SIAM J. Matrix Anal. Appl.
Vol. 36, No. 4, pp. 1465–1488
© 2015 Society for Industrial and Applied Mathematics

CONVEX RELAXATIONS FOR PERMUTATION PROBLEMS*

F. FOGEL†, R. JENATTON‡, F. BACH§, AND A. D'ASPREMONT¶

Abstract. Seriation seeks to reconstruct a linear order between variables using unsorted, pairwise similarity information. It has direct applications in archeology and shotgun gene sequencing, for example. We write seriation as an optimization problem by proving the equivalence between the seriation and combinatorial 2-SUM problems on similarity matrices (2-SUM is a quadratic minimization problem over permutations). The seriation problem can be solved exactly by a spectral algorithm in the noiseless case, and we derive several convex relaxations for 2-SUM to improve the robustness of seriation solutions in noisy settings. These convex relaxations also allow us to impose structural constraints on the solution and hence solve semisupervised seriation problems. We derive new approximation bounds for some of these relaxations and present numerical experiments on archeological data, Markov chains, and DNA assembly from shotgun gene sequencing data.

Key words. seriation, C1P, convex relaxation, quadratic assignment problem, shotgun DNA sequencing

AMS subject classifications. 06A07, 90C27, 90C25, 92D20

DOI. 10.1137/130947362

1. Introduction.

We study optimization problems written over the set of permutations. While the relaxation techniques discussed in what follows are applicable to a much more general setting, most of the paper is centered on the *seriation* problem: we are given a similarity matrix between a set of n variables and assume that the variables can be ordered along a chain, where the similarity between variables decreases with their distance within this chain. The seriation problem seeks to reconstruct this linear ordering based on unsorted, possibly noisy, pairwise similarity information.

This problem has its roots in archeology [38] and also has direct applications in, e.g., envelope reduction algorithms for sparse linear algebra [3], in identifying interval graphs for scheduling [17], or in shotgun DNA sequencing where a single strand of genetic material is reconstructed from many cloned shorter reads (i.e., small, fully

*Received by the editors December 2, 2013; accepted for publication (in revised form) February 11, 2015; published electronically November 3, 2015. A subset of these results was presented at Advances in Neural Information Processing Systems 26 (NIPS 2013).

http://www.siam.org/journals/simax/36-4/94736.html

†CMAP, École Polytechnique, 91128 Palaiseau, France (fajwel.fogel@cmap.polytechnique.fr). This author's research was supported by a European Research Council starting grant (project SIPA), the MSR-Inria Joint Centre, the chaire *Économie des nouvelles données*, the *data science* joint research initiative with the *fonds AXA pour la recherche*, and a gift from Société Générale Cross Asset Quantitative Research.

‡CMAP, École Polytechnique, 91128 Palaiseau, France, and Amazon Labs, Berlin, Germany (r.jenatton@criteo.com). This author's research was supported by a European Research Council starting grant (project SIPA), the MSR-Inria Joint Centre, the chaire *Économie des nouvelles données*, the *data science* joint research initiative with the *fonds AXA pour la recherche*, and a gift from Société Générale Cross Asset Quantitative Research.

§INRIA, SIERRA Project-Team, 75214 Paris, France, and D.I., École Normale Supérieure, 75230 Paris, France (francis.bach@ens.fr). This author's research was supported by a European Research Council starting grant (project SIERRA).

¶CNRS, UMR 8548, 75005 Paris, France, and D.I., École Normale Supérieure, 75230 Paris, France (aspremon@ens.fr). This author's research was supported by a European Research Council starting grant (project SIPA), the MSR-Inria Joint Centre, the chaire *Économie des nouvelles données*, the *data science* joint research initiative with the *fonds AXA pour la recherche*, and a gift from Société Générale Cross Asset Quantitative Research.

sequenced sections of DNA) [18, 31]. With shotgun gene sequencing applications in mind, many references focused on the *consecutive ones property* (C1P) problem, which seeks to permute the rows of a binary matrix so that all the ones in each column are contiguous. In particular, [17] studied further connections to interval graphs and [24] crucially showed that a solution to the C1P problem can be obtained by solving the seriation problem on the squared data matrix. We refer the reader to [10, 44, 28] for a much more complete survey of applications.

On the algorithmic front, the seriation problem was shown to be NP-complete in [19]. Archeological examples are usually small scale, and earlier references such as [38] used greedy techniques to reorder matrices. Similar techniques were and are still used to reorder genetic data sets. More general ordering problems were studied extensively in operations research, mostly in connection with the quadratic assignment problem (QAP), for which several convex relaxations were derived in, e.g., [27, 46]. Since a matrix is a permutation matrix iff it is both orthogonal and doubly stochastic, much work also focused on producing semidefinite relaxations to orthogonality constraints [32, 42]. These programs are convex and can be solved using conic programming solvers, but the relaxations are usually very large and scale poorly. More recently, however, a spectral algorithm that exactly solves the seriation problem in a noiseless setting was produced in [2]. They show that for similarity matrices computed from serial variables (for which a total order exists) the ordering of the second eigenvector of the Laplacian (a.k.a. the Fiedler vector) matches that of the variables, in results that are closely connected to those obtained on the interlacing of eigenvectors for Sturm–Liouville operators. A lot of work has focused on the minimum linear arrangement problem or 1-SUM, with the authors of [12, 13, 6] and [37, 14, 7] producing semidefinite relaxations with nearly dimension independent approximation ratios. While these relaxations form semidefinite programs that have an exponential number of constraints, they admit a polynomial-time separation oracle and can be solved using the ellipsoid method. The latter algorithm being extremely slow, these programs have very little practical impact. Finally, seriation is also directly related to the manifold learning problem [45], which seeks to reconstruct a low dimensional manifold based on local metric information. Seriation can be seen as a particular instance of that problem, where the manifold is unidimensional but the similarity information is not metric.

Our contribution here is twofold. First, we explicitly write seriation as an optimization problem by proving the equivalence between the seriation and combinatorial 2-SUM problems on similarity matrices. 2-SUM, defined in, e.g., [19], is a quadratic minimization problem over permutations. Our result shows in particular that 2-SUM is polynomially solvable for matrices coming from serial data. This quadratic problem was mentioned in [2], but no explicit connection was established between combinatorial problems like 2-SUM and seriation. While this paper was under review, a recent working paper by [26] has extended the results in Propositions 2.11 and 2.12 here to show that the QAP problem $Q(A, B)$ is solved by the spectral seriation algorithm when A is a similarity matrix (satisfying the Robinson assumption detailed below) and B is a Toeplitz dissimilarity matrix (e.g., $B_{ij} = (i - j)^2$ in the 2-SUM problem discussed here).

Second, we derive several new convex relaxations for the seriation problem. Our simplest relaxation is written over the set of doubly stochastic matrices and appears to be more robust to noise than the spectral solution in a number of examples. Perhaps more importantly, it allows us to impose additional structural constraints to solve semisupervised seriation problems. We also briefly outline a fast algorithm for

projecting on the set of doubly stochastic matrices, which is of independent interest. In the appendix, we also produce a semidefinite relaxation for the seriation problem using the classical lifting argument in [40, 30] written on a nonconvex quadratic program (QP) formulation of the combinatorial 2-SUM problem. Based on randomization arguments in [34, 9] for the MaxCut and k-dense-subgraph problems, we show that this relaxation of the set of permutation *matrices* achieves an approximation ratio of $O(\sqrt{n})$. We also recall how several other relaxations of the minimum linear arrangement (MLA) problem, written on permutation *vectors*, can be adapted to get nearly dimension independent $O(\sqrt{\log n})$ approximation ratios by forming (exponentially large but tractable) semidefinite programs. While these results are of limited practical impact because of the computational cost of the semidefinite programs they form, they do show that certain QAP instances written on Laplacian matrices, such as the seriation problem considered here, are much simpler to approximate than generic QAP problems. They also partially explain the excellent empirical performance of our relaxations in the numerical experiments of section 5.

