(\kappa \circ x^M) = 0$. The steady states provide useful information about the dynamics of the biological system under study, and this is a key topic in the theory of reaction networks; see [Fei19] for an introduction to the field.

To remove redundancies arising when N does not have full rank n, we make a choice of matrix $C \in \mathbb{R}^{s\times m}$ with $s := \operatorname{rk}(N)$ and $\ker(C) = \ker(N)$, and consider the ***steady state system***

$$f := C(\kappa \circ x^M) \in \mathbb{R}[\kappa, x^{\pm}]^s. \tag{2.2}$$

The results in this work do not depend on the specific matrix C, so a choice is implicitly made throughout. We write f_κ for the system (2.2) evaluated at a fixed $\kappa \in \mathbb{R}^m_{>0}$, and whose zeros are the steady states for this κ. In this case the ***positive steady state variety*** of the system f_κ is the (semialgebraic) set

$$\mathbb{V}_{>0}(f_\kappa) := \{x \in \mathbb{R}^n_{>0} : C(\kappa \circ x^M) = 0\} = \{x \in \mathbb{R}^n_{>0} : N(\kappa \circ x^M) = 0\}.$$

We denote the set of parameter values for which $\mathbb{V}_{>0}(f_\kappa)$ is nonempty by

$$\mathcal{Z} := \{\kappa \in \mathbb{R}^m_{>0} : \mathbb{V}_{>0}(f_\kappa) \neq \varnothing\}.$$

By letting $\mathbb{R}^*$ and $\mathbb{C}^*$ denote the set of real and complex numbers excluding 0, we will also consider the *real* and *complex steady state varieties*, given by

$$\mathbb{V}_{\mathbb{R}^*}(f_\kappa) := \{x \in (\mathbb{R}^*)^n : f_\kappa(x) = 0\} \qquad \text{and} \qquad \mathbb{V}_{\mathbb{C}^*}(f_\kappa) := \{x \in (\mathbb{C}^*)^n : f_\kappa(x) = 0\}.$$

Note that $\mathbb{V}_{>0}(f_\kappa) = \mathbb{V}_{\mathbb{C}^*}(f_\kappa) \cap \mathbb{R}^n_{>0}$.

The vector subspace $\mathrm{im}(N)$ is called the ***stoichiometric subspace***. The trajectory of system (2.1) with initial condition x^0 is confined to the linear subspace $x^0 + \mathrm{im}(N)$. We will therefore also be interested in the positive steady states constrained to these linear subspaces. Specifically, we consider a fixed matrix $L \in \mathbb{R}^{d\times n}$ with full rank $d := n - s$ (recall $s = \mathrm{rk}(N) = \mathrm{rk}(C)$), whose rows form a basis of the left kernel of N. Then, we consider the (open) ***stoichiometric compatibility classes***

$$\mathcal{P}_b := \{x \in \mathbb{R}^n_{>0} : Lx - b = 0\}, \qquad b \in \mathbb{R}^d.$$

Such a matrix L is called a ***matrix of conservation laws***.

Many questions about the dynamics of (2.1), and in particular about the steady states, are studied in the restriction to the stoichiometric compatibility classes. Two steady states in the same class are said to be ***stoichiometrically compatible***. Studying the set of all positive steady states in the class $\mathcal{P}_b$ can be done via the square polynomial system

$$F = \begin{pmatrix} C(\kappa \circ x^M) \\ Lx - b \end{pmatrix} \in \mathbb{R}[\kappa, b, x^{\pm}]^n, \tag{2.3}$$

which we refer to as the ***augmented steady state system*** (by L). We write $F_{\kappa,b}$ for the system (2.3) when κ and b have been fixed. The set of positive zeros of $F_{\kappa,b}$ is by construction the intersection of $\mathbb{V}_{>0}(f_\kappa)$ and $\mathcal{P}_b$. Analogous to $\mathcal{Z}$, we consider

$$\mathcal{Z}_{\mathrm{cc}} := \{(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d : \mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b \neq \varnothing\}.$$

Note here that $\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b \neq \varnothing$ requires that $b \in L(\mathbb{R}^n_{>0})$, that is, b belongs to the positive cone on the columns of L. We remark that as L has full rank, $L(\mathbb{R}^n_{>0})$ is a d-dimensional cone.

Again, the choice of L is implicit throughout, as it does not affect the conclusions of this work. The sets $\mathbb{V}_{>0}(F_{\kappa,b})$, $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$ and $\mathbb{V}_{\mathbb{R}^*}(F_{\kappa,b})$ are defined analogously to f.

Example 2.2. For the network in Example 2.1, the ODE system (2.1) is built out of the matrices

$$N = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}, \qquad M = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

The rank of N is $s = 3$, as the bottom two rows are linearly dependent. Hence we can choose

$$C = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{bmatrix} \qquad \text{and} \qquad L = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}.$$

With this in place, the steady state system becomes

$$f = (\kappa_1 - \kappa_2 x_1 + \kappa_3 x_1 x_2 - \kappa_4 x_1 x_3 + \kappa_5 x_4, -\kappa_3 x_1 x_2 + \kappa_6 x_4, -\kappa_4 x_1 x_3 + \kappa_5 x_4 + \kappa_6 x_4),$$

while the augmented steady state system has the additional entry $x_3 + x_4 - b$.

In general, for a polynomial system $g = (g_1, \ldots, g_\ell)$ in the variables $x = (x_1, \ldots, x_n)$, a zero x^* of g is said to be ***degenerate*** if the Jacobian matrix $J_g(x) = \left(\frac{\partial g_i}{\partial x_j}\right)_{i,j}$ of g does not have full rank when evaluated at x^*.

A steady state $x^* \in \mathbb{V}_{>0}(f_\kappa)$ is said to be ***degenerate*** if it is degenerate as a zero of F_{κ,Lx^*}. This is a weaker property than being a degenerate zero of f_κ, as it requires $J_{f_\kappa}(x^*)$ to be singular on $\mathrm{im}(N)$ (and hence sometimes one refers to degeneracy *with respect to* $\mathrm{im}(N)$).

The key to be able to make conclusions about the steady state varieties is the bijection

$$\begin{aligned} \phi\colon \ker(C) \times (\mathbb{C}^*)^n &\to \{(\kappa, b, x) \in \mathbb{C}^m \times \mathbb{C}^d \times (\mathbb{C}^*)^n : F(\kappa, b, x) = 0\} \\ (w, h) &\mapsto (w \circ h^M, Lh^{-1}, h^{-1}), \end{aligned} \tag{2.4}$$

where h^{-1} is taken componentwise, and the image is the (complex) ***incidence variety*** of the system F. Importantly, ϕ also restricts to a bijection onto the positive incidence variety:

$$(\ker(C) \cap \mathbb{R}^m_{>0}) \times \mathbb{R}^n_{>0} \to \{(\kappa, b, x) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d \times \mathbb{R}^n_{>0} : F(\kappa, b, x) = 0\}.$$

This bijection allows to derive the following well-known proposition, which is handy for studying the set of all Jacobian matrices of f and F for all parameter values and zeros. In preparation for that, we define the matrices

$$\begin{aligned} Q_f(w) &:= C\,\mathrm{diag}(w)M^\top \in \mathbb{C}^{s\times n}, \quad w \in \mathbb{C}^m, \\ Q_F(w,h) &:= \begin{bmatrix} C\,\mathrm{diag}(w)M^\top\,\mathrm{diag}(h) \\ L \end{bmatrix} \in \mathbb{C}^{n\times n}, \quad w \in \mathbb{C}^m,\ h \in (\mathbb{C}^*)^n. \end{aligned} \tag{2.5}$$

Proposition 2.3. *Consider a reaction network with stoichiometric matrix $N \in \mathbb{Z}^{n\times m}$, kinetic matrix $M \in \mathbb{Z}^{n\times m}$, steady state system f as in (2.2) for $C \in \mathbb{R}^{s\times n}$ of full rank $s = \mathrm{rk}(N)$ such that $\ker(N) = \ker(C)$, and augmented steady state system F as in (2.3) for a full rank matrix $L \in \mathbb{R}^{d\times n}$ with $LN = 0$ and $d = n - s$. Then the following holds:*

(i) $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing \iff \mathcal{Z} \neq \varnothing \iff \mathcal{Z}_{\mathrm{cc}} \neq \varnothing$.

(ii) The set $\mathcal{Z}$ is connected.

(iii) For $(\kappa, b, x) = \phi(w, h)$ it holds that

$$J_{f_\kappa}(x) = Q_f(w)\,\mathrm{diag}(h) \quad \textit{and} \quad J_{F_{\kappa,b}}(x) = Q_F(w,h).$$

Proof. (i) For $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing \Rightarrow \mathcal{Z} \neq \varnothing$ apply ϕ to some $(w, h) \in (\ker(C) \cap \mathbb{R}^m_{>0}) \times \mathbb{R}^n_{>0}$. The rest of the implications are immediate from definition. (ii) $\mathcal{Z}$ is the image of the connected set $(\ker(C) \cap \mathbb{R}^m_{>0}) \times \mathbb{R}^n_{>0}$ by the projection of the continuous map ϕ in (2.4) onto $\mathbb{R}^m_{>0}$. (iii) Follows from a direct computation of the Jacobian matrices, see [FHPE24, Proposition 3.2]. □

It is common in the literature to say that a network is ***consistent*** (or dynamically nontrivial) if the equivalent statements in Proposition 2.3(i) hold. Proposition 2.3(iii) is telling us that the set of Jacobian matrices $J_{f_\kappa}(x)$ for $\kappa \in \mathbb{C}^m$ and $x \in \mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ and the set of matrices $Q_f(w)\,\mathrm{diag}(h)$ for $w \in \ker(C)$ and $h \in (\mathbb{C}^*)^n$ agree. Furthermore, the equality of sets restricts when values of κ and x are chosen positive. An analogous equality of matrix sets is derived for F.

We conclude this section by reminding the reader of a series of well-known definitions from the theory of reaction networks that will be used later on:

- The ***linkage classes*** of a reaction network are its connected components as a digraph. Its ***strong linkage classes*** are the maximal strongly connected subdigraphs. Among the strong linkage classes, one distinguishes the ***terminal strong linkage classes***, as those for which there is no edge from a node inside of the class to a node outside of it, in the original network.
- A network is ***weakly reversible*** if all connected components of the underlying digraph are strongly connected.
- The ***deficiency*** of a given network is the nonnegative integer $\delta = c - \ell - s$, where c is the number of complexes and ℓ is the number of linkage classes.
- A network is called ***conservative*** if $\ker(N^\top) \cap \mathbb{R}^n_{>0} \neq \varnothing$, that is, the row span of L contains a positive vector (equivalently the Euclidean closure of $\mathcal{P}_b$ is a compact set for all b [BI64]).

- If $M \in \mathbb{Z}_{\geq 0}^{n\times m}$ (so that zeros of f, F can be considered in $\mathbb{R}^n_{\geq 0}$), a network is said to ***lack relevant boundary steady states*** if there does not exist $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$ and $x \in \mathbb{R}^n_{\geq 0} \setminus \mathbb{R}^n_{>0}$ such that $F_{\kappa,b}(x) = 0$ and $\mathcal{P}_b \cap \mathbb{R}^n_{>0} \neq \varnothing$.
- A network is called ***injective*** if f_κ is injective as a map $\mathcal{P}_b \to \mathbb{R}^s$ for all $(\kappa, b) \in \mathbb{R}^m_{>0} \times L(\mathbb{R}^n_{>0})$ (see [CF05, FW12, MFR+15]).

3. Nondegeneracy, consistency, dimension and finiteness

This section is devoted to the key theorems about generic nonemptiness and dimension of the positive steady state variety and its intersection with the stoichiometric compatibility classes. The results are a consequence of the general theorems proven in [FHPE24], which are rewritten here for this more restricted context and adapted to the reaction network language. In particular, in [FHPE24] the dimension and consistency of a type of parametric systems called (augmented) vertically parametrized systems is studied. The steady state and the augmented steady state systems are both of this type. A version of the main result there can be found in Section 6. The main take home message is that the existence of a nondegenerate zero is enough to ensure that the varieties behave "nicely".

We say that a property holds for ***generic*** parameters in a set $\mathcal{A} \subseteq \mathbb{R}^\ell$ when this property holds in a nonempty Zariski open subset of $\mathcal{A}$. When $\mathcal{A}$ is $\mathbb{R}^m_{>0}$ or $\mathbb{R}^m_{>0} \times \mathbb{R}^d$, this implies that the property holds in a set with nonempty Euclidean interior, and outside a subset of $\mathcal{A}$ of Lebesgue measure zero, and hence is robust against small perturbations of the parameter values.

Given a reaction network with kinetic matrix M, the steady state system f has s linearly independent entries and n variables, while the augmented steady state system has n entries and variables. One could therefore expect that $\dim(\mathbb{V}_{>0}(f_\kappa)) = n - s$ and $\dim(\mathbb{V}_{>0}(F_{\kappa,b})) = 0$ as semialgebraic sets. This is certainly the case in typical examples, but it is not hard to construct examples where the expectation does not hold true.

3.1. Generic consistency and dimension.

The following theorems give us precise tools to describe the geometry of $\mathbb{V}_{>0}(f_\kappa)$ and $\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b$ for generic choices in $\mathcal{Z}$ and $\mathcal{Z}_{\text{cc}}$.

Theorem 3.1 (Expected dimension of steady state varieties)**.** *Consider a reaction network with stoichiometric matrix $N \in \mathbb{Z}^{n\times m}$ and kinetic matrix $M \in \mathbb{Z}^{n\times m}$, and consider the steady state system f as in (2.2) for $C \in \mathbb{R}^{s\times n}$ of full rank $s = \operatorname{rk}(N)$ such that $\ker(N) = \ker(C)$. Assume $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. The following statements are equivalent:*

(i) $\operatorname{rk}(C \operatorname{diag}(w) M^\top) = s$ *for some* $w \in \ker(C)$.

(ii) f_κ *has a nondegenerate zero in* $(\mathbb{C}^*)^n$ *for some* $\kappa \in \mathbb{C}^m$.

(iii) For generic $\kappa \in \mathcal{Z}$*, all zeros of* f_κ *in* $(\mathbb{C}^*)^n$ *are nondegenerate.*

(iv) $\mathcal{Z}$ *has nonempty Euclidean interior in* $\mathbb{R}^m_{>0}$.

(v) $\mathcal{Z}$ *is not contained in a hypersurface of* $\mathbb{R}^m$.

(vi) $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ *has pure dimension* $n - s$ *for at least one* $\kappa \in \mathbb{C}^m$.

Furthermore, the following holds:

- *If the equivalent statements hold, then* $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ *and* $\mathbb{V}_{\mathbb{R}^*}(f_\kappa)$ *have pure dimension* $n - s$ *for generic* $\kappa \in \mathcal{Z}$*, and the same is true for* $\mathbb{V}_{>0}(f_\kappa)$ *as a semialgebraic set. Additionally, all zeros of* f_κ *in* $\mathbb{R}^n_{>0}$ *are nondegenerate for generic* $\kappa \in \mathcal{Z}$.
- *If the equivalent statements do not hold, then* $\mathbb{V}_{\mathbb{C}^*}(f_\kappa) = \varnothing$ *for generic* $\kappa \in \mathbb{R}^m_{>0}$*, and when not empty,* $\dim \mathbb{V}_{\mathbb{C}^*}(f_\kappa) > n - s$ *and all zeros of* f_κ *are degenerate.*
- *The ideal generated by* f_κ *in* $\mathbb{C}[x^{\pm}]$ *is radical for generic* $\kappa \in \mathbb{C}^m$.

The proof of this result can be found in Section 6. A steady state system with $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$ and satisfying any of the equivalent conditions (i)-(vi) in Theorem 3.1 is said to be ***nondegenerate***. The term hints at the existence of nondegenerate zeros as well as the fact that the dimension of the varieties of f are generically as expected.

Example 3.2. With the matrices given in Example 2.2, and for $w = (1,1,1,2,1,1) \in \ker(C) \cap \mathbb{R}^m_{>0}$, it holds that

$$\operatorname{rk}(C \operatorname{diag}(w) M^\top) = 3.$$

Hence condition (i) in Theorem 3.1 is satisfied, and we conclude that f is nondegenerate. Furthermore, Theorem 3.1 tells us that the network has positive steady states for κ in a subset of $\mathbb{R}^6_{>0}$ with nonempty Euclidean interior (Theorem 3.1(iv)), and that for generic κ in this set, $\mathbb{V}_{>0}(f_\kappa), \mathbb{V}_{\mathbb{C}^*}(f_\kappa), \mathbb{V}_{\mathbb{R}^*}(f_\kappa)$ have dimension 1 and all zeros of f_κ are nondegenerate.

Example 3.3. For the following reaction network with mass-action kinetics

$$3X_1 \xrightarrow{\kappa_1} 2X_1 \qquad 2X_1 + X_2 \xrightarrow{\kappa_2} 2X_1 \qquad X_1 + X_3 \xrightarrow{\kappa_3} X_1$$
$$2X_2 \xrightarrow{\kappa_4} 2X_2 + X_3 \qquad X_2 + X_3 \xrightarrow{\kappa_5} X_1 + X_2 + X_3 \qquad 2X_3 \xrightarrow{\kappa_6} X_2 + 2X_3,$$

we have $s = 3$, and the defining matrices are

$$C = N = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad M = \begin{bmatrix} 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 \end{bmatrix}.$$

Any $w \in \ker(C)$ is of the form $w = (u_1, u_2, u_3, u_3, u_1, u_2)$ for some $u_1, u_2, u_3 \in \mathbb{C}$. But then

$$\det(C \operatorname{diag}(w) M^\top) = \det \begin{bmatrix} -3u_1 & u_1 & u_1 \\ -2u_2 & -u_2 & 2u_2 \\ -u_3 & 2u_3 & -u_3 \end{bmatrix} = 0.$$

Hence condition (i) in Theorem 3.1 does not hold. We conclude that f is degenerate and none of the statements (i)-(vi) hold. In particular, any positive steady state is degenerate, and the positive steady state variety is generically empty.

Theorem 3.4 (Finiteness of the number of stoichiometrically compatible steady states)**.** *Consider a reaction network with stoichiometric matrix $N \in \mathbb{Z}^{n\times m}$ and kinetic matrix $M \in \mathbb{Z}^{n\times m}$. Consider the augmented steady state system F as in (2.3) for $C \in \mathbb{R}^{s\times n}$ of full rank $s = \operatorname{rk}(N)$ such that $\ker(N) = \ker(C)$ and $L \in \mathbb{R}^{d\times n}$ with $LN = 0$ and $d = n - s$. Assume $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. The following statements are equivalent:*

(i) For some $w \in \ker(C)$ and $h \in (\mathbb{C}^*)^n$ it holds that

$$\operatorname{rk} \begin{bmatrix} C \operatorname{diag}(w) M^\top \operatorname{diag}(h) \\ L \end{bmatrix} = n.$$

(ii) $F_{\kappa,b}$ has a nondegenerate zero in $(\mathbb{C}^*)^n$ for some $(\kappa, b) \in \mathbb{C}^m \times \mathbb{C}^d$.

(iii) For generic $(\kappa, b) \in \mathcal{Z}_{\mathrm{cc}}$, all zeros of $F_{\kappa,b}$ in $(\mathbb{C}^*)^n$ are nondegenerate.

(iv) $\mathcal{Z}_{\mathrm{cc}}$ has nonempty Euclidean interior in $\mathbb{R}^m_{>0} \times \mathbb{R}^d$.

(v) $\mathcal{Z}_{\mathrm{cc}}$ is not contained in a hypersurface of $\mathbb{R}^m \times \mathbb{R}^d$.

(vi) $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$ is nonempty and finite for at least one $(\kappa, b) \in \mathbb{C}^m \times \mathbb{C}^d$.

Furthermore:

- *If the equivalent statements are satisfied, then for generic $(\kappa, b) \in \mathcal{Z}_{\mathrm{cc}}$, it holds that*

$$0 < \#(\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b) < \infty$$

  and all points of $\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b$ are nondegenerate steady states.
- *The ideal generated by $F_{\kappa,b}$ in $\mathbb{C}[x^{\pm}]$ is radical for generic $(\kappa, b) \in \mathbb{C}^m \times \mathbb{C}^d$.*

The proof is analogous to that of Theorem 3.1, and both can be found in Section 6, where additional properties of f and F are given. We say that the augmented steady state system F (or the network) is ***nondegenerate*** if $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$ and F satisfies any of the equivalent conditions (i)-(vi) in Theorem 3.4.

Theorem 3.4 tells us that $\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b$ is either generically empty (Theorem 3.4(v) does not hold), or generically finite. We obtain the following consequence.

Corollary 3.5. *The set* $\{(\kappa, b) \in \mathbb{R}^m_{>0} \times L(\mathbb{R}^n_{>0}) : \#(\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b) = \infty\}$ *is contained in a proper algebraic variety, and hence always has empty Euclidean interior in* $\mathbb{R}^m_{>0} \times L(\mathbb{R}^n_{>0})$.

It is straightforward to see that condition (i) in Theorem 3.4 implies condition (i) in Theorem 3.1. This gives rise to the following corollary. However, the converse is not necessarily true, as shown by Example 3.7.

Corollary 3.6. *If the augmented steady state system is nondegenerate, then so is the steady state system.*

Example 3.7. Consider the reaction network

$$X_1 + X_2 \xrightarrow{\kappa_1} X_1 \qquad X_2 \xrightarrow{\kappa_2} 2X_2.$$

The steady states are described by the single parametric polynomial

$$f = -\kappa_1 x_1 x_2 + \kappa_2 x_2 = x_2(-\kappa_1 x_1 + \kappa_2).$$

We obtain that $\mathbb{V}_{>0}(f_\kappa) = \{(x_1, x_2) \in \mathbb{R}^2_{>0} : x_1 = \frac{\kappa_2}{\kappa_1}\}$ for all $\kappa \in \mathbb{R}^2_{>0}$, hence $\mathcal{Z} = \mathbb{R}^2_{>0}$ and hence f is nondegenerate (as condition (iv) in Theorem 3.1 holds). As the stoichiometric compatibility classes are defined by the equation $x_1 = b$ for $b > 0$, $\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b \neq \varnothing$ only if $\frac{\kappa_2}{\kappa_1} = b$, and hence $\mathcal{Z}_{\text{cc}}$ has empty Euclidean interior. We conclude that F is degenerate as condition (iv) in Theorem 3.4 does not hold.

Theorem 3.1 and Theorem 3.4 concern the *generic* behavior of the zero sets, but they do not preclude pathological behaviors from arising for specific choices of parameters. Examples of what these behaviors can be are given next.

Example 3.8. The reaction network

$$2X_1 + X_2 \xrightarrow{\kappa_1} 3X_1 \qquad X_1 + 2X_2 \xrightarrow{\kappa_2} 2X_1 + X_2 \qquad X_1 + X_2 \xrightarrow{\kappa_3} 2X_2$$

gives rise to the system $f = x_1 x_2(\kappa_1 x_1 + \kappa_2 x_2 - \kappa_3)$ and the sets $\mathcal{P}_b$ are defined by the equation $x_1 + x_2 = b$, giving that F is generically finite, hence nondegenerate. Nevertheless, for $\kappa = (1, 1, 1)$, $\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b = \varnothing$ if $b \neq 1$, whereas $\#(\mathbb{V}_{>0}(f_\kappa) \cap \mathcal{P}_b) = \infty$ for $b = 1$. This only happens nongenerically, namely for parameters of the form $((\kappa_1, \kappa_1, \kappa_1 b), b)$ with $\kappa_1, b > 0$.

Example 3.9. For the reaction network

$$3X_1 + X_2 \xrightarrow{\kappa_1} 4X_1 \qquad 2X_1 + X_2 \xrightarrow{\kappa_2} 3X_2 \qquad X_1 + X_2 \xrightarrow{\kappa_3} 2X_1,$$

we have $n = 2$, $s = 1$, and the steady state system is $f = x_1 x_2(\kappa_1 x_1^2 - 2\kappa_2 x_1 + \kappa_3)$, which is generically 1-dimensional. For parameters satisfying $\kappa_2^2 = \kappa_1 \kappa_3$, all zeros of f_κ are degenerate but $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ and $\mathbb{V}_{\mathbb{R}^*}(f_\kappa)$ still have pure dimension 1. Hence, degeneracy of all steady states for a choice of parameters does not necessarily imply that the dimension of the steady state variety is higher than expected for that parameter value.

Example 3.10. For the network

$$3X_1 + X_2 \xrightarrow{\kappa_1} 4X_1 + 2X_2 \qquad X_1 + X_2 \xrightarrow{\kappa_2} 2X_1 + X_2 \xrightarrow{\kappa_3} X_1$$
$$X_1 + 3X_2 \xrightarrow{\kappa_4} 2X_1 + 4X_2 \qquad X_1 + X_2 \xrightarrow{\kappa_5} X_1 + 2X_2 \xrightarrow{\kappa_6} X_2,$$

with stoichiometric and reactant matrices

$$C = \begin{bmatrix} 1 & 1 & -1 & 1 & 0 & -1 \\ 1 & 0 & -1 & 1 & 1 & -1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 3 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 3 & 1 & 2 \end{bmatrix},$$

the steady state system f is degenerate. For $\kappa = (1, 2, 2, 1, 2, 2)$ we have

$$f_\kappa = \begin{pmatrix} x_1 x_2(x_1^2 + 2 - 2x_1 + x_2^2 - 2x_2) \\ x_1 x_2(x_1^2 - 2x_1 + x_2^2 + 2 - 2x_2) \end{pmatrix} = \begin{pmatrix} x_1 x_2((x_1 - 1)^2 + (x_2 - 1)^2) \\ x_1 x_2((x_1 - 1)^2 + (x_2 - 1)^2) \end{pmatrix},$$

which has finite $V_{>0}(f_\kappa)$. Hence, statement (vi) in Theorem 3.1 and Theorem 3.4 cannot be replaced by the existence of a choice of parameters for which the set of positive zeros is finite.

Remark 3.11 (Boundary steady states)**.** The results in this section concern the steady states with nonzero coordinates. By [FHPE24, Theorem 3.19], the results on generic dimension extend from $(\mathbb{C}^*)^n$ to $\mathbb{C}^n$ if all the polynomials in f_κ have a constant term involving a different parameter. This scenario arises when the network includes all *inflow reactions*, that is, reactions of the form $0 \to X_i$ for all species X_i (in particular $s = n$). This is the case for *Continuous-flow stirred-tank reactors (CFSTR)*, see [Fei19, Section 4.2.1]. For these networks, the existence of a nondegenerate zero of f_κ ensures that the set of nonnegative steady states in $\mathbb{R}^n_{\geq 0}$ has generically dimension 0.

3.2. **Computationally deciding on nondegeneracy.** Checking condition (i) in Theorems 3.1 and 3.4 can be done computationally as follows. Let $G \in \mathbb{R}^{m\times(m-s)}$ be a Gale dual matrix to C, in the sense that $\ker(C) = \operatorname{im}(G)$. Recall the matrices $Q_f(w)$ and $Q_F(w, h)$ from (2.5).

Consistency: Nonemptiness of $\ker(C) \cap \mathbb{R}^m_{>0}$ is equivalent to the feasibility of the system $\{Cx = 0, x \geq \mathbb{1}\}$, which can be checked with linear programming.

Nondegeneracy of f: To check condition (i) of Theorem 3.1, we pick a random $u \in \mathbb{R}^{m-s}$ (for some appropriate distribution), and compute $\operatorname{rk}(Q_f(Gu))$ with exact arithmetic. If the rank is s, we conclude that f is nondegenerate. If not, we view u as indeterminate, compute the s-minors of $Q_f(Gu)$ in $\mathbb{R}[u_1, \dots, u_{m-s}]$, and use that f is degenerate if and only if all minors are zero.

Nondegeneracy of F: Similarly, condition (i) in Theorem 3.4 can be checked by first computing $\operatorname{rk}(Q_F(Gu, h))$ for random $(u, h) \in \mathbb{R}^{m-s} \times (\mathbb{R}^*)^n$ with exact arithmetic. If the rank is n, it follows that F is nondegenerate; if not, we compute $\det(Q_F(Gu, h))$ as a polynomial in $\mathbb{R}[u_1, \dots, u_{m-s}, h_1^{\pm}, \dots, h_n^{\pm}]$, and use that F is degenerate if and only if the determinant is zero.

A Julia implementation of these computations, based on the computer algebra package `Oscar.jl` [OSC24] and the reaction network theory package `Catalyst.jl` [LMI+23], can be found in the GitHub repository

`https://github.com/oskarhenriksson/generic-geometry-of-steady-state-varieties`.

As a demonstration of the applicability of our implementation, we apply it to the networks in the database ODEbase [LSR22], modeling all of them with mass-action kinetics for simplicity. Out of 628 networks, we found that precisely 368 are consistent. Among these, 6 networks have degenerate steady state system f. For the other 362 networks, both f and F are nondegenerate. The results of these computations are available in the aforementioned GitHub repository.

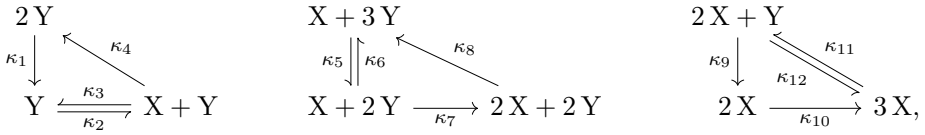
The largest network in the database that admits a nondegenerate positive steady state is `BIOMD0000000014`, with $n = 86$, $m = 300$ and $d = n - s = 9$, for which applying the nondegeneracy checks takes less than 2 seconds in Julia*.

3.3. **Network-theoretic conditions that ensure nondegeneracy.** In this section we go through a number of properties that play a central role in the reaction network theory literature, and prove that they imply nondegeneracy of the network, and hence the equivalent properties in Theorem 3.1 and Theorem 3.4.

Weakly reversible networks are quite well understood; for example, they are known to admit positive steady states for all choices of $(\kappa, b) \in \mathbb{R}^m_{>0} \times L(\mathbb{R}^n_{>0})$ [Bor19], and are conjectured to display strong dynamical behaviors such as persistence [Fei87] or bounded trajectories [And11]. It was also earlier believed that the number of steady states in each stoichiometric compatibility class was finite for this type of networks, a fact that was disproven in [BCY20] with the following example (see also [KD24] for further examples).

*All computations were run on a MacBook Air with an Apple M2 chip and 16 GB of RAM.

Example 3.12. In [BCY20], the authors considered the network



with $s = n = 2$, and fine tuned the reaction rate constants in such a way that the two equations defining the steady states had a common factor and the other factors did not admit positive zeros. They obtained $\dim \mathbb{V}_{\mathbb{C}^*}(f_\kappa) = \dim \mathbb{V}_{>0}(f_\kappa) = 1$, and hence infinitely many steady states in the only stoichiometric compatibility class $\mathbb{R}^n_{>0}$. With this trick, they illustrated that even for weakly reversible reaction networks, $\dim \mathbb{V}_{>0}(f_\kappa)$ could be larger than expected for some choice of parameter values.

Motivated by the example above, the authors of [BCY20] posed the following question:

> *Is it possible for a weakly reversible network to have infinitely many positive steady states [in each positive stoichiometric compatibility class] for each choice of reaction rate constants in a [Euclidean] open set of the parameter space $\mathbb{R}^m_{>0}$?*

We answered the question in the negative in Corollary 3.5 for *any* network. So the condition on weak reversibility is superfluous in answering the question. However, for weakly reversible networks, we give the following strengthened answer.

Corollary 3.13. *Suppose that a reaction network satisfies any of the following conditions:*

(i) It is weakly reversible and has mass-action kinetics.

(ii) $M \in \mathbb{Z}^{n\times m}_{\geq 0}$, it is conservative, and lacks relevant boundary steady states.

Then $\mathcal{Z}_{\text{cc}} = \mathbb{R}^m_{>0} \times L(\mathbb{R}^n_{>0})$ and hence the network is nondegenerate.

Proof. The fact that $\mathcal{Z}_{\text{cc}} = \mathbb{R}^m_{>0} \times L(\mathbb{R}^n_{>0})$ follows for case (i), from [Bor19], and for case (ii), by a standard Brouwer fixed point argument, using that the Euclidean closure of $\mathcal{P}_b$ is compact and there are no steady states at the boundary (see e.g. [FH24, Proposition 6.5]). The statement now follows from condition (iv) in Theorem 3.4. □

Two other important classes of networks for which our results have implications are those that are injective and those that satisfy the conditions of Feinberg's Deficiency One Theorem [Fei95, Theorem 4.2].

The *Deficiency One Theorem* requires that, when all linkage classes of a given network are considered separately, their deficiencies add up to exactly the deficiency of the network, none of them being higher than 1. Additionally, each linkage class should have no more than one terminal strong linkage class. These properties do not guarantee the existence of a positive steady state, so we need to require the network to be consistent.

Corollary 3.14. *Suppose that a reaction network satisfies* $\ker(N) \cap \mathbb{R}^m_{>0} \neq \varnothing$ *and any of the following conditions:*

(i) It is injective.

(ii) It fulfils the criteria of the Deficiency One Theorem.

Then $\mathcal{Z}_{\text{cc}}$ has nonempty Euclidean interior and hence the network is nondegenerate.

Proof. By Proposition 2.3(i), $\mathcal{Z}_{\text{cc}} \neq \varnothing$. Any steady state is guaranteed to be nondegenerate in case (i) by [CF10, Sec. 6], see also [FW12, Corollary 5.12], and in case (ii) by [Fei95, Theorem 4.3]. The result now follows since condition (i) in Theorem 3.4 holds. □

3.4. **The kinetic and stoichiometric subspace.** We close the section with a discussion on how the dimension of the *kinetic subspace* is related to nondegeneracy.

For a fixed choice of reaction rate constants $\kappa \in \mathbb{R}^m_{>0}$, let Σ_κ be the coefficient matrix of $N(\kappa \circ x^M)$ as a Laurent polynomial system in $\mathbb{R}[x^{\pm}]^n$. The *kinetic subspace* is

$$S_\kappa := \operatorname{im}(\Sigma_\kappa),$$

and satisfies that the trajectories of (2.1) are contained in parallel translates of S_κ. In fact, this is the minimal vector subspace with this property. It clearly holds that $S_\kappa \subseteq \operatorname{im}(N)$, but the inclusion might be strict. The stoichiometric subspace $S = \operatorname{im}(N)$ depends only on N, while the kinetic subspace also depends on the value of κ and the kinetic matrix M.

In [FH77], it is shown that if all linkage classes contain a unique terminal strong linkage class, then S and S_κ agree for all κ. Moreover, inequality of these spaces implies all steady states are degenerate; this is immediate as if $S_\kappa \subsetneq S$, then the entries of f_κ are linearly dependent (see also [Fei19, Section 3.A.1]). The converse is not true, as it is possible to find degenerate systems f with $S_\kappa = S$ for all κ (e.g., Example 3.3).

Corollary 3.15. *If $S_\kappa \subsetneq S$ for generic $\kappa \in \mathcal{Z}$, then f and F are both degenerate. In particular, the network is degenerate, $\mathcal{Z}$ has empty Euclidean interior, and*

$$\#(V_{\mathbb{C}^*}(f_\kappa) \cap \mathcal{P}_b) = \infty \qquad \text{for all } \ (\kappa, b) \in \mathcal{Z}.$$

Proof. If $S_\kappa \subsetneq S$ for generic $\kappa \in \mathcal{Z}$, then all zeros of f_κ are degenerate generically for $\kappa \in \mathcal{Z}$. Thus condition (iii) in Theorem 3.1 cannot hold, and f, and hence F, are degenerate. The rest of the conclusions follow from Theorem 3.4. □

Example 3.16. Consider the following network from [Fei19, Example 3.A.2]

$$X_2 \xleftarrow{\kappa_1} X_1 \xrightarrow{\kappa_2} X_3 \qquad\qquad X_2 + X_3 \xrightarrow{\kappa_3} 2X_1$$

with mass-action kinetics. This network has defining matrices

$$N = \begin{bmatrix} -1 & -1 & 2 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad \Sigma_\kappa = \begin{bmatrix} -\kappa_1 - \kappa_2 & 2\kappa_3 \\ \kappa_1 & -\kappa_3 \\ \kappa_2 & -\kappa_3 \end{bmatrix}$$

and $s = 2$. The first row of Σ_κ is linearly dependent to the two bottom rows, and, if the two bottom rows are linearly independent, that is, $\kappa_1 \neq \kappa_2$, then f_κ cannot have a zero in $(\mathbb{C}^*)^2$. We conclude that Σ_κ needs to have rank 1 if $\kappa \in \mathcal{Z}$ and hence $S_\kappa \subsetneq S$ for generic $\kappa \in \mathcal{Z}$. By Corollary 3.15, f is degenerate, and so is the network.

In [Fei19, Page 123] it is stated that *"when the kinetic subspace is smaller than the stoichiometric subspace, a positive equilibrium in a stoichiometric compatibility class will usually be accompanied by an infinite number of them"*. This was made precise in Corollary 3.15.

Finally, the results from [FH77] regarding the stoichiometric and kinetic subspaces allow us to relate degeneracy and deficiency δ. Let ℓ and t denote the number of linkage classes and terminal strong linkage classes, respectively.

Corollary 3.17. *For any network such that $t - \ell > \delta$, taken with mass-action kinetics, both f and F are degenerate. In particular, the network with mass-action kinetics is degenerate.*

Proof. This follows from Corollary 3.15, as our assumption implies $S_\kappa \subsetneq S$ for all $\kappa \in \mathbb{R}^m_{>0}$ by statement (ii) in the theorem in Section 6 of [FH77]. □

4. Generic absolute concentration robustness

In this section we use our new understanding of the steady state varieties to strenghten previous results on Absolute Concentration Robustness (ACR) and to clarify its generic behavior.

A reaction network with a choice of power-law kinetics is said to have ACR for a certain species X_i if it is consistent and, for each $\kappa \in \mathbb{R}^m_{>0}$, the concentration of X_i attains a unique value for all positive steady states in $\mathbb{V}_{>0}(f_\kappa)$ [SF10]. The weaker notion of *local* ACR was introduced in [PEF22], and refers to the concentration of a species X_i attaining only a *finite* number of possible values for all positive steady states in $\mathbb{V}_{>0}(f_\kappa)$ for each $\kappa \in \mathbb{R}^m_{>0}$.

Example 4.1. Consider the following network with mass-action kinetics from [SF10] describing the phosphorylation and dephosphorylation mechanism for the enzyme isocitrate dehydrogenase:

$$X_1 + X_2 \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} X_3 \xrightarrow{\kappa_3} X_1 + X_4 \qquad X_3 + X_4 \underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} X_5 \xrightarrow{\kappa_6} X_2 + X_3\,.$$

For $\kappa \in \mathbb{R}^6_{>0}$, the positive steady state variety is

$$\mathbb{V}_{>0}(f_\kappa) = \left\{\left(x_1, x_2, \tfrac{\kappa_1}{\kappa_2+\kappa_3}x_1x_2, \tfrac{\kappa_3}{\kappa_4}\left(1+\tfrac{\kappa_5}{\kappa_6}\right), \tfrac{\kappa_1\kappa_3}{\kappa_6(\kappa_2+\kappa_3)}x_1x_2\right) : x_1, x_2 \in \mathbb{R}_{>0}\right\} \subset \mathbb{R}^5_{>0},$$

from where it is clear that $x_4 = \frac{\kappa_3}{\kappa_4}(1+\frac{\kappa_5}{\kappa_6})$ is constant for any $x \in \mathbb{V}_{>0}(f_\kappa)$. The network has ACR for X_4, but not for X_1, as the steady state value of x_1 does not only depend on κ.

Example 4.2. For the network in Example 3.9, the value of x_1 at steady state is a positive root of the polynomial $\kappa_1 x_1^2 - 2\kappa_2 x_1 + \kappa_3$, and hence attains one or two values whenever $\mathbb{V}_{>0}(f_\kappa) \neq \varnothing$. Hence the network displays local ACR for X_1.

Many works have explored ACR, both from the experimental and the theoretical points of view, see for example [ASBL99, CGK20, GPGH$^+$25, KPMD$^+$12, MST22, PEF22, SF10]. However, no general procedure allows to detect it in practice, outside of specific contexts or under restrictive hypotheses. Previous work in the literature has attempted to understand ACR at the level of the ideal generated by f_κ, often for fixed values of κ. One of the first sources of this is [PM11], and a more recent account of this approach can be found in [GPGH$^+$25]. However, understanding the positive part of an algebraic variety for *all* parameter values is a challenging problem as pathologies can easily arise for specific values of κ (such as the variety having real dimension lower than expected). Several of these pathologies are discussed in [GPGH$^+$25].

As illustrated by Theorem 3.1 and Theorem 3.4, the behavior of the positive variety for *generic* parameter values is nonpathological and resembles that of the complex variety. This leads us to introduce the concepts of *generic ACR* and *generic local ACR*, and we show that these properties can be characterized in a more satisfactory way.

In what follows, we let $\pi_i \colon \mathbb{C}^n \to \mathbb{C}$ denote the projection onto the x_i-coordinate.

Definition 4.3. For a reaction network assume that the steady state system f is nondegenerate. For a fixed $\kappa \in \mathcal{Z}$, we say that $\mathbb{V}_{>0}(f_\kappa)$ has

- ***local ACR*** for X_i if $\#\pi_i(\mathbb{V}_{>0}(f_\kappa)) < \infty$,
- ***ACR*** for X_i if $\#\pi_i(\mathbb{V}_{>0}(f_\kappa)) = 1$.

We furthermore say that the network has

- ***generic local ACR*** for X_i if $\#\pi_i(\mathbb{V}_{>0}(f_\kappa)) < \infty$ for generic $\kappa \in \mathcal{Z}$,
- ***generic ACR*** for X_i if $\#\pi_i(\mathbb{V}_{>0}(f_\kappa)) = 1$ for generic $\kappa \in \mathcal{Z}$.

If the properties hold for all $\kappa \in \mathcal{Z}$, then we say that the network has local ACR or ACR for X_i respectively.

The following theorem settles that a linear algebra condition arising from [PEF22, Theorem 5.3], the study of elimination ideals over the coefficient field $\mathbb{C}(\kappa)$, or the study of the complex counterpart of $\mathbb{V}_{>0}(f_\kappa)$, all completely characterize generic (local) ACR.

Theorem 4.4. *Consider a reaction network with stoichiometric matrix $N \in \mathbb{Z}^{n\times m}$ and kinetic matrix $M \in \mathbb{Z}^{n\times m}$. Let $f = (f_1, \dots, f_s)$ be the steady state system (2.2), and let $C \in \mathbb{R}^{s\times m}$ be of full rank s with $\ker(N) = \ker(C)$. Suppose f is nondegenerate. The following are equivalent:*

(i) The network has generic local ACR for X_i.

(ii) $\pi_i(\mathbb{V}_{\mathbb{C}^*}(f_\kappa))$ is a finite set for generic $\kappa \in \mathbb{C}^m$.

(iii) $\operatorname{rk}(C \operatorname{diag}(w) M_{\setminus i}^\top) \leq s - 1$ for all $w \in \ker(C)$, where $M_{\setminus i}$ is M without the i-th row.

(iv) For $I := \langle f_1, \dots, f_s\rangle \subseteq \mathbb{C}(\kappa)[x^\pm]$, the elimination ideal $I \cap \mathbb{C}(\kappa)[x_i^\pm]$ is generated by one nonconstant polynomial.

Furthermore:

- *If the network does not have generic local ACR for X_i, then, for generic $\kappa \in \mathcal{Z}$, $\mathbb{V}_{>0}(f_\kappa)$ has no (local) ACR for X_i.*
- *If the network has generic local ACR for X_i and for a specific $\kappa \in \mathcal{Z}$ every irreducible component of $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ that intersects $\mathbb{R}^n_{>0}$ contains a nondegenerate zero of f_κ, then $\mathbb{V}_{>0}(f_\kappa)$ has local ACR for X_i. In particular, the network has local ACR for X_i if $\operatorname{rk}(C \operatorname{diag}(v) M^\top) = s$ for* all *$v \in \ker(C) \cap \mathbb{R}^m_{>0}$.*

The proof of Theorem 4.4 is given in Section 6.2, as it relies on a general result on augmented vertically parametrized systems. We emphasize first, that condition (iii) of Theorem 4.4 can easily be rejected by taking one random choice of parameter value, and second, that conditions (ii) and (iv) of Theorem 4.4 refer to the complex variety $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$; hence generic local ACR cannot occur if the equivalent property does not arise over $\mathbb{C}^*$.

The following corollary is a direct consequence of Theorem 4.4.

Corollary 4.5. *Consider a reaction network for which the steady state system f is nondegenerate. If the set of $\kappa \in \mathbb{R}^m_{>0}$ for which $\mathbb{V}_{>0}(f_\kappa)$ has local ACR for X_i has nonempty Euclidean interior, then the network has generic local ACR for X_i, and hence any of the equivalent statements (ii)-(iv) in Theorem 4.4 holds.*

Remark 4.6. Suppose $M \in \mathbb{Z}^{n\times m}_{\geq 0}$. In this case, we can replace $I = \langle f_1, \dots, f_s\rangle \subseteq \mathbb{C}(\kappa)[x^\pm]$ in Theorem 4.4 with the saturation ideal

$$\langle f_1, \dots, f_s\rangle_{\mathbb{C}(\kappa)[x]} : (x_1 \cdots x_n)^\infty \subseteq \mathbb{C}(\kappa)[x]$$

as this ideal equals $I \cap \mathbb{C}(\kappa)[x]$ and monomials are units in $\mathbb{C}(\kappa)[x^\pm]$.

In [GPGH+25, Proposition 3.8], the ideal $\langle f_{\kappa,1}, \dots, f_{\kappa,s}\rangle \cap \mathbb{C}[x_i] \subseteq \mathbb{C}[x_i]$ is studied for fixed values of $\kappa \in \mathbb{R}^m_{>0}$, as a sufficient condition for $\mathbb{V}_{>0}(f_\kappa)$ to have ACR. It is also given as a sufficient condition that the elimination ideal of the saturated ideal is generated by a polynomial of the form $x_i - \alpha$ for some α. We settle here that the study of the generator of the elimination ideal after saturating $\langle f_{\kappa,1}, \dots, f_{\kappa,s}\rangle$ completely characterizes whether (local) ACR arises generically.

Corollary 4.7. *Consider a reaction network such that the steady state system f is nondegenerate. The following statements are equivalent:*

(i) The network has generic ACR for X_i.

(ii) For $I := \langle f_1, \dots, f_s\rangle \subseteq \mathbb{C}(\kappa)[x^\pm]$, the elimination ideal $I \cap \mathbb{C}(\kappa)[x_i^\pm]$ is generated by one polynomial that has exactly one positive root for generic $\kappa \in \mathcal{Z}$.

Proof. Immediate from Theorem 4.4. □

Example 4.8. In Example 4.2, it holds that

$$\{C \operatorname{diag}(w) M^\top : w \in \ker(C)\} = \{[-2w_1 + 2w_2 \quad 0] : w_1, w_2 \in \mathbb{C}\}.$$

We see that condition (i) in Theorem 3.1 holds, so f is nondegenerate. Removal of the first column gives a matrix of rank 0. Hence, condition (ii) in Theorem 4.4 is satisfied, and we conclude that the network has generic local ACR for X_1.

Remark 4.9. Clearly, having generic ACR is necessary for the network to have ACR (for nondegenerate steady state systems). Theorem 4.4 gives also that lack of generic local ACR implies that generically, there is no ACR. We illustrate here that other relations between these concepts might not hold.

We exploit an easy source of examples for ACR, which arise from networks with full rank $s = n$. When the number of steady states is finite, uniqueness of steady states readily gives ACR. Although these are not the interesting cases in applications, they allow us to understand what phenomena are not to be expected. In particular:

(1) *Generic (local) ACR does not imply (local) ACR*: The steady state system

$$f = (-\kappa_1 x_1 x_2 + \kappa_2 x_2^2 + \kappa_3 x_2, \kappa_1 x_1 x_2 - \kappa_4 x_2^2 - \kappa_5 x_2),$$

which can be seen to arise from a network with mass-action kinetics, has positive zeros in a set with nonempty Euclidean interior $\mathcal{Z}$, and for generic $\kappa \in \mathcal{Z}$, there is only one such zero, namely $(\frac{\kappa_2\kappa_5 - \kappa_3\kappa_4}{\kappa_1(\kappa_2 - \kappa_4)}, \frac{\kappa_5 - \kappa_3}{\kappa_2 - \kappa_4})$. Thus, the network has generic ACR for X_1 and X_2. However, when $\kappa_2 = \kappa_4$ and $\kappa_3 = \kappa_5$, the zero set consists of the points $(\frac{\kappa_2 x_2 + \kappa_3}{\kappa_1}, x_2)$ for $x_2 > 0$, and hence there is no ACR (nor local ACR).

(2) *Absence of generic ACR does not imply that, generically, there is no ACR*: The network with one species

$$3X_1 \xrightarrow{\kappa_1} 2X_1 \qquad 2X_1 \xrightarrow{\kappa_2} 3X_1 \qquad X_1 \xrightarrow{\kappa_3} 0 \qquad 0 \xrightarrow{\kappa_4} X_1$$

has steady state system $f = -\kappa_1 x_1^3 + \kappa_2 x_1^2 - \kappa_3 x_1 + \kappa_4$, which is an arbitrary degree three polynomial with coefficients of fixed and alternating sign. Hence, $\mathcal{Z}$ has nonempty Euclidean interior. Furthermore, in a nonempty Euclidean open subset of $\mathcal{Z}$, f has exactly one positive root, so there is ACR for X_1. However, also in a nonempty Euclidean open subset of $\mathcal{Z}$, f has three positive roots, and hence there is no ACR. As both presence and absence of ACR occur in nonempty Euclidean open sets, none of the properties arise generically for this network.

(3) *Absence of generic local ACR does not preclude local ACR for a specific κ*: The network with two species

$$\begin{array}{lll} 3X_1 + X_2 \xrightarrow{\kappa_1} 4X_1 & 2X_1 + X_2 \xrightarrow{\kappa_2} X_1 + 2X_2 & \\ X_1 + 3X_2 \xrightarrow{\kappa_3} 2X_1 + 2X_2 & X_1 + 2X_2 \xrightarrow{\kappa_4} 3X_2 & X_1 + X_2 \xrightarrow{\kappa_5} 2X_1 \end{array}$$

has $s = 1$ and the steady state system is

$$f = x_1 x_2 (\kappa_1 x_1^2 - \kappa_2 x_1 + \kappa_3 x_2^2 - \kappa_4 x_2 + \kappa_5)\,.$$

For generic κ in $\mathcal{Z}$, the zero set is a nonlinear curve, and hence, there is no generic local ACR. However, for $\kappa = (1, 2, 1, 2, 2)$, we have $f_\kappa = (x_1 - 1)^2 + (x_2 - 1)^2$, $\mathbb{V}_{>0}(f_\kappa)$ consists of one (degenerate) point, and has trivially ACR for X_1 and X_2.

The GitHub repository of this paper (cf. Section 3.2) also includes code for checking the rank condition in Theorem 4.4. When applying the condition to the networks in the database ODEbase, we found 48 consistent nondegenerate networks that have generic local ACR but not full rank (which would trivially imply generic local ACR). For this check, we excluded species that do not appear in the reactant or product of any of the reactions.

Example 4.10. Network `BIOMD0000000167` satisfies the conditions for generic local ACR in X_9. In fact, we have generic ACR, since

$$\kappa_2\kappa_4^2\kappa_6^2\kappa_7\kappa_{10}(\kappa_{11}^2 + 2\kappa_{11}\kappa_{13} + \kappa_{13}^2)x_9^2 - \kappa_1\kappa_3^2\kappa_5^2\kappa_8\kappa_9(\kappa_{12}^2 + 2\kappa_{12}\kappa_{14} + \kappa_{14}^2) \in \langle f \rangle \cap \mathbb{C}(\kappa)[x_9^{\pm}]$$

which clearly has a unique positive root for all $\kappa \in \mathbb{R}^{14}_{>0}$.

5. Nondegenerate multistationarity

A network is said to be ***multistationary*** if there exists $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$ such that

$$2 \leq \#\mathbb{V}_{>0}(F_{\kappa,b})\,.$$

We say that the network admits ***nondegenerate multistationary*** if in addition the elements in $\mathbb{V}_{>0}(F_{\kappa,b})$ are nondegenerate steady states. The following theorem, which is strengthening of [CFMW17, Theorem 1], shows that, for a special type of networks, multistationarity and nondegenerate multistationarity go hand in hand. The concept of ***dissipative*** networks, which generalizes conservative networks, refers to networks for which, given κ and b, there exists a compact set C such that the trajectories of (2.1) in $\mathcal{P}_b$ eventually remain in C; see [CFMW17] for details. Recall the matrix $Q_F(w,h)$ given in (2.5).

Theorem 5.1. *Consider a reaction network with kinetic matrix $M \in \mathbb{Z}^{n\times m}_{\geq 0}$. Assume that the network lacks relevant boundary steady states and is dissipative, and let F be the augmented steady state system. Assume that $\det(Q_F(w,h))$ attains both positive and negative signs for $(w,h) \in (\ker(N) \cap \mathbb{R}^m_{>0}) \times \mathbb{R}^n_{>0}$. Then the network admits at least three nondegenerate positive steady states for some choice of parameters.*

Proof. On the one hand, by [CFMW17, Theorem 1], there exists $\varepsilon \in \{\pm 1\}$ (depending on $\mathrm{rk}(N)$ and a choice of a specific order of the rows of $J_{F_{\kappa,b}}(x)$) such that the network is multistationary if

$$\mathrm{sign}(\det(J_{F_{\kappa^*,Lx^*}}(x^*)) = \varepsilon \tag{5.1}$$

for some $\kappa^* \in \mathbb{R}^m_{>0}$ and $x^* \in \mathbb{V}_{>0}(f_\kappa)$. More specifically, [CFMW17, Theorem 1] states that if (5.1) holds, the network has more than two steady states for this κ^* and $b = Lx^*$, and if all steady states are nondegenerate, then there is an odd number of them.

By Proposition 2.3 and hypothesis, as $\det(Q_F(w,h))$ takes both signs, there exists $(w,h) \in (\ker(N) \cap \mathbb{R}^m_{>0}) \times \mathbb{R}^n_{>0}$ such that the sign of $\det(Q_F(w,h))$ is ε. Then $(\kappa^*, b^*, x^*) := \phi(w,h)$ (with ϕ as in (2.4)) is such that x^* is a nondegenerate steady state for the reaction rate constant κ^* and there is multistationarity for b^*.

On one hand, the implicit function theorem provides a Euclidean ball $B \subseteq \mathbb{R}^m_{>0} \times \mathbb{R}^d$ containing (κ^*, b^*), such that for all $(\kappa, b) \in B$, $F_{\kappa,b}$ has a nondegenerate zero x and the sign of $\det(J_{F_{\kappa,b}}(x))$ is preserved, that is, is ε.

On the other hand, Theorem 3.4 and the existence of a nondegenerate steady state tells us that all steady states are nondegenerate for all (κ, b) in a nonempty Zariski open subset $\mathcal{U} \subseteq \mathcal{Z}_{\mathrm{cc}}$. We conclude that for any parameter point (κ, b) in the intersection $\mathcal{U} \cap B$, which is nonempty, all steady states are nondegenerate and at least one steady state satisfies (5.1). Hence there is an odd number of them and at least two, giving the statement. □

Next, we give conditions that ensure that a network with the capacity for multistationarity also has the capacity for nondegenerate multistationarity. This can be seen as a version of the Nondegeneracy Conjecture from [JS17, JTZ24, SdW19]. We note that, trivially, degenerate networks cannot exhibit nondegenerate multistationarity.

Theorem 5.2. *Suppose that a network has at least two isolated positive steady states for a given parameter value $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^d$, and that furthermore one of the following conditions holds:*

(i) The network is nondegenerate and $\mathbb{V}_{>0}(F_{\kappa',b'})$ is finite for all (κ', b') in an open neighborhood of (κ, b) in $\mathbb{R}^m_{>0} \times \mathbb{R}^d$.

(ii) At least one of the positive steady states for (κ, b) is an isolated point in $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$.

Then there also exists a choice of parameters such that the network has at least two nondegenerate positive steady states.

Proof. Consider the parametrization of $\mathcal{Z}_{\mathrm{cc}}$, obtained by composing the restriction of ϕ in (2.4) to the positive real orthant with the projection map:

$$\varphi\colon (\ker(N)\cap\mathbb{R}^m_{>0})\times\mathbb{R}^n_{>0}\to\mathbb{R}^m\times\mathbb{R}^d,\quad (v,h)\mapsto(v\circ h^M, Lh^{-1}),$$

where $d=n-s$. By the definition of ϕ, the preimage of (κ,b) by φ is in one-to-one correspondence with the positive zeros of $F_{\kappa,b}$. An easy dimension count shows that both the domain and codomain of φ are differential manifolds of dimension $m+d$. By definition, $\mathrm{im}(\varphi)=\mathcal{Z}_{\mathrm{cc}}$.

Note that the network is nondegenerate, so in particular, $\mathcal{Z}_{\mathrm{cc}}$ is a full-dimensional semialgebraic set. The nondegeneracy is an assumption in case (i). In case (ii), the theorem of dimension of fibers [Mum76, Theorem 3.13, Corollary 3.15] says that if $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$ has an isolated point, then $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa',b'})$ is nonempty and finite for generic $(\kappa',b')\in\mathbb{C}^m\times\mathbb{C}^d$, so the network is nondegenerate by Theorem 3.4.

The set $\mathcal{D}_{\mathrm{cc}}$ of parameter values for which there is a degenerate steady state coincides with the set of critical values of φ, and the Zariski closure $H\subseteq\mathbb{R}^m\times\mathbb{R}^d$ of $\mathcal{D}_{\mathrm{cc}}$ is a proper algebraic variety under the assumption of nondegeneracy [FHPE24, Proposition 3.4]. So

$$T:=\varphi^{-1}(H)\subseteq(\ker(N)\cap\mathbb{R}^m_{>0})\times\mathbb{R}^n_{>0}$$

is a proper Zariski closed subset.

By assumption, there exists a point $(\kappa,b)\in\mathbb{R}^m_{>0}\times\mathbb{R}^d$, such that its fiber contains at least two isolated points $\xi_1,\xi_2\in\varphi^{-1}(\kappa,b)$. If we are in case (ii), we can in addition assume that ξ_1 corresponds to an isolated point in $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$.

In order to have something to prove, we assume that at least one of ξ_1 and ξ_2 corresponds to a degenerate steady state, which implies that $(\kappa,b)\in H$. We want to show that there exist $(\kappa',b')\in\mathbb{R}^m_{>0}\times\mathbb{R}^d$ such that

$$1<\#\varphi^{-1}(\kappa',b')<\infty\qquad\text{and}\qquad(\kappa',b')\notin H. \tag{5.2}$$

Let U_1 and U_2 be disjoint open balls in $(\ker(N)\cap\mathbb{R}^m_{>0})\times\mathbb{R}^n_{>0}$ around ξ_1 and ξ_2, respectively, such that they only contain these preimages of (κ,b). Then $(\kappa,b)\in\varphi(U_1)\cap\varphi(U_2)$. Both $\varphi(U_1)$ and $\varphi(U_2)$ have nonempty Euclidean interior, as any ball around (κ,b) contains points not in H.

If $V:=(\varphi(U_1)\cap\varphi(U_2))^\circ\neq\varnothing$, that is, the intersection has nonempty interior, then for all $(\kappa',b')\in V$, the fiber $\varphi^{-1}(\kappa',b')$ contains at least two distinct points: one in U_1 and one in U_2. Any $(\kappa',b')\in V\setminus H$ now satisfies (5.2).

If the above does not hold, that is $(\varphi(U_1)\cap\varphi(U_2))^\circ=\varnothing$, then necessarily (κ,b) is neither in $\varphi(U_1)^\circ$ nor $\varphi(U_2)^\circ$. To complete the proof, it is enough to show that φ is not injective on $U_1\setminus T$ by using that (κ,b) is a boundary point of $\varphi(U_1)$.

We first show that there exists an open neighborhood $U\subseteq U_1$ of ξ_1, such that the fibers of $\varphi_{|U}$ are finite. This is trivial in case (i). For case (ii), we use the fact that the map $\xi\mapsto\dim_\xi(\varphi_{\mathbb{C}}^{-1}(\varphi_{\mathbb{C}}(\xi)))$, where $\dim_\xi$ denotes the local dimension at ξ and $\varphi_{\mathbb{C}}$ is the extension of φ to $\mathbb{C}$ as in (6.1) below, is upper semi-continuous [Sta24, 02FZ]. As by assumption, ξ_1 is isolated in $\varphi_{\mathbb{C}}^{-1}(\kappa,b)$, it holds that $\dim_{\xi_1}(\varphi_{\mathbb{C}}^{-1}(\kappa,b))=0$ and hence, there is an open neighborhood $U\subseteq U_1$ of ξ_1 such that $\varphi_{|U}^{-1}(\varphi_{|U}(\xi))$ is finite for all $\xi\in U$.

Aiming for a contradiction, we assume that φ is injective on $U\setminus T$. As all fibers of $\varphi_{|U}$ are finite, the Main Theorem in [BOT06] tells us that $\varphi_{|U}$ is injective and hence an open map by the theorem of invariance of domain. Hence, ξ_1 is mapped to a point in the interior of $\varphi(U_1)$, thus contradicting the fact that (κ,b) is a boundary point of $\varphi(U_1)$. This concludes the proof. □

We remark that case (ii) of Theorem 5.2 implies that the network is nondegenerate and holds if $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$ is finite. For nondegenerate networks, case (i) holds if the sets $\mathbb{V}_{>0}(F_{\kappa',b'})$ are finite for all parameter choices (as required for the original Nondegenerate Conjecture in [JS17].) Note also that (i) can be replaced by $\mathbb{V}_{>0}(F_{\kappa',b'})$ being finite *and nonempty* for all (κ',b') in an open neighborhood of (κ,b) in $\mathbb{R}^m_{>0}\times\mathbb{R}^d$ (which already implies nondegeneracy of the network).

Example 5.3. To understand some of the phenomena behind assumptions (i) and (ii) in Theorem 5.2, we consider the network with $s = n = 2$ and stoichiometric and kinetic matrix

$$N = \begin{bmatrix} 1 & -4 & 2 & -6 & 11 & -4 & 12 & -14 & 1 & -6 & 15 & -18 & 10 & 0 \\ 1 & -4 & 2 & -6 & 11 & -4 & 12 & -14 & 1 & -6 & 15 & -18 & 0 & 10 \end{bmatrix},$$

$$M = \begin{bmatrix} 5 & 4 & 3 & 3 & 3 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 2 & 1 & 3 & 2 & 1 & 5 & 4 & 3 & 2 & 1 & 1 \end{bmatrix},$$

which gives rise to a degenerate steady state system f (hence all positive steady states are degenerate). Taking $\kappa = (1, \ldots, 1)$ gives

$$f_\kappa = \begin{pmatrix} x_1\, x_2\, ((x_1-1)^2 + (x_2-1)^2)\, ((x_1-1)^2 + (x_2-2)^2) \\ x_1\, x_2\, ((x_1-1)^2 + (x_2-1)^2)\, ((x_1-1)^2 + (x_2-2)^2) \end{pmatrix} \tag{5.3}$$

for which $\mathbb{V}_{>0}(f_\kappa) = \{(1,1),(1,2)\}$, and hence has two (isolated) points. Therefore, we have multistationarity but not nondegenerate multistationarity. Note that condition (ii) of Theorem 5.2 is not satisfied, as $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ is infinite for this κ. The second part of condition (i) of Theorem 5.2 is not satisfied either, as $\mathbb{V}_{>0}(f_\kappa)$ is infinite for any $\kappa = (1, \ldots, 1, a, a)$ with $0 < a < 1$.

Remark 5.4. One might expect that the existence of two isolated positive steady states for some parameter value (κ, b) in a nondegenerate network is sufficient for nondegenerate multistationarity. However, some additional assumptions, like assumptions (i) or (ii) in Theorem 5.2, are needed to guarantee that multiple nondegenerate steady states arise for some small perturbation of (κ, b).

To illustrate this, we consider a modification of Example 5.3 with matrices

$$N = \begin{bmatrix} 1 & -4 & 2 & -6 & 11 & -4 & 12 & -14 & 1 & -6 & 15 & -18 & 0 & 10 & 0 \\ 1 & -4 & 2 & -6 & 11 & -4 & 12 & -14 & 1 & -6 & 15 & 0 & -18 & 0 & 10 \end{bmatrix} \in \mathbb{Z}^{2\times 15},$$

$$M = \begin{bmatrix} 5 & 4 & 3 & 3 & 3 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 2 & 1 & 3 & 2 & 1 & 5 & 4 & 3 & 2 & 2 & 1 & 1 \end{bmatrix} \in \mathbb{Z}^{2\times 15}_{\geq 0}.$$

The matrix N has full rank, the steady state system f_κ is nondegenerate, and $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ has generically four elements.

An easy computation shows that if either $\kappa_{12} = \kappa_{13}$ or $\kappa_{14} = \kappa_{15}$, but not simultaneously, then there are no positive steady states. If $\kappa_{12} = \kappa_{13}$ and $\kappa_{14} = \kappa_{15}$, then $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ has dimension 1 (hence all zeros are degenerate), and for $\kappa^* = (1, \ldots, 1)$, f_{κ^*} coincides with (5.3), and there are thus two degenerate and isolated positive steady states. Analogously to Example 5.3, assumptions (i) and (ii) of Theorem 5.2 are not satisfied.

