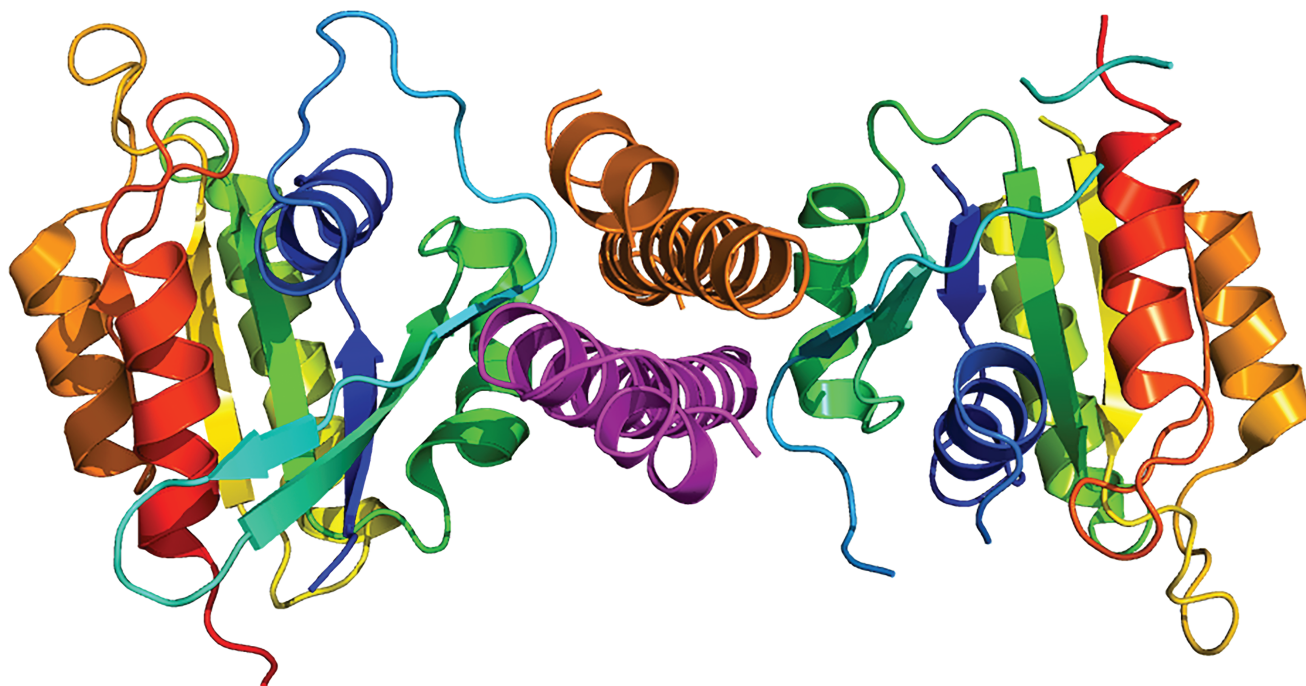


---

# Algebraic Geometry Tools in Systems Biology



*Alicia Dickenstein*

---

Many models in the sciences and engineering are expressed as solution sets to systems of polynomial equations, that is, as affine algebraic varieties. This is a basic notion in algebraic geometry, a vibrant area of mathematics which is particularly good at counting (solutions, tangencies, obstructions, etc.), giving structure to interesting sets (varieties with special properties, moduli spaces, etc.) and, principally, understanding structure. Starting in the 1980s with the development of computer algebra systems, and increasingly over the last years, ideas and methods from algebraic geometry are being applied to a great number of new areas (both in mathematics and in other disciplines including biology, computer science, physics, chemistry, etc.).

---

*Alicia Dickenstein is a professor of mathematics at the University of Buenos Aires and a researcher at CONICET, Argentina. Her email address is [alidick@dm.uba.ar](mailto:alidick@dm.uba.ar).*

*For permission to reprint this article, please contact:*  
[reprint-permission@ams.org](mailto:reprint-permission@ams.org).

DOI: <https://doi.org/10.1090/noti2188>

The aim of this note is to give a glimpse of how methods and concepts from algebraic geometry (in particular, from computational and real algebraic geometry) can be used to analyze standard models in molecular biology. These models occur in systems and synthetic biology, which focus on understanding the design principles of living systems. The past ten years have experienced an intense activity in the field and a rapidly growing literature. In turn, this application has challenged the current theory, mainly in the realm of real algebraic geometry.

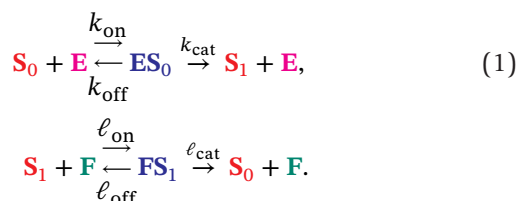
## Some Background on Chemical Reaction Networks

In chemical reaction network theory, and in particular in the study of (bio)chemical reaction networks in biochemistry, the input consists of a finite number of chemical species together with binding and unbinding reactions among them. The goal is to understand the evolution of the concentrations of these chemical species over time. This evolution is usually modeled by means of an autonomous system of ordinary differential equations

depending on kinetic parameters. In particular, a standard modelization is by means of what is called mass-action kinetics, that we briefly recall below. We refer the readers to the surveys [6, 7] and Chapter 5 in the book [5], as well as the book [9], for the history, basic definitions, examples, and important results, with an extensive bibliography. Together with Elisenda Feliu, we expect to complete an introductory book for students by 2021.

Before giving the general definitions, we introduce an important biological mechanism. It is a small network, but we will see that even in this basic case the number of variables and parameters involved makes it difficult to deal with.

**The phosphorylation cycle.** Below is the directed graph (or *digraph*) associated with the phosphorylation of a protein (considered the substrate and denoted by  $S_0$ ) by means of a kinase enzyme (another protein denoted by  $E$ ), which “helps” the substrate to acquire a phosphate group (not modeled) in a *binding site* and get transformed into a phosphorylated substrate (denoted by  $S_1$ ) via an intermediate species (denoted by  $ES_0$ ):



The second connected component of this digraph represents the emission of the phosphate group from  $S_1$ , which is then transformed back to  $S_0$  by means of a phosphatase enzyme (another protein denoted by  $F$ ), via another intermediate species (denoted by  $FS_1$ ).

