INTERFACE

rsif.royalsocietypublishing.org

Research

Cite this article: Gross E, Davis B, Ho KL, Bates DJ, Harrington HA. 2016 Numerical algebraic geometry for model selection and its application to the life sciences. *J. R. Soc. Interface* **13**: 20160256.
http://dx.doi.org/10.1098/rsif.2016.0256

Received: 31 March 2016
Accepted: 19 September 2016

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biomathematics, systems biology, computational biology

Keywords:
model validation, polynomial optimization, maximum-likelihood, chemical reaction networks, parameter estimation

Author for correspondence:
Heather A. Harrington
e-mail: harrington@maths.ox.ac.uk

Electronic supplementary material is available online at http://dx.doi.org/10.6084/m9.figshare.c.3491538

THE ROYAL SOCIETY PUBLISHING

Numerical algebraic geometry for model selection and its application to the life sciences

Elizabeth Gross[1], Brent Davis[2], Kenneth L. Ho[3], Daniel J. Bates[2] and Heather A. Harrington[4]

[1]Department of Mathematics, San José State University, San José, CA 95112, USA
[2]Department of Mathematics, Colorado State University, Fort Collins, CO 80523, USA
[3]Department of Mathematics, Stanford University, Stanford, CA 94305, USA
[4]Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

HAH, 0000-0002-1705-7869

Researchers working with mathematical models are often confronted by the related problems of parameter estimation, model validation and model selection. These are all optimization problems, well known to be challenging due to nonlinearity, non-convexity and multiple local optima. Furthermore, the challenges are compounded when only partial data are available. Here, we consider polynomial models (e.g. mass-action chemical reaction networks at steady state) and describe a framework for their analysis based on optimization using numerical algebraic geometry. Specifically, we use probability-one polynomial homotopy continuation methods to compute all critical points of the objective function, then filter to recover the global optima. Our approach exploits the geometrical structures relating models and data, and we demonstrate its utility on examples from cell signalling, synthetic biology and epidemiology.

1. Introduction

Across the physical, biological and social sciences, mathematical models are formulated and studied to better understand real-world phenomena. Often, multiple models are developed to explore alternative hypotheses. It then becomes necessary to choose between different models, for example, based on their fit with noisy experimental data. This is the problem of model selection, a fundamental scientific problem with practical implications [1–3].

When dealing with deterministic models in the life sciences, the standard approach to model selection is to first estimate all model parameters and hidden variables from the data, then select the model with the smallest best-fit error, up to some penalty on model complexity [4,5]. Thus, at its core, model selection is intimately tied to parameter estimation. For example, if we have a model described by a system of equations $f(a, x) = 0$ in the parameters a and variables x with measurable 'outputs' $z = g(x)$, then parameter estimation can be written as the following least-squares optimization problem:

$$\min_{a,x,z} \|z - y\|^2 \quad \text{s.t.} \quad \begin{cases} f(a, x) = 0 \\ z = g(x), \end{cases} \tag{1.1}$$

where y denotes the observed data, i.e. measured outputs. Unless f and g are convex, solving (1.1) is a non-convex problem, which can be challenging as standard local solvers run the risk of getting trapped in local minima (especially in high dimensions). This can be mitigated somewhat with techniques such as simulated annealing [6,7] or convex relaxation that has been successful for model invalidation [8–10], but there is generally no guarantee that a global minimum will be found.

© 2016 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

When f and g are polynomial, however, problem (1.1) can be solved globally by finding all roots of an associated polynomial system. In this case, ideas from computational algebra and algebraic geometry can be effective; see, e.g. [11–14] for applications of Gröbner bases in systems biology and [15] for applications of algebraic geometry to statistical inference. Such symbolic methods tend to be computationally expensive, which limits their use in practice and are bypassed here. Thus, although they provide a solution in principle, new algorithms and techniques are yet desired.

In this paper, we aim to fill this gap by proposing a framework for global parameter estimation for polynomial deterministic models using numerical algebraic geometry (NAG), a suite of tools for numerically approximating the solution sets of multivariate polynomial systems via adaptive multi-precision, probability-one polynomial homotopy continuation [16,17]. This is a deterministic method, so it will produce the same results (up to numerical error) every time. Unlike other approaches, there is no sense of 'simulations' or sampling required for this method. Our approach scales well in dimension relative to classical symbolic methods [18], and, while it comes with a higher computational cost than standard local optimization, it has a probability-one guarantee to recover the global optima, solving problem (1.1) in the strong sense. This allows us to reason rigorously about model fit and to address the related problems of model selection and parameter estimation from a maximum-likelihood perspective. We demonstrate our techniques on examples from biology, where polynomial models often arise as the steady-state descriptions of mass-action chemical reaction networks. Although some limitations remain, we believe that this work achieves its primary purpose of introducing NAG as a valuable complement to existing tools for model evaluation and analysis. Additionally, this paper highlights specific challenges that arise when using polynomial methods for model inference, such as dealing with positivity constraints and non-isolated solutions, and provides guidance for tackling these challenges.

The remainder of the paper is organized as follows. In §2, we state precisely the problems with which we are concerned: model validation, model selection, and parameter and hidden variable estimation. We then present the NAG algorithms for solving each problem. Finally, we showcase our approach on a few examples, including cell death activation, synthetic biocircuits, human immunodeficiency virus (HIV) progression and protein modification.

2. Problem statement

Consider a model whose dynamics are described by the system of polynomial differential equations

$$\dot{x} = f(a, x), \tag{2.1}$$

where $a = (a_1, \ldots, a_k)$ are parameters (e.g. rate constants in a deterministic mechanistic model, such as a chemical reaction network with mass-action kinetics), $x = (x_1, \ldots, x_n)$ are variables, and $f = (f_1, \ldots, f_r)$ are polynomials in x and a with measurable outputs $z = g(x)$, where $z = (z_1, \ldots, z_m)$, $m \leq n$ and $g = (g_1, \ldots, g_m)$ are polynomials in x. While the parameters $a_1, \ldots, a_k$ are treated as fixed variables in our exposition, we separate them from $x_1, \ldots, x_n$ to respect how such variables are treated differently in experimental and computational settings.

In algebraic geometry, a *variety* is a solution set of a system of polynomial equations; we use this terminology for our next two definitions. The *real model variety* is the solution set of the system

$$f(a, x) = 0, \tag{2.2}$$
$$z - g(x) = 0, \tag{2.3}$$

that is,

$$(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} := \{(a, x, z) \in \mathbb{R}^{k+n+m} : f(a, x) = 0, z - g(x) = 0\},$$

corresponding to the steady states of the model. Now, consider, for simplicity, the case of a single data point $y = (y_1, \ldots, y_m)$. (See the electronic supplementary material for multiple data points.) The *real data variety* is then the affine linear space

$$(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} := \{(a, x, z) \in \mathbb{R}^{k+n+m} : z_i = y_i, \forall i = 1, \ldots, m\},$$

with $\dim(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} = k + n$. We consider the case when the data include some extrinsic (measurement) noise; we assume the errors $\{\epsilon_1, \ldots, \epsilon_m\}$ on the observed data variables are uncorrelated random variables and each error ϵ_i is normally distributed with known variance σ_i (which can be obtained by instrument calibration).