The paper is organized as follows. In section 2, we show how to decompose similarity matrices formed in the C1P problem as conic combinations of CUT matrices, i.e., elementary block matrices. This allows us to connect the solutions of the seriation and 2-SUM minimization problems on these matrices. In section 3, we use these results to write convex relaxations of the seriation problem by relaxing the set of permutation matrices as doubly stochastic matrices in a QP formulation of the 2-SUM minimization problem. Section 4 briefly discusses first order algorithms solving the doubly stochastic relaxation and details in particular a block coordinate descent algorithm for projecting on the set of doubly stochastic matrices. Finally, section 5 describes applications and numerical experiments on archeological data, Markov chains, and DNA assembly problems. In the appendix, we describe larger semidefinite relaxations of the 2-SUM QP and obtain $O(\sqrt{n})$ approximation bounds using randomization arguments. We also detail several direct connections with the MLA problem.

Notation. We use the notation $\mathcal{P}$ for both the set of permutations of $\{1,\dots,n\}$ and the set of permutation matrices. The notation π will refer to a permuted vector $(1,\dots,n)^T$, while the notation Π (in a capital letter) will refer to the corresponding matrix permutation, which is a $\{0,1\}$ matrix such that $\Pi_{ij}=1$ iff $\pi(i)=j$. Moreover, y_π is the vector with coefficients $(y_{\pi(1)},\dots,y_{\pi(n)})$, and hence $\Pi y = y_\pi$ and $\Pi^T y_\pi = y$. This also means that $A\Pi^T$ is the matrix with coefficients $A_{i\pi(j)}$, and $\Pi A\Pi^T$ is the matrix with coefficients $A_{\pi(i)\pi(j)}$. For a vector $y\in\mathbb{R}^n$, we write $\mathbf{var}(y)$ as its variance, with $\mathbf{var}(y)=\sum_{i=1}^n y_i^2/n-(\sum_{i=1}^n y_i/n)^2$, and we also write $y_{[u,v]}\in\mathbb{R}^{v-u+1}$ as the vector $(y_u,\dots,y_v)^T$. Here, $e_i\in\mathbb{R}^n$ is the ith Euclidean basis vector and $\mathbf{1}$ is the vector of ones. Recall also that the matrix product can be written in terms of outer products, with $AB=\sum_i A_{(i)}B^{(i)}$, with $A_{(i)}$ (resp., $B^{(i)}$) the ith column (resp., row) of A (resp., B). For a matrix $A\in\mathbb{R}^{m\times n}$, we write $\mathbf{vec}(A)\in\mathbb{R}^{mn}$ as the vector formed by stacking up the columns of A. We write $\mathbf{I}$ as the identity matrix and $\mathbf{S}_n$ as the set of symmetric matrices of dimension n, $\|\cdot\|_F$ denotes the Frobenius norm, $\lambda_i(X)$ denotes the ith eigenvalue (in increasing order) of X, and $\|X\|_\infty=\|\mathbf{vec}(X)\|_\infty$.

2. Seriation, 2-SUM, and C1P. Given a symmetric, binary matrix A, we will focus on variations of the following 2-SUM combinatorial minimization problem, studied in, e.g.. [19] and written as

$$\begin{array}{ll}\text{minimize} & \sum_{i,j=1}^n A_{ij}(\pi(i)-\pi(j))^2\\ \text{subject to} & \pi\in\mathcal{P},\end{array} \tag{1}$$

where $\mathcal{P}$ is the set of permutations of the vector $(1,\dots,n)^T$. This problem is used, for example, to reduce the envelope of sparse matrices and is shown in [19, Thm. 2.2] to be NP-complete. When A has a specific structure, in [2] it is shown that a related matrix ordering problem used for seriation can be solved explicitly by a spectral algorithm. However, the results in [2] do not explicitly link spectral ordering and the optimum of (1). The main objective of this section is to show the equivalence between the 2-SUM and seriation problems for certain classes of matrices A. In particular, for some instances of A related to seriation and the C1P problem, we will show below that the spectral ordering directly minimizes the objective of problem (1). We first focus on binary matrices and then extend our results to more general unimodal matrices.

Let $A \in \mathbf{S}_n$, and consider the following generalization of the 2-SUM minimization problem:

$$\begin{array}{ll} \text{minimize} & f(y_\pi) \triangleq \sum_{i,j=1}^n A_{ij}(y_{\pi(i)} - y_{\pi(j)})^2 \\ \text{subject to} & \pi \in \mathcal{P} \end{array} \tag{2}$$

in the permutation variable π, where $y \in \mathbb{R}^n$ is a given weight vector. The classical 2-SUM minimization problem (1) is a particular case of problem (2) with $y_i = i$. The main point of this section is to show that if A is the permutation of a similarity matrix formed from serial data, then minimizing (2) recovers the correct variable ordering. To do this, we simply need to show that when A is correctly ordered, a monotonic vector y solves (2), since reordering y is equivalent to reordering A. Our strategy is to first show that we can focus on matrices A that are sums of simple CUT matrices, i.e., symmetric block matrices with a single constant block (see [16]). We then show that all minimization problems (2) written on CUT matrices have a common optimal solution, where y_π is monotonic.

2.1. Similarity, C1P, and unimodal matrices. We begin by introducing a few definitions on R-matrices (i.e., similarity matrices), the C1P, and unimodal matrices following [2].

DEFINITION 2.1 (R-matrices). *We say that the matrix $A \in \mathbf{S}_n$ is an R-matrix (or Robinson matrix) iff it is symmetric and satisfies $A_{i,j} \leq A_{i,j+1}$ and $A_{i+1,j} \leq A_{i,j}$ in the lower triangle, where $1 \leq j < i \leq n$.*

Another way to write the R-matrix conditions is to impose $A_{ij} \leq A_{kl}$ if $|i-j| \leq |k-l|$ is off-diagonal, i.e., the coefficients of A decrease as we move away from the diagonal (cf. Figure 1). In that sense, R-matrices are similarity matrices between variables organized on a *chain*, i.e., where the similarity A_{ij} is monotonically decreasing with the distance between i and j on this chain. We also introduce a few definitions related to the C1P problem and its unimodal extension.

DEFINITION 2.2 (P-matrices). *We say that the $\{0,1\}$-matrix $A \in \mathbb{R}^{n\times m}$ is a P-matrix (or Petrie matrix) iff for each column of A the ones form a consecutive sequence.*

As in [2], we will say that A is *pre-R* (resp., *pre-P*) iff there is a permutation Π such that $\Pi A \Pi^T$ is an R-matrix (resp., ΠA is a P-matrix). Based on [24], we also define a generalization of P-matrices called (appropriately enough) Q-matrices, i.e., matrices with unimodal columns.

DEFINITION 2.3 (Q-matrices). *We say that a matrix $A \in \mathbb{R}^{n\times m}$ is a Q-matrix iff each column of A is unimodal, i.e., the coefficients of each column increase to a maximum, then decrease.*

Note that R-matrices are symmetric Q-matrices. We call a matrix A *pre-Q* iff there is a permutation Π such that ΠA is a Q-matrix. Next, again based on [24], we

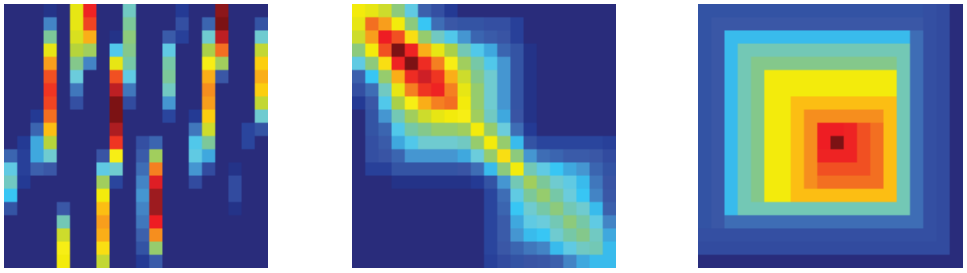


FIG. 1. *A sample Q-matrix (see Definition* 2.3*), which has unimodal columns* (left)*; its "circular square" $A \circ A^T$ (see Definition* 2.4*), which is an R-matrix* (center)*; and a matrix $a \circ a^T$ where a is a unimodal vector* (right).

define the *circular product* of two matrices.

DEFINITION 2.4. *Given $A, B^T \in \mathbb{R}^{n\times m}$, their circular product $A \circ B$ is defined as*

$$(A \circ B)_{ij} = \sum_{k=1}^{m} \min\{A_{ik}, B_{kj}\}, \quad i, j = 1, \ldots, n;$$

note that when A is a symmetric matrix, $A \circ A$ is also symmetric.

