SIAM J. DISCRETE MATH.
Vol. 34, No. 4, pp. 2510–2531

SOLUTIONS TO THE MINIMUM VARIANCE PROBLEM USING DELAUNAY TRIANGULATION*

ALEXANDER STRANG†

Abstract. We consider the problem of minimizing the variance of a distribution supported on a finite set of points Ω in $\mathbb{R}^n$ given the expected value of the distribution. This produces the distribution with the least uncertainty in X, in the l_2 sense, given the support and the mean. We show that, for general norms, the support of the solution must be small in the sense that it does not contain any points from Ω in the interior of its convex hull. We then show that, under an appropriate choice of norm on the covariance, the solution is given by evaluating the tent functions associated with a Delaunay triangulation of the support at the mean. Moreover, when the Delaunay triangulation is not unique the space of solutions is the space of convex combinations of solutions associated with each possible triangulation. Solutions to special cases are presented, along with a special solution on the lattice which simultaneously minimizes three natural choices of norm.

Key words. covariance, convex optimization, linear programming, triangulation

AMS subject classifications. 62J10, 52B25, 90C05, 32B25

DOI. 10.1137/19M1301837

1. Introduction.

We consider solutions to the following constrained optimization problem.

The minimum variance problem (MVP). *Let Ω be a finite set of points in $\mathbb{R}^n$. Let conv(Ω) be the space of all convex combinations of the vertices of Ω. Let $\bar{x}$ be a point in conv(Ω). Let $\rho(V)$ be a norm on symmetric positive definite matrices. Find the space of probability distributions p with support supp(p) $= S \subseteq \Omega$ and mean $\mathbb{E}_p[X] = \bar{x}$ which minimize $\rho(\mathbb{V}_p[X])$ where $\mathbb{V}_p[X] = \mathbb{E}_p[(X - \bar{x})(X - \bar{x})^T]$ is the covariance of the distribution.*

The MVP is motivated by moment closure problems for discrete space stochastic processes such as chemical reaction networks. In a moment closure approximation the time evolution of a probability distribution is approximated by the time evolution of its low order moments. The dynamics of the low order moments are typically coupled to higher order moments, so they usually cannot be simulated exactly without simulating all moments. In a moment closure approximation the higher moments are approximated as functions of the lower order moments, thereby producing a finite set of coupled evolution equations for the lower order moments. These equations can be used to simulate the dynamics of the low order moments. Sample applications of moment closure to chemical reaction networks can be found in [21], [26], [33], [40], and a review is provided in [19]. Moment closure can be used to create fast approximate simulation algorithms for systems of chemical reactions [40]. These algorithms are important since many essential biological processes that occur at the cellular scale are accurately modeled using stochastic chemical reaction networks [12], [29], [30], [33]. Reaction networks are usually studied using approximate Monte Carlo simulation since both analysis and exact simulation are often intractable. Moment closure eases the computational burden by estimating only the relevant low order moments.

*Received by the editors November 25, 2019; accepted for publication (in revised form) September 22, 2020; published electronically December 8, 2020.
https://doi.org/10.1137/19M1301837
Funding: This work was supported by NSF grant DEB-1654989.
†Statistics, University of Chicago, Chicago, IL 60615 USA (alexstrang@uchicago.edu).

A fundamental problem in moment closure is to understand the range of admissible higher order moments, given the lower order moments. If the assumed moments are too large or too small given the geometry of the reaction network, then approximate simulation algorithms based on those assumptions can fail. A motivating example, based on the algorithms presented in [29], [30], and [40], is provided in Appendix A. In the example, fatal errors arise from underestimation of the variance.

The goal of the MVP is to find a distribution with minimal covariance given the mean. This MVP is distinct from other well studied minimum variance methods (cf. [2] and [15]).

In this paper we prove the following. The MVP is a convex optimization problem for all choices of ρ, solutions exist for all continuous ρ, and the covariance of the solutions is unique if ρ obeys the strict triangle inequality (see Theorem 2.1). Examples of norms considered are the trace $\rho(V) = \text{trace}(V)$, induced two norm $||V||_2$, and Frobenius norm $||V||_F$. Under reasonable assumptions on ρ the supports of solutions to the MVP are always small in the sense that they do not contain any point from Ω in the interior of their convex hull (see Theorem 2.2). For $\rho(V) = \text{trace}(V)$, we show that the MVP is a linear programming problem and the solution to the MVP is given by evaluating piecewise linear basis functions associated with a Delaunay triangulation of Ω (see Theorem 2.3). When this triangulation is not unique the space of solutions to the MVP is the space of convex combinations of solutions on each possible triangulation. In section 3 we consider a series of special cases. These are chosen to illustrate the general solution and its limitations when applied to highly symmetric Ω as may arise from chemical reaction networks. An alternative representation of the solution space on integer lattices is proposed in Lemma 3.1 and is applied to find a special solution which simultaneously minimizes all of the singular values of the covariance (3.5).

2. Results. The MVP is a constrained optimization problem with both inequality and equality constraints. Let $p(x)$ denote the probability $X = x$ for any $x \in \Omega$. Then the constraints are as follows.

1. Nonnegativity: $p(x) \geq 0$.
2. Normalization: $\sum_{x\in\Omega} p(x) = 1$.
3. Mean: $\sum_{x\in\Omega} x_i p(x) = \bar{x}_i$ for any $i \in [1, 2, \ldots, n]$.

The cost function to minimize is

$$f(p) = \rho(\mathbb{V}_p[X]). \tag{2.1}$$

Let $\Omega_p(\bar{x})$ denote the domain defined by the three constraints. Note that $\Omega_p(\bar{x})$ depends on the position of the mean, and the boundaries of $\Omega_p(\bar{x})$ encode the supports of possible solutions. For general $\bar{x}$ all of these supports include at least $n+1$ vertices. If $\bar{x} = x \in \Omega$, then $\Omega_p(\bar{x})$ will include the delta distribution at x. If $\bar{x}$ lies in the convex combination of $m < n + 1$ points, then $\Omega_p(x)$ will include a distribution supported on those points. Here we focus on solutions to the general case since those solutions extend to the special cases.

The domain $\Omega_p(\bar{x})$ is the intersection of a unit hypercube (nonnegativity), which is a polytope, with an affine subspace (normalization and mean). The intersection of a polytope with a affine subspace is always a polytope [36], so the domain is a polytope. Since $\Omega_p(\bar{x})$ is a polytope, the MVP can be recast geometrically.

The MVP (geometric). *Let Ω be a finite set of points in $\mathbb{R}^n$ with a point $\bar{x} \in$ conv(Ω). Let $\Omega_p(\bar{x})$ be the set of all coefficients of convex combinations of the vertices*

of Ω that add up to $\bar{x}$. Find the set of coefficients p which minimize $f(p)$ over the polytope $\Omega_p(\bar{x})$.

THEOREM 2.1 (existence, uniqueness, and convexity). *The MVP is convex for all norms ρ. If $\rho(V)$ is a continuous function from $\mathbb{R}^{n^2} \to \mathbb{R}$, then there always exist solutions to the MVP. If $\rho(V)$ is continuous and satisfies the strict triangle inequality, $\rho(V+W) < \rho(V) + \rho(W)$ for all W not proportional to V, then all solutions to the MVP have the same covariance.*

Proof. All ρ are convex in V since all norms obey the triangle inequality. Since the mean of the distribution is fixed, the covariance is linear in the probability at each vertex of Ω. Then, for all choices of ρ, the cost function is the composition of a linear function of the probabilities, $\mathbb{V}_p[X]$, with a convex function. It follows that the cost function $f(p)$ is convex. The domain defined by the constraints is bounded, closed, and convex. Therefore the MVP is always a convex optimization problem.

The covariance is linear in the probabilities, so it is a continuous function of the probabilities. The composition of continuous functions is continuous, so the cost function is continuous in the distribution if ρ is continuous. By the extreme value theorem, any continuous function on a bounded closed domain achieves a minimum on that domain; therefore, a solution to the MVP always exists.

Suppose that ρ obeys the strict triangle inequality and there are two solutions to the MVP with covariances V and W that are not mutually proportional. Consider a convex combination of these solutions. The covariance of the combination is the convex combination of the covariances since the distributions share the same mean. Then, by the strict triangle inequality, $\rho(tV + (1-t)W) < |t|\rho(V) + |1-t|\rho(W)$. But $\rho(V) = \rho(W)$ since V and W are covariances corresponding to global minima, so $\rho(tV + (1-t)W) < \rho(V)$ for $t \in (0,1)$. This contradicts the claim that V and W minimize ρ. It follows that all solutions to the MVP must share proportional covariances. Now suppose that $W = \lambda V$ for $\lambda > 1$. Then $\rho(W) = |\lambda|\rho(V) > \rho(V)$ so W cannot be the covariance of a solution to the MVP. Thus, if ρ obeys the strict triangle inequality, the covariances of all solutions to the MVP are identical. ∎

The solution to a constrained optimization problem must satisfy the Karush–Kuhn–Tucker (KKT) conditions. The KKT conditions generalize the method of Lagrange multipliers to constrained optimization problems with both equality and inequality constraints [18], [20]. The KKT conditions are sufficient for optimality in this case since the optimization problem is convex with affine equality constriants.

Assuming the norm ρ is differentiable the KKT conditions are as follows.