Using this geometric framework, the problems of (1) model validation, (2) model selection and (3) parameter estimation can be described precisely in terms of the real algebraic varieties $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$.

2.1. Problem 1: model validation

For model validation, we want to determine whether a deterministic polynomial model $\mathcal{M}$ is compatible with the data according to a given significance level α. Using the noise assumptions from above, each deterministic model $\mathcal{M}$ gives rise to an associated statistical model. Specifically, given a deterministic system $\dot{x} = f(x, a)$ with an observation y made at steady state, the statistical model under consideration is

$$Y_i = z_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i), \quad 1 \leq i \leq m, \tag{2.4}$$
$$f(x, a) = 0 \tag{2.5}$$

and

$$z - g(x) = 0, \tag{2.6}$$

where x, a, z are all unknown, and σ_i is known for all i.

The compatibility question is akin to significance testing and asking whether the model is a 'good fit' for the data. A natural goodness-of-fit statistic is

$$d^2 := \min \sum_{i=1}^{m} \frac{(z_i - Y_i)^2}{\sigma_i^2} \text{ subject to } (a, x, z) \in (\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}. \tag{2.7}$$

When the variances differ, d^2 is the minimum-squared weighted Euclidean distance between $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$. When all variances are equal to one, the value of (2.7) is just the minimum-squared distance. For the remainder of the text, we assume the latter, knowing that we can simply rescale.

Optimization problem (2.7) can be derived directly from the log-likelihood function

$$\log L(a, x, z|y) = \sum_{i=1}^{m} \left(\frac{1}{2} \log 2\pi\sigma_i^2 - \frac{(z_i - y_i)^2}{2\sigma_i^2} \right) \tag{2.8}$$

as demonstrated in §2.1 of the electronic supplementary material. Indeed, minimizing the argument of (2.7) over $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ is equivalent to maximizing $\log L(a, x, z|y)$ over $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$. This provides a bridge to standard statistical model selection tools such as AIC [19] and BIC [20]. Note that a and x do not appear in (2.8), instead, they appear in the constraints of the problem.

In model validation, our null hypothesis is that the data y are generated from the statistical version of the deterministic model $\mathcal{M}$. If the data are generated from the model $\mathcal{M}$ with normally distributed extrinsic noise, then the distribution function of d^2 is dominated by that of the chi-squared distribution with m degrees of freedom, χ^2_m, where m is the number of measurable outputs. Thus, if p_α is the upper α-percentile for χ^2_m, then $\Pr(d^2 \geq p_\alpha) \leq \Pr(U \geq p_\alpha) = \alpha$, where $U \sim \chi^2_m$. We reject the model $\mathcal{M}$ as *incompatible*, i.e. we reject the null hypothesis, if the observed value of d^2 is greater than p_α; otherwise, we say that the model $\mathcal{M}$ is *compatible*.

If the real model and data varieties intersect, that is, $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R} \neq \emptyset$, then $d^2 = 0$, and we also say that the model is compatible with the data. If there are restrictions on (a,x,y), for example, if all parameters and variables are required to be non-negative, then finding d^2 becomes a *constrained optimization* problem (see electronic supplementary material).

2.2. Problem 2: model selection

For model selection, we are given a set of models $\{\mathcal{M}_1, \ldots, \mathcal{M}_s\}$ and want to determine the model of best fit. In this setting, we use the statistic d^2 to make a selection, either by choosing the model with the smallest value of d^2 or by using d^2 in conjunction with a complexity penalty, similar to the Bayesian or Akaike information criteria [1].

If the statistic d^2 evaluates to zero for all (or even multiple) models, then we are unable to make a selection between the models. This can be remedied by designing experiments that yield more informative data. For example, measuring more variables can reduce the dimension of $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R}$; the most informative situation is when $(\mathcal{V}_{\mathcal{M}_i})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R} = \emptyset$ for all models. This indeterminacy can also be resolved by taking multiple measurements and minimizing the joint squared distance (see the electronic supplementary material).

2.3. Problem 3: parameter estimation

Parameter estimation can be achieved by finding the point $(\hat{a}, \hat{x}, \hat{z}) \in (\mathcal{V}_\mathcal{M})_\mathbb{R}$ that minimizes the value of (2.7). The point $(\hat{a}, \hat{x}, \hat{z})$ is the maximum-likelihood estimate under the given noise assumptions. The parameter estimate is then $\hat{a}$; the estimate of the hidden variables, $\hat{x}$ and the estimate of the 'de-noised' outputs, $\hat{z}$. Of course, if the data and model varieties intersect, then there will be one or more (possibly infinite) choices for $(\hat{a}, \hat{x}, \hat{z})$. Otherwise, it is a matter of solving a polynomial system that yields the point(s) on $(\mathcal{V}_\mathcal{M})_\mathbb{R}$ nearest $(\mathcal{V}_\mathcal{D})_\mathbb{R}$. This is described in more detail in §3.

3. Geometry

In each of the problems stated above, we seek to minimize the distance between $(\mathcal{V}_\mathcal{M})_\mathbb{R}$ and $(\mathcal{V}_\mathcal{D})_\mathbb{R}$ or to find the intersection of these two sets. Standard methods for solving nonlinear optimization problems are local in nature, i.e. only guaranteed to converge to a local minimum which may or may not be the global minimum.

However, using NAG, we find *all* local extrema over $\mathbb{C}$, necessarily including the global minimum. Owing to this global benefit, NAG has been used before in statistical inference in the field of algebraic statistics [21,22].

Let $\mathcal{V}_\mathcal{M} \subseteq \mathbb{C}^{k+n+m}$ be the (complex) Zariski closure of $(\mathcal{V}_\mathcal{M})_\mathbb{R}$ and $\mathcal{V}_\mathcal{D} \subseteq \mathbb{C}^{k+n+m}$ be the (complex) Zariski closure of $(\mathcal{V}_\mathcal{D})_\mathbb{R}$. We refer to $\mathcal{V}_\mathcal{M}$ and $\mathcal{V}_\mathcal{D}$ as the *model variety* and *data variety*, respectively.

The problem of determining the intersection of $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}$ is simply a matter of solving the polynomial system obtained by taking the union of the polynomials defining $\mathcal{V}_\mathcal{M}$ and the polynomials defining $\mathcal{V}_\mathcal{D}$. This is handled directly by NAG. If the intersection is nonempty and positive-dimensional (complex curves, surfaces, etc.), then real points can be found using the polynomial homotopy method described in [23], a method based on symbolic algorithms in [24] and, more classically, on the decision method in [25].

In the case that $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}$ is empty, the problem of finding the points on the varieties nearest one another can also be stated in terms of a polynomial system, on which we then call NAG solvers to find solutions. A well-known necessary condition for local extrema is given by the Fritz John conditions, related to Lagrange multipliers. In the main text, we assume that $r + m = \operatorname{codim} \mathcal{V}_\mathcal{M}$; however, when this is not the case, the number of equations can be reduced (see electronic supplementary material).

