The Fastest Mixing Markov Process on a Graph and a Connection to a Maximum Variance Unfolding Problem*

Jun Sun[†]
Stephen Boyd[†]
Lin Xiao[‡]
Persi Diaconis[§]

Abstract. We consider a Markov process on a connected graph, with edges labeled with transition rates between the adjacent vertices. The distribution of the Markov process converges to the uniform distribution at a rate determined by the second smallest eigenvalue λ_2 of the Laplacian of the weighted graph. In this paper we consider the problem of assigning transition rates to the edges so as to maximize λ_2 subject to a linear constraint on the rates. This is the problem of finding the fastest mixing Markov process (FMMP) on the graph. We show that the FMMP problem is a convex optimization problem, which can in turn be expressed as a semidefinite program, and therefore effectively solved numerically. We formulate a dual of the FMMP problem and show that it has a natural geometric interpretation as a maximum variance unfolding (MVU) problem, i.e., the problem of choosing a set of points to be as far apart as possible, measured by their variance, while respecting local distance constraints. This MVU problem is closely related to a problem recently proposed by Weinberger and Saul as a method for "unfolding" high-dimensional data that lies on a low-dimensional manifold. The duality between the FMMP and MVU problems sheds light on both problems, and allows us to characterize and, in some cases, find optimal solutions.

Key words. Markov process, fast mixing, second smallest eigenvalue, semidefinite programming, dimensionality reduction

AMS subject classifications. 60J25, 65F15, 68Q32, 90C22, 90C46

DOI. 10.1137/S0036144504443821

1. The Problem.

The fastest mixing Markov chain problem was proposed and studied by Boyd, Diaconis, and Xiao in [4]. In that problem the mixing rate of a discrete-time Markov chain on a graph was optimized over the set of transition probabilities on the edges of a given graph. In this paper, we discuss its continuous-time counterpart.

1.1. The Laplacian and Mixing Rate. Let $G = (V, E)$ be an undirected connected graph with n vertices, labeled $1, \ldots, n$, and m edges. We will use two schemes to label the edges of the graph. If vertices i and j are adjacent $(i \sim j)$, we denote the associated edge by an unordered pair $\{i, j\}$. On other occasions we will label the edges with a single index that runs from 1 to m. We consider a symmetric Markov process on the graph. The state space is V, and the edges E represent the allowed transitions. Each edge $\{i, j\}$ is labeled with the transition rate $w_{ij} \geq 0$ between the associated vertices. We will also give the set of edge transition rates as a vector $w = (w_1, \ldots, w_m) \in \mathbf{R}^m$, where the subscript refers to the edges described using a single index.

Let $\pi(t) \in \mathbf{R}^n$ denote the distribution of the state at time t. Its evolution is given by the heat equation

$$\frac{d\pi(t)}{dt} = -L\pi(t),$$

where $L \in \mathbf{R}^{n \times n}$ is the (weighted) *Laplacian*, defined as

$$L_{ij} = \begin{cases} -w_{ij}, & i \sim j, \ i \neq j, \\ 0, & i \nsim j, \ i \neq j, \\ \sum_{j \sim i} w_{ij}, & i = j. \end{cases}$$

The solution is given by the action of the semigroup e^{-tL} on the initial condition, i.e.,

$$\pi(t) = e^{-tL}\pi(0).$$

The Laplacian L is symmetric and positive semidefinite (PSD). Its smallest eigenvalue is 0, with associated eigenvector $\mathbf{1} = (1, \ldots, 1)$. We order the eigenvalues of L in increasing order:

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

If the subgraph consisting of edges with positive rates is connected, then $\lambda_2 > 0$ (i.e., λ_1 is isolated). Conversely, if $\lambda_2 > 0$, the subgraph of edges with positive rates is connected. For a survey of the properties of the Laplacian, see, e.g., [19]. In the context of the Laplacian, we refer to the rates w_{ij} as *weights*, since they can be thought of as weights on the edges of the graph.

From $L\mathbf{1} = 0$ it follows that the uniform distribution $\mathbf{1}/n$ is an equilibrium distribution of the Markov process. If $\lambda_2 > 0$, it is the unique equilibrium distribution, and $\pi(t) \to \mathbf{1}/n$ as $t \to \infty$ for any initial distribution. Moreover, the rate of convergence of the distribution $\pi(t)$ to uniform is determined by λ_2; for example, we have

$$\sup_{\pi(0)} \|\pi(t) - (1/n)\mathbf{1}\|_{\mathrm{tv}} \leq (1/2)n^{1/2}e^{-\lambda_2 t},$$

where $\| \cdot \|_{\mathrm{tv}}$ is the *total variation distance* between two distributions (see, e.g., [8]). (The total variation distance is the maximum difference in probability assigned by the two distributions, over any subset of vertices.) Thus, the larger the second Laplacian eigenvalue λ_2, the faster the Markov process mixes.

1.2. The Fastest Mixing Markov Process Problem. We now turn to the question of how to choose the rates w so that the Markov process mixes as quickly as possible. Since λ_2 is a positive homogeneous function of w, it can be made as large as we like by scaling w by a large positive factor. To make the problem of maximizing λ_2 sensible,

we must somehow limit the rates. Perhaps the simplest way to do this is to impose a limit on a positive weighted sum of the rates,

$$\sum_{\{i,j\}\in E} d_{ij}^2 w_{ij} \leq 1,$$

where $d_{ij}^2 > 0$ represents a relative cost on edge $\{i,j\}$.

We can now state the problem. We wish to assign rates to the edges, subject to this constraint, to get the fastest mixing Markov process (FMMP) on the given graph. This can be posed as the following constrained optimization problem:

$$
\begin{aligned}
\text{maximize} \quad & \lambda_2(w) \\
\text{subject to} \quad & \sum d_{ij}^2 w_{ij} \leq 1, \\
& w \geq 0.
\end{aligned}
$$

(1)

Here the optimization variable is $w \in \mathbf{R}^m$, and the problem data are the graph and the weights d_{ij}; the notation $w \geq 0$ is meant elementwise. We call (1) the FMMP problem.

We make several simple observations about the FMMP problem. First, it always has a solution, since the objective is continuous and the constraint set is compact. Moreover, the inequality constraint $\sum d_{ij}^2 w_{ij} \leq 1$ always holds with equality at the solution, since λ_2 is positive homogeneous. In particular, we can replace the inequality by equality. The optimal value of the problem is always positive, since we can choose all rates to be small and positive, which results in $\lambda_2 > 0$ (since the graph is connected). Finally, we note that if w^* is optimal, the subgraph associated with $w_{ij}^* > 0$ must be connected.

The *absolute algebraic connectivity problem*, described by Fiedler [9, 10], is a special case of the FMMP problem, with $d_{ij} = 1/\sqrt{m}$ for all $\{i,j\} \in E$. The absolute algebraic connectivity problem is a very interesting topic in algebraic graph theory; the fact that it is a special case of the FMMP problem also gives it several physical interpretations (discussed in the next section).