Remark that when A, B are $\{0, 1\}$ matrices, $\min\{A_{ik}, B_{kj}\} = A_{ik}B_{kj}$, so the circular product matches the regular matrix product AB. Similarly, a $\{0, 1\}$ matrix with the C1P is also unimodal.

2.2. Seriation on CUT matrices. We now introduce CUT matrices (named after the CUT decomposition in [16] whose definition is slightly more flexible) and first study the seriation problem on these simple block matrices. The motivation for this definition is that if A is a P-, Q-, or R-matrix, then $A \circ A^T$ can be decomposed as a sum of CUT matrices. This is illustrated in Figure 1 and means that we can start by studying problem (2) on CUT matrices.

DEFINITION 2.5. *For $u, v \in [1, n]$, we call $CUT(u, v)$ the matrix such that*

$$CUT(u, v)_{ij} = \begin{cases} 1 & \text{if } u \le i \le v \text{ and } u \le j \le v, \\ 0 & \text{otherwise;} \end{cases}$$

i.e., $CUT(u, v)$ is symmetric and block diagonal and has one square block equal to one.

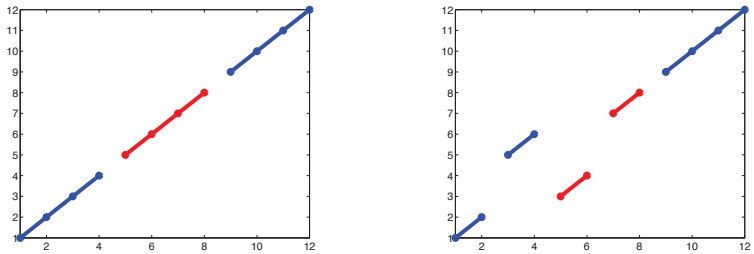
We first show that the objective of (2) has a natural interpretation when A is a CUT matrix as the variance of a subset of y under a uniform probability measure.

LEMMA 2.6. *Suppose $A = CUT(u, v)$; then*

$$f(y) = \sum_{i,j=1}^{n} A_{ij}(y_i - y_j)^2 = (v - u + 1)^2 \, \mathbf{var}(y_{[u,v]}).$$

Proof. We can write $\sum_{ij} A_{ij}(y_i - y_j)^2 = y^T L_A y$, where $L_A = \mathrm{diag}(A\mathbf{1}) - A$ is the Laplacian of A, which is a block matrix with a single nonzero block equal to $(v - u + 1)\delta_{\{i=j\}} - 1$ for $u \le i, j \le v$. ∎

This last lemma shows that solving the seriation problem (2) for CUT matrices amounts to finding a subset of y of size $(u-v+1)$ with minimum variance. This is the key to all the results that follow. As illustrated in Figure 2, for CUT matrices and of course conic combinations of CUT matrices, monotonic sequences have lower variance than sequences where the ordering is broken, and the results that follow make this explicit. We now show a simple technical lemma about the impact of switching two coefficients in y on the objective of problem (2) when A is a CUT matrix.



$\mathbf{var}(y_{[5,8]}) = \mathbf{1.6}$ $\mathbf{var}(y_{[5,8]}) = \mathbf{5.6}$

FIG. 2. *Objective values of 2-SUM problem (2) when $A = CUT(5,8)$ and $y_i = i$, $i = 1,\dots,12$. We plot the permuted values $y_{\pi(i)}$ against i, linking consecutive values of y both inside and outside the interval $[5,8]$. The solution on the left, where the values of $y_{[5,8]}$ are consecutive, has $\mathbf{var}(y_{[5,8]}) = 1.6$, while $\mathbf{var}(y_{[5,8]}) = 5.6$ on the right, where there is a gap between y_6 and y_7. Minimizing the 2-SUM objective for CUT matrices, i.e., the variance of a subset of the coefficients of y, tends to pull the coefficients in this subset together.*

LEMMA 2.7. *Let $A \in \mathbf{S}_n$, let $y \in \mathbb{R}^n$, and let $f(\cdot)$ be the objective of problem (2). Suppose we switch the values of y_j and y_{j+1} calling the new vector z; then we have*

$$f(y) - f(z) = 4 \sum_{\substack{i=1 \\ i\neq j,\, i\neq j+1}}^{n} \left(\frac{y_j + y_{j+1}}{2} - y_i \right) (y_{j+1} - y_j)(A_{ij+1} - A_{ij}).$$

Proof. Because A is symmetric, we have

$$\begin{aligned}
(f(y) - f(z))/2 = & \sum_{i\neq j,\, i\neq j+1} A_{ij}(y_i - y_j)^2 + \sum_{i\neq j,\, i\neq j+1} A_{ij+1}(y_i - y_{j+1})^2 \\
& - \sum_{i\neq j,\, i\neq j+1} A_{ij}(y_i - y_{j+1})^2 - \sum_{i\neq j,\, i\neq j+1} A_{ij+1}(y_i - y_j)^2 \\
= & \sum_{i\neq j,\, i\neq j+1} 2A_{ij}(y_j - y_{j+1}) \left(\frac{y_j + y_{j+1}}{2} - y_i \right) \\
& + \sum_{i\neq j,\, i\neq j+1} 2A_{ij+1}(y_{j+1} - y_j) \left(\frac{y_j + y_{j+1}}{2} - y_i \right),
\end{aligned}$$

which yields the desired result. ∎

The next lemma characterizes optimal solutions of problem (2) for CUT matrices and shows that they split the coefficients of y in disjoint intervals.

LEMMA 2.8. *Suppose $A = CUT(u,v)$, and write $w = y_\pi$ as the optimal solution to (2). If we say that $I = [u,v]$ and I^c is its complement, then*

$$w_j \notin [\min(w_I), \max(w_I)] \quad \text{for all } j \in I^c;$$

in other words, the coefficients in w_I and w_{I^c} belong to disjoint intervals.

Proof. Without loss of generality, we can assume that the coefficients of w_I are sorted in increasing order. By contradiction, suppose that there is a w_j such that $j \in I^c$ and $w_j \notin [w_u, w_v]$. Suppose also that w is larger than the mean of coefficients inside I, i.e., $w_j \geq \sum_{i=u+1}^{v} w_i/(v-u)$. This, combined with our assumption that $w_j \leq w_v$ and Lemma 2.7, means that switching the values of w_j and w_v will decrease the objective by

$$4\sum_{i=u}^{v-1}\left(\frac{w_j + w_v}{2} - y_i\right)(w_v - w_j),$$

which is positive by our assumptions on w_j and the mean, which contradicts optimality. A symmetric result holds if w_j is smaller than the mean. ∎

This last lemma shows that when A is a CUT matrix, the monotonic vector $y_i = ai + b$, for $a, b \in \mathbb{R}$ and $i = 1, \ldots, n$, is always an optimal solution to the 2-SUM problem (2), since all subvectors of y of a given size have the same variance. This means that, when y is a permutation of $y_i = ai + b$, all minimization problems (2) written on CUT matrices have *a common optimal solution*, where y_π is monotonic.

2.3. Ordering P-, Q-, and R-matrices. Having shown that all 2-SUM problems (2) written on CUT matrices share a common monotonic solution, this section now shows how to decompose the square of P-, Q-, and R-matrices as a sum of CUT matrices and then links the reordering of a matrix with that of its square $A \circ A^T$. We will first show a technical lemma proving that if A is a Q-matrix, then $A \circ A^T$ is a conic combination of CUT matrices. The CUT decomposition of P- and R-matrices will then naturally follow, since P-matrices are just $\{0,1\}$ Q-matrices, and R-matrices are symmetric Q-matrices.