Proposition 3.1. (Equations given by Fritz John conditions). *Let $r + m = \operatorname{codim} \mathcal{V}_\mathcal{M}$. Let $f(a,x) = 0$, $z - g(x) = 0$ be defined on a Zariski open set of $\mathcal{V}_\mathcal{M}$, and define $h(a,x,z) = \{f(a,x), z - g(x)\}$ for simplicity of notation below. If $(a,x,z) \in (\mathcal{V}_\mathcal{M})_\mathbb{R}$ is a local minimum of*

$$\sum_{i=1}^{m} (z_i - y_i)^2, \tag{3.1}$$

then there exists $\lambda := (\lambda_0, \lambda_1, \ldots, \lambda_{r+m}) \in \mathbb{P}^{r+m}$, such that (a,x,z,λ) is a solution to the system

$$f(a,x) = 0, \tag{3.2}$$

$$z - g(x) = 0 \tag{3.3}$$

and

$$\lambda_0 \begin{bmatrix} 0 \\ z - y \end{bmatrix} + \sum_{i=1}^{r+m} \lambda_i \nabla_{a,x,z} h_i(a,x,z) = 0, \tag{3.4}$$

where $\mathbb{P}^{r+m}$ refers to projective space, a slight generalization of complex affine space, and $\nabla_{a,x,z}$ refers to the vector consisting of all first-order derivatives with respect to all a, x, and z.

Solving this system with NAG will provide us with all local extrema of (3.1) over $\mathcal{V}_\mathcal{M}$, from which we may easily select the pair of nearest points. In fact, in the examples below, we use minor variations on this theme for computational efficiency; see the corresponding sections of the electronic supplementary material for details.

The geometry of zero sets of systems such as (3.2)–(3.4) is described in [26]. In particular, Draisma *et al.* [26] explore the *Euclidean distance degree* (ED degree) of a variety, which is the number of critical points of the squared distance to a general point outside the variety. The number of complex solutions to (3.2)–(3.4) when y is generic is the ED degree of a projection of the model variety, along the data subspace. The ED degree captures the complexity of any method that solves a squared distance minimization problem *exactly* and plays a key role in the methods discussed here. For example, the ED degree is the number of paths that need to be tracked when using the parameter homotopies described in §3.1.

Table 1. Algorithm 1: model validation.

input	model $\mathcal{M}$, data $\mathcal{D} = \{y\}$, tolerance α
output	*yes* or *no*
1	If $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D} = \emptyset$, go to step 3.
2	If $\dim(\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}) \geq 0$ and $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R} \neq \emptyset$, return *yes*,
	else, go to step 3.
3	Find a pair $((\hat{a},\hat{x},\hat{z}),(\hat{a},\hat{x},y)) \in (\mathcal{V}_\mathcal{M})_\mathbb{R} \times (\mathcal{V}_\mathcal{D})_\mathbb{R}$ that minimizes (3.1) (using NAG software such as Bertini or PHCpack).
4	If $\|\|\hat{z} - y\|\|^2 < p_\alpha$, return *yes*; else *no*.

3.1. Numerical algebraic geometry

Given a polynomial system F consisting of r polynomials and N variables, NAG packages, such as `Bertini` [16], `PHCpack` [27], `HOM4PS-2.0` [28], use polynomial homotopy continuation to provide *probability-one* numerical methods for finding approximations of all isolated complex solutions of $F = 0$ (points) as well as *witness points* on all positive-dimensional irreducible components of the solution set of $F = 0$. These methods are probability-one in that the computations include some randomization, and this randomization will yield theoretically correct results so long as the random numbers are not chosen from some measure zero set in the parameter spaces of potential choices [17,29].

If $x \in \mathbb{R}^N$ is a *real* solution of $F = 0$, it is either isolated among the complex solutions or it lies on a positive-dimensional complex irreducible component. In the former case, the methods of NAG will find x and recognize it as real. In the latter case, x can be difficult to uncover.

However, for our purposes, we usually only need to verify the existence of a real solution. In this case, we can find witness points on all positive-dimensional components and then use the procedure in §2.1 of [23] to verify the existence of real points.

Finally, there is a setting in which a particular method from NAG is especially powerful. If a parametrized polynomial system needs to be solved multiple times for varying parameter values, the *parameter homotopy* [17,29–31] can greatly reduce the computation time. Refer to the electronic supplementary material for more details, particularly §§4.1 and 4.4.

3.2. Algorithms

We present three related algorithms to address model validation, model selection and parameter estimation.

The aim of the first algorithm, algorithm 1, is to find the pair of points that minimize the distance between $(\mathcal{V}_\mathcal{M})_\mathbb{R}$ and $(\mathcal{V}_\mathcal{D})_\mathbb{R}$. If $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R} = \emptyset$, then this is obtained by solving (3.2)–(3.4); otherwise, additional techniques are required.

We also note simply that these techniques are indeed probability-one algorithms. The computations will necessarily be carried out in finite time and the steps proceed linearly (no loops), so the methods will necessarily finish.

3.2.1 Algorithm 1: model validation

The computation of the intersection $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}$ in steps 1 and 2 (table 1) can be determined in several ways. The simplest way is by considering dimensions: if $\dim(\mathcal{V}_\mathcal{M}) + \dim(\mathcal{V}_\mathcal{D})$ exceeds the ambient dimension, then they will almost surely intersect. If the ambient dimension is larger than the sum of the variety dimensions, then they typically do not intersect. To compute the intersection (or check to see if it is nonempty), one could substitute $y - g(x)$ for (3.3) in the system (3.2)–(3.3).

In step 2, if $\dim(\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}) = 0$, then the intersection of the two varieties consists of finitely many points; the condition $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R} \neq \emptyset$ indicates that at least one of the points is real, which is straightforward to determine. If $\dim(\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}) > 0$, to check if $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R} \neq \emptyset$, one needs more sophisticated methods, such as those in [23]. In particular, such methods will return a point in the intersection, if such a real point exists; if the point is smooth on $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}$, then we can also conclude that the dimension of $(\mathcal{V}_\mathcal{M})_\mathbb{R} \cap (\mathcal{V}_\mathcal{D})_\mathbb{R}$ is equal to the dimension of $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}$.

To find the pair $((\hat{a},\hat{x},\hat{z}),(\hat{a},\hat{x},y))$ in step 3, one may solve the polynomial system (3.2)–(3.4). If there is a positive-dimensional set of (complex) extremal points, then the procedures in [23] could be used to determine if the set contains a real point.

If there are constraints on the variables or parameters, for example, if we seek to minimize (3.1) over the non-negative part of $(\mathcal{V}_\mathcal{M})_\mathbb{R}$, then the algorithm is updated as follows: if the $\dim(\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}) = 0$ or $\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D} = \emptyset$, then the Fritz John equations as described in §3.1 of the electronic supplementary material should be used, while if $\dim(\mathcal{V}_\mathcal{M} \cap \mathcal{V}_\mathcal{D}) > 0$, or if there is a positive-dimensional set of extremal points at step 3, then the algorithm should return *possibly*. There are methods [32–34] for finding real points, curves and surfaces within complex components of dimension 2 or less, but little is known about higher dimensions.