In this scheme we have represented six *chemical species* ( $S_0, S_1, E, F, ES_0, FS_1$ ) and six so-called *complexes* formed with these species. Complexes are at the nodes of the digraph (1) corresponding to the interactions among species ( $S_0 + E, ES_0, S_1 + E, S_1 + F, FS_1, S_0 + F$ ). The six directed edges represent the chemical reactions and are labeled with positive constants termed *reaction rate constants*. Once we give an order to the species (for instance, as we enumerated them above), we can identify the complexes with nonnegative integer vectors in  $\mathbb{Z}^6$  (that is,  $e_1 + e_3 = (1, 0, 1, 0, 0, 0)$ ,  $e_5, e_2 + e_3, e_2 + e_4, e_6, e_1 + e_4$ , respectively). There are two monomolecular complexes consisting of each of the intermediate species  $ES_0, FS_1$  and four bimolecular complexes. Enzymes  $E, F$  speed up the transformation of the other proteins without being incorporated in the final products of the process. This is crucial in the transport of energy and the regulation of metabolism in the body.

As we mentioned, we are interested in understanding the evolution of the concentrations of the chemical

species. These concentrations are usually denoted with lowercase letters ( $s_0, s_1, e, f, es_0, fs_1$ , respectively) or just with  $x_1, \dots, x_6$ . However, it is good to keep the original notation when studying (bio)chemical reaction networks, so as to have quick visual information of the role played by each of the (bio)chemical species. Mass-action kinetics associated to this network with positive rate constants  $k_{\text{on}}, \dots, \ell_{\text{cat}}$  yields the following polynomial autonomous system of ordinary differential equations:

$$\begin{aligned} \frac{dx_1}{dt} &= -k_{\text{on}} x_1 x_3 + k_{\text{off}} x_5 + \ell_{\text{cat}} x_6, \\ \frac{dx_2}{dt} &= -\ell_{\text{on}} x_2 x_4 + k_{\text{cat}} x_5 + \ell_{\text{cat}} x_6, \\ \frac{dx_3}{dt} &= -k_{\text{on}} x_1 x_3 + (k_{\text{off}} + k_{\text{cat}}) x_5, \\ \frac{dx_4}{dt} &= -\ell_{\text{on}} x_2 x_4 + (\ell_{\text{off}} + \ell_{\text{cat}}) x_6, \\ \frac{dx_5}{dt} &= k_{\text{on}} x_1 x_3 - (k_{\text{off}} + k_{\text{cat}}) x_5, \\ \frac{dx_6}{dt} &= \ell_{\text{on}} x_2 x_4 - (\ell_{\text{off}} + \ell_{\text{cat}}) x_6. \end{aligned} \quad (2)$$

This is a system of the form  $\frac{dx}{dt} = f(x)$ , with  $f = (f_1, \dots, f_6)$  and any  $f_i \in \mathbb{R}[x_1, \dots, x_6]$  is a polynomial with real coefficients. It is straightforward to check that

$$f_1 + f_2 + f_5 + f_6 = 0, \quad f_3 + f_5 = 0, \quad f_4 + f_6 = 0.$$

As we will argue later, for any initial value  $x(0)$  there is a solution curve  $x = (x_1, \dots, x_6)$  defined for all  $t \geq 0$ . This implies that there exist constants  $T = (T_1, T_2, T_3)$  such that the following linear conservation laws are valid for any non-negative  $t$ :

$$x_1 + x_2 + x_5 + x_6 = T_1, \quad x_3 + x_5 = T_2, \quad x_4 + x_6 = T_3.$$

These three linear conservation constants (or *total amounts*) are usually denoted by  $S_{\text{tot}}, E_{\text{tot}}, F_{\text{tot}}$ , respectively, because they represent the total amounts of substrate, kinase, and phosphatase. Moreover, if  $x(0) \in \mathbb{R}_{>0}$ , we have that  $T \in \mathbb{R}_{>0}^3$ . Thus,  $\mathbb{R}^6$  is foliated by the parallel translates of the 3-dimensional linear subspace

$$S = \{x \in \mathbb{R}^6 : x_1 + x_2 + x_5 + x_6 = x_3 + x_5 = x_4 + x_6 = 0\},$$

which contain the trajectories.

We can easily describe in this case the (*algebraic variety* of) constant solutions  $V(f) = \{f_1(x) = \dots = f_6(x) = 0\}$ . The points of  $V(f)$  are known as the *steady states* of the system and can be described by the vanishing of the three following *binomials*:

$$\{x_6 - K_2 x_1 x_3, x_5 - K_1 x_1 x_3, x_2 x_4 - K_0 x_1 x_3\}, \quad (3)$$

where  $K_0, K_1, K_2$  are rational functions of the given six reaction rate constants  $\kappa = (k_{\text{on}}, k_{\text{off}}, \dots, \ell_{\text{cat}})$  which are defined over the whole positive orthant, i.e., for all  $\kappa \in \mathbb{R}_{>0}^6$ .

For instance,

$$K_1 = \frac{k_{\text{on}}}{(k_{\text{off}} + k_{\text{cat}})}$$

is the inverse of the so-called Michaelis–Menten constant.

Mass-action kinetics systems enjoy the following important property: both the positive orthant and the nonnegative orthant ( $\mathbb{R}_{>0}^6$  and  $\mathbb{R}_{\geq 0}^6$  in this case) are forward invariant for the dynamics. This means that if at time 0 all the concentrations are positive (resp., nonnegative), they will remain positive (resp., nonnegative) for any positive time. This is something desirable if we are to model the evolution of the concentrations of chemical species. The trajectory  $x = (x_1, \dots, x_6)$  starting at a point  $x(0) \in \mathbb{R}_{\geq 0}^6$  will thus be confined to the so-called *stoichiometric compatibility class*, or *S-class* for short:

$$P_{x(0)} = (S + x(0)) \cap \mathbb{R}_{\geq 0}^6.$$

A major object of study, as we will explain later, is the intersection  $V(f) \cap P_{x(0)}$  for a positive initial value, that is, the steady states in each *S-class* containing a positive point. To understand this intersection, one could parametrize the linear variety  $S + x(0)$  and replace this parametrization into the equations defining  $V(f)$ , but this will produce a loss in sparsity. Instead, in this case we can easily find a rational parametrization of  $V(f) \cap \mathbb{R}_{>0}^6$  from the binomial description (3):

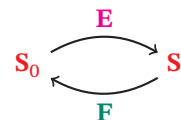
$$\begin{aligned} \mathbb{R}_{>0}^3 &\longrightarrow V(f) \cap \mathbb{R}_{>0}^6 x_4 \\ (x_1, x_3, x_4) &\longmapsto (x_1, K_0 \frac{x_1 x_3}{x_4}, x_3, x_4, K_1 x_1 x_3, K_2 x_1 x_3). \end{aligned} \quad (4)$$

So, any concentration vector at steady state can be rationally described in terms of the concentrations of the substrate  $S_0$  and the concentrations of the enzymes  $E$  and  $F$  at steady state. In this case, there is a single positive steady state in any *S-class* which intersects the positive orthant, for *any* choice of the nine positive parameters  $(\kappa, T)$ , but even in this simple case, this is not an easy computation to do without further tools. The intersection  $V(f) \cap \mathbb{R}_{>0}^6 \cap P_{x(0)}$  can then be described as follows:

$$\{x \in \mathbb{R}_{>0}^6 : x_1 x_4 + K_0 x_1 x_3 + (K_1 + K_2) x_1 x_3 x_4 - T_1 x_4 = x_3 + K_1 x_1 x_3 - T_2 = x_4 + K_2 x_1 x_3 - T_3 = 0\}.$$

Our point of view is that we have a 9-dimensional family of ODE systems (composed of a 6-dimensional family of steady state varieties and a 3-dimensional family of translates of  $S$ ) and we aim at developing tools to understand the different qualitative behaviors as we move the parameters, based on the *structure of the reaction network*. Of course, this study is also useful to understand quantitatively each particular system.