LEMMA 2.9. *Suppose $A \in \mathbb{R}^{n\times m}$ is a Q-matrix; then $A \circ A^T$ is a conic combination of CUT matrices.*

Proof. Suppose $a \in \mathbb{R}^n$ is a unimodal vector; let us show that the matrix $M = a \circ a^T$ with coefficients $M_{ij} = \min\{a_i, a_j\}$ is a conic combination of CUT matrices. Let $I = \{j : j = \operatorname{argmax}_i a_i\}$; then I is an index interval $[I_{\min}, I_{\max}]$ because a is unimodal. Denoting $\bar{a} = \max_i a_i$ and $b = \max_{i\in I^c} a_i$ (with $b = 0$ when $I^c = \emptyset$), the deflated matrix

$$M^- = M - (\bar{a} - b)\ CUT(I_{\min}, I_{\max})$$

can be written as $M^- = a^- \circ (a^-)^T$ with

$$a^- = a - (\bar{a} - b)v,$$

where $v_i = 1$ iff $i \in I$. By construction, $|\operatorname{argmax} M^-| > |I|$, i.e., the size of $\operatorname{argmax} M$ increases by at least one, so this deflation procedure ends after at most n iterations. This shows that $a \circ a^T$ is a conic combination of CUT matrices when a is unimodal. Now, we have $(A \circ A^T)_{ij} = \sum_{k=1}^{n} w_k \min\{A_{ik}, A_{jk}\}$, so $A \circ A^T$ is a sum of n matrices of the form $\min\{A_{ik}, A_{jk}\}$ where each column is unimodal; hence we get the desired result. ∎

This last result also shows that, when A is a Q-matrix, $A \circ A^T$ is an R-matrix as a sum of CUT matrices, which is illustrated in Figure 1. We now recall the central result in [24, Thm. 1] showing that for Q-matrices, reordering $A \circ A^T$ also reorders A.

THEOREM 2.10 (see [24, Thm. 1]). *Suppose $A \in \mathbb{R}^{n\times m}$ is pre-Q; then ΠA is a Q-matrix iff $\Pi(A \circ A^T)\Pi^T$ is an R-matrix.*

We use these last results to show that at least for some vectors y, if C is a Q-matrix, then the 2-SUM problem (2) written on $A = C \circ C^T$ has a monotonic solution y_π.

PROPOSITION 2.11. *Suppose $C \in \mathbb{R}^{n\times m}$ is a pre-Q matrix and $y_i = ai + b$ for $i = 1, \ldots, n$ and $a, b \in \mathbb{R}$ with $a \neq 0$. Let $A = C \circ C^T$; if Π is such that $\Pi A \Pi^T$ is an R-matrix, then the corresponding permutation π solves the combinatorial minimization problem* (2).

Proof. If $C \in \mathbb{R}^{n\times m}$ is pre-Q, then Lemma 2.9 and Theorem 2.10 show that there is a permutation Π such that $\Pi(C \circ C^T)\Pi^T$ is a sum of CUT matrices (hence an R-matrix). Now all monotonic subsets of y of a given length have the same variance, and hence Lemmas 2.6 and 2.8 show that π solves problem (2). ∎

We now show that when the R-constraints are strict, the converse is also true; i.e., for matrices that are the square of Q-matrices, if y_π solves the 2-SUM problem (2), then π makes A an R-matrix. In the next section, we will use this result to reorder pre-R matrices (with noise and additional structural constraints) by solving convex relaxations to the 2-SUM problem.

PROPOSITION 2.12. *Suppose A is a pre-R matrix that can be written as $A = C \circ C^T$, where $C \in \mathbb{R}^{n\times m}$ is a pre-Q matrix, $y_i = ai + b$ for $i = 1, \ldots, n$, and $a, b \in \mathbb{R}$ with $a \neq 0$. Suppose, moreover, that A has strict R-constraints, i.e., the rows/columns of A are strictly unimodal after reordering. If the permutation π solves the 2-SUM problem* (2)*, then the corresponding permutation matrix Π is such that $\Pi A \Pi^T$ is an R-matrix.*

Proof. We can assume that A is an R-matrix without loss of generality. We will show that the identity is optimal for 2-SUM and that it is the unique such solution; hence solving 2-SUM solves seriation. Lemma 2.9 shows that A is a conic combination of CUT matrices. Moreover, by Proposition 2.11 the identity matrix solves problem (2). Following the proof of Proposition 2.11, the identity matrix is also optimal for each seriation subproblem on the CUT matrices of A.

Now remark that since the R-constraints are strict on the first column of A, there must be $n-2$ CUT matrices of the form $A_i = CUT(1, i)$ for $i = 2, \ldots, n-1$ in the decomposition of A (otherwise, there would be some index $k > 1$ for which $A_{1k} = A_{1k+1}$, which would contradict our strict unimodal assumption). Following the previous remarks, the identity matrix is optimal for all the seriation subproblems in A_i, which means that the variance of all the corresponding subvectors of y_π, i.e., $(y_{\pi(1)}, y_{\pi(2)})$, $(y_{\pi(1)}, y_{\pi(2)}, y_{\pi(3)}), \ldots, (y_{\pi(1)}, \ldots, y_{\pi(n-1)})$, must be minimized. Since these subvectors of y_π are monotonically embedded, up to a permutation of $y_{\pi(1)}$ and $y_{\pi(2)}$, Lemma 2.8 shows that this can only be achieved for contiguous $y_{\pi(i)}$, that is, for π equal to the identity or the reverse permutation. Indeed, to minimize the variance of $(y_{\pi(1)}, \ldots, y_{\pi(n-1)})$, we have to choose $\pi(n) = n$ or $\pi(n) = 1$. Then to minimize the variance of $(y_{\pi(1)}, \ldots, y_{\pi(n-2)})$ we have to choose $\pi(n-1) = n-1$ or $\pi(n-1) = 2$. Thus we get by induction $\pi(i) = i$ or $\pi(i) = n - i + 1$ for $i = 3, \ldots, n$. Finally, there are only two permutations left for $y_{\pi(1)}$ and $y_{\pi(2)}$. Since $A_{31} < A_{32}$, we have to choose $(y_{\pi(3)} - y_{\pi(1)})^2 > (y_{\pi(3)} - y_{\pi(2)})^2$, and the remaining ambiguity on the order of $y_{\pi(1)}$ and $y_{\pi(2)}$ is removed. ∎

These results shows that if A is pre-R and can be written as $A = C \circ C^T$ with C pre-Q, then the permutation that makes A an R-matrix also solves the 2-SUM problem (2). Conversely, when A is pre-R (strictly), the permutation that solves (2) reorders A as an R-matrix. Since [2] show that sorting the Fiedler vector also orders A as an R-matrix, Proposition 2.11 gives a polynomial time solution to the 2-SUM

problem (2) when A is pre-R with $A = C \circ C^T$ for some pre-Q matrix C. Note that the strict monotonicity constraints on the R-matrix can be somewhat relaxed (we only need one strictly monotonic column plus two more constraints), but requiring strict monotonicity everywhere simplifies the argument.

3. Convex relaxations. In the sections that follow, we will use the combinatorial results derived above to produce convex relaxations of optimization problems written over the set of permutation matrices. We mostly focus on the 2-SUM problem in (2); however, many of the results below can be directly adapted to other objective functions. We detail several convex approximations, some new and some taken from the computer science literature, ranked by increasing numerical complexity. Without loss of generality, we always assume that the weight matrix A is nonnegative (if A has negative entries, it can be shifted to become nonnegative, with no impact on the permutation problem). The nonnegativity assumption is in any case natural since A represents a similarity matrix in the seriation problem.

3.1. Spectral ordering. We first recall classical definitions from spectral clustering and briefly survey the spectral ordering results in [2] in the noiseless setting.

DEFINITION 3.1. *The Fiedler value of a symmetric, nonnegative matrix A is the smallest nonzero eigenvalue of its Laplacian $L_A = \mathrm{diag}(A\mathbf{1}) - A$. The corresponding eigenvector is called the Fiedler vector and is the optimal solution to*

$$\begin{array}{ll} \textit{minimize} & y^T L_A y \\ \textit{subject to} & y^T \mathbf{1} = 0, \|y\|_2 = 1 \end{array} \tag{3}$$

in the variable $y \in \mathbb{R}^n$.

We now recall the main result from [2], which shows how to reorder pre-R matrices in a noise-free setting.

PROPOSITION 3.2 (see [2, Thm. 3.3]). *Suppose $A \in \mathbf{S}_n$ is a pre-R-matrix, with a simple Fiedler value whose Fiedler vector v has no repeated values. Suppose that Π is a permutation matrix such that the permuted Fielder vector Πv is strictly monotonic; then $\Pi A \Pi^T$ is an R-matrix.*

We now extend the result of Proposition 2.11 to the case where the weights y are given by the Fiedler vector.

PROPOSITION 3.3. *Suppose $A \in \mathbf{S}^{n\times n}$ is an R-matrix and y is its Fiedler vector. Then the identity permutation solves the 2-SUM problem* (2).