We show next that any perturbation of κ^* yields a system without nondegenerate real zeros. To see this, we need to assume $\kappa_{12} \neq \kappa_{13}$ and $\kappa_{14} \neq \kappa_{15}$, in which case the polynomial $f_{\kappa,1} - f_{\kappa,2}$ yields $x_2 = \frac{5(\kappa_{14}-\kappa_{15})}{9(\kappa_{12}-\kappa_{13})}$ at any zero. By inserting this expression into $f_{\kappa,1}$ and clearing denominators and the factor $x_1\, x_2$, we obtain a degree 4 polynomial in x_1. We reparametrize the polynomial using $\kappa_{12} = \epsilon + \kappa_{13}$ and $\kappa_{14} = a \cdot \epsilon + \kappa_{15}$ with $\epsilon \neq 0$ and $a > 0$ (so x_2 is positive), such that it becomes:

$$\begin{aligned} g := \epsilon^4 \Big(& 6561\kappa_1 x_1^4 - 26244\kappa_2 x_1^3 + (4050\, a^2\kappa_3 - 21870\, a\, \kappa_4 + 72171\kappa_5) x_1^2 \\ & + (-8100\, a^2\kappa_6 + 43740\, a\, \kappa_7 - 91854\kappa_8) x_1 + 625\, a^4\kappa_9 \\ & - 6750\, a^3\kappa_{10} + 30375\, a^2\kappa_{11} - 65610\, a\, \kappa_{13} + 65610\kappa_{15} \Big). \end{aligned}$$

At $\kappa_i = 1$, for $i \in \{1, \ldots, 11, 13, 15\}$, g has the roots

$$1 \pm (\tfrac{5a}{9} - 2)\,\mathrm{I}, \qquad 1 \pm (\tfrac{5a}{9} - 1)\mathrm{I},$$

i.e., either 4 complex roots, or 2 complex roots and a double positive root, independently of ϵ. Hence, no choice of $\epsilon \neq 0$ and $a > 0$ leads to nondegenerate positive steady states. If $a \notin \{\frac{9}{5}, \frac{18}{5}\}$, then any small perturbation of the κ_i for $i \in \{1, \dots, 11, 13, 15\}$ yields a polynomial with 4 complex simple roots as well. If a perturbation for $a \in \{\frac{9}{5}, \frac{18}{5}\}$ yielded positive real simple roots, then the same would be true after perturbing a, which we already shown is not the case.

This reparameterization shows that any small perturbation of κ^* will yield a system with no nondegenerate real zero. In particular, this illustrates that extra assumptions such as (i) and (ii) in Theorem 5.2 are necessary if one aims at obtaining nondegenerate multistationarity by perturbing the given isolated zeros.

The system f_κ in this discussion admits, however, choices of parameters for which there are two nondegenerate positive steady states, but these are "far" from κ^* and hence there existence is presumably independent of the existence of two zeros for κ^*.

6. Proof of Theorem 3.1, Theorem 3.4 and Theorem 4.4

In this final section we give the full theorem on augmented vertically parametrized systems from [FHPE24] for completeness, and use it to derive Theorem 3.1, Theorem 3.4 and Theorem 4.4.

An augmented vertically parametrized system is one of the form

$$g = \Big(C(\kappa \circ x^M),\, Lx - b\Big) \in \mathbb{C}[\kappa, b, x^{\pm}]^{s+\ell}, \qquad \operatorname{rk}(C) = s, \quad s \leq n, \quad 0 \leq \ell \leq n - s.$$

With $\ell = 0$, the steady state system is of this form (there is no linear part), and with $\ell = n - s$ so is the augmented steady state system. The complex incidence variety

$$\mathcal{I}_g := \{(\kappa, b, x) \in \mathbb{C}^m \times \mathbb{C}^\ell \times (\mathbb{C}^*)^n : g(\kappa, b, x) = 0\}$$

is nonsingular and admits a parametrization

$$\phi\colon \ker(C) \times (\mathbb{C}^*)^n \to \mathbb{C}^m \times \mathbb{C}^\ell \times (\mathbb{C}^*)^n, \quad (w, h) \mapsto (w \circ h^M, Lh^{-1}, h^{-1}), \tag{6.1}$$

analogous to (2.4). We let

$$\mathcal{Z}_g = \{(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell : \mathbb{V}_{>0}(g_{\kappa,b}) \neq \varnothing\}.$$

The following theorem is [FHPE24, Theorem 3.7] applied (with the notation of loc. cit.) to $\mathcal{A} = \mathbb{R}^m_{>0} \times \mathbb{R}^\ell$ and $\mathcal{X} = \mathbb{R}^n_{>0}$, combined with the considerations at the start of Section 3.5 and Remark 2.4 in [FHPE24]. Additionally, we add the condition (setZC), which is equivalent to (setZ) by [FHPE24, Lemma 2.6]. At the core of the proof is that degenerate zeros of g correspond to critical values of the projection of ϕ onto parameter space [FHPE24, Proposition 3.3].

Theorem 6.1. *For a real augmented vertically parametrized system*

$$g = (C(\kappa \circ x^M),\, Lx - b) \in \mathbb{R}[\kappa, b, x^{\pm}]^{s+\ell} \quad \text{with } s \leq n \text{ and } 0 \leq \ell \leq n - s,$$

assume that $\mathcal{I}_g \cap (\mathbb{R}^m_{>0} \times \mathbb{R}^\ell \times \mathbb{R}^n_{>0}) \neq \varnothing$. Consider the following statements:

- (deg1) *$g_{\kappa,b}$ has a nondegenerate zero in $(\mathbb{C}^*)^n$ for some $(\kappa, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$.*
- (degX1) *$g_{\kappa,b}$ has a nondegenerate zero in $\mathbb{R}^n_{>0}$ for some $(\kappa, b) \in \mathbb{R}^m_{>0} \times \mathbb{R}^\ell$.*
- (degXG) *There exists a nonempty Zariski open subset $\mathcal{U} \subseteq \mathcal{I}_g$ such that for all $(\kappa, b, x) \in \mathcal{U} \cap (\mathbb{R}^m_{>0} \times \mathbb{R}^\ell \times \mathbb{R}^n_{>0})$, x is a nondegenerate zero of $g_{\kappa,b}$.*
- (degAll) *For generic $(\kappa, b) \in \mathcal{Z}_g$, all zeros of $g_{\kappa,b}$ in $(\mathbb{C}^*)^n$ are nondegenerate.*
- (setE) *$\mathcal{Z}_g$ has nonempty Euclidean interior in $\mathbb{R}^m_{>0} \times \mathbb{R}^\ell$.*
- (setZ) *$\mathcal{Z}_g$ is Zariski dense in $\mathbb{C}^m \times \mathbb{C}^\ell$.*
- (setZC) *$\{(\kappa, b) \in \mathbb{C}^m_{>0} \times \mathbb{C}^\ell : \mathbb{V}_{\mathbb{C}^*}(g_{\kappa,b}) \neq \varnothing\}$ is Zariski dense in $\mathbb{C}^m \times \mathbb{C}^\ell$.*
- (dim1) *$\mathbb{V}_{\mathbb{C}^*}(g_{\kappa,b})$ has pure dimension $n - s - \ell$ for at least one $(\kappa, b) \in \mathcal{Z}_g$.*
- (dimG) *$\mathbb{V}_{\mathbb{C}^*}(g_{\kappa,b})$ has pure dimension $n - s - \ell$ for generic $(\kappa, b) \in \mathcal{Z}_g$.*
- (real) *For generic $(\kappa, b) \in \mathcal{Z}$, $\mathbb{V}_{\mathbb{R}^*}(g_{\kappa,b})$ has pure dimension $n - s - \ell$, and $\mathbb{V}_{>0}(g_{\kappa,b})$ has dimension $n - s - \ell$ as a semialgebraic set.*

(rad) *$g_{\kappa,b}$ generates a radical ideal in $\mathbb{C}[x^{\pm}]$ for generic $(\kappa, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$, and g generates a radical ideal in $\mathbb{C}(\kappa, b)[x^{\pm}]$.*

(reg) *For generic $(\kappa, b) \in \mathbb{C}^m \times \mathbb{C}^\ell$, $\mathbb{V}_{\mathbb{C}^*}(g_{\kappa,b})$ is a nonsingular complex algebraic variety (in particular, different irreducible components do not intersect).*

Then the following holds:

- *The statements* (deg1), (degX1), (degXG), (degAll), (setE), (setZ), (setZC), (dimG), (dim1) *are all equivalent to the condition*

$$\operatorname{rk} \begin{bmatrix} C \operatorname{diag}(w) M^\top \operatorname{diag}(h) \\ L \end{bmatrix} = s + \ell \quad \text{for some } (w, h) \in \ker(C) \times (\mathbb{C}^*)^n.$$

- *Any of the statements mentioned above implies* (real) *and* (reg).
- *The statement* (rad) *holds, independently of the other statements.*

6.1. **Proof of Theorem 3.1 and Theorem 3.4.** We apply Theorem 6.1 to the steady state system f, where $\ell = 0$ and $\mathcal{Z} = \mathcal{Z}_f$ for Theorem 3.1, or to the augmented steady state system F with $\ell = n - s$ and $\mathcal{Z}_F = \mathcal{Z}_{\text{cc}}$ for Theorem 3.4. In both cases, the condition $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$ implies $\mathcal{Z} \neq \varnothing$ and $\mathcal{Z}_{\text{cc}} \neq \varnothing$ by Proposition 2.3, and hence $\mathcal{I}_f \cap (\mathbb{R}^m_{>0} \times \mathbb{R}^n_{>0}) \neq \varnothing$ and $\mathcal{I}_F \cap (\mathbb{R}^m_{>0} \times \mathbb{R}^d \times \mathbb{R}^n_{>0}) \neq \varnothing$. Therefore, we can apply Theorem 6.1.

The equivalence between (i)–(vi) in Theorem 3.1 and in Theorem 3.4 correspond to the first bullet point of Theorem 6.1 by using (deg1), (degAll), (setE), (setZ), (dim1).

For Theorem 3.1, we now have that if the equivalent statements hold, Theorem 6.1 gives that (dimG) and (real) hold, and so does the first bullet point. For the second bullet point, as (dim1) and (deg1) do not hold, we obtain the statement about dimension and degeneracy. As (setZC) does not hold, the complex varieties are nonempty only for parameters in a proper Zariski closed subset, hence generically empty.

The last bullet point of both Theorem 3.1 and Theorem 3.4 agree with that of Theorem 6.1.

For Theorem 3.4, the first bullet point follows from (setZ), (real) and (degAll). □

6.2. **Proof of Theorem 4.4.** We first show (ii) ⇔ (iii) ⇔ (iv). As f is nondegenerate, $\mathcal{Z}$ has nonempty Euclidean interior and is Zariski dense in $\mathbb{C}^m$. Let H be the augmented vertical system constructed by appending $x_i - c$ to f, and let $\mathcal{Z}_H \subseteq \mathbb{R}^m_{>0} \times \mathbb{R}$ be the subset of parameters (κ, c) such that $H_{\kappa,c}$ has a zero in $\mathbb{R}^n_{>0}$. For each $\kappa \in \mathcal{Z}$, let c_κ be the x_i value of some $x \in \mathbb{V}_{>0}(f_\kappa)$, which is nonempty by hypothesis. Then $(\kappa, c_\kappa) \in \mathcal{Z}_H$ as H has a positive zero. Hence we have a dominant map of varieties

$$\rho \colon \overline{\mathcal{Z}_H} \to \overline{\mathcal{Z}} = \mathbb{C}^m, \qquad (\kappa, c) \mapsto \kappa,$$

where the overline refers to the Zariski closure in complex spaces. By construction, for $\kappa \in \mathbb{C}^m$, $\rho^{-1}(\kappa) = \pi_i(\mathbb{V}_{\mathbb{C}^*}(f_\kappa))$ if nonempty. By the theorem of the dimension of fibers [Sha13, Theorem 1.25], we conclude that the fibers of ρ have generically dimension $\dim(\overline{\mathcal{Z}_H}) - m$.

Condition (iii) is exactly the failure of the rank condition in Theorem 6.1, namely that the matrix

$$\begin{bmatrix} C \operatorname{diag}(w) M^\top \operatorname{diag}(h) \\ e_i \end{bmatrix} \in \mathbb{C}^{(s+1)\times n},$$

with e_i is the canonical row vector with 1 in the i-th entry and zero otherwise, has rank at most s for all $w \in \ker(C) \subseteq \mathbb{C}^m$ and $h \in (\mathbb{C}^*)^n$. Hence, (iii) holds if and only if (setZC) does not hold, that is, $\overline{\mathcal{Z}_H}$ is a proper Zariski closed subset of $\mathbb{C}^m \times \mathbb{C}$, that is, has dimension at most m. This in turn holds if and only if the fibers of ρ have generically dimension 0, hence if and only if $\pi_i(\mathbb{V}_{\mathbb{C}^*}(f_\kappa))$ has generically dimension 0, giving the equivalence with (ii).

For the equivalence between (ii) and (iv), let I_κ denote the specialization of I to κ. The closure theorem gives us that $\mathbb{V}_{\mathbb{C}^*}(I_\kappa \cap \mathbb{C}[x_i^{\pm}]) = \overline{\pi_i(\mathbb{V}_{\mathbb{C}^*}(f_\kappa))}$ for all $\kappa \in \mathbb{C}^m$ [CLO15]. So $\pi_i(\mathbb{V}_{\mathbb{C}^*}(f_\kappa))$ is finite if and only if $I_\kappa \cap \mathbb{C}[x_i^{\pm}]$ is not the zero ideal.

Given $G \subseteq \mathbb{C}(\kappa)[x^{\pm}]$ a Gröbner basis of I, it specializes to a Gröbner basis of I_κ for generic κ. This gives that $I_\kappa \cap \mathbb{C}[x_i^{\pm}]$ is generated by one polynomial for generic κ if and only if $I \cap \mathbb{C}(\kappa)[x_i^{\pm}]$ is, and from this the equivalence follows.

As $\mathcal{Z}$ is Zariski dense, we readily have that (ii) implies (i). The proof of the equivalences is complete if we show (i) implies (iii). Given $\kappa \in \mathbb{R}^m_{>0}$, let V^{nd}_κ be the union of the irreducible components of $\mathbb{V}_{\mathbb{C}^*}(f_\kappa)$ that contain a nondegenerate positive zero of f_κ. Then [PEF22, Theorem 5.3] states that (iii) is equivalent to $V^{\mathrm{nd}}_\kappa \cap \mathbb{R}^n_{>0}$ having local ACR for X_i for all $\kappa \in \mathbb{R}^m_{>0}$ for which $V^{\mathrm{nd}}_\kappa \neq \varnothing$. By Theorem 3.1, all zeros of f_κ are nondegenerate generically for $\kappa \in \mathcal{Z}$ and hence, generically, $\mathbb{V}_{>0}(f_\kappa) = V^{\mathrm{nd}}_\kappa \cap \mathbb{R}^n_{>0}$. With this in place the equivalence between (i) and (iii) follows.

We now show the bullet points of the statement. For the first bullet point, if (i) does not hold, neither does (ii), and then $\pi_i(\mathbb{V}_{\mathbb{C}^*}(f_\kappa))$ equals $\mathbb{C}^*$ minus a finite number of points for generic $\kappa \in (\mathbb{C}^*)^m$. As $\mathcal{Z}$ is Zariski dense, it follows that $\pi_i(\mathbb{V}_{>0}(f_\kappa))$ also is infinite for generic $\kappa \in \mathcal{Z}$ and hence, generically, there is no local ACR.

For the first part of the second bullet point, the extra assumption implies $\mathbb{V}_{>0}(f_\kappa) = V^{\mathrm{nd}}_\kappa \cap \mathbb{R}^n_{>0}$ for the given κ. Then, the specialized version [PEF22, Theorem 4.11] of [PEF22, Theorem 5.3] gives that $\mathbb{V}_{>0}(f_\kappa)$ has local ACR for X_i if and only if the matrix $J_{f_\kappa}(x)_{\setminus i}$ obtained by removing the i-th column of $J_{f_\kappa}(x)$, has rank at most $s-1$ for all $x \in \mathbb{V}_{>0}(f_\kappa)$. But this is guaranteed by (iii) and Proposition 2.3. The second part follows again from [PEF22, Theorem 5.3], as the condition ensures $\mathbb{V}_{>0}(f_\kappa) = V^{\mathrm{nd}}_\kappa \cap \mathbb{R}^n_{>0}$ for all $\kappa \in \mathcal{Z}$ by Proposition 2.3. □

References

[And11] D. F. Anderson. Boundedness of trajectories for weakly reversible, single linkage class reaction systems. *J. Math. Chem.*, 49:2275–2290, 2011.

[ASBL99] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397(6715):168–171, 1999.

[BCY20] B. Boros, G. Craciun, and P. Y. Yu. Weakly reversible mass-action systems with infinitely many positive steady states. *SIAM J. Appl. Math*, 80(4):1936–1946, 2020.

[BI64] A. Ben-Israel. Notes on linear inequalities, I: The intersection of the nonnegative orthant with complementary orthogonal subspaces. *J. Math. Anal. Appl.*, 9:303–314, 1964.

[Bor19] B. Boros. Existence of positive steady states for weakly reversible mass-action systems. *SIAM J. Math. Anal.*, 51(1):435–449, 2019.

[BOT06] A. Blokh, L. Oversteegen, and E. D. Tymchatyn. On almost one-to-one maps. *Trans. Amer. Math. Soc.*, 358(11):5003–5014, 2006.

[BP18] M. Banaji and C. Pantea. The inheritance of nondegenerate multistationarity in chemical reaction networks. *SIAM J. Appl. Math.*, 78:1105–1130, 2018.

[CF05] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks. I. The injectivity property. *SIAM J. Appl. Math.*, 65(5):1526–1546, 2005.

[CF06] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: extensions to entrapped species models. *Syst. Biol. (Stevenage)*, 153:179–186, 2006.

[CF10] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: Semiopen mass action systems. *SIAM J. Appl. Math.*, 70(6):1859–1877, 2010.

[CFMW17] C. Conradi, E. Feliu, M. Mincheva, and C. Wiuf. Identifying parameter regions for multistationarity. *PLOS Comput. Biol.*, 13(10):e1005751, 2017.

[CFRS07] C. Conradi, D. Flockerzi, J. Raisch, and J. Stelling. Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. *Proc. Nat. Acad. Sci.*, 104(49):19175–80, 2007.

[CFW20] D. Cappelletti, E. Feliu, and C. Wiuf. Addition of flow reactions preserving multistationarity and bistability. *Math. Biosc.*, 320(108295), 2020.

[CGK20] D. Cappelletti, A. Gupta, and M. Khammash. A hidden integral structure endows absolute concentration robust systems with resilience to dynamical concentration disturbances. *J. R. Soc. Interface*, 17:20200437, 2020.

[CLO15] D. A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer International Publishing, 2015.

[Fei72] M. Feinberg. Complex balancing in general kinetic systems. *Arch. Rational. Mech. Anal.*, 49(3):187, 1972.

[Fei87] M. Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—I. the deficiency zero and deficiency one theorems. *Chem. Eng. Sci.*, 42(10):2229–2268, 1987.

[Fei95] M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rational. Mech. Anal.*, 132(4):311, 1995.

[Fei19] M. Feinberg. *Foundations of chemical reaction network theory*, volume 202 of *Applied Mathematical Sciences*. Springer, Cham, 2019.

[FH77] M. Feinberg and F. J. M. Horn. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. *Arch. Rational. Mech. Anal.*, 66(1):83–97, 1977.

[FH24] E. Feliu and O. Henriksson. Toricity of vertically parametrized systems with applications to reaction network theory, 2024. Preprint: `arXiv:2411.15134`.

[FHPE24] E. Feliu, O. Henriksson, and B. Pascual-Escudero. Generic consistency and nondegeneracy of vertically parametrized systems, 2024. Preprint: `arXiv:2304.02302`.

[FW12] E. Feliu and C. Wiuf. Preclusion of switch behavior in networks with mass-action kinetics. *Appl. Math. Comput.*, 219(4):1449–1467, 2012.

[FW13] E. Feliu and C. Wiuf. Simplifying biochemical models with intermediate species. *J. R. Soc. Interface*, 10:20130484, 2013.

[GES05] K. Gatermann, M. Eiswirth, and A. Sensse. Toric ideals and graph theory to analyze Hopf bifurcations in mass action systems. *J. Symbolic Comput.*, 40(6):1361–1382, 2005.

[GPGH+25] L. D. García Puente, E. Gross, H. A. Harrington, M. Johnston, N. Meshkat, M. Pérez Millán, and A. Shiu. Absolute concentration robustness: Algebra and geometry. *J. Symb. Comput.*, 128:102398, 2025.

[GW64] C. M. Guldberg and P. Waage. Studier i affiniteten. *Videnskabs-Selskabet i Christiania*, pages 35–45, 1864.

[HJ72] F. Horn and R. Jackson. General mass action kinetics. *Arch. Rational. Mech. Anal.*, 47:81–116, 1972.

[HT79] V. Hárs and J. Tóth. On the inverse problem of reaction kinetics. In *Colloquia Mathematica Societatis János Bolyai 30., Qualitative Theory of Differential Equations, Szeged (Hungary)*, pages 363–379, 1979.

[JEKF18] H. Ji, P. Ellison, D. Knight, and M. Feinberg. CRNToolbox version 2.35, 2018. available at `http://crnt.osu.edu/CRNTWin`.

[JS13] B. Joshi and A. Shiu. Atoms of multistationarity in chemical reaction networks. *J. Math. Chem.*, 51(1):153–178, 2013.

[JS15] B. Joshi and A. Shiu. A survey of methods for deciding whether a reaction network is multistationary. *Math. Model. Nat. Phenom.*, 10(5):47–67, 2015.

[JS17] B. Joshi and A. Shiu. Which small reaction networks are multistationary? *SIAM J. Appl. Dyn. Syst.*, 16(2):802—-833, 2017.

[JTZ24] Y. Jiao, X. Tang, and X. Zeng. Multistability of small zero-one reaction networks, 2024. Preprint: `arXiv:2406.11586`.

[KD24] S. Kothari and A. Deshpande. Endotactic and strongly endotactic networks with infinitely many positive steady states. *J. Math. Chem.*, 62(6):1454–1478, 2024.

[KPMD+12] R. L. Karp, M. Pérez Millán, T. Dasgupta, A. Dickenstein, and J. Gunawardena. Complex-linear invariants of biochemical networks. *J. Theoret. Biol.*, 311:130–138, 2012.

[LK99] M. Laurent and N. Kellershohn. Multistability: a major means of differentiation and evolution in biological systems. *Trends Biochem. Sciences*, 24(11):418–422, 1999.

[LMI+23] T. E. Loman, Y. Ma, V. Ilin, S. Gowda, N. Korsbo, N. Yewale, C. Rackauckas, and S. A. Isaacson. Catalyst: Fast and flexible modeling of reaction networks. *PLOS Comput. Biol.*, 19(10):1–19, 10 2023.

[LSR22] C. Lüders, T. Sturm, and O. Radulescu. ODEbase: a repository of ODE systems for systems biology. *Bioinf. Adv.*, 2(1), 2022.

[MFR+15] S. Müller, E. Feliu, G. Regensburger, C. Conradi, A. Shiu, and A. Dickenstein. Sign conditions for injectivity of generalized polynomial maps with applications to chemical reaction networks and real algebraic geometry. *Found. Comput. Math.*, 16(1):69–97, 2015.

[MST22] N. Meshkat, A. Shiu, and A. Torres. Absolute concentration robustness in networks with low-dimensional stoichiometric subspace. *Vietnam J. Math.*, 50:623–651, 2022.

[Mum76] D. Mumford. *Algebraic Geometry I, Complex Projective Varieties*. Classics in Mathematics. Springer Verlag, 1976.

[OSC24] Oscar – open source computer algebra research system, version 1.0.0, 2024.

[PEF22] B. Pascual-Escudero and E. Feliu. Local and global robustness at steady state. *Math. Methods Appl. Sci.*, 45(1):359–382, 2022.

[PM11] M. Pérez Millán. *Métodos algebraicos para el estudio de redes bioquímicas*. PhD Thesis, 2011. Available at https://bibliotecadigital.exactas.uba.ar/collection/tesis/document/tesis_n5103_PerezMillan.

[PMDSC12] M. Pérez Millán, A. Dickenstein, A. Shiu, and C. Conradi. Chemical reaction systems with toric steady states. *Bull. Math. Biol.*, 74:1027–1065, 2012.

[SdW19] A. Shiu and T. de Wolff. Nondegenerate multistationarity in small reaction networks. *Discrete Contin. Dyn. Syst. Ser. B*, 24(6):2683–2700, 2019.

[SF10] G. Shinar and M. Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–91, 2010.

[Sha13] I. R. Shafarevich. *Basic Algebraic Geometry 1: Varieties in Projective Space*. Springer, Heidelberg, third edition, 2013.

[Sta24] The Stacks Project. Available at https://stacks.math.columbia.edu, 2024.

[Vol72] A. I. Vol'pert. Differential equations on graphs. *Math. USSR-Sb*, 17:571–582, 1972.

Authors' addresses:

Elisenda Feliu, University of Copenhagen efeliu@math.ku.dk
Oskar Henriksson, University of Copenhagen oskar.henriksson@math.ku.dk
Beatriz Pascual-Escudero, Universidad Politécnica de Madrid beatriz.pascual@upm.es

D

Toricity of vertically parametrized systems with applications to reaction network theory

Elisenda Feliu
Department of Mathematical Sciences
University of Copenhagen

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Publication details

Preprint: https://doi.org/10.48550/arXiv.2411.15134 (2024)

TORICITY OF VERTICALLY PARAMETRIZED SYSTEMS WITH APPLICATIONS TO REACTION NETWORK THEORY

ELISENDA FELIU AND OSKAR HENRIKSSON

Abstract. In this paper, we present new necessary conditions and sufficient conditions for the (positive parts of) the varieties of vertically parametrized systems to admit monomial parametrizations. The conditions are based on a combination of polyhedral geometry and previously known results about injectivity of monomial maps. The motivation arises from the study of steady state varieties of reaction networks, as toricity simplifies the determination of multistationarity substantially.

1. Introduction

Toric varieties are central objects in combinatorial algebraic geometry, and appear naturally in many applications, including equation solving [Tel22], statistics [GMS06], phylogenetics [SS05], and chemical reaction network theory [CDSS09]. Much is known about the geometry of toric varieties (see, e.g., [CLS11] for an overview), but effectively deciding if a given variety is toric from an implicit description remains a hard problem, that has recently been approached from the point of view of, e.g., symbolic computation [GIR+20] and Lie theory [MP23, KV24].

In this paper, we study the following notion of ***parametric toricity*** over $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$: We say that a parametrized system $F \in \mathbb{R}[\kappa_1, \dots, \kappa_m, x_1^{\pm}, \dots, x_n^{\pm}]^s$ (with parameters κ and variables x) displays toricity over $\mathbb{G}$ if there is an exponent matrix $A \in \mathbb{Z}^{d\times n}$ such that all nonempty zero sets $\mathbb{V}_{\mathbb{G}}(F_\kappa) \subseteq \mathbb{G}^n$ for $\kappa \in \mathbb{G}^m$ admit a monomial parametrization of the form

$$\mathbb{G}^d \to \mathbb{G}^n, \quad t \mapsto \alpha_\kappa \circ t^A$$

with $\alpha_\kappa \in \mathbb{G}^n$ depending on κ. In this case, we write $\mathbb{V}_{\mathbb{G}}(F_\kappa) = \alpha_\kappa \circ \mathcal{T}_A^{\mathbb{G}}$ where $\mathcal{T}_A^{\mathbb{G}} = \{t^A : t \in \mathbb{G}^d\}$, and we view $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ as a coset of the multiplicative group $\mathcal{T}_A^{\mathbb{G}}$. Detecting this type of parametric toricity can formulated as a quantifier elimination problem [RS21a], which gives an algorithm for solving the problem over $\mathbb{R}^*$ and $\mathbb{R}_{>0}$, albeit at a substantial computational cost. Another general but computationally intense approach for $\mathbb{G} = \mathbb{R}_{>0}$ is taken in [SF19a, RS21b] which give sufficient conditions in terms of Gröbner bases and comprehensive Gröbner systems.

Here, we focus on the case when F is ***vertically parametrized*** in the language of [HR22, FHPE24], in the sense that it can be written in the form

$$F = C(\kappa \circ x^M) \in \mathbb{R}[\kappa_1, \dots, \kappa_m, x_1^{\pm}, \dots, x_n^{\pm}]^s$$

where each row of $C \in \mathbb{R}^{s\times m}$ encodes a linear combination of m monomials with exponents given by the columns of $M \in \mathbb{Z}^{n\times m}$, scaled by the parameters $\kappa = (\kappa_1, \dots, \kappa_m)$ through component-wise multiplication $\circ$. For example, the following is a vertically parametrized system:

$$F = \begin{bmatrix} 2\kappa_2\, x_1x_2^2 - \kappa_3\, x_1^2x_2^3 \\ \kappa_1\, x_2 - \kappa_2\, x_1x_2^2 + 2\kappa_3\, x_1^2x_2^3 \end{bmatrix} \quad \text{where} \quad C = \begin{bmatrix} 0 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Vertically parametrized systems describe the steady states of reaction networks (our motivating scenario), and the critical points of multivariate polynomials. They include sparse polynomial systems of fixed support, but the framework allows also for fixing the sign of the coefficients and ratios between coefficients of the same monomial. For an overview of the algebraic-geometric properties of vertically parametrized systems, see [FHPE24].

In the setting of reaction network theory, the parametric notion of toricity described above is motivated by the fact that multistationarity can be established through a simple sign condition [MDSC12, MFR+15]. In this context, several other sufficient conditions for toricity are known, for instance the existence of a *partitioning kernel basis* of the coefficient matrix C detects the existence of a binomial generating set, which is obtained by taking linear combinations of the original generators. This criterion completely characterizes binomiality for homogeneous ideals [MDSC12, CK15]. Other routes to checking toricity involve one of the oldest themes in reaction network theory, namely deficiency theory [Fei95, CDSS09, Bor12], and their extension through the use of network operations that preserve the steady states [Joh22, BiMCS22, HSSY23].

In this work, we approach toricity over $\mathbb{R}_{>0}$ by considering the intermediate scenarios that arise when $\mathbb{V}_{>0}(F_\kappa)$ is a finite union of cosets of the form $\alpha \circ \mathcal{T}_A^{>0}$, in which case we talk about ***local toricity***, or an infinite union of such cosets, in which case we simply have $\mathcal{T}_A^{>0}$***-invariance***. It turns out that, under mild conditions, finding a maximal rank matrix A for which the system displays $\mathcal{T}_A^{\mathbb{G}}$-invariance does not depend on the choice of $\mathbb{G}$ and reduces to linear algebra computations (Theorem 4.5, Algorithm 4.7). This generalizes the study of quasi-homogeneity [GKZ94, Chapter 6.1] and is explained in Section 4. With this in place, local toricity for generic parameter values κ can be fully characterized and reduced, again, to linear algebra arguments (Theorem 5.3). For toricity for all κ, we give sufficient conditions over $\mathbb{G} = \mathbb{R}_{>0}$ that combine several approaches within real algebraic geometry or homotopy continuation computations. All the strategies are the content of Section 5 and Section 6, and are gathered in Algorithm 7.6.

After having established the general theory, we go back to reaction networks in Section 7 and Section 8. In particular, we study the problems of multistationarity and absolute concentration robustness under the hypothesis of toricity, provide a reduction of reaction networks that reduces the computational cost, and apply our algorithms to biochemically relevant networks from the database ODEbase [LSR22]. The latter shows that our conditions are often enough to completely decide upon toricity for realistic networks.

Notation. The cardinality of a set S is denoted by $\#S$. For $n \in \mathbb{Z}_{\geq 0}$, we let $[n] = \{1, \dots, n\}$. For a field $\Bbbk$, we denote $\Bbbk \setminus \{0\}$ by $\Bbbk^*$. We write $\circ\colon \Bbbk^n \times \Bbbk^n \to \Bbbk^n$ for component-wise multiplication. The vector of all ones is denoted by $\mathbb{1}$ and the zero vector of size d by $0_d \in \mathbb{R}^d$. For $A \in \mathbb{R}^{n\times m}$, we write A_i for the ith column, and A_{i*} for the ith row. The *support* of $v \in \mathbb{R}^n$ is the set $\operatorname{supp}(v) = \{i \in [n] : v_i \neq 0\}$. For $x \in \mathbb{R}^n_{>0}$, we let x^{-1} be defined component-wise, and $x^A \in \mathbb{R}^m$ be defined by $(x^A)_j = x_1^{a_{1j}} \cdots x_n^{a_{nj}}$ for $j \in [m]$.

Acknowledgements. This work has been supported by the Novo Nordisk Foundation, with grant reference NNF20OC0065582, and by the European Union under the Grant Agreement number 101044561, POSALG. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or European Research Council (ERC). Neither the European Union nor ERC can be held responsible for them.

2. Vertically parametrized systems

Throughout, we work with (Laurent) polynomials with coefficients in $\Bbbk = \mathbb{R}$ or $\mathbb{C}$. We consider **vertically parametrized systems** (or **vertical systems** for short), which by definition are parametric systems of the form

$$F = C(a \circ x^M) \in \Bbbk[\kappa_1, \dots, \kappa_m, x_1^{\pm}, \dots, x_n^{\pm}]^s, \tag{2.1}$$

consisting of $s \leq n$ polynomials with parameters $\kappa = (\kappa_1, \dots, \kappa_m)$ and variables $x = (x_1, \dots, x_n)$, encoded by a coefficient matrix $C \in \Bbbk^{s\times m}$ and an exponent matrix $M \in \mathbb{Z}^{n\times m}$. The component-wise product $\kappa \circ x^M$ indicates that the monomial encoded by the ith column of M is scaled by κ_i, while the rows of C give linear combinations of the scaled monomials. An important feature is that F is linear in the parameters and that each parameter accompanies always the same monomial (though a monomial can be accompanied by different parameters; this is achieved by considering repeated columns in M).

We consider also **augmented vertically parametrized systems** of the form

$$\Big(C(\kappa \circ x^M),\ Lx - b\Big) \in \Bbbk[\kappa, b, x^{\pm}]^{s+d}, \qquad s + d \leq n\,,$$

where we also include d affine linear equations, encoded by a coefficient matrix $L \in \Bbbk^{d\times n}$ and parametric constant terms $b = (b_1, \ldots, b_d)$. Geometrically, this corresponds to intersecting the variety given by the vertical system $C(\kappa \circ x^M)$ by a parallel translate of $\ker(L)$. Vertical systems can be seen as augmented vertical systems with $d = 0$.

For an augmented vertical system F, we denote the specialization at $(\kappa, b) \in \Bbbk^{m+d}$ by

$$F_{\kappa,b} := F(\kappa, b, \cdot) \in \Bbbk[x^{\pm}]^{s+d}$$

We are interested in zeros over $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$, given by

$$\mathbb{V}_{\mathbb{G}}(F_{\kappa,b}) = \{x \in \mathbb{G}^n : F_{\kappa,b}(x) = 0\}\,, \tag{2.2}$$

and the set of parameters for which $\mathbb{V}_{\mathbb{G}}(F_{\kappa,b})$ is not empty:

$$\mathcal{Z}_{\mathbb{G}} = \{(\kappa, b) \in \mathbb{G}^m \times \Bbbk^d : F_{\kappa,b}(x) \neq \varnothing\}\,. \tag{2.3}$$

We implicitly consider the ground field to be $\Bbbk = \mathbb{R}$ if $\mathbb{G} = \mathbb{R}^*$ or $\mathbb{R}_{>0}$, and $\Bbbk = \mathbb{C}$ if $\mathbb{G} = \mathbb{C}^*$.

Augmented vertical systems are the focus of study in [FHPE24], where the dimension and nonemptiness of the sets $\mathbb{V}_{\mathbb{G}}(F_{\kappa,b})$ is characterized in terms of nondegenerate zeros. We review these results next, as they play an important role later on.

For an augmented vertical system $F \in \Bbbk[\kappa, b, x^{\pm}]^{s+d}$ with defining matrices $C \in \Bbbk^{s\times m}$ of full row rank, $M \in \mathbb{Z}^{n\times m}$ and $L \in \Bbbk^{d\times n}$, we say that a zero $x \in \mathbb{G}^n$ of $F_{\kappa,b}$ is ***nondegenerate*** if the Jacobian $J_{F_{\kappa,b}}(x)$ has rank $s + d$. A simple equation-counting argument gives that

$$\dim(\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})) \geq n - s - d \tag{2.4}$$

for all $(\kappa, b) \in (\mathbb{C}^*)^m \times \mathbb{C}^d$ such that $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b}) \neq \varnothing$. Additionally, if for a given $(\kappa, b) \in (\mathbb{C}^*)^m \times \mathbb{C}^d$, all irreducible components contain a nondegenerate zero, then the variety $\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})$ has pure dimension $n - s - d$, and (2.4) holds with equality (cf. [CLO15, §9.6, Thm. 9]). We point out that nondegenerate zeros are in particular nonsingular points.

Definition 2.1. Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and $F \in \Bbbk[\kappa, b, x^{\pm}]^{s+d}$ be an augmented vertical system. We say that F is ***generically consistent over*** $\mathbb{G}$ if $\mathcal{Z}_{\mathbb{G}}$ is Zariski dense in $\mathbb{C}^m$. Otherwise we say that F is ***generically inconsistent over*** $\mathbb{G}$.

A summary of the main conclusions from [FHPE24] is given in the next proposition.

Proposition 2.2. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and $F \in \Bbbk[\kappa, b, x^{\pm}]^{s+d}$ be an augmented vertical system with defining matrices $C \in \Bbbk^{s\times m}$ of rank s, $M \in \mathbb{Z}^{n\times m}$, and $L \in \Bbbk^{d\times n}$. Assume that $\ker(C) \cap \mathbb{G}^m \neq \varnothing$. Then F is generically consistent over $\mathbb{G}$ if and only if*

$$\operatorname{rk}\left(\begin{bmatrix} C\operatorname{diag}(w)M^\top \operatorname{diag}(h) \\ L \end{bmatrix}\right) = s + d \quad \text{for some } w \in \ker(C) \text{ and } h \in (\mathbb{C}^*)^n. \tag{2.5}$$

Furthermore, the following holds:

(i) If (2.5) holds, then $\mathcal{Z}_{\mathbb{G}}$ has nonempty Euclidean interior and there exists a nonempty Zariski open subset $\mathcal{U} \subseteq \mathcal{Z}_{\mathbb{G}}$ such that for all $(\kappa, b) \in \mathcal{U}$,

$$\dim(\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})) = \dim(\mathbb{V}_{\mathbb{G}}(F_{\kappa,b})) = n - s - d,$$

and all zeros of $F_{\kappa,b}$ in $(\mathbb{C}^*)^n$ are nondegenerate. If in addition (2.5) holds for all $w \in \ker(C) \cap \mathbb{G}^m$ and $h \in \mathbb{G}^n$, then we can take $\mathcal{U} = \mathcal{Z}_{\mathbb{G}}$.

(ii) If (2.5) does not hold, then for all $(\kappa, b) \in \mathcal{Z}_{\mathbb{G}}$, $\dim(\mathbb{V}_{\mathbb{C}^*}(F_{\kappa,b})) > n - s - d$ and all zeros of $F_{\kappa,b}$ in $(\mathbb{C}^*)^n$ are degenerate.

Proof. This follows from Theorem 3.7, together with Propositions 2.11, 3.2 and 3.11, and Remark 3.6 in [FHPE24]. □

As condition (2.5) does not depend on $\mathbb{G}$, generic consistency over $\mathbb{G}$ is equivalent to generic consistency over $\mathbb{C}^*$, as long as $\ker(C) \cap \mathbb{G}^m \neq \varnothing$. Condition (2.5) characterizes the existence of a nondegenerate zero of $F_{\kappa,b}$ in $(\mathbb{C}^*)^n$ for some $(\kappa, b) \in (\mathbb{C}^*)^m \times \mathbb{C}^d$, and the condition $\ker(C) \cap \mathbb{G}^m \neq \varnothing$ ensures that this also holds after restricting to $\mathbb{G}$. When F is a vertical system, then (2.5) reduces to

$$\operatorname{rk}(C \operatorname{diag}(w) M^\top) = s \quad \text{for some } w \in \ker(C). \tag{2.6}$$

Example 2.3. As the main running example, we consider $\mathbb{G} = \mathbb{R}_{>0}$, and let

$$C = \begin{bmatrix} -1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}, \qquad M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

This gives rise to the vertically parametrized system

$$F = \begin{bmatrix} -\kappa_1 x_1 x_2 + \kappa_2 x_3 + \kappa_3 x_3 \\ -\kappa_1 x_1 x_2 + \kappa_2 x_3 + \kappa_6 x_5 \\ \kappa_4 x_3 x_4 - \kappa_5 x_5 - \kappa_6 x_5 \end{bmatrix} \in \mathbb{R}[\kappa_1, \dots, \kappa_6, x_1^{\pm}, \dots x_5^{\pm}]^3. \tag{2.7}$$

The vectors $v_1 = (1,0,1,1,0,1)$, $v_2 = (0,0,0,1,1,0)$ and $v_3 = (1,1,0,0,0,0)$ form a basis of $\ker(C)$, hence $\ker(C) \cap \mathbb{R}^6_{>0} \neq \varnothing$. By writing $w \in \ker(C)$ as $w = \mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3$, the matrix in (2.6) takes the form

$$C \operatorname{diag}(w) M^\top = \begin{bmatrix} -\mu_1 - \mu_3 & -\mu_1 - \mu_3 & \mu_1 + \mu_3 & 0 & 0 \\ -\mu_1 - \mu_3 & -\mu_1 - \mu_3 & \mu_3 & 0 & \mu_1 \\ 0 & 0 & \mu_1 + \mu_2 & \mu_1 + \mu_2 & -\mu_2 - \mu_1 \end{bmatrix},$$

which has rank 3 for $\mu_1 = \mu_2 = \mu_3 = 1$. Hence F is generically consistent and has positive zeros for parameters in a set $\mathcal{U} \subseteq \mathbb{R}^6_{>0}$ with nonempty Euclidean interior. Furthermore, $\mathbb{V}_{>0}(F_\kappa)$ has dimension 2 for $\kappa \in \mathcal{U}$.

Remark 2.4 (Freely parametrized systems)**.** As discussed in [FHPE24, §3.6], vertical systems include ***freely parametrized systems***, obtained by fixing the support and letting all coefficients vary freely. Given support sets $\mathcal{S}_1, \dots, \mathcal{S}_s \subseteq \mathbb{Z}^n$, the corresponding freely parametrized system is the vertical system given by

$$C = \begin{bmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_s \end{bmatrix} \quad \text{with} \quad C_i = [1 \ \cdots \ 1] \in \mathbb{C}^{1 \times \#\mathcal{S}_i}, \quad \text{and} \quad M = [M_1 \ \cdots \ M_s], \tag{2.8}$$

where the columns $M_i \in \mathbb{Z}^{n \times \#\mathcal{S}_i}$ are the elements of $\mathcal{S}_i$ (in some fixed order). Restricting to $\mathbb{G} = \mathbb{R}_{>0}$ allows us to consider systems with fixed support and coefficients with fixed sign by specifying the signs of the C_i in the construction above (coefficients of free sign are included by repeating the monomial with opposite signs). For instance, the vertical system

$$F = (\kappa_1 - \kappa_2) x_1^3 x_2^2 + \kappa_3 x_2^4 - 2\kappa_4 x_1^6 \quad \in \mathbb{R}[\kappa_1, \kappa_2, \kappa_3, \kappa_4, x_1^{\pm}, x_2^{\pm}] \tag{2.9}$$

can be thought of as a generic system with support $x_1^3 x_2^2$, x_2^3 and x_1^6, where the coefficient of $x_1^3 x_2^2$ may take arbitrary signs, and the coefficients of x_2^3 and x_1^6 are fixed to $+$ and $-$, respectively. (We will revisit this system in Example 6.6.)

Conversely, any vertical system for which the rows of C have pairwise disjoint support can be interpreted as a freely parametrized system, where the sign of some coefficients may be fixed if $\mathbb{G} = \mathbb{R}_{>0}$.

3. Toricity and cosets

In this work, we study whether $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ displays various forms of *toricity*, which we now proceed to define. We will often abbreviate "$\mathbb{R}_{>0}$" by ">0" in superscripts and subscripts.

Definition 3.1. The $\mathbb{G}$-***torus*** and ***torus*** associated with $A \in \mathbb{Z}^{d\times n}$ are, respectively,

$$\mathcal{T}_A^{\mathbb{G}} = \{t^A : t \in \mathbb{G}^d\} \subseteq \mathbb{G}^n, \qquad \mathcal{T}_A = \mathcal{T}_A^{\mathbb{C}^*} = \{t^A : t \in (\mathbb{C}^*)^d\} \subseteq \mathbb{C}^n\,.$$

The Zariski closure of $\mathcal{T}_A^{\mathbb{G}}$ in $(\mathbb{C}^*)^n$ is $\mathcal{T}_A$ and $\mathcal{T}_A^{\mathbb{G}} \subseteq \mathcal{T}_A \cap \mathbb{G}^n$. We have equality in the latter for $\mathbb{G} = \mathbb{R}_{>0}$ and $\mathbb{C}^*$. For $\mathbb{G} = \mathbb{R}^*$, the reverse inclusion might not hold. We view $\mathcal{T}_A^{\mathbb{G}}$ as a multiplicative subgroup of $\mathbb{G}^n$, so that each coset $\alpha \circ \mathcal{T}_A^{\mathbb{G}}$ for $\alpha \in \mathbb{G}^n$ is the image of the monomial map $\mathbb{G}^d \to \mathbb{C}^n$ given by $t \mapsto \alpha \circ t^A$.

Remark 3.2. Whenever $\mathrm{row}_{\mathbb{Z}}(A) = \mathrm{row}_{\mathbb{Z}}(A')$, it holds $\mathcal{T}_A^{\mathbb{G}} = \mathcal{T}_{A'}^{\mathbb{G}}$. For $\mathbb{G} = \mathbb{R}_{>0}$ or $\mathbb{C}^*$, it suffices that $\mathrm{row}_{\mathbb{Q}}(A) = \mathrm{row}_{\mathbb{Q}}(A')$. Furthermore, if $\alpha' \in \alpha \circ \mathcal{T}_A^{\mathbb{G}}$, then $\alpha \circ \mathcal{T}_A^{\mathbb{G}} = \alpha' \circ \mathcal{T}_A^{\mathbb{G}}$.

In the special case where $\mathbb{G} = \mathbb{R}_{>0}$, it is a well-known fact in the reaction networks literature (see, e.g., [Fei95], [MFR+15, §3.2]) that

$$\alpha \circ \mathcal{T}_A^{>0} = \{x \in \mathbb{R}^n_{>0} : \log(x) - \log(\alpha) \in \ker(A)\} = \{x \in \mathbb{R}^n_{>0} : x^B = \alpha^B\}\,,$$

where $B \in \mathbb{Z}^{n\times(n-\mathrm{rk}(A))}$ is a matrix whose columns form a basis for $\ker_{\mathbb{Q}}(A)$. Cosets of this form are sometimes called *log-parametrized* sets in the context of reaction networks [HM23] and *log-linear* or *log-affine* sets in the context of algebraic statistics, see, e.g. [Sul18, Chapter 6–7].

Definition 3.3. For $A \in \mathbb{Z}^{d\times n}$, $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$, and a set $X \subseteq \mathbb{G}^n$, we say that:

- X is $\mathcal{T}_A$-***invariant over*** $\mathbb{G}$ if
$$X \circ \mathcal{T}_A^{\mathbb{G}} \subseteq X,$$
that is, $x \circ t^A \in X$ for all $x \in X$ and $t \in \mathbb{G}^d$. In this case, X is a union of $\mathcal{T}_A^{\mathbb{G}}$-cosets. We denote the set of these cosets by $X/\mathcal{T}_A^{\mathbb{G}}$, and the number of cosets by $\#(X/\mathcal{T}_A^{\mathbb{G}})$.
- X is ***locally*** $\mathcal{T}_A$-***toric over*** $\mathbb{G}$ if $X \neq \varnothing$, X is $\mathcal{T}_A$-invariant over $\mathbb{G}$ and $\#(X/\mathcal{T}_A^{\mathbb{G}}) < \infty$.
- X is $\mathcal{T}_A$-***toric over*** $\mathbb{G}$ if it $\mathcal{T}_A$-invariant over $\mathbb{G}$ and $\#(X/\mathcal{T}_A^{\mathbb{G}}) = 1$.

Definition 3.4. For $A \in \mathbb{Z}^{d\times n}$, $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$, and a vertical system $F \in \Bbbk[\kappa, x^{\pm}]^s$, we say:

- F is $\mathcal{T}_A$-***invariant over*** $\mathbb{G}$ if $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is $\mathcal{T}_A$-invariant for all $\kappa \in \mathbb{G}^m$, that is,
$$F(\kappa, x \circ t^A) = 0, \qquad \text{for all} \quad \kappa \in \mathbb{G}^m,\ x \in \mathbb{V}_{\mathbb{G}}(F_\kappa),\ \text{and } t \in \mathbb{G}^d\,. \tag{3.1}$$
- F is ***generically locally*** $\mathcal{T}_A$-***toric over*** $\mathbb{G}$ if F is generically consistent over $\mathbb{G}$ and $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is locally $\mathcal{T}_A$-toric over $\mathbb{G}$ for generic $\kappa \in \mathcal{Z}_{\mathbb{G}}$.
- F is ***generically*** $\mathcal{T}_A$-***toric over*** $\mathbb{G}$ if F is generically consistent over $\mathbb{G}$ and $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is $\mathcal{T}_A$-toric over $\mathbb{G}$ for generic $\kappa \in \mathcal{Z}_{\mathbb{G}}$.
- F is ***locally*** $\mathcal{T}_A$-***toric over*** $\mathbb{G}$ if F is generically consistent over $\mathbb{G}$ and $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is locally $\mathcal{T}_A$-toric for all $\kappa \in \mathcal{Z}_{\mathbb{G}}$.
- F is $\mathcal{T}_A$-***toric over*** $\mathbb{G}$ if F is generically consistent over $\mathbb{G}$ and $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is $\mathcal{T}_A$-toric for all $\kappa \in \mathcal{Z}_{\mathbb{G}}$.

If $\mathbb{G}$ is omitted, then it is implicitly assumed that $\mathbb{G} = \mathbb{C}^*$.

In the light of Definition 3.4, we take the following two-step approach to detecting or precluding toricity over $\mathbb{G}$ for a vertical system F:

(1) Find a maximal-rank matrix A such that F is $\mathcal{T}_A$-invariant over $\mathbb{G}$, i.e., satisfies (3.1). Under mild assumptions, this comes down to a linear algebra condition, which we discuss in Section 4.

(2) Bound $\#(\mathbb{V}_{\mathbb{G}}(F_\kappa)/\mathcal{T}_A^{\mathbb{G}})$ by applying concepts such as nondegeneracy and injectivity; this is the topic of Sections 5 and 6.

A key observation is that the study of $\mathcal{T}_A$-invariance can be reduced to the case $\mathbb{G} = \mathbb{C}^*$.

Theorem 3.5. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and consider a vertical system $F \in \Bbbk[\kappa, x^{\pm}]^s$ as in (2.1). Let $\mathcal{K} \subseteq \mathbb{G}^m$ be a Euclidean open set such that $\mathbb{V}_{\mathbb{G}}(F_\kappa) \neq \varnothing$ for some $\kappa \in \mathcal{K}$. For $A \in \mathbb{Z}^{d\times n}$, the following are equivalent:*

(i) $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is $\mathcal{T}_A$-invariant over $\mathbb{G}$ for all $\kappa \in \mathcal{K}$.

(ii) F is $\mathcal{T}_A$-invariant over $\mathbb{G}$.

(iii) F is $\mathcal{T}_A$-invariant (over $\mathbb{C}^*$).

Proof. The implications (iii) ⇒ (ii) ⇒ (i) are clear. For the implication (i) ⇒ (iii), form the complex incidence variety

$$\mathcal{E} = \{(\kappa, x) \in (\mathbb{C}^*)^m \times (\mathbb{C}^*)^n : F(\kappa, x) = 0\}\,.$$

Taking Zariski closures in the complex torus, we have $\overline{\mathbb{G}^d} = (\mathbb{C}^*)^d$ and $\overline{\mathcal{E} \cap (\mathcal{K} \times \mathbb{G}^n)} = \mathcal{E}$; this is obvious in the complex case, whereas in the real case, it follows by smoothness of $\mathcal{E}$ [FHPE24, Thm. 3.1] together with denseness of a Euclidean open subset of the real part of a smooth complex variety defined by polynomials with real coefficients (see [PEF22, Thm. 6.5] and [BCR98, Prop. 3.3.6]). With this in place, for the map

$$\Phi\colon (\mathbb{C}^*)^d \times \mathcal{E} \to (\mathbb{C}^*)^m \times (\mathbb{C}^*)^n\,, \quad (t, (\kappa, x)) \mapsto (\kappa, t^A \circ x)\,.$$

we have using (i) that

$$\begin{aligned}\Phi((\mathbb{C}^*)^d \times \mathcal{E}) &= \Phi\left(\overline{\mathbb{G}^d \times (\mathcal{E} \cap (\mathcal{K} \times \mathbb{G}^n))}\right) \subseteq \overline{\Phi\left(\mathbb{G}^d \times (\mathcal{E} \cap (\mathcal{K} \times \mathbb{G}^n))\right)} \\ &= \overline{\Phi(\mathbb{G}^d \times (\mathcal{E} \cap (\mathcal{K} \times \mathbb{G}^n))} \subseteq \overline{\mathcal{E} \cap (\mathcal{K} \times \mathbb{G}^n)), \subseteq \mathcal{E}}\,,\end{aligned}$$

which is equivalent to (iii) by (3.1). □

For common parameter sets $\mathcal{K}$, checking that $\mathbb{V}_{\mathbb{G}}(F_\kappa) \neq \varnothing$ for some $\kappa \in \mathcal{K}$ boils down to a simple computation, as the next lemma indicates.

Lemma 3.6. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and F a vertical system defined by matrices $C \in \Bbbk^{s\times n}$ of rank s and $M \in \mathbb{Z}^{n\times m}$. We have that (i) ⇒ (ii) for the following statements:*

(i) $\mathbb{V}_{\mathbb{G}}(F_\kappa) \neq \varnothing$ for some $\kappa \in \mathcal{K}$.

(ii) $\ker(C) \cap \mathbb{G}^m \neq \varnothing$.

Furthermore, if $\mathcal{K}$ is such that

$$\ker(C) \cap \mathbb{G}^m \subseteq \{\kappa \circ x^M : \kappa \in \mathcal{K}, x \in \mathbb{G}^n\}\,, \tag{3.2}$$

then it also holds that (ii) ⇒ (i). In particular, (i) ⇔ (ii) when $\mathcal{K} = \mathbb{G}^m$.

Proof. (i)⇒(ii): If $x \in \mathbb{V}_{\mathbb{G}}(F_\kappa)$ for $\kappa \in \mathcal{K}$, then $\kappa \circ x^M \in \ker(C) \cap \mathbb{G}^m$. (ii)⇒(i): Let $v \in \ker(C) \cap \mathbb{G}^m$. By (3.2), there exists $\kappa \in \mathcal{K}$ and $x \in \mathbb{G}^n$, such that $\kappa \circ x^M = v$, which implies $x \in \mathbb{V}_{\mathbb{G}}(F_\kappa)$. Finally if $\mathcal{K} = \mathbb{G}^m$, for any $v \in \ker(C) \cap \mathbb{G}^m$, we have $v = v \circ \mathbb{1}^M$. □

Remark 3.7. When $\mathbb{G} = \mathbb{R}_{>0}$, it is well known that any binomial ideal $I \subseteq \mathbb{R}[x]$ gives a toric vanishing locus $\mathbb{V}_{>0}(I)$ [CIK19, Prop. 5.2]. An alternative computational sufficient condition for toricity of a real vertical system F over $\mathbb{R}_{>0}$ when $M \in \mathbb{Z}^{n\times m}_{\geq 0}$ is therefore to compute a reduced Gröbner basis (with respect to any monomial ordering) for the ideal $\langle F\rangle \subseteq \mathbb{R}(\kappa)[x]$ and verify that it specializes for each $\kappa \in \mathbb{R}^m_{>0}$ [SF19b, Rk. 2.4].

Example 3.8. The polynomials in the vertical system F in Example 2.3 generate a binomial ideal in $\mathbb{R}(\kappa)[x]$, and for each $\kappa \in \mathbb{R}^6_{>0}$, the positive zero locus admits the parametrization

$$\mathbb{R}^2_{>0} \to \mathbb{V}_{>0}(F_\kappa)\,, \quad (t_1, t_2) \mapsto \left(t_1, t_2, \tfrac{\kappa_1}{\kappa_2+\kappa_3}\, t_1t_2, \tfrac{\kappa_3(\kappa_5+\kappa_6)}{\kappa_4\kappa_6}, \tfrac{\kappa_1\kappa_3}{\kappa_6(\kappa_2+\kappa_3)}\, t_1t_2\right)\,.$$

This implies in particular that $\mathcal{Z}_{>0} = \mathbb{R}^6_{>0}$ and that F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$ with

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \in \mathbb{Z}^{2\times 5}.$$

Example 3.9. A simple example to show that generic toricity does not imply toricity arises by considering the linear vertical system

$$F = \begin{bmatrix} \kappa_1 x_1 - \kappa_2 x_2 \\ \kappa_1 x_1 - \kappa_3 x_2 \end{bmatrix} \quad \in \mathbb{C}[\kappa_1, \kappa_2, \kappa_3, x_1, x_2]^2.$$

As the zero set is generically one point, F is generically $\mathcal{T}_A$-toric with A the empty matrix. However, when $\kappa_2 = \kappa_3 = 1$, the number of $\mathcal{T}_A$-cosets is infinite as $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ is one dimensional.

Example 3.10. The vertical system defined by the polynomial

$$F = -2\kappa_1 x_1^9 - \kappa_2 x_1^3 x_2^4 + 2\kappa_3 x_2^6 + 2\kappa_4 x_1^6 x_2^2$$

is $\mathcal{T}_A$-invariant over $\mathbb{R}_{>0}$ for $A = [2 \quad 3]$, but the number of $\mathcal{T}_A^{>0}$-cosets of $\mathbb{V}_{>0}(F_\kappa)$ varies with κ: there are three for $\kappa = (0.01, 3, 1, 1)$ and one for $\kappa = (0.01, 1, 1, 1)$ (see Figure 6.1).

Example 3.11. Consider the vertical system with $s = 1$

$$F = -\kappa_1 x_1^2 x_3 + \kappa_2 x_1 x_2^2 + \kappa_3 x_2^2 x_3 \quad \in \mathbb{R}[\kappa_1, \kappa_2, \kappa_3, x_1^{\pm}, x_2^{\pm}, x_3^{\pm}]\,.$$

Since F_κ is homogeneous in x for all $\kappa \in \mathbb{R}^3_{>0}$, it is clear that $\mathbb{V}_{>0}(F_\kappa)$ is a ruled surface, which in our language implies that F is $\mathcal{T}_A$-invariant over $\mathbb{R}_{>0}$ for $A = [1 \quad 1 \quad 1]$ and that $\#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0}) = \infty$. Hence F is not locally $\mathcal{T}_A$-toric.

Remark 3.12. We only focus on toricity of the zeros with nonzero or positive coordinates as the behavior on the coordinate hyperplanes might vary drastically. If such solutions are of interest, one may perform a systematic case-by-case analysis, where different combinations of variables are set to zero, and the resulting system is studied with the methods of this paper.

4. Characterization of toric invariance

We give now a characterization of $\mathcal{T}_A$-invariance of F in terms of $\ker(C)$. A basic sufficient condition for (3.1) to hold is that every polynomial in F is *quasihomogeneous* with weights given by $A \in \mathbb{Z}^{d\times n}$, in the sense that there exists $B \in \mathbb{Z}^{d\times s}$ such that

$$F(x \circ t^A) = t^B \circ F(x) \qquad \text{in } \Bbbk[\kappa][t^{\pm}, x^{\pm}]$$

[HL12, Appendix A]. This happens if and only if the rows of A are perpendicular to the affine hull of the Newton polytope of each of the s polynomials in F [GKZ94, Prop. 6.1.2(a)], i.e.,

$$AM_j = AM_{j_0} \quad \text{for all } j, j_0 \in \operatorname{supp}(C_{i*}) \text{ and all } i \in [s]\,. \tag{4.1}$$

This is only a sufficient condition for invariance, and the goal of this section is to partition the columns of M by an equivalence relation $\sim$ on $[m]$, depending on $\ker(C)$, such that the property $AM_j = AM_{j_0}$ whenever $j \sim j_0$ completely characterizes $\mathcal{T}_A$-invariance.

Definition 4.1. Given a matrix $C \in \mathbb{C}^{s\times m}$ with $\ker(C) \cap (\mathbb{C}^*)^m \neq \varnothing$, the ***fundamental equivalence relation*** on $[m]$ is generated by

$$j \sim j_0 \quad \text{if } j, j_0 \text{ both belong to a circuit of the matroid on the columns of } C\,.$$

The ***fundamental partition*** on $[m]$ is given by the equivalence classes of the fundamental equivalence relation. For a vertical system $F = C(\kappa \circ x^M)$, the fundamental partition and equivalence relation are by definition those of C.

Recalling that the ***elementary vectors*** of $\ker(C)$ are the nonzero vectors of $\ker(C)$ with minimal support [Roc69], the circuits of the matroid on the columns of C are by definition the supports of the elementary vectors of $\ker(C)$. Finding these circuits can be an expensive calculation. However, by the following lemma it is enough to find a basis for $\ker(C)$ consisting of elementary vectors, which can be done by applying Gaussian elimination to any basis for $\ker(C)$.

We write the fundamental partition as $[m] = \rho_1 \sqcup \cdots \sqcup \rho_\theta$ and for $v \in \mathbb{C}^m$, we let $v^{(i)} \in \mathbb{C}^m$, for all $i \in [\theta]$, be the unique vectors with $v = v^{(1)} + \cdots + v^{(\theta)}$ and $\operatorname{supp}(v^{(i)}) \subseteq \rho_i$.

Lemma 4.2. *Let $C \in \mathbb{C}^{s\times m}$ of rank s with $\ker(C) \cap (\mathbb{C}^*)^m \neq \varnothing$, let $[m] = \rho_1 \sqcup \cdots \sqcup \rho_\theta$ be the fundamental partition, and let $\{E_1, \ldots, E_{m-s}\}$ be a basis of elementary vectors for $\ker(C)$.*

(i) $v \in \ker(C)$ if and only if $v^{(i)} \in \ker(C)$ for all $i = 1, \ldots, \theta$.

(ii) The fundamental equivalence relation is generated by

$$j \sim j_0 \quad \text{if } j, j_0 \in \operatorname{supp}(E_i) \text{ for some } i \in [m-s]\,.$$

Proof. By definition, for all $j \in [m-s]$, $\operatorname{supp}(E_j)$ is included in one of the subsets of the fundamental partition. In particular, given $v = \sum_{j=1}^{m-s} \lambda_j E_j \in \ker(C)$, we have that

$$v^{(i)} = \sum_{\operatorname{supp}(E_j)\subseteq\rho_i} \lambda_j E_j. \tag{4.2}$$

From here we conclude that $v^{(i)} \in \ker(C)$. (i) now follows, as the reverse implication is trivial.

To show (ii), let $v \in \ker(C)$ be an elementary vector and let ρ_i be such that $\operatorname{supp}(v) \subseteq \rho_i$. Then by (4.2), $v = v^{(i)} = \sum_{\operatorname{supp}(E_j)\subseteq\rho_i} \lambda_j E_j$. If for an index $j_1 \in \operatorname{supp}(v)$ its equivalence class $\widetilde{j_1}$ by the equivalence relation in the statement does not include $\operatorname{supp}(v)$, then

$$v = \sum_{\operatorname{supp}(E_j)\subseteq\widetilde{j_1}} \lambda_j E_j + \sum_{\operatorname{supp}(E_j)\not\subseteq\widetilde{j_1}} \lambda_j E_j\,,$$

where the two summands are nonzero, have disjoint index sets, and define vectors in $\ker(C)$. This gives a contradiction as v has minimal support. Hence (ii) holds. □

The fundamental partition defined here agrees with the partition in [HADlC+22, Thm. 3.3] given in the context of reaction networks.

Remark 4.3. An immediate consequence of Lemma 4.2(i) is that $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ can be written as the intersection of the varieties $\mathbb{V}_{\mathbb{C}^*}(F_\kappa^{(i)})$ for $i = 1, \ldots, \theta$, where the vertical system $F^{(i)}$ is defined by considering only the columns of C and M with indices in the partition set ρ_i.

Given a vertical system F, we form the ***toric invariance group*** of all row vectors $a \in \mathbb{Z}^n$ for which there is invariance, i.e. we consider $\mathbb{Z}$-submodule of $\mathbb{Z}^n$

$$\mathcal{I} = \{a \in \mathbb{Z}^n : F \text{ is } \mathcal{T}_a\text{-invariant}\}\,. \tag{4.3}$$

We obtain the unique maximal-dimensional positive torus $\mathcal{T}_A$ for which the system is invariant by letting the rows of A form a $\mathbb{Z}$-basis for $\mathcal{I}$ (c.f. Remark 3.2).