3.2.2. Algorithm 2: model selection

In this case, there are several competing models, each with its own polynomial system. The algorithm proceeds much as in algorithm 1, but iterated for the several models under consideration. If a threshold α is provided, then one should first reject models that do not adequately support the data ($d^2 \geq p_\alpha$), then choose the model yielding the minimum value of d^2 (up to some complexity penalty). Various conclusions may be drawn, e.g. *no model adequately fits the data* or *three models adequately fit the data and model $\mathcal{M}_1$ provides the best fit*.

3.2.3. Algorithm 3: parameter estimation

Again, this algorithm is similar to the first. The input consists of only one model $\mathcal{M}$ and data $\mathcal{D}$. It is assumed that there are unknown parameters and the goal is to find values of these parameters producing the best fit between $\mathcal{M}$ and $\mathcal{D}$. The

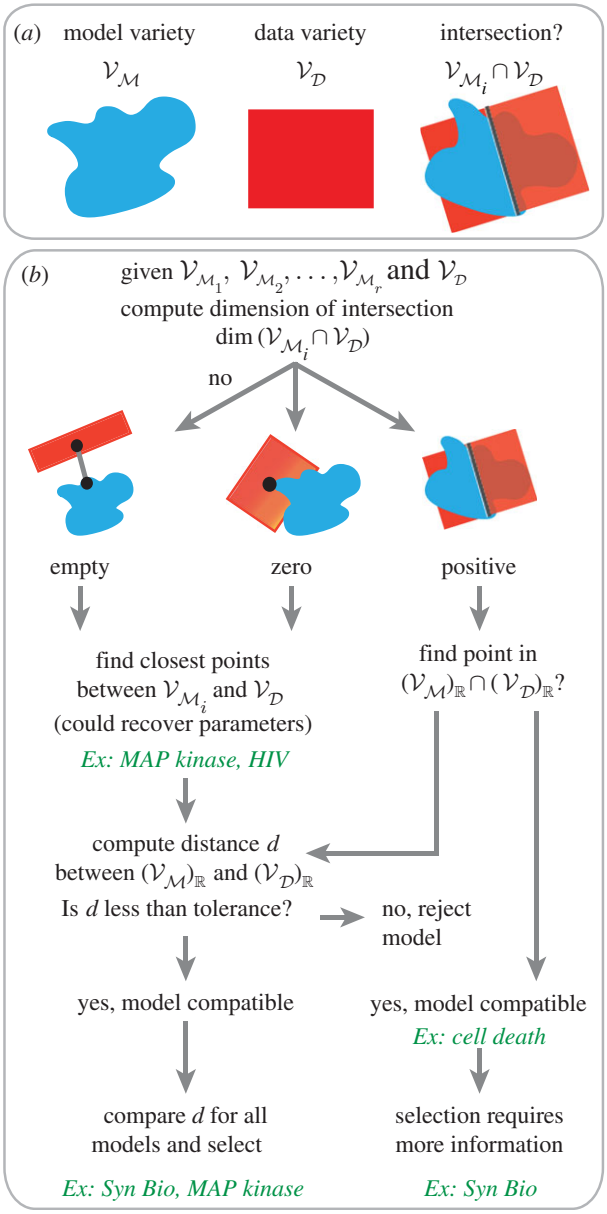

Figure 1. Schematic of numerical algebraic geometry framework corresponding to algorithms 1–3. (*a*) Input to algorithms include model (system of polynomials) translated into a model variety (red), and steady-state data translated into a data variety (blue). (*b*) Flow chart of model compatibility, parameter estimation and model selection methods. Examples (green) are described in §4.

outputs of steps 2 and 3 need to be adjusted appropriately. The output of step 4 is simply $(\hat{a}, \hat{x}, \hat{z})$, because there is no α to be used for rejection. The method also simultaneously estimates hidden/unknown variables and 'de-noised' outputs.

3.2.4. Simple example

To illustrate algorithm 1, consider a simple model with three variables x, y, z and three parameters a, b, c satisfying

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Let $\alpha = 0.1$ and assume, for this example, that the variances on the errors are $\sigma_i^2 = 0.1$ for $i = 1, 2, 3$. The model variety $\mathcal{V}_{\mathcal{M}}$ is simply an ellipsoid (figure 2*a*) where a, b, c describe the principal axes. Suppose we know that $a, b, c = 1$. For the case when the outputs are x and y and their observed values are $x' = 0$, $y' = 0$ and $\alpha = 0.1$, step 2 indicates that

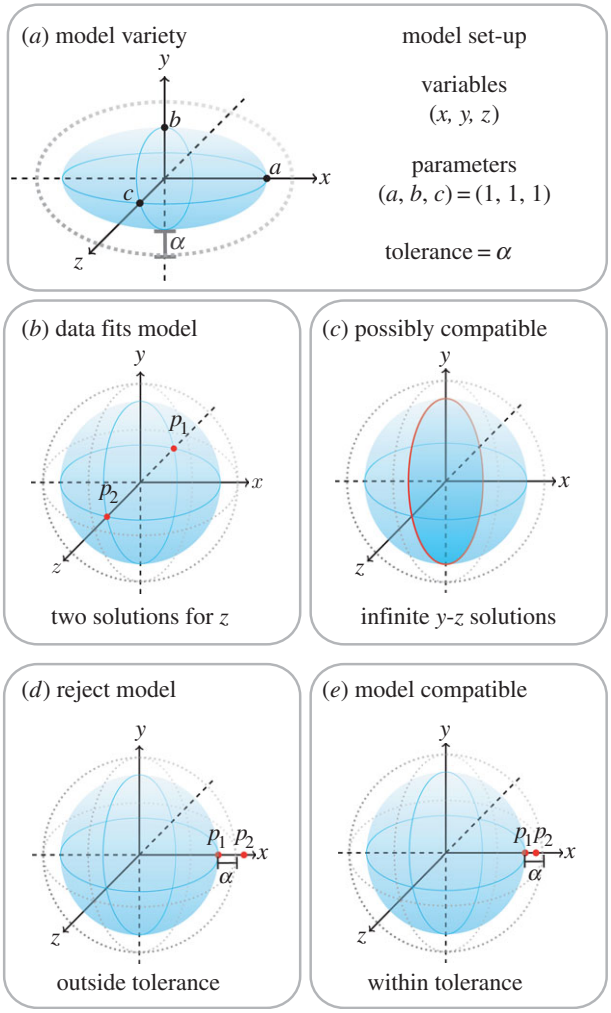

Figure 2. Simple example demonstrating model compatibility following algorithm 1. For ease of illustrating the main idea, we use α instead of $\sqrt{p_\alpha}$ in this figure.

the data *do* fit the model, i.e. the model is compatible with the data. In this case, there are two real points in the zero-dimensional intersection $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ (figure 2*b*). For the same set-up, but with data $x' = 0$, step 3 indicates that these data *possibly* fit this model. Because there is a positive-dimensional intersection (figure 2*c*), it is possibly compatible (depending on constraints imposed by the user). For the same model and α, but different data $x' = 1.7$, $y' = 0$, step 3 yields points $(1, 0, 0)$ and $(1.7, 0, 0)$, so the observed value of d^2 is greater than $p_\alpha = 0.4605$, and the model is rejected (figure 2*d*). However, when the data are $x' = 1.01$, $y' = 0$, algorithm 1 indicates model compatibility (figure 2*e*). Previous algebraic methods that required Gröbner basis calculations would result in an empty ideal and thus those approaches are not useful here.