1. Primal feasibility: The probability distribution $p(x)$ has mean $\bar{x}$ and is supported on Ω.
2. Stationarity: For all $x \in \Omega$,

$$\sum_{i,j} (x - \bar{x})_i \partial_{v_{ij}} \rho(\mathbb{V}_p[X])(x - \bar{x})_j + \lambda^T x + \lambda_0 = \mu(x), \tag{2.2}$$

   where $\partial_{v_{ij}} \rho(\mathbb{V}_p[X])$ denotes the partial derivative of the norm ρ with respect to the ij entry of $\mathbb{V}_p[X]$, λ is an $n \times 1$ vector, λ_0 is a scalar, and $\mu(x)$ is a scalar valued function defined on Ω.
3. Dual feasibility: $\mu(x) \geq 0$ for all $x \in \Omega$.
4. Complementary slackness: For all $x \in \Omega$,

$$\mu(x)p(x) = 0. \tag{2.3}$$

These rules establish a set of conditions on the support S of any distribution p which solves the MVP. In particular, under some conditions on the norm ρ, it can be shown that S is not the support of the solution if there is any point from Ω inside $\text{conv}(S)$. This restricts the support of solutions to the MVP to small subsets of Ω.

THEOREM 2.2 (general requirements on the support). *Assume that ρ is differentiable. Let $A(p)$ be the matrix with entries $a_{ij} = \partial_{v_{ij}}\rho(\mathbb{V}_p[X])$. If $A(p)$ is positive semidefinite for all p and $S = supp(p)$ is the support of a solution to the MVP, then* (i) *$\bar{x} \in conv(S)$,* (ii) *all points in S lie on the boundary of some ellipsoid, and* (iii) *if $x \in \Omega$ but $x \notin S$, then $x \notin conv(S)$.*

Proof. Complementary slackness requires that $\mu(x)p(x) = 0$ for all $x \in \Omega$. If $x \in S$, then, by definition, $p(x) \neq 0$, so $\mu(x) = 0$. Therefore, for all $x \in S$ the left-hand side of the stationarity requirement equals zero. For all x not in S dual feasibility requires $\mu(x) \geq 0$, so the left-hand side of the stationarity requirement is nonnegative. Therefore, the KKT conditions reduce to

$$\begin{aligned} &(x-\bar{x})^T A(p)(x-\bar{x}) + \lambda^T x + \lambda_0 = 0 \text{ for all } x \in S, \\ &(x-\bar{x})^T A(p)(x-\bar{x}) + \lambda^T x + \lambda_0 \geq 0 \text{ for all } x \notin S, \end{aligned} \tag{2.4}$$

where $A(p)$ is the sensitivity matrix $a_{ij}(p) = \partial_{v_{ij}}\rho(\mathbb{V}_p)$. This matrix is the sensitivity of the norm to each entry of the covariance.

Suppose S contains m points. Then, for a given p, the first line of (2.4) is a system of m linear equations in $n+1$ unknowns, $\lambda_0, \lambda_1, \ldots \lambda_n$. Suppose that λ_0, λ is a solution to this system of equations. Let $Q_{p,\lambda,\lambda_0}(x)$ be the quadratic function $(x-\bar{x})^T A(p)(x-\bar{x}) + \lambda^T x + \lambda_0$. Then stationarity requires that all points in the support S lie on the level surface where $Q_{p,\lambda,\lambda_0}(x) = 0$, and that, for all other points $x \in \Omega$, $Q_{p,\lambda_0,\lambda}(x) \geq 0$.

If $A(p)$ is positive definite for all p, then the isosurfaces of Q are ellipsoids. Then all the points $x \in S$ must lie on some ellipsoid, and all of the points $x \in \Omega$ not in S must lie outside this ellipsoid. Since, for a given S, there may be multiple distributions p with mean $\bar{x}$ and multiple solutions λ, λ_0 given p, there may be more than one quadratic function and associated ellipsoid. Note that $A(p)$ fixes the orientation and eccentricity of the ellipsoid, while λ fixes its center and λ_0 its scale.

Let E denote the union of an ellipsoid and its interior. The union of an ellipsoid with its interior is always convex, so if all the points in S lie on the surface of E, then $\text{conv}(S) \subset E$. Thus, if $x \in \Omega$ and $x \in \text{conv}(S)$, then x is necessarily inside of any ellipsoid passing through every point of S. Therefore, S cannot be the support if there is any $x \in \Omega$ that is not in S but is in $\text{conv}(S)$.

More generally, if $A(p)$ is positive semidefinite for all p, then the quadratic function $Q_{p,\lambda,\lambda_0}(x)$ is convex. Then any level surface $Q_{p,\lambda,\lambda_0}(x) = c$ bounds a convex region where $Q_{p,\lambda,\lambda_0}(x) \leq c$. Let E be the region such that $Q_{p,\lambda,\lambda_0}(x) \leq 0$. Then, since E is convex, $\text{conv}(S) \subset E$ for any possible E, so the conclusion follows as in the positive definite case. ∎

A variety of reasonable choices of the norm ρ have $A(p)$ which is necessarily positive semidefinite. Three important examples are as follows.

1. $\rho(V) = \text{trace}(V)$, which is the total variance, $\mathbb{E}\left[||X-\bar{x}||^2\right]$,
2. $\rho(V) = ||V||_2 = \sigma_1(V)$, which is the variance in the distribution when projected onto the direction which maximizes the variance,
3. $\rho(V) = ||V||_F = \sqrt{\sum_{i,j}|v_{ij}|^2}$, which is the Frobenius norm of the covariance.

Note that the two norm is only differentiable if the largest singular value of V is not a repeated singular value [24]. Respectively, these norms set $A(p) = I$, $A(p) = \mathbb{V}_p^{(1)}$ where $\mathbb{V}_p^{(1)}$ is the closest rank one approximation to $\mathbb{V}_p[X]$, and $A(p) = \mathbb{V}_p$ (see Appendix B). All of these norms are natural choices since they are invariant under unitary transformations, so they do not depend on the coordinate system used. All are continuous, so all admit solutions, and the Frobenius norm satisfies the strict triangle inequality, so it produces solutions with a unique minimal covariance.

The sensitivity matrix $A(p)$ depends on p in all but the first case. This is because the total variance is linear in the entries of $\mathbb{V}$ while all the other norms are nonlinear. Using the total variance is natural because it corresponds to the expected value of $||X - \bar{x}||^2$. The fact that $A(p)$ does not depend on p when minimizing the total variance separates the problem of finding possible supports S from distributions p on those supports. This allows us to present a general description of the space of all solution to the MVP with ρ set to the total variance.

2.1. Results for the total variance case ρ = trace. The general description of the solution space when $\rho = \text{trace}$ depends on the following definition.

Delaunay simplex. *A simplex is a Delaunay simplex for the set Ω if it satisfies the circumsphere condition: the interior of the unique sphere passing through all the vertices of the simplex contains no point from Ω* [9].

THEOREM 2.3 (solution space for total variance). *Let $\rho(V) = \text{trace}(V)$. If S are the vertices of a Delaunay simplex such that $\bar{x} \in \text{conv}(S)$, then there exists a unique distribution p with support S that solves the MVP. Moreover, this distribution is given by setting the probability at each $x \in S$ to the value of the tent function $\psi(\bar{x}, x)$ (piecewise linear basis function associated with node x on a Delaunay mesh including the simplex) evaluated at $\bar{x}$. If $\bar{x} \in \text{conv}(\Omega)$, then such a simplex always exists. Let $C(S)$ be the circumsphere of such a simplex. Let $R = C(S) \cap \Omega$. For each simplex with vertices in R containing $\bar{x}$ there exists a unique distribution p supported on the simplex which solves the MVP. The space of all solutions to the MVP is the space of convex combinations of the unique solutions on all such simplices.*

Proof. The first key idea is to show that the stationarity conditions (2.4) imply the circumsphere condition when ρ is set to the total variance. This is trivial since, if ρ is set to the total variance, $A(p) = I$, so the quadratic function $Q_{p,\lambda,\lambda_0}(x)$ always has spherical isosurfaces. Then, requiring that all $x \in S$ are roots of the quadratic function requires that all $x \in S$ lie on some sphere. Let $C(S)$ denote a circumsphere of the set S. If S corresponds to a simplex, then $C(S)$ is unique. The second half of the stationarity requirement requires that there is at least one circumsphere of S such that no points from Ω that are not in S are inside the circumsphere.

This condition is the same condition used to define a Delaunay triangulation [16]. A Delaunay triangulation partitions Ω into a series of simplices each satisfying the property that the circumsphere of the simplex does not contain any other points from Ω in its interior [7]. This is the standard triangulation often used for finite elements and for linear interpolation [5]. Every set of points Ω admits a Delaunay triangulation [6]. Therefore, for any point $\bar{x}$ there exists at least one simplex with vertices S satisfying the circumsphere condition and with $\bar{x} \in \text{conv}(S)$. Although a Delaunay triangulation always exists it is not always unique [16], so there may be multiple solutions to the MVP.

Let S be the vertices of a Delaunay simplex containing $\bar{x}$. Then, since $\bar{x} \in \text{conv}(S)$, there exists at least one distribution p with support S such that $\mathbb{E}_p[X] = \bar{x}$. Since

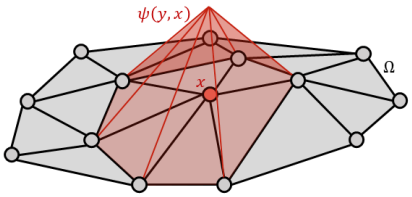

FIG. 1. *The tent function $\psi(y,x)$ for a triangulation of point set Ω associated with vertex x is shown in red.*

conv(S) is a simplex, this distribution is unique. Let x be a vertex of the simplex, and let $p(x|\bar{x})$ be the probability at vertex x given $\bar{x}$. Here we treat $p(x|\bar{x})$ as a function of $\bar{x}$. For any $\bar{x}$, the distribution $p(x|\bar{x})$ is the solution to an $n+1 \times n+1$ set of linear equations. The system of linear equations is always invertible (see Appendix C). Specifically, $p(x|\bar{x})$ is the inverse of a matrix, whose entries are defined in terms of the coordinates of the vertices, multiplied by the vector $[1;\bar{x}]$. It follows that $p(x|\bar{x})$ is an affine function of $\bar{x}$ as long as $\bar{x} \in \text{conv}(S)$ and we restrict to solutions with support S. This affine function is uniquely specified by noting that, if $\bar{x} = x$, then $p(x|\bar{x}) = 1$, and if $\bar{x} = y \in S \setminus x$, then $p(x|\bar{x}) = 0$. Consequently $p(x|\bar{x})$ is the unique affine function of $\bar{x}$ that equals one at $\bar{x} = x$ and zero at any other vertex of the simplex.