We will use the following usual representation of the phosphorylation cycle (1). It will be depicted as in Figure 1.



**Figure 1.** One-site phosphorylation cycle.

Note that the intermediate species and the reactions involving them as well as the six reaction rate constants are omitted.

**General definitions.** The input data of a *chemical reaction network* is a finite labeled digraph  $G = G(X, Y, R, \kappa)$  with  $m$  nodes and  $r$  directed edges, where  $X = \{X_1, \dots, X_n\}$  is a finite (numbered) set of species,  $Y = \{y_1, \dots, y_m\} \subset \mathbb{Z}_{\geq 0}^n$  is a finite (numbered) set of complexes on these species that label the vertices of  $G$ , and  $R$  is the set of edges  $(y_i, y_j)$  of  $G$  (also denoted by  $y_i \rightarrow y_j$ ), which represent the reactions between pairs of complexes. The entries of the positive vector  $\kappa = (\kappa_{ij} : y_i \rightarrow y_j \in R) \in \mathbb{R}_{>0}^r$  label the edges and are called reaction rate constants. Complexes are classically represented as nonnegative integer combinations of the species, as we had in (1), and we also will think of them as vectors in  $\mathbb{Z}_{\geq 0}^n$ .

Our variables  $x = (x_1, \dots, x_n)$  will be the (respective) concentrations of the species, a continuous approximation that makes sense, for example in the phosphorylation cycle where chemical species represent different molecules, when the number of molecules per unit volume is high. When this number is small, stochastic modeling is used instead of the deterministic mass-action ordinary differential equation modeling that we now define. Note that a complex  $y_i$  gives rise to a monomial in the concentrations of the chemical species  $x^{y_i} = x_1^{y_{i1}} x_2^{y_{i2}} \dots x_n^{y_{in}}$ .

**Definition 0.1.** *Mass-action kinetics* specified by a chemical network  $G = G(X, Y, R, \kappa)$  gives the following autonomous system of ordinary differential equations in the concentrations  $(x_1, x_2, \dots, x_n)$  of the species, as functions of time  $t$ :

$$\frac{dx}{dt} = \sum_{y_i \rightarrow y_j \in R} \kappa_{ij} x^{y_i} (y_j - y_i). \quad (5)$$

Then, for each coordinate  $i \in \{1, \dots, n\}$ ,

$$\frac{dx_i}{dt} = f_i(x),$$

where  $f_1, \dots, f_n$  are polynomials in  $\mathbb{R}[x_1, \dots, x_n]$  which carry combinatorial information from  $G$ .

A direct consequence of the form of the differential equations in (5) is what we observed in the case of the phosphorylation cycle. For any trajectory  $x : I \rightarrow \mathbb{R}^n$  defined in an interval  $I$ , the vector  $\frac{dx}{dt}$  lies for all  $t \in I$  in the *stoichiometric subspace*  $S$ . This is the linear subspace of  $\mathbb{R}^n$  generated by the differences  $y_j - y_i$ , which record the net production of each species in the corresponding reaction

$y_i \rightarrow y_j$ :

$$S = \langle y_j - y_i \mid y_i \rightarrow y_j \in R \rangle. \quad (6)$$

Assume  $S$  has dimension  $n - s$  and is defined by the vanishing of the linear forms  $\ell_1, \dots, \ell_s \in S^\perp$ . Thus, any trajectory  $x$  defined in  $I$  is contained in a parallel translate

$$S_T = \{x \in \mathbb{R}^n : \ell_1(x) = T_1, \dots, \ell_s(x) = T_s\}, \quad (7)$$

where  $T = (T_1, \dots, T_s)$  is a vector of total amounts. We have that  $\ell_i(f_1, \dots, f_n) = 0$  for any  $i = 1, \dots, s$ . In general, any linear dependency among  $f_1, \dots, f_n$  gives rise to a linear first integral, which restricts the dynamics of the system to linear varieties and reflects the biochemistry of the network.

Another consequence of the form of the differential equations in (5) is that both the positive orthant  $\mathbb{R}_{\geq 0}^n$  as well as its closure  $\mathbb{R}_{\geq 0}^n$  are forward invariant for the dynamics, for any mass-action system. Thus, any trajectory  $x : I \rightarrow \mathbb{R}^n$  starting at a nonnegative point is contained in an  $S$ -class, which is the intersection of a linear variety  $S_T$  with  $\mathbb{R}_{\geq 0}^n$ . When there exists a linear form  $\ell \in S^\perp$  with all  $n$  coefficients positive,  $S$ -classes are compact and the system is called *conservative*. Like in the case of the phosphorylation cycle, trajectories are defined for any positive time (in this case,  $\ell = (x_1 + x_2 + x_5 + x_6) + (x_3 + x_5) + (x_4 + x_6)$  is such a linear form). All the networks we will consider in this note are conservative, so to ease the notation, we will assume from now on that trajectories are defined for any  $t \geq 0$ . In this case, trajectories starting in the nonnegative orthant are confined to polyhedra, which are compact manifolds with boundary, contained in the boundary of the orthant and carrying combinatorial information from the network  $G$ .

**Definition 0.2.** The *steady state variety*  $V(f)$  of the kinetic system (5) equals the zero set of  $f_1, \dots, f_n$ . Any element of  $V(f)$  is called a steady state of the system. We denote by  $V(f)_{>0}$  (resp.,  $V(f)_{\geq 0}$ ) the positive (resp., the nonnegative) steady states.

Steady states are essential ingredients in the study of the dynamics of (5). Trajectories do not need to converge when  $t \rightarrow \infty$ , but if they do, the limit point is a steady state. It is well known that as polynomials are continuously differentiable functions, if all eigenvalues of the Jacobian  $J(f_1, \dots, f_n)$  of the system at a steady state  $x^* \in V(f)$  have negative real part, then  $x^*$  is locally asymptotically stable, that is, nearby trajectories converge to it. When  $S \neq \mathbb{R}^n$ , there are linear dependencies among the polynomials  $f_1, \dots, f_n$  and so at least  $s$  eigenvalues are equal to 0 (and there might be others). Therefore, we say that  $x^*$  is a *stable* steady state if the remaining  $n - s$  eigenvalues have negative real part. Hopf bifurcations are defined similarly, from the standard definition. Also, a steady state is said to be *nondegenerate* when the rank of  $J(f_1, \dots, f_n)$  equals  $s$ .