Several other closely related problems have been explored by the authors, including the discrete-time counterpart [4, 3], distributed algorithms for resource allocation [32], and distributed algorithms for averaging [31]. We will also see, in section 4.7, a somewhat surprising connection to recent work in machine learning in the area of manifold unfolding [28, 29].

2. Interpretations. In this section we give some simple physical interpretations of the FMMP problem (1). None of this material is used in what follows (except in interpretations of subsequent results), so the reader anxious to get to the main results can skip this section. We note that similar extremal eigenvalue problems have been studied for continuous domains, with the Laplacian matrix replaced by a Laplacian operator; see, e.g., [6, 7].

2.1. Grounded Capacitor RC Circuit. Consider a connected grounded capacitor resistor-capacitor (RC) circuit, as shown in Figure 1. Each node has a grounded unit value capacitor, and nodes i and j are connected by a conductance g_{ij} (which is zero if nodes i and j are not directly connected). Let $q(t)$ denote the vector of charge distribution on the nodes. Then
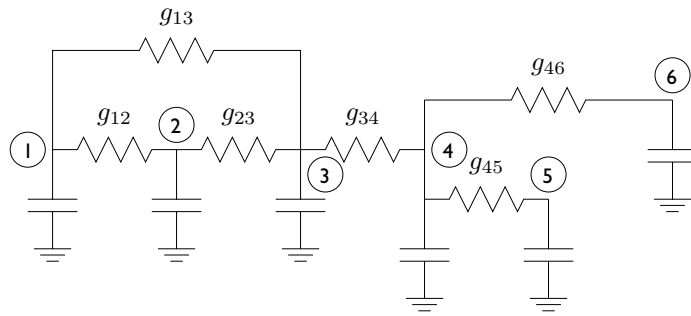
$$\frac{dq}{dt} = -Lq(t),$$

Fig. 1 *Example of a grounded capacitor RC circuit.*

where L is the Laplacian with g_{ij} as the edge weights. When λ_2, the second eigenvalue of the Laplacian, is positive, the charge distribution converges to uniform, i.e., $q_i(t) \to \mathbf{1}^T q(0)/n$. The second Laplacian eigenvalue λ_2 gives the rate at which the charge is equilibrated. In this context, the FMMP problem is to choose the conductances in the circuit, subject to a limit on the weighted total conductance, to make the circuit equilibrate charge as fast as possible.

We can give a simple physical interpretation of the weighted total conductance constraint. Assume the conductors are made of a material with unit resistivity, and let d_{ij} and a_{ij} be the length and cross-sectional area of the conductor between nodes i and j, respectively. The conductance between the nodes is then

$$g_{ij} = \frac{a_{ij}}{d_{ij}},$$

and the total volume of the conductors is

$$\sum d_{ij} a_{ij} = \sum d_{ij}^2 g_{ij}.$$

Thus, we can interpret the FMMP problem as the problem of choosing the conductor cross-sectional areas, subject to a unit total volume, so as to make the circuit equilibrate charge as quickly as possible.

For a related application of semidefinite programming to optimizing RC circuits, see [27].

2.2. Isolated Thermal System. We can give a similar interpretation for a thermal system. We consider a thermal system consisting of n unit thermal masses, connected by some thermal conductances but otherwise isolated. The total heat is constant and asymptotically distributes itself evenly across the masses, with a rate that depends on the second Laplacian eigenvalue λ_2. We can imagine that the conductances are, say, rods with unit thermal conductivity, given lengths d_{ij} and cross-sectional areas a_{ij}. The problem is to choose the cross-sectional areas, subject to a total volume constraint, so that the system equilibrates heat as rapidly as possible.

2.3. Mass-Spring System. Consider a mass-spring system such as the one shown in Figure 2. The unit masses move horizontally without friction, with some pairs connected by springs with spring constant k_{ij} between adjacent nodes i and j. (The spring constant is the ratio of force to displacement.) The vector of displacements $x(t)$ satisfies the wave equation
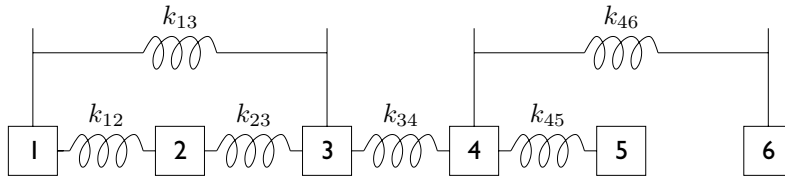
$$\frac{d^2 x}{dt^2} = -Lx(t),$$

Fig. 2 *Example of a mass-spring system.*

where L is the Laplacian with k_{ij} as the edge weights. The solution has the form

$$x_i(t) = a_i + bt + \sum_{j=2}^{n} c_{ij} \cos(\lambda_j^{1/2} t + \phi_{ij}),$$

where a_i, b, c_{ij}, and ϕ_{ij} depend on the initial displacements and velocities, and $\lambda_2, \ldots, \lambda_n$ are the Laplacian eigenvalues. Assuming $\lambda_2 > 0$, the motion of each mass consists of a (common) constant velocity component $a_i + bt$ and a sum of sinusoidal oscillations with frequencies given by the square roots of the Laplacian eigenvalues, which are the *natural frequencies* of the system. The slowest or *fundamental natural frequency* is $\lambda_2^{1/2}$, and the associated *fundamental period* is $2\pi\lambda_2^{-1/2}$. In this example, then, the second Laplacian eigenvalue is related to the fundamental period of the system; the larger λ_2 is, the shorter the fundamental period. The FMMP problem is to choose the stiffness of each spring, subject to a total weighted stiffness constraint, so as to make the fundamental period as fast as possible.

Here too we can relate the optimization variables to cross-sectional areas. Assume that the springs are bars with unit Young's modulus, and let d_{ij} and a_{ij} be the length and cross-sectional area of the bar connecting mass i and mass j. Then its stiffness is

$$k_{ij} = \frac{a_{ij}}{d_{ij}},$$

and the total volume of the bars is

$$\sum d_{ij} a_{ij} = \sum d_{ij}^2 k_{ij}.$$

Thus, the FMMP problem is to allocate a unit volume of material to the bars so as to minimize the fundamental period.

Of course the same type of analysis holds for an inductor-capacitor (LC) circuit.

3. Convexity and Semidefinite Program Formulation.

3.1. Concavity of λ_2. The second Laplacian eigenvalue λ_2 is a concave function of the weights w. This can be seen in several ways; for example, from the variational characterization of eigenvalues [15, section 4.2],

$$\lambda_2(w) = \inf_{\|u\|=1, \ \mathbf{1}^T u = 0} u^T L(w) u = \inf_{\|u\|=1, \ \mathbf{1}^T u = 0} \sum_{\{i,j\} \in E} w_{ij} (u_i - u_j)^2.$$

This shows that $\lambda_2(w)$ is the pointwise infimum of a family of linear functions of w, and therefore is concave (see [5, section 3.2]).

Since the constraints in the FMMP problem are linear, we see that it is a convex optimization problem (since the objective is concave and is to be maximized). General background on eigenvalue optimization can be found in, e.g., [21, 17].