So, if we fix a Delaunay triangulation, then $p(x|\bar{x})$ is a continuous piecewise linear function of $\bar{x}$ that is affine for $\bar{x}$ in the interior of each simplex, one at $\bar{x} = x$, and zero if $\bar{x}$ equals any other node in Ω. These are exactly the conditions which define the tent functions (piecewise linear basis functions) $\psi(y,x)$ on the triangulation. The tent functions are the standard basis functions used for linear interpolation. An example tent function is illustrated in Figure 1.

It follows that, if we fix a Delaunay triangulation $\mathcal{T}$, then for each vertex in Ω there is a unique tent function $\psi(y,x)$, and the MVP is solved by letting the probability at each node x equal the value of the tent function associated with node x evaluated at $\bar{x}$:

$$p(x|\bar{x}) = \psi(\bar{x},x). \tag{2.5}$$

The tent functions can be evaluated by projection as illustrated in Figure 2. Let x be a vertex in S, let $u(S \setminus x)$ be the unit normal vector to the affine subspace containing $S \setminus x$ (the facet of the simplex opposite x), and let $c(S \setminus x)$ be the centroid of the face. Then the height of the simplex is the length between x and its projection onto the affine subspace containing $S \setminus x$, which equals $u(S \setminus x)^T(x - c(S \setminus x))$. Define the height function $h_x(y) = u(S \setminus x)^T(y - c(S \setminus x))$, which evaluates the height of the simplex if we replace node x with node y. Then the tent function (2.5) and associated probability are simply the ratio of the heights at x and $\bar{x}$. Note that the ratio of the heights of two simplices that only differ by a vertex is also the ratio of their volumes. The ratio of volumes is the standard equation for computing barycentric coordinates (a distribution of weights on the nodes of the simplex that average to a given point in the interior) [38]:

$$p(x|\bar{x}) = \psi(\bar{x},x) = \frac{\text{Vol}((S \setminus x) \cup \bar{x})}{\text{Vol}(S)} = \frac{h_x(\bar{x})}{h_x(x)} = \frac{u(S \setminus x)^T(\bar{x} - c(S \setminus x))}{u(S \setminus x)^T(x - c(S \setminus x))}. \tag{2.6}$$

Therefore, any Delaunay simplex with vertices S such that $\bar{x} \in \text{conv}(S)$ admits a solution to the MVP, the MVP has a unique solution supported on S, and at least

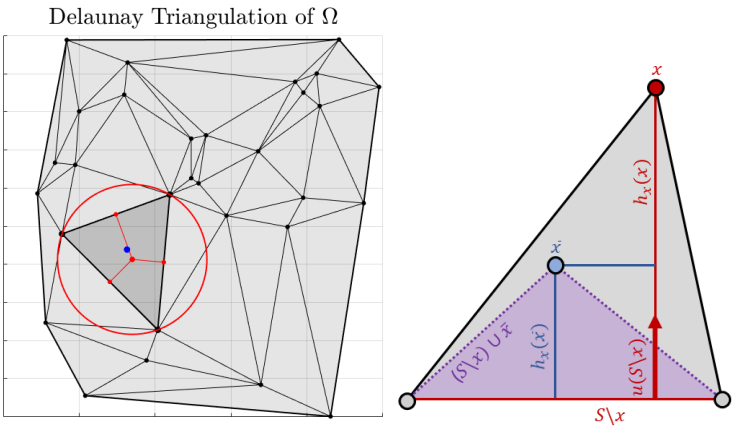

FIG. 2. *The left panel shows a Delaunay triangulation of* 30 *randomly distributed points drawn uniformly over the box* $[0,1]\times[0,1]$. *The points,* Ω *are shown in black, along with the triangulation. The mean* $\bar{x}$ *is the blue point inside the shaded triangle, and the simplex containing the mean is the grey shaded triangle. The bisectors of each side of the triangle are shown in red. Their intersection is the center of the circumcircle of the triangle. The circumcircle is shown in red. Note that no point from* Ω *is inside the circumcircle; thus this triangle supports a solution to the MVP. The right panel shows a triangle containing a point* $\bar{x}$ *illustrated with a blue shaded circle. The red vertex is* x. *The vertical red line is the height* $h_x(x)$ *from the edge of the simplex* $S\setminus x$ *to* x, *and the blue line is the height* $h_x(\bar{x})$ *from the edge* $S\setminus x$ *to* $\bar{x}$. *These can be computed via a dot product with the unit normal* $u(S\setminus x)$. *The probability at* x *given* $\bar{x}$ *is the ratio of the height of* $\bar{x}$ *to the full height of the triangle* $h_x(x)$. *This is equivalent to the fraction of the grey triangle contained in the purple triangle, which is given by replacing* x *with* $\bar{x}$.

one such simplex exists. We now show that the space of all solutions to the MVP is the space of convex combinations of solutions associated with these simplices. First we provide a recipe for finding these simplices.

Suppose that we start with a Delaunay triangulation of Ω. Then, for any $\bar{x}$, we have a unique simplex satisfying the KKT conditions. The circumsphere of any n-dimensional simplex is unique with center given by the intersection of the bisectors of the faces. Denote this circumsphere $C(S)$. Now let $R = C(S)\cap\Omega$. All solutions to the MVP have support contained in R, so the circumsphere associated with a particular solution can be used to find the support of all possible solutions. The proof follows.

The cost function of the MVP, $f(p) = \rho(\mathbb{V}_p[X])$, is linear in p for all p with mean $\bar{x}$, and the domain is a convex polytope. Therefore the MVP is a linear programming problem. It follows that the convex combination of any set of solutions is also a solution. The support of the convex combination of a set of solutions is the union of the supports of each solution. This means that, if there are solutions with different supports, then there must also be a solution with support equal to the union of the supports of each separate solution. In order for the union to be a solution it must also satisfy the circumsphere condition, so all points in the support must lie on at least one shared circumsphere. Since there is always a solution on at least one simplex, and since the associated circumsphere is unique, the support of all solutions lie on a unique circumsphere. Therefore all solutions to the MVP have support contained in $R = C(S)\cap\Omega$.

Given R, the MVP can be reduced to a new MVP restricted to the vertices in R. On R the stationarity requirement is satisfied, so all distributions on R with mean $\bar{x}$ are solutions to the MVP. The space of all possible distributions supported on a

subset of R with mean $\bar{x}$ is a convex polytope. It remains to show that the vertices of this polytope (the extreme solutions) are necessarily solutions on simplices. The constraints define a set of linear equations. The set of linear equations has a nontrivial nullspace if the support includes more than $n+1$ vertices (see Appendix C). Therefore, if p is supported on more than $n+1$ vertices it is not a vertex of the solution space. It follows that all the vertices of the solution space (all extreme solutions) are solutions on Delaunay simplices. ∎

COROLLARY 2.4. *If the points x in Ω are in general position, then the solution to the MVP with $\rho =$ trace is unique, and if they are drawn randomly from a continuous joint distribution, then the solution is almost surely unique.*

Proof. If $x \in \Omega$ are in general position, then no set of more than $n+1$ points lie on the same circumsphere (and the Delaunay triangulation is unique [7]). If the points x are drawn from a continuous joint distribution, then they are almost surely in general position. ∎

The problem of finding a Delaunay triangulation is well studied, as Delaunay triangulations have rich applications in finite elements, computer aided design, medical imaging, and computer graphics. Reviews of triangulation methods are provided in [16], [27], [32]. Multiple fast triangulation algorithms exist. The divide and conquer algorithm has been shown to be the fastest triangulation algorithm [34] and works in arbitrary dimension [4]. Divide and conquer typically runs in $\mathcal{O}(|\Omega| \log(|\Omega|))$ time [16], [22]. Parallelized divide and conquer algorithms designed for use in computer graphics are particularly efficient and have been shown to handle billions of points and process millions of simplices per second [23]. This suggests that, even if Ω is large, it may be possible to find $p(x|\bar{x})$ for all x before specifying $\bar{x}$, thereby solving the MVP for all possible $\bar{x}$.

Note that, if $\bar{x}$ is fixed, then we only need to find one simplex satisfying the circumsphere condition, so we only need a Delaunay triangulation of a subset of Ω. Therefore it is reasonable to try triangulating a neighborhood of $\bar{x}$ first. If we find a Delaunay triangulation of Ω in a neighborhood of $\bar{x}$, and the neighborhood contains the circumsphere of the simplex containing $\bar{x}$, then the tent function solution on the simplex is necessarily a solution to the MVP on all of Ω. It follows that the problem can be solved on local neighborhoods of $\bar{x}$, so the actual computational cost for a fixed $\bar{x}$ will depend on the cost of finding a triangulation of a sufficiently large neighborhood of $\bar{x}$, not all of Ω. This means that the problem could be solved numerically even if Ω is a countable set of points, provided the number of points from Ω in any compact neighborhood is finite. In this context, one of the incremental methods for constructing a Delaunay triangulation [16] may be more efficient than a divide and conquer algorithm.

3. Special cases. In special cases when Ω is highly regular, there may be many Delaunay simplices containing $\bar{x}$. For example, if Ω is a cubic lattice, then the circumsphere condition requires that p is supported on the hypercube containing $\bar{x}$ (see Lemma 3.1). The polytope of solutions supported on the hypercube contains many different extreme solutions as the hypercube supports at least $2^{n-1}n!+2^n$ simplices, of which at least 2^{n-1} contain $\bar{x}$. Ironically, this implies that, when Ω is highly regular, the space of solutions to the MVP can be difficult to explore. Here we will consider a series of special cases in order to show the power of the general solution, as well as its limitations when Ω is highly symmetric. We focus on regular lattices in particular, since they are frequently the underlying state space of reaction networks [21], [26], [33].