4. Results

We demonstrate our methods on problems in biomedicine: cell death activation, synthetic biology, epidemiology and multisite phosphorylation. Each of the forthcoming applications can be written as a mass-action chemical reaction network, which has the form $\dot{x} = f(a, x)$, studied at steady state: $f(a, x) = 0$ as in the problem statement. Throughout these real-world examples, we emphasize the pivotal computations in these methods, such as determining the dimension

of the intersection $V_{\mathcal{M}} \cap V_{\mathcal{D}}$ and finding points in the intersection (figure 1*b*). Our first two examples, cell death signalling and genetic toggle biocircuits, demonstrate how to handle positive-dimensional intersections using two different approaches, whereas the remaining two examples, HIV and MAP kinase signalling, highlight analysis of zero-dimensional and empty intersections. In the following examples, we are interested in results that can be interpreted biologically; therefore, we restrict our analysis to non-negative real solutions. We also compare our method results to the squeeze-and-breathe parameter estimation optimization algorithm [35].

4.1. Cell death activation

We demonstrate model compatibility (algorithm 1) on an example from receptor-mediated programmed cell death, which is initiated by the activation of death receptors upon the detection of extracellular death ligands [36–39]. We consider, in particular, the 'cluster' model of [12], which was inspired by crystallographic data [40] and describes the recruitment of receptors by ligands into local self-activating clusters capable of bistability.

The cluster model is a system of three degree-four polynomials in the form of (2.2) in three variables (representing various receptor states) and six rate parameters, supplemented by ligand and receptor conservations (see the electronic supplementary material). We assume that we can measure the total ligand and receptor concentrations, which may be considered experimental inputs, as well as the concentration of active receptors. We do not assume access to the concentrations of other individual receptor states nor to any of the rate parameters.

A steady-state data point was simulated from the model with all parameters and initial concentrations drawn independently and identically from the lognormal distribution $\ln \mathcal{N}(0,4)$, then combined and corrupted with i.i.d. noise from $\mathcal{N}(0,0.1)$ to obtain y. The real model and real data varieties intersect in the positive orthant with a distance zero and hence the model is indeed compatible with the data. As detailed in §4.1 of the electronic supplementary material, these runs required a total of approximately 15 s (averaged over 20 runs). The ED degree defined in §3 is 5 for this problem.

4.2. Synthetic biology and experimental design

We demonstrate an example from synthetic biology with excess intersection ($\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) > 0$). A goal in synthetic biology is to design or modify existing biological systems with new features according to specific design criteria. Reverse engineering of biological systems often includes modules (such as feedback loops), and how these are interconnected are described by different circuits (models). Understanding differences between biocircuit implementations is crucial; therefore, we compare three bistable biocircuits models analysed in [41]: monomer–dimer toggle circuit ($\mathcal{M}_1$), dimer–dimer toggle circuit ($\mathcal{M}_2$) and single-operator gene circuit ($\mathcal{M}_3$), which were initially presented in [42–44]. The model variables include genes (X_i) and proteins (P_i), where $i = 1$ for $\mathcal{M}_3$ or $i = 1, 2$ for $\mathcal{M}_1$, $\mathcal{M}_2$, and their species complexes (e.g. P_iP_i, $X_jP_iP_i$). These variables interact following mass-action kinetics and form systems of polynomial differential equations where $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ have 7, 8 and 6 model variables, respectively, and 10, 12 and 9 kinetic parameters, respectively. The models can be reduced (given in the electronic supplementary material) by assuming that the total amount of gene 1 ($X_{1_{\text{tot}}}$) and gene 2 ($X_{2_{\text{tot}}}$) is conserved.

Suppose that the total amounts $X_{1_{\text{tot}}}$ and $X_{2_{\text{tot}}}$ and specific protein synthesis and degradation parameters $k_{\text{bas}_1}, k_{\text{bas}_2}, k_{\text{deg}_1}$ and k_{deg_2} are known. Because protein concentrations are often measurable, we assume that our data are P_1, P_2, and their complexes P_1P_1 and P_2P_2. The aim is to select the best model $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ given the data. We simulate steady-state data from the dimer–dimer toggle model ($\mathcal{M}_2$) and add Gaussian noise from $\mathcal{N}(0, 0.1)$. We find that all three have positive-dimensional intersections, where the dimension of the intersections are

$$\dim((\mathcal{V}_{\mathcal{M}_1})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) = 3,$$
$$\dim((\mathcal{V}_{\mathcal{M}_2})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) = 4$$
and
$$\dim((\mathcal{V}_{\mathcal{M}_3})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) = 3.$$

Clearly, all three models are compatible with the data, thus one can only select a 'best fit' model using data-independent measures, e.g. number of parameters, dimension, etc.

In fact, the dimensions of the model and data varieties can help us design more informative experiments for model selection. This dimension calculation provides guidance towards the minimal number of additional variable and parameter values that must be measured to ensure $\mathcal{V}_{\mathcal{M}_1} \cap \mathcal{V}_{\mathcal{D}} = \emptyset$. For example, because $\dim((\mathcal{V}_{\mathcal{M}_1})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) = 3$, at least four more variables and parameter values must be known. Because we get positive-dimensional intersections, the model is non-identifiable for steady-state data.

Suppose we can experimentally measure four forward biochemical reaction rate constants (e.g. k_{cF}, k_{kF}, k_{nF} and k_{kR}), then $\mathcal{V}_{\mathcal{M}_1}$ is cut down by four dimensions and does not intersect $\mathcal{V}_{\mathcal{D}}$. We get similar results for the model varieties associated with $\mathcal{M}_2$ and $\mathcal{M}_3$ provided we measure rate constants specific to these models (see electronic supplementary material). Now that all the intersections are empty, we run algorithm 2 and find that the sums of squares (equation (3.7)) for each model are as follows: $d_1^2 = 0.2262$, $d_2^2 = 7.34 \times 10^{-7}$ and $d_3^2 = 0.3040$. Therefore, we select the $\mathcal{M}_2$ model, which is indeed the true model. As described in the electronic supplementary material, solving the zero-dimensional system for the monomer–dimer toggle circuit, $\mathcal{M}_1$, took 1 min and 3 s, solving the system for the dimer–dimer toggle circuit, $\mathcal{M}_2$, took 1 min and 43 s, and solving the system for the single-operator positive feedback circuit, $\mathcal{M}_3$ took 0.092 s. The ED degrees as described in §3 are, respectively, 3, 4 and 4.