3.2. An Alternate Formulation. Here we give an alternate formulation of the FMMP problem which will be more convenient in what follows. Since the objective λ_2 and the constraint function $\sum d_{ij}^2 w_{ij}$ in the FMMP problem are positive homogeneous, we can just as well minimize the constraint function subject to a minimum on λ_2:

$$
\begin{aligned}
\text{minimize} \quad & \sum d_{ij}^2 w_{ij} \\
\text{subject to} \quad & \lambda_2(w) \geq 1, \\
& w \geq 0.
\end{aligned}
\tag{2}
$$

As in the original FMMP problem (1), the solution here will always satisfy the inequality constraint with equality. (The solutions of the two FMMP formulations are the same, up to a positive scaling.) In this alternate formulation, we want to find rates on the edges so as to minimize the total weighted cost, subject to the constraint that the mixing rate (as measured by λ_2) exceeds 1. We will also refer to problem (2) as the FMMP problem. Like the original formulation, it is a convex optimization problem.

3.3. Semidefinite Program (SDP) Formulation. We can transform the FMMP problem (2) into an SDP by observing that

$$
\lambda_2(w) \geq 1 \iff L(w) \succeq I - (1/n)\mathbf{1}\mathbf{1}^T,
$$

where $\succeq$ denotes matrix inequality. Using this, we can express the FMMP problem as

$$
\begin{aligned}
\text{minimize} \quad & \sum d_{ij}^2 w_{ij} \\
\text{subject to} \quad & L(w) \succeq I - (1/n)\mathbf{1}\mathbf{1}^T, \\
& w \geq 0.
\end{aligned}
\tag{3}
$$

This problem has a linear objective, a linear matrix inequality (LMI) constraint, and a set of nonnegativity constraints, and so is an SDP. We refer to (3) as the *primal SDP*. While problem (3) is not in one of the so-called standard forms for an SDP, it is easily transformed to a standard form; see, e.g., [5, 26].

One immediate consequence of the SDP formulation (3) is that we can numerically solve the FMMP problem efficiently, using standard algorithms for an SDP (e.g., [23, 2, 1, 11, 24]; see [13] for a comprehensive list of current SDP software). For m no more than 1000 or so, interior-point methods can be used to solve the FMMP problem in a minute or so (or much less) on a small personal computer. Far larger problems can be solved using subgradient-type methods and by exploiting the problem structure; see, e.g., [5, 4]. More sophisticated methods for solving large scale problems of this type are discussed in [22, 14, 20]. However, we won't pursue numerical methods for the FMMP problem in this paper.

3.4. The Dual Problem. The *dual* of the SDP (3) is (the SDP)

$$
\begin{aligned}
\text{maximize} \quad & \mathbf{Tr}(I - (1/n)\mathbf{1}\mathbf{1}^T)X \\
\text{subject to} \quad & X_{ii} + X_{jj} - X_{ij} - X_{ji} \leq d_{ij}^2, \quad \{i,j\} \in E, \\
& X \succeq 0,
\end{aligned}
\tag{4}
$$

where the symmetric matrix $X \in \mathbf{R}^{n \times n}$ is the optimization variable, and the problem data are the graph and the weights d_{ij}. (See [5] for more on SDP duality.)

Before proceeding, we simplify this problem a bit. If X is feasible, then

$$\tilde{X} = (I - (1/n)\mathbf{1}\mathbf{1}^T)X(I - (1/n)\mathbf{1}\mathbf{1}^T)$$

is also feasible, since $\tilde{X} \succeq 0$,

$$\tilde{X}_{ii} + \tilde{X}_{jj} - \tilde{X}_{ij} - \tilde{X}_{ji} = X_{ii} + X_{jj} - X_{ij} - X_{ji} \leq d_{ij}^2, \quad \{i,j\} \in E.$$

Moreover, $\mathbf{Tr}\,\tilde{X}$ is equal to the objective $\mathbf{Tr}(I - (1/n)\mathbf{1}\mathbf{1}^T)X$:

$$\begin{aligned}
\mathbf{Tr}\,\tilde{X} &= \mathbf{Tr}(I - (1/n)\mathbf{1}\mathbf{1}^T)X(I - (1/n)\mathbf{1}\mathbf{1}^T) \\
&= \mathbf{Tr}(I - (1/n)\mathbf{1}\mathbf{1}^T)^2 X \\
&= \mathbf{Tr}(I - (1/n)\mathbf{1}\mathbf{1}^T)X.
\end{aligned}$$

(Here we use $\mathbf{Tr}(AB) = \mathbf{Tr}(BA)$ and $(I - (1/n)\mathbf{1}\mathbf{1}^T)^2 = (I - (1/n)\mathbf{1}\mathbf{1}^T)$.) Evidently, $\tilde{X}$ always satisfies $\mathbf{1}^T \tilde{X} \mathbf{1} = 0$.

Conversely, if $\tilde{X}$ satisfies

$$\tilde{X}_{ii} + \tilde{X}_{jj} - \tilde{X}_{ij} - \tilde{X}_{ji} \leq d_{ij}^2, \quad \{i,j\} \in E, \qquad \mathbf{1}^T \tilde{X} \mathbf{1} = 0, \qquad \tilde{X} \succeq 0,$$

then it also is feasible for the problem (4), and its objective there is equal to $\mathbf{Tr}\,\tilde{X}$. It follows that we can use $\tilde{X}$ as the variable in the dual, in place of X.

Using $\tilde{X}$ as the variable and adding this constraint (and relabeling $\tilde{X}$ as X) yields the SDP

(5)
$$\begin{aligned}
\text{maximize} \quad & \mathbf{Tr}\,X \\
\text{subject to} \quad & X_{ii} + X_{jj} - X_{ij} - X_{ji} \leq d_{ij}^2, \quad \{i,j\} \in E, \\
& \mathbf{1}^T X \mathbf{1} = 0, \quad X \succeq 0.
\end{aligned}$$

We will refer to this version of the SDP dual as the *dual SDP*. The primal SDP (3) and the dual SDP (5) form a pair of duals, in the sense that the following duality results hold.

- *Weak duality.* For any primal feasible w and any dual feasible X, we have

$$\sum d_{ij}^2 w_{ij} \geq \mathbf{Tr}\,X.$$

Thus, any dual feasible X gives a lower bound on the optimal value of the primal FMMP problem. (Similarly, any primal feasible w gives an upper bound on the optimal value of the dual problem.) We can give a simple direct derivation of this fact: For any primal-dual feasible pair (w, X),

$$\begin{aligned}
\sum d_{ij}^2 w_{ij} - \mathbf{Tr}\,X &= \sum [d_{ij}^2 - (X_{ii} + X_{jj} - X_{ij} - X_{ji})]w_{ij} \\
&\quad + \mathbf{Tr}(L(w) - (I - (1/n)\mathbf{1}\mathbf{1}^T))X \\
&\geq 0.
\end{aligned}$$

The difference between the primal and dual objective values, i.e., the left-hand side in the equation above, is the *gap* for a feasible pair (w, X). If the gap is zero, then w is optimal for the primal, and X is optimal for the dual. In other words, zero gap is sufficient for optimality.