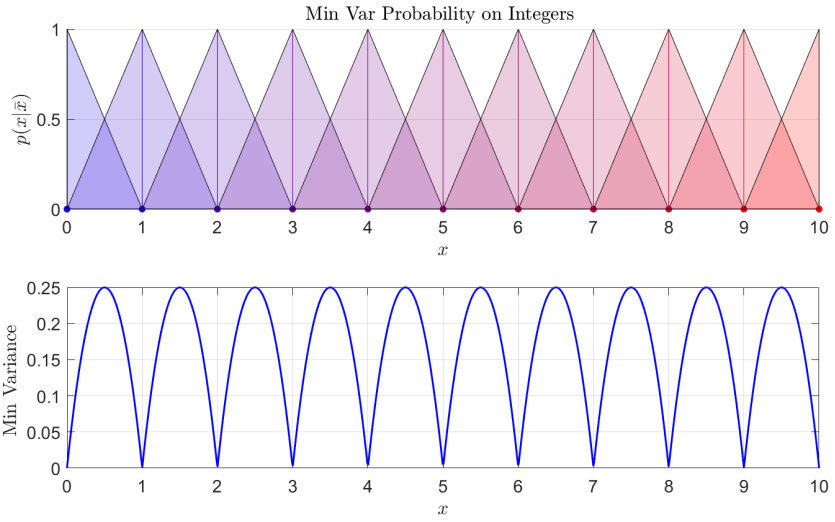

FIG. 3. *The upper panel shows the tent functions associated with the solution to the MVP on an integer lattice. The bottom panel shows the minimum variance as a function of* $\bar{x}$.

3.1. One dimension. Suppose that the points $x \in \Omega$ are all on the real line. Then the norms are equivalent, and the solution to the MVP using total variance is equivalent to simply minimizing the variance. Let $\lfloor y \rfloor$ be the largest vertex in Ω less than or equal to y and $\lceil y \rceil$ be the smallest vertex in Ω larger or equal to than y. Theorem 2.3 implies that the minimum variance distribution is unique since the Delaunay triangulation of Ω consists simply of the intervals between adjacent entries of Ω. Evaluating the tent functions on the intervals gives

$$p(x|\bar{x}) = \frac{1}{\lceil \bar{x} \rceil - \lfloor \bar{x} \rfloor} \begin{bmatrix} \lceil \bar{x} \rceil - \bar{x} \text{ if } x = \lfloor \bar{x} \rfloor \\ \bar{x} - \lfloor \bar{x} \rfloor \text{ if } x = \lceil \bar{x} \rceil \\ 0 \text{ if } x \neq \lfloor \bar{x} \rfloor \text{ or } \lceil \bar{x} \rceil \end{bmatrix}. \tag{3.1}$$

The associated variance is

$$\mathbb{V}(\bar{x}) = \frac{(\lceil \bar{x} \rceil - \bar{x})(\bar{x} - \lfloor \bar{x} \rfloor)}{\lceil \bar{x} \rceil - \lfloor \bar{x} \rfloor}. \tag{3.2}$$

If $\Omega = \mathbb{Z}$, then the solution to the MVP is $p(x|\bar{x}) = \max\{1 - |x - \bar{x}|, 0\}$. The solution and associated variances are illustrated in Figure 3.

3.2. Two-dimensional integer lattice. Consider a two-dimensional integer lattice $\Omega = \mathbb{Z}^2$. Then, for $\bar{x}$ in general position, there are two Delaunay simplices containing $\bar{x}$. This can be seen by noting that each $\bar{x}$ is contained inside a unit square and that the square can be subdivided into four triangles. General $\bar{x}$ lie in two of these triangles. All of these triangles have circumcenter equal to the center of the square, with radius $\sqrt{2}/2$. The corresponding circumspheres are all identical and include all four corners of the square.

It follows that, for general $\bar{x}$, the space of solutions is the space of convex combinations of the unique solutions associated with the two triangles containing $\bar{x}$. To simplify the notation let $y = \bar{x} - \lfloor \bar{x} \rfloor$. Then y is contained in a unit square with

corners $[0,0],[1,0],[1,1],[0,1]$. Each of the four triangles is specified by picking three of the four corners. Index the triangles by the central corner vertex. These triangles come in pairs. The first pair each include the diagonal from $[0,0]$ to $[1,1]$ and the second pair include the diagonal from $[1,0]$ to $[0,1]$. If $y_1+y_2<1$, then y is contained in the $[0,0]$ triangle; if $y_1+y_2>1$, then y is contained in the $[1,1]$ triangle. Similarly, if $y_1-y_2>0$, then y is contained in the $[1,0]$ triangle, and if $y_1-y_2<0$, then y is contained in the $[0,1]$ triangle. The barycentric distributions on each triangle are given by

(3.3)
$$\begin{aligned}
&[0,0]\text{ triangle}: p(0,0|y)=1-(y_1+y_2),\quad p(1,0|y)=y_1,\quad p(0,1|y)=y_2,\\
&[1,1]\text{ triangle}: p(1,1|y)=(y_1+y_2)-1,\quad p(1,0|y)=1-y_2,\quad p(0,1|y)=1-y_1,\\
&[1,0]\text{ triangle}: p(0,0|y)=1-y_1,\quad p(1,0|y)=y_1-y_2,\quad p(1,1|y)=y_2,\\
&[0,1]\text{ triangle}: p(0,0|y)=1-y_2,\quad p(0,1|y)=y_2-y_1,\quad p(1,1|y)=y_1.
\end{aligned}$$

The covariance for each triangle is

(3.4)
$$\begin{aligned}
&[0,0]\text{ triangle}: \mathbb{V}(y)=\begin{bmatrix}(1-y_1)y_1 & -y_1y_2\\ -y_1y_2 & (1-y_2)y_2\end{bmatrix},\\
&[1,1]\text{ triangle}: \mathbb{V}(y)=\begin{bmatrix}(1-y_1)y_1 & -(1-y_1)(1-y_2)\\ -(1-y_1)(1-y_2) & (1-y_2)y_2\end{bmatrix},\\
&[1,0]\text{ triangle}: \mathbb{V}(y)=\begin{bmatrix}(1-y_1)y_1 & (1-y_1)y_2\\ (1-y_1)y_2 & (1-y_2)y_2\end{bmatrix},\\
&[0,1]\text{ triangle}: \mathbb{V}(y)=\begin{bmatrix}(1-y_1)y_1 & y_1(1-y_2)\\ y_1(1-y_2) & (1-y_2)y_2\end{bmatrix}.
\end{aligned}$$

Then, for a given $\bar{x}$, the associated solutions to the MVP are given by convex combinations of the two distributions associated with the two triangles containing $\bar{x}$, and the associated covariance is the convex combination of the covariances associated with each triangle. Notice that all four possible covariance functions have the same diagonal. All four have the same diagonal since all four triangles have the same projection onto the coordinate axes.

When projected onto the coordinate axes the integer lattice becomes $\mathbb{Z}$. The distributions provided above all have marginal distributions which equal the solution to the MVP in the one-dimensional case. It follows that these distributions minimize not only the trace of the covariance but also each diagonal entry.

3.3. General integer lattice. This observation extends to integer lattices in arbitrary dimension. Suppose $\Omega=\mathbb{Z}^n$. Then the projection of Ω onto any coordinate axis is $\mathbb{Z}$. Since the total variance is the sum of the variation in each marginal, if there is a joint distribution on Ω whose marginals are the solutions to the MVP when projected onto each coordinate axis, then this joint distribution is necessarily a solution to the MVP on the full space.

LEMMA 3.1 (MVP on lattice restricts to hypercube). *Suppose Ω is an integer lattice or a subset of an integer lattice such that the intersection of the unit hypercube containing $\bar{x}$ and Ω produces a polytope containing $\bar{x}$ in the interior of its convex hull. Then any distribution supported on this intersection is a solution to the MVP using the total variance, and any distribution whose support includes vertices outside this hypercube is not a solution.*

Proof. Any general $\bar{x}$ is contained in some unique unit hypercube whose vertices are in the lattice. The corners of this hypercube are given by rounding each of the coordinates up or down. Since $\bar{x}$ is inside the hypercube, there must exist a convex

combination of the corners which equals $\bar{x}$. The weights of this convex combination are a joint probability distribution on the full space with mean $\bar{x}$. In order for the mean to be $\bar{x}$, the expected value of the jth marginal must equal $\bar{x}_j$ for all $j \in [1, \dots n]$. Moreover, when projected onto the jth coordinate axis the corners of the hypercube map to $\lfloor \bar{x} \rfloor, \lceil \bar{x} \rceil$. Therefore the marginals of this joint distribution are only supported on the two integers above and below $\bar{x}_j$. There is only one marginal distribution with mean $\bar{x}_j$ supported only on $\lfloor \bar{x}_j \rfloor$ and $\lceil \bar{x}_j \rceil$. This marginal distribution is the solution to the MVP on the one-dimensional line, so by restricting the support of the solution to the hypercube we automatically produce a solution which minimizes the variance of every marginal distribution. This minimizes each diagonal entry of the covariance, thus minimizes its trace.