Another important dynamical property is the existence of at least two stable positive steady states in the same

$S$ -class, that is, with the same total amounts  $T$  for some  $T$ . In this case, the system is said to be *bistable*. Bistability is interpreted as the mathematical counterpart of biological decision making, even if the dynamical systems are deterministic. A necessary property for the existence of bistability is the occurrence of more than one positive steady state in the same  $S$ -class. When this is the case, the system is called *multistationary*.

Note that this is not reflected in the notation, but the polynomials  $f_1, \dots, f_n$  depend on the parameters  $\kappa$ . We can look at (5) as a family of polynomial autonomous ODE's. In fact, we will view the total amounts  $T$  in (7) also as parameters.

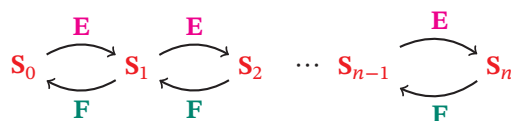
**Definition 0.3.** A reaction network  $G$  is said to have the *capacity for multistationarity* if there exists a choice of the  $r + s$  parameters  $(\kappa, T)$  for which the cardinality of the intersection of  $V(f)$  with the positive points in  $S_T$  (or with the positive points in the corresponding  $S$ -class  $S_T \cap \mathbb{R}_{\geq 0}^n$ ) is at least two. If  $x^*, x^{**}$  are distinct positive points in  $V(f) \cap S_T$  for some value of  $(\kappa, T)$ , we say that they are two *stoichiometrically compatible* positive steady states (denoted by *scpps*) and that  $(\kappa, T)$  are *multistationarity parameters*.

## Some Important Biochemical Reaction Networks

We quickly summarize some important signaling pathways that model reactions inside the cell.

**0.1.  $n$ -site phosphorylation cycle.** Phosphorylation of a substrate can occur in many binding sites and there are many mechanisms modeling a multiple phosphorylation-dephosphorylation cycle. Multisite phosphorylation plays important regulatory roles in the cell and it is related to multiple disorders. There is an important body of work on the mathematics of phosphorylation systems, which belong to the so-called post-translational modification systems, consisting of diverse configurational transformations of the proteins after they have been produced. Phosphorylation of a substrate typically activates it, which means that it can “transmit information” by participating in other reactions.

Similar to the case of a single phosphorylation site we considered before for  $n = 1$ , the *sequential distributive  $n$ -site cycle* is represented in Figure 2.



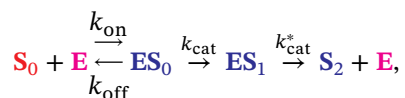
**Figure 2.**  $n$ -site phosphorylation cycle.

This reaction network consists of  $3n + 3$  variables,  $4n + 2$  nodes, and  $6n$  reactions (and so  $6n$  reaction rate constants). For any  $n$ ,  $\dim S^\perp = 3$ , so there are always three total

conservation constants  $S_{\text{tot}}, E_{\text{tot}}, F_{\text{tot}}$ . Thus, the total number of parameters is  $6n + 3$ . We will explain below why we can prove general results for these networks even as the number of variables and the number of parameters grow linearly with  $n$ .

The nice mathematical paper [20] initiated a series of further developments, with clever ad-hoc computations that involve an algebraic perspective. Many results are known on multisite sequential distributive phosphorylations, collecting work developed through several articles in our references and the references therein, as well as other very recent work that we could not include in the bibliography due to a limitation on the number of items. However, many questions are still open for the sequential distributive  $n$ -site cycle, even for  $n = 2$ . We detail some of these questions together with some known results (unluckily, without precise references due to size limitation in the bibliography):

1. When  $n \geq 2$ , the  $n$ -site system has the capacity for multistationarity.
2. When  $n = 2$ , the multistationarity region is described by means of a suite of recent techniques from real algebraic geometry to determine the signs of a multivariate polynomial over the positive orthant.
3. It is not known, even for  $n = 2$ , if the system can have oscillatory solutions (but this is not possible for many simplified models). Oscillations do exist in mixed mechanisms, where for instance the dephosphorylation is *processive* instead of *distributive*, which means that the first (or the second) component of the digraph is instead assumed to be of the form:



where  $ES_0, ES_1$  are intermediate species. Note that the species  $S_1$  of the singly phosphorylated phosphate do not occur in this component.

4. There are quite explicit open sets in parameter space where the number of scps is  $n$  for  $n$  odd and  $n + 1$  for  $n$  even.
5. Roughly “half” of these  $n$  or  $n + 1$  scps can be stable (and the other “half” unstable).
6. An upper bound for the number of scps is  $2n - 1$ , and it is conjectured that indeed this bound is attained, but this has only been demonstrated for  $n \leq 4$ .

**0.2. ERK pathways.** The Ras-Raf-MEK-ERK pathway is a cascade of phosphorylation of proteins in the cell that communicates a signal from a receptor on the surface of the cell to the DNA in the nucleus. In a cascade, the phosphorylated substrate in one layer acts as an enzyme in the next layer. ERK stands for extracellular signal-regulated kinase, which is activated and then controls downstream

responses, such as cell division. Figure 3 shows a partial reaction network of this pathway, where  $E = \text{Ras}$ , and  $S_0 = \text{Raf}$ ,  $P_0 = \text{MEK}$ ,  $R_0 = \text{ERK}$  are the unphosphorylated forms of the proteins, which can be phosphorylated in two sites each. Mutations in these proteins were studied in relation to the occurrence of cancer. The number of variables of

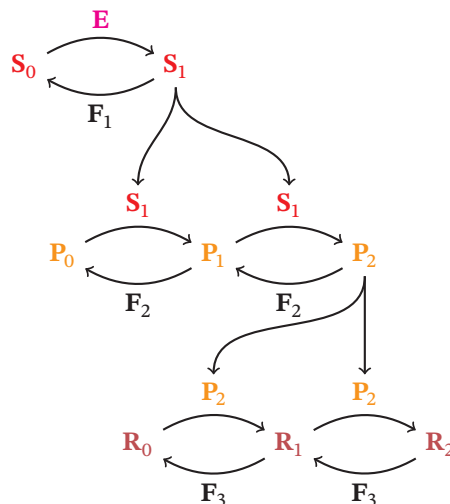
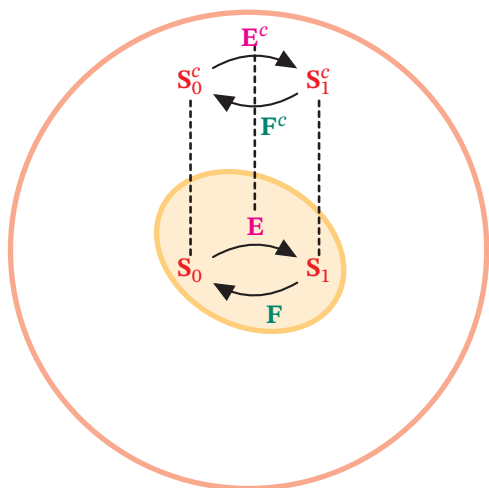