Proof. The combinatorial problem (2) can be rewritten as

$$\begin{array}{ll} \text{minimize} & y^T \Pi^T L_A \Pi y \\ \text{subject to} & \Pi \in \mathcal{P}, \end{array}$$

which is also equivalent to

$$\begin{array}{ll} \text{minimize} & z^T L_A z \\ \text{subject to} & z^T \mathbf{1} = 0, \|z\|_2 = 1, z = \Pi y, \Pi \in \mathcal{P}, \end{array}$$

since y is the Fiedler vector of A. By dropping the constraints $z = \Pi y, \Pi \in \mathcal{P}$, we can relax the last problem into (3), whose solution is the Fiedler vector of A. Note that the optimal value of problem (2) is thus an upper bound on that of its relaxation (3), i.e., the Fiedler value of A. This upper bound is attained by the Fiedler vector, i.e., the optimum of (3); therefore the identity matrix is an optimal solution to (2). ∎

Using the fact that the Fiedler vector of an R-matrix is monotonic [2, Thm. 3.2], the next corollary immediately follows.

COROLLARY 3.4. *If A is a pre-R matrix such that $\Pi^T A \Pi$ is an R-matrix, then π is an optimal solution to problem* (2) *when y is the Fiedler vector of A sorted in increasing order.*

The results in [2] thus provide a polynomial time solution to the R-matrix ordering problem in a noise-free setting (extremal eigenvalues of dense matrices can be computed by randomized polynomial time algorithms with complexity $O(n^2 \log n)$ [25]). While [2] also shows how to handle cases where the Fiedler vector is degenerate, these scenarios are highly unlikely to arise in settings where observations on A are noisy, and we refer the reader to [2, sect. 4] for details.

3.2. QP relaxation. In most applications, A is typically noisy and the pre-R assumption no longer holds. The spectral solution is stable when the magnitude of the noise remains within the spectral gap (i.e., in a perturbative regime [43]). Beyond that, while the Fiedler vector of A can still be used as a heuristic to find an approximate solution to (2), there is no guarantee that it will be optimal.

The results in section 2 made the connection between the spectral ordering in [2] and the 2-SUM problem (2). In what follows, we will use convex relaxations to (2) to solve matrix ordering problems in a noisy setting. We also show in section 3.2.3 how to incorporate a priori knowledge on the true ordering in the formulation of the optimization problem to solve semisupervised seriation problems. Numerical experiments in section 5 show that semisupervised seriation solutions are sometimes significantly more robust to noise than the spectral solutions ordered from the Fiedler vector.

3.2.1. Permutations and doubly stochastic matrices. We write $\mathcal{D}_n$ as the set of doubly stochastic matrices, i.e., $\mathcal{D}_n = \{X \in \mathbb{R}^{n \times n} : X \geqslant 0, X\mathbf{1} = \mathbf{1}, X^T\mathbf{1} = \mathbf{1}\}$. Note that $\mathcal{D}_n$ is convex and polyhedral. Classical results show that the set of doubly stochastic matrices is the convex hull of the set of permutation matrices. We also have that $\mathcal{P} = \mathcal{D} \cap \mathcal{O}$, i.e., a matrix is a permutation matrix iff it is both doubly stochastic and orthogonal. The fact that $L_A \succeq 0$ means that we can directly write a convex relaxation to the combinatorial problem (2) by replacing $\mathcal{P}$ with its convex hull $\mathcal{D}_n$, to get

$$\begin{array}{ll} \text{minimize} & g^T \Pi^T L_A \Pi g \\ \text{subject to} & \Pi \in \mathcal{D}_n, \end{array} \tag{4}$$

where $g = (1, \ldots, n)$, in the permutation matrix variable $\Pi \in \mathcal{P}$. By symmetry, if a vector Πy minimizes (4), then the reverse vector also minimizes (4). This often has a significant negative impact on the quality of the relaxation, and we add the linear constraint $e_1^T \Pi g + 1 \leq e_n^T \Pi g$ to break symmetries, which means that we always pick solutions where the first element comes before the last one. Because the Laplacian L_A is positive semidefinite, problem (4) is a convex quadratic program in the variable $\Pi \in \mathbb{R}^{n \times n}$ and can be solved efficiently. To produce approximate solutions to problem (2), we then generate permutations from the doubly stochastic optimal solution to the relaxation in (4) (we will describe an efficient procedure to do so in section 3.2.4).

The results of section 2 show that the optimal solution to (2) also solves the seriation problem in the noiseless setting when the matrix A is of the form $C \circ C^T$ with C a Q-matrix and y is an affine transform of the vector $(1, \ldots, n)$. These results also hold empirically for small perturbations of the vector y, and to improve robustness

to noisy observations of A, we average several values of the objective of (4) over these perturbations, solving

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(Y^T\Pi^T L_A \Pi Y)/p \\ \text{subject to} & e_1^T\Pi g + 1 \leq e_n^T \Pi g, \\ & \Pi\mathbf{1} = \mathbf{1},\ \Pi^T\mathbf{1} = \mathbf{1},\ \Pi \geq 0 \end{array} \tag{5}$$

in the variable $\Pi \in \mathbb{R}^{n\times n}$, where $Y \in \mathbb{R}^{n\times p}$ is a matrix whose columns are small perturbations of the vector $g = (1, \ldots, n)^T$. Solving (5) is roughly p times faster than individually solving p versions of (4).

3.2.2. Regularized QP relaxation. In the previous section, we have relaxed the combinatorial problem (2) by relaxing the set of permutation matrices into the set of doubly stochastic matrices. As the set of permutation matrices $\mathcal{P}$ is the intersection of the set of doubly stochastic matrices $\mathcal{D}$ and the set of orthogonal matrices $\mathcal{O}$, i.e., $\mathcal{P} = \mathcal{D} \cap \mathcal{O}$, we can add a penalty to the objective of the convex relaxed problem (5) to force the solution to get closer to the set of orthogonal matrices. Since a doubly stochastic matrix of Frobenius norm $\sqrt{n}$ is necessarily orthogonal, we would ideally like to solve

$$\begin{array}{ll} \text{minimize} & \frac{1}{p}\mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) - \frac{\mu}{p}\|\Pi\|_F^2 \\ \text{subject to} & e_1^T\Pi g + 1 \leq e_n^T \Pi g, \\ & \Pi\mathbf{1} = \mathbf{1},\ \Pi^T\mathbf{1} = \mathbf{1},\ \Pi \geq 0, \end{array} \tag{6}$$

with μ large enough to guarantee that the global solution is indeed a permutation. However, this problem is not convex for any $\mu > 0$ since its Hessian is not positive semidefinite. Note that the objective of (5) can be rewritten as $\mathrm{Vec}(\Pi)^T(YY^T \otimes L_A)\mathrm{Vec}(\Pi)/p$; so the Hessian here is $YY^T \otimes L_A - \mu I \otimes I$ and is never positive semidefinite when $\mu > 0$ since the first eigenvalue of L_A is always zero. Instead, we propose a slightly modified version of (6), which has the same objective function up to a constant and is convex for some values of μ. Recall that the Laplacian matrix L_A is always positive semidefinite with at least one eigenvalue equal to zero corresponding to the eigenvector $\mathbf{1}/\sqrt{n}$ (strictly one if the graph is connected), and let $P = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$.