For the converse, suppose that p has support S including nodes outside the unit hypercube. Then at least one projection of the support onto a coordinate axis includes three vertices instead of two. Therefore the corresponding marginal is not a solution to the one-dimensional MVP, so the corresponding diagonal entry of the variance is not minimized. Since there exist solutions to the MVP with every diagonal entry of the covariance minimized, all solutions to the MVP must minimize all of the diagonal entries of the covariance. It follows that, if S includes a vertex outside the unit hypercube containing $\bar{x}$, then p is not a solution to the MVP. ∎

Lemma 3.1 presents the half-space representation of the solution polytope for integer lattices, while Theorem 2.3 presents the vertex representation for generic Ω. The half-space representation is equivalent to the general vertex representation since all vertices of the hypercube are a distance $\sqrt{n}/2$ from its center, so all simplices of the hypercube lie on the same circumsphere and are Delaunay. Therefore, the extreme solutions to the MVP are each associated with a simplex drawn from the unit hypercube containing $\bar{x}$. A review of triangulations of the unit hypercube is provided in [5] and [6].

Unfortunately, the set of all simplices of the unit hypercube is quite large. It is so large that there are many more simplices and triangulations to consider than nodes in the hypercube. For example, in three dimensions there are 74 different triangulations of only 8 vertices [6]. As a consequence, the half-space representation (Lemma 3.1) is much more economical than the vertex representation.

To get a sense for the number of simplices of the unit hypercube we briefly recall some simple methods for generating such simplices. Represent the hypercube as a graph with edges between nodes that only differ by one coordinate. Color the edges according to which coordinate they change. Then any tree that uses exactly one edge of each color corresponds to a unique simplex (see Figure 4). To start counting simplices, consider the two simplest tree topologies: paths and stars.

Consider paths first. Pick a vertex x of the hypercube. The coordinates of x can be represented with ones and zeros. The coordinates of the opposite corner of the hypercube, x^*, are given by swapping all the coordinates of x from one to zero, or zero to one. Pick a permutation σ of n elements. Then the vertices of a simplex $S(\sigma, x)$ are specified by swapping the coordinates of x sequentially, in the order specified by σ, until $x = x^*$. For any x there are $n!$ different σ, and each produces a unique simplex. These simplices form the standard triangulation of the unit hypercube (see the top row of figure Figure 5) [25], [35], [11]. There are 2^{n-1} distinct triangulations of this form since there are 2^{n-1} pairs of opposite corners of the hypercube. Therefore there are at least $2^{n-1}n!$ simplices of the hypercube, and 2^{n-1} simplices containing any $\bar{x}$.

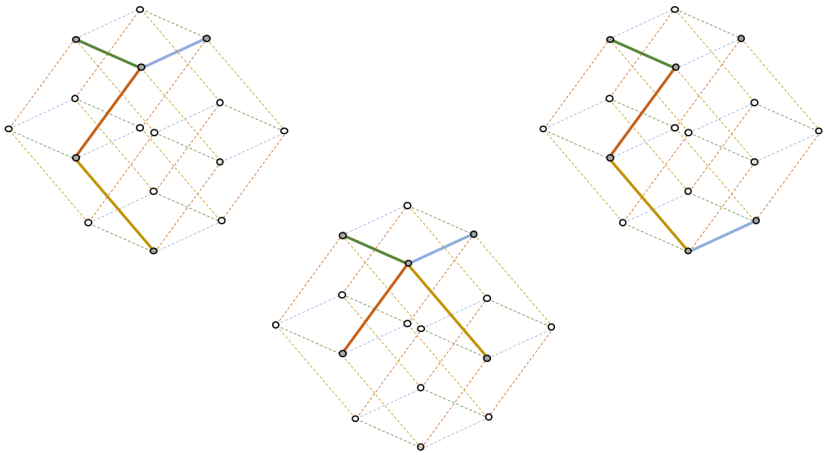

FIG. 4. *Triangulations of the hypercube generated by trees. Three different simplices of the four-dimensional hypercube are shown here. The edges of the hypercube are colored according to which coordinate changes when crossing the edge. These colors correspond to the orientation of the edges. Each tree uses each edge color exactly once. The two trees on the right both generate simplices associated with paths or stars. The tree on the left has a different topology. The number and variety of such topologies increases the higher the dimension.*

Consider stars next. Pick a vertex of the simplex x. Then consider all vertices of the hypercube neighboring x (see the bottom row of Figure 5). There are 2^n simplices in this class, and these simplices are necessarily distinct from the previous set since they never include two opposite corners. Therefore there are at least $2^{n-1}n!$ more simplices of the hypercube than vertices.

Even $2^{n-1}n! + 2^n$ is a dramatic underestimate of the number of simplices. There are many tree topologies other than paths and stars (see Figure 4), and there are also simplices that are not associated with a tree. For example, in three dimensions there are eight simplices in the corners of the cube, each associated with stars. Pick four of these eight simplices, chosen so that they do not overlap, and remove them from the cube. This process leaves another simplex which is not associated with a tree since all of its edges are diagonals (see Figure 5) [5].

Thus, even in fairly low dimensions, the set of extreme solutions to the MVP will likely be intractably large. Then it is natural to seek a particular solution. Here we show that there is a simple distribution which minimizes any unitarily invariant norm of the covariance simultaneously.

LEMMA 3.2 (special solution on integer lattice). *Suppose* $\Omega = \mathbb{Z}^n$ *and* $\bar{x} \in conv(\Omega)$ *is in general position. Then*

$$p(x, \bar{x}) = \prod_{j=1}^{n} \begin{bmatrix} (\lceil \bar{x}_j \rceil - \bar{x}_j) \text{ if } x_j = \lfloor \bar{x}_j \rfloor \\ (\bar{x}_j - \lfloor \bar{x}_j \rfloor) \text{ if } x_j = \lceil \bar{x}_j \rceil \\ 0 \text{ else} \end{bmatrix} \tag{3.5}$$

is the unique distribution which simultaneously minimizes the total variance, two norm of the covariance, Frobenius norm of the covariance, and every singular value of the covariance.

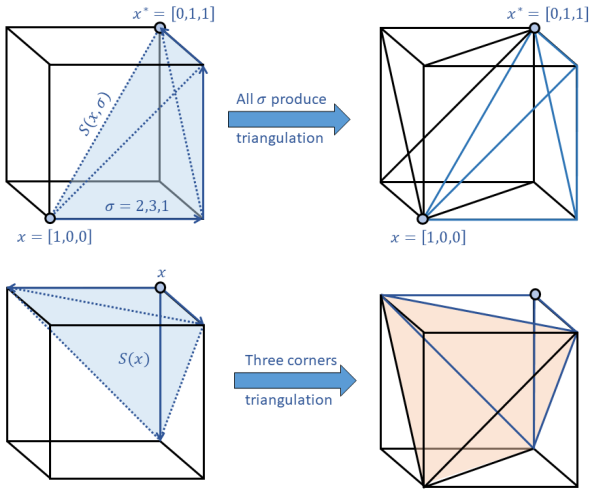

FIG. 5. *Triangulations of the cube. The top left panel shows a simplex generated by moving along a path between opposite corners of the cube specified by a specific permutation. The top right panel shows the triangulation generated by holding the two opposite corners fixed and considering all such paths. The bottom left panel shows the simplex generated by picking a corner and all its neighbors. The bottom right shows a triangulation generated by picking a series of these simplices. Note that this leaves an interior simplex which is not of the first or second kind considered, as all of the edges of this (orange shaded) simplex are diagonals.*

Proof. By construction (3.5) equals the product of the marginal distributions which solve the MVP on each coordinate direction. It follows that this distribution is only supported on the unit hypercube containing $\bar{x}$, so it solves the MVP for the total variance. Since this distribution is a product of marginal distributions, each coordinate of X is independent of the others, so the corresponding covariance is diagonal with diagonal entries equal to $\mathbb{V}_{jj}(\bar{x}) = (\lceil \bar{x} \rceil - \bar{x})(\bar{x} - \lfloor \bar{x} \rfloor)$ (see (3.2)). The square of each off-diagonal entry is minimized since the covariance is diagonal, and each diagonal entry is minimized since this distribution is supported on the unit hypercube. Therefore the Frobenius norm is also minimized. This minimal covariance is unique since the Frobenius norm satisfies the strict triangle inequality.

To see that the l_2 norm is minimized, return to the KKT conditions. These require that, when the largest singular value of the covariance is not repeated, all of the points in the support lie on a pair of parallel hyperplanes normal to the corresponding singular vector, and that no other point from Ω lies in-between the pair of hyperplanes. Since the covariance is diagonal its singular vectors are all coordinate directions, and its singular values are its diagonal entries. For general $\bar{x}$ one of these entries is larger than the rest, so for general $\bar{x}$ the required singular vector is simply the canonical unit vector associated with the coordinate which maximizes $\mathbb{V}_{jj}(\bar{x})$. Then the necessary hyperplanes are the hyperplanes containing the faces of the hypercube perpendicular to this coordinate direction. Since Ω is an integer lattice there is no vertex of Ω between the hyperplanes. Therefore, (3.5) satisfies the KKT conditions for the l_2 norm, so it solves the MVP using the two norm.

To see that each singular value is minimized we extend the argument used for the l_2 norm. If, instead of trying to minimize the first singular value, we attempted to minimize the jth singular value, then we would arrive at the same KKT condition used for the l_2 norm, only with the first singular vector replaced with the jth singular

vector. Then, using the same argument as above, for $\bar{x}$ in general position (3.5) satisfies the KKT conditions for each singular value, so it minimizes each singular value of the covariance. ∎

COROLLARY 3.3. *The special solution* (3.5) *simultaneously minimizes all unitarily invariant matrix norms ρ of the covariance.*

Proof. Any unitarily invariant matrix norm can be expressed as a symmetric gauge function ϕ evaluated on the singular values [17], [37]. A symmetric gauge function is a norm on $\mathbb{R}^n$ which is invariant under permutations and sign changes. If σ, σ' are both vectors in $\mathbb{R}^n$ with entries arranged in nonascending order, then σ' dominates σ if, for each partial sum, $\sum_{i=1}^{j} \sigma_i \leq \sum_{i=1}^{j} \sigma'_i$. Note that if $\sigma_i \leq \sigma'_i$ for all i, then σ' dominates σ. If σ' dominates σ, then $\phi(\sigma) \leq \phi(\sigma')$. Then, since (3.5) minimizes each singular value of the covariance, (3.5) also minimizes any unitarily invariant norm evaluated at the covariance. ∎

The special solution (3.5) is equivalent to the generalized barycentric coordinates proposed by Warren for generic convex polytopes [38]. The generalized barycentric coordinates for an arbitrary convex polytope with vertices S are a set of unique rational functions (one for each vertex of the polytope) that are nonnegative for all $\bar{x} \in \text{conv}(S)$, satisfy the normalization and mean constraints and are of minimal degree. This minimal degree equals the number of facets of the polytope minus the dimension of the space [38]. The distribution (3.5) is a polynomial order n. The n-dimensional hypercube has $2n$ facets, so the barycentric coordinates are rational functions of degree $2n - n = n$. The generalized barycentric coordinates are unique, so this distribution sets the probability at each node of the hypercube to the corresponding barycentric coordinate.