Figure 3. Three-layer cascade.

this network (including the concentrations of the 10 intermediate species not depicted) is equal to 20, 21, or 22, depending on how many of the phosphatases  $F_1, F_2, F_3$  are equal, and there are accordingly five, six, or seven total amounts. Moreover, the number of reaction rate constants is 30. These are really big numbers to deal with in a nonlinear system. It is known that the associated system has the capacity for multistationarity and there are oscillatory solutions. Algebraic-geometric methods can give further insight on this important signaling cascade. We refer the reader to the article [16] for a review of key biological questions. The emergence of multistationarity and oscillations in sub-networks of a different model of ERK regulation where only one layer of the cascade is considered but the two phosphorylation sites are distinguished has been studied in [15].

**Cellular compartments.** Figure 4 features a simplified network which is composed of two modules which are monostationary acting on different cellular compartments. Both modules in the nucleus and the cytoplasm of the cell are the phosphorylation cycles with one binding site that we discussed before. Moreover, it is assumed that there could be linear reactions between similar species, like  $S^c \leftrightarrow S$  (or  $S^c \rightarrow S$ ), but also between corresponding intermediate species (which are not depicted) through linear shuttling reactions. This “coupling” of biological systems, spatially organized, can produce emergent complex behavior, in particular, multistationarity. Several mathematical



**Figure 4.** A compartment model.

tools are used in [18] to analyze these models and identify minimal subnetworks which are bistable, but these tools also require an intensive parameter sampling and simulation.

There are many other important mechanisms in biochemistry usually modeled with mass-action kinetics, including other compartmental models, different types of cascades, models used in immunology introduced by McKeithan in a 1995 *PNAS* article and mathematically studied later by Sontag, or bifunctional enzyme models in bacteria introduced by Shinar and Feinberg in a 2010 *Science* paper. These last models show an interesting property called Absolute Concentration Robustness because the concentration of one of the regulatory proteins has a constant value at any positive steady state, independently of the  $S$ -class.

**Mass-action kinetics modeling in other frameworks.** The basic question of which autonomous polynomial dynamical systems  $dx/dt = f(x)$  in  $n$  variables  $(x_1, \dots, x_n)$  come from a chemical reaction network under mass-action kinetics has been answered by Hász and Tóth in 1979. The necessary and sufficient condition is that each monomial with negative coefficient in each  $f_i$  is divisible by  $x_i$ . This restriction is satisfied by the oscillatory Lotka-Volterra equations and other standard models in ecology, but not by the “chaotic” Lorenz equations  $\frac{dx_1}{dt} = \alpha x_2 - \alpha x_1$ ,  $\frac{dx_2}{dt} = \gamma x_1 - x_2 - x_1 x_3$ ,  $\frac{dz}{dt} = x_1 x_2 - \beta x_3$ , where  $\alpha, \beta, \gamma$  are positive, due to the existence of the negative term  $-x_1 x_3$  in  $f_2$ .

Many models of population dynamics in epidemiology, like the basic SIR model, can also be realized as arising from a reaction network with mass-action kinetics, as well as models for gene regulation.

When all the complexes are monomolecular, that is, consists of a single species, we have that  $m = n$  and the

resulting mass-action dynamics system (5) associated to the corresponding digraph  $G = (X, X, R, \kappa)$  is a linear system:

$$\frac{dx}{dt} = -\mathcal{L}(G) \cdot x, \quad (8)$$

where  $\mathcal{L}(G) \in \mathbb{R}^{n \times n}$  is the Laplacian matrix of  $G$ :

$$(-\mathcal{L}(G))_{ji} = \begin{cases} k_{ij}, & X_i \rightarrow X_j \in R, \\ 0, & X_i \rightarrow X_j \notin R, i \neq j, \\ -\sum_{X_i \rightarrow X_\ell \in R} k_{i\ell}, & i = j. \end{cases}$$

The description of the kernel of this matrix is a beautiful match of linear algebra with combinatorics, which goes back to Tutte, known as the Matrix-Tree Theorem [19]. When  $G$  is strongly connected, that is, when for each pair of nodes in  $G$  there exists a *directed* path joining them, the Laplacian matrix has rank  $n-1$  and there is a positive vector generating its kernel, where each entry is combinatorially expressed as a sum of certain monomials in the reaction rate constants  $\kappa$  with coefficient 1. For many nonlinear mass-action kinetics systems, taking advantage of this positivity with respect to certain embedded linear networks yields interesting explicit results. However, the global dynamic behavior of nonlinear mass-action kinetic systems is far from the linear case and mostly an open question.

## Some Results

### Some recent work on biochemical reaction networks.

We summarize now some recent results obtained using algebro-geometric techniques about special families of systems usual in biochemistry.

Together with Mercedes Pérez Millán, we recognized in many relevant biochemical networks, including all those we briefly presented in this note, what we called a *MESSI structure*. The main ingredient is the existence of a partition of the sets of species into a subset of *intermediate* species (which could be empty) and several nonempty subsets of *core* species, together with axioms that restrict the conformation of the complexes (which are at most bimolecular) and the kind of reactions that are allowed. This partition is in general natural, according to the function of the different species in the network, and features several underlying linearities (which are not independent, due to the interaction of the intermediate species).

Recognizing this structure allowed us to prove general results based on the combinatorics of digraphs associated to the original digraph. For instance, we give sufficient conditions to ensure the existence of a *rational parametrization of the steady state varieties* like the one in (4), that can be obtained algorithmically. As we mentioned, the fact that the coordinates of this rational function are well defined over the positive orthant ultimately relies on the Matrix-Tree Theorem. It is important to note that a “generic” algebraic variety does not admit a rational parametrization,

so this is a very special property that allows us to reduce the number of variables. For instance, in the case of the phosphorylation cycle we are led to study polynomials in three variables for any value of  $n$ . In the case of cascades, however, the dimension of the steady state variety  $V(f)$  increases with each layer. We can also give easily checkable combinatorial conditions on  $G$  and its associated digraphs that ensure *persistence* of the system, which means that any trajectory starting in the positive orthant must remain at a fixed positive distance (depending on the trajectory) from the coordinate hyperplanes in  $\mathbb{R}^n$ . All MESSI systems are conservative and it is possible to find linear forms in  $S^\perp$  with coefficients 0, 1 like those in (6), and give sufficient conditions for these forms to be a basis of linear relations among the polynomials  $f_1, \dots, f_n$ . Our work was built on several previous articles, in particular, the pioneering article [19] by Thomson and Gunawardena and the interesting article [10] by Feliu and Wiuf on the elimination of intermediate species, plus previous axiomatizations by Gnacadja.