Proposition 4.4. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and F a vertical system with defining matrices $C \in \Bbbk^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, and suppose that $\ker(C) \cap \mathbb{G}^m \neq \varnothing$. Let $A \in \mathbb{Z}^{d\times n}$. Then F is $\mathcal{T}_A$-invariant over $\mathbb{G}$ if and only if $\operatorname{row}_{\mathbb{Z}}(A) \subseteq \mathcal{I}$.*

Proof. This is a direct consequence of Theorem 3.5, Lemma 3.6, and the discussion above. □

Theorem 4.5. *Let F be a vertical system with defining matrices $C \in \mathbb{C}^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, and let $\sim$ be the fundamental equivalence relation. Suppose that $\ker(C) \cap (\mathbb{C}^*)^m \neq \varnothing$. Then for a row vector $a \in \mathbb{Z}^n$, the following are equivalent:*

(i) $a \in \mathcal{I}$.

(ii) Whenever $j \sim j_0$, it holds that $aM_j = aM_{j_0}$.

Proof. By (3.1), $a \in \mathcal{I}$ if and only if $(\kappa \circ x^M) \circ t^{aM} \in \ker(C)$ for all $t \in \mathbb{C}^*$ if $\kappa \circ x^M \in \ker(C)$. As (3.2) holds with equality for $\mathcal{K} = (\mathbb{C}^*)^m$ and $\mathbb{G} = \mathbb{C}^*$, (i) is equivalent to

$$\text{(i')} \quad C(v \circ t^{aM}) = 0 \quad \text{for all } v \in \ker(C) \cap (\mathbb{C}^*)^m \text{ and } \ t \in \mathbb{C}^* .$$

(ii) $\Rightarrow$ (i'): Let $[m] = \rho_1 \sqcup \cdots \sqcup \rho_\theta$ be the fundamental partition. By (ii), for each $i \in [\theta]$ there exists $\alpha_i \in \mathbb{C}$ such that $aM_j = \alpha_i$ for all $j \in \rho_i$. We write $v \in \ker(C)$ as $v = v^{(1)} + \cdots + v^{(\theta)}$ and use Lemma 4.2(i) to obtain

$$C(v \circ t^{aM}) = \sum_{i=1}^{\theta} C(v^{(i)} \circ t^{aM}) = \sum_{i=1}^{\theta} t^{\alpha_i} C v^{(i)} = 0.$$

(i') $\Rightarrow$ (ii): As $\ker(C) \cap (\mathbb{C}^*)^m \neq \varnothing$, $\ker(C) \cap (\mathbb{C}^*)^m$ is Zariski dense in $\ker(C)$ and hence (i') holds for all $v \in \ker(C)$. In particular, given an elementary vector $v \in \ker(C)$, $C(v \circ t^{aM}) = 0$ for all $t \in \mathbb{C}^*$. This can only happen if all coefficients of $C(v \circ t^{aM})$ are zero. Thus, for any $j_0 \in \operatorname{supp}(v)$, the coefficient of $t^{aM_{j_0}}$ must be zero:

$$\sum_{\substack{j \in \operatorname{supp}(v) \\ aM_j = aM_{j_0}}} C_j v_j = 0 .$$

We now obtain a nonzero vector $w \in \ker(C)$ by setting $w_j = v_j$ if $aM_j = aM_{j_0}$ and 0 otherwise. Since v has minimal support among nonzero vectors in $\ker(C)$, we conclude that $w = v$ and hence $aM_j = aM_{j_0}$ for all $j \in \operatorname{supp}(v)$ and (ii) follows. □

Let $[m] = \rho_1 \sqcup \cdots \sqcup \rho_\theta$ be the fundamental partition. Theorem 4.5 is saying that

$$\mathcal{I} = \bigcap_{i=1}^{\theta} \operatorname{span}_{\mathbb{Z}}\{M_j - M_{j_0} : j, j_0 \in \rho_i\}^{\perp}.$$

Geometrically, it means that F is $\mathcal{T}_A$-invariant if and only if the rows of A are perpendicular to each of the (possibly overlapping) polytopes generated by the columns of M with indices in ρ_i for $i \in [\theta]$. This is a weaker condition than the rows being perpendicular to each of the polytopes generated by the columns of M in each entry of F, which is the geometric interpretation of quasihomogeneity in (4.1).

Remark 4.6. If the rows of the coefficient matrix C have pairwise disjoint support, the corresponding vertical system can be interpreted as a freely parametrized system for some fixed supports $\mathcal{S}_1, \ldots, \mathcal{S}_s \subseteq \mathbb{Z}^n$ and freely varying coefficients (with fixed signs of some coefficients if $\mathbb{G} = \mathbb{R}_{>0}$), c.f. Remark 2.4. In this case, the sets of the fundamental partition are $\mathcal{S}_1, \ldots, \mathcal{S}_s$ the and toric invariance corresponds to quasihomogeneity of the system, i.e.

$$\mathcal{I} = \bigcap_{i=1}^{s} \operatorname{span}_{\mathbb{Z}} \{v - w : v, w \in \mathcal{S}_i\}^{\perp} .$$

Theorem 4.5, Proposition 4.4 and Lemma 3.6 lead to Algorithm 4.7 for finding a maximal rank matrix A for which F displays toric invariance over $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$.

Algorithm 4.7 (Toric invariance)**.**
Input: Matrix $C \in \Bbbk^{s \times m}$ of full rank s, matrix $M \in \mathbb{Z}^{n \times m}$
Output: Rank d matrix $A \in \mathbb{Z}^{d \times n}$ such that $F = C(\kappa \circ x^M)$ is $\mathcal{T}_A$-invariant over $\mathbb{G}$.

Find a basis $\{E_1, \ldots, E_{m-s}\}$ for $\ker(C)$ consisting of elementary vectors
if (Union of supports of $E_1, \ldots, E_{m-s}$) $\neq [m]$ **then**
    return "The varieties $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ are empty for all $\kappa \in \mathbb{G}^m$"
if $\mathbb{G} = \mathbb{R}_{>0}$ and (Interior of polyhedral cone $\ker(C) \cap \mathbb{R}^m_{\geq 0}$ is empty) **then**
    return "The varieties $\mathbb{V}_{>0}(F_\kappa)$ are empty for all $\kappa \in \mathbb{R}^m_{>0}$"
Find the fundamental partition $\{\rho_1, \ldots, \rho_\theta\}$ from the supports of $\{E_1, \ldots, E_{m-s}\}$
For each $i \in [\theta]$, pick $j_0 \in \rho_i$, and construct a matrix B with columns $M_j - M_{j_0}$ for all $j \in \rho_i \setminus \{j_0\}$
if $\mathbb{G} = \mathbb{R}_{>0}$ or $\mathbb{C}^*$ **then**
    Find a $\mathbb{Z}$-matrix A with rows a basis for $\ker_{\mathbb{Q}}(B^\top)$
else
    Find a $\mathbb{Z}$-matrix A with rows a basis for $\ker_{\mathbb{Z}}(B^\top)$
return A

Example 4.8. For the vertical system over $\mathbb{C}^*$ with matrices

$$C = \begin{bmatrix} -3 & 3 & -3 & 3 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 6 & 3 & 3 & 0 & 1 & 0 \\ 0 & 2 & 2 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix},$$

$\ker(C)$ admits a basis consisting of the four elementary vectors $(1,1,0,0,0,0)$, $(-1,0,1,0,0,0)$, $(1,0,0,1,0,0)$ and $(0,0,0,0,1,1)$, whose supports generate the fundamental partition $[m] = \{1,2,3,4\} \sqcup \{5,6\}$. The matrix B in Algorithm 4.7, after choosing the indices 1 and 5 in each set of the partition as j_0, is

$$B = \begin{bmatrix} -3 & -3 & -6 & -1 \\ 2 & 2 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix},$$

and a basis for the $\mathbb{Z}$-kernel of $B^\top$ consists of the vector $(10, 15, 2)$. Hence, F is $\mathcal{T}_a$-invariant for $a = (10, 15, 2)$, and there is no invariance for a higher dimensional torus.

Geometrically, $\mathcal{I}$ consists of the tuples $a \in \mathbb{Z}^3$ that are perpendicular to both the dashed orange and green line segments in Figure 4.1. In contrast, quasihomogeneity with weights in a matrix A corresponds to the stronger condition of the rows of A being perpendicular to the whole polytope. In this case there is only trivial quasihomogeneity with all weights zero.

Example 4.9. In Example 2.3, the elementary vectors v_1, v_2, v_3 give that the fundamental partition has a single block and toric invariance agrees with quasihomogeneity. We find that $\mathcal{I} = \operatorname{span}_{\mathbb{Z}}\{(1,1,-1,0,0), (0,0,0,1,0), (0,0,1,0,-1)\}^\perp$ and a $\mathbb{Z}$-basis is given by the rows of

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

We conclude that the 2-dimensional torus $\mathcal{T}_A^{>0}$ is the maximal-dimensional torus for which F is invariant over $\mathbb{R}_{>0}$. Note that we recover the same matrix A given in Example 3.8.

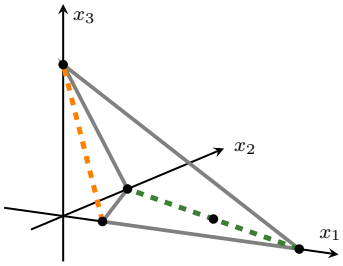


Figure 4.1

Example 4.10. Toric invariance for a nonzero matrix A does not imply that the varieties $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ are (generically) nonempty. To see this, consider the vertical system

$$F = \begin{bmatrix} \kappa_1 x_1^2 - \kappa_2 x_1 x_2 \\ \kappa_3 x_1^2 - \kappa_2 x_1 x_2 \end{bmatrix}.$$

It is straightforward to see that if $\kappa_1 \neq \kappa_3$, then $\mathbb{V}_{\mathbb{C}^*}(F_\kappa) = \varnothing$. The toric invariance group is $\mathcal{I} = \operatorname{span}_{\mathbb{Z}}\{(1,1)\}$, showing that when $\kappa_1 = \kappa_3$, the variety $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ is $\mathcal{T}_A$-invariant for $A = [1\ 1]$. Indeed, in this case the variety is defined by the equation $\kappa_1 x_1 - \kappa_2 x_2 = 0$.

Remark 4.11. The group $\mathcal{I}$ describes the largest-dimensional torus for which $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ is invariant for *all* $\kappa \in (\mathbb{C}^*)^m$, and by Theorem 3.5, it also describes the largest-dimensional $\mathbb{G}$-torus for which $\mathbb{V}_{\mathbb{G}^*}(F_\kappa)$ is invariant for all $\kappa \in \mathcal{K}$, as long as $\mathcal{K} \subseteq \mathbb{G}^m$ is a Euclidean open parameter set for which $\mathbb{V}_{\mathbb{G}^*}(F_\kappa)$ is nonempty for some κ. However, a smaller subfamily might display invariance under a larger-dimensional torus if the parameter space is a Euclidean closed set. For instance, with $\mathbb{G} = \mathbb{R}_{>0}$, we have that $\mathcal{I} = \{0\}$ for the vertical system

$$F = (\kappa_1 - \kappa_2) x_1 x_2 + \kappa_3 x_2^2 - \kappa_4 x_1^3 .$$

However, for $\mathcal{K} = \mathbb{V}(\kappa_1 - \kappa_2) \cap \mathbb{R}^4_{>0}$, it holds that $\mathbb{V}_{>0}(F_\kappa)$ is invariant for $A = [2 \quad 3]$.

Remark 4.12. In [MR24, Thm. 5], $\mathbb{V}_{>0}(F_\kappa)$ is expressed as the union of toric cosets (Z_c in loc. cit.). Adapted to our setting, the start point of that work is a decomposition of $\ker C \cap \mathbb{R}^m_{>0}$ as a direct product of cones. By choosing the decomposition given by the fundamental partition, their expression agrees, as expected, with ours. An analogous approach to parametrize $\mathbb{V}_{>0}(F_\kappa)$, but without the decomposition of the cone, can be found in [BBH24, §2.4].

5. Local toricity

At this point we have provided a simple algorithm to find a maximal-rank matrix A for which a vertical system F is $\mathcal{T}_A$-invariant, but this is not sufficient for toricity. In this section we show that generic local toricity can be completely characterized, and sufficient conditions for local toricity for all parameter values can be determined.

A useful observation is that if F is $\mathcal{T}_A$-invariant, then

$$\dim(\mathbb{V}_{\mathbb{C}^*}(F_\kappa)) \geq \dim(\mathcal{T}_A) = \operatorname{rk}(A)\,, \tag{5.1}$$

for all $\kappa \in \mathcal{Z}_{\mathbb{C}^*}$. If in addition $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ is locally $\mathcal{T}_A$-toric, then $\dim(\mathbb{V}_{\mathbb{C}^*}(F_\kappa)) = \operatorname{rk}(A)$.

Lemma 5.1. *Let F be a vertical system with defining matrices $C \in \mathbb{C}^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d. If F is generically consistent over $\mathbb{C}^*$, then $s + d \leq n$.*

Proof. This follows from (5.1), as $\dim(\mathbb{V}_{\mathbb{C}^*}(F_\kappa)) = n - s$ for some κ by Proposition 2.2. □

In Example 4.10, where F was generically inconsistent, we had $s + \operatorname{rk}(A) > n$, hence the inequality in Lemma 5.1 does not necessarily hold in general. The following useful lemma will be at the core of several results below.

Lemma 5.2. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$, and $F = C(\kappa \circ x^M)$ be a vertical system defined by $C \in \Bbbk^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, which is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d. Assume that for a fixed $\kappa \in \mathbb{G}^m$, $\mathbb{V}_{\mathbb{G}}(F_\kappa) \neq \varnothing$ and all irreducible components of $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ that intersect $\mathbb{G}^n$ contain a nondegenerate zero of F_κ in $\mathbb{G}^n$. Then $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is locally $\mathcal{T}_A$-toric over $\mathbb{G}$ if and only if $n = s + d$.*

Proof. As F is $\mathcal{T}_A$-invariant, then it is also $\mathcal{T}_A$-invariant over $\mathbb{G}$ by Theorem 3.5. Let $\bigcup_{i=1}^{\ell} Y_i$ be the union of the irreducible components of $\mathbb{V}_{\mathbb{C}^*}(F_\kappa)$ that intersect $\mathbb{G}^n$. The condition on nondegeneracy guarantees that $\dim(Y_i) = n - s$ for all i and that $\overline{\mathbb{V}_{\mathbb{G}}(F_\kappa)} = \bigcup_{i=1}^{\ell} Y_i$, where the overline denotes the Zariski closure in $(\mathbb{C}^*)^n$. For a coset $\alpha \circ \mathcal{T}_A^{\mathbb{G}}$ with $\alpha \in \mathbb{V}_{\mathbb{G}}(F_\kappa)$, we have

$\alpha \circ \mathcal{T}_A$ must be contained in Y_i for some $i \in [\ell]$ by irreducibility. As $\dim(\alpha \circ \mathcal{T}_A) = d$, if $d = n - s$, then $\alpha \circ \mathcal{T}_A = Y_i$ and local toricity follows. Conversely, if $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ is locally $\mathcal{T}_A$-toric over $\mathbb{G}$, then $\mathbb{V}_{\mathbb{G}}(F_\kappa) = \bigsqcup_{i=1}^{p} \alpha_i \circ \mathcal{T}_A^{\mathbb{G}}$ for some $\alpha_i \in \mathbb{V}_{\mathbb{G}}(F_\kappa)$, $i \in [p]$. Hence,

$$n - s = \dim(\bigcup_{i=1}^{\ell} Y_i) = \dim\left(\overline{\bigsqcup_{i=1}^{p} \alpha_i \circ \mathcal{T}_A^{\mathbb{G}}}\right) = \dim\left(\bigsqcup_{i=1}^{p} \alpha_i \circ \mathcal{T}_A\right) = d\,. \qquad \square$$

Theorem 5.3. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and F a vertical system with defining matrices $C \in \Bbbk^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, such that $\ker(C) \cap \mathbb{G}^m \neq \varnothing$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d such that $s + d \leq n$. The following statements are equivalent:*

(i) The augmented vertical system $(C(\kappa \circ x^M), Ax - b)$ is generically consistent over $\mathbb{C}^*$.

(ii) F is generically consistent over $\mathbb{C}^*$ and $n = s + d$.

(iii) F is generically locally $\mathcal{T}_A$-toric over $\mathbb{G}$.

(iv) F is generically locally $\mathcal{T}_A$-toric over $\mathbb{C}^*$.

Proof. (i) $\Rightarrow$ (ii): Clear as (i) implies (2.5), which in turn implies (2.6) and hence (ii).

(ii) $\Rightarrow$ (iii): As $\ker(C) \cap \mathbb{G}^m \neq \varnothing$, Proposition 2.2 gives that F is generically consistent over $\mathbb{G}$. By Proposition 2.2(i), there exists a nonempty Zariski open subset $\mathcal{U} \subseteq \mathcal{Z}_{\mathbb{G}}$ such that for all $\kappa \in \mathcal{U}$, all zeros of F_κ in $\mathbb{G}^n$ are nondegenerate. Hence (iii) follows from Lemma 5.2.

(iii) $\Rightarrow$ (iv): If F is generically locally $\mathcal{T}_A$-toric over $\mathbb{G}$, then it is generically consistent over $\mathbb{G}$ (and hence over $\mathbb{C}^*$) by definition, and $n = s + d$ by Lemma 5.2 and Proposition 2.2(i), which also give that F is generically locally $\mathcal{T}_A$-toric over $\mathbb{C}^*$.

(iv) $\Rightarrow$ (i): By hypothesis, there is a nonempty Zariski open subset $\mathcal{U} \subseteq (\mathbb{C}^*)^m$ such that for all $\kappa \in \mathcal{U}$, we have $0 < \#(\mathbb{V}_{\mathbb{C}^*}(F_\kappa)/\mathcal{T}_A) \leqslant \ell < \infty$, and in particular, $n - s = \dim(\mathbb{V}_{\mathbb{C}^*}(F_\kappa)) = d$.

Let $\mathcal{Z}(H) \subseteq (\mathbb{C}^*)^m \times \mathbb{C}^d$ be the subset of parameters (κ, b) for which the augmented vertical system $H = (C(\kappa \circ x^M), Ax - b)$ has zeros in $(\mathbb{C}^*)^n$, and let $\mathcal{C} \subseteq (\mathbb{C}^*)^n \times \mathbb{C}^d$ be the subset of parameters (α, b) for which $\alpha \circ \mathcal{T}_A$ and $\mathbb{V}_{\mathbb{C}^*}(Ax - b)$ have nonempty intersection. We show below that $\mathcal{C}$ is Zariski dense in $(\mathbb{C}^*)^n \times \mathbb{C}^d$. Consider the map

$$\psi\colon (\ker(C) \cap (\mathbb{C}^*)^m) \times (\mathbb{C}^*)^n \times \mathbb{C}^d \to (\mathbb{C}^*)^m \times \mathbb{C}^d, \qquad (w, \alpha, b) \mapsto (w \circ (\alpha^{-1})^M, b)\,.$$

For $(\kappa, b) \in \mathcal{Z}(H)$, as the system H has at least one zero $\alpha \in (\mathbb{C}^*)^n$, it follows that $\psi(\kappa \circ \alpha^M, \alpha, b) = (\kappa, b)$. Hence $\mathcal{Z}(H) \subseteq \mathrm{im}(\psi)$.

For any $(w, \alpha, b) \in (\ker(C) \cap (\mathbb{C}^*)^m) \times \mathcal{C}$, there is at least one point $x^* \in (\mathbb{C}^*)^n$ for which $Ax^* = b$ and $x^* \in \alpha \circ \mathcal{T}_A$. Setting $\kappa = w \circ (\alpha^{-1})^M$, we have $F_\kappa(\alpha) = 0$. As F is $\mathcal{T}_A$-invariant we also have $F_\kappa(x^*) = 0$. Hence $H_{\psi(w,\alpha,b)}$ has at least the zero x^* and thus $\psi(w, \alpha, b) \in \mathcal{Z}(H)$.

This shows that $\psi((\ker(C) \cap (\mathbb{C}^*)^m) \times \mathcal{C}) = \mathcal{Z}(H)$. Taking Zariski closures gives that $\overline{\mathrm{im}(\psi)} = \overline{\mathcal{Z}(H)}$. We now apply the theorem of dimension of fibers to conclude that for all (κ, b) in a nonempty Zariski open subset $\mathcal{U}' \subseteq \overline{\mathcal{Z}(H)}$, it holds that

$$\dim(\psi^{-1}(\kappa, b)) = m + 2d - \dim(\overline{\mathcal{Z}(H)}).$$

(Here, we use that $\ker(C) \cap (\mathbb{C}^*)^m$ has dimension $m - s = m - n + d$.) We now argue that $\dim(\psi^{-1}(\kappa, b)) = d$ for generic $(\kappa, b) \in \mathcal{Z}(H)$. To see this, we note that $\kappa = w \circ (\alpha^{-1})^M$ for some $w \in \ker(C) \cap (\mathbb{C}^*)^m$ and $\alpha \in (\mathbb{C}^*)^n$ if and only if α is a zero of F_κ. Hence,

$$\begin{aligned}\dim(\psi^{-1}(\kappa, b)) &= \dim(\{(w, \alpha) \in (\ker(C) \cap (\mathbb{C}^*)^m) \times (\mathbb{C}^*)^n : w \circ (\alpha^{-1})^M = \kappa\}) \\ &= \dim(\{(\kappa \circ \alpha^M, \alpha) \in (\ker(C) \cap (\mathbb{C}^*)^m) \times (\mathbb{C}^*)^n : F_\kappa(\alpha) = 0\}) = d\end{aligned}$$

for all $\kappa \in \mathcal{U}$ and $b \in \mathbb{C}^d$. Hence the fiber has dimension d for all $(\kappa, b) \in (\mathcal{U} \times \mathbb{C}^d) \cap \mathcal{Z}(H)$, hence generically in $\mathcal{Z}(H)$ (the intersection is nonempty as the projection of $\mathcal{Z}(H)$ onto $(\mathbb{C}^*)^m$ contains $\mathcal{U}$ by definition).

For any $(\kappa, b) \in (\mathcal{U} \times \mathbb{C}^d) \cap \mathcal{U}' \cap \mathcal{Z}(H) \neq \varnothing$, we have shown that

$$d = \dim(\psi^{-1}(\kappa, b)) = m + 2d - \dim(\overline{\mathcal{Z}(H)}) \quad \Rightarrow \quad \dim(\overline{\mathcal{Z}(H)}) = m + d.$$

This means that $\mathcal{Z}(H)$ is Zariski dense in $(\mathbb{C}^*)^m \times \mathbb{C}^d$. Thus H is generically consistent.

All that is left to complete the proof is to show that $\mathcal{C}$ is Zariski dense in $(\mathbb{C}^*)^n \times \mathbb{C}^d$. Note that $\mathcal{C}$ is the set of parameters (α, b) for which the vertical system $G = A(\alpha \circ t^A) - b$ in the variables $t = (t_1, \ldots, t_d)$ has zeros in $\mathbb{G}$. The defining matrices are the full row rank matrices

$$C_G = [\ A \mid -\mathrm{id}_d\ ] \in \mathbb{Z}^{d\times(n+d)}, \qquad M_G = [\ A \mid 0_d\ ] \in \mathbb{Z}^{d\times(n+d)}.$$

Clearly, $\ker(C_G) \cap (\mathbb{C}^*)^{n+d} \neq \varnothing$. By Proposition 2.2 and (2.6), $\mathcal{C}$ is Zariski dense if and only if

$$\mathrm{rk}(C_G \,\mathrm{diag}(w) M_G^\top) = d \quad \text{for some } w \in \ker(C_G).$$

An easy computation shows that

$$C_G \,\mathrm{diag}(w) M_G^\top = A \,\mathrm{diag}(w') A^\top$$

where $w' = (w_1, \ldots, w_n)$, if $(w_1, \ldots, w_{n+d}) \in \ker(C_G)$. By taking $w' = (1, \ldots, 1)$ and extending it to a vector $w \in \ker(C_G)$, we have

$$\mathrm{rk}(C_G \,\mathrm{diag}(w) M_G^\top) = \mathrm{rk}(AA^\top) = \mathrm{rk}(A) = d,$$

and hence $\mathcal{C}$ is Zariski dense. □

Example 5.4. For the vertical system F in Example 3.11, Algorithm 4.7 gives that the maximal-dimensional torus for which we have invariance is defined by $A = [1 \quad 1 \quad 1]$. As $s + d = 2 < 3 = n$, F is not generically locally $\mathcal{T}_A$-toric. As F is generically consistent, $\mathbb{V}_{>0}(F_\kappa)$ is a union of infinitely many lines whenever not empty.

Theorem 5.3 completely characterizes generic local toricity, once a maximal-rank matrix A for which F is invariant has been found. Conditions (i) and (ii) illustrate that generic local toricity does not rely on $\mathbb{G}$, as long as $\ker(C) \cap \mathbb{G}^m \neq \varnothing$. In Proposition 5.5 below, which gives a sufficient condition for local toricity, the choice of $\mathbb{G}$ might be relevant.

Proposition 5.5. *Let $\mathbb{G} \in \{\mathbb{R}_{>0}, \mathbb{R}^*, \mathbb{C}^*\}$ and F a vertical system with defining matrices $C \in \Bbbk^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, such that $\ker(C) \cap \mathbb{G}^m \neq \varnothing$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d with $n = s + d$. If*

$$\mathrm{rk}(C \,\mathrm{diag}(w) M^\top) = s \quad \textit{for all } w \in \ker(C) \cap \mathbb{G}^m, \tag{5.2}$$

then F is locally $\mathcal{T}_A$-toric over $\mathbb{G}$ and $\dim(\mathbb{V}_{\mathbb{G}}(F_\kappa)) = n - s$ for all $\kappa \in \mathbb{G}^m$.

Proof. By the last statement in Proposition 2.2(i), all zeros of F_κ in $\mathbb{G}^n$ are nondegenerate for all $\kappa \in \mathbb{G}^m$, so he statement follows from Lemma 5.2. □

It might seem that condition (5.2) in Proposition 5.5 is very strict, but it applies to surprisingly many realistic reaction networks, as we will see in Section 8. In that application, we have $\mathbb{G} = \mathbb{R}_{>0}$ and (5.2) can be checked using the parametrization of $\ker(C) \cap \mathbb{R}^m_{>0}$ given by the generators of the polyhedral cone $\ker(C) \cap \mathbb{R}^m_{\geq 0}$, as the next example illustrates.

Example 5.6. For Example 2.3 (with $n = 5$ and $s = 3$), we have

$$\ker(C) \cap \mathbb{R}^6_{>0} = \{\lambda_1(1,0,1,1,0,1) + \lambda_2(0,0,0,1,1,0) + \lambda_3(1,1,0,0,0,0) : \lambda \in \mathbb{R}^3_{>0}\}.$$

From this we obtain

$$C \,\mathrm{diag}(E\lambda) M^\top = \begin{bmatrix} \lambda_1+\lambda_3 & \lambda_1+\lambda_3 & -\lambda_3 & 0 & -\lambda_1 \\ 0 & 0 & \lambda_1 & 0 & -\lambda_1 \\ 0 & 0 & \lambda_1+\lambda_2 & \lambda_1+\lambda_2 & -\lambda_1-\lambda_2 \end{bmatrix},$$

which for instance has the 3×3 minor $(\lambda_1 + \lambda_3)\lambda_1(\lambda_1 + \lambda_2)$ given by columns $1, 3, 4$. Therefore, (5.2) holds and we conclude that F is locally $\mathcal{T}_A$-toric with $\dim(\mathbb{V}_{>0}(F_\kappa)) = 2$ for all $\kappa \in \mathbb{R}^6_{>0}$. For $\mathbb{G} = \mathbb{R}^*$, letting $\lambda_1 = -\lambda_2$, (5.2) fails, and hence Proposition 5.5 is not informative. We still have local toricity generically, as we have that over $\mathbb{R}_{>0}$ (c.f. Theorem 5.3).

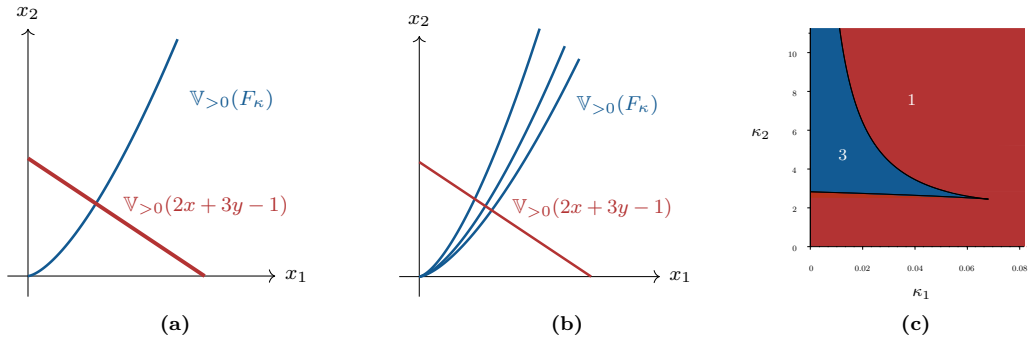


Figure 6.1. (a)-(b) Positive zero locus $\mathbb{V}_{>0}(F_\kappa)$ for the system in Example 3.10 and geometric interpretation of the coset-counting system, for different parameter values: (a) $\kappa = (0.01, 3, 1, 1)$ and (b) $\kappa = (0.01, 1, 1, 1)$. (c) A cylindrical algebraic decomposition for the coset counting system, for $\kappa_3 = \kappa_4 = 1$.

The results of this section yield a procedure to detect (generic) local toricity, when Algorithm 4.7 returns a matrix $A \in \mathbb{Z}^{d\times n}$. Namely if $s+d < n$, then we readily conclude that F is not generically locally $\mathcal{T}_A$-toric over $\mathbb{G}$. Otherwise, all we need is to check generic consistency, which can be verified by computing $r := \mathrm{rk}(C\,\mathrm{diag}(w)M^\top)$ for a randomly generated $w \in \ker(C)$ and deciding whether $r = s$. If this is not the case, then one should verify symbolically that the rank r is smaller for all w. Finally, local toricity can be certified if $\mathrm{rk}(C\,\mathrm{diag}(w)M^\top) = s$ for all $w \in \ker(C) \cap \mathbb{G}^m$. These steps are incorporated in Algorithm 7.6 below, given for $\mathbb{G} = \mathbb{R}_{>0}$.

6. Counting the number of cosets in $\mathbb{R}^n_{>0}$

After one has verified that F is locally $\mathcal{T}_A$-toric, the next question is to decide how many cosets there are. In the setting when $\mathbb{G} = \mathbb{R}^*$ or $\mathbb{C}^*$, the number of cosets can sometimes be found by counting the number of intersection points between $\mathbb{V}_{\mathbb{G}}(F_\kappa)$ and a certain toric variety, as discussed in [HL12, §4.2]. We focus now instead on the case $\mathbb{G} = \mathbb{R}_{>0}$, so $F \in \mathbb{R}[\kappa, x^{\pm}]^s$, where it turns out that one can always find the number of cosets by counting the number of intersections with a linear variety. To see this, consider the augmented vertical system

$$H = (C(\kappa \circ x^M), Ax - b) \in \mathbb{R}[\kappa, b, x^{\pm}]^{s+d}. \tag{6.1}$$

By Theorem 5.3, H is generically consistent over $\mathbb{R}_{>0}$ if and only if F is generically locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$ (in which case also $n = s + d$). But even more is true: the zeros of the system H count the number of $\mathcal{T}_A$-cosets. We call (6.1) the ***coset counting system***.

Proposition 6.1. *Let F be a vertical system with defining matrices $C \in \mathbb{R}^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, such that $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ and let H be the coset counting system (6.1). Then, for a given $\kappa \in \mathbb{R}^m_{>0}$ and any $b \in A(\mathbb{R}^n_{>0})$, there is a bijection of sets*

$$\mathbb{V}_{>0}(H_{\kappa,b}) \to \mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0}\,, \quad x \mapsto x \circ \mathcal{T}_A^{>0}\,.$$

Proof. Note that F is $\mathcal{T}_A$-invariant over $\mathbb{R}_{>0}$ by Theorem 3.5. The statement follows from the classical result (in some settings known as *Birch's theorem*) that for any $x, x^* \in \mathbb{R}^n_{>0}$, the coset $x \circ \mathcal{T}_A^{>0}$ intersects the translated subspace $x^* + \ker(A)$ exactly once; see, e.g., [Fei95, Prop. 5.1 and B.1] and [Bor12, Lem. 3.15] for a proof. □

Example 6.2. We noticed that the vertical system F in Example 3.10 is $\mathcal{T}_A$-invariant over $\mathbb{R}_{>0}$ for $A = [2 \quad 3]$ and that there are finitely many cosets. The exact number of cosets is found counting the number of points in $\mathbb{V}_{>0}(F_\kappa) \cap \mathbb{V}(2x + 3y - b)$ for any $b > 0$, see Figure 6.1(a-b).

If F is generically locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$, then the coset counting system has generically a finite number of zeros and hence the generic cardinality of the number of cosets is bounded above by the mixed volume by Bernstein's theorem [Ber75] (see also [CLO05, §7.5], as well as [GH21] where Bernstein's theorem is applied to a similar system). Sharper bounds that take into account the dependencies among the parametric coefficients arise from Newton–Okunkov bodies [OW24] and tropical methods [HR22, HHR24]. This type of bounds can be quite far from the number of positive solutions, as they count solutions over $\mathbb{C}^*$, but if the bound is 1, then we conclude that F is generically $\mathcal{T}_A$-toric.

We discuss next two approaches to assert toricity, that is, to confirm that the coset counting system has at most one solution for all κ.

Injectivity. The study of chemical reaction networks has focused to a great extent on multistationarity, which is equivalent to the system (6.1) having two or more positive zeros for a given κ and b, for a specific matrix A derived from the reaction network. Many of the techniques used to study this type of systems in reaction network theory work also for a more general matrix A, and hence can be used to give bounds on $\#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0})$.

One of the simplest methods arising from this theory is *injectivity with respect to a linear subspace* S, which by definition means that the function $F_\kappa(x) = C(\kappa \circ x^M)$ is injective on the positive part of cosets of S (see, e.g., [WF13] and [MFR+15]). This gives rise to the following sufficient computational criterion for the number of $\mathcal{T}_A^{>0}$-cosets to be one.

Theorem 6.3. *Let F be a vertical system with defining matrices $C \in \mathbb{R}^{s\times m}$ be of rank s and $M \in \mathbb{R}^{n\times m}$ with $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{(n-s)\times n}$ of rank $n-s$. For $\mu = (\mu_1, \ldots, \mu_m)$ and $\alpha = (\alpha_1, \ldots, \alpha_n)$, form the matrix*

$$\mathcal{L}_{\mu,\alpha} := \begin{bmatrix} C \operatorname{diag}(\mu) M^\top \operatorname{diag}(\alpha) \\ A \end{bmatrix}.$$

If $\det(\mathcal{L}_{\mu,\alpha})$ is a nonzero polynomial in $\mathbb{R}[\mu, \alpha]$, with all nonzero coefficients having the same sign, then F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$.

Proof. The implication (i) $\Rightarrow$ (ii) of Theorem 5.3 together with Proposition 2.2 give that F is generically consistent over $\mathbb{R}_{>0}$. By [MFR+15, Thm. 2.13], the assumption on $\det(\mathcal{L}_{\mu,\alpha})$ is equivalent to the polynomial map F_κ being injective on $(x^* + \ker(A)) \cap \mathbb{R}^n_{>0}$ for all $x^* \in \mathbb{R}^n_{>0}$ and $\kappa \in \mathbb{R}^m_{>0}$. This implies that for the coset counting system H in (6.1), $\#\mathbb{V}_{>0}(H_{\kappa,b}) \leq 1$ for all κ and b, and the statement follows from Proposition 6.1. □

We note that in [MFR+15] the determinant criterion in Theorem 6.3 is also phrased in terms of sign vectors. A concrete scenario is shown later in Proposition 7.1.

Example 6.4. The vertical system F in Example 2.3 satisfies the conditions in Theorem 6.3 for the matrix A that we found in Example 4.9, since the corresponding polynomial is

$$\begin{aligned}\det(\mathcal{L}_{\mu,\alpha}) = &- \alpha_1\alpha_3\alpha_4\mu_1\mu_3\mu_4 - \alpha_1\alpha_4\alpha_5\mu_1\mu_4\mu_6 - \alpha_2\alpha_3\alpha_4\mu_1\mu_3\mu_4 \\ &- \alpha_2\alpha_4\alpha_5\mu_1\mu_4\mu_6 - \alpha_3\alpha_4\alpha_5\mu_2\mu_4\mu_6 - \alpha_3\alpha_4\alpha_5\mu_3\mu_4\mu_6\,.\end{aligned}$$

As all coefficients of this polynomial have the same sign, F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$.

A numerical algebraic geometry approach. In practice, it often turns out that $\#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0})$ is constant with respect to $\kappa \in \mathbb{R}^m_{>0}$. When this is the case, $\#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0})$ can be determined by solving the coset counting system (6.1) for any fixed $\kappa \in \mathbb{R}^m_{>0}$, and count the number of positive solutions. This can be done with numerical algebraic geometry combined with certification of positivity using interval arithmetic, e.g. using the Julia package `HomotopyContinuation.jl` [BT18, BRT23], provided that one can guarantee that the algorithm has not missed any solutions (e.g. by comparing the number of solutions over $\mathbb{C}^*$ with the mixed volume or specific bounds for vertical systems [HR22, HHR24].)

The following is a sufficient criterion for $\#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0})$ to be constant for all $\kappa \in \mathbb{R}^m_{>0}$.

Proposition 6.5. *Let F be a vertical system with defining matrices $C \in \mathbb{R}^{s\times m}$ of rank s and $M \in \mathbb{Z}_{\geq 0}^{n\times m}$ with $\ker(C)\cap\mathbb{R}^m_{>0} \neq \varnothing$, and suppose that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{(n-s)\times n}$ of rank $n-s$. For $b \in A(\mathbb{R}^n_{>0})$ fixed, suppose the following hold:*

(i) $\operatorname{row}(A) \cap \mathbb{R}^n_{>0} \neq \varnothing$.

(ii) $\mathbb{V}_{\mathbb{R}}(F_\kappa) \cap \mathbb{V}_{\mathbb{R}}(Ax-b) \cap (\mathbb{R}^n_{\geq 0} \setminus \mathbb{R}^n_{>0}) = \varnothing$ *for all* $\kappa \in \mathbb{R}^m_{>0}$.

(iii) $\operatorname{rk}\left(\begin{bmatrix} C\operatorname{diag}(w)M^\top \operatorname{diag}(h) \\ A \end{bmatrix}\right) = n$ *for all* $w \in \ker(C)\cap\mathbb{R}^n_{>0}$ *and* $h \in \mathbb{R}^n_{>0}$.

Then $\#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0})$ is constant with respect to $\kappa \in \mathbb{R}^m_{>0}$, and in particular $\mathcal{Z}_{>0} = \mathbb{R}^m_{>0}$.

Proof. By Proposition 2.2(i), condition (iii) gives that the coset counting system H in (6.1) has a finite number of zeros in $\mathbb{R}^n_{>0}$ for all $\kappa \in \mathbb{R}^m_{>0}$ and the fixed b, and these are all nondegenerate. Statement (i) ensures that $\mathcal{P} := \mathbb{V}_{\mathbb{R}}(Ax-b)\cap\mathbb{R}^n_{\geq 0}$ is compact [BI64]. Let $B \in \mathbb{R}^{n\times s}$ be a matrix of rank s whose columns form a basis of $\ker(A)$. For $\kappa \in \mathbb{R}^m_{>0}$, consider the map

$$\varphi\colon \mathcal{P} \to \mathcal{P}, \qquad x \mapsto x + BC(\kappa \circ x^M)\,.$$

This map is well defined as $\varphi(x) \in \mathbb{R}^n_{\geq 0}$ if $x \in \mathbb{R}^n_{\geq 0}$ and further $A(x + BC(\kappa \circ x^M)) = Ax = b$. From Brouwer's fixed point theorem we conclude that there exist $x^* \in \mathcal{P}$ such that $x^* = \varphi(x^*)$ and hence $x^* \in \mathbb{V}_{\mathbb{R}}(F_\kappa)\cap\mathcal{P}$ (as B has full rank). Using (ii), it must be the case that $x^* \in \mathbb{R}^n_{>0}$. Hence, for every $\kappa \in \mathbb{R}^m_{>0}$, the coset counting system $H_{\kappa,b}$ has positive zeros.

We now form the (positive) incidence variety

$$\mathcal{E} := \{(\kappa, x) \in \mathbb{R}^m_{>0} \times \mathbb{R}^n_{>0} : H_{\kappa,b}(x) = 0\}\,,$$

and consider the projection to parameter space $\pi\colon \mathcal{E} \to \mathbb{R}^m_{>0}$. By the above, π is surjective and has finite fibers. As all zeros of $H_{\kappa,b}$ in $\mathbb{R}^n_{>0}$ are nondegenerate, [FHPE24, Prop. 3.3] gives further that π has no critical points.

Our goal is to show that $\#\pi^{-1}(\kappa)$ is constant for all κ. As $\mathbb{R}^m_{>0}$ is connected, it suffices to show that the cardinality is locally constant. Given $\kappa^* \in \mathbb{R}^m_{>0}$, write $\pi^{-1}(\kappa^*) = \{(\kappa^*, x_1), \ldots, (\kappa^*, x_\ell)\}$ for $x_1, \ldots, x_\ell \in \mathbb{V}_{>0}(F_{\kappa^*})$. The absence of critical points for π allows us to find, for each $i \in [\ell]$, an open neighborhood $U_i \subseteq \mathcal{E}$ of (κ^*, x_i) and an open neighborhood V_i of κ^* such that $\pi_{|U_i}\colon U_i \to V_i$ is a homeomorphism. These open sets can be chosen such that $U_i \cap U_j = \varnothing$ for $i \neq j$. Furthermore, by letting $V = \cap_{i=1}^\ell V_i$ and replacing U_i by $(\pi_{|U_i})^{-1}(V)$, $\pi(U_i) = V$ for all $i \in [\ell]$. Now, form the closed subset $Q = \mathcal{E} \setminus (U_1 \cup \cdots \cup U_\ell) \subseteq \mathcal{E}$, which is closed also in $\mathbb{R}^m_{>0} \times \mathcal{P}$. Since $\mathcal{P}$ is compact, the projection $\mathbb{R}^m_{>0} \times \mathcal{P} \to \mathbb{R}^m_{>0}$ is a closed map, and hence $\pi(Q)$ is closed in $\mathbb{R}^m_{>0}$. We now form the open neighborhood $W := V \setminus \pi(Q) \subseteq \mathbb{R}^m_{>0}$ of κ^*, which has the property that $\#\pi^{-1}(\kappa) = \ell$ for all $\kappa \in W$. □

Condition (i) in Proposition 6.5 can be verified by deciding whether the polyhedral cone $\operatorname{row}(A)\cap\mathbb{R}^n_{\geq 0}$ has nonempty interior. Condition (ii) can efficiently be checked with SAT-SMT solvers [BFT17]. Alternatively, a sufficient conditions for (ii) can be obtained from the theory of siphons developed in chemical reaction network theory [ADLS07, SS10].

Example 6.6. We show now that F in (2.9) is toric over $\mathbb{R}_{>0}$. The defining matrices are

$$C = \begin{bmatrix} 1 & -1 & 1 & -2 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 3 & 3 & 0 & 6 \\ 2 & 2 & 4 & 0 \end{bmatrix},$$

and Algorithm 4.7 returns the maximal-rank matrix $A = \begin{bmatrix} 2 & 3 \end{bmatrix}$ for which F is $\mathcal{T}_A$-invariant.

It is immediate to see that Proposition 6.5(i) and (ii) hold. For (iii), we find the generators of the polyhedral cone $\ker(C)\cap\mathbb{R}^4_{\geq 0}$ and obtain the parametrization

$$\ker(C)\cap\mathbb{R}^4_{>0} = \{\lambda_1(2,0,0,1) + \lambda_2(1,1,0,0) + \lambda_3(0,0,2,1) + \lambda_4(0,1,1,0) : \lambda \in \mathbb{R}^4_{>0}\}.$$

This allows us to parametrize the set of matrices in Proposition 6.5(iii) with $\lambda \in \mathbb{R}^4_{>0}$ and their determinant becomes

$$\det \begin{bmatrix} (-6\lambda_1 - 3\lambda_4 - 12\lambda_3)\, h_1 & (4\lambda_1 + 2\lambda_4 + 8\lambda_3)\, h_2 \\ 2 & 3 \end{bmatrix} = -(9h_1 + 4h_2)(2\lambda_1 + 4\lambda_3 + \lambda_4).$$

As this polynomial does not vanish for any $\lambda \in \mathbb{R}^4_{>0}$ and $h \in \mathbb{R}^2_{>0}$, Proposition 6.5(iii) holds. Proposition 6.5 tells us that it is enough to find the positive zeros of the coset counting system

$$H = ((\kappa_1 - \kappa_2)x_1^3 x_2^2 + \kappa_3 x_2^4 - 2\kappa_4 x_1^6, 2x_1 + 3x_2 - 5) \tag{6.2}$$

for any choice of $\kappa \in \mathbb{R}^4_{>0}$. Using `HomotopyContinuation.jl` with $\kappa = (1,1,1,1)$, we verify that H has a unique zero in $\mathbb{R}^2_{>0}$, and hence F is $\mathcal{T}_A$-toric (and $\mathcal{Z}_{>0} = \mathbb{R}^4_{>0}$). In particular, we conclude that for each $\kappa \in \mathbb{R}^4_{>0}$, the positive zero locus admits the monomial parametrization

$$\mathbb{R}_{>0} \to \mathbb{V}_{>0}(F_\kappa)\,, \quad t \mapsto (\alpha_1 t^2, \alpha_2 t^3)\,,$$

where $\alpha = (\alpha_1, \alpha_2)$ is the unique positive zero of (6.2). We note that for this example, the ideal $\langle F_\kappa \rangle \subseteq \mathbb{R}[x]$ is not binomial whenever $\kappa_1 \neq \kappa_2$. We also point out that F is not toric over $\mathbb{R}^*$, since $\mathbb{V}_{\mathbb{R}^*}(F_\kappa)$ has two irreducible components for, e.g., $\kappa = (1,1,1,1)$.

7. Applications

7.1. **Multiple positive solutions of linear sections of vertical systems.** As we pointed out in the introduction, a motivation to study toricity arises from the determination of multistationarity in reaction networks. In our language, this translates into determining whether an augmented vertical system has multiple positive zeros for some choice of parameter values. When the vertical part of the augmented vertical system is toric, then the theory of reaction networks has provided a criterion to answer this problem. We expand on this application here.

The following proposition is a consequence and extension of [MDSC12, §5] and [MFR+15, §3] (see also [SF19b, §2]), where we also allow multiple cosets. For a set $P \subseteq \mathbb{R}^n$, we let $\operatorname{sign}(P)$ denote the set of tuples in $\{0,+,-\}^n$ obtained by taking the sign of all elements of P component-wise.

Proposition 7.1. *Let F be a vertical system with defining matrices $C \in \mathbb{R}^{s\times m}$ and $M \in \mathbb{Z}^{n\times m}$ such that $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d. Let $B \in \mathbb{R}^{(n-d)\times n}$ be a matrix whose columns form a basis for $\ker(A)$ and let $L \in \mathbb{R}^{(n-s)\times n}$. For $\alpha = (\alpha_1, \ldots, \alpha_{n-d})$ form the matrix*

$$\Gamma_\alpha := \begin{bmatrix} B^\top \operatorname{diag}(\alpha) \\ L \end{bmatrix}.$$

Then for the following statements it holds (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (iv):

- *(i)* $\ker(\Gamma_\alpha) \neq \{0\}$ *for some* $\alpha \in \mathbb{R}^n_{>0}$.
- *(ii)* $\operatorname{sign}(\ker(B^\top)) \cap \operatorname{sign}(\operatorname{im}(L)^\perp) \neq \{0\}$.
- *(iii) The map $x \mapsto x^B$ is not injective on some coset $(x^* + \operatorname{im}(L)^\perp) \cap \mathbb{R}^n_{>0}$ for $x^* \in \mathbb{R}^n_{>0}$.*
- *(iv) The augmented vertical system $(C(\kappa \circ x^M), Lx - b)$ has at least two zeros in $\mathbb{R}^n_{>0}$ for some $\kappa \in \mathbb{R}^m_{>0}$ and $b \in \mathbb{R}^{n-s}$.*

Furthermore, if F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$, then also (iv) $\Rightarrow$ (iii).

Proof. The equivalence of (i), (ii) and (iii) follows from (the proof of) Theorem 2.13, Proposition 3.9 and Corollary 3.11 in [MFR+15]. Note that in loc. cit. x^B refers to the monomial map with exponents given by the *rows* of B.

The implication from (iv) to (iii) is straightforward under the assumption of $\mathcal{T}_A$-toricity. To show the implication from (iii) to (iv), under the assumption of $\mathcal{T}_A$-invariance over $\mathbb{R}_{>0}$, let $x, y \in \mathbb{R}^n_{>0}$ with $x \neq y$, $x - y \in \mathrm{im}(L)^\perp$ and $x^B = y^B$. The latter together with Remark 3.2 gives that $y \in x \circ \mathcal{T}_A$. Choose $v \in \ker(C) \cap \mathbb{R}^m_{>0}$ and let κ such that $v = \kappa \circ x^B$. Then $F_\kappa(x) = 0$, and the $\mathcal{T}_A$-invariance gives $F_\kappa(y) = 0$. By letting $b := Lx = Ly$, x and y are zeros the system in (iv) and (iv) holds. □

Remark 7.2. After finding the invariance group $\mathcal{I}$ and building a matrix A of maximal rank such that F is $\mathcal{T}_A$-invariant, if Proposition 7.1(i) holds, then we readily decide that the system in Proposition 7.1(iv) has more than one positive root.

Proposition 7.1(i) can be checked computationally by computing the $n \times n$ minors of Γ_α, and checking if they all simultaneously vanish for some $\alpha \in \mathbb{R}^n_{>0}$. If $d = n - s$, then Γ_α is square and it becomes sufficient to check whether $\det(\Gamma_\alpha)$ vanishes for some $\alpha \in \mathbb{R}^n_{>0}$. This is true precisely if $\det(\Gamma_\alpha)$, viewed as a polynomial in $\mathbb{R}[\alpha]$, is either zero or has two nonzero terms with different signs [MFR+15, Thm. 2.13]. Proposition 7.1(ii) can be checked by deciding the feasibility of a system of linear inequalities, see [MFR+15, §4].

Example 7.3. For Example 2.3, which we know is toric, we take

$$L = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ -2 & 1 & -1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

This gives

$$\det(\Gamma_\alpha) = \det\left(\begin{bmatrix} -\alpha_1 & -\alpha_2 & 0 & 0 & \alpha_5 \\ 0 & 0 & 0 & \alpha_4 & 0 \\ -\alpha_1 & -\alpha_2 & \alpha_3 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ -2 & 1 & -1 & 1 & 0 \end{bmatrix}\right) = -\alpha_1\alpha_3\alpha_4 - \alpha_1\alpha_4\alpha_5 - 2\,\alpha_2\alpha_3\alpha_4 - \alpha_2\alpha_4\alpha_5 - \alpha_3\alpha_4\alpha_5\,.$$

As this determinant does not vanish for $\alpha \in \mathbb{R}^5_{>0}$, Proposition 7.1(i) holds and hence the augmented system in Proposition 7.1(iv) does not have multiple positive zeros.

7.2. **Constant coordinates.** A second question of interest in the context of reaction network theory is to determine whether the set of positive zeros of a vertical system F are contained in a translate of a coordinate hyperplane for all positive parameter values. In that context, the problem has been studied with various algebraic techniques in the past [Mil11, KPMD+12, PEF22, GPGH+25], but just as with multistationarity, the analysis is considerably simplified under the assumption of toricity. The following result is an immediate consequence of the definitions.

Proposition 7.4. *Let F be a vertical system with defining matrices $C \in \mathbb{R}^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$ with $\ker(C) \cap \mathbb{R}^m_{>0} \neq \varnothing$. Assume that F is $\mathcal{T}_A$-invariant for a matrix $A \in \mathbb{Z}^{d\times n}$ of rank d. The following holds:*

(i) If for all $\kappa \in \mathbb{R}^m_{>0}$, $\mathbb{V}_{>0}(F_\kappa)$ is contained in a finite union of translates of the coordinate hyperplane $\{x_i = 0\}$, then $A_i = 0_d$.

(ii) If $A_i = 0_d$ and F is locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$, then for all $\kappa \in \mathbb{R}^m_{>0}$, $\mathbb{V}_{>0}(F_\kappa)$ is contained in a finite union of translates of the coordinate hyperplane $\{x_i = 0\}$. If in addition F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$, then there is only one such hyperplane.

Example 7.5. In Example 2.3, the fact that there is $\mathcal{T}_A$-toricity and the fourth column of A is zero reveals, by Proposition 7.4, that for all $\kappa \in \mathbb{R}^6_{>0}$, $\mathbb{V}_{>0}(F_\kappa)$ is contained in a translate of the coordinate hyperplane $\{x_4 = 0\}$. In particular, the 4-th coordinate of points in $\mathbb{V}_{>0}(F_\kappa)$ attains always the same value.

7.3. **Algorithm for $\mathbb{G} = \mathbb{R}_{>0}$.** Given matrices $C \in \mathbb{R}^{s\times m}$ of rank s and $M \in \mathbb{Z}^{n\times m}$, we have presented several results to address the (generic) toricity of the zero sets of the associated vertical system F. We gather these in Algorithm 7.6.

Algorithm 7.6 (Summary for $\mathbb{R}_{>0}$)**.**
Input: Matrices $C \in \mathbb{R}^{s\times m}$ of rank s, $M \in \mathbb{Z}^{n\times m}$
Output: Whether $F = C(\kappa \circ x^M)$ is generically consistent, (generically) locally $\mathcal{T}_A$-toric, or (generically) $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$

Run Algorithm 4.7. Proceed if a maximal-rank matrix $A \in \mathbb{Z}^{d\times n}$ is returned (hence $\ker(C)\cap\mathbb{R}^m_{>0} \neq \varnothing$
 # Decide generic consistency
Generate a random $w \in \ker(C)$ and compute $r := \mathrm{rk}(C\,\mathrm{diag}(w)M^\top)$
if $r < s$ and $\mathrm{rk}(C\,\mathrm{diag}(w)M^\top) < s$ for all $w \in \ker(C)$ **then**
  return "F is not generically consistent over $\mathbb{R}_{>0}$"
if $s + d < n$ **then**
  return F is not generically locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$
 # At this point we know F is generically locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$
 # We proceed to study toricity and number of cosets
if All coefficients of $\det(\mathcal{L}_{\mu,\alpha})$ have the same sign **then**
  return F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$
Find $mv :=$ mixed volume of the coset counting system
if $\mathrm{rk}(C\,\mathrm{diag}(w)M^\top) = s$ for all $w \in \ker(C)\cap\mathbb{R}^m_{>0}$ **then**
  if $mv = 1$ **then**
    return F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$
  if Interior of polyhedral cone $\mathrm{row}(A)\cap\mathbb{R}^n_{\geq 0}$ is nonempty and Proposition 6.5(ii) holds **then**
    Set $r :=$ number of solutions in $\mathbb{R}^n_{>0}$ of the coset counting system for a random choice of $\kappa \in \mathbb{R}^m_{>0}$
    if $r = 1$ **then**
      return F is $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$
    return F is locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$ with r cosets
  return F is locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$ with constant number of cosets and at most mv
if $mv = 1$ **then**
  return F is generically $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$
return F is generically locally $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$ with at most mv cosets

8. Reaction-network-theoretic perspectives

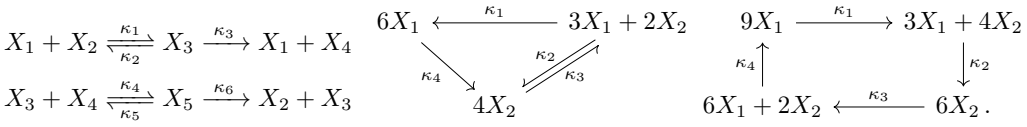
In this section we focus on the motivating scenario, namely that of reaction networks. We introduce them in Section 8.1 and adapt our algorithms for toricity to this setting. We exploit the special structure of the reaction networks to simplify some computations, and apply our algorithms to the networks of the database ODEbase [LSR22], to illustrate their usability. Finally, we compare our criteria to previous results on toricity.

8.1. **Reaction networks.** Our main motivation for studying systems of the form (2.1) comes from chemical reaction network theory, which we now give a quick introduction to. For a more thorough introduction, we refer the reader to [Dic16, Fei19].

A ***reaction network*** on an ordered set $\mathcal{S} = \{X_1, \ldots, X_n\}$ of species is a collection of m reactions between formal nonnegative linear combinations of the species (called complexes):

$$\sum_{i=1}^{n} \alpha_{ij} X_i \longrightarrow \sum_{i=1}^{n} \beta_{ij} X_i\,, \qquad j \in [m]\,, \tag{8.1}$$

where $\alpha_{ij}, \beta_{ij} \in \mathbb{Z}_{\geq 0}$. The net production of the species in each of the reactions is encoded by the stoichiometric matrix $N = (\beta_{ij} - \alpha_{ij}) \in \mathbb{Z}^{n\times m}$.



(a) Network for Example 2.3 **(b)** Network for (2.9) **(c)** Network for Example 3.10

Figure 8.1. (a) is a model of the IDHKP-IDH system in bacterial cell [SF10]; (b) is a variation of the classical triangle network that appears in several places in the literature (e.g. [HJ72, Eq. 7-2], [CDSS09, Ex. 1], [MDSC12, Ex. 2.3]); (c) is a variation of a classical network studied in [HJ72, §7].

The concentration of the respective species is denoted by a vector $x = (x_1, \dots, x_n)$ of nonnegative real numbers. Under common assumptions, these concentrations vary according to an autonomous ordinary differential equation system of the form

$$\frac{dx}{dt} = N(\kappa \circ x^M)\,, \tag{8.2}$$

where $M \in \mathbb{Z}^{n\times m}$ is called the kinetic matrix and $\kappa = (\kappa_1, \dots, \kappa_m) \in \mathbb{R}^m_{>0}$ is a vector of rate constants, which typically are viewed as unknown parameters. The main example of this construction arises under the mass-action assumption, where $M = (\alpha_{ij})$ is the reactant matrix consisting of the coefficients of the left-hand sides of the reactions. In this case, $\mathbb{R}^n_{\geq 0}$ is forward-invariant by the ODE system (8.2).

Letting $s = \mathrm{rk}(N)$, we choose a matrix $C \in \mathbb{R}^{s\times n}$ of rank S such that $\ker(C) = \ker(N)$. Then, the steady states of (8.2) are the zeros of the vertically parameterized system

$$F = C(\kappa \circ x^M) \in \mathbb{R}[\kappa, x^{\pm}].$$

We call any such system ***the steady state system*** (note that a choice of C is implicitly made). One is particularly interested in steady states with *strictly positive* entries.

As trajectories of (8.2) are confined in parallel translates of $\mathrm{im}(N)$, by letting $L \in \mathbb{R}^{(n-s)\times n}$ be any matrix whose rows form a basis for the left-kernel of N, trajectories are confined in the sets of the form

$$\{x \in \mathbb{R}^n_{\geq 0} : Lx - b = 0\}, \qquad b \in \mathbb{R}^{n-s},$$

which are called ***stoichiometric compatibility classes***. Positive steady states in stoichiometric compatibility classes are then the positive zeros of the augmented vertically parametrized system

$$(C(\kappa \circ x^M), Lx - b) \in \mathbb{R}[\kappa, b, x^{\pm}]. \tag{8.3}$$

Many of the examples of vertical systems in the previous sections are steady state systems of reaction networks with mass-action kinetics. Some examples are given in Figure 8.1. As already hinted in the previous sections, deciding upon two algebraic questions have been central in the study of reaction networks:

- The network is said to have the capacity for ***multistationarity***, if system (8.3) admits at least two positive zeros for some choice of $\kappa \in \mathbb{R}^m_{>0}$ and $b \in \mathbb{R}^{n-s}$.
- A network is said to have ***absolute concentration robustness*** (or ***ACR*** for short) with respect to a variable x_i if $\pi_i(\mathbb{V}_{>0}(F_\kappa))$ consists of at most a single point for all $\kappa \in \mathbb{R}^m_{>0}$, where $\pi_i \colon \mathbb{R}^n \to \mathbb{R}$ denotes the canonical projection onto the ith factor.

Multistationarity might imply that trajectories might converge to different steady states for the same parameter values, and has been associated with robust cell decision making.

ACR means that the concentration x_i at steady state is independent from initial conditions of the system. Because of this, ACR is believed to be a mechanism that contributes to the remarkable robustness many biological systems display to changes in their environment [SF10]. A weaker notion is that of ***local ACR***, where one instead requires the projection to be a finite set [PEF22].

Proposition 7.1 and Proposition 7.4 give criteria to decide upon these properties when the system F displays some form of toricity. For example, in Example 7.3 and Example 7.5 we verified that the network in Figure 8.1(a) is not multistationary (the matrix L given in Example 7.3 defines the stoichiometric compatibility classes) and has ACR with respect to X_4.

Example 8.1. Consider the following network studied in [SC13]:

$$X_1 + X_2 \rightleftharpoons X_3 \longrightarrow X_4 + X_5 \qquad X_7 + X_8 \rightleftharpoons X_2$$
$$X_4 + X_5 \rightleftharpoons X_6 \longrightarrow X_1 + X_5 \qquad X_5 + X_6 \rightleftharpoons X_9\,.$$

We apply Algorithm 7.6 and conclude that the steady state system is $\mathcal{T}_A$-toric with

$$A = \begin{bmatrix} 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 & -1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

In particular, it passes the injectivity test. (It also satisfies the conditions for having constant number of cosets, and a certified numerical computation reveals that that number is 1.) We readily see by Proposition 7.4 that the network does not have ACR. We build the matrix Γ_α in Proposition 7.1 and verify that $\det(\Gamma_\alpha)$ has both positive and negative coefficients. We conclude that the network has the capacity for multistationarity.

8.2. **Network reduction and toricity.** For a reaction network with mass-action kinetics, with stoichiometric matrix N and reactant matrix M, it turns out that the search for toric invariance of the steady state system can be simplified by removing *input-1 intermediates*.

We begin by recalling the notion of intermediates in reaction network theory. For a more detailed presentation we refer to [FW13, SF19a, SF19b]. Given a network with set of species $\mathcal{S}$, a ***choice of intermediates*** is a partition $\mathcal{S} = \mathcal{X} \sqcup \mathcal{Y}$ of the set of species into a set of ***non-intermediates*** $\mathcal{X}$ and a set of ***intermediates*** $\mathcal{Y}$, with the following properties:

(i) Each species $Y \in \mathcal{Y}$ only appears in monomolecular complexes (i.e. complexes where the coefficients sum to 1).

(ii) For each $Y \in \mathcal{Y}$ there exists a sequence of reactions

$$c \to Y_1 \to \cdots \to Y_i \to Y \to Y_{i+1} \to \cdots \to Y_r \to c' \tag{8.4}$$

with $Y_1, \ldots, Y_r \in \mathcal{Y} \setminus \{Y\}$ (there might be repetitions) and c, c' are complexes in the non-intermediates $\mathcal{X}$.

The non-intermediate complex c in (8.4) is called an ***input complex*** of Y. An ***input-1 intermediate*** has by definition a unique input complex. Note that there might be several possible choices of intermediates for a given network. One of the key ideas in the theory of intermediates is that some properties of the network are preserved in a simpler ***reduced network*** defined as follows: one removes the intermediates, and all reactions involving intermediates, and adds a reaction $c \to c'$ for every sequence of reactions as (8.4) [FW13].

By letting x, y denote the vectors of concentrations of the species in $\mathcal{X}$ and $\mathcal{Y}$ respectively, the key idea is that condition (i) ensures that, with mass-action kinetics, the entries of y appear linearly in the ODE system (8.2). Then condition (ii) ensures that the steady state system has a unique zero in y, which in addition is a polynomial in x with coefficients being rational functions in κ with all coefficients positive. Plugging the expressions of y at steady state into the remaining ODE equations (for x), one obtains the ODE system for the reduced network for a choice of rate constants given as a vector of rational functions $\varphi(\kappa)$ in the original rate constants.

When Y_i is an input-1 intermediate, the expression takes the form $y_i = \psi_i(\kappa)x^c$, where c is the vector of coefficients of the unique input of Y_i. In that case, it turns out that the networks share important toricity properties.

As a convention, we order the species such that the vector of concentrations is (x, y), that is, so that the non-intermediates come before the intermediates. We use a tilde to denote quantities and objects that correspond to the reduced network.