PROPOSITION 3.5. *The optimization problem*

$$\begin{array}{ll} \textit{minimize} & \frac{1}{p}\mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) - \frac{\mu}{p}\|P\Pi\|_F^2 \\ \textit{subject to} & e_1^T\Pi g + 1 \leq e_n^T \Pi g, \\ & \Pi\mathbf{1} = \mathbf{1},\ \Pi^T\mathbf{1} = \mathbf{1},\ \Pi \geq 0 \end{array} \tag{7}$$

is equivalent to problem (6)*; their objectives differ by a constant. Furthermore, when* $\mu \leq \lambda_2(L_A)\lambda_1(YY^T)$*, this problem is convex.*

Proof. Let us first remark that

$$\begin{aligned} \|P\Pi\|_F^2 &= \mathbf{Tr}(\Pi^T P^T P \Pi) = \mathbf{Tr}(\Pi^T P \Pi) \\ &= \mathbf{Tr}(\Pi^T (I - \mathbf{1}\mathbf{1}^T/n)\Pi) = \mathbf{Tr}(\Pi^T\Pi - \mathbf{1}\mathbf{1}^T/n) \\ &= \mathbf{Tr}(\Pi^T\Pi) - 1, \end{aligned}$$

where we used the fact that P is the (symmetric) projector matrix onto the orthogonal complement of $\mathbf{1}$ and Π is doubly stochastic (so $\Pi\mathbf{1} = \Pi^T\mathbf{1} = \mathbf{1}$). We deduce that problem (7) has the same objective function as (6) up to a constant. Moreover, it is

convex when $\mu \leq \lambda_2(L_A)\lambda_1(YY^T)$ since the Hessian of the objective is given by

$$-\mathcal{A} = \frac{1}{p} YY^T \otimes L_A - \frac{\mu}{p} \mathbf{I} \otimes P \tag{8}$$

and the eigenvalues of $YY^T \otimes L_A$, which are equal to $\lambda_i(L_A)\lambda_j(YY^T)$ for all i, j in $\{1, \ldots, n\}$, are all superior or equal to the eigenvalues of $\mu\mathbf{I} \otimes P$ which are all smaller than μ. □

To have μ strictly positive, we need YY^T to be definite, which can be achieved w.h.p. by setting p higher than n and sampling independent vectors y. The key motivation for including several monotonic vectors y in the objective of (7) is to increase the value of $\lambda_1(YY^T)$. The higher this eigenvalue, the stronger the effect of the regularization term in (7), which in turn improves the quality of the solution (all of this being somewhat heuristic of course). The problem of generating good matrices Y with both monotonic columns and high values of $\lambda_1(YY^T)$ is not easy to solve, however, and hence we use randomization to generate Y.

3.2.3. Semisupervised problems. The QP relaxation above allows us to add structural constraints to the problem. For instance, in archeological applications, one may specify that observation i must appear before observation j, i.e., $\pi(i) < \pi(j)$. In gene sequencing applications, one may constrain the distance between two elements (e.g., mate reads), which would be written as $a \leq \pi(i) - \pi(j) \leq b$, and introduce an affine inequality on the variable Π in the QP relaxation of the form $a \leq e_i^T \Pi g - e_j^T \Pi g \leq b$. Linear constraints could also be extracted from a reference gene sequence. More generally, we can rewrite problem (7) with n_c additional linear constraints as follows:

$$\begin{array}{ll} \text{minimize} & \frac{1}{p}\mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) - \frac{\mu}{p}\|P\Pi\|_F^2 \\ \text{subject to} & D^T\Pi g + \delta \leq 0, \\ & \Pi\mathbf{1} = \mathbf{1},\ \Pi^T\mathbf{1} = \mathbf{1},\ \Pi \geq 0, \end{array} \tag{9}$$

where D is a matrix of size $n \times (n_c + 1)$ and δ is a vector of size n_c. The first column of D is equal to $e_1 - e_n$ and $\delta_1 = 1$ (to break symmetry).

3.2.4. Sampling permutations from doubly stochastic matrices. This procedure is based on the fact that a permutation can be defined from a doubly stochastic matrix S by the order induced on a monotonic vector. A similar argument was used in [4] to round orthogonal matrices into permutations. Suppose we generate a *monotonic* random vector v and compute Sv. To each v we can associate a permutation Π such that ΠSv is monotonically increasing. If S is a permutation matrix, then the permutation Π generated by this procedure will be constant; if S is a doubly stochastic matrix but not a permutation, then it might fluctuate. Starting from a solution S to problem (7), we can use this procedure to sample many permutation matrices Π, and we pick the one with lowest cost $g^T\Pi^T L_A \Pi g$ in the combinatorial problem (2). We could also project S on permutations using the Hungarian algorithm, but this proved more costly and less effective in our experiments.

4. Algorithms. The convex relaxation in (9) is a quadratic program in the variable $\Pi \in \mathbb{R}^{n \times n}$, which has dimension n^2. For reasonable values of n (around a few hundred), interior point solvers such as MOSEK [1] solve this problem very efficiently (the experiments in this paper were performed using this library). Furthermore, most pre-R matrices formed by squaring pre-Q matrices are very sparse, which considerably speeds up linear algebra. However, first-order methods remain the only alternative

for solving (9) beyond a certain scale. We briefly discuss below the implementation of two classes of methods: the conditional gradient (a.k.a. Frank–Wolfe) algorithm, and accelerated gradient methods. Alternatively, an extended formulation of the permutahedron using only $O(n \log n)$ variables and constraints was produced in [20], which can be used to write QP relaxations of 2-SUM with only $O(n \log n)$ variables. While the constant in these representations is high, more practical formulations are available with $O(n \log^2 n)$ variables. This formulation was tested by [29] while this paper was under review, and combined with an efficient interior point solver (GUROBI) it provides significant speedup.

4.1. Conditional gradient. Solving (9) using the conditional gradient algorithm in, e.g., [15] requires minimizing an affine function over the set of doubly stochastic matrices at each iteration. This amounts to solving a classical transportation (or matching) problem for which very efficient solvers exist [36].

4.2. Accelerated smooth optimization. On the other hand, solving (9) using accelerated gradient algorithms requires solving a projection step on doubly stochastic matrices at each iteration [35]. Here too, exploiting structure significantly improves the complexity of these steps. Given some matrix Π_0, the Euclidean projection problem is written as

$$\begin{array}{lll} & \text{minimize} & \frac{1}{2}\|\Pi - \Pi_0\|_F^2 \\ (10) & \text{subject to} & D^T \Pi g + \delta \leq 0, \\ & & \Pi \mathbf{1} = \mathbf{1},\ \Pi^T \mathbf{1} = \mathbf{1},\ \Pi \geq 0 \end{array}$$

in the variable $\Pi \in \mathbb{R}^{n \times n}$, with parameter $g \in \mathbb{R}^n$. The dual is written as

$$\begin{array}{lll} & \text{maximize} & -\frac{1}{2}\|x\mathbf{1}^T + \mathbf{1}y^T + Dzg^T - Z\|_F^2 - \mathbf{Tr}(Z^T \Pi_0) \\ (11) & & +x^T(\Pi_0 \mathbf{1} - \mathbf{1}) + y^T(\Pi_0^T \mathbf{1} - \mathbf{1}) + z(D^T \Pi_0 g + \delta) \\ & \text{subject to} & z \geq 0,\ Z \geq 0 \end{array}$$

in the variables $Z \in \mathbb{R}^{n \times n}$, $x, y \in \mathbb{R}^n$, and $z \in \mathbb{R}^{n_c}$. The dual optimizes over decoupled linear constraints in $(z,\ Z)$, while x and y are unconstrained.

Each subproblem is equivalent to computing a conjugate norm and can be solved in closed form. This means that, with independent constraints (D full rank), at each iteration, explicit formulas are available to update variables block by block in the dual Euclidean projection problem (11) over doubly stochastic matrices (cf. Algorithm 1). Problem (11) can thus be solved very efficiently by block coordinate ascent, whose convergence is guaranteed in this setting [5], and a solution to (10) can be reconstructed from the optimum in (11).

The detailed procedure for block coordinate ascent in the dual Euclidean projection problem (11) is described in Algorithm 1. We perform block coordinate ascent until the duality gap between the primal and the dual objectives is below the required precision. Warm-starting the projection step in both the primal and the dual objectives provided a very significant speedup in our experiments.

5. Applications and numerical experiments. We now study the performance of the relaxations detailed above in some classical applications of seriation. Other applications not discussed here include social networks, sociology, cartography, ecology, operations research, and psychology [28].

In most of the examples below, we will compare the performance of the spectral solution to that of the QP relaxation in (7) and the semisupervised seriation QP in (9).

Algorithm 1. Projection on doubly stochastic matrices.

Input: A matrix $Z \in \mathbb{R}_+^{n \times n}$, a vector $z \in \mathbb{R}_+^{n_c}$, two vectors $x, y \in \mathbb{R}^n$, a target precision ϵ, a maximum number of iterations N.