3.4. Integer lattice with a conservation constraint. In some cases subsets of the integer lattice admit unique solutions to the MVP that can be easily identified using Lemma 3.1. For example, suppose $n = 3$, and Ω is the intersection of $\mathbb{Z}^3$ with the conservation constraint $x_1 + x_2 + x_3 = m \in \mathbb{Z}$. This is a natural constraint that arises in a variety of chemical reaction networks (for example, see [29], [30]). Then the intersection of the unit hypercube containing $\bar{x}$ with Ω is a simplex on the subspace $x_1 + x_2 + x_3 = m$. These simplices form a hexagonal triangulation of the subspace. It follows that the MVP admits a unique solution which is given by the tent functions on this hexagonal triangulation. These are

$$p(x|\bar{x}) = \max\{1 - ||x - \bar{x}||_\infty, 0\}. \tag{3.6}$$

Therefore, even though the symmetry of the integer lattice allows for a high-dimensional space of solutions in the general case, it also allows the problem to be reframed in terms of the intersection of Ω with a unit hypercube, which can allow for elegant solutions in some special cases.

3.5. Hexagonal lattice. Consider a hexagonal lattice in $\mathbb{R}^2$ with horizontal rows. Then the vertices are organized in alternating rows. The horizontal coordinate of all vertices in the even rows are integers, and the horizontal coordinate of all vertices in the odd rows are half-integers. The rows are separated by height $\sqrt{3}/2$.

The solution on the hexagonal lattice is unique since it admits a unique Delaunay triangulation into equilateral triangles (see Figure 6). This can be seen by noting that a Delaunay triangulation of the integer lattice in $\mathbb{R}^3$ intersected with a conservation constraint $x_1 + x_2 + x_3 = m$ is a hexagonal lattice on the surface $x_1 + x_2 + x_3 = m$. It

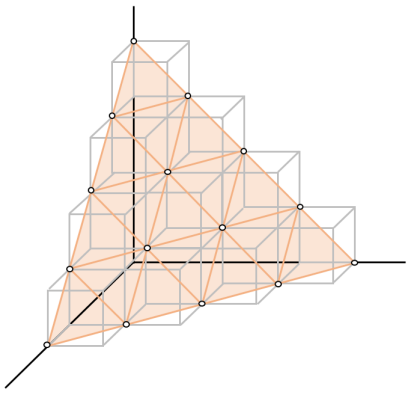

FIG. 6. *The grey cubes represent the integer lattice in three dimensions. The shaded orange triangle is the plane such that* $x_1 + x_2 + x_3 = 4$. *Notice that the intersection of each cube with the plane forms a regular hexagonal lattice.*

follows that the solution to the MVP on a hexagonal lattice can be recovered directly from the solution (3.6) if the lattice is embedded on the appropriate subspace in $\mathbb{R}^3$. This is accomplished by letting $y(x) = [x_1 - x_2/\sqrt{3}, 2x_2/\sqrt{3}, -x_1 - x_2/\sqrt{3}]$, as this maps the corners of the equilateral triangle $[0,0], [1/2, \sqrt{3}/2], [1,0]$ to the vertices $[0,0,0], [0,1,-1], [1,0,-1]$ and conserves $y_1 + y_2 + y_3 = 0$. Then

$$p(x|\bar{x}) = \max\{1 - ||y(x) - y(\bar{x})||_\infty, 0\}. \tag{3.7}$$

3.6. Reaction networks and general lattices. At the cellular scale, systems of chemical reactions are stochastic processes. The number of particles involved in the reactions are small and integer valued, so the noise associated with the timing of individual reaction events cannot be ignored. The stochastic modeling of discrete space reaction networks is an important area of research in molecular biology, because many essential cellular processes depend on systems of chemical reactions [1], [39].

A reaction network is a discrete space continuous-time Markov chain embedded in $\mathbb{R}^n$, where n is the number of chemical species of interest. The state of the system is denoted $X(t) \in \mathbb{Z}^n$. The network is specified by a set of reactions $\mathcal{R}$. Each reaction $r_k \in \mathcal{R}$ is associated with a stoichiometry vector $s_k \in \mathbb{R}^n$ and a reaction propensity $\lambda_k(x,t)$. The stoichiometry describes how the state of the system changes after a reaction. If reaction r_k occurs, then $X(t)$ jumps to $X(t) + s_k$. The reaction propensities are the expected rate of reaction r_k given $X(t) = x$. The state space Ω is all $x \in \mathbb{Z}^n$ that can be reached from an initial state $X(0)$ by a sequence of reactions with nonzero propensities [1].

This state space can be recovered from a linear transformation of a subset of an integer lattice. Let $\mathbb{Z}^{|\mathcal{R}|}$ be an integer lattice, and let $R(t) \in \mathbb{Z}^{|\mathcal{R}|}$ be a count of how many times each reaction has occurred at time t. Then $X(t) = X(0) + SR(t)$, where S is the stoichiometry matrix (matrix whose columns are the stoichiometry vectors). The count $R(t)$ is itself a reaction network where the stoichiometry vector of the kth reaction is the kth column of the identity matrix, and the propensity functions are $\lambda_k(X(0) + Sr, t)$. Let $\Omega_{\mathcal{R}}$ be the subset of the $|\mathcal{R}|$-dimensional integer lattice that can be reached by a sequence of reactions with nonzero propensities. Then $\Omega = S\Omega_{\mathcal{R}}$. Thus the state space of every reaction network is a linear transformation of a subset of an integer lattice.

This observation motivates a study of Ω that are linear transformations of subsets of integer lattices, where the coefficients of the linear transformation are integer valued. These are examples of subsets of general lattices, which are countable sets of points in $\mathbb{R}^n$ spanned by integer combinations of a set of representative vectors [3].

Let $\Omega = S\Omega_{\mathcal{R}}$, where $\Omega_{\mathcal{R}}$ is a subset of $\mathbb{Z}^{|\mathcal{R}|}$. If $S \in \mathbb{R}^{n\times|\mathcal{R}|}$, then $\Omega \subset \text{range}(S) + X(0) \subseteq \mathbb{R}^n$. The linear transformation maps each unit cube in the reaction space to a parallelepiped on the affine subspace $\text{range}(S) + X(0)$.

The state space Ω is a subset of the lattice given by S acting on $\mathbb{Z}^{|\mathcal{R}|}$. Delaunay triangulations of lattices are well-studied; for example, see [8] or [3]. In some cases the Delaunay triangulation of the lattice can be recovered directly from a triangulation of $\mathbb{Z}^{|\mathcal{R}|}$. For example, if the linear transformation is a distortion and the distortion is sufficiently small, then the standard triangulation of the unit cube maps to a Delaunay triangulation of the lattice [8]. More generally, the translation invariance of the lattice implies that if a triangulation of a single parallelepiped can be found whose simplices satisfy the circumsphere condition on the lattice, then a Delaunay triangulation of the entire lattice is given by triangulating each parallelepiped equivalently. This means that a Delaunay triangulation of the lattice can be found by computing the triangulation of a single parallelepid.

Let $\mathcal{T}$ be a Delaunay triangulation of the lattice. If T is a simplex in $\mathcal{T}$, then the interior of the circumsphere of T does not contain any vertices from the lattice, so it must not contain any vertices of Ω. Therefore, if the vertices of T are in Ω, then T is also a Delaunay simplex of Ω. Let P be a parallelepiped that is the image of a unit hypercube in $\mathbb{Z}^{|\mathcal{R}|}$ under the transformation S. If all of the vertices of P are in Ω, then each simplex in the triangulation of P given by $\mathcal{T}$ is a Delaunay simplex. This gives an automatic method for triangulating most of Ω directly from a triangulation of a single parallelepiped in the lattice. This triangulation can then be used to solve the MVP. Parallelepipeds on the boundary must be treated separately, as the triangulation of the lattice may include vertices of the parallelepiped that are not included in Ω.

Alternatively, since the state $X(t)$ can be recovered directly from the reaction counts $R(t)$, the process $X(t)$ can be simulated by simulating $R(t)$ instead. Then it is important to understand the moment structure of $R(t)$, not $X(t)$. The state space of interest in this case is $\Omega_{\mathcal{R}}$ which is always a subset of the integer lattice. Then the special solution (3.5) solves the MVP on each unit hypercube in $\Omega_{\mathcal{R}}$. This approach is attractive since the distribution can be computed analytically from the expected reaction counts $\bar{r}$, and it simultaneously minimizes all of the singular values of the covariance in $R(t)$. Moreover, it can be mapped directly to a distribution for X by multiplying by S. Unlike direct application of the MVP to Ω, which only accounts for the location of the vertices of the reaction network, this approach accounts for the edges of the reaction network. It treats vertices of Ω as close if they are connected by a single reaction, rather than if they are close using the Euclidean norm, so it accounts for the topology of the reaction network.