- *Strong duality.* There exists a primal-dual feasible pair (w^*, X^*) with zero gap, i.e.,

$$\sum d_{ij}^2 w_{ij}^* = \mathbf{Tr}\, X^*.$$

In particular, w^* is primal optimal and X^* is dual optimal. This means that optimal values of the primal and dual problems are the same. This strong duality result follows from Slater's condition [5, 26].

3.5. Optimality Conditions. From duality, we obtain the so-called Karush–Kuhn–Tucker (KKT) optimality conditions: A pair (w^*, X^*) is primal-dual optimal if and only if its elements are feasible and have zero gap, i.e., they satisfy the following.

- *Primal feasibility*:

$$w^* \geq 0, \qquad L(w^*) \succeq I - (1/n)\mathbf{1}\mathbf{1}^T.$$

- *Dual feasibility*:

$$\mathbf{1}^T X^* \mathbf{1} = 0, \qquad X^* \succeq 0, \qquad X_{ii}^* + X_{jj}^* - X_{ij}^* - X_{ji}^* \leq d_{ij}^2, \quad \{i,j\} \in E.$$

- *Complementary slackness on edges*:

$$\left(d_{ij}^2 - (X_{ii}^* + X_{jj}^* - X_{ij}^* - X_{ji}^*)\right) w_{ij}^* = 0, \quad \{i,j\} \in E.$$

- *Matrix complementary slackness*:

$$L(w^*)X^* = X^*.$$

Complementary slackness on the edges means that for each edge, if the primal constraint is not tight, i.e., $w_{ij}^* > 0$, then the dual constraint is tight, i.e.,

$$X_{ii}^* + X_{jj}^* - X_{ij}^* - X_{ji}^* = d_{ij}^2.$$

The matrix complementary slackness condition means that the range of X^* lies in the eigenspace of $L(w^*)$ associated with λ_2 (which is 1), i.e.,

$$(6) \qquad\qquad (L(w^*) - I)\, X^* = 0.$$

This means that if p is the multiplicity of λ_2 for a primal optimal $L(w^*)$, and q is the rank of a dual optimal X^*, then we must have $p \geq q$. This holds for *any* primal optimal w^* and *any* dual optimal X^*, which leads to the following result.

RESULT 1. *Let p^* denote the minimum multiplicity of the second eigenvalue of $L(w^*)$ over all primal optimal solutions w^*, and let q^* denote the maximum rank of X^* over all dual optimal solutions X^*. Then $p^* \geq q^*$.*

A simple and interesting special case is $p^* = 1$, as follows.

RESULT 2. *Suppose there is a primal optimal solution w^* whose second Laplacian eigenvalue is isolated (i.e., $p^* = 1$). Then the solution of the dual problem is unique and given by*

$$X^* = c^* u u^T,$$

where c^* is the optimal value (of the primal and dual problems), and u is the (normalized) eigenvector associated with the second Laplacian eigenvalue of $L(w^*)$, i.e., $L(w^*)u = u$. This last condition can be expressed as

$$\sum_{j \sim i} w_{ij}(u_i - u_j) = u_i, \quad i = 1, \ldots, n.$$

We note that Result 2 corresponds exactly with the case in which the objective λ_2 in the FMMP problem is differentiable at an optimal point (see, e.g., [16]).

4. A Geometric Dual.

4.1. A Maximum Variance Unfolding Problem. In this section we transform the dual SDP (5) to an equivalent problem that has a simple geometric interpretation. Since the dual variable X is symmetric and PSD, we can express it as

$$
(7) \qquad X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}
$$

for some set of n points $x_i \in \mathbf{R}^n$, i.e., we consider X as a Gram matrix. This factorization is not unique: multiplying x_i by any orthogonal matrix yields another set of vectors with the same Gram matrix. Conversely, any two factorizations are related by multiplication with an orthogonal matrix. We will refer to the set of points $x_1, \ldots, x_n$ as a *configuration*.

In terms of x_i, we have

$$
X_{ii} + X_{jj} - X_{ij} - X_{ji} = \| x_i - x_j \|^2,
$$

the square of the Euclidean distance between x_i and x_j. The condition $\mathbf{1}^T X \mathbf{1} = 0$ implies

$$
\sum_{i=1}^n x_i = 0,
$$

i.e., the points are centered at the origin.

Finally, the dual objective $\mathbf{Tr}\, X$ corresponds to

$$
\mathbf{Tr}\, X = \sum_{i=1}^n \| x_i \|^2,
$$

the variance of the set of points, and thus gives a measure of how spread out the points are. Since $\sum_{i=1}^n x_i = 0$, we can express the dual objective as

$$
\mathbf{Tr}\, X = \sum_{i=1}^n \| x_i \|^2 = \frac{1}{2n} \sum_{i,j=1}^n \| x_i - x_j \|^2,
$$

the sum of square distances between all pairs of points. Note that these interpretations hold for any choice of the x_i, i.e., for any factorization of a dual feasible X.

Now we change variables in the dual SDP (5), using the new variables $x_1, \ldots, x_n$. This gives the following problem:

$$
(8) \qquad \begin{aligned} \text{maximize} \quad & \sum_i \| x_i \|^2 \\ \text{subject to} \quad & \| x_i - x_j \| \le d_{ij}, \quad \{i,j\} \in E, \\ & \sum_i x_i = 0, \end{aligned}
$$

with variables $x_1, \ldots, x_n \in \mathbf{R}^n$. An optimal $x_1^*, \ldots, x_n^*$ will be called an *optimal configuration.* This problem is equivalent to the dual SDP (5): the optimal objective values are the same, and we can obtain optimal solutions of each one from the other using (7).

Problem (8) has a simple geometric interpretation. The goal is to position n points in $\mathbf{R}^n$ as far apart as possible, measured by their variance, while respecting upper bounds on the distance between some pairs of points. We call (8) the *maximum variance unfolding* (MVU) problem. (The term "unfolding" will be explained below.) After this manuscript was accepted for publication, the authors found the recent work [12], which gives the same interpretation of problem (8) (with d_{ij} all the same) as the dual of the absolute algebraic connectivity problem, and discusses many interesting properties of the optimal solution.

4.2. MVU with PSD Constraint. The MVU problem (8) is clearly invariant under orthogonal transformations: If $x_1, \ldots, x_n$ is a feasible (or optimal) configuration, then so is $Qx_1, \ldots, Qx_n$, where Q is orthogonal. Since the MVU problem (8) is invariant under orthogonal transformations, we can if we wish add a constraint that forces a particular choice (or orientation) of the configuration. One obvious choice corresponds to the unique symmetric PSD square root of X, i.e.,

$$
\begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = X^{1/2}.
$$

In other words, we can add the symmetry and PSD constraint

$$
\begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \succeq 0
$$

to the MVU problem to obtain

$$
\begin{aligned}
& \text{maximize} && \sum_i \|x_i\|^2 \\
& \text{subject to} && \|x_i - x_j\| \leq d_{ij}, \quad \{i,j\} \in E, \\
& && \sum_i x_i = 0, \qquad (x_i)_j = (x_j)_i, \quad i,j = 1, \ldots, n, \\
& && [x_1 \ \cdots \ x_n] \succeq 0.
\end{aligned}
\tag{9}
$$

This problem, which we call the *PSD MVU problem*, is equivalent to the MVU problem (8) in the following sense. The optimal values of the two problems are the same; any optimal solution of the PSD MVU problem is also optimal for the original MVU problem without the symmetry constraint; and any optimal solution of the original MVU problem can be transformed by an orthogonal matrix to give an optimal solution of the PSD MVU problem.