We can also give sufficient conditions to ensure the existence of an explicit *monomial* parametrization, like in the case of the phosphorylation cycle. Thus,  $V(f)_{>0}$  is a translate of a toric variety for any value of  $\kappa$ , and we can algorithmically decide the capacity for multistationarity. This is based on results in [13] that take into account the relation of the oriented matroid associated with the exponents of the monomials in the parametrization and the oriented matroid associated with a basis of linear forms of  $S^\perp$ . Moreover, it is possible to construct (some) multistationarity parameters and exhibit pairs of stoichiometrically compatible positive steady states.

A different approach to identifying multistationarity parameters is presented in [4], based on Brouwer degree theory combined with the existence of parametrizations. It is possible to quite explicitly describe open regions in the space of parameters  $\kappa$  for which either  $(\kappa, T)$  is monostationary for any  $T$ , or there exist some (a priori undetermined) parameters  $T$  for which  $(\kappa, T)$  is multistationary. This is based on the sign of a *critical function*. The use of parametrizations to reduce the number of variables and structural reductions of the reaction network is also applied to determine the stability and the existence of Hopf bifurcations, like for instance in [15]. The beautiful Routh-Hurwitz theorem gives conditions on the signs of particular determinants built with the coefficients of a univariate polynomial, to deduce the fact that all its roots have negative real part, without computing (which is in general not possible) or approximating the roots. This is particularly useful for studying the characteristic polynomials of the Jacobians of the polynomials  $f_1, \dots, f_k$  in (5) depending on  $\kappa$  at a steady state (which is in general implicitly defined) and also to ensure the existence of Hopf bifurcations via

a similar implicit criterion due to Yang, which again does not need the computation of the roots. The question of inheritance of oscillation in chemical reaction networks with inflow and outflow reactions ( $0 \rightarrow X_i \rightarrow 0$  for any  $i = 1, \dots, n$ ) is studied in the nice paper [1].

To give a general framework to find open regions of multistationarity jointly in the space of all parameters  $(\kappa, T)$  polyhedral methods have been introduced in [3]. These are based on degeneration methods *à la Viro*. The theoretical results are adapted to make them amenable to effective computations in a variety of specific models of biochemical systems for which there exist explicit parametrizations of the corresponding steady state varieties. Once the intersections with the  $S$ -classes  $V(f) \cap \mathbb{R}_{>0}^n \cap P_{\kappa(0)}$  in the positive orthant are described as the zero set of  $s$  sparse polynomials in  $s$  variables, where  $s$  is the codimension of the  $S$ -classes (similarly to the case of a single phosphorylation), we record the configuration  $A$  of the exponents occurring in them and the matrix  $C$  of coefficients of these polynomials. The sought open multistationarity regions arise from simplices that lie in a single *regular (mixed) subdivision* of  $A$ , which are jointly *decorated* by  $C$ . This is applied to give open multistationarity regions for the sequential distributive  $n$ -site phosphorylation cycle for any  $n$ . We refer to [11] for a simple example and basic definitions, together with an explicit application to cascades with any number of layers. These degeneration tools are also applied to describe open regions where there are  $n$  (or  $n + 1$ ) scps for any  $n$  in [12], where we also provide a basic systematic implementation in a computer algebra system.

A different line of attack is presented for instance in [14], like in some other previous papers, where tools from numerical algebraic geometry are used to describe the “geography” of the space of parameters in particular enzymatic systems. Robustness of multistationarity is measured in terms of the size and the shape of the region where the property holds, with an extensive sampling of parameter points in an 8-dimensional space. Under realistic molecular assumptions, the size of the regions of interest is very small. We will return to these ideas shortly.

It would be very useful to detect and study underlying mathematical structures in other signaling pathways in systems biology.

**Some theoretical results on polynomial equations that emerged from this study.** The question of multistationarity in dynamical systems arising from chemical reaction networks is a question about the number of positive solutions of systems of sparse real polynomials. This is effectively computable for a given system since the work of Alfred Tarski (published in 1951) on quantifier elimination over the reals. There are many recent advances in the implementation of these algorithms, but the complexity is in general high and so the computations turn out to be

unfeasible for parametric systems, even for networks of relatively small size.

Nevertheless, the importance of the biochemical applications has led to theoretical advances on sharp lower and upper bounds for the number of positive solutions of systems of  $n$  real polynomial systems in  $n$  variables with a fixed monomial structure. One first upper bound is clearly the Bernshtein-Kushnirenko-Khovanskii bound for the number of complex solutions with nonzero coordinates for systems of Laurent polynomials in terms of the mixed volume of their Newton polytopes. In the univariate case, this bound is just the degree of the polynomial. The simple Descartes' rule of signs from 1637 gives a better upper bound, which is sharp for instance for polynomials having all roots real, and which implies that a real polynomial with  $k$  monomial terms cannot have more than  $k - 1$  positive roots, independently of the size of the exponents (that is, independently of its degree). A breakthrough result by Khovanskii proved that also in the multivariate case, the number of positive roots cannot exceed a bound that only depends on the number of monomials (and the number of variables), but not on the degrees. This bound is far from sharp. There are some sharper particular results, but the general question of finding a tight upper bound is out of reach for the moment.

No general multivariate Descartes' rule of signs is known (not even a conjectural one). We recognized a partial multivariate Descartes' result giving conditions on a sparse (generalized) polynomial system to have at most one positive solution, in terms of oriented matroids associated to the exponents and to the coefficients of the system. In fact, part of this result was hidden in a previous paper by Craciun, Garcia-Puente, and Sottile in their study of control points for toric patches in geometric modeling. Our work was built on results on chemical reaction networks done by chemical engineers. In fact, in many fields of science the analysis of parametrized systems by means of sign vectors has a long history, for instance in market models in economics or in the well-posedness of the input-output relation of an electronic circuit in electronics.