This solution is also attractive since it can be adopted without picking an arbitrary answer to the question, Which norm is most appropriate when applying the MVP to reaction lattices? Moment closure simulation generally involves a coupled set of update equations, in which the dynamics of the lower order moments are coupled to the higher order moments. It would be natural to try to minimize the variance in order to minimize this coupling. Then the sense in which the variance should be minimized is informed by the coupling terms. For first-order reaction networks the coupling terms have the generic form $\mathbb{V}b$ where the vector b depends on the reaction propensities, stoichiometries, and which reaction occurs. Then we might seek to minimize the

maximum of $||\mathbb{V}b||_p$ over possible b and consequently minimize the induced l_p norm of V. Alternatively, we may wish to minimize the expectation of $||\mathbb{V}b||$ over the possible values of b, which would require minimizing a different norm of the covariance. The question of which norm to minimize is avoided if we use the special solution (3.5) to minimize the variance in the reaction history. The special solution minimizes every singular value of the covariance, so, by Corollary 3.3, it simultaneously minimizes every unitarily invariant norm of the covariance.

4. Discussion. This work allows for the efficient computation of lower bounds on the covariance given a mean state and discrete support. We plan to use the methods described above to compute lower bounds on the covariance of the state of a reaction network given its expected state. These bounds will act as guard-rails in approximate simulation algorithms based on moment closure.

The MVP could also be approached with classical convex optimization methods. The solution presented here for the total variance case has a number of advantages over these techniques. First, once the triangulation is determined, the tent function based solution provides an explicit analytic solution, while an iterative method will only provide a numerical approximation. Second, the tent function solution expresses the solution as a function of $\bar{x}$, so it can be applied for any mean. In contrast, an iterative method would only provide a solution for a particular choice of $\bar{x}$. This offers a more limited understanding of any particular solution since it offers no insight into the sensitivity of the solution to the position of the mean. It is also more expensive if the MVP is meant to be used as a step in a simulation algorithm in which $\bar{x}$ changes over time. Third, the tent function based solution offers a deeper geometric understanding of the problem and solution space. By considering all possible triangulations, it is possible to express all possible solutions to the MVP using tent functions. In contrast, an iterative method would only return a single solution.

That said, convex optimization techniques could play an important role in solving the MVP if they prove to be faster than the tent function method. The efficiency of the tent function solution depends on the efficiency of computing a triangulation of a sufficiently large neighborhood of $\bar{x}$. The cost of finding such a triangulation depends largely on the effectiveness of heuristics for choosing the sequence of vertices added to the neighborhood. Similarly, iterative techniques can be made much faster by constraining the solution to a plausible support, either at initialization or over a sequence of neighborhoods. This can greatly reduce the cost of each step by limiting the number of variables optimized. Thus the practical efficiency of both methods depends largely on heuristics for picking an appropriate sequence of neighborhoods. These heuristics can be guided by Theorem 2.2. A study of the time complexity of each method, and neighborhood heuristics, is saved for future work.

This work could also be extended by considering the solutions under other norms. For example, if $\mathbb{E}[(X-\bar{x})^T W(X-\bar{x})]$ is minimized for some positive semidefinite weight matrix W with symmetric decomposition $W = R^T R$, then, under the change of coordinates $y = Rx$, minimizing $\mathbb{E}[(X-\bar{x})^T W(X-\bar{x})]$ is equivalent to minimizing the trace of the covariance of Y. Therefore the solution to the MVP on Ω using the weighted norm can be recovered directly from the solutions to the MVP on $R\Omega$ when using the trace norm. An alternative family of relevant matrix norms are the Schatten p-norms. These are the l_p norm of the singular values of the covariance. All of the Schatten p-norms have sensitivity matrices $A(p)$ that are positive semidefinite, so all satisfy the conditions of Theorem 2.2. The Schatten infinity norm is the induced

two norm of the covariance, the Schatten two norm is the Frobenius norm, and the Schatten one norm is the trace of the covariance since the covariance is positive semidefinite. It would be interesting to study how solutions using the Schatten p-norms depend on p. Classical convex optimization techniques will likely be required to solve the MVP for generic norms. Future work will seek efficient solution algorithms for the general norms as well as theoretical guarantees on the support of solutions.

Appendix A. Motivating moment closure example.

A reaction network is a discrete-space continuous-time stochastic process $X(t)$, which is defined by a set of reactions $\mathcal{R}$, a propensity for each reaction $\lambda_k(x,t)$, and a stoichiometry s_k for each reaction. The propensity function, $\lambda_k(x,t)$, is the expected rate at which reaction $r_k \in \mathcal{R}$ occurs given state $X(t) = x$. The precise reaction times are random and exponentially distributed.

Reaction networks are widely used to model systems of chemical reactions when the number of particles involved is small and to model small populations [12]. Many reaction networks are studied via Monte Carlo simulation. Exact simulation methods, such as those in [14] and [13], generally require simulation of each individual reaction event. Simulating every reaction is inefficient if the phenomena of interest evolves much more slowly than the rate of the fastest occurring reactions.

Approximation algorithms sacrifice accuracy for efficiency. There are a plethora of available approximation schemes based on different assumptions. Some methods simplify the model by lumping states or trimming reactions [10]. Others use time scale separation to separate fast evolving quantities from slow evolving quantities [28]. Alternatively, if the number of particles is sufficiently large, then diffusion approximations, in which the discrete-space process is replaced with a Langevin stochastic differential equation (SDE), are commonly employed.

Stochastic shielding is an approximation method introduced by Schmandt and Galán [29]. The stochastic shielding approximation simplifies a simulation by reducing the number of sources of variation in the model. Only the reactions which contribute the most uncertainty to the relevant (measured) components of $X(t)$ are simulated stochastically. The rest are approximated deterministically.

The stochastic shielding approximation has been shown to be both fast and accurate for appropriate models. A rigorous mathematical foundation has been developed for the approximation when applied to Langevin SDEs [30], [31]. An equivalent foundation does not exist for reaction network models without an additional discrete time tau-leaping approximation as in [29]. Naive generalization of the method can produce impossible trajectories with negative particle counts. More careful analysis reveals that, in order to properly extend stochastic shielding to discrete space, the stochastic shielding method should be viewed as a moment closure approximation on a hidden conditional distribution as in [40]. The low order moments of this conditional distribution are coupled to the higher order moments. The naive extension fails because it neglects these couplings, which is equivalent to assuming that the conditional distribution is always a delta distribution. If the domain Ω of $X(t)$ is discrete, then, for general $\bar{x}$, the corresponding distribution cannot be a delta distribution. The naive stochastic shielding method fails because it chooses higher order moments which are impossible given the geometry of the domain. In one-dimensional models it can be shown that, if the variance is set to its minimum possible value (3.2), then the simulated trajectories remain nonnegative. If the variance is any smaller, then simulated trajectories may become negative. Thus, the errors in the naive stochastic shielding algorithm can be traced directly to underestimation of the variance.

Appendix B. Differentiating matrix norms.

B.1. Trace and total variance. Let $\rho(V) = \text{trace}(V)$. Then $\rho(V) = \sum_j v_{jj}$ so $\partial_{v_{ij}}\rho(V) = 0$ if $i \neq j$ and equals 1 if $i = j$. Therefore $A(V)$ equals the identity I.

B.2. Frobenius norm. Minimizing the Frobenius norm is equivalent to minimizing one half the Frobenius norm squared. This sets the cost function to $\rho(V) = \frac{1}{2}\sum_{ij} v_{ij}^2$. Clearly, $\partial_{v_{ij}}\rho(V) = v_{ij}$, so $A(V) = V$.

B.3. Two norm. Minimizing the two norm is the same as minimizing one half the two norm squared. This sets the cost function to $\rho(V) = \frac{1}{2}||V||_2^2$. Then, by the chain rule, $\partial_{v_{ij}}\rho(V) = ||V||_2\partial_{v_{ij}}||V||_2$ provided the latter derivative exists.

Let $\sigma_1(V)$ be the largest singular value of V so that $||V||_2 = \sigma_1(V)$. Then, since V is symmetric positive semidefinite it has SVD $V = U(V)\Sigma(V)U(V)^T$ which coincides with its eigenvalue decomposition since $U(V)$ is orthonormal. Therefore,

$$Vu_1(V) = \sigma_1(V)u_1(V). \tag{B.1}$$

Any eigenvalue and corresponding eigenvector of a real symmetric matrix is continuously differentiable if the eigenvalue is simple (not a repeated eigenvalue) [24]. In this case, we can differentiate $\sigma_1(V)$ and $u_1(V)$ in the entries of V.

Differentiating with respect to v_{ij},

$$[\partial_{v_{ij}}V]u_1(V) + V\partial_{v_{ij}}u_1(V) = [\partial_{v_{ij}}\sigma_1(V)]u_1(V) + \sigma_1(V)\partial_{v_{ij}}u_1(V). \tag{B.2}$$

Take an inner product on both sides with $u_1(V)$. Then, since V is symmetric, $u_1(V)^TV = \sigma_1(V)u_1(V)^T$ leaving,

$$\begin{aligned} &u_1(V)^T[\partial_{v_{ij}}V]u_1(V) + \sigma_1(V)u_1(V)^T\partial_{v_{ij}}u_1(V) \\ &\quad = [\partial_{v_{ij}}\sigma_1(V)]u_1(V)^Tu_1(V) + \sigma_1(V)u_1(V)^T\partial_{v_{ij}}u_1(V). \end{aligned} \tag{B.3}$$

Canceling the repeated term and noting that $u_1(V)^Tu_1(V) = ||u_1(V)||_2^2 = 1$ leaves

$$u_1(V)^T[\partial_{v_{ij}}V]u_1(V) = [\partial_{v_{ij}}\sigma_1(V)]. \tag{B.4}$$

Now, $\partial_{v_{ij}}V = e_ie_j^T$ where e_k is the kth canonical basis vector, so the left-hand side equals $u_1(V)_iu_1(V)_j$. Therefore $\partial_{v_{ij}}||V||_2 = \partial_{v_{ij}}\sigma_1(V) = u_1(V)_iu_1(V)_j$ which matches the result in [24]: $A(V)_{ij} = \partial_{v_{ij}}\rho(V) = \sigma_1(V)u_1(V)_iu_1(V)_j$. Therefore,

$$A(V) = \sigma_1(V)u_1(V)u_1(V)^T = V^{(1)}, \tag{B.5}$$

where $V^{(1)}$ is the rank one approximation to V.