1: Set $k = 0$.
2: **while** duality gap $> \epsilon$ and $k \leq N$ **do**
3: Update dual variables

$$\begin{cases} Z = \max\{\mathbf{0},\ x\mathbf{1}^T + \mathbf{1}y^T + Dzg^T - \Pi_0\}, \\ x = \frac{1}{n}(\Pi_0\mathbf{1} - (y^T\mathbf{1}+1)\mathbf{1} - Dzg^T\mathbf{1} + Z\mathbf{1}), \\ y = \frac{1}{n}(\Pi_0^T\mathbf{1} - (x^T\mathbf{1}+1)\mathbf{1} - gz^TD\mathbf{1} + Z^T\mathbf{1}), \\ z = \frac{1}{\|g\|_2^2}\max\{0,\ (D^TD)^{-1}(D^T(Z+\Pi_0)g + \delta - D^Txg^T\mathbf{1} - D^T\mathbf{1}g^Ty)\}. \end{cases}$$

4: Set $k = k + 1$.
5: **end while**

Output: A doubly stochastic matrix Π.

In the semisupervised experiments, we randomly sample pairwise orderings either from the true order information (if known) or from noisy ordering information. We use a simple symmetric Erdös–Rényi model for collecting these samples, so that a pair of indices (i, j) is included with probability p, with orderings sampled independently. In [11] it is shown that there is a sharp phase transition in the connectivity of the sampled graphs, with the graphs being almost surely disconnected when $p < \frac{(1-\epsilon)\log n}{n}$ and almost surely connected when $p > \frac{(1+\epsilon)\log n}{n}$ for $\epsilon > 0$ and n large enough. Above that threshold, i.e., when $O(n \log n)$ pairwise orders are specified, the graph is fully connected, so the full variable ordering is specified *if the ordering information is noiseless*. Of course, when the samples include errors, some of the sampled pairwise orderings could be inconsistent, so the total order is not fully specified.

5.1. Archeology. We reorder the rows of the Hodson's Munsingen dataset (as provided by [22] and manually ordered by [24]) to date 59 graves from 70 recovered artifact types (under the assumption that graves from similar periods contain similar artifacts). The results are reported in Table 1. We use a fraction of the pairwise orders in [24] to solve the semisupervised version. Note that the original data contains errors, so Kendall's ordering cannot be fully consistent. In fact, we will see that the semisupervised relaxation actually improves on Kendall's manual ordering.

In Figure 3 the first plot on the left shows the row ordering on 59×70 graves by artifacts matrix given by Kendall, the middle plot is the Fiedler solution, and the plot on the right is the best QP solution from 100 experiments with different Y (based on the combinatorial objective in (2)). The quality of these solutions is detailed in Table 1.

5.2. Markov chains. Here, we observe many *disordered* samples from a Markov chain. The mutual information matrix of these variables must be decreasing with $|i-j|$ when ordered according to the true generating Markov chain (this is the "data processing inequality" in [8, Thm. 2.8.1]); hence the mutual information matrix of these variables is a pre-R matrix. We can thus recover the order of the Markov chain by solving the seriation problem on this matrix. In the following example, we try to recover the order of a Gaussian Markov chain written as $X_{i+1} = b_iX_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma_i^2)$. The results are presented in Table 2 on 30 variables. We test

TABLE 1

Performance metrics (median and stdev over 100 *runs of the QP relaxation) for Kendall's* τ*, Spearman's* ρ *ranking correlations (large values are good), the objective value in* (2)*, and the number of R-matrix monotonicity constraint violations (small values are good), comparing Kendall's original solution with that of the Fiedler vector, the seriation QP in* (7)*, and the semisupervised seriation QP in* (9) *with* 0.1% *and* 47.5% *pairwise ordering constraints specified. Note that the semisupervised solution actually improves on both Kendall's manual solution and on the spectral ordering.*

	[24]	Spectral	QP Reg	QP Reg + 0.1%	QP Reg + 47.5%
Kendall τ	1.00	0.75	0.73±0.22	0.76±0.16	0.97±0.01
Spearman ρ	1.00	0.90	0.88±0.19	0.91±0.16	1.00±0.00
Comb. obj.	38520	38903	41810±13960	43457±23004	37602±775
# R-constr.	1556	1802	2021±484	2050±747	1545±43

FIG. 3. *The Hodson's Munsingen dataset: row ordering given by Kendall* (left)*, Fiedler solution* (center)*, and best unsupervised QP solution from* 100 *experiments with different* Y *based on the combinatorial objective* (right)*.*

TABLE 2

Kendall's τ *between the true Markov chain ordering, the Fiedler vector, the seriation QP in* (7)*, and the semisupervised seriation QP in* (9) *with varying numbers of pairwise orders specified. We observe the (randomly ordered) model covariance matrix (no noise), the sample covariance matrix with enough samples so the error is smaller than half of the spectral gap, and then a sample covariance computed using many fewer samples so the spectral perturbation condition fails.*

	No noise	Noise within spectral gap	Large noise
True	1.00±0.00	1.00±0.00	1.00±0.00
Spectral	1.00±0.00	0.86±0.14	0.41±0.25
QP Reg	0.50±0.34	0.58±0.31	0.45±0.27
QP + 0.2%	0.65±0.29	0.40±0.26	0.60±0.27
QP + 4.6%	0.71±0.08	0.70±0.07	0.68±0.08
QP + 54.3%	0.98±0.01	0.97±0.01	0.97±0.02

performance in a noise-free setting where we observe the randomly ordered model covariance, in a noisy setting with enough samples (6000) to ensure that the spectral solution stays in a perturbative regime, and finally using many fewer samples (60) so the spectral perturbation condition fails. In Figure 4, the first plot on the left shows the true Markov chain order, the middle plot is the Fiedler solution, and the plot on the right is the best QP solution from 100 experiments with different Y (based on the combinatorial objective).

5.3. Gene sequencing. In next generation shotgun genome sequencing experiments, DNA strands are cloned about 10 to 100 times before being decomposed into very small subsequences called "reads," each of them 50 to a few hundred base pairs long. Current machines can only accurately sequence these small reads, which must then be reordered by "assembly" algorithms, using the overlaps between reads. These short reads are often produced in pairs, starting from both ends of a longer sequence of known length, and hence a rough estimate of the distance between these "mate

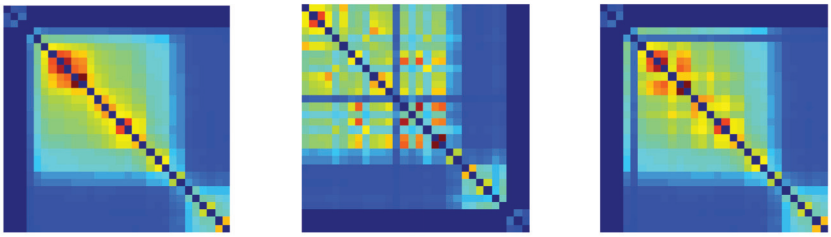


FIG. 4. *Markov Chain experiments: true Markov chain order* (left), *Fiedler solution* (center), *and best unsupervised QP solution from* 100 *experiments with different* Y *based on the combinatorial objective* (right).

pairs" of reads is known, giving additional structural information on the semisupervised assembly problem.

Here, we generate artificial sequencing data by (uniformly) sampling reads from chromosome 22 of the human genome from NCBI; then we store k-mer hit versus read in a binary matrix C (a k-mer is a fixed sequence of k base pairs). If the reads are ordered correctly and have identical length, this matrix has the C1P, and hence we solve the C1P problem on the $\{0,1\}$-matrix whose rows correspond to k-mers hits for each read; i.e., the element (i,j) of the matrix is equal to one if k-mer j is present in read i. The corresponding pre-R matrix obtained, CC^T, which measures overlap between reads, is extremely sparse, as it is approximately band-diagonal with roughly constant bandwidth b when reordered correctly, and computing the Fiedler vector can be done with complexity $O(bn \log n)$ w.h.p. using the Lanczos method [25], as it amounts to computing the second largest eigenvector of $\lambda_n(L)\mathbf{I} - L$, where L is the Laplacian of the matrix. In our experiments, computing the Fiedler vector from 250000 reads takes a few seconds using `eigs` from MATLAB on a standard desktop machine.