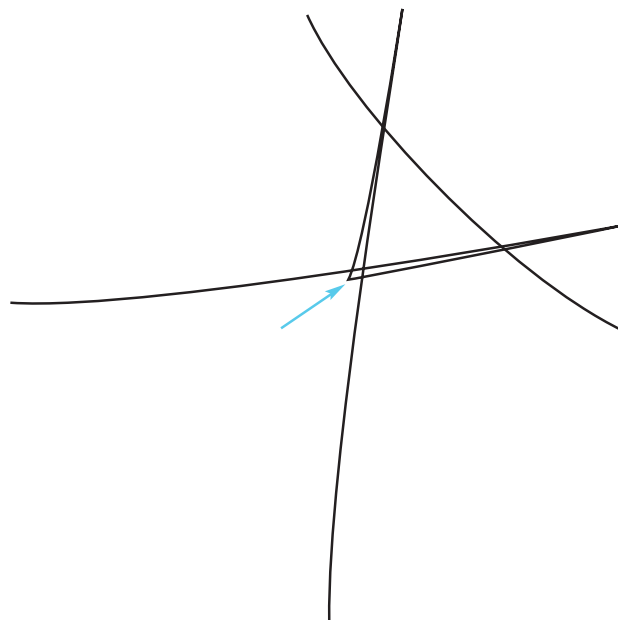
There are sufficient conditions for the existence of a positive solution when no more than one exists for any value of the parameters, based on degree theory. A very intricate result has been obtained in the realm of chemical reaction networks by Boros in 2019, building on work of Nachman and Deng, showing the existence of positive steady states for weakly reversible mass-action systems. It is mostly an open question to find good lower bounds for the number of positive solutions of general systems of sparse real polynomials. Known results include for instance the study of some specific systems by Sottile and Soprunova in 2006, and the use of polyhedral and combinatorial methods by Bihan, Santos, and Spaenlehauer in a beautiful paper from

2018. The obtained lower bounds, based on degenerations, are in general not sharp, as it is deduced from our paper [2] on upper bounds via a new Descartes' rule of signs for polynomial systems supported on circuits (with  $n + 2$  monomials in  $n$  variables).

Moreover, the question about the precise number of positive steady states for parametric systems with fixed monomial structure can in theory be "solved" via the computation of the finitely many chambers in the complement of the discriminant of the system (in fact, plus other resultant hypersurfaces corresponding to systems with a solution with a zero coordinate). We give now a concrete example extracted from [8] not related to biochemistry, but which is useful to visualize what we mentioned about the regions with many positive roots being very small. We considered the following system of two polynomials in two variables  $(x_1, x_2)$  depending on two parameters  $(a, b)$ :

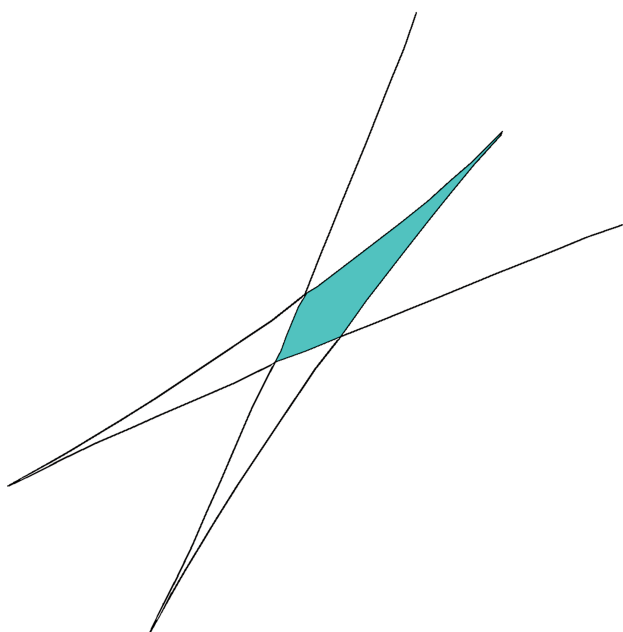
$$\begin{aligned} f_1(a, b, x, y) &= x^6 + ay^3 - y = 0, \\ f_2(a, b, x, y) &= y^6 + bx^3 - x = 0. \end{aligned} \quad (9)$$

The associated discriminant  $D \in \mathbb{Z}[a, b]$  satisfies by definition the following property:  $D(a, b) = 0$  whenever (9) has a degenerate solution. As we move the parameters, the only way of changing the number of *real* roots of the system is when two complex conjugate roots merge, that is, when we traverse the discriminant locus  $\{D = 0\}$ . We can compute  $D$  via elimination of variables, for instance via a Gröbner basis computation. It is a polynomial of degree 90 with 58 monomials and huge integer coefficients! But  $\{D = 0\}$  can be rationally parametrized and Figure 5 shows a picture on a logarithmic scale of this curve in the  $(a, b)$  plane.



**Figure 5.** There is a "hidden" chamber near the arrow.

It is easy to see that for any choice of  $(a, b)$ , the only solution to system (9) on the coordinate axes is the origin, and this is a nondegenerate common root. Then, the signs of the real solutions must also be constant in each chamber in the complement of  $\{D = 0\}$ . It is then enough to pick one point in each of the finitely many chambers and approximate its solutions. Taking parameters in each of the visible chambers, we get either one or three positive roots. But after amplifying the picture close to the arrow in Figure 5 around 1700 times, a new chamber is visible in Figure 6. One point that lies inside is  $(\frac{44}{31}, \frac{44}{31})$  and we get that the associated system (9) has five positive solutions! The volume of this small region with the maximal number five of solutions (an upper bound known a priori by previous work by Li, Rojas, and Wang) is smaller than  $5.701 \cdot 10^{-7}$ .



**Figure 6.** The colored chamber is visible after amplifying around 1700 times.

Both theoretical results and computational methods (symbolic as well as numeric) in (real) algebraic geometry have to be streamlined to be able to get a hand on all the “discriminant complement” chambers associated to families of real polynomials arising in biological models, to predict the number of scps. It is also a hard task to predict from the structure of the network the possible number of scps which are stable. More generally, the ultimate goal is to partition the parameter space into chambers containing parameters whose associated ODE systems have the same qualitative dynamic behavior.

## A Small Summary

We gave a quick overview of recent results and some challenges in the algebro-geometric study of biochemical

reaction networks modeling signaling pathways in systems biology. These networks give rise to quite diverse, but not general, mass-action dynamical systems. Understanding their structure allows us to prove theoretical results for classes of models and to do computations for systems with a large number of variables and parameters. These results on the qualitative analysis of families of models can be used to guide experimental design and in synthetic biology.

Traditionally, these systems are analyzed by means of numerical simulations combined with parameter inference or parameter sampling. However, many parameters are difficult (or even provably impossible) to estimate, and parameter values might vary depending on the environment and individual. Accordingly, algebraic geometry and computational algebra tools are useful for understanding the mathematical properties of networks and families of steady state varieties, providing a global and qualitative approach to study the systems for all parameter values.

There are mathematical tools to address many of the open questions, but for networks of interest the responses tend to be too complex to be understood or computed. This application has thus challenged current methods, mainly in the domain of real algebraic geometry, and has given rise to purely theoretical results on polynomial equations. New advances will require the combination of tools from different areas in mathematics besides algebraic geometry (and its computational aspects), including dynamical systems, discrete mathematics, data science, etc. and of course, also from biochemistry.