Proposition 8.2. *For a reaction network consider a choice of intermediates $\mathcal{S} = \mathcal{X} \sqcup \mathcal{Y}$ with $\mathcal{X} = \{X_1, \dots, X_n\}$ and $\mathcal{Y} = \{Y_1, \dots, Y_\ell\}$ consisting of input-1 intermediates. Let F be the steady state system. Let $B \in \mathbb{Z}_{\geq 0}^{n \times \ell}$ be the matrix where the ith column is the coefficient vector in $\mathcal{X}$ of the unique input complex of the ith intermediate. Then the following holds:*

(i) There are rational maps $\psi\colon \mathbb{R}^m_{>0} \to \mathbb{R}^\ell_{>0}$ and $\varphi\colon \mathbb{R}^m_{>0} \to \mathbb{R}^{\tilde{m}}_{>0}$ such that we have a bijection

$$\Phi_\kappa\colon \mathbb{V}_{>0}(\tilde{F}_{\varphi(\kappa)}) \to \mathbb{V}_{>0}(F_\kappa), \quad x \mapsto (x, \psi(\kappa) \circ x^B).$$

(ii) If $\mathbb{V}_{>0}(F_\kappa)$ is $\mathcal{T}_A$-invariant over $\mathbb{R}_{>0}$ for $A \in \mathbb{Z}^{d \times m}$, then $A = [\tilde{A}|\tilde{A}B]$ for some $\tilde{A} \in \mathbb{Z}^{d \times n}$.

(iii) With the notation in (ii), $\mathbb{V}_{>0}(\tilde{F}_{\varphi(\kappa)})$ is $\mathcal{T}_{\tilde{A}}$-invariant over $\mathbb{R}_{>0}$ if and only if $\mathbb{V}_{>0}(F_\kappa)$ is $\mathcal{T}_A$-invariant over $\mathbb{R}_{>0}$. Furthermore, Φ_κ descends to a bijection

$$\mathbb{V}_{>0}(\tilde{F}_{\varphi(\kappa)})/\mathcal{T}_{\tilde{A}}^{>0} \to \mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0}.$$

(iv) If $\tilde{F}$ is (generically/locally) $\mathcal{T}_{\tilde{A}}$-toric over $\mathbb{R}_{>0}$ then F is (generically/locally) $\mathcal{T}_A$-toric over $\mathbb{R}_{>0}$. The reverse implication holds also if φ surjective.

Proof. Statement (i) is shown in [FW13], see also [SF19a, SF19b]. To see statement (ii), let $y \in \mathbb{V}_{>0}(F_\kappa)$ and write it as $y = \Phi_\kappa(x)$ for the unique $x \in \mathbb{V}_{>0}(\tilde{F}_{\varphi(\kappa)})$. By writing $A = [\tilde{A}|\tilde{A}']$ with $\tilde{A} \in \mathbb{Z}^{d \times n}$, we have

$$\Phi_\kappa(x) \circ t^A = (x, \psi(\kappa) \circ x^B) \circ (t^{\tilde{A}}, t^{\tilde{A}'}) = (x \circ t^{\tilde{A}}, \psi(\kappa) \circ x^B \circ t^{\tilde{A}'}).$$

By hypothesis, $\Phi_\kappa(x) \circ t^A \in \mathbb{V}_{>0}(F_\kappa)$, hence it belongs to the image of Φ_κ. Therefore

$$\psi(\kappa) \circ x^B \circ t^{\tilde{A}'} = \psi(\kappa) \circ (x \circ t^{\tilde{A}})^B = \psi(\kappa) \circ x^B \circ t^{\tilde{A}B},$$

and as this holds for all $t \in \mathbb{R}^d_{>0}$, we must have that $\tilde{A}' = \tilde{A}B$, giving (ii).

Statement (iii) is now a direct consequence of the equality $\Phi_\kappa(x \circ t^{\tilde{A}}) = \Phi_\kappa(x) \circ t^A$. Finally, (iv) follows from (iii) as $\#(\mathbb{V}_{>0}(\tilde{F}_{\varphi(\kappa)})/\mathcal{T}_{\tilde{A}}^{>0}) = \#(\mathbb{V}_{>0}(F_\kappa)/\mathcal{T}_A^{>0})$. □

Remark 8.3. Surjectivity of φ in Proposition 8.2(i) corresponds to the realization condition being satisfied for input-1 intermediates by [SF19b, Prop. 5.3]. In loc. cit. several scenarios where this holds are given. In particular, it holds in the common scenario where intermediates appear in isolated motifs of the form $c \longleftrightarrow Y_1 \longleftrightarrow \dots \longleftrightarrow Y_\ell \longrightarrow c'$, with $\longleftrightarrow$ being either $\longrightarrow$ or $\rightleftharpoons$. We conjecture that φ is surjective whenever all intermediates are input-1.

Proposition 8.2 has important practical consequences: the reduced network is smaller as it has both less variables and reactions. Hence the computational cost for checking toricity is lower, sometimes dramatically lower. Additionally, it might be the case that some checks are inconclusive for the original network, but succeed for the reduced network. An example of this is given below in Example 8.5, where the injectivity test from Theorem 6.3 fails for the original network but it is passed for the reduced network.

Example 8.4. For Figure 8.1(a), one possible choice of intermediates is $\mathcal{X} = \{X_1, X_2, X_3, X_4\}$ and $\mathcal{Y} = \{X_5\}$, and the only input complex of X_5 is $X_3 + X_4$ (so the matrix B in Proposition 8.2 is $(0\ 0\ 1\ 1)^\top$ and φ is surjective, see Remark 8.3). The reduced network is

$$X_1 + X_2 \rightleftharpoons X_3 \longrightarrow X_1 + X_4 \qquad X_3 + X_4 \longrightarrow X_2 + X_3\,.$$

The maps φ, ψ and $\Phi_{\tilde{\kappa}}$ from Proposition 8.2 are

$$\varphi(\kappa) = \left(\kappa_1, \kappa_2, \kappa_3, \tfrac{\kappa_4\kappa_6}{\kappa_5+\kappa_6}\right), \qquad \psi(\kappa) = \tfrac{\kappa_4}{\kappa_5+\kappa_6}, \qquad \Phi_\kappa(x_1, \dots, x_4) = \left(x_1, \dots, x_4, \tfrac{\kappa_4}{\kappa_5+\kappa_6} x_3 x_4\right).$$

Algorithm 7.6 tells us that the steady state system $\tilde{F}$ of the reduced network is $\mathcal{T}_{\tilde{A}}$-toric, hence the original steady state system (2.7) is $\mathcal{T}_A$-toric, with

$$\tilde{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \qquad \text{and} \qquad A = [\ \tilde{A} \mid \tilde{A}B\ ] = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

(In fact, the steady state system of the reduced network is binomial.) This is in accordance with what we saw in Example 2.3.

Example 8.5. The classical network from Shinar and Feinberg's work on ACR [SF10] contains three input-1 intermediates. The original and reduced networks are respectively:

$$\begin{aligned} X_1 &\rightleftharpoons X_2 \rightleftharpoons X_3 \longrightarrow X_4 \\ X_4 + X_5 &\rightleftharpoons X_6 \longrightarrow X_2 + X_7 \\ X_3 + X_7 &\rightleftharpoons X_8 \longrightarrow X_3 + X_5 \\ X_1 + X_7 &\rightleftharpoons X_9 \longrightarrow X_1 + X_5 \end{aligned} \qquad \begin{aligned} X_1 &\rightleftharpoons X_2 \rightleftharpoons X_3 \longrightarrow X_4 \\ X_4 + X_5 &\longrightarrow X_2 + X_7 \\ X_3 + X_7 &\longrightarrow X_3 + X_5 \\ X_1 + X_7 &\longrightarrow X_1 + X_5\,. \end{aligned}$$

Applying Algorithm 4.7 to the reduced network, we conclude that the steady state systems of these two networks are torically invariant with respect to the following matrices, respectively:

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad \tilde{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}.$$

Algorithm 7.6 tells us that $\tilde{F}$ is $\mathcal{T}_{\tilde{A}}$-toric, as it passes the injectivity test. Proposition 8.2 allows us to conclude that F is $\mathcal{T}_A$-toric. In this case, however, F does not pass the injectivity test.

8.3. **Case study: Networks from ODEbase.** To illustrate the applicability of our results for realistic networks, we have applied our algorithms on the networks from the repository of biological and biomedical models BioModels [N+06], using the preprocessing performed in the database ODEbase [LSR22] (which, among other data about the networks, contains stoichiometric matrices and reactant matrices).

In our analysis, we work under the assumption of mass-action kinetics for all models (regardless of the exact kinetic model registered in BioModels). We have considered all 69 nonlinear networks in ODEbase that satisfy

$$m \leq 100\,, \qquad n - \operatorname{rk}(N) > 0\,, \qquad \text{and} \qquad \ker(N) \cap \mathbb{R}^m_{>0} \neq \varnothing\,.$$

For each such network we have applied Algorithm 7.6 to the steady state system. (If the network had input-1 intermediates, we computed a matrix $\tilde{A}$ with invariance for the reduced network, and attempted to prove $\mathcal{T}_{\tilde{A}}$-toricity for the reduced network, before proceeding with Step 8.)

A Julia implementation of the algorithms is available in the GitHub repository

`https://github.com/oskarhenriksson/toric-vertically-parametrized-systems`.

The implementation relies on the packages `Oscar.jl` [OSC24] for polyhedral and symbolic computations, `HomotopyContinuation.jl` [BT18] for certified numerical solving of the coset counting system, and `Graphs.jl` [FBS+21] for working with the reaction graph.

The repository also contains output of the computations for each of the analyzed networks. We here report some summarized data:

- For 38 networks, we rule out (local) toricity.
- For 31 networks, we verify local toricity, and for 30 of them, we verify toricity. This, in particular, includes the largest network among the 155 analyzed network, `250`, with 93 reactions involving 47 species.
- For one of the networks with local toricity (`835`), none of our conditions for toricity are satisfied, and the mixed volume bound is 46 (but we the steady state ideal is actually binomial for all positive rate constants).
- We verify capacity for multistationarity for 2 networks, and preclude multistationarity for 27 of them. We verify local ACR for 15 networks, and ACR for 14 of them.

8.4. **Other flavors of toricity.** In this final section, we view our results in the context of some previous approaches to determine toricity in reaction networks. In this subsection, the reaction networks are taken with mass-action kinetics, and hence M is the reactant matrix.

Conservation laws and deficiency theory. Some classical monostationarity results can be viewed as special cases of our results, for A a matrix satisfying $AN = 0$. In this case, translated subspaces of the form $x^* + \ker(A)$ are stoichiometric compatibility classes, so given invariance, toricity precisely corresponds to monostationarity, c.f. Proposition 6.1.

Toric invariance with respect to the matrix L defining stoichiometric compatibility classes corresponds to partitions of the reactions into linkage classes (connected components of the network viewed as a digraph) and connects to the well-studied complex balancing steady states.

Let r be the number of complexes of a given reaction network and m the number of columns of the exponent matrix M of the steady state system. We introduce the ***linkage class partition*** of $[m] = \gamma_1 \sqcup \cdots \sqcup \gamma_\ell$ where two indices are in the same subset if the corresponding reactions belong to the same linkage class. Recall that the ***deficiency*** of the network is $\delta := r - s - \ell \geq 0$.

Proposition 8.6. *Consider a network with ℓ linkage classes, its steady state system F, and $L \in \mathbb{R}^{(n-s)\times n}$ a matrix defining the stoichiometric compatibility classes.*

(i) If the fundamental partition of the steady state system is finer than the linkage class partition, then F is $\mathcal{T}_L$-invariant.

(ii) Statement (i) holds if the network is connected.

(iii) Statement (i) holds if there is a direct sum decomposition $\mathrm{im}(N) = \mathrm{im}(N_1) \oplus \cdots \oplus \mathrm{im}(N_\ell)$, where N_i is the stoichiometric matrix of the ith linkage class.

(iv) If the network is weakly reversible with deficiency zero or satisfies the conditions of the deficiency one theorem from [Fei95], then F is locally $\mathcal{T}_L$-toric.

Proof. Let $Y \in \mathbb{R}^{n\times r}$ be the matrix whose columns are the coefficients of all complexes that appear in the network in some chosen order. The columns of M are among the columns of Y. Let $C_G \in \mathbb{Z}^{r\times m}$ be the incidence matrix of the network seen as a directed graph: the entry (i, j) is $1, -1, 0$ if the ith complex is on the right, left, or does not occur in the jth reaction. Let $[m] = \gamma_1 \sqcup \ldots \gamma_\ell$ be the linkage class partition. For each $k \in [\ell]$, construct the vector $u_k \in \mathbb{Z}^m$ with 1 for the indices in γ_k and zero otherwise. These vectors generate $\ker(C_G^\top)$.

It is easy to see that $N = YC_G$, hence $0 = LYC_G$ by hypothesis, and the rows of LY belong to the left-kernel of C_G. In particular columns of LY corresponding to complexes in the same linkage class are all equal. It follows that a row a of L satisfies $aY_i = aY_j$ if $i, j \in \gamma_k$ for some k. Statement (i) now follows from Theorem 4.5, using that the fundamental partition is finer than the linkage class partition.

For (ii), if the network is connected, then $\ell = 1$, hence (i) holds. For (iii), the condition is equivalent to $\ker(N) = \ker(N_1) \oplus \cdots \oplus \ker(N_\ell)$. Hence, the support of an elementary vector of $\ker(N)$ is completely included in a subset of the linkage class partition and (i) holds.

For (iv), if the deficiency is zero, then $\ker(N) = \ker(C_G)$, see e.g. [Fei95, Lem. 6.1.4]. After a suitable reordering of the complexes and reactions, C_G is a block diagonal matrix, which gives that (iii) applies. Condition (iii) holds under the setting of the deficiency one theorem by hypothesis. We note also that the property in Proposition 5.5 holds for networks of deficiency zero and in the setting of the deficiency one theorem [Fei19, Sections 15.2 and 17.1], from where we conclude that F is locally $\mathcal{T}_L$-toric. □

The results of Proposition 8.6 are not new, but are reproven here using our approach. It is well known that for networks with deficiency zero (whose steady states are called *complex balanced*) satisfy that $\mathbb{V}_{>0}(F_\kappa) = x^*_\kappa \circ \mathcal{T}_L$, for some $x^*_\kappa \in \mathbb{R}^n_{>0}$, e.g. [CDSS09]. Statement (iii) is discussed also in [Bor13, §3.1]. The deficiency one theorem guarantees in addition exactly one coset for each $\kappa \in \mathbb{R}^m_{>0}$ and hence toricity with respect to L.

Networks with binomial steady state ideals. One of the other main works on toricity in reaction networks gives conditions to guarantee that the complex variety $\mathbb{V}_{\mathbb{C}}(F_\kappa)$ is cut out by binomials [MDSC12]. These networks are called *networks with toric steady states*. The approach in [MDSC12] takes a reaction network with mass-action kinetics and fixed κ, and writes the system $N(\kappa \circ x^M)$ as

$$F_\kappa = \Sigma_\kappa x^Y,$$

where $\Sigma_\kappa \in \mathbb{Q}(\kappa)^{n\times p}$ is the coefficient matrix and $Y \in \mathbb{Z}^{n\times p}$ has exactly one column per reactant complex of the network (the matrix Σ_κ in [MDSC12] might have extra zero columns, but these are irrelevant for the results under discussion). Note that $m \geq p$.

Condition 3.1 in [MDSC12] ask for the existence of a basis $b^1, \ldots, b^d \in \mathbb{R}^p_{\geq 0}$ for $\ker(\Sigma_\kappa)$ such that their supports $I_1, \ldots, I_d$ form a partition of $[p]$. When this is the case, then the system admits toricity with respect to the maximal-rank matrix A such that $AY_i = AY_j$ whenever i, j belong to the same subset I_k (see [MDSC12, Thm 3.11]). This construction resembles Theorem 4.5. To understand the connection, we need first to assume that the partition is independent of $\kappa \in \mathbb{R}^m_{>0}$. Then, the vectors $b^1, \ldots, b^d$ are rational functions in κ, and by multiplying by the denominators if necessary, we can assume they are polynomial and hence continuous functions in $\mathbb{R}^m_{\geq 0}$.

Let $\iota\colon [m] \to [p]$ where $\iota(i)$ is the index of the column of Y that has M_i as column. The hypothesis of [MDSC12] gives then that

$$AM_i = AM_j \quad \text{if } \iota(i), \iota(j) \in I_k \text{ for some } k \in [d]. \tag{8.5}$$

Note that ι is surjective and ι^{-1} induces a partition of $[m]$. The connection between [MDSC12] and this work stems from the fact that the fundamental partition is finer than that induced by ι^{-1}, which we show next.

Let $K_\kappa \in \mathbb{R}^{m\times p}$ be the matrix such that $\kappa \circ x^M = K_\kappa x^Y$, and more explicitly, $(K_\kappa)_{i,\iota(i)} = \kappa_i$ and zero otherwise. It follows easily that $\Sigma_\kappa = NK_\kappa$, and that K_κ has rank p for all $\kappa \in \mathbb{R}^m_{>0}$. By construction, it holds that $v_{\kappa,j} := K_\kappa\, b^j \in \ker(N)$ for all $\kappa \in \mathbb{R}^m_{>0}$, and $\operatorname{supp}(v_{\kappa,j}) = \iota^{-1}(I_j)$. By continuity, if some entries of κ are set to zero, the vector $v_{\kappa,j}$ still belongs to $\ker(N)$.

As $K_\kappa(1, \ldots, 1)^\top = \kappa$, any vector in $\ker(N)$ belongs to $\operatorname{im}(K_\kappa)$ for some $\kappa \in \mathbb{R}^m$. Hence

$$\ker(N) = \bigcup_{\kappa\in\mathbb{R}^m} \operatorname{im}(K_\kappa) \cap \ker(N) = \bigcup_{\kappa\in\mathbb{R}^m} K_\kappa(\ker(\Sigma_\kappa))\,,$$

where in the last equality we use that K_κ has maximal column rank. Using that $b^1, \ldots, b^d$ form a basis for $\ker(\Sigma_\kappa)$, and that the vectors $v_{\kappa,j}$, $v_{\kappa,i}$ have disjoint support if $i \neq j$, we obtain that any elementary vector of $\ker(N)$ has support included in one of $\operatorname{supp}(v_{\kappa,j}) = \iota^{-1}(I_j)$. This implies that the fundamental partition is finer than that induced by ι^{-1} as desired.

References

[ADLS07] D. Angeli, P De Leenheer, and E. Sontag. A Petri net approach to persistence analysis in chemical reaction networks. In *Biology and control theory: current challenges*, volume 357 of *Lect. Notes Control Inf. Sci.*, pages 181–216. Springer, Berlin, 2007.

[BBH24] M. Banaji, B. Boros, and J. Hofbauer. Bifurcations in planar, quadratic mass-action networks with few reactions and low molecularity. *Nonlinear Dyn.*, (Online), 2024.

[BCR98] J. Bochnak, M. Coste, and M. Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, Berlin, 1998.

[Ber75] D. N. Bernshtein. The number of roots of a system of equations. *Funct. Anal. Appl.*, 9:183–185, 1975.

[BFT17] C. Barrett, P. Fontaine, and C. Tinelli. The SMT-LIB standard: Version 2.6. Technical report, Department of Computer Science, The University of Iowa, 2017.

[BI64] A. Ben-Israel. Notes on linear inequalities, 1: The intersection of the nonnegative orthant with complementary orthogonal subspaces. *J. Math. Anal. Appl.*, (9):303–314, 1964.

[BiMCS22] L. Brustenga i Moncusí, G. Craciun, and M.-S. Sorea. Disguised toric dynamical systems. *J. Pure Appl. Algebra*, 226(8):107035, 2022.

[Bor12] B. Boros. Notes on the deficiency-one theorem: Multiple linkage classes. *Math. Biosci.*, 235(1):110–122, 2012.

[Bor13] B. Boros. *On the Positive Steady States of Deficiency-One Mass Action Systems*. PhD thesis, Eötvös Loránd University, 2013. Available at https://web.cs.elte.hu/~bboros/bboros_phd_thesis.pdf.

[BRT23] P. Breiding, K. Rose, and S. Timme. Certifying zeros of polynomial systems using interval arithmetic. *ACM Trans. Math. Softw.*, 49(1), 2023.

[BT18] P. Breiding and S. Timme. `HomotopyContinuation.jl`: A package for homotopy continuation in Julia. In J. H. Davenport, M. Kauers, G. Labahn, and J. Urban, editors, *Mathematical Software – ICMS 2018*, pages 458–465, Cham, 2018. Springer International Publishing.

[CDSS09] G. Craciun, A. Dickenstein, A. Shiu, and B. Sturmfels. Toric dynamical systems. *J. Symbolic Comput.*, 44:1551–1565, 2009.

[CIK19] C. Conradi, A. Iosif, and T. Kahle. Multistationarity in the space of total concentrations for systems that admit a monomial parametrization. *Bull. Math. Biol.*, 81(10):4174–4209, 2019.

[CK15] C. Conradi and T. Kahle. Detecting binomiality. *Adv. Appl. Math.*, 71:52–67, 2015.

[CLO05] D. A. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Graduate Texts in Mathematics. Springer New York, 2005.

[CLO15] D. A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer International Publishing, 2015.

[CLS11] D. A. Cox, J. B. Little, and H. K. Schenck. *Toric Varieties*. Graduate studies in mathematics. American Mathematical Society, 2011.

[Dic16] A. Dickenstein. Biochemical reaction networks: an invitation for algebraic geometers. In *Mathematical Congress of the Americas 2013*, volume 656 of *Contemporary Mathematics*, page 65–83. American Mathematical Society, 2016.

[FBS+21] J. Fairbanks, M. Besançon, S. Simon, J. Hoffiman, N. Eubank, and S. Karpinski. JuliaGraphs/Graphs.jl: an optimized graphs package for the Julia programming language, 2021.

[Fei95] M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Ration. Mech. Anal.*, 132(4):311–370, 1995.

[Fei19] M. Feinberg. *Foundations of Chemical Reaction Network Theory*. Applied Mathematical Sciences. Springer International Publishing, 2019.

[FHPE24] E. Feliu, O. Henriksson, and B. Pascual-Escudero. Generic consistence and nondegeneracy of vertically parametrized systems. Preprint: arXiv:2304.02302v4, 2024.

[FW13] E. Feliu and C. Wiuf. Simplifying biochemical models with intermediate species. *J. R. Soc. Interface*, 10(87):20130484, 2013.

[GH21] E. Gross and C. Hill. The steady-state degree and mixed volume of a chemical reaction network. *Adv. Appl. Math.*, 131(C), 2021.

[GIR+20] D. Grigoriev, A. Iosif, H. Rahkooy, T. Sturm, and A. Weber. Efficiently and effectively recognizing toricity of steady state varieties. *Math. Comput. Sci.*, 2020.

[GKZ94] I. M. Gelfand, M. M. Kapranov, and A.V. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Mathematics (Birkhäuser). Springer, 1994.

[GMS06] D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *Ann. Statist.*, 34(3):1463–1492, 2006.

[GPGH+25] L. D. García Puente, E. Gross, H. A. Harrington, M. Johnston, N. Meshkat, M. Pérez Millán, and A. Shiu. Absolute concentration robustness: Algebra and geometry. *J. Symb. Comput.*, 128:102398, 2025.

[HADlC+22] B. S. Hernández, D. A. Amistas, R. J. L. De la Cruz, L. L. Fontanil, A. A. de los Reyes V, and E. R. Mendoza. Independent, incidence independent and weakly reversible decompositions of chemical reaction networks. *MATCH*, 87:367–396, 2022.

[HHR24] P. A. Helmnick, O. Henriksson, and Y. Ren. A tropical method for solving parametrized polynomial systems. Preprint: arXiv:2409.13288, 2024.

[HJ72] F. Horn and R. Jackson. General mass action kinetics. *Arch. Ration. Mech. Anal.*, 47:81–116, 1972.

[HL12] E. Hubert and G. Labahn. Rational invariants of scalings from Hermite normal forms. *Proceedings of the International Symposium on Symbolic and Algebraic Computation, ISSAC*, pages 219–226, 07 2012.

[HM23] B. S. Hernández and E. R. Mendoza. Positive equilibria of power law kinetics on networks with independent linkage classes. *J. Math. Chem.*, 61(3):630–651, 2023.

[HR22] P. A. Helminck and Y. Ren. Generic root counts and flatness in tropical geometry. Preprint: arXiv:2206.07838v2, 2022.

[HSSY23] S. J. Haque, M. Satriano, M.-S. Sorea, and P. Y. Yu. The disguised toric locus and affine equivalence of reaction networks. *SIAM J. Appl. Dyn. Syst.*, 22(2):1423–1444, 2023.

[Joh22] M. D. Johnston. Analysis of mass-action systems by split network translation. *J. Math. Chem.*, 60(1):195–218, 2022.

[KPMD+12] R. L. Karp, M. Pérez Millán, T. Dasgupta, A. Dickenstein, and J. Gunawardena. Complex-linear invariants of biochemical networks. *J. Theor. Biol.*, 311:130–138, 2012.

[KV24] T. Kahle and J. Vill. Efficiently deciding if an ideal is toric after a linear coordinate change. Preprint: arXiv:2408.14323, 2024.

[LSR22] C. Lüders, T. Sturm, and O. Radulescu. ODEbase: a repository of ODE systems for systems biology. *Bioinf. Adv.*, 2(1), 2022.

[MDSC12] M. Pérez Millán, A. Dickenstein, A. Shiu, and C. Conradi. Chemical reaction systems with toric steady states. *Bull. Math. Biol.*, 74:1027–1065, 2012.

[MFR+15] S. Müller, E. Feliu, G. Regensburger, C. Conradi, A. Shiu, and A. Dickenstein. Sign conditions for injectivity of generalized polynomial maps with applications to chemical reaction networks and real algebraic geometry. *Found. Comput. Math.*, 16(1):69–97, 2015.

[Mil11] M. Pérez Millán. *Métodos algebraicos para el estudio de redes bioquímicas*. PhD thesis, Universidad de Buenos Aires, 2011. Available at https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis_n5103_PerezMillan.pdf.

[MP23] A. Maraj and A. Pal. Symmetry lie algebras of varieties with applications to algebraic statistics. Preprint: arXiv:2309.10741, 2023.

[MR24] S. Müller and G. Regensburger. Parametrized systems of polynomial inequalities with real exponents via linear algebra and convex geometry. Preprint: arXiv:2306.13916v3, 2024.

[N+06] N. Le Novère et al. BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, 34:D689–D691, 2006.

[OSC24] Oscar – open source computer algebra research system, version 1.2.0, 2024. Available at https://oscar.computeralgebra.de.

[OW24] N. K. Obatake and E. Walker. Newton-okounkov bodies of chemical reaction systems. *Adv. Appl. Math.*, 155:102672, 2024.

[PEF22] B. Pascual-Escudero and E. Feliu. Local and global robustness at steady state. *Math. Methods Appl. Sci.*, 45(1):359–382, 2022.

[Roc69] R. T. Rockafellar. The elementary vectors of a subspace of R^N. In *Combinatorial Mathematics and its Applications (Proc. Conf., Univ. North Carolina, Chapel Hill, N.C., 1967)*, volume No. 4 of *University of North Carolina Monograph Series in Probability and Statistics*, pages 104–127. Univ. North Carolina Press, Chapel Hill, NC, 1969.

[RS21a] H. Rahkooy and T. Sturm. Parametric toricity of steady state varieties of reaction networks. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing*, pages 314–333, Cham, 2021. Springer International Publishing.

[RS21b] H. Rahkooy and T. Sturm. Testing binomiality of chemical reaction networks using comprehensive Gröbner systems. In François Boulier, Matthew England, Timur M. Sadykov, and Evgenii V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing*, pages 334–352, Cham, 2021. Springer International Publishing.

[SC13] R. Straube and C. Conradi. Reciprocal enzyme regulation as a source of bistability in covalent modification cycles. *J. Theor. Biol.*, 330:56–74, 2013.

[SF10] G. Shinar and M. Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327:1389–1391, 2010.

[SF19a] A. Sadeghimanesh and E. Feliu. Gröbner bases of reaction networks with intermediate species. *Adv. Appl. Math.*, 107:74–101, 2019.

[SF19b] A. Sadeghimanesh and E. Feliu. The multistationarity structure of networks with intermediates and a binomial core network. *Bull. Math. Biol.*, 81:2428–2462, 2019.

[SS05] B. Sturmfels and S. Sullivant. Toric ideals of phylogenetic invariants. *J. Comput. Biol.*, 12(4):457–481, 2005.

[SS10] A. Shiu and B. Sturmfels. Siphons in chemical reaction networks. *Bull. Math. Biol.*, 72(6):1448–1463, 2010.

[Sul18] S. Sullivant. *Algebraic statistics*, volume 194 of *Grad. Stud. Math.* Providence, RI: American Mathematical Society (AMS), 2018.

[Tel22] S. Telen. Introduction to toric geometry. Preprint: arXiv:2203.01690, 2022.

[WF13] C. Wiuf and E. Feliu. Power-law kinetics and determinant criteria for the preclusion of multistationarity in networks of interacting species. *SIAM J. Appl. Dyn. Syst.*, 12(4):1685–1721, 2013.

Authors' addresses:

Elisenda Feliu, University of Copenhagen `efeliu@math.ku.dk`
Oskar Henriksson, University of Copenhagen `oskar.henriksson@math.ku.dk`

E

Moment varieties from inverse Gaussian and gamma distributions

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Lisa Seccia
Mathematical Institute
University of Neuchâtel

Teresa Yu
Department of Mathematics
University of Michigan

Publication details

Published in *Algebraic Statistics* **15**, 2 (2024)
DOI: 10.2140/astat.2024.15.329

Algebraic Statistics
Vol. 15, No. 2, 2024
https://doi.org/10.2140/astat.2024.15.329

MOMENT VARIETIES FROM INVERSE GAUSSIAN AND GAMMA DISTRIBUTIONS

OSKAR HENRIKSSON, LISA SECCIA AND TERESA YU

Motivated by previous work on moment varieties for Gaussian distributions and their mixtures, we study moment varieties for two other statistically important two-parameter distributions: the inverse Gaussian and gamma distributions. In particular, we realize the moment varieties as determinantal varieties and find their degrees and singularities. We also provide computational evidence for algebraic identifiability of mixtures, and study the identifiability degree and Euclidean distance degree.

1. Introduction

Suppose X is a univariate random variable from a distribution that depends on finitely many parameters $\theta = (\theta_1, \ldots, \theta_n)$, and we want to determine θ from observed data. One approach to this problem is the *method of moments*.

The r-th *moment* of X is defined as the expected value $m_r(\theta) := \mathbb{E}(X^r)$. For many distributions, the moments $m_r(\theta)$ are polynomials in the parameters θ (see, e.g., Table 2 of [10]). This makes it possible to estimate θ by computing the sample moments $\widetilde{m}_r := \frac{1}{N}\sum_{i=1}^{N} x_i^r$ for some large number N many observations, and then solving the system of polynomials

$$m_r(\theta) = \widetilde{m}_r, \quad r \in [d] \tag{1-1}$$

for some $d \in \mathbb{N}$. The law of large numbers implies that we can expect arbitrarily good approximations of θ among the solutions of this system, if N is large enough. In particular, the method of moments gives rise to a consistent estimator in many cases (see, e.g., [35, Theorem 9.6]).

For each $d \in \mathbb{N}$, one can define the *d-th moment variety* $\mathcal{M}_d \subseteq \mathbb{P}^d$ as the Zariski closure of the image of the map

$$\mathbb{C}^n \to \mathbb{P}^d, \quad \theta \mapsto [m_0(\theta) : m_1(\theta) : \cdots : m_d(\theta)].$$

This family of projective varieties provides a starting point for the algebraic-geometric perspective on the method of moments. We are particularly interested in the notions of *algebraic* and *rational identifiability* of the parameters, which correspond to understanding when the system (1-1) has finitely many complex solutions or a unique solution, and therefore to understanding when the parameters θ can be recovered from the moments. Such notions of identifiability can be interpreted as understanding generic fibers of the map $\mathbb{C}^n \to \mathbb{P}^d$, and can therefore be studied with techniques from algebra and geometry.

MSC2020: primary 13P25, 62R01; secondary 13C40, 14N05.
Keywords: method of moments, moment varieties, identifiability, determinantal varieties.

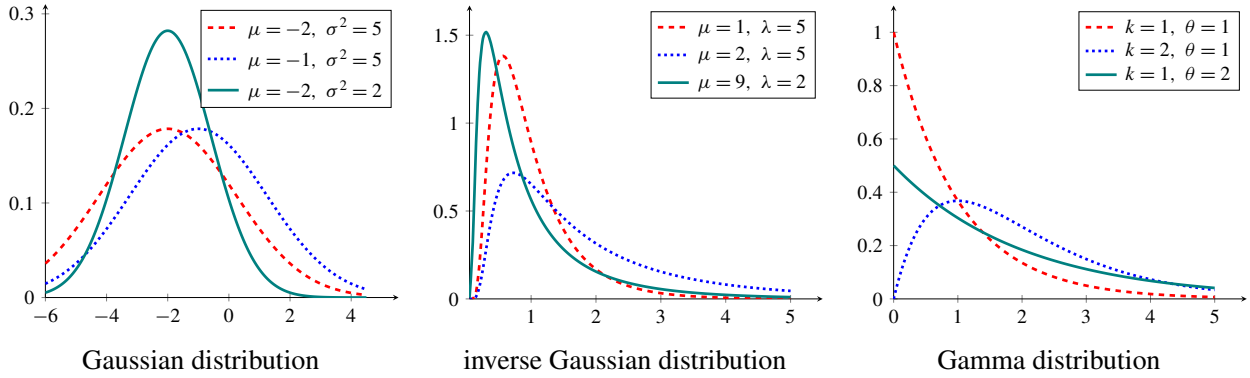


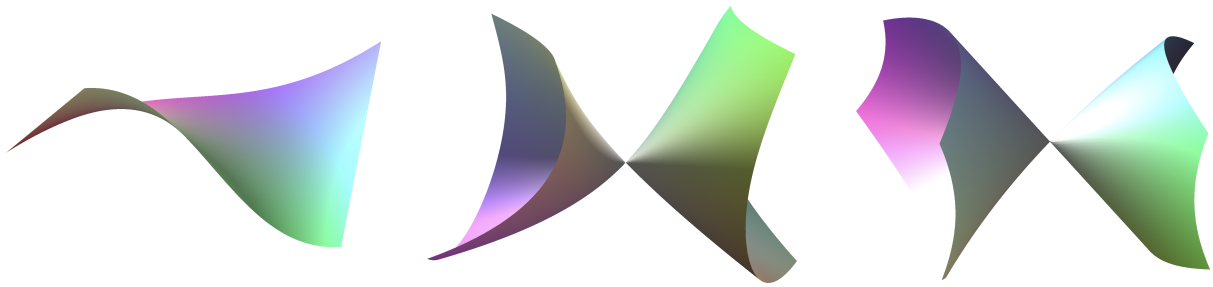
Figure 1. Probability density functions for three distributions.

An important tool for studying moment varieties is the concept of *cumulants* and their associated varieties $\mathcal{K}_d$ (see Definition 2.2). There is an isomorphism of affine varieties $\mathcal{M}_d \cap \{m_0 \neq 0\} \cong \mathcal{K}_d$, and passing from moments to cumulants often gives rise to simpler varieties (see, e.g., [2, §2] and the more general discussion in [14, §4–5]). Cumulants also have the additional advantage of being additive over independent variables. However, moments behave simpler with respect to certain other statistical operations; in particular, moment varieties of mixtures are secant varieties, which is a fact that we will return to in Section 5. Furthermore, the cumulant variety only captures an open dense patch of the moment variety, and therefore might miss important geometric aspects of the moment variety such as singularities.

Previous work on moment varieties. The algebraic study of moments and cumulant varieties goes back to Pearson in 1894 [33], who studied the mixtures of two univariate Gaussians. The case of Gaussians and Gaussian mixtures was later revisited from a nonlinear algebra perspective in [2], as well as in the subsequent papers [4; 1; 28] that addressed various identifiability questions, and (in the latter case) approximation methods based on numerical algebraic geometry. In [2], the authors study moments as projective varieties. The main advantage of studying moments from the perspective of projective geometry is that it allows for the study of secant varieties, which correspond to the moment varieties for mixtures of distributions. In general, algebraic and even rational identifiability for distributions listed in [10, Table 2] is clear, but it is unknown for mixtures of these distributions. Other distributions for which moment and cumulant varieties have been studied include uniform distributions on polytopes [27], Dirac and Pareto distributions [23], and mixtures of products [5].

Main contributions. In this paper, we study the moment varieties for the inverse Gaussian and gamma distributions, which are both two-parameter distributions used in a wide array of applications. See Figure 1 for examples of density functions, and see Figure 2 for an illustration of the moment varieties for $d = 3$.

Our main results (Theorems 3.1 and 4.1) give the homogeneous prime ideals for these moment varieties. We show that these ideals are *determinantal ideals*, in the sense that they are generated by all maximal minors of a matrix. Interestingly, both these varieties turn out to be Cohen–Macaulay, which raises Problem 4.11 on Cohen–Macaulayness of moment varieties.



Gaussian distribution Inverse Gaussian distribution Gamma distribution

Figure 2. Illustration of the $\{m_0 = 1\}$ patch of $\mathcal{M}_3$ for three distributions.

Using these results, we find the degrees and Hilbert series for these moment varieties, and give Gröbner bases for their ideals. In particular, Proposition 4.4 addresses a conjecture stated in [22] on the Hilbert series for the moment variety of the gamma distribution. We also identify the singular loci of the moment varieties (Propositions 3.6 and 4.6).

Additionally, we study the exponential and chi-squared distributions, which are one-parameter specializations of the gamma distribution, and we show that their moment varieties are rational normal curves (Propositions 4.8 and 4.9).

Using the moments-to-cumulants transformation, we prove that the cumulant varieties of the inverse Gaussian are (scaled) Veronese varieties (Proposition 3.8); this has previously been shown to be the case for the gamma distribution [22, Proposition 3.2.1]. These simple geometric models, together with previous results on the projective geometry of the corresponding moment varieties, could be the starting point for addressing the question of identifiability of mixtures.

We conclude the paper by providing computational evidence in Section 5.1 towards a conjecture on algebraic identifiability for mixtures of inverse Gaussian distributions and mixtures of gamma distributions (Conjecture 5.2). We also report on numerical estimations of the identifiability degree (Section 5.2), and compute some Euclidean distance degrees (Section 5.3).

Applications of commutative algebra techniques. Moment varieties and other varieties arising in algebraic statistics provide interesting yet concrete algebraic structures to which one can apply techniques from commutative algebra. In this paper, the most important techniques are from the theories of determinantal ideals and Cohen–Macaulay rings, and we provide necessary general background and references in Section 2, as well as further references as needed. We refer the reader to [7; 12; 18] for further general background on these topics. The Cohen–Macaulay property is particularly important in many of our results. Cohen–Macaulay varieties can be seen as "mildly singular" varieties with good homological properties and for which intersection multiplicity behaves well. For instance, they are equidimensional varieties with no embedded components. We also use polarization and Stanley–Reisner theory from combinatorial commutative algebra in Section 4. We hope that our proof techniques inspire further applications of commutative algebra to algebraic statistics.

Outline of the paper. The rest of the paper is organized as follows. In Section 2, we provide preliminary definitions regarding moment varieties, as well as some background on the commutative algebra results and techniques that we will use. In Section 3, we study moment varieties for the inverse Gaussian distribution. In Section 4, we study moment varieties for the gamma distribution, as well as for certain one-parameter specializations. In Section 5 of the paper, we outline a number of interesting directions for future work on the inverse Gaussian and gamma distributions.

Code for the computational experiments presented in this paper is publicly available at

https://github.com/oskarhenriksson/moment-varieties-inverse-gaussian-and-gamma.

Notation and conventions. Throughout the paper, $\mathcal{V}(I)$ denotes the variety associated to a homogeneous ideal I; we consider this variety as a projective variety unless otherwise specified. For a variety V in a given projective or affine space, we use $\mathcal{I}(V)$ to denote the vanishing ideal. Finally, for a matrix H with entries in a ring, we use $I_t(H)$ to denote the ideal generated by all $(t \times t)$-minors of H.

2. Preliminaries

In this section, we provide some background on moment varieties, as well as preliminaries on the techniques and results from commutative algebra that will be used in the rest of the paper.

2.1. ***Moment and cumulant varieties.*** Let X be a univariate random variable parametrized by

$$\theta = (\theta_1, \dots, \theta_n).$$

The *moment-generating function* is given by

$$M(t, \theta) := \mathbb{E}(e^{tX}) = \sum_{r=0}^{\infty} \frac{m_r(\theta) t^r}{r!},$$

and we recall from the Introduction that the *d-th moment variety* $\mathcal{M}_d \subseteq \mathbb{P}^d$ is defined as the Zariski closure of the image of the map

$$\mathbb{C}^n \to \mathbb{P}^d, \quad \theta \mapsto [m_0(\theta) : m_1(\theta) : \cdots : m_d(\theta)].$$

It is typically the case that if d is large enough, the map $\mathbb{C}^n \to \mathcal{M}_d$ is generically one-to-one (meaning that the system (1-1) has a unique complex solution for generic $\widetilde{m} \in \mathcal{M}_d$), and we say that we have *rational identifiability*. A weaker condition for the map is *algebraic identifiability*, meaning that it is generically finite-to-one (in the sense that the system (1-1) has finitely many complex solutions for generic $\widetilde{m} \in \mathcal{M}_d$). By the theorem of the dimension of fibers, we have algebraic identifiability if and only if $\dim(\mathcal{M}_d) = n$.

For many univariate classical distributions, such as those listed in [10, Table 2], rational identifiability follows immediately from the polynomials $m_r(\theta)$. However, it is generally much more difficult to know when k-mixtures of such distributions are algebraically or rationally identifiable from their moments, and we discuss this further in Section 5.

Example 2.1. The univariate Gaussian distribution is parametrized by the mean μ and variance σ^2, and its moment-generating function is given by

$$M(t,\mu,\sigma^2)=\sum_{r=0}^{\infty}\frac{m_r(\mu,\sigma^2)}{r!}t^r=\exp(t\mu)\cdot\exp\left(\tfrac{1}{2}\sigma^2t^2\right).$$

The polynomials in μ,σ^2 for the moments of order up to 4 are

$$m_0=1,\quad m_1=\mu,\quad m_2=\mu^2+\sigma^2,\quad m_3=\mu^3+3\mu\sigma^2,\quad m_4=\mu^4+6\mu^2\sigma^2+3\sigma^4.$$

These polynomials parametrize the fourth moment variety $\mathcal{M}_4\subseteq\mathbb{P}^4$ of the univariate Gaussian distribution. By [2, Proposition 2], the homogeneous prime ideal of $\mathcal{M}_4$ is the ideal $I\subseteq\mathbb{C}[m_0,\ldots,m_4]$ generated by the 3×3 minors of the matrix

$$\begin{pmatrix}0 & m_0 & 2m_1 & 3m_2\\ m_0 & m_1 & m_2 & m_3\\ m_1 & m_2 & m_3 & m_4\end{pmatrix}.$$

Definition 2.2. The r-th *cumulant* $k_r(\theta)$ of the probability distribution X is defined by the cumulant-generating function

$$K(t,\theta):=\log(M(t,\theta))=\sum_{r=1}^{\infty}\frac{k_r(\theta)t^r}{r!},$$

and the d-th *cumulant variety* $\mathcal{K}_d$ is defined as the Zariski closure of the image of the map

$$\mathbb{C}^n\to\mathbb{C}^d,\quad \theta\mapsto(k_1(\theta),\ldots,k_d(\theta)).$$

The relation $K(t,\theta)=\log(M(t,\theta))$ gives rise to a nonlinear change of coordinates

$$\mathbb{C}^d\to\mathbb{C}^d,\quad (m_1,\ldots,m_d)\mapsto(m_1,m_2-m_1^2,m_3-3m_2m_1+2m_1^3,\ldots)$$

that expresses the moments of order $\le d$ as polynomials in the cumulants of order $\le d$, and vice versa. It restricts to an isomorphism of affine varieties $\mathcal{M}_d\cap\{m_0\neq0\}\to\mathcal{K}_d$, and hence a birational map $\mathcal{M}_d\dashrightarrow\mathcal{K}_d$, and is an example of a *Cremona transformation* in the language of [14]. For a more extensive background on the statistical properties of cumulants, we refer to [29, §2].

2.2. ***Background on commutative algebra.*** We collect here the main commutative algebra results that will be used throughout the paper. For the rest of this subsection, $\dim(S)$ for a ring S denotes its Krull dimension.

We begin by recalling some fundamental properties of Cohen–Macaulay rings. For a more comprehensive introduction to this class of rings, we refer the reader to [8, Chapter 2]. Cohen–Macaulay rings are important because they exhibit good homological behavior that simplifies their algebraic and geometric structure. These rings have a well-behaved dimension theory (they are defined by having depth equal to Krull dimension), which ensures they have the following desirable properties.

Proposition 2.3. *Let R be a Noetherian Cohen–Macaulay ring. Then*:

(i) *R is equidimensional, i.e., all the minimal prime ideals of R have same dimension.*

(ii) *R has no embedded primes. In particular, it is unmixed, i.e., all the associated prime ideals of R have same dimension.*

(iii) $\dim(R/I) = \dim R - \mathrm{ht}(I)$ *for any ideal $I \subset R$, where* $\mathrm{ht}(I)$ *denotes the height or codimension of I.*

(iv) *r is a non-zero-divisor in R if and only if* $\dim(R/(r)) = \dim R - 1$.

Cohen–Macaulay rings are ubiquitous in algebra and geometry. They frequently arise as coordinate rings of many important varieties, such as smooth algebraic varieties and certain rings of invariants.

Example 2.4. The following are some standard examples and nonexamples of Cohen–Macaulay rings:

- Regular (local) rings are Cohen–Macaulay. In particular, a standard graded polynomial ring over a field is Cohen–Macaulay.
- Complete intersection rings are Cohen–Macaulay.
- Let $R = \mathbb{C}[x, y]/(x^2y, x)$. Its associated primes are (x, y) and (x). In particular, R has an embedded component, and so condition (ii) above fails. Therefore, R is not Cohen–Macaulay.
- Let $R = \mathbb{C}[x, y, z]/(xy, xz)$. Its associated primes are (x) and (y, z). In this case R, has no embedded component, but it is not equidimensional and so condition (i) above fails. Therefore, R is not Cohen–Macaulay.

In this paper, we often work with *determinantal rings*, i.e., rings of the form $R = S/I$, where S is a standard graded polynomial ring and $I = I_t(H)$ is an ideal generated by the t-minors of a $k \times \ell$ matrix H with entries in S. The following result is key for proving that moment varieties of inverse Gaussian and gamma distributions have determinantal realizations (Theorems 3.1 and 4.1).

Proposition 2.5 [7, Corollary 3.4.10]. *Let $R = S/I$ be a determinantal ring, where $I = I_t(H)$ with H a matrix of size $k \times \ell$. If I has codimension $(k - t + 1)(\ell - t + 1)$, then R is Cohen–Macaulay.*

For computational purposes, it is often useful to have a Gröbner basis of the homogeneous ideals defining our moment varieties. In Propositions 3.4 and 4.4, we prove that the natural generators of these determinantal ideals form Gröbner bases with respect to a suitable term order. To do so, we use some standard results from commutative algebra that we collect in the next two lemmas.

Lemma 2.6 [7, Proposition 1.4.7]. *Let S be a standard graded polynomial ring and let $I \subseteq S$ be a homogeneous ideal. Consider the ideal $J = (\mathrm{in}(f_1), \ldots, \mathrm{in}(f_r))$, where $f_1, \ldots, f_r$ are homogeneous elements of I. Then $J = \mathrm{in}(I)$ if the following conditions hold*:

- $\dim(S/I) = \dim(S/J)$.
- *J is unmixed.*
- $\deg(J) = \deg(I)$.

Lemma 2.7 [19, §15.1.1]. *Let I be a monomial ideal and let M be a minimal generator of I of degree s. Write $I = I' + (M)$ for an ideal I'. Then, there is a short exact sequence of graded modules and degree 0 maps*

$$0 \to S/(I' : M)[-s] \to S/I' \to S/I \to 0. \tag{2-1}$$

Since the Hilbert series is additive in short exact sequences, this shows that

$$\mathrm{HS}_{S/I}(t) = \mathrm{HS}_{S/I'}(t) - \mathrm{HS}_{S/(I':M)[-s]}(t). \tag{2-2}$$

A special class of determinantal ideals that arise in our study of moment varieties are those coming from 1-generic (Hankel) matrices (see e.g., Propositions 4.8 and 4.9). The notion of 1-generic matrix was introduced by Eisenbud in [18]. A matrix H is called 1-*generic* if it has no generalized entries which are zero, meaning that no entries are identically zero after arbitrary scalar row and column operations. Examples of 1-generic matrices are generic matrices and Hankel matrices (up to some unit coefficients). The ideals of maximal minors of these matrices are well-understood, and the following result of Eisenbud shows that such ideals are prime and of expected codimension.

Proposition 2.8 [18, Theorem 1]. *Let H be a 1-generic matrix with entries in a ring S, and of dimension $k \times \ell$ matrix with $k \leq \ell$. Then $I_k(H)$ is prime and of codimension $\ell - k + 1$. In particular, $I_k(H)$ is Cohen–Macaulay.*

3. Inverse Gaussian moment varieties

We now study moment varieties of the *inverse Gaussian* distribution, which is also sometimes called the *Wald* distribution. We show that the ideals of its moment varieties are determinantal. Based on this, we use results from the theory of determinantal ideals to find degrees and Gröbner bases, and we compute the singular loci of the moment varieties.

Despite being less known than the classical Gaussian distribution, the inverse Gaussian has significant applications in modeling different phenomena. While the Gaussian distribution is widely known for its symmetry and versatility in modeling various data, the inverse Gaussian distribution offers a valuable alternative for situations where asymmetry and right-skewed tails are more appropriate (see Figure 1), such as in modeling lifetime phenomena. We refer the reader to [34] for further background.

Similar to the Gaussian distribution, the inverse Gaussian is defined by two parameters, namely μ and λ. However, its density function is supported on $(0, +\infty)$ and it is given by

$$f_{\mu,\lambda}(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\, \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right).$$

Here, $\mu > 0$ represents the mean while $\lambda > 0$ is called *shape parameter* since it affects how peaked and right-tailed the density function is. Unlike the Gaussian distribution, λ does not directly coincide with the variance, though these two quantities are related by the following formula:

$$\mathrm{Var}(X) = \frac{\mu^3}{\lambda}.$$

The main motivation behind the definition of this distribution comes from Brownian motion. In this context, the inverse Gaussian works as a dual for the Gaussian distribution, in the sense that the inverse Gaussian models the first passage time of the Brownian motion to a certain fixed level, whereas the Gaussian models the Brownian motion's level at a fixed time [30, Chapter 13].

The moment-generating function of the inverse Gaussian is given by

$$M(t) = \exp\left(\frac{\lambda}{\mu}\left(1 - \sqrt{1 - \frac{2\mu^2 t}{\lambda}}\right)\right),$$

and the moments are given by the recursive formula

$$m_0 = 1, \quad m_1 = \mu, \quad m_i = \frac{2i-3}{\lambda}\mu^2 m_{i-1} + \mu^2 m_{i-2} \quad \text{for } i \geq 2. \tag{3-1}$$

Although these functions are rational in λ, one could equivalently parametrize the inverse Gaussian distribution using $1/\lambda$ to obtain moment functions that are polynomial in the parameters for the distribution.

The above system of equations parametrizes the moment variety $\mathcal{M}_d^{\mathrm{IG}}$ of the inverse Gaussian of degree d. Using this parametrization we prove that, for fixed d, the moment variety has a determinantal realization and it is Cohen–Macaulay.

Theorem 3.1. *Let $d \geq 3$. The homogeneous prime ideal of the inverse Gaussian moment variety $\mathcal{M}_d^{\mathrm{IG}}$ is generated by $\binom{d-1}{3}$ cubics and $\binom{d-1}{2}$ quartics, given by the maximal minors of the following $(3 \times d)$-matrix*:

$$H_d^{\mathrm{IG}} = \begin{pmatrix} m_0^2 & m_0 & m_1 & m_2 & m_3 & \cdots & m_{d-2} \\ 0 & m_1 & 3m_2 & 5m_3 & 7m_4 & \cdots & (2d-3)m_{d-1} \\ m_1^2 & m_2 & m_3 & m_4 & m_5 & \cdots & m_d \end{pmatrix}.$$

Furthermore, $\mathcal{M}_d^{\mathrm{IG}}$ is Cohen–Macaulay.

Proof. The vector $(\mu^2, \mu^2/\lambda, -1) \neq 0$ is in the left kernel of the matrix H_d^{IG}, so the 3-minors of H_d^{IG} vanish on $\mathcal{M}_d^{\mathrm{IG}}$. Thus,

$$J_d := I_3(H_d^{\mathrm{IG}}) \subseteq \mathcal{I}(\mathcal{M}_d^{\mathrm{IG}}).$$

This gives that $\dim(\mathcal{V}(J_d)) \geq \dim(\mathcal{M}_d^{\mathrm{IG}}) = 3$ as affine varieties. Let $\mathrm{in}(J_d)$ denote the initial ideal of J_d with respect to the reverse lexicographic ordering; then we have that

$$\mathrm{in}(J_d) \supseteq (m_1^4, m_2^3, \ldots, m_{d-2}^3), \tag{3-2}$$

so $\dim(\mathcal{V}(J_d)) = \dim(\mathcal{V}(\mathrm{in}(J_d))) \leq d + 1 - (d-2) = 3$. Thus, we have proved that

$$\dim(\mathcal{V}(J_d)) = \dim(\mathcal{M}_d^{\mathrm{IG}}) = 2$$

as projective varieties. This shows that the ring $R = S/J_d$, where $S = \mathbb{C}[m_0, \ldots, m_d]$, has expected Krull dimension $d + 1 - (d - 3 + 1) = 3$, so it is Cohen–Macaulay by Proposition 2.5. In particular, the ideal J_d has no embedded components by Proposition 2.3.

We want to prove that J_d is prime, so that $J_d = \mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$. This is equivalent to proving that R is a domain. Let $\Delta = m_0^2 m_1$ be the 2-minor in the upper-left corner, and consider the localization R_Δ. To prove that R is a domain, it is enough to prove that Δ is a non-zero-divisor in R, and that R_Δ is a domain.

To prove that Δ is a non-zero-divisor in R, we prove that m_0 and m_1 are both non-zero-divisors in R. Since R is Cohen–Macaulay, by Proposition 2.3 it suffices to prove that

$$\dim(R/(m_0)) = \dim(R/(m_1)) = \dim(R) - 1 = 2.$$

Observe that

$$R/(m_0) \cong \frac{\mathbb{C}[m_1, \ldots, m_d]}{I_3((H_d^{\mathrm{IG}})|_{m_0=0})},$$

where

$$(H_d^{\mathrm{IG}})|_{m_0=0} = \begin{pmatrix} 0 & 0 & m_1 & m_2 & \cdots & m_{d-2} \\ 0 & m_1 & 3m_2 & 5m_3 & \cdots & (2d-3)m_{d-1} \\ m_1^2 & m_2 & m_3 & m_4 & \cdots & m_d \end{pmatrix}.$$

If we consider the 3-minor on the first three columns, we get $m_1 = 0$. Shifting to the next adjacent minor we get $m_2 = 0$. Thus, iterating, we eventually find that $\mathcal{V}(I_3(H_d^{\mathrm{IG}})) \cap \{m_0 = 0\}$ is given by the following projective curve:

$$m_0 = m_1 = \cdots = m_{d-3} = m_{d-2} = 0.$$

So, $R/(m_0)$ has Krull dimension 2. A similar argument shows that $\mathcal{V}(I_3(H_d^{\mathrm{IG}})) \cap \{m_1 = 0\}$ is a union of two projective curves:

$$m_0 = m_1 = \cdots = m_{d-3} = m_{d-2} = 0,$$
$$m_1 = m_2 = \cdots = m_{d-3} = m_{d-1} = 0.$$

So, $R/(m_1)$ has Krull dimension 2. This proves that Δ is a non-zero-divisor in R.

It remains to prove that R_Δ is a domain. Since Δ is a non-zero-divisor and R is Cohen–Macaulay, we get that $\dim(R_\Delta) = \dim(R) = 3$. In fact, in order to see that $\dim(R_\Delta) = \dim(R)$, it is sufficient to note that Δ is not nilpotent. This is because R is a finitely generated $\mathbb{C}$-algebra and therefore the nilradical of R is equal to the Jacobson radical of R. In particular, if Δ is not nilpotent, then it is not in some maximal ideal $\mathfrak{m}$ of R. Since R is Cohen–Macaulay, the length of a maximal chain of prime ideals ending at $\mathfrak{m}$ is $\dim(R) = 3$, and since $\Delta \notin \mathfrak{m}$, this chain is preserved in R_Δ. Thus, $\dim(R_\Delta) = \dim(R) = 3$. We now claim that R_Δ is a domain. Note that for $3 \le i \le d$, one can inductively use the 3-minor of H_d^{IG} on columns $1, 2, i$ to see that $\overline{m_i} \in R_\Delta$ can be expressed as

$$\overline{m_i} = \bar{f}/\overline{m_0^2 m_1} \quad \text{for some } f \in \mathbb{C}[m_0, m_1, m_2].$$

Thus, R_Δ is a finitely generated $\mathbb{C}$-algebra:

$$R_\Delta = \mathbb{C}\big[\overline{m_0}, \overline{m_1}, \overline{m_2}, \overline{1/m_0^2 m_1}\big] \subseteq R.$$

In particular, we get an isomorphism $\mathbb{C}[w, x, y, z]/(w^2xz - 1) \cong R_\Delta$, induced by the $\mathbb{C}$-algebra homomorphism

$$\varphi : \mathbb{C}[w, z, y, z] \to R_\Delta, \quad w \mapsto \overline{m_0}, \quad x \mapsto \overline{m_1}, \quad y \mapsto \overline{m_2}, \quad z \mapsto \overline{1/m_0^2 m_1}.$$

If $(w^2xz - 1) \subsetneq \ker(\varphi)$, then $\dim \mathcal{V}(\ker(\varphi)) \leq 2$ since $(w^2xz - 1)$ is prime. This is because the height of $\ker(\varphi)$ would need to be strictly greater than the height of $(w^2xz - 1)$, and $\mathbb{C}[w, x, y, z]$ is Cohen–Macaulay so $\dim \mathcal{V}(I) + \mathrm{ht}(I) = 4$ for every ideal I. However, $\dim(R_\Delta) = 3$. Thus it must be that $\ker(\varphi) = (w^2xz - 1)$, and therefore R_Δ is a domain, which completes the proof. □

As an immediate consequence of the previous result, we get two corollaries on important invariants of $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$. These corollaries concern the theory of free resolutions, Eagon–Northcott complexes, and (Castelnuovo–Mumford) regularity. We just recall here some basic facts about these topics and refer the interested reader to [12, Chapter 2.C] for further details.

The regularity of a homogeneous ideal I is a measure of its computational complexity. If I is generated in a single degree D, having a *linear resolution* means that I has regularity equal to D. This is a desirable property for an ideal, since it means that the ideal is "computationally simple". When I is a homogeneous ideal generated in different degrees, the closest notion to linearity is *componentwise linearity*, meaning that for each degree D, the ideal generated by homogeneous elements of I of degree D, i.e., $\langle I_D \rangle$, has a linear resolution (see [26]). In this case, the regularity is equal to the maximum degree of the minimal generators of I.

Corollary 3.2. *The ideal $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$ has a componentwise linear (minimal) free resolution given by the Eagon–Northcott complex. In particular*, $\mathrm{reg}(\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})) = 4$.

Proof. This follows directly from [32, Proposition 4.5] and the above discussion. □

Corollary 3.3. *The degree is given by* $\deg(\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})) = (d-1)^2$.

Proof. Since the ideal is Cohen–Macaulay and of codimension $d - 2$, the degree of the variety is the elementary symmetric polynomial of degree $d-2$ in d unknowns, evaluated at $e_1 = 2, e_2 = 1, \ldots, e_d = 1$, where e_i is the degree of the i-th column's entries. This yields $\deg(\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})) = (d-1)^2$. The previous formula is known as the Thom–Porteous–Giambelli formula. We refer the reader to [21] for further details on the degrees of determinantal varieties. □

It turns out that the determinantal realization of $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$ from Theorem 3.1 provides a Gröbner basis with respect to the reverse lexicographic ordering.

Proposition 3.4. *The 3×3 minors of H_d^{IG} form a Gröbner basis for $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$ with respect to any antidiagonal term order (for example, the reverse lexicographic ordering). Furthermore, the Hilbert series of $S/\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$ is given by*

$$\frac{1 + (d-2)t + \binom{d-1}{2}t^2 + \binom{d-1}{2}t^3}{(1-t)^3}.$$

Proof. Let $I = \mathrm{in}(\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}}))$, and let $J := (\mathrm{in}([i, j, k]) \mid 1 \le i \le j \le k \le d)$ where $[i, j, k]$ represents the 3-minor of H_d^{IG} on columns i, j and k. Then J is given by

$$J = (m_2, \ldots, m_{d-2})^3 + m_1^2(m_1, \ldots, m_{d-2})^2.$$

We want to apply Lemma 2.6 to conclude that $J = I$. Note that J is a primary ideal since every variable m_i that divides a generator of J appears with some power $m_i^{p_i} \in J$ [18, Excercise 3.6], and $\sqrt{J} = (m_1, \ldots, m_{d-2})$. In particular, J is unmixed and

$$\dim(S/J) = \dim(S/\sqrt{J}) = 3 = \dim(S/I).$$

We are left to show that $\deg(J) = \deg(I)$. We already know that $\deg(I) = (d-1)^2$ by Corollary 3.3. As for $\deg(J)$, we will apply Lemma 2.7 to the ideal J and a monomial generator M of degree 4, i.e., $M \in \{m_1^4, m_1^3 m_2, m_1^2 m_2^2, \ldots, m_1^2 m_{d-2}^2\}$.

Iteratively applying Lemma 2.7, we get

$$\mathrm{HS}_{S/J}(t) = \mathrm{HS}_{S/(m_2,\ldots,m_{d-2})^3}(t) - \binom{d-1}{2} \mathrm{HS}_{S/(m_2,\ldots,m_{d-2})[-4]}(t), \tag{3-3}$$

where the factor $\binom{d-1}{2}$ appears in the previous identity because we are applying (2-1) to each of the $\binom{d-1}{2}$ degree 4 monomials arising from a 3-minor of H_d^{IG} involving the first column. Since

$$S/(m_2, \ldots, m_{d-2}) \cong \mathbb{C}[m_0, m_1, m_{d-1}, m_d],$$

we have

$$\mathrm{HS}_{S/(m_2,\ldots,m_{d-2})[-4]}(t) = \frac{t^4}{(1-t)^4}. \tag{3-4}$$

As for $A := S/(m_2, \ldots, m_{d-2})^3$, we can consider its quotient by the regular sequence given by $m_0,\ m_1,\ m_{d-1},\ m_d$, that is,

$$\bar{A} = \frac{S}{\big((m_2, \ldots, m_{d-2})^3 + (m_0, m_1, m_{d-1}, m_d)\big)}.$$

Since $S/(m_2, \ldots, m_{d-2})^3$ is Cohen–Macaulay of Krull dimension 4, it follows that $\bar{A}$ is a 0-dimensional ring, usually called the *Artinian reduction* of A. The Hilbert series of $\bar{A}$ is given by the Hilbert polynomial of $\mathbb{C}[m_2, \ldots, m_{d-2}]/(m_2, \ldots, m_{d-2})^3$, which is

$$HS_{\bar{A}}(t) = 1 + (d-3)t + \binom{d-1}{2}t^2.$$

Since the numerator of the Hilbert series remains unchanged when taking the Artinian reduction (one can also see this by applying Lemma 2.7 for each of the linear generators m_0, m_1, m_{d-1}, m_d), we get

$$HS_{S/(m_2,\ldots,m_{d-2})^3}(t) = \frac{1 + (d-3)t + \binom{d-1}{2}t^2}{(1-t)^4}. \tag{3-5}$$

Hence, by identity (3-3),

$$\mathrm{HS}_{S/J}(t) = \frac{1 + (d-3)t + \binom{d-1}{2} - \binom{d-1}{2}t^4}{(1-t)^4} = \frac{1 + (d-2)t + \binom{d-1}{2}t^2 + \binom{d-1}{2}t^3}{(1-t)^3}.$$

This yields

$$\deg(J) = 1 + (d-2) + 2\binom{d-1}{2} = (d-1)^2 = \deg(I).$$

Applying Lemma 2.6, we get that $J = \mathrm{in}(\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}}))$, which concludes the proof.

Since the Hilbert series does not change when taking the initial ideal, we have also found the Hilbert series of $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$. □

Example 3.5. For $d = 3$, we have the following principal homogeneous prime ideal:

$$\mathcal{I}(\mathcal{M}_3^{\mathrm{IG}}) = (-m_0^2 m_1 m_3 + 3m_0^2 m_2^2 - 3m_0 m_1^2 m_2 + m_1^4).$$

Its Hilbert series is $(1+t+t^2+t^3)/(1-t)^3$, and its degree is 4.