We compare our results using the squeeze-and-breathe evolutionary optimization algorithm [35]. This method uses Monte Carlo simulation based on an initial parameter 'prior' to find local minima of the sum of squared errors (d^2) using derivative-free optimization. At the end of each iteration, it computes a 'posterior' from the best local minima, which is then used as a 'prior' in the next iteration. The sample–optimize–recompute cycle continues until convergence. One advantage is that the local optimization steps allow it to explore beyond the 'prior' (i.e. the true parameter value could lie outside of it) [35], which is ideal, because the NAG method does not restrict to certain ranges of the parameter space. Moreover, squeeze-and-breathe has good success finding the true parameter (global minimum) in biological models [45] and therefore is a suitable choice to compare with our NAG method. As before, we take data from $\mathcal{M}_2$ and estimate

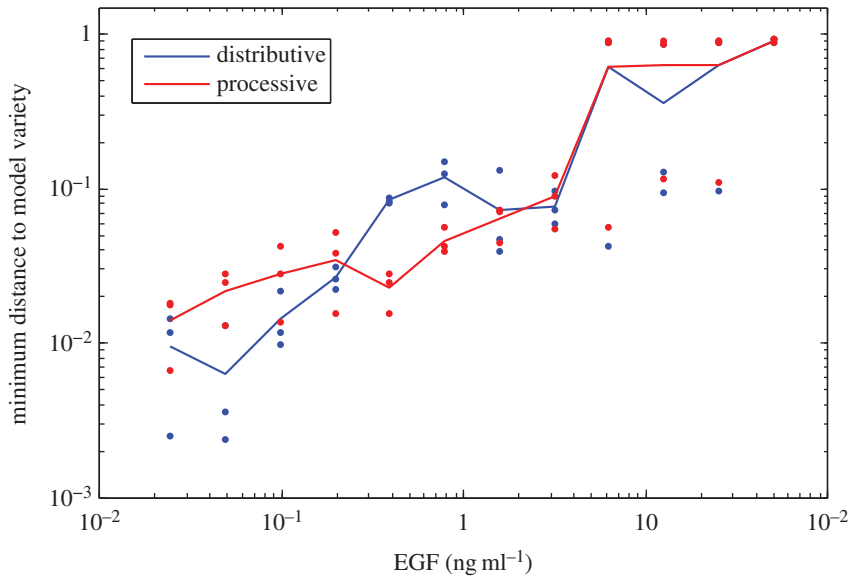

Figure 3. MAP kinase model selection using 36 data points from three model variables (aggregate phosphoforms of ERK) taken in triplicate at each of 12 EGF stimulation levels.

parameters based on the previous experimental design analysis. We estimate parameters k_{nR}, k_{kR} for $\mathcal{M}_1$, k_{kR}, k_{iR}, k_{nR} for $\mathcal{M}_2$ and k_{kR} for $\mathcal{M}_3$. We then get the sum of squared errors for the best-fit parameter of each model: $d_1^2 = 0.9640$, $d_2^2 = 0.001$, $d_3^2 = 0.3898$. The method takes approximately 10 min to run for each model and also correctly selects $\mathcal{M}_2$.

4.3. Human immunodeficiency virus progression

We demonstrate parameter estimation (algorithm 3) on an example coming from epidemiology. We use a model that includes long-term HIV dynamics from initial viremia, latency and virus increase [46], based on [47]. In the model (see the electronic supplementary material), the HIV virus inhibits the CD4$^+$ T cell population while promoting macrophage proliferation, which, in turn, houses the replicating virus. As macrophages proliferate, the virus reservoir increases, so the model offers a description of HIV patient progression to acquired immunodeficiency syndrome. Model variables x are uninfected CD4$^+$ T cells (T), infected CD4$^+$ T cells (T_i), uninfected macrophages (M), infected macrophages (M_i) and HIV virus population (V).

Hernandez-Vargas *et al*. show that the model can have two real equilibria, one of which is stable, representing patients that are 'long-term non-progressors' [46]. The parameters a are $(s_1, s_2, k_1, \ldots, k_6, \delta_1, \ldots, \delta_5)$, where s_i represents synthesis of T cells and macrophages, k_i are rate constants describing interactions between variables x, and δ_i represents natural death. For this example, $y = x$. We estimated the natural death of the virus, parameter δ_5, using the long-term non-progressors steady-state value (table 3 of [46]) and adding noise to each variable $\sim \mathcal{N}(0, 1)$. By algorithm 3, the data variety and model variety do not intersect. We find the closest point and estimate $\delta_5^* = 2.99876$ (true value of $\delta_5 = 3$, which was obtained by conferring with the authors of [46]). This run took 48 s, as described in the electronic supplementary material. The relevant ED degree is 16.

4.4. Multisite phosphorylation with experimental data

We examine phosphorylation mechanisms of cellular signalling with experimental data, and demonstrate model selection (algorithm 2) and parameter estimation (algorithm 3). We focus on phosphorylation, a key cellular regulatory mechanism that has been the subject of extensive study, both experimentally and theoretically ([48] and references therein). An area of interest is the mechanism by which a kinase phosphorylates a two-site substrate, either distributively, where the kinase can add at most one phosphate before dissociating, or processively, where it can add both phosphates in sequence. The MAPK/ERK pathway is a well-known system for studying phosphorylation, whereby MEK (kinase) phosphorylates ERK (its substrate). Aoki *et al*. [49] showed experimental evidence, while working with polynomial models, that the mammalian MAPK/ERK pathway acts distributively *in vitro* but processively *in vivo*.

We compare these distributive and processive models against the *in vivo* data reported in the same study. The distributive model consists of 12 molecular species and 17 mass-action reactions, whereas the processive model has 14 species and 18 reactions (species correspond to variables, each reaction corresponds to a parameter). The data take the form of 36 concentration measurements of three aggregate phosphoforms over a range of 12 EGF stimulation levels. All model parameters are calibrated using *in vitro* estimates by [49], except the parameter k_1 representing EGF loading, which we estimate for both models (see electronic supplementary material).

Next, we perform model selection by running algorithm 2 on each data point individually and select independently for each run the preferred model (more details in the electronic supplementary material). Under low EGF stimulation, the best model estimates are nearly identical with a slight preference for distributive. At high EGF stimulation, the models are identical with no preference for one model over the other. These results can be justified by noting that the main distinction between distributivity and processivity is nonlinear switching behaviour (i.e. a sigmoidal response curve), and this occurs only at intermediate stimulations. However, at medium EGF stimulation (figure 3), there is a preference for the processive model, which supports the findings in [49]. As detailed in §4.4.3 of the electronic supplementary material, all runs for the distributive model took

approximately 77 s, whereas the larger processive model required about 131 s (both averaged over 20 runs). It is worth noting that each of the 36 parameter homotopy runs in each of the two cases took less than 2% of the time of a regular, non-parameter homotopy run (less than 1% for the processive model). The ED degree for this problem is 20 for both models.