## References

- [1] Murad Banaji, *Inheritance of oscillation in chemical reaction networks*, Appl. Math. Comput. **325** (2018), 191–209, DOI 10.1016/j.amc.2017.12.012. MR3759136
- [2] Frédéric Bihan and Alicia Dickenstein, *Descartes’ rule of signs for polynomial systems supported on circuits*, Int. Math. Res. Not. IMRN **22** (2017), 6867–6893, DOI 10.1093/imrn/rnw199. MR3737324
- [3] Frédéric Bihan, Alicia Dickenstein, and Magalí Giaroli, *Lower bounds for positive roots and regions of multistationarity in chemical reaction networks*, J. Algebra **542** (2020), 367–411, DOI 10.1016/j.jalgebra.2019.10.002. MR4031126
- [4] Carsten Conradi, Elisenda Feliu, Maya Mincheva, and Carstein Wiuf, *Identifying parameter regions for multistationarity*, PLoS Comput. Biol. **13**(10) (2017), e1005751.
- [5] David A. Cox, *Applications of polynomial systems*, with contributions by Carlos D’Andrea, Alicia Dickenstein, Jonathan Hauenstein, Hal Schenck, and Jessica Sidman, CBMS Regional Conference Series in Mathematics, volume 134, Amer. Math. Soc., Providence, RI, 2020.
- [6] Alicia Dickenstein, *Biochemical reaction networks: an invitation for algebraic geometers*, Mathematical Congress of the Americas, Contemp. Math., vol. 656, Amer. Math. Soc., Providence, RI, 2016, pp. 65–83, DOI 10.1090/conm/656/13076. MR3457596

- [7] Alicia Dickenstein, *Algebra and geometry in the study of enzymatic cascades*, World Women in Mathematics 2018, Proceedings of the First World Meeting for Women in Mathematics (WM)<sup>2</sup>, Association for Women in Mathematics Series, volume 20, Springer, Cham, 2019, pp. 57–81.
- [8] Alicia Dickenstein, J. Maurice Rojas, Korben Rusek, and Justin Shih, *Extremal real algebraic geometry and  $\mathcal{A}$ -discriminants* (English, with English and Russian summaries), *Mosc. Math. J.* **7** (2007), no. 3, 425–452, 574, DOI 10.17323/1609-4514-2007-7-3-425-452. MR2343141
- [9] Martin Feinberg, *Foundations of chemical reaction network theory*, Applied Mathematical Sciences, vol. 202, Springer, Cham, 2019. MR3890056
- [10] Elisenda Feliu and Carsten Wiuf, *Simplifying biochemical models with intermediate species*, *J. Royal Soc. Interface* (2013), 10:20130484.
- [11] Magalí Giaroli, Frédéric Bihan, and Alicia Dickenstein, *Regions of multistationarity in cascades of Goldbeter-Koshland loops*, *J. Math. Biol.* **78** (2019), no. 4, 1115–1145, DOI 10.1007/s00285-018-1304-0. MR3922967
- [12] Magalí Giaroli, Rick Rischter, Mercedes P. Millán, and Alicia Dickenstein, *Parameter regions that give rise to  $2\lfloor \frac{n}{2} \rfloor + 1$  positive steady states in the  $n$ -site phosphorylation system*, *Math. Biosci. Eng.* **16** (2019), no. 6, 7589–7615, DOI 10.3934/mbe.2019381. MR4034263
- [13] Stefan Müller, Elisenda Feliu, Georg Regensburger, Carsten Conradi, Anne Shiu, and Alicia Dickenstein, *Sign conditions for injectivity of generalized polynomial maps with applications to chemical reaction networks and real algebraic geometry*, *Found. Comput. Math.* **16** (2016), no. 1, 69–97, DOI 10.1007/s10208-014-9239-3. MR3451424
- [14] Kee-Myoung Nam, Benjamin M. Gyori, Silviana V. Amethyst, Daniel J. Bates, and Jeremy Gunawardena, *Robustness and parameter geography in post-translational modification systems*. <https://doi.org/10.1101/862003>.
- [15] Nida Obatake, Anne Shiu, Xiaoxian Tang, and Angélica Torres, *Oscillations and bistability in a model of ERK regulation*, *J. Math. Biol.* **79** (2019), no. 4, 1515–1549, DOI 10.1007/s00285-019-01402-y. MR4019931
- [16] Aleena L. Patel and Stanislav Y. Shvartsman, *Outstanding questions in developmental ERK signaling*, *Development* **145** (2018), no. 14, dev143818, DOI 10.1242/dev.143818.
- [17] Mercedes Pérez Millán and Alicia Dickenstein, *The structure of MESSI biological systems*, *SIAM J. Appl. Dyn. Syst.* **17** (2018), no. 2, 1650–1682, DOI 10.1137/17M1113722. MR3811780
- [18] Debora Tenenbaum, Juan Ignacio Marrone, Hernán E. Grecco, and Alejandra C. Ventura, *Robustness in spatially driven bistability in signaling systems*, *Sci. Rep.* **10** (2020), 5591.
- [19] Matthew Thomson and Jeremy Gunawardena, *The rational parameterisation theorem for multisite post-translational modification systems*, *J. Theoret. Biol.* **261** (2009), no. 4, 626–636, DOI 10.1016/j.jtbi.2009.09.003. MR2974153

- [20] Liming Wang and Eduardo D. Sontag, *On the number of steady states in a multiple futile cycle*, *J. Math. Biol.* **57** (2008), no. 1, 29–52, DOI 10.1007/s00285-007-0145-z. MR2393207



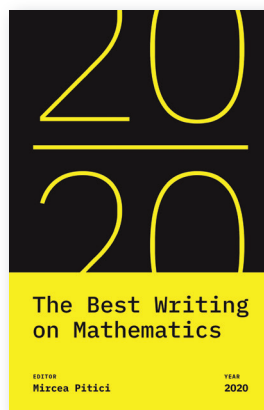
Alicia Dickenstein

### Credits

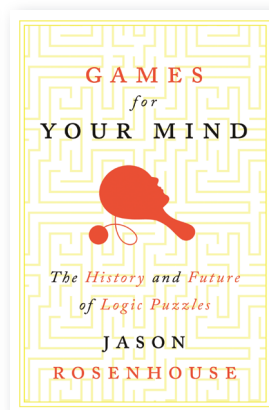
Opening photo is courtesy of Wikimedia.

Figures 1–6 are courtesy of the author.

Author photo is courtesy of Raúl Gómez.



The year's finest mathematical writing from around the world



A lively and engaging look at logic puzzles and their role in recreation, mathematics, and philosophy

Available in print, ebook, and audio



PRINCETON UNIVERSITY PRESS