The only difference between the MVU problem (8) and the PSD MVU problem (9) is that in the PSD MVU problem, we have forced a particular orientation for the configuration. We will see in section 4.4 that this choice gives us a nice interpretation of the optimality conditions.

4.3. FMMP-MVU Duality. The MVU problem is *not* a convex optimization problem, although it is equivalent to one. However, from the SDP duality results, we do have the following.

- *Weak duality.* For any w feasible for the primal FMMP problem (2), and any $x_1, \ldots, x_n$ feasible for the associated MVU problem (8), we have

$$
\sum d_{ij}^2 w_{ij} \geq \sum \|x_i\|^2.
$$

In other words, the variance of any feasible configuration is a lower bound on the optimal value of the FMMP problem.

- *Strong duality.* There exist a primal FMMP feasible w^* and an MVU feasible configuration $x_1^*, \ldots, x_n^*$, for which

$$\sum d_{ij}^2 w_{ij}^* = \sum \|x_i^*\|^2.$$

In particular, w^* is an optimal set of edge rates for the FMMP problem, and $x_1^*, \ldots, x_n^*$ is an optimal configuration for the MVU problem.

In this sense, the FMMP problem (2) and the associated MVU problem (8) can be considered as (strong) duals.

4.4. Optimality conditions revisited. We can restate the optimality conditions for the FMMP and its SDP dual in terms of the MVU problem. A set of edge rates w^* and a configuration $x_1^*, \ldots, x_n^*$ are primal-dual optimal if and only if they satisfy the following.

- *Primal feasibility:*

$$w^* \geq 0, \qquad \lambda_2(w^*) \geq 1.$$

- *Dual feasibility:*

$$\sum x_i^* = 0, \qquad \|x_i^* - x_j^*\| \leq d_{ij}, \quad \{i,j\} \in E.$$

- *Complementary slackness on edges:*

$$w_{ij}^*(d_{ij} - \|x_i^* - x_j^*\|) = 0, \quad \{i,j\} \in E.$$

- *Matrix complementary slackness:*

$$L(w^*) \begin{bmatrix} x_1^{*T} \\ \vdots \\ x_n^{*T} \end{bmatrix} = \begin{bmatrix} x_1^{*T} \\ \vdots \\ x_n^{*T} \end{bmatrix}.$$

The edge complementary slackness condition means that for each edge, if the optimal rate w_{ij}^* is strictly positive, then the associated distance in the optimal configuration, $\|x_i^* - x_j^*\|$, is at its maximum value d_{ij}.

The matrix complementary slackness condition can be given an interesting interpretation. Let q be any unit vector, and define $z = (q^T x_1^*, \ldots, q^T x_n^*)$, which is the vector of q-coordinates of the configuration. Then we have

$$L(w^*)z = L(w^*) \begin{bmatrix} x_1^{*T} \\ \vdots \\ x_n^{*T} \end{bmatrix} q = \begin{bmatrix} x_1^{*T} \\ \vdots \\ x_n^{*T} \end{bmatrix} q = z.$$

Thus, for any direction q, the coordinates of an optimal configuration, when nonzero, form an eigenvector of $L(w^*)$ associated with the second eigenvalue.

We can give a very simple interpretation of the matrix complementary slackness condition for the PSD MVU problem (9). Using the symmetry constraint, the matrix complementary slackness condition simplifies to

$$L(w^*)x_i^* = x_i^*, \quad i = 1, \ldots, n,$$

i.e., each point in the optimal configuration $x_1^*, \ldots, x_n^*$ lies in the eigenspace of $L(w^*)$ associated with its second eigenvalue. This can also be concluded from the fact that the range of X^* is contained in the eigenspace of the second Laplacian eigenvalue of $L(w^*)$ (see (6)).

For future use, we note that the matrix complementary slackness condition can be expressed as

$$\sum_{j \sim i} w_{ij}^*(x_i^* - x_j^*) = x_i^*, \quad i = 1, \ldots, n.$$

(This is the case with and without the symmetry constraint.)

4.5. Some Rank and Dimension Inequalities. The (affine) dimension of a configuration $x_1, \ldots, x_n$ is the rank of the associated (Gram) matrix X. Therefore we can restate our rank inequalities from Result 1 in section 3.5 in terms of the MVU problem.

RESULT 3. *Let p denote the multiplicity of the second eigenvalue of $L(w^*)$ for some primal FMMP optimal w^*. Then any configuration that is optimal for the associated MVU problem has an affine dimension not exceeding p.*

Since this result holds for any primal optimal w^*, it holds with $p = p^*$, the minimum multiplicity over all primal optimal solutions. This result shows that if an FMMP instance has an optimal solution with low multiplicity of λ_2, then any optimal solution of the associated MVU is *guaranteed* to have a low affine dimension.

RESULT 4. *Suppose there is a primal optimal solution w^* whose second Laplacian eigenvalue is isolated (i.e., $p^* = 1$), with normalized eigenvector u. Then the optimal configuration of the associated MVU problem is unique, up to multiplication by an orthogonal matrix, and has affine dimension 1, i.e., its points all lie on a line. The optimal configuration for the MVU with symmetry constraints is given by*

$$x_i^* = c^* u_i u, \quad i = 1, \ldots, n,$$

where c^* is the optimal value of the FMMP (and MVU) problem.

There seem to be some interesting connections between the results in this section and some well-known functions of graphs, the Lovász number [18] and the Colin de Verdière parameter [25]. In connection with the Lovász number, suppose we label the vertices of the graph with unit vectors in a d-dimensional Euclidean space, such that vectors associated with nonadjacent vertices are orthogonal. The Lovász number is a lower bound on the embedding dimension d. The Colin de Verdière parameter is related to the maximum multiplicity of the second smallest eigenvalue of the weighted Laplacian of the graph, which plays an important role in certain graph embedding problems.