For $d = 4$, we have the following homogeneous prime ideal:

$$\mathcal{I}(\mathcal{M}_4^{\mathrm{IG}}) = \begin{pmatrix} -m_0^2 m_1 m_3 + 3m_0^2 m_2^2 - 3m_0 m_1^2 m_2 + m_1^4, \\ -3m_0^2 m_2 m_4 + 5m_0^2 m_3^2 - 5m_1^3 m_3 + 3m_1^2 m_2^2, \\ -m_0^2 m_1 m_4 + 5m_0^2 m_2 m_3 - 5m_0 m_1^2 m_3 + m_1^3 m_2, \\ -3m_0 m_2 m_4 + 5m_0 m_3^2 + m_1^2 m_4 - 6m_1 m_2 m_3 + 3m_2^3 \end{pmatrix}.$$

Its Hilbert series is $(1+2t+3t^2+3t^3)/(1-t)^3$, and its degree is 9.

We end our discussion about the algebraic properties of $\mathcal{M}_d^{\mathrm{IG}}$ by computing its singular locus.

Proposition 3.6. *The singular locus of $\mathcal{M}_d^{\mathrm{IG}}$ is given by the line $m_0 = m_1 = \cdots = m_{d-2} = 0$ and the point $m_1 = m_2 = \cdots = m_d = 0$ in $\mathbb{P}^d$.*

Proof. We begin by noting that the open affine patch $\mathcal{M}_d^{\mathrm{IG}} \cap \{m_0 m_1 \neq 0\}$ is included in the smooth locus. To see this, note that the 3-minors involving the first two columns of H_d^{IG} together give rational expressions for $m_3, m_4, \ldots, m_d$ in the variables m_0, m_1, m_2, with denominators that are monomials in m_0 and m_1, which gives an isomorphism of varieties $(\mathbb{C}^*)^2 \times \mathbb{C} \cong \mathcal{M}_d^{\mathrm{IG}} \cap \{m_0 m_1 \neq 0\}$. The complement of this affine patch in $\mathcal{M}_d^{\mathrm{IG}}$ is the union of two projective lines:

$$\mathcal{L}_1 : m_0 = m_1 = \cdots = m_{d-2} = 0, \quad \mathcal{L}_2 : m_1 = m_2 = \cdots = m_{d-1} = 0.$$

The singular locus of $\mathcal{M}_d^{\mathrm{IG}}$ must therefore be contained in $\mathcal{L}_1 \cup \mathcal{L}_2$.

Let $\mathcal{J}$ be the $\binom{d}{3} \times (d+1)$ Jacobian of the maximal minors of H_d^{IG}. The singular locus is precisely the set of points where the rank of $\mathcal{J}$ is less than $\mathrm{codim}(\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})) = d-2$.

If we evaluate $\mathcal{J}$ at $\mathcal{L}_1$, we get a matrix with just $d-3$ nonzero entries. To see this, note that the only minors that can give a nonzero contribution to the Jacobian are those that have a term that is at most linear in the variables $m_0, \ldots, m_{d-2}$. The only such maximal minor arises from picking column indices $\{i, d-1, d\}$ for $i \in \{2, \ldots, d-2\}$, and its only contribution to the Jacobian will come from the term $(2d-3)m_{d-1}^2 m_{i-2}$.

If, on the other hand, we evaluate $\mathcal{J}$ at $\mathcal{L}_2 \setminus \mathcal{L}_1 = \mathcal{L}_2 \cap \{m_0 \neq 0\}$, the only minors of H_d^{IG} that make a nonzero contribution to the Jacobian, are those that have total degree at most 1 in the variables $m_1, m_2, \ldots, m_{d-1}$. This corresponds to the column indices $\{1, i, d\}$ for $i \notin \{1, d\}$ (which gives a minor

containing the term $(2i-3)m_0^2 m_i m_d$) or $\{2, i, d\}$ for $i \notin \{1, 2, d\}$ (which gives a minor containing the term $(2i-3)m_0 m_i m_d$). From this, we see that $\mathcal{J}$ evaluated at $\mathcal{L}_2 \cap \{m_0 \neq 0\}$ has rank $d-2$ for $m_d \neq 0$, and vanishes completely for $m_d = 0$. □

Remark 3.7. Based on Corollary 3.3, it might be tempting to believe that $\mathcal{M}_d^{\mathrm{IG}}$ is a Roman surface (a generic projection of a Veronese variety). However, the above proposition shows that this cannot be the case, since the singular locus of a Roman surface consists of three points, while the singular locus of $\mathcal{M}_d^{\mathrm{IG}}$ is given by a curve and a point.

Before closing this section, we briefly turn our attention to the cumulants of the inverse Gaussian distribution. The cumulant-generating function is

$$K(t) = \log(M(t)) = \frac{\lambda}{\mu}\left(1 - \sqrt{1 - \frac{2\mu^2 t}{\lambda}}\right),$$

from which we obtain the following formula for the cumulants (where !! denotes the double factorial):

$$\kappa_r = \frac{(2r-3)!!\,\mu^{2r-1}}{\lambda^{r-1}}.$$

Similar to the moment variety, the ideal of the cumulant variety $\mathcal{K}_d^{\mathrm{IG}} \subseteq \mathbb{C}^d$ can be realized as a determinantal ideal.

Proposition 3.8. *The prime ideal $\mathcal{I}(\mathcal{K}_d^{\mathrm{IG}})$ is generated by the $\binom{d-1}{2}$ quadrics given by the 2-minors of the matrix*

$$K_d = \begin{pmatrix} -\kappa_1 & \kappa_2 & 3\kappa_3 & \cdots & (2d-3)\kappa_{d-1} \\ \kappa_2 & \kappa_3 & \kappa_4 & \cdots & \kappa_d \end{pmatrix}.$$

Furthermore, $\mathcal{I}(\mathcal{K}_d^{\mathrm{IG}})$ is Cohen–Macaulay, its degree is $d-1$, and has a Gröbner basis given by the 2-minors of K_d with respect to any antidiagonal term order.

Proof. The recursive relation $\kappa_r = \frac{\mu^2}{\lambda}(2r-3)\kappa_{r-1}$ gives that $(\mu^2/\lambda, -1) \in \ker(K_d)$. We conclude that $J_d := I_2(K_d) \subseteq \mathcal{I}(\mathcal{K}_d^{\mathrm{IG}})$. The matrix H_d^{IG} is a 1-generic Hankel matrix, so $J_d = I_2(K_d)$ is prime with the expected dimension $d - ((d-1) - 2 + 1) = 2$ by Proposition 2.8. In particular, it is Cohen–Macaulay. Since we have a parametrization $(\mathbb{C}^*) \times \mathbb{C} \to \mathcal{K}_d^{\mathrm{IG}}$ with algebraic identifiability, we also know that $\mathcal{I}(\mathcal{K}_d^{\mathrm{IG}})$ is prime of dimension 2. We conclude that $\mathcal{I}(\mathcal{K}_d^{\mathrm{IG}}) = I_2(K_d)$. Similarly to $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$, the degree of $\mathcal{I}(\mathcal{K}_d^{\mathrm{IG}})$ can be computed via Thom–Porteous–Giambelli's formula. The Gröbner basis of determinantal ideals of symmetric matrices, and in particular Hankel matrices, was given by Conca (see [15; 16]). □

Since the ideal obtained above is homogeneous, we note that the cumulant variety is a cone. In particular, it is a (scaled) rational normal curve when viewed as a projective curve in $\mathbb{P}^{d-1}$. An interesting question is whether any features of $\mathcal{K}_d^{\mathrm{IG}}$ can be pulled back by the birational moments-to-cumulants map $\mathcal{M}_d^{\mathrm{IG}} \dashrightarrow \mathcal{K}_d^{\mathrm{IG}}$. For example, we hope that this simple geometric model given by the cumulants could help prove nondefectiveness of secants of moment varieties, which corresponds in statistics to algebraic identifiability of mixtures of inverse Gaussian distributions (see Section 5.1 for further details).

4. Gamma moment varieties

In this section, we study moment varieties $\mathcal{M}_d^\Gamma$ for the gamma distribution. We find their defining ideals and use this to find their degrees and singular loci. We also discuss statistically important special cases of the gamma distribution whose moment varieties are rational normal curves embedded in $\mathcal{M}_d^\Gamma$.

4.1. ***Gamma distribution.*** The gamma distribution is commonly used to model physical and economic processes, especially in relation to arrival or waiting times. The density function is supported on $(0, +\infty)$, and involves the gamma function Γ (see Figure 1). The distribution has two parameters, and there are two commonly used parametrizations:

(1) The shape-scale parametrization is given by a shape parameter $k > 0$ and a scale parameter $\theta > 0$. The density function is given by

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}.$$

(2) The other parametrization is given by a shape parameter $\alpha > 0$ and a rate parameter $\beta > 0$. This parametrization is related to the previous one via $\alpha = k$ and $\beta = 1/\theta$.

The moment-generating function is $M(t) = (1 - \theta t)^{-k}$, and the moments are given by

$$m_i = \theta^i \prod_{j=0}^{i-1} (k + j) = \theta m_{i-1}(k + (i - 1)).$$

Let $\mathcal{M}_d^\Gamma$ denote the d-th moment variety of the gamma distribution. The affine part $\mathcal{M}_d^\Gamma \cap \{m_0 \neq 0\}$ has previously been studied in [22, §3.2]. Here, we study the full projective variety. We begin by finding its defining ideal. The following proof is similar to the proof of the defining ideal for the inverse Gaussian moment variety (Theorem 3.1), and so we omit some details.

Theorem 4.1. *Let $d \geq 3$. The homogeneous prime ideal of the gamma moment variety $\mathcal{M}_d^\Gamma$ is generated by the $\binom{d}{3}$ cubics given by the maximal minors of the following $(3 \times d)$-matrix*:

$$H_d^\Gamma = \begin{pmatrix} 0 & m_1 & 2m_2 & 3m_3 & \cdots & (d-1)m_{d-1} \\ m_0 & m_1 & m_2 & m_3 & \cdots & m_{d-1} \\ m_1 & m_2 & m_3 & m_4 & \cdots & m_d \end{pmatrix}.$$

Furthermore, $\mathcal{M}_d^\Gamma$ is Cohen–Macaulay.

Proof. Notice that the vector $(k\theta, \theta, -1) \neq 0$ is in the left kernel of the matrix H_d^Γ, so the 3-minors of H_d^Γ vanish on $\mathcal{M}_d^\Gamma$. Thus,

$$J_d := I_2(H_d^\Gamma) \subseteq \mathcal{I}(\mathcal{M}_d^\Gamma).$$

Then, $\mathcal{V}(J_d) \supseteq \mathcal{M}_d^\Gamma$ and $\dim(\mathcal{V}(J_d)) \geq \dim(\mathcal{M}_d^\Gamma) = 2$ as projective varieties. On the other hand, if we fix an antidiagonal term order, then

$$(m_i m_j m_k \mid 0 < i \leq j < k < d) \subseteq \operatorname{in}(J_d).$$

We observe that $\sqrt{(m_i m_j m_k \mid 0 < i \le j < k < d)} = (m_i m_j \mid i \ne j, 0 < i, j < d)$, so

$$\dim(\mathcal{V}(J_d)) = \dim(\mathcal{V}(\mathrm{in}(J_d))) \le \dim(\mathcal{V}(m_i m_j \mid i \ne j, 0 < i, j < d)) = 2.$$

Thus, we have proved that $\dim(\mathcal{V}(J_d)) = \dim(\mathcal{M}_d^\Gamma) = 2$ as projective varieties. This shows that the ring $R = S/J_d$ has expected Krull dimension $d + 1 - (d - 3 + 1) = 3$, so it is Cohen–Macaulay by Proposition 2.5. In particular, the ideal J_d has no embedded components by Proposition 2.3.

We now show that J_d is prime. Let $\Delta = m_0 m_1$ be the 2-minor in the upper-left corner (after multiplying by -1), and consider the localization R_Δ. As in the inverse Gaussian case, it is sufficient to prove that Δ is a non-zero-divisor in R, and that R_Δ is a domain.

To prove that Δ is a non-zero-divisor in R, we prove that m_0 and m_1 are both non-zero-divisors in R. By Proposition 2.3, since R is Cohen–Macaulay, it suffices to prove that

$$\dim(R/(m_0)) = \dim(R/(m_1)) = \dim(R) - 1 = 2.$$

Observe that

$$R/(m_0) \cong \frac{\mathbb{C}[m_1, \ldots, m_d]}{I_3((H_d^\Gamma)|_{m_0=0})},$$

where

$$(H_d^\Gamma)|_{m_0=0} = \begin{pmatrix} 0 & m_1 & 2m_2 & 3m_3 & \cdots & (d-1)m_{d-1} \\ 0 & m_1 & m_2 & m_3 & \cdots & m_{d-1} \\ m_1 & m_2 & m_3 & m_4 & \cdots & m_d \end{pmatrix}.$$

If we consider the 3-minor on the first three columns, we get $m_1^2 m_2 = 0$. Thus we have two possibilities, either $m_1 = 0$ or $m_2 = 0$. Then, for $i = 2, \ldots, d-2$, the i-th antidiagonal is given by $(i+1)m_i^2 m_{i+1}$, and we inductively see that either m_i or m_{i+1} must be equal to zero. This shows that $\mathcal{V}(I_3(H_d^\Gamma)) \cap \{m_0 = 0\}$ is a union of $d - 1$ curves given by

$$m_0 = m_1 = \cdots = \widehat{m_i} = \cdots = m_{d-1} = 0, \qquad i = 1, \ldots, d-1,$$

where the hat denotes omission. Thus, $R/(m_0)$ has affine dimension 2. A similar argument shows that $R/(m_1)$ has affine dimension 1. This shows that Δ is a non-zero-divisor.

The proof that R_Δ is a domain is similar to the inverse Gaussian case. One can use the 3-minors of H_d^Γ to see that $R_\Delta = \mathbb{C}[\overline{m_0}, \overline{m_1}, \overline{m_2}, \overline{1/m_0 m_1}] \subseteq R$, and so $R_\Delta \cong \mathbb{C}[w, x, y, z]/(wxz - 1)$. Since $(wxz - 1)$ is prime, this shows that R_Δ is a domain, which completes the proof. □

By Theorem 4.1, we have that $\mathcal{I}(M_d^\Gamma)$ is a Cohen–Macaulay ideal generated by the maximal minors of a matrix with linear entries. Therefore, $\mathcal{I}(M_d^\Gamma)$ has a linear minimal free resolution given by the Eagon–Northcott complex, and the degree of the gamma moment variety can be calculated via [12, Proposition 2.15].

Corollary 4.2. *The ideal $\mathcal{I}(\mathcal{M}_d^\Gamma)$ has a linear (minimal) free resolution given by the Eagon–Northcott complex. In particular, the regularity of the ideal is $\mathrm{reg}(\mathcal{I}(\mathcal{M}_d^\Gamma)) = 3$ and the degree of the gamma moment surface is $\deg(\mathcal{I}(\mathcal{M}_d^\Gamma)) = \binom{d}{2}$.*

We now show that the 3-minors of H_d^Γ form a Gröbner basis for $\mathcal{I}(\mathcal{M}_d^\Gamma)$ with respect to the reverse lexicographic order. The outline of the proof is as in the inverse Gaussian case. Let J denote the ideal generated by the initial terms of the minors. We show in the following two lemmas that J is unmixed, and that $\deg(J) = \binom{d}{2}$. We then apply Lemma 2.6 to conclude that J is indeed the initial ideal of $\mathcal{I}(\mathcal{M}_d^\Gamma)$.

To show that J is unmixed, we apply the technique of polarization to obtain a squarefree monomial ideal (see [20] for background on polarization). We then use Stanley–Reisner theory to study the associated primes of the polarization.

Lemma 4.3. *The ideal J generated by initial terms of the 3-minors of H_d^Γ is unmixed.*

Proof. Define the ideal J' of $\mathbb{C}[m_1, \dots, m_{d-1}]$ by eliminating m_0 and m_d:

$$J' = J \cap \mathbb{C}[m_1, \dots, m_{d-1}].$$

If J' is unmixed, then J is as well. We show that every associated prime of J' has height $d-2$. Since J' is a monomial ideal, its associated primes are also monomial. We therefore show that if $P \in \operatorname{Ass}(J')$, then P is of the form

$$P = (m_{i_1}, \dots, m_{i_{d-2}} \mid 1 \le i_1 < \cdots < i_{d-2} \le d-1).$$

Let $\mathcal{P}(J')$ be the polarization of J', so

$$\begin{aligned}\mathcal{P}(J') &= (m_{i,1}m_{i,2}m_{j,1},\ m_{i,1}m_{j,1}m_{k,1} \mid 1 \le i < j < k \le d-1) \\ &\subseteq S = \mathbb{C}[m_{i,1}, m_{i,2}, m_{d-1,1} \mid 1 \le i \le d-2].\end{aligned}$$

Suppose Q is an associated prime of $\mathcal{P}(J')$, and define the set $A \subseteq [d-1]$ to be

$$A = \{i \mid m_{i,j} \in Q\}.$$

By the correspondence between associated primes of an ideal and those of its polarization, to show that J' is unmixed, we therefore want to show that $\#A = d-2$ [20, Corollary 2.6].

Let Δ denote the Stanley–Reisner complex associated to J'. Then a prime ideal $Q \subseteq S$ is an associated prime of $\mathcal{P}(J')$ if and only if the variables of S that are not generators of Q form a facet of Δ.

First, suppose for contradiction that $\#A = d-1$. Let F be the facet corresponding to Q. Then F must only involve variables $m_{i,j}$ with $1 \le i \le d-2$, and it cannot contain both $m_{i,1}$ and $m_{i,2}$ for any i. Since F is a facet, the number of i such that $m_{i,1} \in F$ cannot be greater than 2; otherwise, this would correspond to the generator of $\mathcal{P}(J'_d)$ given by $m_{i,1}m_{j,1}m_{k,1}$. We then have the following cases:

(1) Suppose $m_{i,1}, m_{j,1} \in F$ with $i < j$. This implies that

$$F = \{m_{i,1}, m_{j,1}, m_{k_1,2}, \dots, m_{k_{d-3},2} \mid k_\ell \in [d-2] \setminus \{i, j\}\}.$$

Then, F cannot be a facet, as $F \cup \{m_{j,2}\}$ is a face of Δ.

(2) If there is at most one element of the form $m_{i,1}$ in F, then F cannot be a facet, as it is contained in $F \cup \{m_{d-1,1}\}$, which is also a face of Δ.

Now suppose for contradiction that $\#A < d-2$, and let F be the corresponding facet to Q. Then one of the following must be true:

(1) $m_{i,1}, m_{i,2}, m_{d-1,1} \in F$ for some $1 \le i \le d-2$, or

(2) $m_{i,1}m_{i,2}, m_{j,1}, m_{j,2} \in F$ for some $1 \le i < j \le d-2$.

In both cases, we obtain a contradiction that F is a face of Δ, since $\mathcal{P}(J')$ has generators of the form $m_{i,1}m_{i,2}m_{j,1}$ for all $1 \le i < j \le d-1$.

We therefore see that any associated prime of J' has height $d-2$, and so J is unmixed. □

We now prove that J is indeed the initial ideal by computing the degree of J; the argument is similar to the proof of Proposition 3.4. This result also addresses [22, Conjecture 3.2.5].

Proposition 4.4. *The 3-minors of H_d^Γ form a Gröbner basis for the homogeneous prime ideal of $\mathcal{M}_d^\Gamma$ with respect to any antidiagonal term order. Moreover, the Hilbert series of $S/\mathcal{I}(\mathcal{M}_d^\Gamma)$ is given by*

$$\frac{1+(d-2)t+\binom{d-1}{2}t^2}{(1-t)^3}.$$

Proof. Let J be the ideal generated by initial terms of 3-minors of H_d^Γ. We have that $J \subseteq \mathrm{in}(\mathcal{I}(\mathcal{M}_d^\Gamma))$, and Lemma 4.3 shows that J is unmixed. Therefore, to show equality of the ideals, it suffices by Lemma 2.6 to show that the degrees of the two ideals are equal.

We iteratively apply Lemma 2.7. Each time, we remove a minimal generator of J of the form $m_i^2 m_j$, with $1 \le i < j \le d-1$; there are $\binom{d-1}{2}$ such generators. Then the colon ideal is always of the form $(m_{i_1}, \ldots, m_{i_{d-3}})$, and the final J' ideal that we end up with is

$$J' = (m_i m_j m_k \mid 1 \le i < j < k \le d-1).$$

Thus, the Hilbert series of J is given by

$$\begin{aligned}\mathrm{HS}_{S/J}(t) &= \mathrm{HS}_{S/J'}(t) - \binom{d-1}{2}t^3 \mathrm{HS}_{S/(m_1,\ldots,m_{d-3})}(t)\\ &= \frac{1+(d-3)t+\binom{d-2}{2}t^2}{(1-t)^4} - \frac{\binom{d-1}{2}t^3}{(1-t)^4}\\ &= \frac{1+(d-2)t+\binom{d-1}{2}t^2}{(1-t)^3}.\end{aligned}$$

Substituting $t=1$ in the numerator, we see that

$$\deg(J) = 1+(d-2)+\binom{d-1}{2} = \binom{d}{2}.$$

By Corollary 4.2, this is also the degree of $\mathcal{I}(\mathcal{M}_d^\Gamma)$, and so this completes the proof. □

Example 4.5. For $d=3$, we have the following principal homogeneous prime ideal:

$$\mathcal{I}(\mathcal{M}_3^\Gamma) = (-m_0m_1m_3 + 2m_0m_2^2 - m_1^2m_2).$$

Its Hilbert series is $(1+t+t^2)/(1-t)^3$, and its degree is 3.

For $d = 4$, we have the following homogeneous prime ideal:

$$\mathcal{I}(\mathcal{M}_4^\Gamma) = \begin{pmatrix} -m_0m_1m_3 + 2m_0m_2^2 - m_1^2m_2, \\ -m_0m_1m_4 + 3m_0m_2m_3 - 2m_1^2m_3, \\ -2m_0m_2m_4 + 3m_0m_3^2 - m_1m_2m_3, \\ -m_1m_2m_4 + 2m_1m_3^2 - m_2^2m_3 \end{pmatrix}.$$

Its Hilbert series is $(1 + 2t + 3t^2)/(1 - t)^3$, and its degree is 6.

We now describe the singular locus of $\mathcal{M}_d^\Gamma$.

Proposition 4.6. *The singular locus of $\mathcal{M}_d^\Gamma$ is given by two points in $\mathbb{P}^d$, defined by the ideals*

$$(m_0, m_1, \ldots, m_{d-1}) \quad \textit{and} \quad (m_1, m_2, \ldots, m_d).$$

Proof. Analogously to the proof of Proposition 3.6, one can construct an isomorphism

$$(\mathbb{C}^*)^2 \times \mathbb{C} \cong \mathcal{M}_d^\Gamma \cap \{m_0m_1 \neq 0\}$$

by considering the maximal minors involving the first two columns of H_d^Γ. The complement of this affine open patch in $\mathcal{M}_d^\Gamma$ is the union of $2(d-1)$ projective curves:

- $m_0 = m_1 = \cdots \widehat{m_i} \cdots = m_{d-2} = m_{d-1} = 0$ for $i \in \{1, \ldots, d-1\}$, and
- $m_1 = m_2 = \cdots \widehat{m_i} \cdots = m_{d-1} = m_d = 0$ for $i \in \{2, \ldots, d\}$.

Let $\mathcal{J}$ denote the Jacobian of the maximal minors of H_d^Γ. If we evaluate $\mathcal{J}$ at one of the curves of the first kind described above, then the only minors that make a contribution to $\mathcal{J}$ are those that contain a term that is at most linear in $m_0, \ldots, \widehat{m_i}, \ldots, m_{d-1}$. This happens precisely for column indices $\{i, i+1, j\}$ for $j \notin \{i, i+1\}$, which gives a minor with a term $im_i^2m_{j+1}$, and $\{i+1, j, d\}$ for $j \notin \{i+1, d\}$, which gives a minor with a term divisible by $m_im_jm_d$. Hence, the resulting evaluated matrix $\mathcal{J}$ vanishes completely when $m_i = 0$, and has rank at least $d-2$ if $m_i \neq 0$. The situation is analogous when $\mathcal{J}$ is evaluated on a curve of the second kind. □

We conclude this section by pointing out that, similar to the cumulant variety of the inverse Gaussian distribution, the cumulant variety of the gamma distribution is a cone. This result first appeared in [22, Proposition 3.2.1]. We include it here for the sake of completeness.

Proposition 4.7. *The ideal $\mathcal{I}(\mathcal{K}_d^\Gamma)$ is generated by the $\binom{d-1}{2}$ quadrics given by the 2-minors of the matrix*

$$K_d = \begin{pmatrix} \kappa_1 & \kappa_2 & \frac{\kappa_3}{2} & \cdots & \frac{\kappa_{d-1}}{(d-2)!} \\ \kappa_2 & \frac{\kappa_3}{2} & \frac{\kappa_4}{3!} & \cdots & \frac{\kappa_d}{(d-1)!} \end{pmatrix}.$$

Furthermore, $\mathcal{I}(\mathcal{K}_d^\Gamma)$ is Cohen–Macaulay, its degree is $d-1$, and its Gröbner basis is given by the 2-minors that generate the ideal.

Similar to the case of the inverse Gaussian distribution, having this simpler geometric model given by the cumulants could help in understanding the geometry of secants of moment varieties of the gamma distribution.

4.2. ***Exponential and chi-squared distributions.*** We now consider two important special cases of the gamma distribution: the exponential and chi-squared distributions. Both distributions are given by one parameter, and have one-dimensional moment varieties. We show that the moment varieties are in fact rational normal curves.

The *exponential distribution* is the distribution of the time between events in a process, where events occur continuously and independently at a constant average rate $\lambda > 0$. The only parameter is λ, and the moment-generating function is

$$M(t) = (1 - \lambda t)^{-1}.$$

The moments are given by

$$m_i = i!\,\lambda^i = i\lambda m_{i-1}.$$

This is a specialization of the gamma distribution with shape $k = 1$ and rate $\theta = \lambda$ under the shape-scale parametrization.

Let $\mathcal{M}_d^{\exp}$ denote the d-th moment variety of the exponential distribution. The following result shows that the variety is a rational normal curve.

Proposition 4.8. *The homogeneous ideal $\mathcal{I}(\mathcal{M}_d^{\exp})$ is generated by the maximal minors of*

$$H_d^{\exp} = \begin{pmatrix} m_0 & 2m_1 & 3m_2 & 4m_3 & \cdots & dm_{d-1} \\ m_1 & m_2 & m_3 & m_4 & \cdots & m_d \end{pmatrix}.$$

In particular, $\mathcal{M}_d^{\exp}$ is a rational normal curve.

Proof. Notice that the vector $(\lambda, -1) \neq 0$ is in $\ker(H_d^{\exp})$, so the 2-minors of $H_d^{\exp}$ vanish on $\mathcal{M}_d^{\exp}$. Thus,

$$J_d := I_2(H_d^{\exp}) \subseteq \mathcal{I}(\mathcal{M}_d^{\exp}).$$

Moreover, $H_d^{\exp}$ is a 1-generic Hankel matrix, so $J_d = I_2(H_d^{\exp})$ is prime of expected codimension $d-1$ by Proposition 2.8. In particular, it defines a Cohen–Macaulay ring of dimension 2. As projective varieties,

$$\dim(\mathcal{V}(J_d)) = 1 = \dim(\mathcal{M}_d^{\exp}).$$

Hence, $\mathcal{V}(J_d) = \mathcal{M}_d^{\exp}$. Since J_d is prime, we get $J_d = \mathcal{I}(\mathcal{M}_d^{\exp})$. □

We now consider the *chi-squared distribution*, which is the distribution of a sum of the squares of $k \geq 1$ independent standard Gaussian variables. The only parameter is k, and the moment-generating function is

$$M(t) = (1 - 2t)^{-k/2}.$$

The moments are given by

$$m_i = k(k+2)\cdots(k+2i-2) = m_{i-1}(k+2i-2) = m_{i-1}k + m_{i-1}(2i-2).$$

This is a specialization of the gamma distribution with shape $k/2$ and scale 2.

Let $\mathcal{M}_d^{\chi}$ denote the moment variety. Following the proof of Proposition 4.8, we attempt to find generators of the ideal $\mathcal{I}(\mathcal{M}_d^{\chi})$ by using the recursive formula given above, which gives that the maximal

minors of

$$H_d^\chi = \begin{pmatrix} 0 & 2m_1 & 4m_2 & 6m_3 & \cdots & (2d-2)m_{d-1} \\ m_0 & m_1 & m_2 & m_3 & \cdots & m_{d-1} \\ m_1 & m_2 & m_3 & m_4 & \cdots & m_d \end{pmatrix}$$

vanish on $\mathcal{M}_d^\chi$. This matrix is similar to the one for the gamma distribution, and the dimension of the projective variety $\mathcal{V}(I_3(H_d^\chi))$ is $(d+1)-(d-3+1)-1=2$. However, the dimension of the projective variety $\mathcal{M}_d^\chi$ is 1. Thus, the 3-minors of H_d^χ are not sufficient to generate the ideal of $\mathcal{M}_d^\chi$.

Instead, we follow the technique used in [22, §3.3]. Here, the author uses a change in coordinates from moments to powers of the parameter to obtain the defining ideal of the Poisson moment variety. Let $s_{d,i}$ denote the (signed) Stirling numbers of the first kind, and $S_{d,i}$ denote the Stirling numbers of the second kind. Recall that the Stirling numbers of the first kind satisfy

$$x(x-1)(x-2)\cdots(x-n+1) = \sum_{k=0}^{n} s_{n,k} x^k,$$

and that they are related to the Stirling numbers of the second kind by the following property. Let $(f_n)_{n=0}^\infty$ and $(g_n)_{n=0}^\infty$ be two sequences of complex numbers; then

$$g_d = \sum_{i=0}^{d} S_{d,i} f_i = \sum_{i=1}^{d} S_{d,i} f_i \quad \Longleftrightarrow \quad f_d = \sum_{i=0}^{d} s_{d,i} g_i = \sum_{i=1}^{d} s_{d,i} g_i.$$

Note that $s_{n,0} = S_{n,0} = 0$ for all $n \geq 1$. Using the formula for the d-th moment, we see that

$$m_d = \sum_{i=0}^{d} (-2)^{d-i} s_{d,i} k^i,$$

and so applying the relationship between Stirling numbers of the first and second kind with $f_d = m_d/(-2)^d$ and $g_d = (-k/2)^d$, we have that

$$k^d = \sum_{i=0}^{d} (-2)^{d-i} S_{d,i} m_i.$$

Proposition 4.9. *Let $d \geq 2$. The homogeneous prime ideal of the chi-squared moment variety $\mathcal{M}_d^\chi$ is generated by the 2×2 minors of the $(2 \times d)$-matrix*

$$H_d^\chi = \begin{pmatrix} m_0 & m_1 & m_2 - 2m_1 & \cdots & \sum_{i=0}^{d-1} (-2)^{d-i-1} S_{d-1,i} m_i \\ m_1 & m_2 - 2m_1 & m_3 - 6m_2 + 4m_1 & \cdots & \sum_{i=0}^{d} (-2)^{d-i} S_{d,i} m_i \end{pmatrix}.$$

In particular, $\mathcal{M}_d^\chi$ is a rational normal curve.

Proof. Consider the algebra homomorphism $\varphi : \mathbb{C}[m_0, \ldots, m_d] \to \mathbb{C}[x_0, \ldots, x_d]$ defined by the linear map

$$m_j \mapsto \begin{cases} \sum_{i=0}^{j} (-2)^{j-i} s_{j,i} x_i & \text{if } j \geq 1, \\ x_0 & \text{if } j = 0. \end{cases}$$

This linear map can be given by an upper triangular matrix and is invertible, with inverse

$$x_j \mapsto \begin{cases} \sum_{i=0}^{j} (-2)^{j-i} S_{j,i} m_i & \text{if } j \geq 1, \\ m_0 & \text{if } j = 0. \end{cases}$$

Thus, φ is an algebra isomorphism, and induces an isomorphism of varieties $\Phi : \mathbb{P}^d \to \mathbb{P}^d$, where the coordinates of the domain are given by the x_i's, and the coordinates of the codomain are given by the m_i's.

Under this map, a point $[1 : m_1 : \cdots : m_d] \in \mathcal{M}_d^\chi \cap \{m_0 = 1\}$ is mapped to $[1 : x_1 : x_1^2 : \cdots : x_1^d]$. Thus, the homogenous prime ideal of the projective variety $\Phi(\mathcal{M}_d^\chi)$ is generated by the 2×2 minors of the matrix

$$H_d' = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{d-1} \\ x_1 & x_2 & x_3 & \cdots & x_d \end{pmatrix},$$

and we conclude that the homogeneous prime ideal of $\mathcal{M}_d^\chi$ is given by $\varphi^{-1}(I_2(H_d')) = I_2(H_d^\chi)$. □

An interesting direction for future work is to classify algebraic and geometric properties of moment varieties for polynomial distributions. For example, in [22], the author studies moment varieties for the Poisson distribution, which is also a one-parameter distribution. These varieties were shown to be rational normal curves. Given our results on the moment varieties for the exponential and chi-squared distributions, this raises the following problem.

Problem 4.10. Classify projective curves that can appear as moment varieties of one-parameter distributions.

It is also worth noticing that all moment varieties studied in this paper are Cohen–Macaulay. Moreover, to the best of our knowledge, there is no known distribution for which the moment varieties fail to have this property (for example, those studied in [2; 22] are all Cohen–Macaulay). This leads to the following problem.

Problem 4.11. Classify all Cohen–Macaulay moment varieties from polynomial distributions.

5. Future directions

In this section, we discuss two possible directions for the future algebraic study of moment varieties arising from inverse Gaussian and gamma distributions: identifiability of mixtures, and the complexity of the optimization problem that arises in the overdetermined scenario.

5.1. ***Algebraic identifiability of mixtures.*** It is often the case that real-world datasets exhibit multimodal behavior or heterogeneity. To model these datasets, statisticians use *mixtures of distributions* which, thanks to their inherent flexibility, allow modeling complex and diverse patterns within data. Mixtures of distributions are simply convex combinations of probability distributions (see Figure 3 for an example).

Geometrically, the moment variety of a distribution that is the mixture of k independent observations of a distribution with moment variety $\mathcal{M}_d$ corresponds to the k-th *secant variety* of $\mathcal{M}_d$, which we denote by $\mathrm{Sec}_k(\mathcal{M}_d)$ (see, e.g., [6] for an overview on secant varieties). If the moments depend on n parameters $\theta = (\theta_1, \ldots, \theta_n)$, it holds that $\mathrm{Sec}_k(\mathcal{M}_d)$ is given by the projective closure in $\mathbb{P}^d$ of the image of the map

$$(\mathbb{C}^n)^k \times \mathbb{C}^{k-1} \to \mathbb{C}^d, \quad \left((\theta^{(1)}, \ldots, \theta^{(k)}), \alpha\right) \mapsto \left(\sum_{i=1}^{k-1} \alpha_i m_r(\theta^{(i)}) + \left(1 - \sum_{i=1}^{k-1} \alpha_i\right) m_r(\theta^{(k)})\right)_{r=1,\ldots,d}. \quad (5\text{-}1)$$

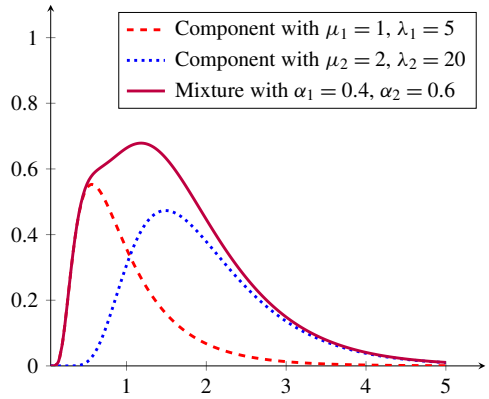


Figure 3. Probability density for a mixture of two inverse Gaussians.

The dimension of $\mathrm{Sec}_k(\mathcal{M}_d)$ is bounded above by the dimension of the domain of (5-1), so we have

$$\dim(\mathrm{Sec}_k(\mathcal{M}_d)) \leq \min\{d, nk+k-1\}.$$

If equality holds we say that $\mathrm{Sec}_k(\mathcal{M}_d)$ is *nondefective*; otherwise, we say that $\mathcal{M}_d$ is *k-defective*. In particular, if $\mathrm{Sec}_k(\mathcal{M}_d)$ is nondefective and $d \geq nk+k-1$, then the parameters of each component, as well as the mixing coefficients, are algebraically identifiable from the $nk+k-1$ first moments. This motivates the study of *nondefectiveness* of $\mathrm{Sec}_k(\mathcal{M}_d)$. It was proven in [4] that $\mathrm{Sec}_k(\mathcal{M}_d)$ is nondefective for all $k, d \geq 2$ in the Gaussian case, and a natural question is whether the same is true for the distributions considered in this paper.

For the exponential and chi-squared distributions, nondefectiveness (and, in fact, rational identifiability) follows directly from the determinantal realizations found in Section 4.2.

Proposition 5.1. *For k-mixtures of the exponential distribution or the chi-squared distribution, we have rational identifiability from the first* $2k-1$ *moments.*

Proof. Propositions 4.8 and 4.9 give that both $\mathcal{M}_d^{\exp}$ and $\mathcal{M}_d^{\chi}$ are linearly isomorphic to the rational normal curve in $\mathbb{P}^d$. The result now follows from the fact that a general point on the k-th secant variety of the rational normal curve in $\mathbb{P}^d$ lies on a unique k-secant if $2k \leq d+1$ (since any $d+1$ points on such a curve are linearly independent [25, Example 1.14]). □

The cases of the inverse Gaussian and gamma distributions are more complicated. An elementary first observation one can make is that the dimension of $\mathrm{Sec}_k(\mathcal{M}_d)$ is bounded below by the rank of the Jacobian of the parametrization (5-1) at any point in the domain. By computing this rank for randomly chosen points with rational entries in exact arithmetic in `Maple`, we are able to computationally verify nondefectiveness for all $d, k \leqslant 100$. Based on this, we conjecture the following.

Conjecture 5.2. $\mathrm{Sec}_k(\mathcal{M}_d^{\mathrm{IG}})$ and $\mathrm{Sec}_k(\mathcal{M}_d^{\Gamma})$ are nondefective for all $d \geq 2$ and $k \geq 2$.

A helpful step towards a proof of this conjecture is Terracini's classification of k-defective surfaces (see, e.g., [4, Theorem 8; 13, Theorem 1.3]), which establishes that the only k-defective surfaces are either cones or quadratic Veronese embeddings of a rational normal surface in $\mathbb{P}^k$. Similar to in [4], we can rule out the latter possibility.

Proposition 5.3. *Let $\mathcal{M}_d$ be either $\mathcal{M}_d^{\mathrm{IG}}$ or $\mathcal{M}_d^{\Gamma}$. If $\mathcal{M}_d$ is k-defective, then it is contained in a cone over a curve, with apex a linear space of dimension at most $k-2$.*

Proof. Assume for a contradiction that $\mathcal{M}_d$ is not contained in such a cone. Then, by Terracini's classification of defective surfaces, it should be the quadratic Veronese embedding of a rational normal surface in $\mathbb{P}^k$, and in particular have at most one singular point. But this is a contradiction, since the singular locus of $\mathcal{M}_d$ contains a line in the case of the inverse Gaussian distribution by Proposition 3.6, and two points in the case of the gamma distribution by Proposition 4.6. □

Ruling out the possibility of $\mathcal{M}_d$ being contained in a cone of the type described in Proposition 5.3 is ongoing work with Kristian Ranestad.

5.2. ***Identifiability degree.*** Suppose Conjecture 5.2 is true, so that $\dim(\mathrm{Sec}_k(\mathcal{M}_d))$ equals $\min\{3k-1, d\}$ for both the inverse Gaussian and gamma distribution. Then, for $d = 3k-1$, we have *generically finite fibers* of the parametrization (5-1). The generic cardinality of these fibers is the *identifiability degree*, which measures the complexity of solving the moment equations. Identifiability degrees for mixtures of Gaussians have previously been studied with numerical algebraic geometry tools [3], and here we take a similar approach.

Table 1 contains numerical bounds on the identifiability degree up to the label-swapping symmetry for $d = 3k-1$. The numbers were computed using monodromy in `HomotopyContinuation.jl` (with 10 loops without new solutions as the stopping condition), and certified as lower bounds with the techniques of [9]. When numerically feasible, the pseudo-Segre trace test from [3, §3.3] was used to verify completeness of the solution set, using a tolerance of $\varepsilon = 10^{-12}$.

Many open questions in this direction remain. Apart from computing the very large identifiability degree for $k = 4$ in the inverse Gaussian case, one can also investigate the case where some parameters are fixed, or impose equality of some of the parameters of the components of the mixtures. This has been a successful line of research in the Gaussian case [28; 1]. Another possible direction is to better understand the smaller identifiability degrees already found in Table 1 from the point of view of [33]. For instance, we propose the following problem.

k	2	3	4
Gaussian	9	225	$\geq 10\,350$*
Gamma	≥ 9*	≥ 242*	$\geq 13\,327$
Inverse Gaussian	≥ 24*	≥ 1637	$\geq 20\,000$

Table 1. Identifiability degrees for k-mixtures of three distributions, up to label-swapping symmetry. An asterisk is used to mark bounds for which a trace test indicates sharpness with high probability. The monodromy calculation for $k = 4$ for the inverse Gaussian was manually terminated, and is therefore not expected to be sharp. The three values for the Gaussian distribution were computed in the works [33], [2], and [3], respectively.

Problem 5.4. Find explicit univariate polynomials of degree 9 and 24, analogous to the "Pearson polynomial" found in [33], that account for the identifiability degrees for $k = 2$ in Table 1, e.g., using a similar computer-algebra strategy as in [2, §3].

Finally, we also note that numerical experiments for $k \leq 4$ indicate that a single additional polynomial is enough to decrease the generic number of solutions given in Table 1 to one. This would be in line with [2, Conjecture 15] in the Gaussian case, and we therefore conjecture the following.

Conjecture 5.5. For k-mixtures of the inverse Gaussian distribution or the gamma distribution, we have rational identifiability from the first $3k$ moments.

A natural first step could be to prove the more modest claim of rational identifiability from $3k + 2$ moments, for instance by attempting to adapt the techniques previously employed to prove this in the Gaussian setting [28, Theorem 4.4].

5.3. ***Euclidean distance degree.*** For noisy sample moments, the moment equations might not have any statistically meaningful solutions. In that case, a common approach is to instead solve the optimization problem

$$\min_{\theta \in \mathbb{C}^n} \|m(\theta) - \widetilde{m}\|^2, \tag{5-2}$$

where $m(\theta) = (m_1(\theta), \ldots, m_d(\theta))$ is the first d moments as functions of θ, $\widetilde{m} \in \mathbb{R}^d$ is the first d sample moments, and $d > n$ is such that $\dim(\mathcal{M}_d) = n$. Here, $\|\cdot\|$ denotes the Euclidean norm, but we remark that the computations done in this section can be adapted also to weighted norms $\|\cdot\|_W$ with $\|x\|_W^2 = x^\top W x$ for some (positive semidefinite) matrix $W \in \mathbb{R}^{d \times d}$. Such norms are often considered in the more general framework of the *generalized method of moments* [24].

Compared to maximum likelihood estimation (MLE), which instead minimizes the Kullback–Liebler divergence, the distance-based approach studied here has the advantage of being a *polynomial* optimization problem. Hence, it has a finite *optimization degree*, which is the number of complex critical points for generic sample moments $\widetilde{m}_i$. Understanding this degree is a difficult problem, but can be studied geometrically through the lens of the related notion of the *Euclidean distance degree* (ED degree).

The ED degree of $\mathcal{M}_d$ is the generic number of critical points of the optimization problem

$$\min_{m \in \mathcal{M}_d \cap \{m_0 = 1\}} \|m - \widetilde{m}\|^2. \tag{5-3}$$

This notion was introduced in [17], and has been shown to have a rich geometric meaning in terms of polar degrees [17, Section 5] or Euler characteristics [31]. This might make the ED degree more approachable to compute than the optimization degree of (5-2). Since the optimization degree is bounded below by the product of the identifiability degree and the ED degree of $\mathcal{M}_d$, understanding the ED degree is a natural subproblem.

In this work, we take a first step in this direction by computing the ED degree of $\mathcal{M}_3$ for the inverse Gaussian and gamma distribution. For comparison purposes, we also do this for the Gaussian distribution, which to our knowledge has not been studied from the ED degree perspective before. In all three cases, the affine patch $\{m_0 = 1\}$ is a hypersurface cut out by a single polynomial $f \in \mathbb{C}[m_1, m_2, m_3]$ is given in

Examples 3.5 and 4.5, and [2, Proposition 2], respectively. The ED degree is the number of complex roots of the system

$$f(m) = k\,\nabla f(m) - (m - \widetilde{m}) = 0, \quad m \in \mathbb{C}^3, \quad k \in \mathbb{C} \tag{5-4}$$

for generic parameters $\widetilde{m} \in \mathbb{C}^3$. A standard Gröbner basis calculation in `Oscar.jl` over the field $\mathbb{C}(\widetilde{m}_1, \widetilde{m}_2, \widetilde{m}_3)$ of rational functions in the sample moments proves the following result.

Proposition 5.6. *For the inverse Gaussian distribution*, *the gamma distribution*, *and the Gaussian distribution*, *the ED degree is given by*

$$\text{EDdegree}(\mathcal{M}_3^{\text{IG}}) = 12, \quad \text{EDdegree}(\mathcal{M}_3^{\Gamma}) = 10, \quad \text{EDdegree}(\mathcal{M}_3^{\text{Gaussian}}) = 7.$$

This means that in all three cases, f is generic enough to satisfy the mixed volume bounds given in [11, Theorem 1], but not generic enough to satisfy the bound of [17, Proposition 2.6], which evaluates to 52, 21, and 21, respectively. The determinantal realizations of the three hypersurfaces are also not generic enough in the family of 3×3 matrices to give the value 1 obtained from [17, Example 2.3].

A natural future direction is to compute and analyze the ED degrees for higher values of d, e.g., using numerical algebraic geometry and ideas from the emerging field of *metric algebraic geometry*. It would also be interesting to understand the ED degree of $\text{Sec}_k(\mathcal{M}_d)$; since the implicitization problem for these secant varieties turns out to be very hard in our experience, this is a considerable challenge, even with numerical techniques.

Acknowledgements

The authors thank Bernd Sturmfels for suggesting the problem and for his guidance throughout the project, as well as Carlos Améndola, Oliver Clarke, Francesco Galuppi, Alexandros Grosdos Koutsoumpelias, Julia Lindberg, Lizzie Pratt, Kristian Ranestad, Jose Israel Rodriguez, and Matteo Varbaro for helpful discussions. Part of this research was performed while the authors were visiting the Institute for Mathematical and Statistical Innovation (IMSI) for the Algebraic Statistics long program, which was supported by the National Science Foundation (NSF grant DMS-1929348), and the authors thank the other participants of the program for their feedback and comments. The authors also thank the Max Plank Institute for Mathematics in the Sciences for hosting the first and third authors and providing a productive visit, during which part of this research was also performed. Henriksson was partially supported by the Novo Nordisk project with grant reference number NNF20OC0065582. Seccia was partially supported by SNSF grant TMPFP2-217223. Yu was partially supported by NSF grant DGE-2241144.

References

[1] D. Agostini, C. Améndola, and K. Ranestad, "Moment identifiability of homoscedastic Gaussian mixtures", *Found. Comput. Math.* **21**:3 (2021), 695–724.

[2] C. Améndola, J.-C. Faugère, and B. Sturmfels, "Moment varieties of Gaussian mixtures", *J. Algebr. Stat.* **7**:1 (2016), 14–28.

[3] C. Améndola, J. Lindberg, and J. I. Rodriguez, "Solving parameterized polynomial systems with decomposable projections", 2021. arXiv 1612.08807v2

[4] C. Améndola, K. Ranestad, and B. Sturmfels, "Algebraic identifiability of Gaussian mixtures", *Int. Math. Res. Not.* **2018**:21 (2018), 6556–6580.

[5] Y. Alexandr, J. Kileel, and B. Sturmfels, "Moment varieties for mixtures of products", pp. 53–60 in *Proceedings of the International Symposium on Symbolic & Algebraic Computation* (ISSAC 2023), edited by G. Jeronimo, ACM, New York, 2023.

[6] A. Bernardi, E. Carlini, M. V. Catalisano, A. Gimigliano, and A. Oneto, "The hitchhiker guide to: secant varieties and tensor decomposition", *Mathematics* **6**:12 (2018), art. id. 314.

[7] W. Bruns, A. Conca, C. Raicu, and M. Varbaro, *Determinants, Gröbner bases and cohomology*, Springer, 2022.

[8] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, 2nd ed., Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1998.

[9] P. Breiding, K. Rose, and S. Timme, "Certifying zeros of polynomial systems using interval arithmetic", *ACM Trans. Math. Software* **49**:1 (2023), art. id. 11.

[10] M. Belkin and K. Sinha, "Polynomial learning of distribution families", *SIAM J. Comput.* **44**:4 (2015), 889–911.

[11] P. Breiding, F. Sottile, and J. Woodcock, "Euclidean distance degree and mixed volume", *Found. Comput. Math.* **22**:6 (2022), 1743–1765.

[12] W. Bruns and U. Vetter, *Determinantal rings*, Lecture Notes in Mathematics **1327**, Springer, 1988.

[13] L. Chiantini and C. Ciliberto, "Weakly defective varieties", *Trans. Amer. Math. Soc.* **354**:1 (2002), 151–178.

[14] C. Ciliberto, M. A. Cueto, M. Mella, K. Ranestad, and P. Zwiernik, "Cremona linearizations of some classical varieties", pp. 375–407 in *From classical to modern algebraic geometry*, edited by G. Casnati et al., Springer, 2016.

[15] A. Conca, "Gröbner bases of ideals of minors of a symmetric matrix", *J. Algebra* **166**:2 (1994), 406–421.

[16] A. Conca, "Straightening law and powers of determinantal ideals of Hankel matrices", *Adv. Math.* **138**:2 (1998), 263–292.

[17] J. Draisma, E. Horobe¸t, G. Ottaviani, B. Sturmfels, and R. R. Thomas, "The Euclidean distance degree of an algebraic variety", *Found. Comput. Math.* **16**:1 (2016), 99–149.

[18] D. Eisenbud, "On the resiliency of determinantal ideals", pp. 29–38 in *Commutative algebra and combinatorics* (Kyoto, 1985), edited by M. Nagata and H. Matsumura, Adv. Stud. Pure Math. **11**, North-Holland, Amsterdam, 1987.

[19] D. Eisenbud, *Commutative algebra: with a view toward algebraic geometry*, Graduate Texts in Mathematics **150**, Springer, 1995.

[20] S. Faridi, "Monomial ideals via square-free monomial ideals", pp. 85–114 in *Commutative algebra*, edited by A. Corso et al., Lect. Notes Pure Appl. Math. **244**, Chapman & Hall/CRC, Boca Raton, FL, 2006.

[21] W. Fulton and P. Pragacz, *Schubert varieties and degeneracy loci*, Lecture Notes in Mathematics **1689**, Springer, 1998.

[22] A. Grosdos Koutsoumpelias, *Algebraic methods for the estimation of statistical distributions*, Ph.D. thesis, Osnabrück University, 2020.

[23] A. Grosdos Koutsoumpelias and M. Wageringel, "Moment ideals of local Dirac mixtures", *SIAM J. Appl. Algebra Geom.* **4**:1 (2020), 1–27.

[24] L. P. Hansen, "Large sample properties of generalized method of moments estimators", *Econometrica* **50**:4 (1982), 1029–1054.

[25] J. Harris, *Algebraic geometry: a first course*, Graduate Texts in Mathematics **133**, Springer, 1992.

[26] J. Herzog and T. Hibi, "Componentwise linear ideals", *Nagoya Math. J.* **153** (1999), 141–153.

[27] K. Kohn, B. Shapiro, and B. Sturmfels, "Moment varieties of measures on polytopes", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **21** (2020), 739–770.

[28] J. Lindberg, C. Améndola, and J. I. Rodriguez, "Estimating Gaussian mixtures using sparse polynomial moment systems", 2023. arXiv 2106.15675v2

[29] P. McCullagh, *Tensor methods in statistics*, 2nd ed., Chapman & Hall/CRC, 2018.

[30] A. W. Marshall and I. Olkin, *Life distributions*: *structure of nonparametric*, *semiparametric*, *and parametric families*, Springer, 2007.

[31] L. G. Maxim, J. I. Rodriguez, and B. Wang, "Euclidean distance degree of the multiview variety", *SIAM J. Appl. Algebra Geom.* **4**:1 (2020), 28–48.

[32] U. Nagel and T. Römer, "Criteria for componentwise linearity", *Comm. Algebra* **43**:3 (2015), 935–952.

[33] K. Pearson, "Contributions to the mathematical theory of evolution", *Philos. Trans. R. Soc. A* **185** (1894), 71–110.

[34] V. Seshadri, *The inverse Gaussian distribution*, The Clarendon Press, Oxford University Press, New York, 1993.

[35] L. Wasserman, *All of statistics: a concise course in statistical inference*, Springer, 2004.

Received 2024-02-15. Revised 2024-06-03. Accepted 2024-07-30.

OSKAR HENRIKSSON: oskar.henriksson@math.ku.dk
University of Copenhagen, Copenhagen, Denmark

LISA SECCIA: lisa.seccia@unine.ch
Université de Neuchâtel, Neuchâtel, Switzerland

TERESA YU: twyu@umich.edu
University of Michigan, Ann Arbor, MI, United States

F

Moment varieties of the inverse Gaussian and gamma distributions are nondefective

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Kristian Ranestad
Department of Mathematics
University of Oslo

Lisa Seccia
Mathematical Institute
University of Neuchâtel

Teresa Yu
Department of Mathematics
University of Michigan

Publication details

Preprint: https://doi.org/10.48550/arXiv.2409.18421 (2024)

MOMENT VARIETIES OF THE INVERSE GAUSSIAN AND GAMMA DISTRIBUTIONS ARE NONDEFECTIVE

OSKAR HENRIKSSON, KRISTIAN RANESTAD, LISA SECCIA, TERESA YU

ABSTRACT. We show that the parameters of a k-mixture of inverse Gaussian or gamma distributions are algebraically identifiable from the first $3k-1$ moments, and rationally identifiable from the first $3k+2$ moments. Our proofs are based on Terracini's classification of defective surfaces, careful analysis of the intersection theory of moment varieties, and a recent result on sufficient conditions for rational identifiability of secant varieties by Massarenti–Mella.

1. INTRODUCTION

Secant varieties have played a central role in algebraic geometry since the turn of the 20th century, and have recently also been used to shed light on problems in many areas of applied mathematics, including tensor decomposition [BCCGO18], rigidity theory [CMNT23], and optimization [OTT25].

In this work, we use secant varieties to study the *method of moments* for parameter estimation in statistics. Given a stochastic variable X of a distribution depending on parameters $\theta = (\theta_1, \dots, \theta_n)$, the moments $m_r(\theta) = \mathbb{E}[X^r]$ for $r \in \mathbb{N}$ are often rational functions of the parameters (see, e.g., [BS15]). The goal of the method of moments is to estimate the parameters by solving the system

$$m_r(\theta) = \widehat{m}_r \quad \text{for } r = 1, \dots, d, \tag{1.1}$$

where $\widehat{m}_1, \dots, \widehat{m}_d$ are sample moments computed from a sample of X. Geometrically, this corresponds to studying the fibers of the rational map

$$\mathbb{C}^n \dashrightarrow \mathbb{P}^d, \quad \theta \mapsto [1 : m_1(\theta) : \cdots : m_d(\theta)], \tag{1.2}$$

where the Zariski closure $\mathcal{M}_d$ of the image is the dth *moment variety* of the distribution.

Of fundamental statistical importance is the following identifiability problem: How many moments do we expect to need to include in system (1.1) for it to have finitely many or even a unique solution? We say that we have *algebraic identifiability* if the fibers of $\mathbb{C}^n \dashrightarrow \mathcal{M}_d$ are finite over generic points of $\mathcal{M}_d$ (which is equivalent to $\dim(\mathcal{M}_d) = n$), and we say that we have *rational identifiability* if fibers over generic points of $\mathcal{M}_d$ contain a unique point.

These types of identifiability questions have been studied for many distributions using algebraic techniques. The most well-understood examples come from Gaussian distributions [AFS16, ARS18, AAR21, LAR21, Blo23, BCMO23], and other examples include uniform distributions on polytopes [KSS20], Dirac and Pareto distributions [GKW20], as well as inverse Gaussian distributions and gamma distributions [HSY23].

The connection between moment varieties and secant varieties is the following. Consider a given distribution with n parameters and dth moment variety $\mathcal{M}_d$. Then one can consider a k-mixture of this distribution; it has $kn + k - 1$ parameters, and its dth moment variety is the secant variety $\mathrm{Sec}_k(\mathcal{M}_d)$. This perspective has previously been exploited in the Gaussian case to study the method of moments for their mixtures, starting with the seminal paper [ARS18]. In this paper, we extend and generalize some of those techniques to the gamma and inverse Gaussian distributions, whose moment varieties have more complicated structures [HSY23].

Our first main result concerns algebraic identifiability of k-mixtures. We have algebraic identifiability from the first d moments if and only if

$$\dim(\mathrm{Sec}_k(\mathcal{M}_d)) = kn + k - 1.$$

We approach this through the notion of *non-defectivity* in the theory of secant varieties, where a variety $X \subseteq \mathbb{P}^d$ is said to be *k-nondefective* if $\dim(\mathrm{Sec}_k(X)) = \min\{k\dim(X) + k - 1, d\}$, and *$k$-defective* otherwise.

Theorem 1.1. *Let $\mathcal{M}_d$ be the dth moment variety for the inverse Gaussian or gamma distribution. Then $\mathcal{M}_d$ is k-nondefective for each $d \geq 2$ and $k \geq 2$, in the sense that*

$$\dim(\mathrm{Sec}_k(\mathcal{M}_d)) = \min\{3k - 1, d\}.$$

In particular, we have algebraic identifiability from the first $3k-1$ moments for k-mixtures of the inverse Gaussian distribution and k-mixtures of the gamma distribution.

The defectivity of secant varieties is a well-studied yet difficult problem in algebraic geometry. In particular, the classification of defective Segre–Veronese varieties is a classical problem, with many applications to computer science and statistics, due to such secant varieties being closely related to symmetric tensor decomposition [AH95, Lan12, ABGO24]. Although moment varieties are typically not Segre–Veronese, they are often still determinantal varieties. For example, [AFS16] showed that the dth Gaussian moment variety in $\mathbb{P}^d$ is given by the maximal minors of

$$H_1 = \begin{pmatrix} 0 & x_0 & 2x_1 & 3x_2 & \cdots & (d-1)x_{d-2} \\ x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-1} \\ x_1 & x_2 & x_3 & x_4 & \cdots & x_d \end{pmatrix},$$

and it was shown in [HSY23] that the dth moment variety for the inverse Gaussian in $\mathbb{P}^d$ is defined by the maximal minors of

$$H_2 = \begin{pmatrix} x_0^2 & x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-2} \\ 0 & x_1 & 3x_2 & 5x_3 & 7x_4 & \cdots & (2d-3)x_{d-1} \\ x_1^2 & x_2 & x_3 & x_4 & x_5 & \cdots & x_d \end{pmatrix}.$$

Our strategy for proving Theorem 1.1 is similar to that used in [ARS18] to study k-mixtures of univariate Gaussian distributions, in that it builds on Terracini's classification of defective surfaces, and uses intersection theory to rule out each of the possibilities for defectivity. However, having only linear entries in the matrix H_1 is essential in the argument of [ARS18], as their approach relies on the intersection theory of a smooth surface defined using a Hilbert–Burch matrix. As the matrix H_2 has nonlinear entries, we must develop a more general approach.

With Theorem 1.1 in place, we turn our attention to rational identifiability. For k-mixtures, rational identifiability up to permutation of the mixture components corresponds to proving that a generic point on $\mathrm{Sec}_k(\mathcal{M}_d)$ lies on a unique k-secant. In the theory of secant varieties, this property is commonly referred to as *k-identifiability* of $\mathcal{M}_d$. Sufficient conditions for k-identifiability in terms of non-defectivity was developed in [CM23], and was recently sharpened to conditions that also involve the geometry of the Gauss map in [MM22]. These conditions have been used to show identifiability results in, e.g., rigidity theory [CMNT23], Waring theory [CP24], and for various Gaussian mixture distributions [LAR21, Blo23, BCMO23]. We use it to prove the following result, which is known to be true in the Gaussian case from [LAR21].

Theorem 1.2. *Up to permuting the mixture components, we have rational identifiability from the first $3k+2$ moments for k-mixtures of the inverse Gaussian distribution and k-mixtures of the gamma distribution for any $k \geq 2$.*

Future research directions. Our result on algebraic identifiability is optimal in the sense that it is impossible to have algebraic identifiability from fewer than $3k-1$ moments for dimension reasons. However, for rational identifiability it is still an open question whether it is possible to get rational identifiability from fewer than $3k+2$ moments. Based on numerical experiments, it is conjectured in [LAR21, HSY23] that $3k$ moments is enough in the Gaussian, gamma and inverse Gaussian cases. Another major challenge for the future would be to find effective techniques for proving non-defectivity for determinantal varieties of higher dimensions, where there is no obvious known analog of Terracini's classification.

Organization of the paper. The remainder of the paper is organized as follows. In Section 2, we provide background on the intersection theory needed for our results, and outline our proof strategy for Theorem 1.1. In Sections 3 and 4, we carry out this strategy for the inverse Gaussian and gamma distributions respectively. In Section 5, we deduce Theorem 1.2 on rational identifiability of k-mixtures.

Acknowledgements. The authors are grateful to Elisenda Feliu, Francesco Galuppi, Alexandros Grosdos Koutsoumpelias, Ragni Piene, and Jose Israel Rodriguez for helpful discussions. The authors also thank Bernd Sturmfels and the Max Plank Institute for Mathematics in the Sciences for hosting OH, KR and TY, and providing a productive visit, during which part of this research was performed. Part of this research was also performed while OH, LS, and TY were visiting the Institute for Mathematical and Statistical Innovation (IMSI) for the Algebraic Statistics long program, which was supported by NSF grant DMS-1929348.

LS was partially supported by SNSF grant TMPFP2-217223. TY was partially supported by NSF grant DGE-2241144. OH was partially supported by the Novo Nordisk Foundation grant NNF20OC0065582, as well as the European Union under the Grant Agreement no. 101044561, POSALG. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or European Research Council (ERC). Neither the European Union nor ERC can be held responsible for them.

2. Preliminaries

2.1. **Moment varieties of mixtures.** Real-world datasets often display multimodal behavior or underlying heterogeneity, suggesting that they consist of different subpopulations or patterns. To effectively model such complexity, statisticians often employ *mixtures of a distribution*, which are convex combinations of a given probability distribution.

In this paper, we focus on moment varieties of mixtures. Consider a univariate distribution with parameter space $\Theta \subseteq \mathbb{R}^n$ and probability density function $p\colon \mathbb{R} \times \Theta \to \mathbb{R}$, such that the first d moments $m_1(\theta), \dots, m_d(\theta)$ depend rationally on $\theta \in \Theta$ and parametrize the dth moment variety $\mathcal{M}_d \subseteq \mathbb{P}^d$. Then the *$k$-mixture* of the distribution has parameter space $\Theta^k \times \Delta_{k-1}$ where $\Delta_{k-1} \subseteq \mathbb{R}^k$ is the $(k-1)$-dimensional probability simplex, and its density function is given by

$$f\colon \mathbb{R} \times \Theta^k \times \Delta_{k-1} \to \mathbb{R}, \quad (x, \theta^{(1)}, \dots, \theta^{(k)}, \alpha) \mapsto \sum_{i=1}^{k} \alpha_i p(x, \theta^{(i)}).$$

The dth moment variety of the k-mixture is the Zariski closure of the image of the rational map

$$(\mathbb{C}^n)^k \times \mathcal{V}\Big(\textstyle\sum_{i=1}^k \alpha_i - 1\Big) \dashrightarrow \mathbb{P}^d, \ \ (\theta^{(1)}, \dots, \theta^{(k)}, \alpha) \mapsto \Big[\textstyle\sum_{i=1}^k \alpha_i m_r(\theta^{(i)})\Big]_{r=0,\dots,d}, \tag{2.1}$$

where $\mathcal{V}(\cdot)$ denotes the zero locus of a polynomial. Equivalently, the moment variety of the k-mixture is the kth secant variety $\mathrm{Sec}_k(\mathcal{M}_d)$. Proving that we have algebraic identifiability from the first d moments corresponds to proving that

$$\dim(\mathrm{Sec}_k(\mathcal{M}_d)) = kn + k - 1, \tag{2.2}$$

while proving that we have rational identifiability up to the natural label-swapping action of the symmetric group S_k on $(\mathbb{C}^n)^k \times \mathcal{V}(\sum_{i=1}^k \alpha_i - 1)$ corresponds to proving that a generic point in $\mathrm{Sec}_k(\mathcal{M}_d)$ lies on a unique k-secant.

Previous work on identifiability for mixture distributions has focused on the Gaussian distribution. In [ARS18], the authors prove algebraic identifiability from the first $3k-1$ moments, and in [LAR21], the authors prove rational identifiability from the first $3k+2$ moments.