Appendix C. Uniqueness of barycentric coordinates.

Suppose $S \subset \Omega$ such that $\bar{x} \in \text{conv}(S)$. Suppose $|S| = m$. Index the vertices of S, $x(0), x(2), \ldots x(m-1)$. Represent the distribution $p(x)$ with a vector such that $p_j = p(x(j))$. Then the mean and normalization constraints require that

$$M(S)p = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1(0) & x_1(1) & \ldots & x_1(m-1) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(0) & x_n(1) & \ldots & x_n(m-1) \end{bmatrix} p = \begin{bmatrix} 1 \\ \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{bmatrix}. \tag{C.1}$$

Now suppose S is a simplex. Then $|S| = m = n + 1$ and the n edge vectors $x(j) - x(0)$ are linearly independent. It follows that $M(S)$ is full rank since subtracting the first column from every other column produces the matrix

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ x_1(0) & x_1(1) - x_1(0) & \dots & x_1(m-1) - x_1(0) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(0) & x_n(1) - x_n(0) & \dots & x_n(m-1) - x_n(0) \end{bmatrix}. \tag{C.2}$$

Since $M(S)$ is square and full rank it is invertible. Then, if $\bar{x} \in \text{conv}(S)$, there is a unique distribution $p(x|\bar{x})$ given by $p = M(S)^{-1}[1; \bar{x}]$ which is affine in $\bar{x}$. Therefore $p(x|\bar{x})$ is unique if S is a simplex.

We now show that, if S is not a simplex, and $\bar{x} \in \text{conv}(S)$, then the distribution p is not unique. Let p denote a distribution with mean $\bar{x}$, support S, such that $|S| > n + 1$, and the polytope formed by S is n dimensional. Then p satisfies the linear equation (C.1). However, if $|S| > n + 1$, then M has more columns than rows. If the polytope formed by the vertices of S is n dimensional, then there is a subset of $n + 1$ of the vertices of S that form a simplex. Order the vertices of S so these come first. Then the first $n+1$ columns of $M(S)$ span $\mathbb{R}^{n+1}$, so the remaining $|S| - (n+1)$ columns are not independent of the original columns and the matrix has a nontrivial nullspace.

Let z be a vector in this nullspace. Then $p + \epsilon z$ still satisfies (C.1). Moreover, since it was assumed that p had support S, every entry of p is positive and nonzero. Hence there must be an ϵ small enough that every entry of $p + \epsilon z$ is still positive and nonzero. Therefore $p + \epsilon z$ is a different distribution with the same support and same mean as p. It follows that the barycentric coordinates are not unique if $|S| > n + 1$. Therefore the only supports which form an n-dimensional polytope, but for which the barycentric coordinates are unique, are simplices.

Acknowledgments. I thank Gilbert Strang (MIT) for his encouragement to pursue this project and his patient and careful editing. I also thank Peter Thomas (Case Western Reserve University) for his advice when discussing this project and his help writing the manuscript. I thank Elisabeth Werner and Elizabeth Meckes (Case Western Reserve University) for their advice and assistance finding relevant references.

REFERENCES

[1] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems*, vol. 1, Springer, Cham, 2015.
[2] B. C. Arnold, *Some examples of minimum variance unbiased estimates*, Amer. Statist., 26 (1972), pp. 34–36.
[3] A. Choudhary and A. Ghosh, *Delaunay simplices in diagonally distorted lattices*, Comput. Geom., 81 (2019), pp. 33–44.
[4] P. Cignoni, C. Montani, and R. Scopigno, *Dewall: A fast divide and conquer Delaunay triangulation algorithm in E^d*, Comput.-Aided Des., 30 (1998), pp. 333–341.
[5] C. Dang, *Triangulations and Simplicial Methods*, Springer Science & Business Media, New York, 2012.
[6] J. A. De Loera, J. Rambau, and F. Santos, *Triangulations Structures for Algorithms and Applications*, Springer, Cham, 2010.
[7] B. Delaunay, *Sur la sphere vide*, Izv. Akad. Nauk SSSR, 6 (1934), pp. 793–800.
[8] H. Edelsbrunner and M. Kerber, *Dual complexes of cubical subdivisions of $\mathbb{R}^n$*, Discrete Comput. Geom., 47 (2012), pp. 393–414.

[9] H. Edelsbrunners, *Geometric algorithms*, in Handbook of Convex Geometry, Elsevier, New York, 1993, pp. 699–735.
[10] D. R. Fredkin and J. A. Rice, *On aggregated Markov processes*, J. Appl. Probab., 23 (1986), pp. 208–214.
[11] H. Freudenthal, *Simplizialzerlegungen von beschrankter flachheit*, Ann. Math., 43 (1942), pp. 580–582.
[12] C. Gadgil, C. H. Lee, and H. G. Othmer, *A stochastic analysis of first-order reaction networks*, Bull. Math. Biol., 67 (2005), pp. 901–946.
[13] M. A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A, 104 (2000), pp. 1876–1889.
[14] D. T. Gillespie, *Stochastic simulation of chemical kinetics*, Annu. Rev. Phys. Chem., 58 (2007), pp. 35–55.
[15] J. Hetthéssy and L. Keviczky, *Minimum variance control: A review and outlook*, Period. Polytech. Electr. Engrg., 21 (1977), pp. 31–45.
[16] Ø. Hjelle and M. Dæhlen, *Triangulations and Applications*, Springer Science & Business Media, New York, 2006.
[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK, Cambridge University Press, 1985.
[18] W. Karush, *Minima of functions of several variables with inequalities as side conditions*, in Traces and Emergence of Nonlinear Programming, Springer, Cham, 2014, pp. 217–245.
[19] C. Kuehn, *Moment closure-a brief review*, in Control of Self-Organizing Nonlinear Systems, Springer, Cham, 2016, pp. 253–271.
[20] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., 1951.
[21] U. Kwon, M. Naghnaeian, and D. Del Vecchio, *Approximation of the chemical master equation using conditional moment closure and time-scale separation*, in Proceedings of the 2019 American Control Conference (ACC), IEEE, 2019, pp. 585–592.
[22] G. Leach, *Improving worst-case optimal Delaunay triangulation algorithms*, in Proceedings of the 4th Canadian Conference on Computational Geometry, vol. 2, Citeseer, 1992, p. 15.
[23] S. Lo, 3*D Delaunay triangulation of* 1 *billion points on a pc*, Finite Elem. Anal. Des., 102 (2015), pp. 65–73.
[24] J. R. Magnus, *On differentiating eigenvalues and eigenvectors*, Econom. Theory, 1 (1985), pp. 179–191.
[25] P. S. Mara, *Triangulations for the cube*, J. Combin. Theory, Ser. A, 20 (1976), pp. 170–177.
[26] M. Naghnaeian and D. Del Vecchio, *Robust moment closure method for the chemical master equation*, in Proceedings of the 2017 IEEE Conference on Control Technology and Applications (CCTA), IEEE, 2017, pp. 967–972.
[27] L. Perumal, *New approaches for Delaunay triangulation and optimisation*, Heliyon, 5 (2019), e02319.
[28] C. V. Rao and A. P. Arkin, *Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm*, J. Chem. Phys., 118 (2003), pp. 4999–5010.
[29] N. T. Schmandt and R. F. Galán, *Stochastic-shielding approximation of Markov chains and its application to efficiently simulate random ion-channel gating*, Phys. Rev. Lett., 109 (2012), 118101.
[30] D. R. Schmidt, R. F. Galán, and P. J. Thomas, *Stochastic shielding and edge importance for Markov chains with timescale separation*, PLoS Comput. Biol., 14 (2018), p. e1006206.
[31] D. R. Schmidt and P. J. Thomas, *Measuring edge importance: A quantitative analysis of the stochastic shielding approximation for random processes on graphs*, J. Math. Neurosci., 4 (2014), p. 6.
[32] J. Shewchuk, T. K. Dey, and S.-W. Cheng, *Delaunay Mesh Generation*, Chapman and Hall/CRC, Boca Raton, FL, 2016.
[33] M. Soltani, C. A. Vargas-Garcia, and A. Singh, *Conditional moment closure schemes for studying stochastic dynamics of genetic circuits*, IEEE Trans. Biomed. Circuits Syst., 9 (2015), pp. 518–526.
[34] P. Su and R. L. S. Drysdale, *A comparison of sequential Delaunay triangulation algorithms*, Comput. Geom., 7 (1997), pp. 361–385.
[35] M. J. Todd, *Improving the convergence of fixed-point algorithms*, in Complementarity and Fixed Point Problems, Springer, Cham, 1978, pp. 151–169.
[36] C. D. Toth, J. O'Rourke, and J. E. Goodman, *Handbook of Discrete and Computational Geometry*, CRC Press, Boca Raton, FL, 2017.
[37] J. Von Neumann, *Some matrix-inequalities and metrization of matric space*, Tomsk Univ. Rev., 1 (1937), pp. 286–300.

[38] J. Warren, *Barycentric coordinates for convex polytopes*, Adv. Comput. Math., 6 (1996), pp. 97–108.
[39] D. J. Wilkinson, *Stochastic Modelling for systems Biology*, CRC Press, Boca Raton, FL, 2011.
[40] C. Zechner and H. Koeppl, *Uncoupled analysis of stochastic reaction networks in fluctuating environments*, PLoS Comput. Biol., 10 (2014), e1003942.