4.6. A Mechanics Interpretation. We can give a simple mechanics interpretation of the MVU problem. We consider an n-point system in $\mathbf{R}^n$, with a potential energy

$$U(r) = -r^2/2n$$

between any two points with distance r. (This is a repulsive potential whose strength grows with distance.) The total system potential or energy is then

$$U(x_1, \ldots, x_n) = -\sum_{i<j} \|x_i - x_j\|^2/2n.$$

We assume that for $i \sim j$, points i and j are connected by a rope of length d_{ij}, so that $\|x_i - x_j\| \leq d_{ij}$. Then the problem of finding the constrained minimum energy

configuration is the same as the MVU problem. This mechanics problem is invariant under multiplication by an orthogonal matrix, as well as under translation, so we can assume without loss of generality that $\sum x_i = 0$, in which case we have

$$U(x_1, \ldots, x_n) = -\sum \|x_i\|^2/2.$$

A necessary condition for $x_1, \ldots, x_n$ to be a constrained minimum energy configuration is that the system is in static equilibrium. By differentiating U with respect to x_i, we find that the potential corresponds to a (centripetal) force $F_i = x_i$ on point i. Let $T_{ij} \geq 0$ be the tension in rope $\{i, j\}$. The static equilibrium condition is that the tensions are in balance with the outward centripetal force, i.e.,

$$\sum_{\{i,j\} \in E} T_{ij} \frac{x_i - x_j}{\|x_i - x_j\|} = x_i, \quad i = 1, \ldots, n.$$

The tensions must be zero unless the ropes are taut, i.e.,

$$(d_{ij} - \|x_i - x_j\|)T_{ij} = 0, \quad \{i, j\} \in E.$$

Using this equation, the force balance can be expressed as

$$\sum_{\{i,j\} \in E} \frac{T_{ij}}{d_{ij}}(x_i - x_j) = x_i, \quad i = 1, \ldots, n.$$

By comparing these conditions with the optimality conditions, we see that $T_{ij} = d_{ij}w_{ij}^*$ is a set of tensions that satisfy the static equilibrium condition, where w_{ij}^* is an optimal set of rates for the FMMP problem. Thus, we can interpret the optimal rates for the FMMP problem as tensions per unit length in the ropes that hold together a minimum energy configuration.

4.7. A Connection to Machine Learning. The MVU problem (8) is very close to a problem proposed by Weinberger and Saul in [28, 29] as a heuristic for unfolding data that lies on a manifold. Indeed, the name "maximum variance unfolding" is inspired entirely by their application.

Weinberger and Saul proposed the following problem:

$$
\begin{aligned}
& \text{maximize} && \sum_i \|x_i\|^2 \\
(10) \quad & \text{subject to} && \|x_i - x_j\| = d_{ij}, \quad \{i, j\} \in E, \\
& && \sum_i x_i = 0,
\end{aligned}
$$

with variables $x_1, \ldots, x_n \in \mathbf{R}^n$; the graph, along with the distances d_{ij}, are the problem data. The only difference between this problem and the MVU problem (8) is in the distance constraints: in Weinberger and Saul's problem, they are equalities, whereas in the MVU problem, they are inequalities. Weinberger and Saul's equality constrained MVU problem can be converted to an SDP very similar to our SDP dual of the FMMP, i.e., the problem (5) (or (4); the only difference is that the distance inequalities in our formulation are replaced with equalities in Weinberger and Saul's).

We can give a brief description of the context and ideas behind Weinberger and Saul's problem. The starting point is a set of distances between all pairs of a large set
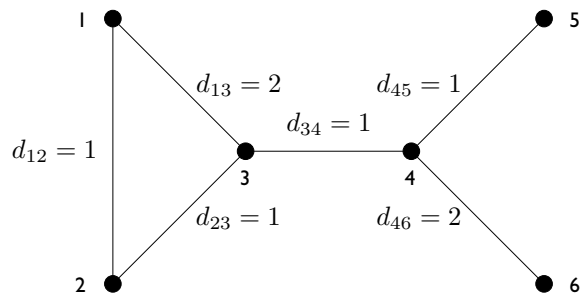
Fig. 3 *A small numerical example.*

of points in $\mathbf{R}^n$. The underlying assumption is that these points lie (approximately) on a manifold of low dimension, and the goal is to discover the manifold (and, in particular, its dimension). To do this, we select a distance threshold and *ignore* all pairwise distances that are larger than the threshold. This results in a sparse graph, whose edges are labeled with the distance d_{ij} between points x_i and x_j. We now try to find a configuration that respects these local distances and has lowest possible affine dimension. This is a hard problem, but Weinberger and Saul proposed maximizing the variance of the configuration, subject to matching the local distances, as a heuristic for solving it. Our result above on rank has immediate ramifications for this method of unfolding, since it gives an upper bound on the affine dimension of the resulting configuration.

We can turn things around as well, and work out a variation on our FMMP problem that is a dual for Weinberger and Saul's problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum d_{ij}^2 w_{ij} \\
\text{subject to} \quad & L(w) \succeq I - (1/n)\mathbf{1}\mathbf{1}^T.
\end{aligned}
$$

This is the same as our FMMP problem, except that we drop the nonnegativity constraints on the edge weights w.

While it seems reasonable, in retrospect, that fast mixing Markov chains and manifold unfolding are related, it was not obvious (at least to us) that the problems are so closely related.

5. Examples.

5.1. A Small Example. We consider the small example with graph and costs d_{ij} shown in Figure 3. The solutions to both the FMMP problem (2) and its dual MVU problem (8) are readily found numerically (and simultaneously) using an SDP solver to solve the SDP (3) (and its dual). For example, `SDPSOL` (Wu and Boyd [30]) gives us the FMMP optimal edge weights

$$
\begin{aligned}
w_{12}^* = 1.776, \quad & w_{23}^* = 3.276, \quad w_{13}^* = 0.362, \\
w_{34}^* = 4.500, \quad & w_{45}^* = 1.500, \quad w_{46}^* = 1.250,
\end{aligned}
$$

with optimal objective value $\sum d_{ij}^2 w_{ij}^* = 17.5$. Along with this primal solution, `SDPSOL` produces a corresponding dual optimal solution X^*, which we can factor to obtain an optimal configuration. The Laplacian associated with the edge weights given above has a second eigenvalue with multiplicity 1, which implies (by our results) that the

Fig. 4   *The path* P_6.

optimal solution of the MVU problem is unique (up to multiplication by an orthogonal matrix) and lies on a line (i.e., has affine dimension 1). This solution is

$$x_1^* = -2.5q, \quad x_2^* = -1.5q, \quad x_3^* = -0.5q, \quad x_4^* = 0.5q, \quad x_5^* = 1.5q, \quad x_6^* = 2.5q,$$

where q is any unit vector. The associated maximum variance is $\sum \|x_i^*\|^2 = 17.5$, the same (as it must be) as the minimum cost for the FMMP problem. This MVU solution was obtained numerically, but is easily verified to be the exact solution.

   The optimal solution to the FMMP problem, however, is *not* unique. The full set of optimal edge rates can be obtained from the optimal configuration of the MVU problem and the optimality conditions, which include the linear equations

$$w_{12}^* + 2w_{13}^* = 2.5, \quad w_{23}^* - w_{12}^* = 1.5, \quad 2w_{13}^* + w_{23}^* = 4,$$
$$w_{34}^* = 4.5, \quad w_{45}^* = 1.5, \quad w_{46}^* = 1.25.$$

It can be verified that any choice of $w_{ij}^* \geq 0$ that satisfies these equations is optimal, so the set of optimal edge rates for the FMMP problem is given by

$$w_{12}^* = a, \quad w_{23}^* = 1.5 + a, \quad w_{13}^* = 1.25 - 0.5a,$$
$$w_{34}^* = 4.5, \quad w_{45}^* = 1.5, \quad w_{46}^* = 1.25,$$

where the parameter a varies between 0 and 2.5.