In what follows, we will study two other distributions that play an important role in statistics: the *inverse Gaussian distribution*, and the *gamma distribution*. Both these distributions are two-dimensional (in the sense that $n = 2$), and we will use results from the theory of secant varieties of surfaces to prove algebraic identifiability from $3k-1$ moments, and rational identifiability from the $3k+2$ moments.

2.2. **Proof strategy.** We now outline our strategy for proving Theorem 1.1 on algebraic identifiability. Recall that for a variety $X \subseteq \mathbb{P}^d$, it holds that

$$\dim(\mathrm{Sec}_k(X)) \leq \min\{k\dim(X) + (k-1), d\},$$

where the upper bound in the right-hand side is called the *expected dimension* of $\mathrm{Sec}_k(X)$. The variety X is said to be *k-nondefective* if the bound is attained, and *k-defective* otherwise.

We will show Theorem 1.1 by proving that $\mathcal{M}_d$ is k-nondefective for all $k \geq 2$ and $d \geq 2$ via the following classical classification result due to Terracini. Here, we use the formulation from [ARS18, Theorem 8] (see also [CC02, Theorem 1.3]).

Theorem 2.1 (Terracini's classification)**.** *Let $X \subseteq \mathbb{P}^d$ be a reduced, irreducible, nondegenerate projective surface. If X is k-defective, then $k \geq 2$ and one of the following two possibilities hold:*

(1) X is the quadratic Veronese embedding of a rational normal surface in $\mathbb{P}^k$ of degree $k-1$.

(2) X is contained in a cone over a curve, with apex a linear space of dimension at most $k-2$.

Furthermore, for general points $p_1, \ldots, p_k$ on X there is a hyperplane section tangent along a curve C that passes through these points. In case (1), the curve C is irreducible; in case (2), the curve C decomposes into k algebraically equivalent curves $C_1, \ldots, C_k$ with $p_i \in C_i$.

For both the inverse Gaussian and gamma distributions, we can rule out case (1) in the Terracini classification based on information about the singular loci. If case (1) were to hold, then X would either be smooth, or singular at only one point, but neither of these hold for our moment varieties [HSY23]. In order to rule out (2), the general strategy will be as follows.

The starting point is to turn the defining parametrization of $\mathcal{M}_d$ by the moments $m_i(\theta)$ into a rational parametrization $\phi\colon \mathbb{P}^2 \dashrightarrow \mathcal{M}_d$ by homogeneous forms $f_i(\theta, s)$ with finitely many indeterminacy points $P_1, \ldots, P_r$, and then form a smooth resolution $\pi\colon \mathcal{S}_d \to \mathbb{P}^d$ of the locus of indeterminacy, such that the parametrization lifts to a morphism $\tilde{\phi}\colon \mathcal{S}_d \to \mathcal{M}_d$ that makes the following diagram commute:

$$\begin{array}{ccc} \mathcal{S}_d & & \\ {\scriptstyle\pi}\downarrow & \searrow^{\tilde{\phi}} & \\ \mathbb{P}^2 & \xrightarrow{\phi} & \mathcal{M}_d\,. \end{array} \tag{2.3}$$

Once we have constructed $\mathcal{S}_d$, we will use intersection-theoretic calculations in the Picard group $\mathrm{Pic}(\mathcal{S}_d)$ of Weil divisors modulo linear equivalence to rule out case (2).

A class that will play a particularly important role is the class H of the strict transform of curves $\mathcal{V}(f_{\mathrm{gen}})$, where f_{gen} is a generic linear combination of the coordinates of the parametrization $\phi = [f_0 : \cdots : f_d]$. This class coincides with the class of the pullback of hyperplane sections of $\mathcal{M}_d \subseteq \mathbb{P}^d$ via $\tilde{\phi}\colon \mathcal{S}_d \to \mathcal{M}_d$. For this class, case (2) in the Terracini classification has the following consequence.

Lemma 2.2 (Lemma 10, [ARS18])**.** *Suppose that $\mathcal{M}_d$ satisfies condition (2) in Theorem 2.1. Then, for any k general points $x_1, \ldots, x_k \in \mathcal{S}_d$, there exist linearly equivalent effective divisors $\mathcal{D}_1 \ni x_1, \ldots, \mathcal{D}_k \ni x_k$, and a hyperplane section of $\mathcal{M}_d$, with pullback $\mathcal{H}$ to $\mathcal{S}_d$, such that*

$$\mathcal{A} = \mathcal{H} - 2\mathcal{D}_1 - \cdots - 2\mathcal{D}_k \tag{2.4}$$

is an effective divisor on $\mathcal{S}_d$.

The strategy is to show that the existence of such divisors $\mathcal{A}$ and $\mathcal{D}_i$ gives rise to a contradiction. We will proceed by casework: the $\mathcal{D}_i$'s are linearly equivalent and therefore their images in $\mathbb{P}^2$ all have the same degree $a \geq 1$, and we will use intersection-theoretic calculations in $\mathrm{Pic}(\mathcal{S}_d)$ to derive a contradiction for any choice of a. One may almost immediately rule out large values of a using the following fact, and thus, the main content of the proofs of Theorems 3.7 and 4.6 is to obtain contradictions for a small, i.e., when $a \in \{1, 2\}$.

Lemma 2.3 (Corollary 7, [ARS18]). *If a surface $X \subseteq \mathbb{P}^d$ is k-defective, then X is k'-defective for some $k' \geq (d-2)/3$.*

Remark 2.4. After reducing to small values of a, we are in some cases able to infer that $d \leq 8$, in which case non-defectivity is known through explicit rank computations from [HSY23, Section 5.1]. Explicit Jacobians that certify the $d \leq 8$ cases can be found in the repository
https://github.com/oskarhenriksson/moment-varieties-inverse-gaussian-and-gamma.

With algebraic identifiability in place, rational identifiability follows from an argument based on the conditions given in [MM22]. This is the subject of Section 5.

2.3. **Construction and properties of the resolution.** We end the section by describing the general structure of the construction of the resolution $\mathcal{S}_d$, and some key facts about the structure of $\mathrm{Pic}(\mathcal{S}_d)$ that are common for both distributions. In Sections 3 and 4, we will describe the particularities of this construction for each of the respective distributions.

The resolution will consist of a sequence of blowups

$$\mathcal{S}_d = \mathcal{S}_{r,\ell_r} \to \cdots \to \mathcal{S}_{i,j} \to \cdots \to \mathbb{P}^2$$

where we, for the ith indeterminacy point P_i, construct a sequence of ℓ_i blowups in the following way (where ℓ_i depends on i and the distribution at hand):

- We start by blowing up the intermediate surface $\mathcal{S}_{i-1,\ell_{i-1}}$ obtained in the previous step (or $\mathbb{P}^2$, when $i=1$) at $P_{i,0} = P_i$. Let $\mathcal{S}_{i,1}$ denote the resulting blowup, $\mathcal{E}_{i,1}$ the exceptional divisor, and $\phi_{i,1} \colon \mathcal{S}_{i,1} \dashrightarrow \mathcal{M}_d$ the lift of the previous map. This map $\phi_{i,1}$ turns out to have a unique indeterminacy point $P_{i,1}$ on $\mathcal{E}_{i,1}$.
- We blow up at $P_{i,1}$. Let $\mathcal{S}_{i,2}$ be the resulting blowup, $\mathcal{E}_{i,2}$ the exceptional divisor, and $\phi_{i,2} \colon \mathcal{S}_{1,2} \dashrightarrow \mathcal{M}_d$ the lift of $\phi_{i,1}$, with unique indeterminacy point $P_{i,2}$ on $\mathcal{E}_{i,2}$.
- Continue in this way, until we blow up at P_{i,ℓ_i-1}, to obtain a surface $\mathcal{S}_{i,\ell_i}$ with exceptional divisor $\mathcal{E}_{i,\ell_i}$, and a lift $\phi_{i,\ell_i} \colon \mathcal{S}_{i,\ell_i} \dashrightarrow \mathcal{M}_d$, such that the map ϕ_{i,ℓ_i} has no further indeterminacy points on $\mathcal{E}_{i,\ell_i}$.

In the jth step for the ith indeterminacy point, we construct $\mathcal{S}_{i,j}$ by picking an affine chart $\mathbb{A}^2$ around $P_{i,j-1}$, which we blow up to $\mathcal{B}_{i,j}$, and we then define $\mathcal{S}_{i,j}$ as the Zariski closure of $\mathcal{B}_{i,j}$ in $\mathcal{S}_{i,j-1} \times \mathbb{P}^1$, and $\pi_{i,j} \colon \mathcal{S}_{i,j} \to \mathcal{S}_{i,j-1}$ as the extension of the blowup map $\mathcal{B}_{i,j} \to \mathbb{A}^2$, so that we get a commutative diagram

$$\begin{array}{ccccc} \mathcal{B}_{i,j} & \hookrightarrow & \mathcal{S}_{i,j} & \subseteq & \mathcal{S}_{i,j-1} \times \mathbb{P}^1 \\ \downarrow & & \downarrow{\scriptstyle \pi_{i,j}} & \searrow{\scriptstyle \phi_{i,j}} & \\ \mathbb{A}^2 & \hookrightarrow & \mathcal{S}_{i,j-1} & \xrightarrow[\phi_{i,j-1}]{} & \mathcal{M}_d. \end{array}$$

We finally let $\mathcal{S}_d = \mathcal{S}_{r,\ell_r}$, and $\tilde{\phi} = \phi_{r,\ell_r}$, and take $\pi \colon \mathcal{S}_d \to \mathbb{P}^2$ to be the composition of all maps $\pi_{i,j}$, giving the commutative diagram (2.3). It follows from the general theory of blowups at points in the projective plane (see, e.g., [Har77, §V.3]) that $\mathrm{Pic}(\mathcal{S}_d)$ is a free abelian group generated by the classes $E_{i,j}$ of the exceptional divisors $\mathcal{E}_{i,j}$ obtained in the construction of $\mathcal{S}_d$, as well as the class L of a line in $\mathbb{P}^2$ pulled back to $\mathcal{S}_d$. Furthermore, the intersection number pairing $\cdot \colon \mathrm{Pic}(\mathcal{S}_d) \times \mathrm{Pic}(\mathcal{S}_d) \to \mathbb{Z}$ is diagonal with

$$L^2 = 1, \qquad E_{i,j}^2 = -1 \text{ for all } i = 1, \ldots, r \text{ and } j = 1, \ldots, \ell_i. \tag{2.5}$$

It furthermore follows from [Har77, Proposition 3.6] that the class $\tilde{C}$ of the strict transform to $\mathcal{S}_d$ of any irreducible curve $\mathcal{C}$ in $\mathbb{P}^2$ can we expressed in terms of these generators as

$$\tilde{C} = \deg(C)L - \sum_{i=1}^{r} \sum_{j=1}^{\ell_i} m_{i,j} E_{i,j}, \tag{2.6}$$

where $m_{i,j}$ is the multiplicity at $P_{i,j-1}$ of the strict transform of C on $\mathcal{S}_{i,j-1}$.

3. The inverse Gaussian distribution

The inverse Gaussian distribution has two parameters μ and λ, and its dth moment variety $\mathcal{M}_d^{\mathrm{IG}} \subseteq \mathbb{P}^d$ is a surface that is the Zariski closure of the image of the map

$$(\mathbb{C}^*) \times \mathbb{C} \to \mathbb{P}^d, \quad (\mu, \lambda) \mapsto [m_0 : \cdots : m_d],$$

where the moments are defined recursively as

$$m_0 = 1, \qquad m_1 = \mu, \qquad m_i = \tfrac{2i-3}{\lambda}\mu^2 m_{i-1} + \mu^2 m_{i-2} \ \text{ for } \ i \geq 2. \tag{3.1}$$

Note that it follows directly from the recursive formula that for $i > 0$,

$$m_i = \frac{\mu^i\, p_{i-1}(\lambda, \mu)}{\lambda^{i-1}},$$

where $p_i(\lambda, \mu)$ is the homogenization of the degree-i Bessel polynomial, with μ as the homogenization variable. For proving non-defectivity, we will use the following basic algebraic and geometric properties of $\mathcal{M}_d^{\mathrm{IG}}$ as a starting point.

Theorem 3.1 (§3, [HSY23]). *Let $d \geq 3$. The homogeneous ideal $\mathcal{I}(\mathcal{M}_d^{\mathrm{IG}})$ is generated by $\binom{d-1}{3}$ cubics and $\binom{d-1}{2}$ quartics, given by the maximal minors of the $(3 \times d)$-matrix*

$$\begin{pmatrix} x_0^2 & x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-2} \\ 0 & x_1 & 3x_2 & 5x_3 & 7x_4 & \cdots & (2d-3)x_{d-1} \\ x_1^2 & x_2 & x_3 & x_4 & x_5 & \cdots & x_d \end{pmatrix}.$$

Furthermore, $\mathcal{M}_d^{\mathrm{IG}}$ has degree $(d-1)^2$. The singular locus of $\mathcal{M}_d^{\mathrm{IG}}$ is given by the line $x_0 = x_1 = \ldots = x_{d-2} = 0$ and the point $x_1 = x_2 = \ldots = x_d = 0$ in $\mathbb{P}^d$.

Our main goal in this section is to rule out case (2) in the Terracini classification, using the strategy outlined in Section 2.2. We begin by homogenizing and clearing denominators in the parametrization (3.1), which gives the following rational map,

$$\phi\colon \mathbb{P}^2 \dashrightarrow \mathcal{M}_d^{\mathrm{IG}}, \quad [\lambda : \mu : s] \mapsto [f_0(\lambda, \mu, s) : f_1(\lambda, \mu, s) : \cdots : f_d(\lambda, \mu, s)], \tag{3.2}$$

where the coordinate functions are given by

$$f_0 = \lambda^{d-1} s^d, \quad f_1 = \lambda^{d-1} s^{d-1} \mu, \quad \ldots \quad f_r = \lambda^{d-r} s^{d-r} \mu^r p_{r-1}(\lambda, \mu), \quad \ldots \quad f_d = \mu^d p_{d-1}(\lambda, \mu).$$

The locus of indeterminacy consists of the following points:

$$P_1 = [0 : 0 : 1], \quad P_2 = [1 : 0 : 0], \quad P_3 = [x_1 : 1 : 0], \quad \ldots \quad P_{d+1} = [x_{d-1} : 1 : 0].$$

where $x_1, \ldots, x_{d-1}$ are the distinct roots of the Bessel polynomial $p_{d-1}(\lambda, 1)$. Note that it follows by [Gro51, Theorem 1] that all roots of each Bessel polynomial are simple. Let f_{gen} be a generic combination of the $d+1$ coordinate functions of ϕ in (3.2), and consider the curve $\mathcal{V}(f_{\mathrm{gen}}) \subseteq \mathbb{P}^2$.

As described in Section 2.2, we construct $\pi\colon \mathcal{S}_d \to \mathbb{P}^2$ by a sequence of blowups. In the case for the inverse Gaussian distribution, we end up needing $\ell_1 = d+1$, $\ell_2 = \cdots = \ell_{d+1} = 1$ blowup steps at the respective indeterminacy points. That this suffices to resolve the indeterminacy locus is proven by the following lemmas. The intersection-theoretic implications of these lemmas that we will use in the rest of this section are collected in Lemma 3.5. We also provide an example of some key steps of the construction in the $d = 4$ case in Example 3.4.

For ease of notation, we will write $\mathcal{E}_i = \mathcal{E}_{i,1}$ for all $i \geq 2$ throughout this section. Note that since the indeterminacy points are isolated points in $\mathbb{P}^2$ and the blowup at a point is birational outside that point, it suffices to independently describe the sequence of blowups over each P_i.

Lemma 3.2. *Let $f_{\mathrm{gen}} = \sum_{k=0}^{d} a_k f_k$ be a linear combination of the coordinate functions of ϕ with general coefficients $(a_0, \ldots, a_d) \in \mathbb{C}^{d+1}$. Then the following holds:*

(1) P_1 is a zero of f_{gen} with multiplicity $d-1$.

(2) The exceptional divisor $\mathcal{E}_{1,1}$ intersects the strict transform of $\mathcal{V}(f_{\mathrm{gen}})$ at a single point $P_{1,1}$ with multiplicity $d-1$; this point corresponds to the tangent direction $\lambda = 0$ at P_1.

(3) *Fix $j \in \{2, \dots, d\}$, and suppose we have already blown up at $P_{1,1}, \dots, P_{1,j-1}$ to obtain $\mathcal{S}_{1,j} \to \mathbb{P}^2$. Then the lift $\phi_{1,j} \colon \mathcal{S}_{1,j} \dashrightarrow \mathbb{P}^d$ has a single new indeterminacy point $P_{1,j}$ on the exceptional divisor $\mathcal{E}_{1,j}$. It is a point on the strict transform of $\mathcal{V}(f_{\text{gen}})$ with multiplicity one.*

(4) *Consider the blowup $\mathcal{S}_{1,d+1}$ at $P_{1,d}$. Then the lift $\phi_{1,d+1} \colon \mathcal{S}_{1,d+1} \dashrightarrow \mathbb{P}^d$ has no new indeterminacy points on $\mathcal{E}_{1,d+1}$.*

Proof. **Part (1):** Consider the affine chart $\mathbb{P}^2 \cap \{s = 1\} \cong \mathbb{A}^2_{(\lambda,\mu)}$, so P_1 is the origin in this chart and the coordinate functions of ϕ are given by

$$f_0 = \lambda^{d-1}, \qquad f_j = \lambda^{d-j}\mu^j p_{j-1}(\lambda, \mu), \quad \text{for } j = 1, \dots, d.$$

The lowest degree terms in λ, μ of these functions are $\lambda^{d-1}, \lambda^{d-1}\mu, \dots, \mu^d$ respectively, and so we see that P_1 is a zero of f_{gen} with multiplicity $d-1$.

Part (2): We continue to work in the affine chart $\mathbb{A}^2_{(\lambda,\mu)}$. The resulting blowup at P_1 is locally given by the coordinates

$$\mathcal{B}_{1,1} = \{((\lambda, \mu), [v_1 : w_1]) \in \mathbb{A}^2 \times \mathbb{P}^1 : \lambda w_1 = \mu v_1\},$$

with blowup morphism given by projection onto $\mathbb{A}^2$. The exceptional divisor $\mathcal{E}_{1,1} \subseteq \mathcal{B}_{1,1}$ is given by $\{(0,0)\} \times \mathbb{P}^1$, and the strict transform of the line $\mathcal{V}(\lambda)$ is given by $\{((0, \mu), [0 : 1]) : \mu \in \mathbb{C}\}$. Recall that $\phi_{1,1} \colon \mathcal{S}_{1,1} \dashrightarrow \mathbb{P}^d$ denotes the lift of ϕ.

Consider the affine chart $\mathcal{B}_{1,1} \cap \{v_1 = 1\} \cong \mathbb{A}^2_{(\lambda,w_1)}$. The restriction to $\mathbb{A}^2_{(\lambda,w_1)}$ of $\phi_{1,1}$ after factoring out the common factor λ^{d-1} is given by

$$(\lambda, w_1) \mapsto \begin{bmatrix} 1 : \lambda w_1 : \cdots \\ \lambda w_1^j p_{j-1}(\lambda, \lambda w_1) : \cdots \\ \lambda w_1^d p_{d-1}(\lambda, \lambda w_1) \end{bmatrix}.$$

These coordinate functions have no common zeros.

Now consider the affine chart $\mathcal{B}_{1,1} \cap \{w_1 = 1\} \cong \mathbb{A}^2_{(\mu,v_1)}$. The restriction of $\phi_{1,1}$ to this chart after factoring out the common factor μ^{d-1} is given by

$$(\mu, v_1) \mapsto \begin{bmatrix} v_1^{d-1} : \mu v_1^{d-1} : \cdots \\ \mu v_1^{d-j} p_{j-1}(\mu v_1, \mu) : \cdots \\ \mu\, p_{d-1}(\mu v_1, \mu) \end{bmatrix}.$$

The coordinate functions have a common zero of multiplicity $d-1$ at $(\mu, v_1) = (0,0)$. In the coordinates of $\mathcal{B}_{1,1}$, this is the point $P_{1,1} = ((0,0), [0 : 1])$, which lies on both the exceptional divisor $\mathcal{E}_{1,1}$ and the strict transform of $\mathcal{V}(\lambda)$. In particular, we see that claim (2) holds.

Parts (3) and (4): *Blowup at $P_{1,1}$*: We first consider the blowup of $\mathcal{S}_{1,1}$ at $P_{1,1}$, which has local coordinates

$$\mathcal{B}_{1,2} = \{((\mu, v_1), [v_2 : w_2]) \in \mathbb{A}^2 \times \mathbb{P}^1 : \mu w_2 = v_1 v_2\}.$$

Consider the affine chart $\mathcal{B}_{1,2} \cap \{w_2 = 1\} \cong \mathbb{A}^2_{(v_1,v_2)}$. Then the restriction of the lift $\phi_{1,2}$ to this chart, after factoring out the common factor v_1^{d-1}, has the zeroth coordinate function given by 1, and so the coordinate functions have no common zeros.

Now consider the affine chart $\mathcal{B}_{1,2} \cap \{v_2 = 1\} \cong \mathbb{A}^2_{(\mu,w_2)}$. The restriction of $\phi_{1,2}$ to this chart, after factoring out the common factor μ^{d-1}, is given by

$$(\mu, w_2) \mapsto \begin{bmatrix} w_2^{d-1} : \mu w_2^{d-1} : \cdots \\ \mu w_2^{d-j} p_{j-1}(\mu w_2, 1) : \cdots \\ \mu\, p_{d-1}(\mu w_2, 1) \end{bmatrix}.$$

To factor out μ^{d-1}, we use the fact that $p_j(\mu^2 w_2, \mu) = \mu\, p_j(\mu w_2, 1)$ for each $j = 1, \ldots, d-1$. Observe that there is a common zero $(\mu, w_2) = (0,0)$ of multiplicity one, and this corresponds to the indeterminacy point $P_{1,2} = ((0,0), [1:0]) \in \mathcal{B}_{1,2}$ on $\mathcal{E}_{1,2}$.

Blowup at $P_{1,2}$: Now we blow up $\mathcal{S}_{1,2}$ at $P_{1,2}$ to obtain $\mathcal{S}_{1,3} \to \mathbb{P}^2$, with local coordinates

$$\mathcal{B}_{1,3} = \{((\mu, w_2), [v_3 : w_3] : \mu w_3 = w_2 v_3\}.$$

and lift $\phi_{1,3}$ of ϕ.

The restriction of $\phi_{1,3}$ to the affine chart $\mathcal{B}_{1,3} \cap \{v_3 = 1\} \cong \mathbb{A}^2_{(\mu, w_3)}$, after factoring out the common factor μ, has zeroth coordinate function given by $\mu^{d-2} w_3^{d-1}$, and dth coordinate function $p_{d-1}(\mu^2 w_3, 1)$, which has a degree zero term. Thus, there are no common zeros in this chart.

Now consider the affine chart $\mathcal{B}_{1,3} \cap \{w_3 = 1\} = \mathbb{A}^2_{(w_2, v_3)}$. The restriction of $\phi_{1,3}$ to this chart, after factoring out the common factor w_2, has coordinate functions

$$(w_2, v_3) \mapsto \begin{bmatrix} w_2^{d-2} : v_3 w_2^{d-1} : \cdots \\ v_3 w_2^{d-k} p_{k-1}(w_2^2 v_3, 1) : \cdots \\ v_3\, p_{d-1}(w_2^2 v_3, 1) \end{bmatrix}.$$

There is a common zero $(w_2, v_3) = (0,0)$ of multiplicity one, and this corresponds to the indeterminacy point $P_{1,3} = ((0,0), [0:1]) \in \mathcal{B}_{1,3}$ on the exceptional divisor $\mathcal{E}_{1,3}$.

Blowup at $P_{1,j}$ for $j = 3, \ldots, d$: Now suppose by induction that we have already blown up at $P_{1,j-1}$ to obtain the surface $\mathcal{S}_{1,j} \to \mathbb{P}^2$, and that there is a indeterminacy point $P_{1,j}$ given by $(0,0) \in \mathbb{A}^2_{(w_2, v_j)} \cong \mathcal{B}_{1,j} \cap \{w_j = 1\}$. Blowup at $P_{1,j}$ to obtain $\mathcal{S}_{1,j+1} \to \mathbb{P}^2$, with local coordinates

$$\mathcal{B}_{1,j+1} = \{((w_2, v_j), [v_{j+1} : w_{j+1}]) \in \mathbb{A}^2 \times \mathbb{P}^1 : w_2 w_{j+1} = v_j v_{j+1}\},$$

Let $\phi_{1,j+1} \colon \mathcal{S}_{1,j+1} \dashrightarrow \mathbb{P}^d$ denote the lift of ϕ.

Consider the affine chart $\mathcal{B}_{1,j+1} \cap \{v_{j+1} = 1\} \cong \mathbb{A}^2_{(w_2, w_{j+1})}$. The restriction of $\phi_{1,j+1}$ to this chart, after factoring out the common factor w_2, has coordinate functions

$$(w_2, w_{j+1}) \mapsto \begin{bmatrix} w_2^{d-j} : w_{j+1} w_2^{d-1} : \cdots \\ w_{j+1} w_2^{d-k} p_{k-1}(w_2^j w_{j+1}, 1) : \cdots \\ w_{j+1}\, p_{d-1}(w_2^j w_{j+1}, 1) \end{bmatrix}.$$

If $j = d$, then there are no common zeros, as the zeroth coordinate function is 1. If $j = 3, \ldots, d-1$, then there is a common zero $(w_2, w_{j+1}) = (0,0)$ of multiplicity one, and this corresponds to the indeterminacy point $P_{1,j+1} = ((0,0), [0:1]) \in \mathcal{B}_{1,j+1}$ on the exceptional divisor $\mathcal{E}_{1,j+1}$ of the blowup.

In the other affine chart $\mathcal{B}_{1,j+1} \cap \{w_{j+1} = 1\} \cong \mathbb{A}^2_{(v_j, v_{j+1})}$, the restriction of $\phi_{1,j+1}$ after factoring out the common factor v_j does not have any common zeros. As in the case of blowing up at $P_{1,2}$, the zeroth coordinate function is $v_j^{d-j} v_{j+1}^{d-j-1}$, while the dth coordinate function is $p_{d-1}(v_j^j v_{j+1}^{j-1}, 1)$, which has a degree zero term.

This completes the proof of parts (3) and (4). □

Lemma 3.3. *Let $f_{\text{gen}} = \sum_{j=0}^d a_j f_j$ be a linear combination of the coordinate functions f_j for ϕ with general coefficients $(a_0, \ldots, a_d) \in \mathbb{C}^d$.*

(1) P_2 is a zero of f_{gen} with multiplicity d. Furthermore, the lift ϕ_2 of ϕ to the blowup $\mathcal{S}_{2,1} \to \mathbb{P}^2$ at P_2 has no new indeterminacy points on the exceptional divisor $\mathcal{E}_2$.

(2) For each $i = 3, \ldots, d+1$, the point $P_i \in \mathbb{P}^2$ is a zero of f_{gen} with multiplicity 1. Furthermore, the lift ϕ_i of ϕ to the blowup $\mathcal{S}_{i,1} \to \mathbb{P}^2$ at P_i has no new indeterminacy points on the exceptional divisor $\mathcal{E}_i$.

Proof. **Part (1):** Consider the affine chart $\mathbb{P}^2 \cap \{\lambda = 1\} \cong \mathbb{A}^2_{(\mu,s)}$. In this chart, P_2 is the origin of $\mathbb{A}^2$, and the coordinate functions of ϕ are given by

$$f_j(\mu, s) = s^{d-j}\mu^j p_{j-1}(1, \mu).$$

The lowest degree terms in μ, s of these functions are all of degree d, and so this proves part (1). Furthermore, these degree d terms of f_0 and f_d are s^d and μ^d respectively, and so the coordinate functions do not have a common tangent direction at the origin. Thus, the strict transform of $\mathcal{V}(f_{\text{gen}})$ does not intersect the exceptional divisor $\mathcal{E}_2$.

Part (2): Recall that $P_i = [x_{i-2} : 1 : 0]$ where x_{i-2} is a simple root of the Bessel polynomial $p_{d-1}(\lambda, 1)$. Consider the affine chart $\mathbb{P}^2 \cap \{\mu = 1\} = \mathbb{A}^2_{(\lambda,s)}$, along with the change of coordinates $\lambda' = \lambda - x_{i-2}$. In this chart, the coordinate functions of ϕ are given by

$$f_0 = (\lambda' + x_{i-2})^{d-1}s^d, \qquad f_j = (\lambda' + x_{i-2})^{d-j}s^{d-j}p_{j-1}(\lambda' + x_{i-2}, 1), \quad \text{for } j = 1, \ldots, d.$$

The lowest degree terms in λ', s of these functions are $s^d, s^{d-1}, \ldots, s, \lambda'$ respectively. Thus, P_i is a zero of f_{gen} with multiplicity 1, and the coordinate functions do not have a common tangent direction at the origin $(\lambda', s) = (0, 0)$. This completes the proof. □

For illustration purposes, we now explicitly carry out the blowups over P_1 in the case $d = 4$.

Example 3.4. Blowing up at P_1: Consider the affine chart $\mathbb{P}^2 \cap \{s = 1\} \cong \mathbb{A}^2$ with coordinates (λ, μ), and where ϕ is given by

$$(\lambda, \mu) \mapsto \left(\lambda^3,\ \lambda^3\mu,\ \lambda^2\mu^2(\lambda + \mu),\ \lambda\mu^3(\lambda^2 + 3\lambda\mu + 3\mu^2),\ \mu^4(15\mu^3 + 15\mu^2\lambda + 6\mu\lambda^2 + \lambda^3)\right).$$

Blowing up at P_1 corresponds to blowing up at $(0, 0)$ in this chart. The resulting surface $\mathcal{S}_{1,1}$ is the closure in $\mathbb{P}^2 \times \mathbb{P}^1$ of

$$\mathcal{B}_{1,1} = \{((\lambda, \mu), [v_1 : w_1]) \in \mathbb{A}^2 \times \mathbb{P}^1 : \lambda w_1 = \mu v_1\} \subseteq \mathbb{P}^2 \times \mathbb{P}^1.$$

Consider the affine chart $\mathcal{B}_{1,1} \cap \{w_1 = 1\}$. Substituting $\lambda = \mu v_1$ and factoring out a common factor μ^3 gives that $\phi_{1,1}$ on this chart is given by

$$(\mu, v_1) \mapsto \begin{bmatrix} v_1^3 : \\ \mu v_1^3 : \\ \mu v_1^2(\mu v_1 + \mu) : \\ \mu v_1(\mu^2 v_1^2 + 3\mu^2 v_1 + 3\mu^2) : \\ \mu(\mu^3 v_1^3 + 6\mu^3 v_1^2 + 15\mu^3 v_1 + 15\mu^3) \end{bmatrix}.$$

Note that the exceptional divisor $\mathcal{E}_{1,1}$ of the blowup is defined by $\mu = 0$, and it intersects the strict transform of $\mathcal{V}(f_{\text{gen}})$ with multiplicity 3 at the point $(0, 0)$. This corresponds to the point $P_{1,1} \in \mathcal{S}_{1,1}$, which gives the tangent direction $\lambda = 0$ of $\mathcal{V}(f_{\text{gen}})$ at P_1.

Blowing up at $P_{1,1}$: The surface $\mathcal{S}_{1,2}$ is defined as the closure in $\mathcal{S}_{1,1} \times \mathbb{P}^1$ of

$$\mathcal{B}_{1,2} = \{((\mu, v_1), [v_2 : w_2]) \in \mathbb{A}^2 \times \mathbb{P}^1 : \mu w_2 = v_1 v_2\}.$$

Taking the affine chart $\{v_2 = 1\}$, substituting $v_1 = \mu w_2$ and factoring out μ^3, we get that $\phi_{1,2}$ on this chart is given by

$$(\mu, w_2) \mapsto \begin{bmatrix} w_2^3 : \\ w_2^3\mu : \\ w_2^2\mu(\mu w_2 + 1) : \\ w_2\mu(\mu^2 w_2^2 + 3\mu w_2 + 3) : \\ \mu(\mu^3 w_2^3 + 6\mu^2 w_2^2 + 15\mu w_2 + 15) \end{bmatrix},$$

with indeterminacy point $(0, 0)$, with multiplicity 1; this corresponds to the point $P_{1,2} \in \mathcal{S}_{1,2}$.

Blowing up at $P_{1,2}$: We blow up at this point to get a surface $\mathcal{S}_{1,3}$ with local coordinates $(\mu, w_2) \times [v_3 : w_3]$. Considering the affine chart $\{w_3 = 1\}$, substituting $\mu = w_2 v_3$ into our coordinate functions and factoring out w_2, we see that $\phi_{1,3}$ on this chart is given by

$$(w_2, v_3) \mapsto \begin{bmatrix} w_2^2 : \\ w_2^3 v_3 : \\ w_2^2 v_3 (w_2^2 v_3 + 1) : \\ w_2 v_3 (w_2^4 v_3^2 + 3 w_2^2 v_3 + 3) : \\ v_3 (w_2^6 v_3^3 + 6 w_2^4 v_3^2 + 15 w_2^2 v_3 + 15). \end{bmatrix},$$

with indeterminacy point $(0,0)$ with multiplicity 1; we call this point $P_{1,3} \in \mathcal{S}_{1,3}$.

Blowing up at $P_{1,3}$: We blow up at this point to get a surface $\mathcal{S}_{1,4}$ with local coordinates $(w_2, v_3) \times [v_4 : w_4]$. Considering the affine chart $\{v_4 = 1\}$, substituting $v_3 = w_2 w_4$ into our coordinate functions, and factoring out w_2, we obtain

$$(w_2, w_4) \mapsto \begin{bmatrix} w_2 \\ w_2^3 w_4 \\ w_2^2 w_4 (w_2^3 w_4 + 1) \\ w_2 w_4 (w_2^6 w_4^2 + 3 w_2^3 w_4 + 3) \\ w_4 (w_2^9 w_4^3 + 6 w_2^6 w_4^2 + 15 w_2^3 w_4 + 15). \end{bmatrix}$$

with indeterminacy point $(0,0)$ of multiplicity 1; we call this point $P_{1,4} \in \mathcal{S}_{1,4}$.

Blowing up at $P_{1,4}$: We blow up at this point to get a surface $\mathcal{S}_{1,5}$ with local coordinates $(w_2, w_4) \times [v_5 : w_5]$. Considering the affine chart $\{v_5 = 1\}$, substituting $w_4 = w_2 w_5$ into our coordinate functions, and factoring out w_2, we finally obtain

$$(w_2, w_5) \mapsto \begin{bmatrix} 1 \\ w_2^3 w_5 \\ w_2^2 w_5 (w_2^4 w_5 + 1) \\ w_2 w_5 (w_2^8 w_5^2 + 3 w_2^4 w_5 + 3) \\ w_5 (w_2^{12} w_5^3 + 6 w_2^8 w_5^2 + 15 w_2^4 w_5 + 15). \end{bmatrix},$$

which lacks indeterminacy points. One can check that at each of the blowup steps above, the other choice of affine chart also does not contain any indeterminacy points for the lift of ϕ. The lift of ϕ is therefore now well-defined over the original point $P_1 \in \mathbb{P}^2$.

The following intersection-theoretic formulas are a direct consequence of Lemma 3.2 and Lemma 3.3 together with (2.6).

Lemma 3.5. *The following formulas hold in* $\mathrm{Pic}(\mathcal{S}_d)$*:*

(1) Let $\mathcal{L}_1 \subseteq \mathbb{P}^2$ be the line through P_1 and P_2. The class of the strict transform of $\mathcal{L}_1$ to $\mathcal{S}_d$ is $L - E_{1,1} - E_2$.

(2) Let $\mathcal{L}_2 \subseteq \mathbb{P}^2$ be the line through P_1 and with tangent direction $\lambda = 0$ (meaning that the strict transform of $\mathcal{L}_2$ under the initial blowup map goes through the point $P_{1,1}$). Then the class of the strict transform of $\mathcal{L}_2$ is $L - E_{1,1} - E_{1,2}$.

(3) Let $\mathcal{L}_3 \subseteq \mathbb{P}^2$ be the line going through the points $P_3, \dots, P_{d+1}$. The class of its strict transform in $\mathrm{Pic}(\mathcal{S}_d)$ is $L - E_3 - \cdots - E_{d+1}$.

(4) The class of the strict transform of $\mathcal{V}(f_{\mathrm{gen}})$ to $\mathcal{S}_d$ for $f_{\mathrm{gen}} = \sum_{i=0}^{d} a_i f_i$ with generic coefficients $(a_0, \dots, a_n) \in \mathbb{C}^n$ is

$$H = (2d-1)L - (d-1)E_{1,1} - (d-1)E_{1,2} - E_{1,3} - \cdots - E_{1,d+1} - dE_2 - E_3 - \cdots - E_{d+1}.$$

As a consequence of Lemmas 3.2, 3.3 and 3.5 we get the following proposition.

Proposition 3.6. *The map $\tilde{\phi} \colon \mathcal{S}_d \to \mathcal{M}_d^{\mathrm{IG}}$ is a birational morphism.*

Proof. On the one hand, by Lemmas 3.2 and 3.3, the map $\tilde{\phi} := \phi_d$ has no indeterminacy points, i.e., it is a morphism. On the other hand, by part (4) of Lemma 3.5, we have that

$$H^2 = (2d-1)^2 - (d-1)^2 - (d-1)^2 - (d-1) - d^2 - (d-1) = (d-1)^2 = \deg(\mathcal{M}_d^{\mathrm{IG}}),$$

where the last equality follows from Theorem 3.1. Since H is the pullback of a hyperplane section of $\mathcal{M}_d^{\mathrm{IG}}$ along $\tilde{\phi}$, we have that $H^2 = \deg(\mathcal{M}_d)\deg(\tilde{\phi})$. Therefore, we conclude that $\deg(\tilde{\phi}) = 1$. Since $\tilde{\phi}$ is dominant, the desired result follows. □

Theorem 3.7. *The moment variety $\mathcal{M}_d^{\mathrm{IG}}$ is k-nondefective for all $k \geq 2$ and $d \geq 2$.*

Proof. Fix $d \geq 2$, and suppose for contradiction that $\mathcal{M}_d^{\mathrm{IG}}$ is k-defective; we may assume that $3k+2 \geq d$ via Lemma 2.3. By Theorem 2.1 and Theorem 3.1, it must be that $\mathcal{M}_d^{\mathrm{IG}}$ is contained in a cone over a curve, i.e., case (2) of Terracini's classification holds, as the singular locus of $\mathcal{M}_d^{\mathrm{IG}}$ is a line [HSY23] and this cannot occur in case (1). Construct the smooth resolution $\pi\colon \mathcal{S}_d \to \mathbb{P}^2$ of the indeterminacy locus of $\phi\colon \mathbb{P}^2 \dashrightarrow \mathcal{M}_d^{\mathrm{IG}}$ via the sequence of blowups as detailed earlier in this section. Recall that H denotes the class of the linear system on $\mathcal{S}_d$ representing hyperplane sections of $\mathcal{M}_d^{\mathrm{IG}} \subseteq \mathbb{P}^d$, pulled back to $\mathcal{S}_d$ via $\tilde{\phi}\colon \mathcal{S}_d \to \mathcal{M}_d^{\mathrm{IG}}$. Then by Lemma 2.2, there exists an effective divisor $\mathcal{A}$ on $\mathcal{S}_d$ with class $A \in \mathrm{Pic}(\mathcal{S}_d)$ and linearly equivalent divisors $\mathcal{D}_1, \ldots, \mathcal{D}_k$ with class $D \in \mathrm{Pic}(\mathcal{S}_d)$ such that H can be expressed as

$$H = A + 2kD.$$

The images of the $\mathcal{D}_i$'s in $\mathbb{P}^2$ all have the same degree $a \geq 1$. Our proof is organized by analyzing the different possibilities for this degree a, and in each case we derive a contradiction. We are then able to conclude that (2) of Theorem 2.1 is not possible, and so $\mathcal{M}_d^{\mathrm{IG}}$ is not k-defective.

Note that a generic enough $\mathcal{H}$ (the pullback of a hyperplane section of $\mathcal{M}_d^{\mathrm{IG}}$ via Lemma 2.2) will not contain any of the exceptional divisors. This implies that none of the curves $\mathcal{D}_i$'s will contain any exceptional divisors, and so D can be represented by a strict transform of a curve in $\mathbb{P}^2$. By (2.6), it then follows that

$$D = aL - b_1E_{1,1} - b_2E_2 - b_3E_{1,2} - \sum_{i=1}^{d-1} c_iE_{1,i+2} - \sum_{i=1}^{d-1} c_i'E_{i+2}, \tag{3.3}$$

where $a = D \cdot L$ is a positive integer giving the degree of the representative's image in $\mathbb{P}^2$, and $b_1, b_2, b_3, c_1, \ldots, c_{d-1}, c_1', \ldots, c_{d-1}'$ are nonnegative. Then

$$0 \leq L \cdot A = L \cdot H - 2kD \cdot L = (2d-1) - 2ka.$$

From this inequality, and using Lemma 2.3, we have $(2ka+1)/2 \leq d \leq 3k+2$, which implies

$$ka + \frac{1}{2} \leq d \leq 3k+2. \tag{3.4}$$

We now proceed by casework on the possibilities for $a \geq 1$. The cases of $a \geq 4$ and $a = 3$ are straightforward consequences from the inequality (3.4), while the cases of $a = 2$ and $a = 1$ require a more careful analysis of the coordinate functions of ϕ.

The case $a \geq 4$: The inequality (3.4) becomes

$$4k + \frac{1}{2} \leq d \leq 3k+2$$

which is a contradiction, since $k \geq 2$.

The case $a = 3$: The inequality (3.4) becomes

$$3k + \frac{1}{2} \leq d \leq 3k+2.$$

Thus, we have only two possibilities for d: either $d = 3k+1$ or $d = 3k+2$.

Assume $d = 3k+1$. Then, we have

$$\begin{aligned} A &= (2d-1-6k)L - (d-1-2kb_1)E_{1,1} - (d-2kb_2)E_2 - (d-1-2kb_3)E_{1,2} - \cdots \\ &= L - (3k-2kb_1)E_{1,1} - (3k+1-2kb_2)E_2 - (3k-2kb_3)E_{1,2} - \cdots . \end{aligned}$$

Since $\mathcal{A}$ is effective and $L \cdot A = 1$, the curve $\pi(\mathcal{A}) \subseteq \mathbb{P}^2$ is a line. Therefore we must have $3k - 2kb_1 \leq 1$, $3k+1-2kb_2 \leq 1$, and $3k - 2kb_3 \leq 1$. Since $k \geq 2$, we get $b_1, b_2, b_3 \geq 2$. We see that $(L - E_{1,1} - E_2) \cdot D < 0$, so the strict transform of the line $\mathcal{L}_1$ is a (fixed) component of the $\mathcal{D}_i$'s, which means that the residual parts are moving conics. We can then reduce to the case $a = 2$, treated below.

If $d = 3k+2$, we obtain

$$\begin{aligned} A &= (2d-1-6k)L - (d-1-2kb_1)E_{1,1} - (d-2kb_2)E_2 - (d-1-2kb_3)E_{1,2} - \cdots \\ &= 3L - (3k+1-2kb_1)E_{1,1} - (3k+2-2kb_2)E_2 - (3k+1-2kb_3)E_{1,2} - \cdots . \end{aligned}$$

Again, since $\mathcal{A}$ is effective and $L \cdot A = 3$, the curve $\pi(\mathcal{A}) \subseteq \mathbb{P}^2$ is a cubic. It will pass through the points P_1, P_2 with multiplicity given by the coefficients above, as well as pass through P_1 with tangent direction $\lambda = 0$. For degree reasons, we must have that

$$3k+1-2kb_1 \leq 3, \quad 3k+2-2kb_2 \leq 3, \quad 3k+1-2kb_3 \leq 3,$$

which implies $b_1, b_3 \geq 1$ and $b_2 \geq 2$. If $b_1 > 1$ or $b_3 > 1$, we could reduce to the case $a = 2$ as above (the plane cubic $\pi(\mathcal{D}_i)$ would need to contain a fixed line between P_1 and P_2). If $b_1 = b_3 = 1$ we obtain $3k+1-2k \leq 3$, which gives $k \leq 2$. Hence, we have reduced to the case when $k = 2$, and $d = 3k+2 = 8$, for which we have already verified computationally that we have non-defectivity (Remark 2.4).

The case $a = 2$: For degree reasons, we can assume that $b_i \leq 1$; otherwise we can reduce to the case $a = 1$. In particular, if $b_i = 2$ for some i, then we would have a conic with a double point, and this would imply that $\pi(\mathcal{D}_i)$ is the union of two lines. Since D is moving, one of these two lines must also be moving and so it is a line through the double point. We could then reduce to the $a = 1$ case.

Furthermore, we can assume $b_1 = b_2 = b_3 = 1$, since otherwise we get $d \leq 8$ (which is covered by Remark 2.4). For example, if $b_1 = 0$, we obtain $A = (2d-1-4k)L - (d-1)E_1 - \cdots$, which implies $d - 1 \leq 2d - 1 - 4k \leq 2d - 1 - 4(d-2)/3$, where in the last inequality we used (3.4). Geometrically, this means that each curve $\mathcal{D}_i$ intersects the exceptional divisors $\mathcal{E}_{1,1}$, $\mathcal{E}_2$ and $\mathcal{E}_{1,2}$. In this case, we have that $\pi(\mathcal{D}_i) \subseteq \mathbb{P}^2$ is given by a quadric of the form

$$g_i(\lambda, \mu, s) = \alpha_i s\lambda + \beta_i \mu^2 + \gamma_i \lambda\mu.$$

In fact, since $\pi(\mathcal{D}_i)$ passes through P_1 and P_2, the monomials s^2 and λ^2 cannot appear in g_i, and since $b_3 = 1$, the tangent direction at P_1 must be $\{\lambda = 0\}$, so the monomial $s\mu$ also cannot appear.

It follows from (2.4) that there exist some linear combination $f_{\text{gen}} = \sum_{i=0}^{d} a_i f_i$ of the coordinate functions with generic coefficients, and a nonzero homogeneous polynomial $h(\lambda, \mu, s)$ corresponding to the plane curve $\pi(\mathcal{A})$ such that

$$f_{\text{gen}} = h(\lambda, \mu, s) \cdot \prod_{i=1}^{k} \left(\alpha_i s\lambda + \beta_i \mu^2 + \gamma_i \lambda\mu\right)^2$$

where $(\alpha_i, \beta_i, \gamma_i) \neq (0,0,0)$ for each i. The contradiction will now follow from a divisibility argument, that relies on the following key observation: The monomials of f_r for $r > 0$ are $\lambda^{d-1-i}\mu^{r+i}s^{d-r}$ for $i \in \{0, \ldots, r-1\}$. Hence, the sets of monomials of $f_0, \ldots, f_d$ are pairwise disjoint, and are of the form $\lambda^\ell s^m \mu^n$ with $\ell + m + n = 2d-1$, $\ell \leq d-1, m \leq d, n \leq 2d-1$. Moreover, we have either $m \leq \ell$, or $m = d$ and $\ell = d-1$.

We may reduce to the case where at least one α_i is nonzero; otherwise, each $\pi(\mathcal{D}_i)$ is reducible and consists of two lines, and we can reduce to the case $a = 1$. If some α_i is nonzero, then we may assume that all the α_i's are nonzero, as the $\mathcal{D}_i$'s are linearly equivalent and move in a one-parameter family by Lemma 2.2.

We therefore see that the monomial $m = s^{2k}\lambda^{2k}$ appears with nonzero coefficient in the expansion of $\prod_{i=1}^{k} g_i^2$. This monomial can only divide the coordinate functions $f_0, \dots, f_{d-2k}$. Let $f_i' = f_i/m$ for each such function divisible by m. Then h must be of the form

$$h = u_0 f_0' + \cdots + u_{d-2k} f_{d-2k}',$$

for some coefficients u_i not all zero (if f_i is not divisible by m, then consider $u_i = 0$ already). Note that f_i' is divisible by s^{d-2k-i}. But if we now consider

$$h \cdot \prod_{i=1}^{k} (\beta_i\mu^2 + \gamma_i\lambda\mu)^2 = u_0 f_0' \prod_{i=1}^{k} (\beta_i\mu^2 + \gamma_i\lambda\mu)^2 + \cdots + u_{d-2k} f_{d-2k}' \prod_{i=1}^{k} (\beta_i\mu^2 + \gamma_i\lambda\mu)^2$$

and recall that each coordinate function is divisible by a distinct power of s, we see that we must have that

$$u_{d-2k} f_{d-2k}' \prod_{i=1}^{k} (\beta_i\mu^2 + \gamma_i\lambda\mu)^2 = v f_d$$

for some $v \in \mathbb{C}$. If the conics $\beta_i\mu^2 + \gamma_i\lambda\mu$ are moving, then we see that $u_{d-2k} = 0$ since f_d is fixed. Otherwise, there exists $\overline{\beta}, \overline{\gamma} \in \mathbb{C}$ such that $\beta_i = \overline{\beta}$ and $\gamma_i = \overline{\gamma}$ for all i, and so the left-hand side with $\mu = 1$ has a root with multiplicity $2k > 1$. However, $p_{d-1}(\lambda, 1)$ and therefore the right-hand side only has simple roots. Thus, we must have that $u_{d-2k} = 0$. We can repeat this argument with $u_{d-2k-1}, \dots, u_0$ in this order to see that all of the u_i's must be zero.

The case $a = 1$: Again, we can assume that $b_i \leq 1$. In fact, since D is moving, at most one of the coefficients in the expansion (3.3) of D can be nonzero (if any two coefficients were equal to one, this would fix either two points or a point and a direction of D).

We begin with the case $b_1 = 1$ (meaning that the representatives of D are strict transforms of lines passing through P_1). Then the projection of the lines $\mathcal{D}_i$ must come from linear forms of the form $g_i(\lambda, \mu, s) = \alpha_i\lambda + \beta_i\mu$, where $(\alpha_i, \beta_i) \neq (0, 0)$. It then follows from Lemma 2.2 that there exists a linear combination $f_{\text{gen}} = \sum_{i=0}^{d} a_i f_i$ for generic coefficients $a_0, \dots, a_d$, and a nonzero homogeneous polynomial $h(\lambda, \mu, s)$ such that

$$f_{\text{gen}} = h(\lambda, \mu, s) \cdot \prod_{i=1}^{k} (\alpha_i\lambda + \beta_i\mu)^2 . \tag{3.5}$$

We now give a divisibility argument similar to that in the $a = 2$ case to show that this is impossible. First, if $\alpha_i = 0$ for all i, the projections of the lines $\mathcal{D}_i$ would pass through both P_1 and P_2. Then both $b_1, b_2 \neq 0$ in (3.3), contradicting the assumption that at most one of the coefficients is nonzero. We may therefore assume that all $\alpha_i \neq 0$. Then λ^{2k} appears with nonzero coefficient in the expansion of $\prod_{i=1}^{k} g_i^2$, and this monomial only divides the coordinate functions $f_0, \dots, f_{d-2k}$. Then, by the same argument as in the $a = 2$ case, this implies that the polynomial $h(\lambda, \mu, s)$ must be zero, which is a contradiction.

The case $b_2 = 1$ (meaning that all representatives of D are strict transforms of lines passing through P_2) is similar (we instead get each $\mathcal{D}_i$ would come from linear forms in μ and s).

Finally, we consider the possibility that $b_1 = b_2 = 0$. In this case,

$$A = H - 2kD = (2d - 1 - 2k)L - (d-1)E_{1,1} - dE_2 - (d-1)E_{1,2} - \cdots .$$

We show that the curve $\mathcal{A}$ on $\mathcal{S}_d$ must have a number of irreducible components that will eventually contradict the condition $d \leq 3k + 2$.

Consider three curves on $\mathcal{S}_d$: the strict transform of the line $\mathcal{L}_1$ between P_1 and P_2, the strict transform of the line $\mathcal{L}_2$ through P_1 whose strict transform passes through $P_{1,2}$, and the strict

transform of the exceptional divisor $\mathcal{E}_{1,1}$ over P_1. By Lemma 4.4, the class of the first in $\mathrm{Pic}(\mathcal{S}_d)$ is $L - E_{1,1} - E_2$, of the second is $L - E_{1,1} - E_{1,2}$ and the third is $E_{1,1} - E_{1,2}$. Now, since we have a negative intersection multiplicity

$$A \cdot (L - E_{1,1} - E_2) = (2d - 1 - 2k) - (d-1) - d = -2k - 1 < 0,$$

we conclude that the strict transform of $\mathcal{L}_1$ is a component of $\mathcal{A}$, appearing with some multiplicity $m > 0$. Then the rest of $\mathcal{A}$ has class

$$A - m(L - E_{1,1} - E_2) = (2d - 1 - 2k - m)L - (d - 1 - m)E_{1,1} - (d - m)E_2 - (d-1)E_{1,2} - \cdots$$

and has intersection multiplicity with the strict transform of $\mathcal{L}_2$ given by

$$(A - m(L - E_{1,1} - E_2)) \cdot (L - E_{1,1} - E_{1,2}) = (2d - 1 - 2k - m) - (d - 1 - m) - (d-1) = 1 - 2k < 0.$$

This is negative, so this strict transform is a component of $\mathcal{A}$ with some multiplicity m'. So the new rest of $\mathcal{A}$ has class

$$\begin{aligned} &A - m(L - E_{1,1} - E_2) - m'(L - E_{1,1} - E_{1,2}) = \\ &\quad (2d - 1 - 2k - m - m')L - (d - 1 - m - m')E_{1,1} - (d - m)E_2 - (d - 1 - m')E_{1,2} - \cdots . \end{aligned}$$

Finally, this has intersection multiplicity with the class $E_{1,1} - E_{1,2}$ given by

$$(A - m(L - E_{1,1} - E_2) - m'(L - E_{1,1} - E_{1,2})) \cdot (E_{1,1} - E_{1,2}) = (d - 1 - m - m') - (d - 1 - m') = -m < 0$$

so the corresponding strict transform appears with some multiplicity $m'' > 0$ in $\mathcal{A}$. The rest $\mathcal{A}'$ of $\mathcal{A}$, after subtracting the three kinds of components we have found, has class

$$\begin{aligned} A' =& A - m(L - E_{1,1} - E_2) - m'(L - E_{1,1} - E_{1,2}) - m''(E_{1,1} - E_{1,2}) \\ &= (2d - 1 - 2k - m - m')L - (d - 1 - m - m' + m'')E_{1,1} - (d - m)E_2 \\ &\quad - (d - 1 - m' - m'')E_{1,2} - \cdots . \end{aligned}$$

Now,

$$A' \cdot (E_{1,1} - E_{1,2}) = d - 1 - m - m' + m'' - (d - 1 - m' - m'') = 2m'' - m \geq 0$$

only if $m'' \geq m/2$. Furthermore

$$A' \cdot (L - E_{1,1} - E_{1,2}) = 1 - 2k + m' \geq 0$$

only if $m' \geq 2k - 1$, and

$$A' \cdot (L - E_{1,1} - E_2) = -2k - m - m' - (-m - m' + m'') - (-m) = -2k + m - m'' \geq 0$$

only if $m - m'' \geq 2k$. But $m'' \geq m/2$ and $m - m'' \geq 2k$ means $m \geq 4k$, so when in addition $m' \geq 2k - 1$, we get

$$0 \leq A' \cdot L = 2d - 1 - 2k - m - m' \leq 2d - 1 - 2k - 4k - 2k + 1 = 2d - 8k,$$

so $d \geq 4k$. So we conclude that $4k \leq d \leq 3k + 2$, which is possible only if $k \leq 2$ and $d \leq 8$, for which non-defectivity was proven computationally (Remark 2.4). □

4. The gamma distribution

From [HSY23, §4], we have the following determinantal realization of the moment variety $\mathcal{M}_d^\Gamma \subseteq \mathbb{P}^d$ for the gamma distribution.

Theorem 4.1 (§4, [HSY23]). *Let $d \geq 3$. The homogeneous prime ideal of the gamma moment variety $\mathcal{M}_d^\Gamma$ is generated by the $\binom{d}{3}$ cubics given by the maximal minors of the $(3 \times d)$-matrix*

$$\begin{pmatrix} 0 & x_1 & 2x_2 & 3x_3 & \cdots & (d-1)x_{d-1} \\ x_0 & x_1 & x_2 & x_3 & \cdots & x_{d-1} \\ x_1 & x_2 & x_3 & x_4 & \cdots & x_d \end{pmatrix}.$$

Furthermore, $\mathcal{M}_d^\Gamma$ has degree $\binom{d}{2}$. The singular locus is given by two points in $\mathbb{P}^d$:

$$x_0 = x_1 = \cdots = x_{d-1} = 0 \quad \text{and} \quad x_1 = x_2 = \cdots = x_d = 0.$$

The moment variety can be parametrized by the rational map

$$\phi\colon \mathbb{P}^2 \dashrightarrow \mathcal{M}_d^\Gamma \subseteq \mathbb{P}^d, \quad [x:y:z] \mapsto [f_0(x,y,z):\cdots:f_d(x,y,z)],$$

where the rth coordinate map is given by

$$f_r(x,y,z) = x^{d-r}\prod_{i=0}^{r-1}(z+iy).$$

This map has finitely many indeterminacy points $P_i = [0:1:-i]$ for $i = 0,\dots,d-1$.

Since the ideal defining $\mathcal{M}_d^\Gamma$ is a determinantal ideal of a matrix with linear entries, the proof strategy used for Gaussian moment varieties in [ARS18] can be applied to show non-defectivity of $\mathcal{M}_d^\Gamma$. In this case, the corresponding Hilbert–Burch matrix would be

$$\begin{pmatrix} y & z & 0 & 0 & \cdots & 0 & 0 \\ 0 & x+y & z & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2x+y & z & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (d-1)x+y & z \end{pmatrix}.$$

However, for illustration purposes, we will instead use the more general strategy outlined in Section 2.2 by constructing a resolution $\pi\colon \mathcal{S}_d \to \mathbb{P}^2$ of the indeterminacy locus of $\phi\colon \mathbb{P}^2 \dashrightarrow \mathcal{M}_d^\Gamma$. We end up needing $\ell_i = d-i$ blowups over the ith indeterminacy point, which we prove in the following lemma. The expressions in the proof are quite involved, and we refer the reader to Example 4.3 below for explicit formulations in the $d = 4$ case. The intersection-theoretic consequences of the construction of $\mathcal{S}_d$ that will be used in the rest of the paper are gathered in Lemma 4.4.

Lemma 4.2. *Fix $i \in \{0,\dots,d-1\}$, and let $j \in \{1,\dots,d-i\}$. Suppose we have blown up at $P_{i,0},\dots,P_{i,j-1}$ to obtain the surface $\mathcal{S}_{i,j} \to \mathbb{P}^2$, along with the exceptional divisor $\mathcal{E}_{i,j}$ and lift $\phi_{i,j}\colon \mathcal{S}_{i,j} \dashrightarrow \mathcal{M}_d$ of $\phi\colon \mathbb{P}^2 \dashrightarrow \mathcal{M}_d$. Let $f_{\text{gen}} = \sum_{k=0}^d a_k f_k$ be a linear combination of the coordinate functions f_k for ϕ, with general coefficients $(a_0,\dots,a_d) \in \mathbb{C}^d$. Then the following hold:*

(1) Let $j = 1$. If $i < d-1$, then the lift $\phi_{i,1}\colon \mathcal{S}_{i,1} \dashrightarrow \mathbb{P}^d$ has a single new indeterminacy point $P_{i,1}$ on $\mathcal{E}_{i,1}$, which does not lie on the strict transform of the line $\mathcal{V}(x)$, and which is a point of multiplicity 1 on the strict transform of the curve $\mathcal{V}(f_{\text{gen}})$. If $i = d-1$, then there are no new indeterminacy points on $\mathcal{E}_{d-1,1}$.

(2) If $1 < j \le d-i-1$, then the lift $\phi_{i,j}\colon \mathcal{S}_{i,j} \dashrightarrow \mathbb{P}^d$ has a single new indeterminacy point $P_{i,j}$ on $\mathcal{E}_{i,j}$, which does not lie on the strict transform of $\mathcal{E}_{i,j-1}$, and which is a point of multiplicity one on the strict transform of the curve $\mathcal{V}(f_{\text{gen}})$.

(3) If $i < d-1$ and $j = d-i$, then the lift $\phi_{i,d-i}\colon \mathcal{S}_{i,d-i} \dashrightarrow \mathbb{P}^d$ has no new indeterminacy points on $\mathcal{E}_{i,d-i}$.

Proof. **Part (1):** Let $j = 1$. Consider the affine chart $\mathbb{P}^2 \cap \{y = 1\} \cong \mathbb{A}^2_{(x,z)}$. Blowing up at $P_{i,0}$ corresponds to blowing up at $(0,-i)$ in this chart. The resulting blowup is given by

$$\mathcal{B}_{i,1} = \{((x,z),[u_1:v_1]) \in \mathbb{A}^2 \times \mathbb{P}^1 : xv_1 = (z+i)u_1\},$$

with blowup morphism given by projection onto $\mathbb{A}^2$. Recall that $\phi_{i,1}\colon \mathcal{B}_{i,1} \dashrightarrow \mathbb{P}^d$ denotes the lift of ϕ. In $\mathcal{B}_{i,1}$, the exceptional divisor $\mathcal{E}_{i,1}$ is given by $\{(0,-i)\} \times \mathbb{P}^1$, and the strict transform of the line $\mathcal{V}(x)$ is given by $\{((0,z),[0:1]) : z \in \mathbb{C}\}$.

Consider the affine chart $\mathcal{B}_{i,1} \cap \{u_1 = 1\} \cong \mathbb{A}^2_{(x,v_1)}$, where the isomorphism is given by $(x,z) \mapsto (x, xv_1 - i)$. The restriction to $\mathbb{A}^2_{(x,v_1)} \dashrightarrow \mathbb{P}^d$ of $\phi_{i,1}$, after factoring out the common

factor x, is given by

$$(x, v_1) \mapsto \left[\begin{array}{l} x^{d-1} : x^{d-2}(xv_1 - i) : \cdots \\ x^{d-k-1}(xv_1 - i)(xv_1 - i + 1) \cdots (xv_1 - i + (k-1)) : \cdots \\ v_1(xv_1 - i)(xv_1 - i + 1) \cdots (xv_1 - 1)(xv_1 + 1) \cdots (xv_1 + d - i - 1) \end{array} \right].$$

The first $d-1$ coordinate functions of $\phi_{i,1}$ (so $k \leq d-2$) always have x as a factor, and the last coordinate function always has v_1 as a factor. If $i = d-1$, the dth function, i.e., when $k = d-1$, is not divisible by either x or by v_1. Then $\phi_{d-1,1}$ has no indeterminacy points. When $i \neq d-1$, the dth function also has v_1 as a factor, so there is a indeterminacy point $P_{i,1} = ((0, -i), [1 : 0])$ on $\mathcal{E}_{i,1}$.

Since the smallest-degree monomial in x and v_1 in the last coordinate function is linear, it follows that $P_{i,1}$ is a point of multiplicity one on the strict transform of the curve $\mathcal{V}(f_{\text{gen}})$.

Now consider the affine chart $\mathcal{B}_{i,1} \cap \{v_1 = 1\} \cong \mathbb{A}^2_{(z,u_1)}$, where the isomorphism is given by $(x, z) \mapsto ((z + i)u_1, z)$. Then the restriction to $\mathbb{A}^2_{(z,u_1)} \dashrightarrow \mathbb{P}^d$ of $\phi_{i,1}$, after factoring out the common factor $(z + i)$, is given by

$$(z, u_1) \mapsto \left[\begin{array}{l} (z + i)^{d-1} u_1^d : \cdots \\ (z + i)^{d-k-1} u_1^{d-k} z(z + 1) \cdots (z + (k-1)) : \cdots \\ z(z + 1) \cdots (z + (i-1))(z + (i + 1)) \cdots (z + (d-1)) \end{array} \right],$$

which has no indeterminacy points on the exceptional divisor $\mathcal{E}_{i,1} \cap \{v_1 = 1\}$. The indeterminacy points $((0, -i'), [0 : 1])$, for $i' \in \{0, \ldots, d-1\} \setminus \{i\}$, do not lie on the exceptional divisor $\mathcal{E}_{i,1}$, and are simply the preimages of the original indeterminacy points $P_{i',0}$. Thus, claim (1) holds.

Part (2) and (3): Suppose $j \in \{2, \ldots, d-i-1\}$, and that we have blown up at $P_{i,0}, \ldots, P_{i,j-2}$ to obtain $\mathcal{S}_{i,j-1} \to \mathbb{P}^2$, and lift $\phi_{i,j-1} \colon \mathcal{S}_{i,j-1} \dashrightarrow \mathbb{P}^d$ of ϕ. Let $\mathcal{B}_{i,j-1} \subseteq \mathbb{A}^2 \times \mathbb{P}^1$ denote local chart of the blowup around $P_{i,j-1}$.