5. Conclusion

The problem of determining whether given real-world data fit one or more given mathematical models is challenging. When a model is defined by algebraic (polynomial) functions, the methods of NAG may be employed to study the geometry underlying the model and data. In particular, these methods are useful for model variables observed at steady state. We demonstrated this numerical and geometric framework for comparing models with experimental dose–response data in MAPK/ERK pathway and highlighted that the intermediate EGF doses were the most informative for model selection, complementing another finding that model selection results can be very sensitive to experimental parameters [50].

Despite the difficulties associated with positive-dimensional components and limitations in analysis, we can reproduce compatible models, and furthermore, can predict additional information, such as measurements of parameters, that are necessary for selecting models. Our geometric investigations of positive-dimensional components may perhaps relate to algebraic analyses for biochemical models, such as model identifiability or matroids for experimental design [13,51,52].

There are further directions to be considered in this vein, aside from making the existing computational methods more efficient. First, there would be great value in developing strictly *real* geometric methods for solving polynomial systems such as those that appear in this article. Some such techniques exist, but only in very special circumstances. Second, there would be much value in developing effective numerical methods for treating inequalities. It should be noted that the methods described in [53] and the references therein will incorporate such constraints, though the cost of such computations restricts their use to relatively low dimensions. Finally, there is likely much to be gained from considering the geometry underlying models not defined by algebraic functions. Algebraic geometry provides very clean, well-understood structures, paving the way for numerical methods. Differential geometry or topology could lead to similarly useful techniques for model selection and parameter estimation.

6. Material and methods

6.1. Numerical algebraic geometry

General references for NAG include [17,29], with the latter doubling as a user manual for the software package Bertini. For computations, we used Bertini 1.4 and Macaulay2 v. 1.6.

6.2. Data generation

Data simulated from cell death activation, synthetic biology and HIV models were performed in Matlab R2014b using ode15s.

Authors' contributions. E.G., K.L.H., D.J.B., H.A.H. designed the study. E.G., B.D., K.L.H., D.J.B., H.A.H. performed experiments and analysis. E.G., B.D., K.L.H., D.J.B., H.A.H. wrote the paper.

Competing interests. We declare we have no competing interests.

Funding. E.G., K.L.H., D.J.B. and H.A.H. acknowledge funding from the American Institute of Mathematics (AIM). E.G. was supported by the US National Science Foundation grant DMS-1304167. B.D. was partially supported by NSF DMS-1115668. K.L.H. acknowledges support from NSF DMS-1203554. D.J.B. gratefully acknowledges partial support from NSF DMS-1115668, NSF ACI-1440467, and the Mathematical Biosciences Institute (MBI). H.A.H. gratefully acknowledges funding from AMS Simons Travel Grant, EPSRC Fellowship EP/K041096/1, King Abdullah University of Science and Technology (KAUST) KUK-C1-013-04 and MPH Stumpf Leverhulme Trust Grant.

Acknowledgements. We thank Kazuhiro Aoki for discussions and providing model files and experimental data files for the MAP kinase example. We thank Vishesh Karwa for helpful feedback on an earlier version of the manuscript. Special thanks to Mariano Beguerisse-Díaz who helped us with squeeze-and-breathe.

References

1. Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. New York, NY: Springer.
2. Chamberlin TC. 1965 The method of multiple working hypotheses: with this method the dangers of parental affection for a favorite theory can be circumvented. *Science* **148**, 754–759. (doi:10.1126/science.148.3671.754)
3. Kirk PDW, Thorne T, Stumpf MPH. 2013 Model selection in systems and synthetic biology. *Curr. Opin. Biotechnol.* **24**, 767–774. (doi:10.1016/j.copbio.2013.03.012)
4. Liepe J, Kirk PDW, Filippi S, Toni T, Barnes CP, Stumpf MPH. 2014 A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nat. Protoc.* **9**, 439–456. (doi:10.1038/nprot.2014.025)
5. Lillacci G, Khammash M. 2010 Parameter estimation and model selection in computational biology. *PLoS Comput. Biol.* **6**, e1000696. (doi:10.1371/journal.pcbi.1000696)
6. Kirkpatrick S, Gelatt CD, Vecchi MP. 1983 Optimization by simulated annealing. *Science* **220**, 671–680. (doi:10.1126/science.220.4598.671)
7. Černý V. 1985 Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *J. Optim. Theory Appl.* **45**, 41–51. (doi:10.1007/BF00940812)
8. Anderson J, Papachristodoulou A. 2009 On validation and invalidation of biological models. *BMC Bioinformatics* **10**, 132. (doi:10.1186/1471-2105-10-132)
9. Mélykúti B, August E, Papachristodoulou A, El-Samad H. 2010 Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Syst. Biol.* **4**, 38. (doi:10.1186/1752-0509-4-38)
10. Rumschinski P, Borchers S, Bosio S, Weismantel R, Findeisen R. 2010 Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC Syst. Biol.* **4**, 69. (doi:10.1186/1752-0509-4-69)
11. Harrington HA, Ho KL, Thorne T, Stumpf MPH. 2012 Parameter-free model discrimination criterion based on steady-state coplanarity. *Proc. Natl Acad. Sci USA* **109**, 15 746–15 751. (doi:10.1073/pnas.1117073109)
12. Ho KL, Harrington HA. 2010 Bistability in apoptosis by receptor clustering. *PLoS Comput. Biol.* **6**, e1000956. (doi:10.1371/journal.pcbi.1000956)
13. MacLean AL, Rosen Z, Byrne HM, Harrington HA. 2015 Parameter-free methods distinguish Wnt pathway models and guide design of experiments.

Proc. Natl Acad. Sci. USA **112**, 2652–2657. (doi:10.1073/pnas.1416655112)