   The intuition provided by the MVU problem allows us to make guesses about how the solution of both the FMMP and the associated MVU changes when the data change. For example, suppose that we increase d_{13} above the value 2 (keeping the other d_{ij} the same). The optimal configuration for the MVU is the same as the one described above, with all points on a line. The distance constraint $\|x_1^* - x_2^*\| \leq d_{13}$, however, is slack at the optimal solution. This in turn implies that any FMMP on the graph assigns zero transition rate to the edge $\{1, 3\}$. This is readily verified numerically, or analytically.

   5.2. Paths. Our second example is a path P_n, with vertices indexed in order on the path, as shown in Figure 4. Boyd et al. solved the problem of finding the fastest mixing discrete-time Markov chain on a path in [3]; not surprisingly, we can do the same here.

   RESULT 5. *The optimal rates on a path are unique and given by*

$$w_{i,i+1}^* = \frac{1}{d_{i,i+1}} \left(\frac{i}{n} \sum_{j=1}^{n-1} (n-j)d_{j,j+1} - \sum_{j=1}^{i-1} (i-j)d_{j,j+1} \right), \quad i = 1, \ldots, n-1.$$

The optimal configuration for the associated MVU problem is also unique (up to multiplication by an orthogonal matrix) and consists of points on a line with distances $d_{i,i+1}$:

$$x_i^* = \frac{1}{n} \left(\sum_{j=1}^{i-1} j d_{j,j+1} - \sum_{j=i}^{n-1} (n-j)d_{j,j+1} \right) u, \quad i = 1, \ldots, n,$$

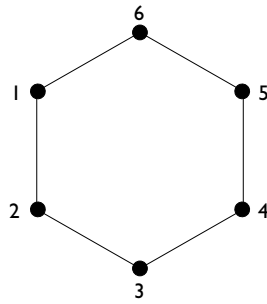*where $u \in \mathbf{R}^n$ is any unit vector.*

Fig. 5 *The cycle* C_6.

In particular, when the edge costs are all 1, the optimal rates form a parabolic curve:

$$w^*_{i,i+1} = \frac{i(n-i)}{2}, \quad i = 1, \ldots, n-1.$$

Proof. We will show directly that the one-dimensional configuration is optimal. For any $x_1, \ldots, x_n$ with $\|x_i - x_{i+1}\| \le d_{i,i+1}$ and $\sum x_i = 0$, we have

$$\sum_{i<j} \|x_i - x_j\|^2 \le \sum_{i<j} (\|x_i - x_{i+1}\| + \cdots + \|x_{j-1} - x_j\|)^2$$

$$\le \sum_{i<j} (d_{i,i+1} + \cdots + d_{j-1,j})^2.$$

This shows that any feasible configuration has a variance smaller than the configuration given above (for which the inequalities become equalities). The optimal rates $w^*_{i,i+1}$ are uniquely determined from the optimal configuration using the optimality conditions.   □

The solution for the path with uniform edge cost was obtained in [9] in the study of the absolute algebraic connectivity problem. In the same paper, and the later one [10], Fiedler gave a complete solution for trees, and showed that the absolute algebraic connectivity for trees is always a rational number.

5.3. Cycles. Our next example is a cycle C_n, as illustrated in Figure 5. We first consider the simple case with all costs on the edges 1, i.e., $d_{ij} = 1$. By symmetry and convexity (see [5, Exercise 4.4]), it is easy to show that an optimal solution to the FMMP problem can, without loss of generality, be assumed to have equal rates on all edges. The eigenvalues of the associated Laplacian are readily found analytically, and we find that the optimal rates

$$w_{ij} = \frac{1}{4\sin^2(\pi/n)}, \quad \{i,j\} \in E,$$

yield $\lambda_2 = 1$.

It is also easy to guess that an optimal configuration for the MVU problem consists of n points evenly spaced on a circle with radius $1/(2\sin(\pi/n))$. However, it is not immediately clear how to prove this directly by symmetry arguments, since the MVU problem (8) is not convex. It is also not obvious how to prove this result using direct arguments like those for a path.

We can, however, establish that this configuration is optimal by verifying that it has a variance equal to the cost of the FMMP for the rates given above, i.e.,

$$\frac{n}{4\sin^2(\pi/n)}.$$

This in fact certifies that this circular configuration, as well as the weights found above, are optimal for the MVU and FMMP, respectively.

The general case with arbitrary weights d_{ij} is more complicated. But we will show the following result.

RESULT 6. *Any optimal configuration of a cycle is two-dimensional.*

Proof. We prove something stronger: The multiplicity of any Laplacian eigenvalue of a cycle does not exceed 2. Suppose u^1, u^2, u^3 are linearly independent eigenvectors associated with an eigenvalue $\lambda > 0$. Take any two adjacent vertices, say, 1 and 2. Let (α, β, γ) be a nonzero triple such that $u = \alpha u^1 + \beta u^2 + \gamma u^3$ vanishes on these two vertices, i.e., $u_1 = 0$ and $u_2 = 0$. Since u is also an eigenvector associated with λ, it should satisfy the equation $L(w)u = \lambda u$, i.e., $\sum_{j \sim i} w_{ij}(u_i - u_j) = \lambda u_i$ for any i.

Since the cycle is connected, without loss of generality, suppose vertex 2 also connects to vertex 3; then $w_{21}(u_2 - u_1) + w_{23}(u_2 - u_3) = \lambda u_2$. Since $u_1 = u_2 = 0$, $w_{23} \neq 0$, we have $u_3 = 0$. By repeating this argument, it is easy to see that u actually vanishes on every vertex, i.e., $u = 0$. This contradicts the assumption that u^1, u^2, u^3 are linearly independent.   □

6. Time-Reversible FMMP. The FMMP problem can be extended to a time-reversible Markov process, with a specified equilibrium distribution π. We can pose this problem as an SDP, and therefore solve it efficiently; its associated dual problem can be interpreted as a type of weighted MVU problem.

The transition rates q_{ij} of a time-reversible Markov process must satisfy the detailed balance equation

$$\pi_i q_{ij} = \pi_j q_{ji}, \quad \{i, j\} \in E,$$

where $(\pi_1, \ldots, \pi_n)$ is the given equilibrium distribution. Let $w_{ij} = \pi_i q_{ij} = \pi_j q_{ji}$ and let $L(w)$ be the Laplacian with w_{ij} as the edge weights. Let $\Pi = \mathrm{diag}(\pi_1, \ldots, \pi_n)$. Then the evolution of the Markov process is given by

$$\frac{d\pi(t)}{dt} = -L\Pi^{-1}\pi(t),$$

where $\pi(t)$ denotes the distribution of the state at time t.

The matrix $L\Pi^{-1}$ is similar to $\Pi^{-1/2}L\Pi^{-1/2}$, which is symmetric PSD. The eigenvector of $\Pi^{-1/2}L\Pi^{-1/2}$ associated with the smallest eigenvalue (which is zero) is $q = (\sqrt{\pi_1}, \ldots, \sqrt{\pi_n})$. The asymptotic rate of convergence to equilibrium is determined by the second eigenvalue of $\Pi^{-1/2}L\Pi^{-1/2}$.