By induction, blowing up at $P_{i,j-1}$ corresponds to blowing up at the origin $(0, 0)$ in the chart $\mathcal{B}_{i,j-1} \cap \{u_{j-1} = 1\} \cong \mathbb{A}^2_{(x,v_{j-1})}$. In these coordinates, the resulting blowup $\mathcal{S}_{i,j}$ is the closure of $\mathcal{B}_{i,j} \subseteq \mathcal{S}_{i,j-1} \times \mathbb{P}^1$, where

$$\mathcal{B}_{i,j} = \{((x, v_{j-1}), [u_j : v_j]) \in \mathbb{A}^2 \times \mathbb{P}^1 : xv_j = v_{j-1}u_j\}.$$

The exceptional divisor $\mathcal{E}_{i,j} \subseteq \mathcal{B}_{i,j}$ is $\{(0, 0)\} \times \mathbb{P}^1$. The strict transform of the previous exceptional divisor $\mathcal{E}_{i,j-1}$ is locally given by $\{((0, v_{j-1}), [0 : 1]) : v_{j-1} \in \mathbb{C}\}$.

Consider the affine chart $\mathcal{B}_{i,j} \cap \{u_j = 1\} = \mathbb{A}^2_{(x,v_j)}$. Here, the lift $\phi_{i,j} \colon \mathcal{S}_{i,j} \dashrightarrow \mathbb{P}^d$ of $\phi_{i,j-1}$, after substituting the appropriate coordinates and factoring x from all coordinate functions, is given by $\phi_{i,j}(x, v_j) = [g_0 : \cdots : g_d]$, where

$$g_k(x, v_j) = \begin{cases} x^{d-j-k} \prod_{-i \leq \ell \leq k-i-1} (x^j v_j + \ell) & 0 \leq k \leq i, \\ x^{d-k} v_j \prod_{\substack{-i \leq \ell \leq k-i-1, \\ \ell \neq 0}} (x^j v_j + \ell) & i + 1 \leq k \leq d. \end{cases}$$

If $i > 0$ and $j = d - i$, i.e., $i = d - j$, then the g_{d-j} coordinate function is of the form

$$\begin{aligned} g_{d-j}(x, v_j) &= (x^j v_j - i)(x^j v_j - (i-1)) \cdots (x^j v_j - i + (d - j - 1)) \\ &= (x^j v_j - i)(x^j v_j - (i-1)) \cdots (x^j v_j - 1), \end{aligned}$$

which is not divisible by either x or v_j. However, note that $g_0 = x^{d-(d-i)} = x^i$, and

$$g_d = v_j \prod_{\substack{-i \leq \ell \leq d-i-1, \\ \ell \neq 0}} (x^{j+1} v_j + \ell).$$

Thus, there are no indeterminacy points of $\phi_{i,j}$ in this case. If $i = 0$ and $j = d$, then we have that $g_0 = 1$, and so there are no indeterminacy points in this case either.

If $2 \leq j \leq d-i-1$, then there is a indeterminacy point $P_{i,j} = ((0,0),[1:0])$ on $\mathcal{E}_{i,j}$, as the coordinate functions $g_0, \ldots, g_{d-j}$ are divisible by x, while the functions $g_{d-j}, \ldots, g_d$ are divisible by v_j. Notice that $P_{i,j}$ does not lie on the strict transform of $\mathcal{E}_{i,j-1}$. Since the monomial in x and v_j of smallest degree of the last coordinate function g_d is linear, it follows that $P_{i,j}$ is a point of multiplicity one on the strict transform of $\mathcal{V}(f_{\text{gen}})$ for a generic linear combination f_{gen} of the coordinate functions.

Now consider the affine chart $\mathcal{B}_{i,j} \cap \{v_j = 1\} = \mathbb{A}^2_{(v_{j-1},u_j)}$. The lift $\phi_{i,j}$, described in local coordinates by making the appropriate substitutions and factoring v_{j-1} from each of the coordinate functions, is given by $(v_{j-1}, u_j) \mapsto [h_0 : \cdots : h_d]$, where

$$h_k(v_{j-1},u_j) = \begin{cases} v_{j-1}^{d-j-k} u_j^{d-j-k+1} \prod_{-i \leq \ell \leq k-i-1} (v_{j-1}^j u_j^{j-1} + \ell) & 0 \leq k \leq i, \\ (v_{j-1}u_j)^{d-k} \prod_{\substack{-i \leq \ell \leq k-i-1, \\ \ell \neq 0}} (v_{j-1}^j u_j^{j-1} + \ell) & i+1 \leq k \leq d. \end{cases}$$

Then, h_0 is a monomial in v_{j-1} and u_j, while h_d is not divisible by either variable. Therefore, there are no indeterminacy points of $\phi_{i,j}$ in this affine chart. This concludes the proof of the claims (2) and (3) of the lemma. □

Example 4.3. For $d = 4$, the rational map $\phi\colon \mathbb{P}^2 \dashrightarrow \mathbb{P}^4$ is given by

$$[x:y:z] \mapsto [x^4 : x^3z : x^2z(z+y) : xz(z+y)(z+2y) : z(z+y)(z+2y)(z+3y)]$$

with indeterminacy points $P_0 = [0:1:0]$, $P_1 = [0:1:-1]$, $P_2 = [0:1:-2]$ and $P_3 = [0:1:-3]$. We will now demonstrate the proof of Lemma 4.2 by resolving the singularity at P_1.

Blowing up at P_1: We blow up at $(0,-1)$ in the chart $\mathbb{P}^2 \cap \{y \neq 0\}$ (isomorphic to $\mathbb{A}^2$ via $(x,z) \mapsto [x:1:z]$), to obtain

$$\mathcal{B}_{1,1} = \{((x,z),[u_1:v_1]) \in \mathbb{A}^2 \times \mathbb{P}^1 : xv_1 = (z+1)u_1\}.$$

The exceptional divisor is given by $\mathcal{E}_{1,1} = \{(0,-1)\} \times \mathbb{P}^1$, and the strict transform of $\mathcal{V}(x)$ is given by $\{((0,z),[0:1]) : z \in \mathbb{A}^1\}$. In the $\mathcal{B}_{1,1} \cap \{u_1 \neq 0\}$ chart, which is isomorphic to $\mathbb{A}^2$ via $(x,v_1) \mapsto ((x, xv_1 - 1),[1:v_1])$, the lift $\phi_{1,1}$ is generically given by

$$(x,v_1) \mapsto \begin{bmatrix} x^3 : \\ x^2\,(xv_1-1) : \\ x^2\,(xv_1-1)\,v_1 : \\ x\,(xv_1-1)\,v_1\,(xv_1+1) : \\ (xv_1-1)\,v_1\,(xv_1+1)\,(xv_1+2) \end{bmatrix}.$$

There is a indeterminacy point $(0,0) \in \mathbb{A}^2$, which corresponds to $((0,-1),[1:0]) \in \mathcal{E}_{1,1}$. On the other hand, in the chart $\mathcal{B}_{1,1} \cap \{v_1 \neq 0\}$, the lift $\phi_{1,1}$ is given by

$$(z,u_1) \mapsto \begin{bmatrix} (z+1)^3\,u_1^4 : \\ (z+1)^2\,u_1^3 z : \\ (z+1)\,u_1^2 z\,(z+1) : \\ u_1 z\,(z+1)\,(z+2) : \\ z\,(z+2)\,(z+3) \end{bmatrix}$$

with no further indeterminacy points on $\mathcal{E}_{1,1}$.

Blowing up at $P_{1,1}$: We now construct $\mathcal{S}_{1,2}$ by blowing up $\mathcal{B}_{1,1} \cap \{u_1 \neq 0\} \cong \mathbb{A}^2$ in the origin, which gives

$$\mathcal{B}_{1,2} = \{((x,v_1),[u_2:v_2] \in \mathbb{A}^2 \times \mathbb{P}^1 : xv_2 = v_1u_2\}.$$

The exceptional divisor is $\mathcal{E}_{1,2} = \{(0,0)\} \times \mathbb{P}^1$, and the strict transform of $\mathcal{E}_{1,1}$ is $\{((0,v_1),[0:1]) : v_1 \in \mathbb{A}^1\}$. In the chart $\mathcal{B}_{1,2} \cap \{u_2 \neq 0\}$, the lift $\phi_{1,2}$ is given by

$$(x,v_2) \mapsto \begin{bmatrix} x^2 : \\ x(x^2v_2-1) : \\ x^2(x^2v_2-1)v_2 : \\ x(x^2v_2-1)v_2(x^2v_2+1) : \\ (x^2v_2-1)v_2(x^2v_2+1)(x^2v_2+2) \end{bmatrix}$$

with a new indeterminacy point $(0,0) \in \mathbb{A}^2$ that corresponds to $((0,0),[1:0]) \in \mathcal{E}_{1,2}$. On the other hand, in the chart $\mathcal{B}_{1,2} \cap \{v_2 \neq 0\}$, the lift is given by

$$(z,v_2) \mapsto \begin{bmatrix} u_2^3v_1^2 : \\ v_1u_2^2(v_1^2u_2-1) : \\ v_1^2u_2^2(v_1^2u_2-1) : \\ v_1u_2(v_1^2u_2-1)(v_1^2u_2+1) : \\ (v_1^2u_2-1)(v_1^2u_2+1)(v_1^2u_2+2) \end{bmatrix}$$

and lacks further indeterminacy points.

Blowing up at $P_{1,2}$: We now construct $\mathcal{S}_{1,3}$ by blowing up $\mathcal{B}_{1,2} \cap \{u_2 \neq 0\} \cong \mathbb{A}^2$ in the origin, which gives

$$\mathcal{B}_{1,3} = \{((x,v_2),[u_3:v_3] \in \mathbb{A}^2 \times \mathbb{P}^1 : xv_3 = v_2u_3\}.$$

The exceptional divisor is $\mathcal{E}_{1,3} = \{(0,0)\} \times \mathbb{P}^1$, and the strict transform of $\mathcal{E}_{1,2}$ is $\{((0,v_2),[0:1]) : v_2 \in \mathbb{A}^1\}$. In the chart $\mathcal{B}_{1,3} \cap \{u_3 \neq 0\}$, the lift $\phi_{1,3}$ is given by

$$(x,v_3) \mapsto \begin{bmatrix} x : \\ x^3v_3-1 : \\ x^2(x^3v_3-1)v_3 : \\ x(x^3v_3-1)v_3(x^3v_3+1) : \\ (x^3v_3-1)v_3(x^3v_3+1)(x^3v_3+2) \end{bmatrix}$$

and lacks further indeterminacy points. Similarly, in the chart $\mathcal{B}_{1,3} \cap \{v_3 \neq 0\}$, the lift is given by

$$(z,v_3) \mapsto \begin{bmatrix} v_2u_3^2 : \\ u_3(u_3^2v_2^3-1) : \\ v_2^2u_3^2(u_3^2v_2^3-1) : \\ v_2u_3(u_3^2v_2^3-1)(u_3^2v_2^3+1) : \\ (u_3^2v_2^3-1)(u_3^2v_2^3+1)(u_3^2v_2^3+2) \end{bmatrix}$$

and lacks further indeterminacy points.

The following intersection-theoretic formulas follow directly from Lemma 4.2 and (2.6).

Lemma 4.4. *Let $E_{i,j}$ denote the class in $\mathrm{Pic}(\mathcal{S}_d)$ of the pullback of $\mathcal{E}_{i,j}$ to $\mathcal{S}_d$ along the composition of appropriate blowup maps, and let L denote the class of the pullback $\mathcal{L}$ of a line in $\mathbb{P}^2$. The following formulas hold:*

(1) The class of the strict transform of the line $\mathcal{V}(x)$ is given by $L - \sum_{i=0}^{d-1} E_{i,1}$.

(2) The class of the strict transform of $\mathcal{E}_{i,j}$ is given by $E_{i,j} - E_{i,j+1}$ for $i \in \{0,\dots,d-1\}$ and $j \in \{1,\dots,d-i-1\}$.

(3) The class of the strict transform of $\mathcal{E}_{i,d-i}$ is given by $E_{i,d-i}$ for $i \in \{0,\dots,d-1\}$.

(4) Let H be the class of the strict transform $\mathcal{H}$ of $\mathcal{V}(f_{\mathrm{gen}}) \subseteq \mathbb{P}^2$ for a generic linear combination f_{gen} of the coordinate functions of ϕ. Then

$$H = dL - \sum_{i=0}^{d-1} \sum_{j=1}^{d-i} E_{i,j}. \tag{4.1}$$

As in the previous section, Lemmas 4.2 and 4.4 imply the following proposition.

Proposition 4.5. *The map $\tilde{\phi}\colon \mathcal{S}_d \to \mathcal{M}_d^\Gamma$ is a birational morphism.*

Proof. On the one hand, by Lemma 4.2, the map $\tilde{\phi} := \phi_{d-1,1}$ has no indeterminacy points, i.e., it is a morphism. On the other hand, by Lemma 4.4, we have that

$$H^2 = d^2 - d - (d-1) - \cdots - 1 = \binom{d}{2} = \deg(\mathcal{M}_d^\Gamma),$$

where the last equality follows from Theorem 4.1. Since H is the pullback of a hyperplane section of $\mathcal{M}_d^\Gamma$ along $\tilde{\phi}$, we have that $H^2 = \deg(\mathcal{M}_d^\Gamma)\deg(\tilde{\phi})$. Therefore, we conclude that $\deg(\tilde{\phi}) = 1$. Since $\tilde{\phi}$ is dominant, the desired result follows. □

We are now ready to state and prove the main theorem of this section.

Theorem 4.6. *The moment variety $\mathcal{M}_d^\Gamma$ is k-nondefective for all $k \geq 2$ and $d \geq 2$.*

Proof. We will use the same proof strategy as for the inverse Gaussian distribution (Theorem 3.7). Fix d, and assume that $\mathcal{M}_d^\Gamma$ is k-defective. Then case (2) of Theorem 2.1 must apply, and so by Lemma 2.2, we have an expression for the class H of the pullback of a general hyperplane section $\mathcal{H}$ of $\mathcal{M}_d^\Gamma$ to $\mathcal{S}_d$.

In particular, there exist linearly equivalent divisors $\mathcal{D}_1, \ldots, \mathcal{D}_k$ of $\mathcal{S}_d$ with class

$$D = aL - \sum_{i=0}^{d-1}\sum_{j=1}^{d-i} b_{i,j}E_{i,j}$$

with coefficients $a > 0$ and $b_{i,j} \geq 0$, such that

$$A = H - 2kD = (d-2ka)L - \sum_{i=0}^{d-1}\sum_{j=1}^{d-i}(1-b_{i,j})E_{i,j}$$

is the class of an effective divisor by Lemma 4.4. As in the inverse Gaussian case, we proceed by casework on the value of a, which is the degree of the projections $\pi(\mathcal{D}_i)$ to $\mathbb{P}^2$. We will almost immediately be able to reduce to the $a = 1$ case, and derive a contradiction there.

We have that $A \cdot L = d - 2ka \geq 0$. Together with Lemma 2.3, we obtain the inequalities

$$2ak \leq d \leq 3k + 2. \tag{4.2}$$

This immediately allows us to rule out the case $a \geq 3$ (since that would give $6k \leq 3k+2$ which is impossible for $k \geq 2$).

In the case $a = 2$, we get $6k \leq 3k + 2$ and the only possibility is $k = 2$ and $d \leq 8$, for which non-defectivity has already been verified computationally (Remark 2.4).

Therefore, we are left to investigate the $a = 1$ case. Here, we have $D = L - \sum_{i=0}^{d-1}\sum_{j=1}^{d-i} b_{i,j}E_{i,j}$, and since D is moving, at most one of the coefficients $b_{i,1}$ is 1, and the coefficients of the classes of the other exceptional divisors are 0. We will now investigate each of the possibilities, and see that each of them leads to a contradiction.

The case $b_{i,1} = 1$ for some $i \in \{0, \ldots, d-1\}$: This means that for each $\ell \in \{1, \ldots, k\}$, the curve $\mathcal{D}_\ell$ is the strict transform of a line in $\mathbb{P}^2$ passing through P_i, which means that it is given by a linear form

$$g_\ell(x, y, z) = \alpha_\ell x + \beta_\ell(iy + z)$$

for some $(\alpha_\ell, \beta_\ell) \neq (0, 0)$. Note that $g_\ell(0, 1, -i) = 0$. Hence, there is a linear combination $f_{\text{gen}} = \sum_{j=0}^{d} a_j f_j$ with $(a_0, \ldots, a_d) \neq 0$, such that

$$f_{\text{gen}} = h(x, y, z)\prod_{\ell=1}^{k}(\alpha_\ell x + i\beta_\ell y + \beta_\ell z)^2 \tag{4.3}$$

for a homogeneous polynomial $h(x, y, z) \neq 0$. From this we can derive a contradiction by a similar divisibility argument as we used for the inverse Gaussian case.

The idea is as follows: We have that f_{gen} is a homogeneous polynomial of degree d, and that $(g_1 \cdots g_k)^2$ is homogeneous of degree $2k$. Thus, h is homogeneous of degree $d - 2k \geq 0$. Note that f_j only involves monomials of the form $x^{d-j}y^r z^s$, so no monomial appears in more than one of the f_j's. Furthermore, recall that if P_i is a root of a coordinate function f_j, then it is a simple root.

If $\alpha_\ell = 0$ for some ℓ, then $\alpha_\ell = 0$ for all ℓ, and so f_{gen} has a root at P_i with multiplicity at least $2k$. But this is not possible, as this would imply that any coordinate functions f_j appearing in f_{gen} also have P_i as a root with multiplicity at least $2k > 1$, contradicting the fact that the P_i's are simple roots of the coordinate functions. Hence, we conclude that the monomial x^{2k} appears as a monomial in $(g_1 \cdots g_k)^2$. However, it only divides $f_0, \ldots, f_{d-2k}$ in the left hand side of (4.3). Thus, we can apply a similar argument to the one applied for the inverse Gaussian in the $a = 2$ case, to conclude that $h(x, y, z) = 0$, which is a contradiction.

The case $b_{i,1} = 0$ for all $i \in \{0, \ldots, d-1\}$: In this case, A has the following form:

$$A = (d - 2k)L - \sum_{i=0}^{d-1} \sum_{j=1}^{d-i} E_{i,j}.$$

Note that $d - 2k > 0$ since $\mathcal{A}$ is effective. We will derive a contradiction by showing that $\mathcal{A}$ has the irreducible effective divisors from Lemma 4.4 as components, and that after removing them in a certain order, we obtain a divisor that is not effective, thereby contradicting the effectiveness of $\mathcal{A}$.

First, we have that

$$A \cdot \left(L - \sum_{i=0}^{d-1} E_{i,1}\right) = (d - 2k) - d = -2k < 0,$$

and so $L - \sum_{i=0}^{d-1} E_{i,1}$ is the class of a divisor that is a component of $\mathcal{A}$. Let $A_1 = A - (L - \sum_{i=0}^{d-1} E_{i,1})$ be the class of the divisor $\mathcal{A}_1$ obtained by removing this component, so

$$A_1 = (d - 2k - 1)L - \sum_{i=0}^{d-2} \sum_{j=2}^{d-i} E_{i,j}.$$

We have that $A_1 \cdot (E_{i,1} - E_{i,2}) = -1$ for $i = 0, \ldots, d-2$, so we can remove each of the components corresponding to these classes $E_{i,1} - E_{i,2}$ to obtain a divisor $\mathcal{A}_1'$, with class in the Picard group

$$\begin{aligned} A_1' &= A_1 - (E_{0,1} - E_{0,2}) - \cdots - (E_{d-2,1} - E_{d-2,2}) \\ &= (d - 2k - 1)L - \sum_{i=0}^{d-2} E_{i,1} - \sum_{i=0}^{d-3} \sum_{j=3}^{d-i} E_{i,j}. \end{aligned}$$

Then,

$$A_1' \cdot \left(L - \sum_{i=0}^{d-1} E_{i,1}\right) = (d - 2k - 1) - (d - 1) = -2k < 0,$$

and so we can once again remove $L - \sum_{i=0}^{d-1} E_{i,1}$ to obtain

$$A_2 = A_1' - \left(L - \sum_{i=0}^{d-1} E_{i,1}\right) = (d - 2k - 2)L + E_{d-1,1} - \sum_{i=0}^{d-3} \sum_{j=3}^{d-i} E_{i,j}.$$

We have that $A_2 \cdot (E_{i,2} - E_{i,3}) = -1$ for $i = 0, \ldots, d-3$, so we can remove each of these components to obtain

$$\begin{aligned} A_2' &= A_2 - (E_{0,2} - E_{0,3}) - \cdots - (E_{d-3,2} - E_{d-3,3}) \\ &= (d - 2k - 2)L + E_{d-1,1} - \sum_{i=0}^{d-3} E_{i,2} - \sum_{i=0}^{d-4} \sum_{j=4}^{d-i} E_{i,j}. \end{aligned}$$

We then have that $A_2' \cdot (E_{i,1} - E_{i,2}) = -1$ for $i = 0, \dots, d-3$, so we can also remove each of the components corresponding to these classes to obtain

$$\begin{aligned} A_2'' &= B_2 - (E_{0,1} - E_{0,2}) - \cdots - (E_{d-3,1} - E_{d-3,2}) \\ &= (d-2k-2)L + E_{d-1,1} - \sum_{i=0}^{d-3} E_{i,1} - \sum_{i=0}^{d-4} \sum_{j=4}^{d-i} E_{i,j}. \end{aligned}$$

Finally, we have that $A_2'' \cdot E_{d-1,1} = -1$, so removing the corresponding components, we obtain

$$A_2''' = (d-2k-2)L - \sum_{i=0}^{d-3} E_{i,1} - \sum_{i=4}^{d-4} \sum_{j=4}^{d-i} E_{i,j}.$$

Now, $A_2'' \cdot (L - \sum_{i=0}^{d-1} E_{i,1}) = (d-2k-2) - (d-2) = -2k < 0$, so we can remove the corresponding component to obtain

$$A_3 = (d-2k-3)L + E_{d-2,1} + E_{d-1,1} - \sum_{i=0}^{d-4} \sum_{j=4}^{d-i} E_{i,j}.$$

We continue in this way, removing effective irreducible components. Eventually, we have removed $d-2k-1$ copies of $\mathcal{L}$, along with many other effective divisors, and have

$$\begin{aligned} A_{d-2k-1} &= (d-2k-(d-2k-1))L + \sum_{i=d-(d-2k-2)}^{d-1} E_{i,1} - \sum_{i=0}^{d-(d-2k-1)-1} \sum_{j=d-2k}^{d-i} E_{i,j} \\ &= L + \sum_{i=2k+2}^{d-1} E_{i,1} - \sum_{i=0}^{2k} \sum_{j=d-2k}^{d-i} E_{i,j}. \end{aligned}$$

One can check that the divisors with classes

$$\begin{aligned} &E_{0,d-2k-1} - E_{0,d-2k}, \dots, E_{2k,d-2k-1} - E_{2k,d-2k}, \\ &E_{0,d-2k-2} - E_{0,d-2k-1}, \dots, E_{2k,d-2k-2} - E_{2k,d-2k-1}, \\ &\vdots \\ &E_{0,1} - E_{0,2}, \dots, E_{2k,1} - E_{2k,1}, \\ &E_{2k+1,1}, \dots, E_{d-1,1} \end{aligned}$$

are components, and remove them in the order above. If $d-2k = 1$, then we are only removing the last line of divisors. We then obtain the divisor $\mathcal{A}'_{d-2k-1}$ with class

$$A'_{d-2k-1} = L - \sum_{i=0}^{2k} E_{i,1} - \sum_{i=0}^{2k-1} \sum_{j=d-2k+1}^{d-i} E_{i,j}.$$

Intersecting this with the strict transform of $\mathcal{V}(x)$, we see that

$$A'_{d-2k-1} \cdot \Big(L - \sum_{i=0}^{d-1} E_{i,1}\Big) = 1 - (2k+1) = -2k < 0,$$

so the strict transform of $\mathcal{V}(x)$ is still a component, and we remove it to get the divisor $\mathcal{A}_{d-2k}$ with class

$$A_{d-2k} = A'_{d-2k-1} - \Big(L - \sum_{i=0}^{d-1} E_{i,1}\Big) = \sum_{i=2k+1}^{d-1} E_{i,1} - \sum_{i=0}^{2k-1} \sum_{j=d-2k+1}^{d-i} E_{i,j}.$$

But $\mathcal{A}_{d-2k}$ is not effective: the divisors whose classes have negative coefficients do not lie over those with positive coefficients. We have thus reached our contradiction. $\square$

5. Rational identifiability

In the previous sections we have seen that the parameters of k-mixtures of the inverse Gaussian or gamma distribution are algebraically identifiable from the first $3k-1$ moments. A natural question is how many further moments we need in order to have *rational identifiability*, in the sense that for generic sample moments for which there is a solution to the moment equations, the solution is unique up to the label swapping symmetry. In the language of algebraic geometry, this corresponds to the problem of *k-identifiability*: Given a k, for what d does it hold that a generic point of $\mathrm{Sec}_k(\mathcal{M}_d)$ lies on a unique k-secant?

Based on numerical experiments, we conjecture that $d \geq 3k$ suffices, but in what follows we will instead prove the more modest claim that $d \geq 3k+2$ suffices. Our proof strategy will be the same as the one used in [LAR21, Section 3] to prove the analogous statement for the Gaussian distribution, namely, to use the following sufficient conditions for k-identifiability.

Theorem 5.1 ([MM22, Theorem 1.5])**.** *Let $X \subseteq \mathbb{P}^d$ be an irreducible nondegenerate variety, and let $k \geq 2$. Then a generic point of $\mathrm{Sec}_k(X)$ lies on a unique k-secant if the following conditions are satisfied:*

(1) $(k+1)\dim(X) + k \leq d$

(2) X is $(k+1)$-nondefective

(3) The Gauss map of X is nondegenerate.

Theorem 5.2. *For k-mixtures of the inverse Gaussian or the gamma distribution, we have rational identifiability from the first $3k+2$ moments.*

Proof. Let $d = 3k+2$. It is immediate from the results from [HSY23] that $\mathcal{M}_d$ is irreducible and nondegenerate, and that condition (1) is satisfied, whereas condition (2) follows from Theorems 3.7 and 4.6.

To prove (3), we use a classical result about Gauss maps of surfaces (see, e.g., [IL03, Theorem 4.3.6]), which says that if the Gauss map of $\mathcal{M}_d$ is degenerate, then $\mathcal{M}_d$ is either a cone over a curve, or the tangential variety of a curve. It follows from the proofs of Theorems 3.7 and 4.6 that $\mathcal{M}_d$ cannot be a cone over a curve. Assume now for a contradiction that $\mathcal{M}_d$ is equal to the tangential variety $\tau(C)$ for some curve C in $\mathbb{P}^d$. Since $\mathcal{M}_d$ is not contained in a plane, C is not a plane curve. Hence, C is contained in the singular locus of $\mathcal{M}_d$. (This follows from the general fact that a nonplanar curve is contained in the singular locus of its tangential variety; see, e.g., [Pie81] for the case of space curve, from which the general case readily follows.) For the gamma distribution, the singular locus is zero-dimensional by Theorem 4.1, so this is impossible. In the inverse Gaussian case, the singular locus is a line and a point by Theorem 3.1, so this would imply that C is a line, which in turn would imply that $\tau(C)$ is a line, which contradicts $\tau(C) = \mathcal{M}_d$. □

References

[AAR21] Daniele Agostini, Carlos Améndola, and Kristian Ranestad. Moment identifiability of homoscedastic Gaussian mixtures. *Found. Comput. Math.*, 21(3):695–724, 2021.

[ABGO24] Hirotachi Abo, Maria Chaira Brambilla, Francesco Galuppi, and Alessandro Oneto. Non-defectivity of Segre–Veronese varieties, 2024. Preprint: `arXiv:2406.20057`.

[AFS16] Carlos Améndola, Jean-Charles Faugère, and Bernd Sturmfels. Moment varieties of Gaussian mixtures. *J. Algebr. Stat.*, 7(1):14–28, 2016.

[AH95] James Alexander and André Hirschowitz. Polynomial interpolation in several variables. *J. Algebraic Geom.*, 4(2):201–222, 1995.

[ARS18] Carlos Améndola, Kristian Ranestad, and Bernd Sturmfels. Algebraic identifiability of Gaussian mixtures. *Int. Math. Res. Not. IMRN*, (21):6556–6580, 2018.

[BCCGO18] Alessandra Bernardi, Enrico Carlini, Maria Virginia Catalisano, Alessandro Gimigliano, and Alessandro Oneto. The hitchhiker guide to: Secant varieties and tensor decomposition. *Mathematics*, 6(12):314, 2018.

[BCMO23] Alexander Taveira Blomenhofer, Alex Casarotti, Mateusz Michałek, and Alessandro Oneto. Identifiability for mixtures of centered Gaussians and sums of powers of quadratics. *Bull. Lond. Math. Soc.*, 55(5):2407–2424, 2023.

[Blo23] Alexander Taveira Blomenhofer. Gaussian mixture identifiability from degree 6 moments, 2023. Preprint: `arXiv:2307.03850`.

[BS15] Mikhail Belkin and Kaushik Sinha. Polynomial learning of distribution families. *SIAM J. Comput.*, 44(4):889–911, 2015.

[CC02] Luca Chiantini and Ciro Ciliberto. Weakly defective varieties. *Trans. Amer. Math. Soc.*, 354(1):151–178, 2002.

[CM23] Alex Casarotti and Massimiliano Mella. From non-defectivity to identifiability. *J. Eur. Math. Soc. (JEMS)*, 25(3):913–931, 2023.

[CMNT23] James Cruickshank, Fatemeh Mohammadi, Anthony Nixon, and Shinichi Tanigawa. Identifiability of points and rigidity of hypergraphs under algebraic constraints, 2023. Preprint: `arXiv:2305.18990`.

[CP24] Alex Casarotti and Elisa Postinghel. Waring identifiability for powers of forms via degenerations. *Proc. Lond. Math. Soc. (3)*, 128(1):Paper No. e12579, 30, 2024.

[GKW20] Alexandros Grosdos Koutsoumpelias and Markus Wageringel. Moment ideals of local Dirac mixtures. *SIAM J. Appl. Algebra Geom.*, 4(1):1–27, 2020.

[Gro51] Emil Grosswald. On some algebraic properties of the Bessel polynomials. *Trans. Am. Math. Soc.*, 71:197–210, 1951.

[Har77] Robin Hartshorne. *Algebraic Geometry*. Graduate Texts in Mathematics. Springer, 1977.

[HSY23] Oskar Henriksson, Lisa Seccia, and Teresa Yu. Moment varieties from inverse Gaussian and gamma distributions, 2023. *Algebraic Statistics*, to appear. Preprint: `arXiv:2312.10433`.

[IL03] Thomas A. Ivey and Joseph M. Landsberg. *Cartan for beginners: differential geometry via moving frames and exterior differential systems*, volume 61. American Mathematical Society Providence, 2003.

[KSS20] Kathlén Kohn, Boris Shapiro, and Bernd Sturmfels. Moment varieties of measures on polytopes. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 21:739–770, 2020.

[Lan12] Joseph M. Landsberg. *Tensors: geometry and applications*, volume 128 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[LAR21] Julia Lindberg, Carlos Améndola, and Jose Israel Rodriguez. Estimating Gaussian mixtures using sparse polynomial moment systems, 2021. Preprint: `arXiv:2106.15675v3`.

[MM22] Alex Massarenti and Massimiliano Mella. Bronowski's conjecture and the identifiability of projective varieties, 2022. Preprint: `arXiv:2210.13524v3`.

[OTT25] Giorgio Ottaviani and Ettore Teixeira Turatti. Generalized identifiability of sums of squares. *J. Algebra*, 661:641–656, 2025.

[Pie81] Ragni Piene. Cuspidal projections of space curves. *Math. Ann.*, 256:95–119, 1981.

Authors' addresses:

Oskar Henriksson, University of Copenhagen `oskar.henriksson@math.ku.dk`
Kristian Ranestad, University of Oslo `ranestad@math.uio.no`
Lisa Seccia, University of Neuchâtel `lisa.seccia@unine.ch`
Teresa Yu, University of Michigan `twyu@umich.edu`

G

3D Genome Reconstruction from Partially Phased Hi-C Data

Diego Cifuentes
School of Industrial Systems Engineering
Georgia Institute of Technology

Jan Draisma
Mathematical Institute
University of Bern

Oskar Henriksson
Department of Mathematical Sciences
University of Copenhagen

Annachiara Korchmaros
Interdisciplinary Center for Bioinformatics
Leipzig University

Kaie Kubjas
Department of Mathematics and Systems Analysis
Aalto University

Publication details

Published in *Bulletin of Mathematical Biology* **86**, 33 (2024)
DOI: 10.1007/s11538-024-01263-7

Bulletin of Mathematical Biology (2024) 86:33
https://doi.org/10.1007/s11538-024-01263-7

Society for Mathematical Biology

ORIGINAL ARTICLE

3D Genome Reconstruction from Partially Phased Hi-C Data

Diego Cifuentes[1] · Jan Draisma[2] · Oskar Henriksson[3] · Annachiara Korchmaros[4] · Kaie Kubjas[5]

Received: 8 July 2023 / Accepted: 22 January 2024 / Published online: 22 February 2024

Abstract

The 3-dimensional (3D) structure of the genome is of significant importance for many cellular processes. In this paper, we study the problem of reconstructing the 3D structure of chromosomes from Hi-C data of diploid organisms, which poses additional challenges compared to the better-studied haploid setting. With the help of techniques from algebraic geometry, we prove that a small amount of phased data is sufficient to ensure finite identifiability, both for noiseless and noisy data. In the light of these results, we propose a new 3D reconstruction method based on semidefinite programming, paired with numerical algebraic geometry and local optimization. The performance of this method is tested on several simulated datasets under different noise levels and with different amounts of phased data. We also apply it to a real dataset from

Kaie Kubjas
kaie.kubjas@aalto.fi

Diego Cifuentes
diego.cifuentes@isye.gatech.edu

Jan Draisma
jan.draisma@math.unibe.ch

Oskar Henriksson
oskar.henriksson@math.ku.dk

Annachiara Korchmaros
annachiara.korchmaros@uni-leipzig.de

1 School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive, NW, Atlanta, GA 30332, USA

2 Mathematisches Institut, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland

3 Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

4 Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany

5 Department of Mathematics and Systems Analysis, Aalto University, P.O. Box 11100, 00076 Aalto, Finland

mouse X chromosomes, and we are then able to recover previously known structural features.

Keywords 3D genome organization · Diploid organisms · Hi-C · Applied algebraic geometry · Numerical algebraic geometry

Mathematics Subject Classification 92E10 · 92-08 · 13P25 · 14P05 · 65H14 · 90C90

1 Introduction

The eukaryotic chromatin has a three-dimensional (3D) structure in the cell nucleus, which has been shown to be important in regulating basic cellular functions, including gene regulation, transcription, replication, recombination, and DNA repair (Uhler and Shivashankar 2017; Wang et al. 2018). The 3D DNA organization is also associated with brain development and function; in particular, it is shown to be misregulated in schizophrenia (Rajarajan et al. 2018; Rhie et al. 2018) and Alzheimer's disease (Nott et al. 2019).

All genetic material is stored in chromosomes, which interact in the cell nucleus, and the 3D chromatin structure influences the frequencies of such interactions. A benchmark tool to measure such frequencies is high-throughput chromosome conformation capture (Hi-C) (Lafontaine et al. 2021). Hi-C first crosslinks cell genomes, which "freezes" contacts between DNA segments. Then the genome is cut into fragments, the fragments are ligated together and then are associated with equally-sized segments of the genome using high-throughput sequencing (Rao et al. 2014). These segments of the genome are called *loci*, and their size is known as *resolution* (e.g., bins of size 1Mb or 50Kb). The result of Hi-C is stored in a matrix called *contact matrix* whose elements are the *contact counts* between pairs of loci.

According to the structure they generate, computational methods for inferring the 3D chromatin structure from a contact matrix fall into two classes: ensemble and consensus methods. In a haploid setting (organisms having a single set of chromosomes), ensemble models such as MCMC5C (Rousseau et al. 2011), BACH-MIX (Hu et al. 2013) and Chrom3D (Paulsen et al. 2017), try to account for structure variations on the genome across cells by inferring a population of 3D structures. On the other hand, consensus methods aim at reconstructing one single 3D structure which may be used as a model for further analysis. In this category, probability-based methods such as PASTIS (Varoquaux et al. 2014; Cauer et al. 2019) and ASHIC (Ye and Ma 2020) model contact counts as Poisson random variables of the Euclidean distances between loci, and distance-based methods such as ChromSDE (Zhang et al. 2013) and ShRec3D (Lesne et al. 2014) model contact counts as functions of the Euclidean distances. An extensive overview of different 3D genome reconstruction techniques is given in Oluwadare et al. (2019).

Most of the methods for 3D genome reconstructions from Hi-C data are for haploid organisms. However, like most mammals, humans are diploid organisms, in which the genetic information is stored in pairs of chromosomes called homologs. Homologous chromosomes are almost identical besides some single nucleotide polymorphisms

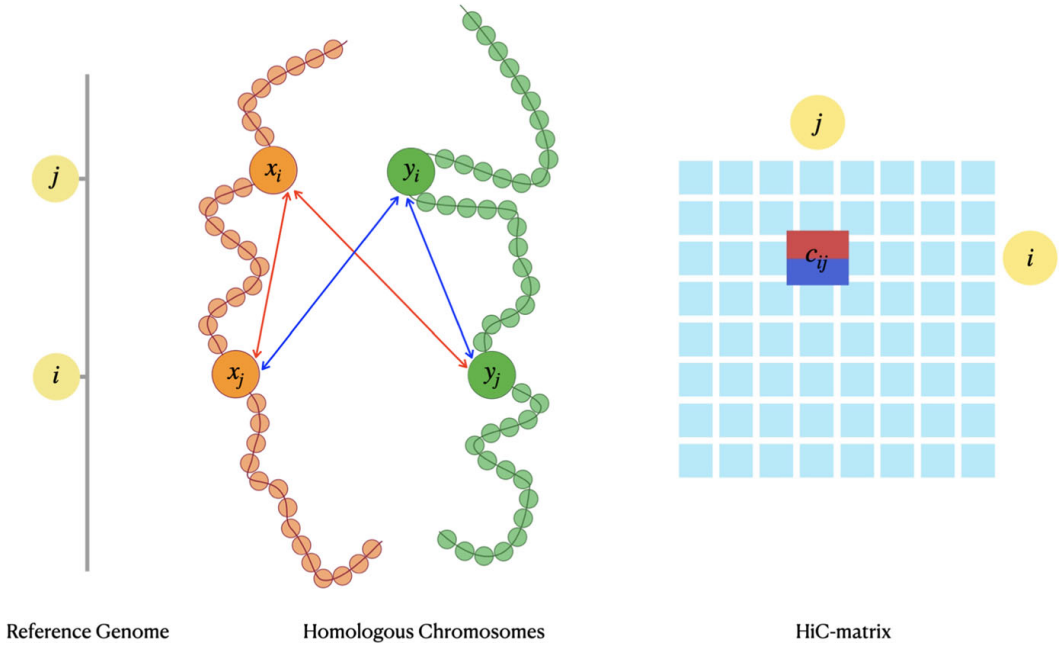


Fig. 1 Ambiguity of phased data. Each entry $c_{i,j}$ of the Hi-C matrix corresponds to four different contacts between the two pairs (x_i, y_i) for locus i and (x_j, y_j) for locus j

(SNPs) (Li et al. 2021). In the case of diploid organisms, the Hi-C data does not generally differentiate between homologous chromosomes. If we model each chromosome as a string of beads, then we associate two beads to each locus $i \in \{1, \ldots, n\}$, one bead for each homolog. Therefore, each observed contact count $c_{i,j}$ between loci i and j represents aggregated contacts of four different types of interactions, more precisely one of the two homologous beads associated to locus i gets in contact with one of the two homologous beads associated to locus j, see Fig. 1. This means that the Hi-C data is *unphased*. *Phased* Hi-C data that distinguishes contacts for homologs is rare. In our setting, we assume that the data is *partially phased*, i.e., some of the contact counts can be associated with a homolog. For example, in the (mouse) Patski (BL6xSpretus) (Deng et al. 2015; Ye and Ma 2020) cell line, 35.6% of the contact counts are phased; while this value is as low as 0.14% in the human GM12878 cell line (Rao et al. 2014; Ye and Ma 2020). Therefore, methods for inferring diploid 3D chromatin structure need to take into account the ambiguity of diploid Hi-C data to avoid inaccurate reconstructions.

Methods for 3D genome reconstruction in diploid organisms have been studied in Tan et al. (2018); Ye and Ma (2020); Cauer et al. (2019); Luo et al. (2020); Belyaeva et al. (2022); Lindsly et al. (2021); Segal (2022). One approach is to phase Hi-C data (Tan et al. 2018; Luo et al. 2020; Lindsly et al. 2021), for example by assigning haplotypes to contacts based on assignments at neighboring contacts (Tan et al. 2018; Lindsly et al. 2021). Cauer et al. (2019) and Ye and Ma (2020) model contact counts as Poisson random variables. To find the optimal 3D chromatin structure, Cauer et al. maximize the associated likelihood function combined with two structural constraints. The first constraint imposes that the distances between neighboring beads are similar, and the second one requires that homologous chromosomes are located in different

regions of the cell nucleus. On the other hand, Ye and Ma first compute the maximum likelihood estimate of model parameters for each of the homologs separately; these estimates are then refined by estimating the distance between the homologs. Belyaeva et al. (2022) show identifiability of the 3D structure when the Euclidean distances between neighboring beads and higher-order contact counts between three or more loci simultaneously are given. Under these assumptions, the 3D reconstruction is obtained by combining distance geometry with semidefinite programming. Segal (2022) applies recently developed imaging technology, in situ genome sequencing (IGS) (Payne et al. 2021), to point out issues in the assumptions made in Tan et al. (2018); Cauer et al. (2019); Belyaeva et al. (2022), and suggests as alternative assumptions that intra-homolog distances are smaller than corresponding inter-homolog distances and intra-homolog distances are similar for homologous chromosomes. IGS (Payne et al. 2021) provides yet another method for inferring the 3D structure of the genome, however, at present the resolution and availability of IGS data is limited.

Contributions In this work, we focus on a distance-based approach for partially phased Hi-C data. In particular, we assume that contacts only for some loci are phased. In the string of beads model, the locations of the pair of beads associated to i-th loci are denoted by $x_i, y_i \in \mathbb{R}^3$. Then homologs are represented by two sequences $x_1, x_2, \ldots, x_n$ and $y_1, x_2, \ldots, y_n$ in $\mathbb{R}^3$; see Fig. 1. Inferring the 3D chromatin structure corresponds to estimating the bead coordinates. Based on Lieberman-Aiden et al. (2009), we assume the power law dependency $c_{i,j} = \gamma d_{i,j}^{\alpha}$, where α is a negative conversion factor, between the distance $d_{i,j}$ and contact count $c_{i,j}$ of loci i and j. Following Cauer et al. (2019), we assume that a contact count between loci is given by the sum of all possible contact counts between the corresponding beads. We call a bead unambiguous if the contacts for the corresponding locus are phased; otherwise, we call a bead ambiguous.

Our first main contribution is to show that for negative rational conversion factors α, knowing the locations of six unambiguous beads ensures that there are generically finitely many possible locations for the other beads, both in the noiseless (Theorem 1) and noisy (Corollary 1) setting. Moreover, we prove finite identifiability also in the fully ambiguous setting when $\alpha = -2$ and the number of loci is at least 12 (Theorem 2). Note that the identifiability does not hold for $\alpha = 2$ as shown in Belyaeva et al. (2022).

Our second main contribution is to provide a reconstruction method when $\alpha = -2$, based on semidefinite programming combined with numerical algebraic geometry and local optimization (Sect. 4). The general idea is the following: We first estimate the coordinates of the unambiguous beads using only the unambiguous contact counts (which precisely corresponds to the haploid setting) using the SDP-based solver implemented in ChromSDE (Zhang et al. 2013). We then exploit our theoretical result on finite identifiability to estimate the coordinates of the ambiguous beads, one by one, by solving several polynomial systems numerically. These estimates are then improved by a local estimation step considering all contact counts. Finally, a clustering algorithm is used to overcome the symmetry $(x_i, y_i) \mapsto (y_i, x_i)$ in the estimation for the ambiguous beads.

The paper is organized as follows. In Sect. 2, we introduce our mathematical model for the 3D genome reconstruction problem. In Sect. 3, we recall identifiability results in the unambiguous setting (Sect. 3.1) and then prove identifiability results in the

partially ambiguous setting (Sect. 3.2) and in the fully ambiguous setting (Sect. 3.3). We describe our reconstruction method in Sect. 4. We test the performance of our method on synthetic datasets and on a real dataset from the mouse X chromosomes in Sect. 5. We conclude with a discussion about future research directions in Sect. 6.

2 Mathematical Model for 3D Genome Reconstruction

In this section we introduce the distance-based model under which we study 3D genome reconstruction. In Sect. 2.1 we give the background on contact count matrices. In Sect. 2.2 we describe a power-law between contacts and distances, which allows to translate the information about contacts into distances.

2.1 Contact Count Matrices

We model the genome as a string of $2n$ beads, corresponding to n pairs of homologous beads. The positions of the beads are recorded by a matrix

$$Z = [x_1, \ldots, x_n, y_1, \ldots, y_n]^T \in \mathbb{R}^{2n\times 3}.$$

The positions x_i and y_i correspond to homologous beads. When convenient, we use the notation $z_1 := x_1, \ldots, z_n := x_n, z_{n+1} := y_1, \ldots, z_{2n} := y_n$. In this notation,

$$Z = [z_1, \ldots, z_n, z_{n+1}, \ldots, z_{2n}]^T \in \mathbb{R}^{2n\times 3}.$$

Let U be the subset of pairs that are unambiguous, i.e., beads in the pair can be distinguished, and let A be the subset of pairs that are ambiguous, i.e., beads in the pair cannot be distinguished. The sets U and A form a partition of $[n]$.

A Hi-C matrix C is a matrix with each row and column corresponding to a genomic locus. Following Cauer et al. (2019), we call these contact counts ambiguous and denote the corresponding contact count matrix by C^A. If parental genotypes are available, then one can use SNPs to map some reads to each haplotype (Deng et al. 2015; Minajigi et al. 2015; Rao et al. 2014). If both ends of a read contains SNPs that can be associated to a single parent, then the contact count is called unambiguous and the corresponding contact count matrix is denoted by C^U. Finally, if only one of the genomic loci present in an interaction can be mapped to one of the homologous chromosomes, then the count is called partially ambiguous and the contact count matrix is denoted by C^P.

The unambiguous count matrix C^U is a $2n \times 2n$ matrix with the first n indices corresponding to $x_1, \ldots, x_n$ and the last n indices corresponding to $y_1, \ldots, y_n$. The ambiguous count matrix C^A is an $n \times n$ matrix and we assume that each ambiguous count is the sum of four unambiguous counts:

$$c^A_{i,j} = c^U_{i,j} + c^U_{i,j+n} + c^U_{i+n,j} + c^U_{i+n,j+n}.$$

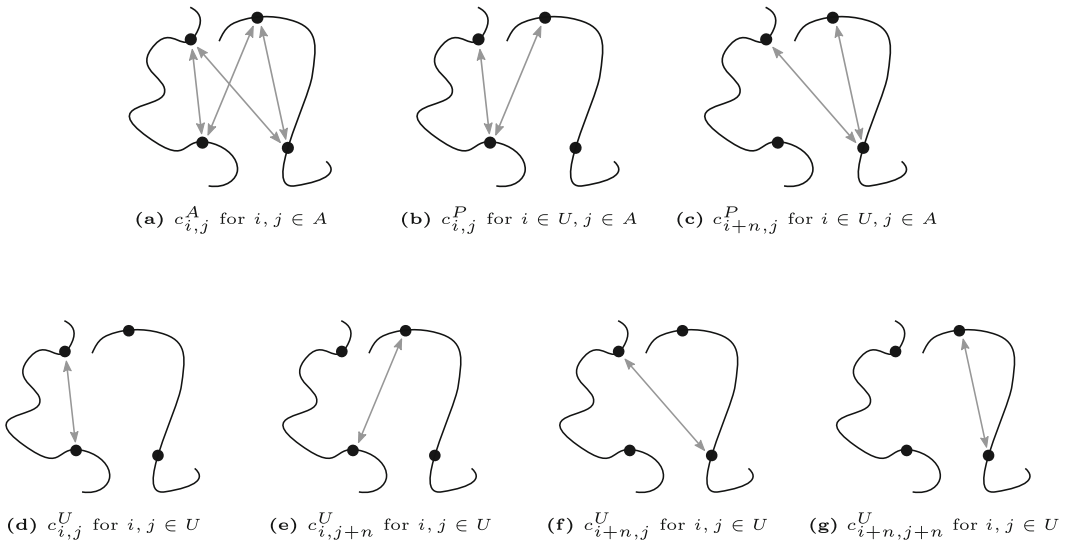


Fig. 2 Seven different types of contacts between the ith and jth locus

The partially ambiguous count matrix C^P is a $2n \times n$ matrix and each partially ambiguous count is the sum of two unambiguous counts:

$$c^P_{i,j} = c^U_{i,j} + c^U_{i,j+n}.$$

2.2 Contacts and Distances

Denoting the distance $\|z_i - z_j\|$ between z_i and z_j by $d_{i,j}$, the power law dependency observed by Lieberman-Aiden et al. (2009) can be written as

$$c^U_{i,j} = \gamma d^{\alpha}_{i,j}, \tag{1}$$

where $\alpha < 0$ is a conversion factor and $\gamma > 0$ is a scaling factor. This relationship between contact counts and distances is assumed in Belyaeva et al. (2022); Zhang et al. (2013), while in Cauer et al. (2019); Varoquaux et al. (2014) the contact counts $c_{i,j}$ are modeled as Poisson random variables with the Poisson parameter being $\beta d^{\alpha}_{i,j}$.

In our paper, we assume that contact counts are related to distances by (1). Similarly to Belyaeva et al. (2022), we set $\gamma = 1$ and in parts of the article $\alpha = -2$. In general, the conversion factor α depends on a dataset and its estimation can be part of the reconstruction problem (Varoquaux et al. 2014; Zhang et al. 2013). Setting $\gamma = 1$ means that we recover the configuration up to a scaling factor. In practice, the configuration can be rescaled using biological knowledge, e.g., the radius of the nucleus.

Our approach to 3D genome reconstruction builds on the power law dependency between contacts and distances between unambiguous beads. We convert the empirical

contact counts to Euclidean distances and then aim to reconstruct the positions of beads from the distances. This leads us to the following system of equations:

$$\begin{cases} c^A_{i,j} = \|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha + \|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha & \forall i, j \in A \\ c^P_{i,j} = \|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha, \ c^P_{i+n,j} = \|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha & \forall i \in U, j \in A \\ c^U_{i,j} = \|x_i - x_j\|^\alpha, \ c^U_{i,j+n} = \|x_i - y_j\|^\alpha, & \\ c^U_{i+n,j} = \|y_i - x_j\|^\alpha, \ c^U_{i+n,j+n} = \|y_i - y_j\|^\alpha & \forall i, j \in U \end{cases} \tag{2}$$

If α is an even integer, then (2) is a system of rational equations.

Determining the points x_i, y_i, where $i \in U$, is the classical Euclidean distance problem: We know the (noisy) pairwise distances between points and would like to construct the locations of points, see Sect. 3.1 for details. Hence after Sect. 3.1 we assume that we have estimated the locations of points x_i, y_i, where $i \in U$, and we would like to determine the points x_i, y_i, where $i \in A$.

3 Identifiability

In this section, we study the uniqueness of the solutions of the system (2) up to rigid transformations (translations, rotations and reflections), or in other words, the identifiability of the locations of beads. We study the unambiguous, partially ambiguous and ambiguous settings in Sects. 3.1, 3.2 and 3.3, respectively.

3.1 Unambiguous Setting and Euclidean Distance Geometry

If all pairs are unambiguous, i.e., $U = [n]$, then constructing the original points translates to a classical problem in Euclidean distance geometry. The principal task in Euclidean distance geometry is to construct original points from pairwise distances between them. In the rest of the subsection, we will recall how to solve this problem. Since pairwise distances are invariant under translations, rotations and reflections (rigid transformations), then the original points can be reconstructed up to rigid transformations. For an overview of distance geometry and Euclidean distance matrices, we refer the reader to Dokmanic et al. (2015), Krislock and Wolkowicz (2012), Liberti et al. (2014) and Mucherino et al. (2012).

The Gram matrix of the points $z_1, \ldots, z_{2n}$ is defined as

$$G = ZZ^T = [z_1, \ldots, z_{2n}]^T \cdot [z_1, \ldots, z_{2n}] \in \mathbb{R}^{2n \times 2n}.$$

Let $\overline{z} = \frac{1}{2n} \sum_{i=1}^{2n} z_i$ and $\tilde{z}_i = z_i - \overline{z}$ for $i = 1, \ldots, 2n$. The matrix $\tilde{Z} = [\tilde{z}_1, \ldots, \tilde{z}_{2n}]^T$ gives the locations of points after centering them around the origin. Let $\tilde{G}$ denote the Gram matrix of the centered point configuration $\tilde{z}_1, \ldots, \tilde{z}_{2n}$.

Let $D_{i,j} = \|z_i - z_j\|^2$ denote the squared Euclidean distance between the points z_i and z_j. The Euclidean distance matrix of the points $z_1, \ldots, z_{2n}$ is defined as $D = (D_{i,j})_{1 \le i,j \le 2n} \in \mathbb{R}^{2n \times 2n}$. To express the centered Gram matrix in terms of the Euclidean distance matrix, we define the geometric centering matrix

$$J = I_{2n} - \frac{1}{2n}\mathbf{1}\mathbf{1}^T,$$

where I_{2n} is the $2n \times 2n$ identity matrix and $\mathbf{1}$ is the vector of ones. The linear relationship between $\tilde{G}$ and D is given by

$$\tilde{G} = -\frac{1}{2} J D J.$$

Therefore, given the Euclidean distance matrix, we can construct the centered Gram matrix for the points $z_1, \ldots, z_{2n}$.

The centered points up to rigid transformations are extracted from the centered Gram matrix $\tilde{G}$ using the eigendecomposition $\tilde{G} = Q \Lambda Q^{-1}$, where Q is orthonormal and Λ is a diagonal matrix with entries ordered in decreasing order $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{2n} \ge 0$. We define $\Lambda_3^{1/2} := [\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}), \mathbf{0}_{3 \times (2n-3)}]^T$ and set $\hat{Z} = Q \Lambda_3^{1/2}$. In the case of noiseless distance matrix D, the Gram matrix $\tilde{G}$ has rank three and the diagonal matrix Λ has precisely three non-zero entries. Hence we could obtain $\hat{Z}$ also from $Q \Lambda^{1/2}$ by truncating zero columns. Using $\Lambda_3^{1/2}$ has the advantage that it gives an approximation for the points also for a noisy distance matrix D. The uniqueness of $\hat{Z}$ up to rotations and reflections follows from Krislock (2010, Proposition 3.2) which states that $AA^T = BB^T$ if and only if $A = BQ$ for some orthogonal matrix Q.

The procedure that transforms the distance matrix to origin centered Gram matrix and then uses eigendecomposition for constructing original points is called classical multidimensional scaling (cMDS) (Cox and Cox 2008). Although cMDS is widely used in practice, it does not always find the distance matrix that minimizes the Frobenius norm to the empirical noisy distance matrix (Sonthalia et al. 2021). Other approaches to solving the Euclidean distance and Euclidean completion problems include non-convex (Fang and O'Leary 2012; Mishra et al. 2011) as well semidefinite formulations (Alfakih et al. 1999; Fazel et al. 2003; Nie 2009; Weinberger et al. 2007; Zhang et al. 2013; Zhou et al. 2020).

3.2 Partially Ambiguous Setting

The next theorem establishes the uniqueness of the solutions of the system (2) in the presence of ambiguous pairs. In particular, it states that there are finitely many possible locations for beads in one ambiguous pair given the locations of six unambiguous beads. The identifiability results in this subsection hold for all negative rational numbers α. In the rest of the paper, we denote the true but unknown coordinates by x^* and the symbol x stands for a variable that we want to solve for. We write $\|\cdot\|$ for the standard inner product on $\mathbb{R}^3$.

Theorem 1 *Let α be a negative rational number. Then for $a^*, b^*, \ldots, f^*, x^*, y^* \in \mathbb{R}^3$ sufficiently general, the system of six equations*

$$\|x - t^*\|^\alpha + \|y - t^*\|^\alpha = \|x^* - t^*\|^\alpha + \|y^* - t^*\|^\alpha \text{ for } t^* = a^*, b^*, \ldots, f^* \quad (3)$$

in the six unknowns $x_1, x_2, x_3, y_1, y_2, y_3 \in \mathbb{R}$ has only finitely many solutions.

Remark 1 The proof will show that this system has only finitely many solutions over the *complex* numbers.

We believe that the theorem holds for general nonzero rational α. Indeed, our argument works, with a minor modification, also for $\alpha > 2$, but for α in the range $(0, 2]$ a refinement of the argument is needed.

Proof First write $Q(x) := x_1^2 + x_2^2 + x_3^2$, so that $\|x\| = \sqrt{Q(x)}$ for $x \in \mathbb{R}^3$. The advantage of Q over $\|x\|$ is that it is well-defined on $\mathbb{C}^3$.

Write $\frac{\alpha}{2} = \frac{m}{n}$ with m, n relatively prime integers, $m \neq 0$, and $n > 0$. Consider the affine variety $X \subseteq (\mathbb{C}^3)^8 \times (\mathbb{C}^2)^6$ consisting of all tuples

$$((a^*, \ldots, f^*, x^*, y^*), (r_{t^*}, s_{t^*})_{t^* = a^*, \ldots, f^*})$$

such that

$$Q(x^* - t^*)^m = r_{t^*}^n \neq 0 \text{ and } Q(y^* - t^*)^m = s_{t^*}^n \neq 0 \text{ for } t^* = a^*, \ldots, f^*.$$

Note that, if x^*, t^* are real, then it follows that

$$Q(x^* - t^*)^m = (\|x^* - t^*\|^\alpha)^n,$$

and similarly for $Q(y^* - t^*)$. Hence if $a^*, \ldots, y^*$ are all real, then the point

$$((a^*, \ldots, f^*, x^*, y^*), (\|x^* - t^*\|^\alpha, \|y^* - t^*\|^\alpha)_{t^*}) \quad (4)$$

is a point in X with real-valued coordinates.

The projection π from X to the open affine subset $U \subseteq (\mathbb{C}^3)^8$ where all $Q(x^* - t^*)$ and $Q(y^* - t^*)$ are nonzero is a finite morphism with fibers of cardinality n^{12}; to see this cardinality note that there are n possible choices for each of the numbers r_{t^*}, s_{t^*}. Each irreducible component of X is a smooth variety of dimension 24.

Consider the map $\psi : X \to (\mathbb{C}^3 \times \mathbb{C}^1)^6$ defined by

$$((a^*, \ldots, f^*, x^*, y^*), (r_{t^*}, s_{t^*})_{t^*}) \mapsto ((t^*, r_{t^*} + s_{t^*}))_{t^*}$$

We claim that for q in some open dense subset of X, the derivative $d_q\psi$ has full rank 24. For this, it suffices to find one point $p \in U$ such that $d_q\psi$ has rank 24 at each of the n^{12} points $q \in \pi^{-1}(p)$. We take a real-valued point $p := (a^*, b^*, \ldots, f^*, x^*, y^*) \in (\mathbb{R}^3)^8$ to be specified later on. Let $q \in \pi^{-1}(p)$. Then, near q, the map ψ factorises

via π and the unique algebraic map $\psi' : U \to (\mathbb{C}^3 \times \mathbb{C}^1)^6$ (defined near p) which on a neighborhood of p in $U \cap (\mathbb{R}^3)^8$ equals

$$\psi'(a, \ldots, f, x, y) = ((t, \xi_{t^*} \cdot Q(x-t)^{\alpha/2} + \eta_{t^*} \cdot Q(y-t)^{\alpha/2}))_{t=a,\ldots,f} \in (\mathbb{C}^3 \times \mathbb{C}^1)^6$$

where ξ_{t^*} and ζ_{t^*} are n-th roots of unity in $\mathbb{C}$ depending on which q is chosen among the n^{12} points in $\pi^{-1}(p)$. The situation is summarised in the following diagram:

$$\begin{array}{ccc} (X, q) & & \\ \downarrow \pi & \searrow^{\psi} & \\ (U, p) & \xrightarrow[\psi']{} & ((\mathbb{C}^3 \times \mathbb{C}^1)^6, \psi(q)). \end{array}$$

Now, $d_q\psi = d_p\psi' \circ d_q\pi$, and since $d_q\pi$ is a linear isomorphism, it suffices to prove that $d_p\psi'$ is a linear isomorphism. Suppose that $(a', \ldots, f', x', y') \in \ker d_p\psi'$. Then, since the map ψ' remembers $a, \ldots, f$, it follows immediately that $a' = \ldots = f' = 0$. On the other hand, by differentiating we find that, for each $t^* \in \{a^*, \ldots, f^*\}$,

$$\begin{aligned} &\xi_{t^*} \cdot (\alpha/2) \cdot Q(x^* - t^*)^{\alpha/2-1} \cdot 2 \cdot \langle x', x^* - t^* \rangle \\ +&\eta_{t^*} \cdot (\alpha/2) \cdot Q(y^* - t^*)^{\alpha/2-1} \cdot 2 \cdot \langle y', y^* - t^* \rangle = 0, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ stands for the standard bilinear form on $\mathbb{C}^3$. In other words, the vector $(x', y') \in \mathbb{C}^6$ is in the kernel of the 6×6-matrix

$$M := \begin{bmatrix} \|x^* - a^*\|^{\alpha-2} \cdot \xi_{a^*} \cdot (x^* - a^*) & \|y^* - a^*\|^{\alpha-2} \cdot \eta_{a^*} \cdot (y^* - a^*) \\ \vdots & \vdots \\ \|x^* - f^*\|^{\alpha-2} \cdot \xi_{f^*} \cdot (x^* - f^*) & \|y^* - f^*\|^{\alpha-2} \cdot \eta_{f^*} \cdot (y^* - f^*) \end{bmatrix}$$

where we have interpreted $a^*, \ldots, f^*, x^*, y^*$ as row vectors. It suffices to show that, for some specific choice of $p = (a^*, \ldots, f^*, x^*, y^*) \in (\mathbb{R}^3)^8$, this matrix is nonsingular *for all* n^{12} *choices of* $((\xi_{t^*}, \eta_{t^*}))_{t^*}$.

We choose $a^*, \ldots, f^*, x^*, y^*$ as the vertices of the unit cube, as follows:

$$\begin{array}{lll} a^* = (1,0,0) & b^* = (0,1,0) & c^* = (0,0,1) \\ c^* = (0,1,1) & d^* = (1,0,1) & f^* = (1,1,0) \\ x^* = (0,0,0) & y^* = (1,1,1). & \end{array}$$

Then the matrix M becomes, with $\beta = \alpha - 2$:

$$\begin{bmatrix} -\xi_{a^*} & 0 & 0 & 0 & 2^{\frac{\beta}{2}} \cdot \eta_{a^*} & 2^{\frac{\beta}{2}} \cdot \eta_{a^*} \\ 0 & -\xi_{b^*} & 0 & 2^{\frac{\beta}{2}} \cdot \eta_{b^*} & 0 & 2^{\frac{\beta}{2}} \cdot \eta_{b^*} \\ 0 & 0 & -\xi_{c^*} & 2^{\frac{\beta}{2}} \cdot \eta_{c^*} & 2^{\frac{\beta}{2}} \cdot \eta_{c^*} & 0 \\ 0 & -(2^{\frac{\beta}{2}} \cdot \xi_{d^*}) & -(2^{\frac{\beta}{2}} \cdot \xi_{d^*}) & \eta_{d^*} & 0 & 0 \\ -(2^{\frac{\beta}{2}} \cdot \xi_{e^*}) & 0 & -(2^{\frac{\beta}{2}} \cdot \xi_{e^*}) & 0 & \eta_{e^*} & 0 \\ -(2^{\frac{\beta}{2}} \cdot \xi_{f^*}) & -(2^{\frac{\beta}{2}} \cdot \xi_{f^*}) & 0 & 0 & 0 & \eta_{f^*} \end{bmatrix}.$$

Now, $\det(M)$ equals

$$-\xi_{a^*} \cdot \xi_{b^*} \cdot \xi_{c^*} \cdot \eta_{d^*} \cdot \eta_{e^*} \cdot \eta_{f^*} + 2^{2+3\beta} \cdot \eta_{a^*} \cdot \eta_{b^*} \cdot \eta_{c^*} \cdot \xi_{d^*} \cdot \xi_{e^*} \cdot \xi_{f^*} + 2^{2\beta} \cdot R \tag{5}$$

where R is a sum of (products of) roots of unity. Now $\alpha < 0$ implies that $\beta < -2$, so that $2 + 3\beta < 2\beta < 0$. Since roots of unity have 2-adic valuation 0, the second term in the expression above is the unique term with minimal 2-adic valuation. Hence $\det(M) \neq 0$, as desired.

It follows that ψ is a dominant morphism from each irreducible component of X into $(\mathbb{C}^3 \times \mathbb{C}^1)^6$, and hence for all q in an open dense subset of X, the fiber $\psi^{-1}(\psi(q))$ is finite. This then holds, in particular, for q in an open dense subset of the real points as in (4). This proves the theorem. □

Remark 2 If $\alpha > 2$, then $\beta > 0$, and hence the unique term with minimal 2-adic valuation in (5) is the first term. This can be used to show that the theorem holds then, as well. The only subtlety is that for positive α, solutions where x or y equal one of the points $a^*, \ldots, f^*$ are not automatically excluded, and these are not seen by the variety X. But a straightforward argument shows that such solutions do not exist for sufficiently general choices of $a^*, \ldots, f^*, x^*, y^*$.