14. Manrai AK, Gunawardena J. 2008 The geometry of multisite phosphorylation. *Biophys. J.* **95**, 5533–5543. (doi:10.1529/biophysj.108.140632)
15. Drton M, Sturmfels B, Sullivant S. 2009 *Lectures on algebraic statistics*, vol. 39 of *Oberwolfach seminars*. Basel, Switzerland: Springer.
16. Bates DJ, Hauenstein JD, Sommese AJ, Wampler CW. 2014 Bertini: software for numerical algebraic geometry. See bertini.nd.edu.
17. Sommese AJ, Wampler CW. 2005 *The numerical solution to systems of polynomials arising in engineering and science*. Singapore: World Scientific.
18. Bates DJ, Decker W, Hauenstein JD, Peterson C, Pfister G, Schreyer F-O, Sommese AJ, Wampler CW. 2014 Comparison of probabilistic algorithms for analyzing the components of an affine algebraic variety. *Appl. Math. Comput.* **231**, 619–633. (doi:10.1016/j.amc.2013.12.165)
19. Akaike H. 1971 Information theory and an extension of the maximum likelihood principle. In *2nd Int Symp. on Information Theory* (eds BN Petrov, F Csáki), pp. 267–281. Budapest, Hungary: Akadémiai Kiadó.
20. Schwarz GE. 1978 Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464. (doi:10.1214/aos/1176344136)
21. Gross E, Petrović S, Verschelde J. 2013 Interfacing with phcpack. *J. Softw. Algebra Geomet.* **5**, 20–25. (doi:10.2140/jsag.2013.5.20)
22. Hauenstein JD, Rodriguez JI, Sturmfels B. 2014 Maximum likelihood for matrices with rank constraints. *J. Algebraic Stat.* **5**, 18–38. (doi:10.18409/jas.v5i1.23)
23. Hauenstein JD. 2013 Numerically computing real points on algebraic sets. *Acta Appl. Math.* **125**, 105–119. (doi:10.1007/s10440-012-9782-3)
24. Rouillier F, Roy M-F, Safey El Din M. 2000 Finding at least one point in each connected component of a real algebraic set defined by a single equation. *J. Complex.* **16**, 716–750. (doi:10.1006/jcom.2000.0563)
25. Seidenberg A. 1954 A new decision method for elementary algebra. *Ann. Math.* **2**, 365–374. (doi:10.2307/1969640)
26. Draisma J, Horobeţ E, Ottaviani G, Sturmfels B, Thomas RR. 2015 The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.* **16**, 99–149. (doi:10.1007/s10208-014-9240-x)
27. Verschelde J. 1999 Algorithm 795: Phcpack: a general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. Math. Softw.* **25**, 251–276. (doi:10.1145/317275.317286)
28. Lee T-L, Li T-Y, Tsai C-H. 2008 Hom4ps-2.0: a software package for solving polynomial systems by the polyhedral homotopy continuation method. *Computing* **83**, 109–133. (doi:10.1007/s00607-008-0015-6)
29. Bates DJ, Hauenstein JD, Sommese AJ, Wampler CW. 2013 *Numerically solving polynomial systems with Bertini*, vol. 25. Philadelphia, PA: SIAM.
30. Li TY, Sauer T, Yorke JA. 1989 The cheater's homotopy: an efficient procedure for solving systems of polynomial equations. *SIAM J. Numer. Anal.* **26**, 1241–1251. (doi:10.1137/0726069)
31. Morgan AP, Sommese AJ. 1989 Coefficient-parameter polynomial continuation. *Appl. Math. Comput.* **29**, 123–160. (doi:10.1016/0096-3003(89)90099-4)
32. Brake DA, Bates DJ, Hao W, Hauenstein JD, Sommese AJ, Wampler CW. 2014 Bertini real: software for one- and two-dimensional real algebraic sets. In *Mathematical software—ICMS 2014* (eds H Hong, C Yap). Lecture Notes in Computer Science, vol. 8592, pp. 175–182. Berlin, Germany: Springer. (doi:10.1007/978-3-662-44199-2_29)
33. Besana GM, Di Rocco S, Hauenstein JD, Sommese AJ, Wampler CW. 2013 Cell decomposition of almost smooth real algebraic surfaces. *Numer. Algorithms* **63**, 645–678. (doi:10.1007/s11075-012-9646-y)
34. Lu Y, Bates DJ, Sommese AJ, Wampler CW. 2007 Finding all real points of a complex curve. *Contemp. Math.* **448**, 183. (doi:10.1090/conm/448/08665)
35. Beguerisse-Díaz M, Wang B, Desikan R, Barahona M. 2012 Squeeze-and-breathe evolutionary Monte Carlo optimization with local search acceleration and its application to parameter fitting. *J. R. Soc. Interface* **9**, 1925–1933. (doi:10.1098/rsif.2011.0767)
36. Fulda S, Debatin KM. 2006 Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene* **25**, 4798–4811. (doi:10.1038/sj.onc.1209608)
37. Meier P, Finch A, Evan G. 2000 Apoptosis in development. *Nature* **407**, 796–801. (doi:10.1038/35037734)
38. Raff M. 1998 Cell suicide for beginners. *Nature* **396**, 119–122. (doi:10.1038/24055)
39. Thomson C. 1995 Apoptosis in the pathogenesis and treatment of disease. *Science* **267**, 1456–1462. (doi:10.1126/science.7878464)
40. Scott FL *et al.* 2009 The Fas-FADD death domain complex structure unravels signalling by receptor clustering. *Nature* **457**, 1019–1022. (doi:10.1038/nature07606)
41. Siegal-Gaskins D, Franco E, Zhou T, Murray RM. 2015 An analytical approach to bistable biological circuit discrimination using real algebraic geometry. *J. R. Soc. Interface* **12**, 20150288. (doi:10.1098/rsif.2015.0288)
42. Gardner TS, Cantor CR, Collins JJ. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342. (doi:10.1038/35002131)
43. Hasty J, Isaacs F, Dolnik M, McMillen D, Collina JJ. 2001 Designer gene networks: towards fundamental cellular control. *Chaos* **11**, 207–220. (doi:10.1063/1.1345702)
44. Siegal-Gaskins D, Mejia-Guerra MK, Smith GD, Grotewold E. 2011 Emergence of switch-like behavior in a large family of simple biochemical networks. *PLoS Comput. Biol.* **7**, e1002039. (doi:10.1371/journal.pcbi.1002039)
45. Beguerisse-Díaz M, Hernández-Gómez MC, Lizzul AM, Barahona M, Desikan R. 2012 Compound stress response in stomatal closure: a mathematical model of ABA and ethylene interaction in guard cells. *BMC Syst. Biol.* **6**, 146. (doi:10.1186/1752-0509-6-146)
46. Hernandez-Vargas EA, Middleton R, Mehta D. 2011 Towards modeling HIV long term behavior. *Proc. 18th IFAC World Congress* **18**, 581–586. (doi:10.3182/20110828-6-it-1002.00685)
47. Hadjiandreou M, Conejeros R, Vassiliadis VS. 2007 Towards a long-term model construction for the dynamic simulation of HIV infection. *Math. Biosci. Eng.* **4**, 489–504. (doi:10.3934/mbe.2007.4.489)
48. Futran AS, Link AJ, Seger R, Shvartsman SY. 2013 ERK as a model for systems biology of enzyme kinetics in cells. *Curr. Biol.* **23**, R972–R979. (doi:10.1016/j.cub.2013.09.033)
49. Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M. 2011 Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proc. Natl Acad. Sci. USA* **108**, 12 675–12 680. (doi:10.1073/pnas.1104030108)
50. Silk D, Kirk PDW, Barnes CP, Toni T, Stumpf MPH. 2014 Model selection in systems biology depends on experimental design. *PLoS Comput. Biol.* **10**, e1003650. (doi:10.1371/journal.pcbi.1003650)
51. Gross E, Harrington HA, Rosen Z, Sturmfels B. 2015 Algebraic systems biology: a case study for the Wnt pathway. (https://arxiv.org/abs/1502.03188)
52. Meshkat N, Sullivant S. 2014 Identifiable reparametrizations of linear compartment models. *J. Symb. Comput.* **63**, 46–67. (doi:10.1016/j.jsc.2013.11.002)
53. Basu S, Pollack R, Roy M-F. 2006 *Algorithms in real algebraic geometry* (*algorithms and computation in mathematics*). Secaucus, NJ: Springer.