The time-reversible FMMP problem can then be formulated as the SDP

(11)
$$\begin{aligned}
\text{minimize} \quad & \sum d_{ij}^2 w_{ij} \\
\text{subject to} \quad & \Pi^{-1/2}L(w)\Pi^{-1/2} \succeq I - qq^T, \\
& w \geq 0.
\end{aligned}$$

The associated dual SDP can be simplified to yield

(12)
$$\begin{aligned}
\text{maximize} \quad & \mathbf{Tr}\, X \\
\text{subject to} \quad & \pi_i^{-1}X_{ii} + \pi_j^{-1}X_{jj} - 2(\pi_i\pi_j)^{-1/2}X_{ij} \leq d_{ij}^2, \quad \{i, j\} \in E, \\
& q^T X q = 0, \quad X \succeq 0.
\end{aligned}$$

To derive a geometric dual for the time-reversible FMMP problem, we let the symmetric matrix $\Pi^{-1/2}X\Pi^{-1/2}$ be the Gram matrix of a configuration of n points $x_1,\ldots,x_n \in \mathbf{R}^n$. Using these variables we can reformulate the dual (12) as

$$
\begin{aligned}
\text{(13)} \qquad \text{maximize} \quad & \sum \pi_i \|x_i\|^2 \\
\text{subject to} \quad & \|x_i - x_j\| \le d_{ij}, \quad \{i,j\} \in E, \\
& \sum \pi_i x_i = 0.
\end{aligned}
$$

The objective here is the variance of the configuration, but with probability π_i on point x_i. (The constraint $\sum \pi_i x_i = 0$ sets the mean of the configuration to be zero.) Thus, the problem is again one of maximizing the variance of a configuration of points, subject to some distance inequalities given by the original graph and costs d_{ij}. The only difference is that the variance is calculated using the probabilities π_i (instead of uniform probabilities).

The time-reversible FMMP problem has a simple interpretation in the context of the physical systems described in section 2. For example, for a grounded capacitor RC circuit, the reversible FMMP problem corresponds to an RC circuit with given capacitors to ground $C_1,\ldots,C_n$.

Acknowledgment. We thank the referees for helpful comments.

REFERENCES

[1] S. Benson and Y. Ye, *DSDP: Dual-Scaling Algorithm for Semidefinite Programming*, Tech. Report ANL/MCS-P851-1000, Argonne National Laboratory, Argonne, IL, 2001.
[2] B. Borchers, *CSDP, a C library for semidefinite programming*, Optim. Methods Softw., 11 (1999), pp. 613–623.
[3] S. Boyd, P. Diaconis, J. Sun, and L. Xiao, *Fastest mixing Markov chain on a path*, Amer. Math. Monthly, 113 (2006), pp. 70–74.
[4] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph*, SIAM Rev., 46 (2004), pp. 667–689.
[5] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004. Available online from www.stanford.edu/~boyd/cvxbook.
[6] S. J. Cox and M. L. Overton, *On the optimal design of columns against buckling*, SIAM J. Math. Anal., 23 (1992), pp. 287–325.
[7] S. J. Cox and P. X. Uhlig, *Where best to hold a drum fast*, SIAM J. Optim., 9 (1999), pp. 948–964.
[8] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab., 6 (1996), pp. 695–750.
[9] M. Fiedler, *Absolute algebraic connectivity of trees*, Linear Multilinear Algebra, 26 (1990), pp. 85–106.
[10] M. Fiedler, *Some minimax problems for graphs*, Discrete Math., 121 (1993), pp. 65–74.
[11] K. Fujisawa, M. Kojima, K. Nakata, and M. Yamashita, *SDPA (SemiDefinite Programming Algorithm) User's Manual. Version* 6.00, Research Reports on Mathematical and Computing Sciences B-308, Tokyo Institute of Technology, Japan, 2002.
[12] F. Göring, C. Helmberg, and M. Wappler, *Embedded in the Shadow of the Separator*, Preprint 2005-12, Fakultät für Mathematik, Technische Universität Chemnitz, Chemnitz, Germany, 2005.
[13] C. Helmberg, *Semidefinite Programming*, www-user.tu-chemnitz.de/~helmberg/semidef.
[14] C. Helmberg, *SBmethod—a C++ Implementation of the Spectral Bundle Method*, Tech. Report ZIB-Report 00-35, Konrad-Zuse-Zentrum für Informationstechnik, Berlin, 2000.
[15] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[16] A. S. Lewis, *Convex analysis on the Hermitian matrices*, SIAM J. Optim., 6 (1996), pp. 164–177.
[17] A. Lewis and M. Overton, *Eigenvalue optimization*, Acta Numerica, 5 (1996), pp. 149–190.

[18] L. Lovász, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory, 25 (1979), pp. 1–7.

[19] B. Mohar, *Some applications of Laplacian eigenvalues of graphs*, in Graph Symmetry: Algebraic Methods and Applications, G. Hahn and G. Sabidussi, eds., NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 497, Kluwer, Dordrecht, The Netherlands, 1997, pp. 227–275.

[20] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2004), pp. 229–251.

[21] M. L. Overton, *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 256–268.

[22] M. L. Overton, *Large-scale optimization of eigenvalues*, SIAM J. Optim., 2 (1992), pp. 88–120.

[23] J. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw., 11-12 (1999), pp. 625–653.

[24] K. Toh, M. Todd, and R. Tutuncu, *SDPT3—a MATLAB software package for semidefinite programming*, Optim. Methods Softw., 11 (1999), pp. 545–581.

[25] H. van der Holst, L. Lovász, and A. Schrijver, *The Colin de Verdière graph parameter*, in Graph Theory and Combinatorial Biology, L. Lovász et al., eds., Bolyai Soc. Math. Stud. 7, János Bolyai Mathematical Society, Budapest, 1999, pp. 29–85.

[26] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.

[27] L. Vandenberghe, S. Boyd, and A. E. Gamal, *Optimizing dominant time constant in RC circuits*, IEEE Trans. Comput. Aided Design, 2 (1998), pp. 110–125.

[28] K. Weinberger and L. Saul, *Unsupervised learning of image manifolds by semidefinite programming*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-04), Washington, D.C., 2004, pp. 988–995.

[29] K. Weinberger, F. Sha, and L. Saul, *Learning a kernel matrix for nonlinear dimensionality reduction*, in Proceedings of the 21st International Conference on Machine Learning (ICML-04), Banff, Canada, 2004, pp. 839–846.

[30] S.-P. Wu and S. Boyd, *SDPSOL: A parser/solver for semidefinite programs with matrix structure*, in Advances in Linear Matrix Inequality Methods in Control, L. E. Ghaoui and S.-I. Niculescu, eds., SIAM, Philadelphia, 2000, pp. 79–91.

[31] L. Xiao and S. Boyd, *Fast linear iterations for distributed averaging*, Systems Control Lett., 53 (2004), pp. 65–78.

[32] L. Xiao and S. Boyd, *Optimal scaling of a gradient method for distributed resource allocation*, to appear in J. Optim. Theory Appl., 129 (2006). Available online from www.stanford.edu/~boyd/fast_redstb.