We now consider the setting when we know locations of seven unambiguous beads. In the special case when $\alpha = -2$, we construct the ideal generated by the polynomials obtained from rational Eqs. (3) for seven unambiguous beads after moving all terms to one side and clearing the denominators. Based on symbolic computations in `Macaulay2` for the degree of this ideal, we conjecture that the location of a seventh unambiguous bead guarantees unique identifiability of an ambiguous pair of beads:

Conjecture 1 *Let $a^*, b^*, c^*, d^*, e^*, f^*, g^*, x^*, y^* \in \mathbb{R}^3$ be sufficiently general. The system of rational equations*

$$\frac{1}{\|t^* - x^*\|^2} + \frac{1}{\|t^* - y^*\|^2} = \frac{1}{\|t^* - x\|^2} + \frac{1}{\|t^* - y\|^2} \text{ for } t^* = a^*, b^*, c^*, d^*, e^*, f^*, g^* \tag{6}$$

has precisely two solutions (x^*, y^*) and (y^*, x^*).

In practice, we only have noisy estimates $a, b, \ldots, f \in \mathbb{R}^3$ of the true positions of unambiguous beads $a^*, b^*, \ldots, f^* \in \mathbb{R}^3$, and we have noisy observations c_t of the true contact counts $c_t^* := \|x^* - t^*\|^\alpha + \|y^* - t^*\|^\alpha$. We aim to find $x, y \in \mathbb{R}^3$ such that

$$\|x - t\|^\alpha + \|y - t\|^\alpha = c_t \text{ for } t = a, b, \ldots, f.$$

We may write $c_t = \|x^* - t\|^\alpha + \|y^* - t\|^\alpha + \epsilon_t$ for some ϵ_t that depends on the noise level. Hence, the above system of equations can be rephrased as

$$\|x - t\|^\alpha + \|y - t\|^\alpha = \|x^* - t\|^\alpha + \|y^* - t\|^\alpha + \epsilon_t \text{ for } t = a, b, \ldots, f. \quad (7)$$

In the following corollary we show that this system has generically finitely many solutions.

Corollary 1 *Let α be a negative rational number. Then for $a, b, \ldots, f, x^*, y^* \in \mathbb{R}^3$ and $\epsilon_a, \epsilon_b, \ldots, \epsilon_f \in \mathbb{R}$ sufficiently general, the system of six equations*

$$\|x - t\|^\alpha + \|y - t\|^\alpha = \|x^* - t\|^\alpha + \|y^* - t\|^\alpha + \epsilon_t \text{ for } t = a, b, \ldots, f \quad (8)$$

in the six unknowns $x_1, x_2, x_3, y_1, y_2, y_3 \in \mathbb{R}$ has only finitely many solutions.

Proof Recall the map $\psi : X \to (\mathbb{C}^3 \times \mathbb{C}^1)^6$ from the proof of Theorem 1 defined by

$$((a, \ldots, f, x^*, y^*), (r_{x^*,t}, s_{y^*,t})_t) \mapsto ((t, r_{x^*,t} + s_{y^*,t}))_t.$$

We showed that ψ is a dominant morphism from each irreducible component of X into $(\mathbb{C}^3 \times \mathbb{C}^1)^6$, and that each irreducible component of X is 24-dimensional. Every solution to (8) is the (x, y)-component of a point in the fiber

$$\psi^{-1}((t, ||x^* - t||^\alpha + ||y^* - t||^\alpha + \epsilon_t))_t.$$

Since this is a fiber over a sufficiently general point, the fiber is finite. □

Corollary 1 will be the basis of a numerical algebraic geometric based reconstruction method in Sect. 4.

3.3 Ambiguous Setting

Finally we consider the ambiguous setting, where one would like to reconstruct the locations of beads only from ambiguous contact counts. It is shown in Belyaeva et al. (2022) that for $\alpha = 2$, one does not have finite identifiability no matter how many pairs of ambiguous beads one considers. We show finite identifiability for the locations of beads given contact counts for 12 pairs of ambiguous beads for $\alpha = -2$ in both the noisy and noiseless setting. We believe that the result might be true for further conversion factors α's, however our proof technique does not directly generalize.

Theorem 2 *Let $\alpha = -2$. Then for $(c_{ij})_{1\le i<j\le 12} \in \mathbb{R}^{66}$ sufficiently general, the system of 66 equations*

$$\|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha + \|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha = c_{ij} \text{ for } 1 \le i < j \le 12 \quad (9)$$

in the 72 unknowns $x_{1,1}, x_{1,2}, x_{1,3}, y_{1,1}, y_{1,2}, y_{1,3}, \ldots, x_{12,1}, x_{12,2}, x_{12,3}, y_{12,1}, y_{12,2}, y_{12,3} \in \mathbb{R}$ has only finitely many solutions up to rigid transformations. In particular, it holds that for sufficiently general $(x_1^*, y_1^*, \ldots, x_{12}^*, y_{12}^*) \in (\mathbb{R}^3)^{24}$, the system

$$\begin{aligned}&\|x_i - x_j\|^\alpha + \|x_i - y_j\|^\alpha + \|y_i - x_j\|^\alpha + \|y_i - y_j\|^\alpha = \\ &\|x_i^* - x_j^*\|^\alpha + \|x_i^* - y_j^*\|^\alpha + \|y_i^* - x_j^*\|^\alpha + \|y_i^* - y_j^*\|^\alpha \text{ for } 1 \le i < j \le 12\end{aligned} \quad (10)$$

has finitely many solutions up to rigid transformation.

Proof As before, we write $Q(x) := x_1^2 + x_2^2 + x_3^2$, so that $\|x\| = \sqrt{Q(x)}$ for $x \in \mathbb{R}^3$. Consider the affine open subset $X \subseteq (\mathbb{C}^3)^{24}$ consisting of all tuples $(x_1^*, y_1^*, \ldots, x_{12}^*, y_{12}^*)$ such that

$$Q(x_i^* - x_j^*) \neq 0, \quad Q(x_i^* - y_j^*) \neq 0, \quad Q(y_i^* - x_j^*) \neq 0 \text{ and } Q(y_i^* - y_j^*) \neq 0 \text{ for } i < j.$$

Consider also the map $\psi : X \to \mathbb{C}^{66}$ defined by

$$(x_1^*, \ldots, y_{12}^*) \mapsto \left(Q(x_i^* - x_j^*)^{-1} + Q(x_i^* - y_j^*)^{-1} + Q(y_i^* - x_j^*)^{-1} + Q(y_i^* - y_j^*)^{-1}\right)_{i<j}.$$

By a computer calculation (with exact arithmetic) we found that at a randomly chosen $q \in X$ with rational coordinates, the derivative $d_q\psi$ had full rank 66. It then follows that for q in some open dense subset of X, $d_q\psi$ has rank 66. Hence ψ is dominant, and for any sufficiently general $c \in \mathbb{C}^{66}$, all irreducible components of the fiber $\psi^{-1}(c)$ have dimension 6. Moreover, each such component C is preserved by the 6-dimensional connected group $G = SO(3, \mathbb{C}) \ltimes \mathbb{C}^3$.

The stabilizer in G of a sufficiently general point in X is zero-dimensional. This follows from a Lie algebra argument: if a point $(x_1^*, y_1^*, \ldots, x_{12}^*, y_{12}^*) \in X$ has a positive-dimensional stabilizer in G, then there is a nonzero element A in the Lie algebra of $SO(3, \mathbb{C})$ that maps all the differences $x_i^* - x_j^*, x_i^* - y_j^*, y_i^* - y_j^*$ to zero. Since A is a skew-symmetric matrix and hence of rank 2, it follows that all points x_i^*, y_j^* lie on a line. The variety of such collinear tuples has dimension 28, so it does not map dominantly to $\mathbb{C}^{66}$. Hence there exists a Zariski open dense subset $V \subseteq \mathbb{C}^{66}$ such that for all $c \in V$, the fiber $\psi^{-1}(c)$ contains no points with positive-dimensional stabilizers in G, and hence $\psi^{-1}(c)$ is a disjoint union of finitely many 6-dimensional G-orbits. Likewise, $\psi^{-1}(V)$ is a Zariski open dense subset of $(\mathbb{C}^3)^{24}$ such that $\psi^{-1}(\psi(q))$ consists of finitely many G-orbits for all $q \in \psi^{-1}(V)$. With this, we have proven the complex analog of the theorem.

To obtain the statement over the real numbers, we note that if $c \in V$ has real-valued coordinates, then a finite number of the G-orbits that make up $\psi^{-1}(c)$ contain a real-valued tuple. If $G \cdot q$ for $q \in (\mathbb{R}^3)^{24}$ is such an orbit, it holds that $(G \cdot q) \cap (\mathbb{R}^3)^{24} = (SO(3, \mathbb{R}) \ltimes \mathbb{R}^3) \cdot q$ whenever the 24 points that make up the tuple q are not coplanar. The set of coplanar configurations form a subset of X of dimension 51, and does therefore not map dominantly to $\mathbb{C}^{66}$. Hence, by shrinking V appropriately, we can assume that no fibers above it contain coplanar configurations. In particular, this means that the real part of the fiber over any real point in V consists of a finitely many orbits under the action of $SO(3, \mathbb{R}) \ltimes \mathbb{R}^3$, as desired. □

Remark 3 A standard numerical algebraic geometry computation with monodromy and the certification techniques of Breiding et al. (2023), using `HomotopyContinuation.jl` (see, e.g., Sturmfels and Telen (2021)), proves that the system (8) generically has more than 1000 complex solutions up to the action of $O(3, \mathbb{C}) \ltimes \mathbb{C}^3$ and the symmetries $(x_i, y_i) \mapsto (y_i, x_i)$ for $i = 1, \ldots, 12$. This constitutes theoretical motivation for working with partially phased data, even if we, in principle, have finite identifiability already from the unphased data.

Remark 4 When $\alpha = 2$, which corresponds to the setting studied in Belyaeva et al. (2022), then computationally we found that for some special choices of $x_1^*, y_1^*, \ldots, x_{12}^*, y_{12}^* \in \mathbb{R}^3$ the rank of the Jacobian matrix in Theorem 2 is 42. This is consistent with the fact that Theorem 2 fails for $\alpha = 2$ (Belyaeva et al. 2022).

4 A New Reconstruction Method

In this section, we outline a new approach to diploid 3D genome reconstruction for partially phased data, based on the theoretical results discussed in subsection 3.2. The method consists of the following main steps:

1. Estimation of the unambiguous beads $\{x_i, y_i\}_{i \in U}$ through semidefinite programming (discussed in Sect. 4.1).
2. A preliminary estimation of the ambiguous beads using numerical algebraic geometry, based on Corollary 1 (discussed in Sect. 4.2).
3. A refinement of this estimation using local optimization (discussed in Sect. 4.3).
4. A final clustering step, where we disambiguate between the estimations (x_i, y_i) and (y_i, x_i) for each $i \in A$, based on the assumption that homolog chromosomes are separated in space (discussed in Sect. 4.4).

In what follows, we will refer to this method by the acronym SNLC (formed from the initial letters in semidefinite programming, numerical algebraic geometry, local optimization and clustering).

4.1 Estimation of the Positions of Unambiguous Beads

As discussed in Sect. 3.1, the unambiguous bead coordinates $\{x_i, y_i\}_{i \in U} = \{z_i\}_{i \in U \cup (n+U)}$ can be estimated with semidefinite programming. More specifically, we

use ChromSDE Zhang (2013, Section 2.1) for this part of our reconstruction, which relies on a specialized solver from Jiang et al. (2014), to solve an SDP relaxation of the optimization problem

$$\min_{\{z_i\}_{i\in U\cup(n+U)}} \sum_{\substack{i,j\in U\cup(n+U)\\ c_{ij}^U\neq 0}} \sqrt{c_{ij}^U}\left(\frac{1}{c_{ij}^U} - \|z_i - z_j\|^2\right)^2 + \lambda \sum_{\substack{i,j\in U\cup(n+U)\\ c_{ij}^U=0}} \|z_i - z_j\|^2 \tag{11}$$

with $\lambda = 0.01$ (cf. Zhang, et al. (2013, Eq. 4)). The terms in the first sum are weighted by the square root for the corresponding contact counts, in order to account for the fact that higher counts can be assumed to be less susceptible to noise.

4.2 Preliminary Estimation Using Numerical Algebraic Geometry

To estimate the coordinates of the ambiguous beads $\{x_i, y_i\}_{i\in A}$, we will use a method based on numerical equation solving, where we estimate the ambiguous bead pairs one by one.

Let x, y be the unknown coordinates in $\mathbb{R}^3$ of a pair of ambiguous beads. We pick six unambiguous beads with already estimated coordinates $a, b, c, d, e, f \in \mathbb{R}^3$. For each $t \in \{a, \ldots, f\}$, let $c_t \in \mathbb{R}$ be the corresponding partially ambiguous counts between t and the ambiguous bead pair (x, y). Clearing the denominators in the system (8), we obtain a system of polynomial equations

$$\|x - t\|^2 + \|y - t\|^2 = c_t\|x - t\|^2\|y - t\|^2 \text{ for } t = a, b, c, d, e, f. \tag{12}$$

By Corollary 1, this system has finitely many complex solutions both in the noiseless and noisy setting, which can be found using homotopy continuation.

We observe that the system (12) generally has 80 complex solutions, and we only expect one pair of solutions $(x, y), (y, x)$ to correspond to an accurate estimation. Naively adding another polynomial arising from a seventh unambiguous bead (as in Conjecture 1) does not work; in the noisy setting this over-determined system typically lacks solutions. Instead, we compute an estimation based on the following two heuristic assumptions:

1. The most accurate estimation should be *approximately real*, in the sense that the max-norm of the imaginary part is below a certain tolerance (in this work, 0.15 was used for the experiments in both Sects. 5.1 and 5.2). The choice of this threshold was made based on analysing the imaginary parts of solutions to (12) for various choices of unambiguous beads, see Fig. 9.
2. The most accurate estimation should be consistent when we change the choice of six unambiguous beads.

Based on these assumptions, we apply the following strategy. We make a number $N \geq 2$, choices of sets of six unambiguous beads, and solve the corresponding N square systems of the form (12). Since larger contact counts can be expected to have

smaller relative noise, we make the choices of beads among the 20 unambiguous beads t that have highest contact count c_t to the ambiguous locus at hand. For each system, we pick out the approximately real solutions, and obtain N sets $\mathcal{S}_1, \ldots, \mathcal{S}_N \subseteq \mathbb{R}^6$ consisting of the real parts of the approximately real solutions. Up to the symmetry $(x, y) \mapsto (y, x)$, we expect these sets to have a unique "approximately common" element. We therefore compute, by an exhaustive search, the tuple $(w_1, \ldots, w_N) \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_N$ that minimizes the sum

$$\left\| w_1 - \frac{w_1 + \cdots + w_N}{N} \right\| + \cdots + \left\| w_N - \frac{w_1 + \cdots + w_N}{N} \right\|,$$

and use $\frac{w_1+\cdots+w_N}{N}$ as our estimation of (x, y). For the computations presented in Sect. 5, we use $N = 5$.

To solve the systems, we use the Julia package *HomotopyContinuation.jl* (Breiding et al. 2018), and follow the two-phase procedure described in Sommese and Wampler (2005, Sect. 7.2). For the first phase, we solve (12) with randomly chosen parameters $a^*, \ldots, f^* \in \mathbb{C}^3$ and $c_{a^*}, \ldots, c_{f^*} \in \mathbb{C}$, using a polyhedral start system (Huber and Sturmfels 1995). We trace 1280 paths in this first phase, since the Newton polytopes of the polynomials appearing in the system (12) all contain the origin, and have a mixed volume of 1280, which makes 1280 an upper bound on the number of complex solutions by Li (1996, Theorem 2.4). For the second phase, we use a straight-line homotopy in parameter space from the randomly chosen parameters $a^*, \ldots, f^* \in \mathbb{C}^3$ and $c_{a^*}, \ldots, c_{f^*} \in \mathbb{C}$, to the values $a, \ldots, f$ and $c_a, \ldots, c_f \in \mathbb{C}$ at hand. We observe that we generally find 80 complex solutions in the first phase, which means 40 orbits with respect to the symmetry $(x, y) \mapsto (y, x)$. By the discussion in Sommese, (2005, Sect. 7.6) it is enough to only trace one path per orbit, so in the end, we only trace 40 paths in the second phase.

Remark 5 If the noise levels are sufficiently high, there could be choices of six unambiguous beads for which the system lacks approximately-real solutions. If this situation is encountered, we try to redraw the six unambiguous beads until we find an approximately-real solution. If this does not succeed within a certain number of attempts (100 in the experiments conducted for this paper), we use the average of the closest neighboring unambiguous beads instead.

4.3 Local Optimization

A disadvantage of the numerical algebraic geometry based estimation discussed in the previous subsection is that it only takes into account "local" information about the interactions for one ambiguous locus at a time, which might make it more sensitive to noise. In our proposed method, we therefore refine this preliminary estimation of $\{x_i, y_i\}_{i \in A}$ further in a local optimization step that takes into account the "global" information of all available data.

The idea is to estimate $\{x_i, y_i\}_{i\in A}$ by solving the optimization problem

$$\min_{\{x_i,y_i\}_{i\in A}} \sum_{i\in U, j\in A} \left(\left(c^P_{i,j} - \frac{1}{\|x_i - x_j\|^2} - \frac{1}{\|x_i - y_j\|^2} \right)^2 + \left(c^P_{i+n,j} - \frac{1}{\|y_i - x_j\|^2} - \frac{1}{\|y_i - y_j\|^2} \right)^2 \right) \tag{13}$$

while keeping the estimates of $\{x_i, y_i\}_{i\in U}$ from the ChromSDE step fixed. We use the quasi-Newton method for unconstrained optimization implemented in the Matlab Optimization Toolbox for this step. The already estimated coordinates of $\{x_i, y_i\}_{i\in A}$ from the numerical algebraic geometry step are used for the initialization.

4.4 Clustering to Break Symmetry

Our objective function remains invariant if we exchange x_i and y_i for any $i \in A$. We can break symmetry by relying on the empirical observation that homologous chromosomes typically are spatially separated in different so-called compartments of the nucleus (Eagen 2018). Let $(\bar{x}_i, \bar{y}_i)_{i=1}^n$ denote the estimates from the previous steps. Our final estimations will be obtained by solving the minimization problem

$$\min_{\{x_i,y_i\}_{i\in A}} \sum_{i=1}^{n-1} g_{i,i+1}(x, y), \text{ with } g_{i,i+1}(x, y) := \left(\|x_i - x_{i+1}\|^2 + \|y_i - y_{i+1}\|^2 \right), \tag{14}$$

where $(x_i, y_i) = (\bar{x}_i, \bar{y}_i)$ for $i \in U$ are fixed, and $(x_i, y_i) \in \{(\bar{x}_i, \bar{y}_i), (\bar{y}_i, \bar{x}_i)\}$ for $i \in A$ are the optimization variables. The optimal solution can be computed efficiently, as explained next.

We first decompose the problem into contiguous chunks of ambiguous beads. Let $(i_1, \ldots, i_L) := U$ be the indices of the unambiguous beads and let $i_0 := 1, i_{L+1} := n$. The optimization problem can be phrased as

$$\min_{\{x_i,y_i\}_{i\in A}} \sum_{\ell=0}^{L} G_\ell(x, y), \quad \text{with} \quad G_\ell(x, y) := \sum_{i=i_\ell}^{i_{\ell+1}-1} g_{i,i+1}(x, y) \tag{15}$$

where there is one summand $G_\ell(x, y)$ for each contiguous chunk of ambiguous beads. Since the summands $G_\ell(x, y)$ do not share any ambiguous bead, we can minimize them independently.

We proceed to describe the optimal solution of the problem. Let

$$s_i = \begin{cases} 1, & \text{if } (x_i, y_i) = (\bar{x}_i, \bar{y}_i) \\ -1, & \text{if } (x_i, y_i) = (\bar{y}_i, \bar{x}_i) \end{cases}, \qquad w_{i,i+1} = (\bar{x}_i - \bar{y}_i)^T (\bar{x}_{i+1} - \bar{y}_{i+1}).$$

The variable s_i indicates whether we keep using $(\bar{x}_i, \bar{y}_i)$ or we reverse it. Note that $s_i = 1$ for $i \in U$. The next lemma gives the optimal assignment of s_i for $i \in A$. This assignment is constructed by using inner products $w_{i,i+1}$.

Lemma 1 *The optimal solution of (14) can be constructed as follows:*

1. *For the last chunk ($\ell = L$) we have*

$$s^*_{i_\ell} = 1, \qquad s^*_{i+1} = \operatorname{sgn}(w_{i,i+1}) s^*_i \quad \text{for } i = i_\ell, i_\ell{+}1, \ldots, i_{\ell+1}{-}1$$

   where $\operatorname{sgn}(\cdot)$ *is the sign function and* $\operatorname{sgn}(0)$ *can be either 1 or* -1.
2. *For the first chunk ($\ell = 0$) we have*

$$s^*_{i_{\ell+1}} = 1, \qquad s^*_i = \operatorname{sgn}(w_{i,i+1}) s^*_{i+1} \quad \text{for } i = i_{\ell+1}{-}1, i_{\ell+1}{-}2, \ldots, i_\ell$$

3. *For any other chunk, let k be the index of the smallest absolute value $|w_{k,k+1}|$, among $i_\ell \le k \le i_{\ell+1} - 1$. The solution is*

$$\begin{aligned} s^*_{i_\ell} &= 1, \qquad & s^*_{i+1} &= \operatorname{sgn}(w_{i,i+1}) s^*_i \quad \text{for } i = i_\ell, i_\ell{+}1, \ldots, k{-}1 \\ s^*_{i_{\ell+1}} &= 1, \qquad & s^*_i &= \operatorname{sgn}(w_{i,i+1}) s^*_{i+1} \quad \text{for } i = i_{\ell+1}{-}1, i_{\ell+1}{-}2, \ldots, k{+}1 \end{aligned}$$

Proof Denoting $\bar{u}_i := \frac{1}{2}(\bar{x}_i + \bar{y}_i)$, $\bar{v}_i := \frac{1}{2}(\bar{x}_i - \bar{y}_i)$, then $x_i = u_i + s_i v_i$, $y_i = u_i - s_i v_i$. Note that

$$\begin{aligned} \|\bar{x}_i\|^2 + \|\bar{y}_i\|^2 + \|\bar{x}_{i+1}\|^2 + \|\bar{y}_{i+1}\|^2 - g_{i,i+1}(x, y) &= 2(x_i^T x_{i+1} + y_i^T y_{i+1}) \\ &= 2(\bar{u}_i + s_i \bar{v}_i)^T (\bar{u}_{i+1} + s_{i+1} \bar{v}_{i+1}) + 2(\bar{u}_i - s_i \bar{v}_i)^T (\bar{u}_{i+1} - s_{i+1} \bar{v}_{i+1}) \\ &= 4(\bar{u}_i^T \bar{u}_{i+1}) + 4(\bar{v}_i^T \bar{v}_{i+1}) s_i s_{i+1} \\ &= 4(\bar{u}_i^T \bar{u}_{i+1}) + w_{i,i+1} s_i s_{i+1} \end{aligned}$$

Since $\bar{x}_i$, $\bar{y}_i$, $\bar{u}_i$, $\bar{v}_i$ are constants, minimizing $g_{i,i+1}(x, y)$ is equivalent to maximizing $w_{i,i+1} s_i s_{i+1}$. Then for each chunk we have to solve the optimization problem

$$\max_{s_i \in \{1,-1\}} \sum_{i=i_\ell}^{i_{\ell+1}-1} w_{i,i+1} s_i s_{i+1} \,, \tag{16}$$

The formulas from the first and last chunk are such that $w_{i,i+1} s^*_i s^*_{i+1} \ge 0$ for all i. This is possible because in these cases only one of the endpoints has a fixed value, and the remaining values are computed recursively starting from such a fixed point. Since all summands are nonnegative, the sum in (16) is maximized.

For the inner chunks, the two endpoints are fixed, so it may not be possible to have that $w_{i,i+1} s^*_i s^*_{i+1} \ge 0$ for all indices. In an optimal assignment we should pick at most one term to be negative, and such a term (if it exists) should be the one with the smallest absolute value $|w_{i,i+1}|$. This leads to the formula from the lemma. □

5 Experiments

In this section, we apply the SNLC scheme described in Sect. 4 to synthetic and real datasets, and compare its performance with the preexisting software packages ASHIC (Ye and Ma 2020) and PASTIS (Cauer et al. 2019). We chose these two reconstruction methods for comparison because they are best suited for our setting. Also Belyaeva et al. (2022) and Tan et al. (2018) have reconstruction methods for diploid organisms, but the former method requires higher-order contact information and the latter method is targeted for single cell data.

All SNLC experiments are done using Julia 1.6.1, with ChromSDE being run in Matlab 2021a, and the Julia package `MATLAB.jl` (v0.8.3) acting as interface between Julia and Matlab. The numerical algebraic geometry part of the estimation procedure is done with `HomotopyContinuation.jl` (v2.5.5) (Breiding et al. 2018). The PASTIS experiments are run in Python 3.8.10, and the ASHIC experiments in Python 3.10.5.

For the PASTIS computations, we fix $\alpha = -2$ to ensure compatibility with the modelling assumptions made in this paper. We run PASTIS without filtering, in order to make it possible to compare RMSD values. Since PASTIS only takes integer inputs, we multiply the theoretical contact counts calculated by (2) by a factor 10^5 and round them to the nearest integer. Following the approach taken in Cauer et al. (2019), we use a coarse grid search to find the optimal coefficients for the homolog separating constraint and bead connectivity constraints. Specifically, we fix a structure simulated with the same method as used in the experiments, and compute the RMSD values for all $\lambda_1, \lambda_2 \in \{1, 10^1, 10^2, \ldots, 10^{12}\}$. In this way, we find that $\lambda_1 = 10^{11}$ and $\lambda_2 = 10^{12}$ give optimal results.

For the ASHIC computations, we use the ASHIC-ZIPM method, which has the lowest distance error rate among the ASHIC's models according to Ye (2020, Fig. 2) and models the contact counts as a zero-inflated Poisson distribution (ZIP) to account for the sparsity of the Hi-C matrix. We run ASHIC without filtering out any loci and with the setting `|aggregate|` to ensure that the coordinates of all beads are estimated.

5.1 Synthetic Data

We conduct a number of experiments where we simulate a single chromosome pair (referred to as X and Y in figures) through Brownian motion with fixed step length, compute unambiguous, partially ambiguous and ambiguous contact counts according to (2), add noise, and then try to recover the structure of the chromosomes through the SNLC scheme described in Sect. 4. Following (Belyaeva et al. 2022), we model noise by multiplying each entry of C^U, C^P and C^A by a factor $1 + \delta$, where δ is sampled uniformly from the interval $(-\varepsilon, \varepsilon)$ for some chosen noise level $\varepsilon \in [0, 1]$.

As a measure of the quality of the reconstruction, we use the minimal root-mean square distance (RMSD) between, on the one hand, the true coordinates $(x_i^*, y_i^*)_{i=1}^n$, and, on the other hand, the estimated coordinates $(x_i, y_i)_{i=1}^n$ after rigid transformations and scaling, i.e., we find the minimum

$$\min_{\substack{R\in O(3)\\ s>0,\ b\in\mathbb{R}^3}} \sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left(\|(sRx_i+b)-x_i^*\|^2+\|(sRy_i+b)-y_i^*\|^2\right)}.$$

This can be seen as a version of the classical Procrustes problem solved in Schönemann (1966), which is implemented in Matlab as the function `procrustes`.

Specific examples of reconstructions of the Brownian motion and helix-shaped chromosomes obtained with SNLC at varying noise levels and 50% of ambiguous beads are shown in Fig. 3. For low noise levels the reconstructions by SNLC and the original structure highly overlap. For higher noise levels the general region occupied by the reconstructions overlaps with the original structure, while the local features become less aligned. Analogous reconstructions obtained with SNLC without the local optimization step are shown in Fig. 6 in Appendix.

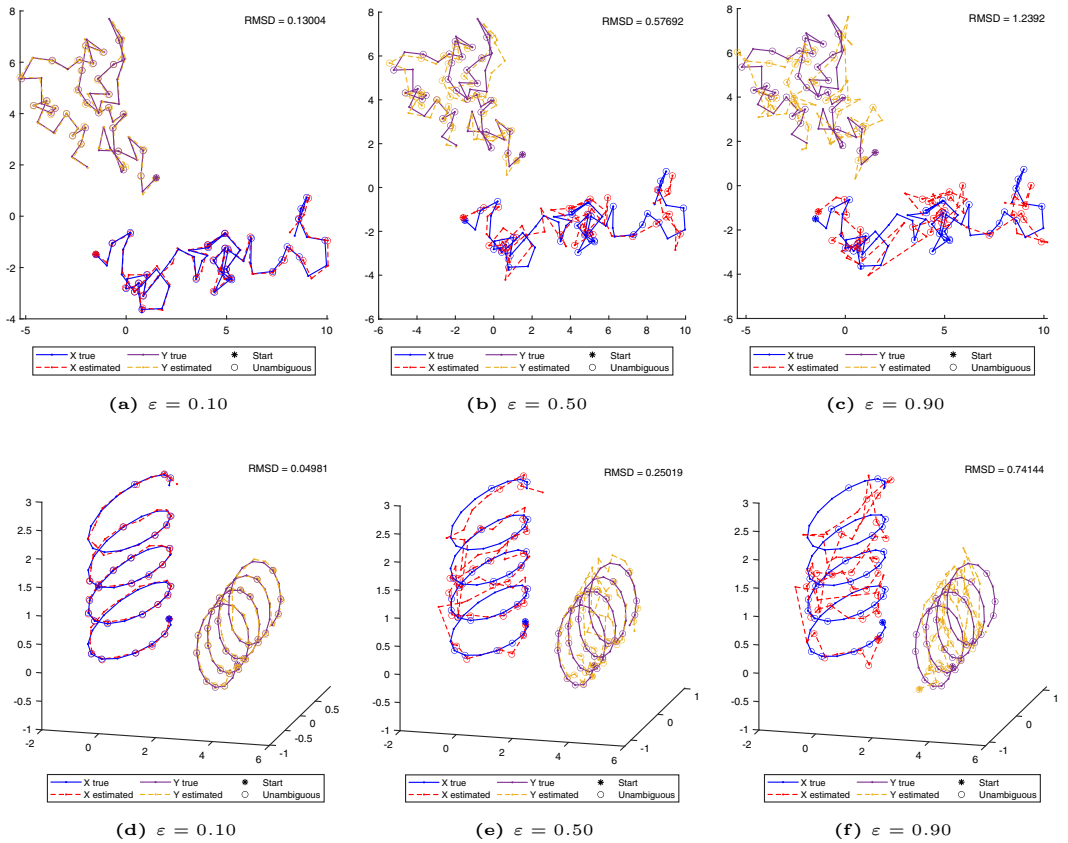
A comparison of how the quality of the reconstruction depends on the noise level and proportion of ambiguous beads for SNLC, ASHIC and PASTIS is done in Fig. 4. We measure the RMSD value between the reconstructed and original 3D structure for different noise levels over 20 runs. The RMSD values obtained by SNLC are consistently lower than the ones obtained by ASHIC and PASTIS. The difference is specially large for low to medium noise levels. While our method outperforms ASHIC and PASTIS in the setting considered in this paper, it is worth mentioning that ASHIC and PASTIS work also in a more general setting, where there might be contacts of all three types (ambiguous, partially ambiguous and unambiguous) between every pair of loci.

5.2 Experimentally Obtained Data

We compute SNLC reconstructions based on the real dataset explored in Cauer et al. (2019), which is obtained from Hi-C experiments on the X chromosomes in the Patski (BL6xSpretus) cell line. The data has been recorded at a resolution of 500 kb, which corresponds to 343 bead pairs in our model.

For some of these pairs, no or only very low contact counts have been recorded. Since such low contact counts are susceptible to high uncertainty and can be assumed to be a consequence of experimental errors, we exclude the 47 loci with the lowest total contact counts from the analysis. To select the cutoff, the loci are sorted according to the total contact counts (see Fig. 7a in Appendix), and the ratios between the total contact counts for consecutive loci are computed. A peak for these ratios is observed at the 47th contact count, as shown in Fig. 7b in Appendix. After applying this filter, we obtain a dataset with 296 loci. Out of these, we consider as ambiguous all loci i for which less than 40% of the total contact count comes from contacts where x_i and y_i were not distinguishable. These proportions for all loci are shown in Fig. 7c in Appendix. For the Patski dataset, we obtain 46 ambiguous loci and 250 unambiguous loci in this way.

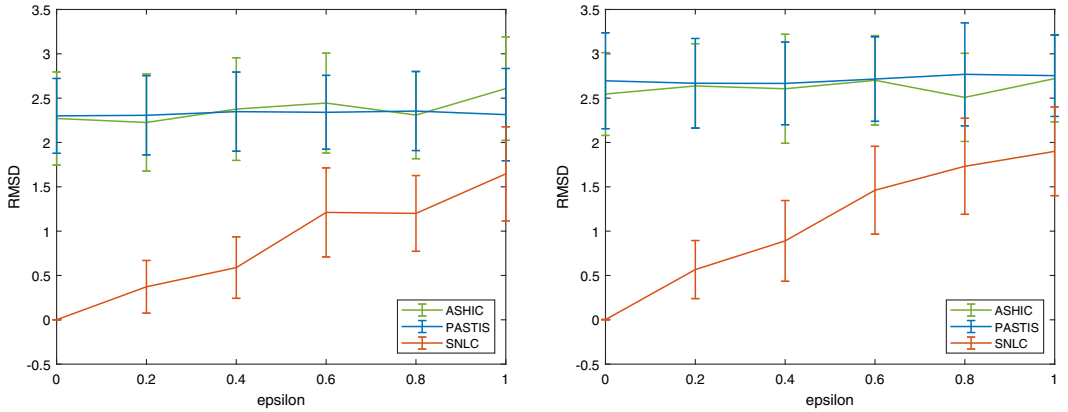
In the Patski dataset, a locus can simultaneously participate in unambiguous, partially ambiguous and ambiguous contacts. To obtain the setting of our paper where loci are partitioned into unambiguous or ambiguous, we reassign the contacts according to



(a) $\varepsilon = 0.10$ (b) $\varepsilon = 0.50$ (c) $\varepsilon = 0.90$

(d) $\varepsilon = 0.10$ (e) $\varepsilon = 0.50$ (f) $\varepsilon = 0.90$

Fig. 3 Examples of reconstructions for varying noise levels, for a chromosome pair with 60 loci, out of which 50% are ambiguous. **a**–**c** Show chromosomes simulated with Brownian motion (projected onto the xy-plane), whereas **d**–**e** show helix-shaped chromosomes (color figure online)



(a) 25% ambiguous loci (b) 50% ambiguous loci

Fig. 4 Comparison between our reconstruction method, ASHIC and PASTIS. The values are the average over 20 runs, with the error bars showing the standard deviation. All experiments took place with 60 loci, with varying levels of noise, as well as varying numbers of ambiguous loci, uniformly randomly distributed over the chromosomes (color figure online)

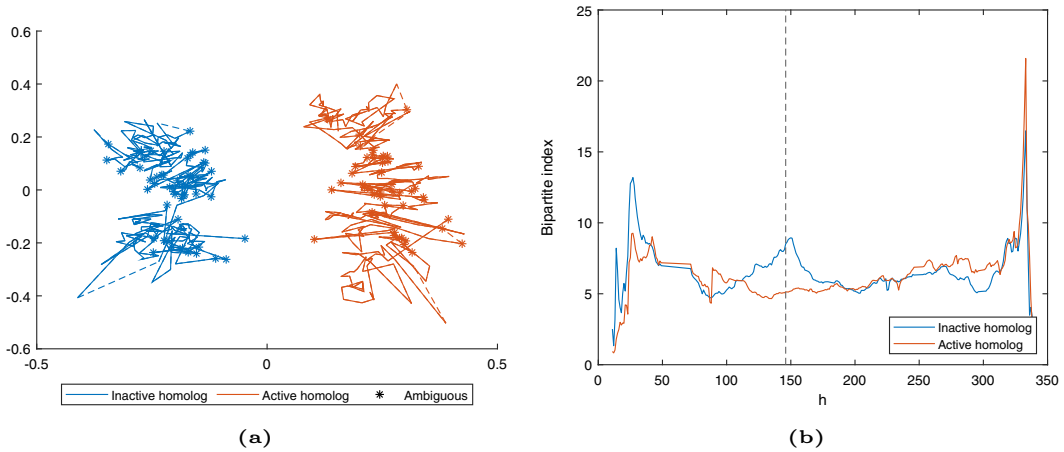


Fig. 5 **a** Reconstruction from a real dataset using our reconstruction method. A dashed line between two beads is used to indicate that there is one or more beads between them, for which we have not given an estimation (due to low contact counts). **b** Bipartite index for the reconstructed chromosomes. The dashed vertical line indicates the known hinge point at locus 146 (color figure online)

whether a locus is unambiguous or ambiguous. Our reassignment method is motivated by the assignment of haplotype to unphased Hi-C reads in Lindsly et al. (2021). The exact formulas are given in Appendix.

The reconstruction obtained via SNLC can be found in Fig. 5a. The logarithmic heatmaps for contact count matrices for original data and the SNLC reconstruction are shown in Fig. 8.

It was discovered in Deng et al. (2015) that the inactive homolog in the Patski X chromosome pair has a bipartite structure, consisting of two superdomains with frequent intra-chromosome contacts within the superdomains and a boundary region between the two superdomains. The active homolog does not exhibit the same behaviour. The boundary region on the inactive X chromosome is centered at 72.8−72.9 MB (Deng et al. 2015) which at the 500 kB resolution corresponds to the bead 146 (Cauer et al. 2019). We show in Fig. 5b that the two chromosomes reconstructed using SNLC exhibit this structure by computing the bipartite index for the respective homologs as in Cauer et al. (2019); Deng et al. (2015). We recall that, in the setting of a single chromosome with beads $z_1, \ldots, z_n \in \mathbb{R}^3$, the bipartite index is defined as the ratio of intra-superdomain to inter-superdomain contacts in the reconstruction:

$$BI(h) = \frac{\frac{1}{h^2}\sum_{i=1}^{h}\sum_{j=1}^{h}\frac{1}{\|z_i - z_j\|^2} + \frac{1}{(n-h)^2}\sum_{i=h+1}^{n}\sum_{j=h+1}^{n}\frac{1}{\|z_i - z_j\|^2}}{\frac{2}{h(n-h)}\sum_{i=1}^{h}\sum_{j=h+1}^{n}\frac{1}{\|z_i - z_j\|^2}}.$$

6 Discussion

In this article we study the finite identifiability of 3D genome reconstruction from contact counts under the model where the distances $d_{i,j}$ and contact counts $c_{i,j}$ between two beads i and j follow the power law dependency $c_{i,j} = d_{i,j}^{\alpha}$ for a conversion factor $\alpha < 0$. We show that if at least six beads are unambiguous, then the locations of the rest of the beads can be finitely identified from partially ambiguous contact counts for rational α satisfying $\alpha < 0$ or $\alpha > 2$. In the fully ambiguous setting, we prove finite identifiability for $\alpha = -2$, given ambiguous contact counts for at least 12 pairs of beads. From Belyaeva et al. (2022) it is known that finite identifiability does not hold in the fully ambiguous setting for $\alpha = 2$. It is an open question whether finite identifiability of 3D genome reconstruction holds for other $\alpha \in \mathbb{R}\backslash\{-2, 2\}$ in the fully ambiguous setting and for rational $\alpha \in (0, 2]$ in the partially ambiguous setting. We conjecture that in the partially ambiguous setting seven unambiguous loci guarantee unique identifiability of the 3D reconstruction for rational $\alpha < 0$ or $\alpha > 2$. When $\alpha = -2$, then one approach to studying the unique identifiability might be via the degree of a parametrized family of algebraic varieties.

After establishing the identifiability, we suggest a reconstruction method for the partially ambiguous setting with $\alpha = -2$ that combines semidefinite programming, homotopy continuation in numerical algebraic geometry, local optimization and clustering. To speed up the homotopy continuation based part, we observe that the parametrized system of polynomial equations corresponding to six unambiguous beads has 40 pairs of complex solutions and we trace one path for each orbit. It is an open question to prove that for sufficiently general parameters the system has 40 pairs of complex solution. This question again reduces to studying the degree of a family of algebraic varieties. While our goal is to highlight the potential of our method, one could further regularize its output and use interpolation for the beads that are far away from the neighboring beads. A future research direction is to explore whether numerical algebraic geometry or semidefinite programming based methods can be proposed also for other conversion factors $\alpha < 0$.

Supplementary information

The code for computations and experiments is available at https://github.com/kaiekubjas/3D-genome-reconstruction-from-partially-phased-HiC-data.

Acknowledgements We thank Anastasiya Belyaeva, Gesine Cauer, AmirHossein Sadegemanesh, Luca Sodomaco, and Caroline Uhler for very helpful discussions and answers to our questions.

Funding Open Access funding provided by Aalto University. Oskar Henriksson and Kaie Kubjas were partially supported by the Academy of Finland Grant No. 323416. Oskar Henriksson was also partially funded by the Novo Nordisk project with grant reference number NNF20OC0065582.

Data Availibility The Patski dataset analyzed in Sect. 5.2 comes from the third-party repository https://noble.gs.washington.edu/proj/diploid-pastis/, and is based on the dataset GSE68992 from the Gene Expression Omnibus, available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68992.

Code Availability The code used for generating the synthetic data discussed in Sect. 5.1 is available in the GitHub repository https://github.com/kaiekubjas/3D-genome-reconstruction-from-partially-phased-HiC-data. This repository also contains the code used for the computations referred to in the discussion preceeding Conjecture 1, and in the proof of Theorem 2.

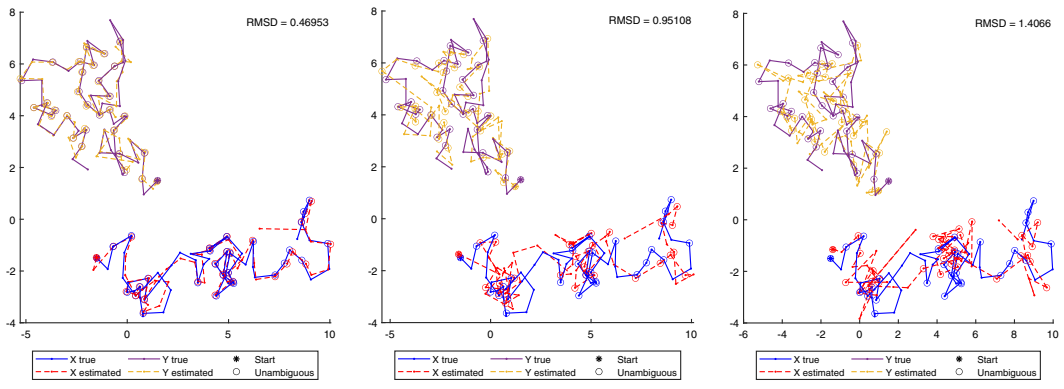
Appendix

In this part of the paper, we include additional details and figures for the experiments in Sect. 5.

Figure 6 shows reconstructions of the same chromosomes as displayed in Fig. 3 but without the local optimization step, indicating that semidefinite programming, numerical algebraic geometry and clustering alone can recover the main features of the 3D structure.

Figure 7 illustrates the preprocessing steps of the real dataset where loci with low contact counts are removed and the rest of the loci are partitioned into unambiguous and ambiguous. The total contact count for the ith locus is defined as the sum of all contacts where it participates:

$$T(i)=\sum_{j\in[n]}\Big(c^A(i,j)+c^P(i,j)+c^P(i+n,j)\Big)+\sum_{j\in[2n]}\Big(c^P(j,i)+c^U(i,j)+c^U(i+n,j)\Big).$$



(a) $\varepsilon = 0.10$ (b) $\varepsilon = 0.50$ (c) $\varepsilon = 0.90$

Fig. 6 SNLC reconstructions, without the local optimization step (color figure online)

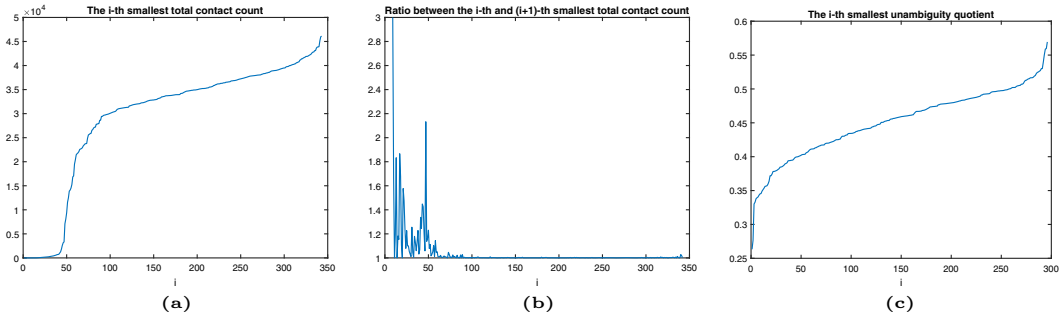


Fig. 7 **a** Total contact counts sorted in increasing order. **b** Ratios between total contact counts. The peak corresponding to the ratio between the 48th and the 47th smallest count is used as a motivation for excluding the 47 loci with smallest total contact from the analysis. **c** Unambiguity quotients for each of the remaining 296 loci, sorted in increasing order. We consider a locus as ambiguous if this ratio is less than 0.4; otherwise, we consider it as unambiguous (color figure online)

Similarly, we define the unambiguity quotient as the proportion of $T(i)$ that consists of contacts where x_i and y_i could be distinguished:

$$UQ(i) = \frac{1}{T(i)} \left(\sum_{j \in [n]} \left(c^P(i, j) + c^P(i+n, j) \right) + \sum_{j \in [2n]} \left(c^U(i, j) + c^U(i+n, j) \right) \right).$$

To obtain the setting of our paper where loci are partitioned into unambiguous or ambiguous, we reassign the contact counts of $\tilde{C}^U$ $\tilde{C}^P$ and $\tilde{C}^A$ of the Patski dataset according to whether a locus is unambiguous or ambiguous. For $i, j \in U$, we define

$$c^U_{i,j} = \tilde{c}^U_{i,j} + \tilde{c}^P_{i,j} \frac{\tilde{c}^U_{i,j}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i,j+n}} + \tilde{c}^P_{j,i} \frac{\tilde{c}^U_{i,j}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i+n,j}} + \tilde{c}^A_{i,j} \frac{\tilde{c}^U_{i,j}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i,j+n} + \tilde{c}^U_{i+n,j} + \tilde{c}^U_{i+n,j+n}},$$

$$c^U_{i,j+n} = \tilde{c}^U_{i,j+n} + \tilde{c}^P_{i,j} \frac{\tilde{c}^U_{i,j+n}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i,j+n}} + \tilde{c}^P_{j+n,i} \frac{\tilde{c}^U_{i,j+n}}{\tilde{c}^U_{i,j+n} + \tilde{c}^U_{i+n,j+n}} + \tilde{c}^A_{i,j} \frac{\tilde{c}^U_{i,j+n}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i,j+n} + \tilde{c}^U_{i+n,j} + \tilde{c}^U_{i+n,j+n}},$$

$$c^U_{i+n,j} = \tilde{c}^U_{i+n,j} + \tilde{c}^P_{i+n,j} \frac{\tilde{c}^U_{i+n,j}}{\tilde{c}^U_{i+n,j} + \tilde{c}^U_{i+n,j+n}} + \tilde{c}^P_{j,i} \frac{\tilde{c}^U_{i+n,j}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i+n,j}} + \tilde{c}^A_{i,j} \frac{\tilde{c}^U_{i+n,j}}{\tilde{c}^U_{i,j} + \tilde{c}^U_{i,j+n} + \tilde{c}^U_{i+n,j} + \tilde{c}^U_{i+n,j+n}},$$

$$c^U_{i+n,j+n} = \tilde{c}^U_{i+n,j+n} + \tilde{c}^P_{i+n,j} \frac{\tilde{c}^U_{i+n,j+n}}{\tilde{c}^U_{i+n,j} + \tilde{c}^U_{i+n,j+n}} + \tilde{c}^P_{j+n,i} \frac{\tilde{c}^U_{i+n,j+n}}{\tilde{c}^U_{i,j+n} + \tilde{c}^U_{i+n,j+n}} +$$

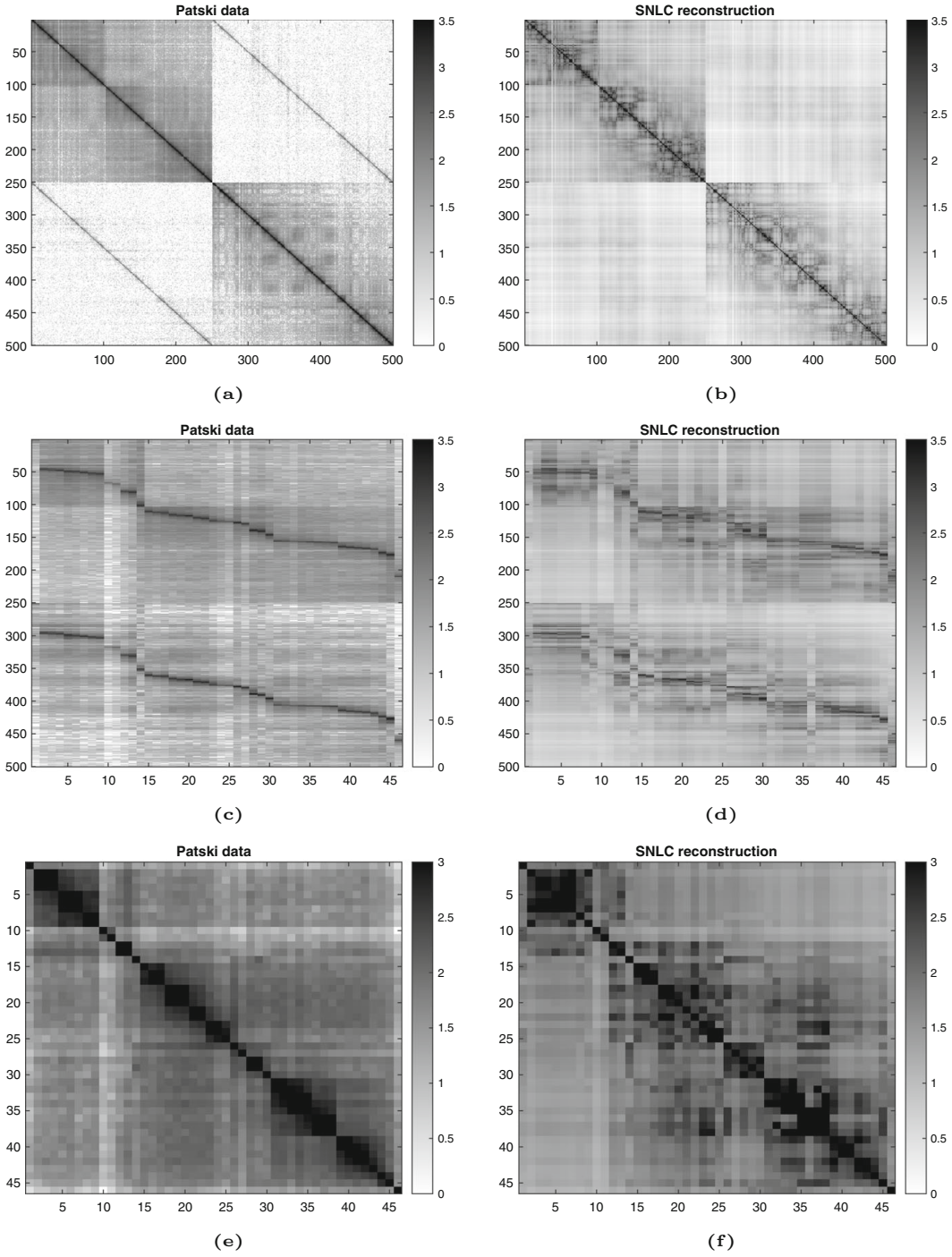


Fig. 8 Logarithmic heat maps for the reassigned contact count matrices obtained from the original Patski dataset and from the SNLC reconstruction: **a** and **b** C^U; **c** and **d** C^P; **e** and **f** C^A. The axis labels correspond to the 500 unambiguous beads, and the 46 ambiguous loci

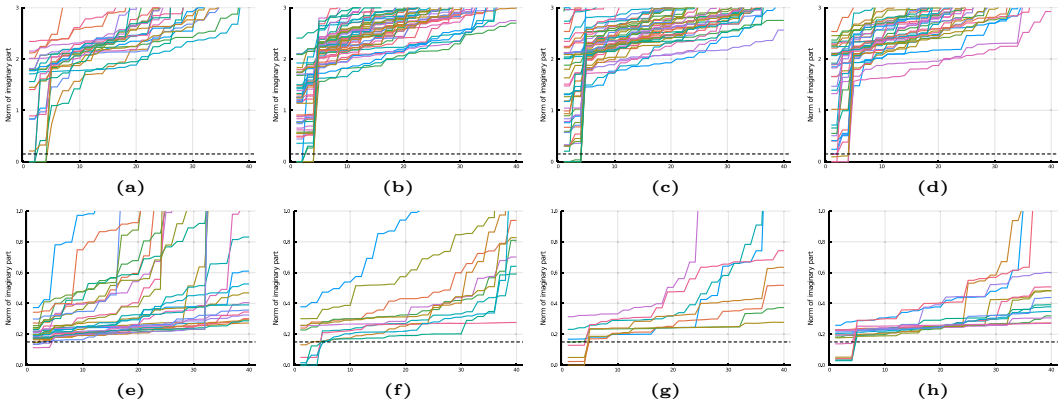


Fig. 9 Max-norm of the imaginary parts encountered in the numerical algebraic geometry estimation of various loci. Each subfigure corresponds to an ambiguous locus: **a**–**d** correspond to the first four loci of the synthetic dataset used in Fig. 3b; **e**–**h** correspond to the first four ambiguous loci of the Patski dataset. Each colored line corresponds to a specific choice of 6 unambiguous beads used in the estimation of the locus. Each line connects 40 points, that record the max-norm of the imaginary part of a solution (up to symmetry) found for the corresponding choice of 6 unambiguous beads. The dashed line at 0.15 corresponds to the choice of threshold for when a solution is considered approximately real. Similar figures for the rest of the ambiguous loci in the respective chromosome pairs can be found in the Github repository (color figure online)

$$+\tilde{c}^A_{i,j}\frac{\tilde{c}^U_{i+n,j+n}}{\tilde{c}^U_{i,j}+\tilde{c}^U_{i,j+n}+\tilde{c}^U_{i+n,j}+\tilde{c}^U_{i+n,j+n}}.$$

For $i \in U$, $j \in A$, we define

$$c^P_{i,j}=\tilde{c}^U_{i,j}+\tilde{c}^U_{i,j+n}+\tilde{c}^P_{i,j}+\tilde{c}^P_{j,i}\frac{\tilde{c}^U_{i,j}}{\tilde{c}^U_{i,j}+\tilde{c}^U_{i+n,j}}+\tilde{c}^P_{j+n,i}\frac{\tilde{c}^U_{i,j+n}}{\tilde{c}^U_{i,j+n}+\tilde{c}^U_{i+n,j+n}}+$$
$$+\tilde{c}^A_{i,j}\frac{\tilde{c}^P_{i,j}}{\tilde{c}^P_{i,j}+\tilde{c}^P_{i+n,j}},$$
$$c^P_{i+n,j}=\tilde{c}^U_{i+n,j}+\tilde{c}^U_{i+n,j+n}+\tilde{c}^P_{i+n,j}+\tilde{c}^P_{j,i}\frac{\tilde{c}^U_{i+n,j}}{\tilde{c}^U_{i,j}+\tilde{c}^U_{i+n,j}}+\tilde{c}^P_{j+n,i}\frac{\tilde{c}^U_{i+n,j+n}}{\tilde{c}^U_{i,j+n}+\tilde{c}^U_{i+n,j+n}}+$$
$$+\tilde{c}^A_{i,j}\frac{\tilde{c}^P_{i+n,j}}{\tilde{c}^P_{i,j}+\tilde{c}^P_{i+n,j}}.$$

Finally, for $i, j \in A$, we define

$$c^A_{i,j}=\tilde{c}^U_{i,j}+\tilde{c}^U_{i,j+n}+\tilde{c}^U_{i+n,j}+\tilde{c}^U_{i+n,j+n}+\tilde{c}^P_{i,j}+\tilde{c}^P_{i+n,j}+\tilde{c}^P_{j,i}+\tilde{c}^P_{j+n,i}+\tilde{c}^A_{i,j}.$$

In Fig. 8 in Appendix, the experimental contact counts from the Patski dataset are compared with the contact counts from the SNLC reconstruction.

Figure 9 shows how the max-norm of the imaginary part of the solutions varies between different instances of the system (12) used for the reconstruction in Fig. 3(b), and for the reconstruction from the Patski data in Fig. 5. A complete set of figures for

these two datasets can be found in the Github repository. Taken together, the figures indicate that a max-norm of 0.15 was an appropriate threshold for approximate realness for both data sets, in the sense that it is low enough to single out solutions that have significantly smaller imaginary parts than the others, while also ensuring that it is possible to find an approximately real solution for each ambiguous locus.

References

Alfakih AY, Khandani A, Wolkowicz H (1999) Solving euclidean distance matrix completion problems via semidefinite programming. Comput Optim Appl 12(1):13–30

Belyaeva A, Kubjas K, Sun LJ, Uhler C (2022) Identifying 3D genome organization in diploid organisms via Euclidean distance geometry. SIAM J Math Data Sci 4(1):204–228

Breiding P, Rose K, Timme S (2023) Certifying zeros of polynomial systems using interval arithmetic. ACM Trans Math Softw 49(1):1–14

Breiding P, Timme S (2018) HomotopyContinuation.jl: A package for homotopy continuation in Julia. In: Davenport JH, Kauers M, Labahn G, Urban J (eds) Mathematical Software—ICMS 2018. Springer, Cham, pp 458–465

Cauer AG, Yardimci G, Vert JP, Varoquaux N, Noble WS (2019) Inferring diploid 3D chromatin structures from Hi-C data. In: 19th International workshop on algorithms in bioinformatics (WABI 2019)

Cox MA, Cox TF (2008) Multidimensional scaling. In: Handbook of data visualization. Springer, Berlin, pp 315–347

Deng X, Ma W, Ramani V, Hill A, Yang F, Ay F, Berletch JB, Blau CA, Shendure J, Duan Z (2015) Bipartite structure of the inactive mouse X chromosome. Genome Biol 16(1):1–21

Dokmanic I, Parhizkar R, Ranieri J, Vetterli M (2015) Euclidean distance matrices: essential theory, algorithms, and applications. IEEE Signal Process Mag 32(6):12–30

Eagen KP (2018) Principles of chromosome architecture revealed by Hi-C. Trends Biochem Sci 43(6):469–478

Fang H-R, O'Leary DP (2012) Euclidean distance matrix completion problems. Optim Methods Softw 27(4–5):695–717

Fazel M, Hindi H, Boyd SP (2003) Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In: Proceedings of the 2003 American control conference, vol 3. IEEE, pp 2156–2162

Hu M, Deng K, Qin Z, Dixon J, Selvaraj S, Fang J, Ren B, Liu JS (2013) Bayesian inference of spatial organizations of chromosomes. PLoS Comput Biol 9(1):1002893

Huber B, Sturmfels B (1995) A polyhedral method for solving sparse polynomial systems. Math Comput 64(212):1541–1555

Jiang K, Sun D, Toh K-C (2014) A partial proximal point algorithm for nuclear norm regularized matrix least squares problems. Math Program Comput 6:1

Krislock N (2010) Semidefinite facial reduction for low-rank Euclidean distance matrix completion. PhD thesis, University of Waterloo. http://hdl.handle.net/10012/5093

Krislock N, Wolkowicz H (2012) Euclidean distance matrices and applications. Handbook on Semidefinite. Conic and Polynomial Optimization. Springer, New York, pp 879–914

Lafontaine DL, Yang L, Dekker J, Gibcus JH (2021) Hi-C 3.0: improved protocol for genome-wide chromosome conformation capture. Curr Protoc 1(7):198

Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J (2014) 3D genome reconstruction from chromosomal contacts. Nat Methods 11(11):1141–1143

Li T-Y, Wang X (1996) The BKK root count in $\mathbb{C}^n$. Math Comput 65(216):1477–1484

Li J, Lin Y, Tang Q, Li M (2021) Understanding three-dimensional chromatin organization in diploid genomes. Comput Struct Biotechnol J 19:3589

Liberti L, Lavor C, Maculan N, Mucherino A (2014) Euclidean distance geometry and applications. SIAM Rev 56(1):3–69

Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293

Lindsly S, Jia W, Chen H, Liu S, Ronquist S, Chen C, Wen X, Stansbury C, Dotson GA, Ryan C (2021) Functional organization of the maternal and paternal human 4D nucleome. IScience 24(12):103452

Luo H, Li X, Fu H, Peng C (2020) HiCHap: a package to correct and analyze the diploid hi-c data. BMC Genomics 21(1):1–13

Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Boukhali M, Haas W et al (2015) A comprehensive Xist interactome reveals Cohesin repulsion and an RNA-directed chromosome conformation. Science 349(6245):1

Mishra B, Meyer G, Sepulchre R (2011) Low-rank optimization for distance matrix completion. In: 2011 50th IEEE conference on decision and control and european control conference. IEEE, pp 4455–4460

Mucherino A, Lavor C, Liberti L, Maculan N (2012) Distance geometry: theory, methods, and applications. Springer, New York

Nie J (2009) Sum of squares method for sensor network localization. Comput Optim Appl 43(2):151–179

Nott A, Holtman IR, Coufal NG, Schlachetzki JC, Yu M, Hu R, Han CZ, Pena M, Xiao J, Wu Y (2019) Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science 366(6469):1134–1139

Oluwadare O, Highsmith M, Cheng J (2019) An overview of methods for reconstructing 3-d chromosome and genome structures from hi-c data. Biol Proced Online 21(1):1–20

Paulsen J, Sekelja M, Oldenburg AR, Barateau A, Briand N, Delbarre E, Shah A, Sørensen AL, Vigouroux C, Buendia B (2017) Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. Genome Biol 18(1):1–15

Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, Yao C-C, Markoulaki S, Earl AS, Labade AS, Jaenisch R (2021) In situ genome sequencing resolves DNA sequence and structure in intact biological samples. Science 371(6532):3446

Rajarajan P, Borrman T, Liao W, Schrode N, Flaherty E, Casiño C, Powell S, Yashaswini C, LaMarca EA, Kassim B et al (2018) Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. Science 362(6420):1

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680

Rhie SK, Schreiner S, Witt H, Armoskus C, Lay FD, Camarena A, Spitsyna VN, Guo Y, Berman BP, Evgrafov OV (2018) Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation. Sci Adv 4(12):8550

Rousseau M, Fraser J, Ferraiuolo MA, Dostie J, Blanchette M (2011) Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. Bioinformatics 12(1):414

Schönemann PH (1966) A generalized solution of the orthogonal Procrustes problem. Psychometrika 31(1):1–10

Segal MR (2022) Can 3D diploid genome reconstruction from unphased Hi-C data be salvaged? NAR Genom Bioinf 4(2):038

Sommese AJ, Wampler CW (2005) Numerical solution of systems of polynomials arising in engineering and science. World Scientific Publishing Company, Singapore

Sonthalia R, Van Buskirk G, Raichel B, Gilbert A (2021) How can classical multidimensional scaling go wrong? Adv Neural Inf Process Syst 34:12304–12315

Sturmfels B, Telen S (2021) Likelihood equations and scattering amplitudes. Algebr Stat 12(2):167–186

Tan L, Xing D, Chang C-H, Li H, Xie XS (2018) Three-dimensional genome structures of single diploid human cells. Science 361(6405):924–928

Uhler C, Shivashankar G (2017) Regulation of genome organization and gene expression by nuclear mechanotransduction. Nat Rev Mol Cell Biol 18(12):717–727

Varoquaux N, Ay F, Noble WS, Vert J-P (2014) A statistical approach for inferring the 3D structure of the genome. Bioinformatics 30(12):26–33

Wang H, Xu X, Nguyen CM, Liu Y, Gao Y, Lin X, Daley T, Kipniss NH, La Russa M, Qi LS (2018) CRISPR-mediated programmable 3D genome positioning and nuclear organization. Cell 175(5):1405–1417

Weinberger KQ, Sha F, Zhu Q, Saul LK (2007) Graph Laplacian regularization for large-scale semidefinite programming. In: Advances in neural information processing systems, pp 1489–1496

Ye T, Ma W (2020) ASHIC: hierarchical Bayesian modeling of diploid chromatin contacts and structures. Nucl Acids Res 48(21):123–123

Zhang Z, Li G, Toh K-C, Sung W-K (2013) Inference of spatial organizations of chromosomes using semi-definite embedding approach and Hi-C data. In: Annual international conference on research in computational molecular biology. Springer, pp 317–332

Zhou S, Xiu N, Qi H-D (2020) Robust Euclidean embedding via EDM optimization. Math Program Comput 12(3):337–387