In practice, besides sequencing errors (handled relatively well by the high coverage of the reads), there are often repeats in long genomes. If the repeats are longer than the k-mers, the C1P assumption is violated and the order given by the Fiedler vector is no longer reliable. On the other hand, handling the repeats is possible using the information given by mate pairs, i.e., reads that are known to be separated by a given number of base pairs in the original genome. This structural knowledge can be incorporated into the relaxation (9). While our algorithm for solving (9) scales only up to a few thousand base pairs on a regular desktop, it can be used to solve the sequencing problem hierarchically, i.e., to refine the spectral solution.

In Figure 5, we show the result of spectral ordering on simulated reads from human chromosome 22. The full R-matrix formed by squaring the reads $\times$ k-mers matrix is too large to be plotted in MATLAB, and we zoom in on two diagonal block submatrices. In the first submatrix, the reordering is good and the matrix has very low bandwidth, and the corresponding gene segment (called contig) is well reconstructed. In the second, the reordering is less reliable, and the bandwidth is larger, so the reconstructed gene segment contains errors.

In Figure 6, we show recovered read position versus true read position for the Fiedler vector and the Fiedler vector followed by semisupervised seriation, where the QP relaxation is applied to groups of reads (contigs) assembled by the spectral solution, on the 250000 reads generated in our experiments. The spectral solution orders most of these reads correctly, which means that the relaxation is solved on a matrix

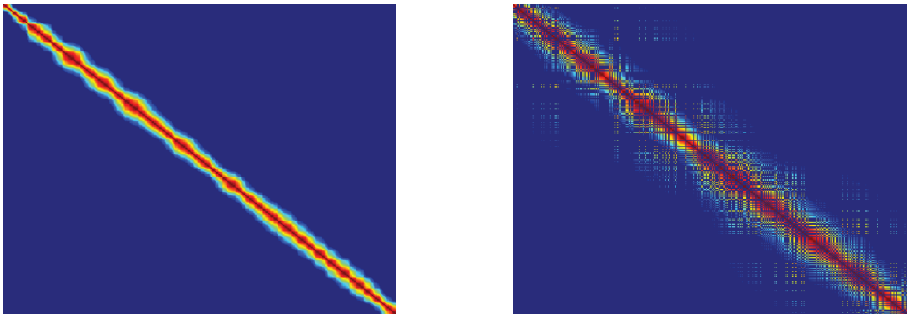


FIG. 5. *We plot the reads × reads matrix measuring the number of common k-mers between read pairs, reordered according to the spectral ordering on two submatrices.*

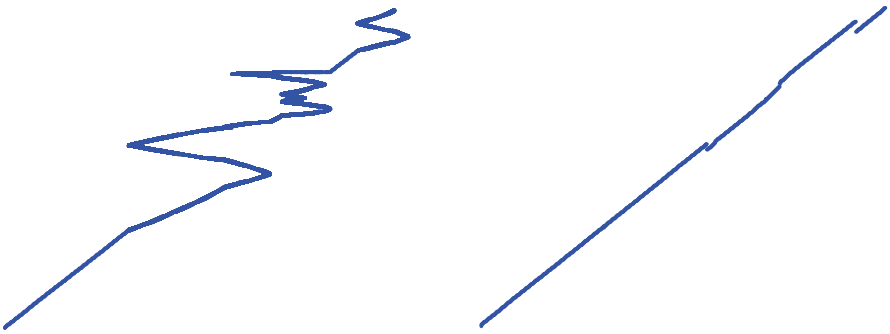


FIG. 6. *We plot the Fiedler and Fiedler+QP read orderings versus true ordering. The semisupervised solution contains many fewer misplaced reads.*

of dimension about 100. We see that the number of misplaced reads significantly decreases in the semisupervised seriation solution. Looking at the correlation between the true positions and the retrieved positions of the reads, both Kendall τ and Spearman ρ are equal to one for Fiedler+QP ordering, while they are equal to 0.87 and 0.96 for Fiedler ordering alone. A more complete description of the assembly algorithm is given in the appendix.

5.4. Generating Y. We conclude by testing the impact of Y on the performance of the QP relaxation in (6) on a simple ranking example. In Figure 7, we generate several matrices $Y \in \mathbb{R}^{n\times p}$ as in section 3.2.2 and compare the quality of the solutions (permutations issued from the procedure described in section 3.2.4) obtained for various values of the number of columns p. On the left, we plot the histogram of the values of $g^T\Pi^T L_A \Pi g$ obtained for 100 solutions with random matrices Y where $p = 1$ (i.e., rank one). On the right, we compare these results with the average value of $g^T\Pi^T L_A \Pi g$ for solutions obtained with random matrices Y with p varying from 1 to $5n$ (sample of 50 random matrices Y for each value of p). The red horizontal line represents the best solution obtained for $p = 1$ over all experiments. By raising the value of $\lambda_1(YY^T)$, larger values of p allow for higher values of μ in Proposition 3.5, which seems to have a positive effect on performance until a point where p is much larger than n and the improvement becomes insignificant. We do not have an intuitive explanation for this behavior at this point.

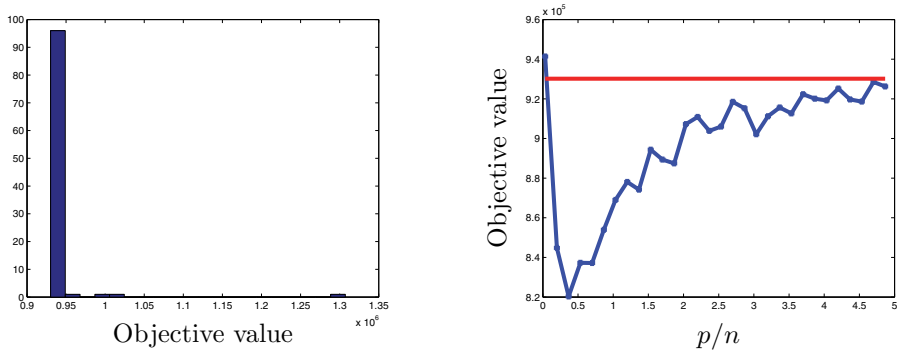


FIG. 7. Left: *we plot the histogram of the values of* $g^T\Pi^T L_A \Pi g$ *obtained for* 50 *solutions with random matrices* Y *where* $p = 1$ *(i.e., rank one).* Right: *we compare these results with the average value of* $g^T\Pi^T L_A \Pi g$ *for solutions obtained with random matrices* Y *with* p *varying from* 1 *to* $5n$. *The red horizontal line represents the best solution obtained for* $p = 1$ *over all experiments.*

6. Appendix. In this appendix, we first briefly detail two SDP based relaxations for the 2-SUM problem, one derived from results in [34, 9] and the other adapted from work on the minimum linear arrangement (MLA) problem in [12, 13, 6] among many others. While their complexity is effectively too high to make them practical seriation algorithms, these relaxations come with explicit approximation bounds which are not yet available for the QP relaxation in section 3.2. These SDP relaxations illustrate a clear tradeoff between approximation ratios and computational complexity: low complexity but unknown approximation ratio for the QP relaxation in (4), high complexity and $\sqrt{n}$ approximation ratio for the first semidefinite relaxation, and very high complexity but excellent $\sqrt{\log n}$ approximation ratio for the second SDP relaxation. The question of how to derive convex relaxations with nearly dimension independent approximation ratios (e.g., $O(\sqrt{\log n})$) and good computational complexity remains open at this point.

We then describe in detail the data sets and procedures used in the DNA sequencing experiments of section 5.

6.1. SDP relaxations and doubly stochastic matrices. Using randomization techniques derived from [34, 9], we can produce approximation bounds for a relaxation of the nonconvex QP representation of (2) derived in (7), namely

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) - \mu\|P\Pi\|_F^2 \\ \text{subject to} & \Pi\mathbf{1} = \mathbf{1},\ \Pi^T\mathbf{1} = \mathbf{1},\ \Pi \geq 0, \end{array}$$

which is a (possibly nonconvex) QP in the matrix variable $\Pi \in \mathbb{R}^{n\times n}$, where $P = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. We now set the penalty $\mu > 0$ sufficiently high to ensure that the objective is concave and the constraint $\|\Pi\| = \sqrt{n}$ is saturated. From Proposition 3.5 above, this means $\mu > \|L_A\|_2\|Y\|_2^2$. The solution of this concave minimization problem over the convex set of doubly stochastic matrices will then be at an extreme point, i.e., a permutation matrix. We first rewrite the above QP as a more classical maximization problem over vectors:

$$\begin{array}{ll} \text{maximize} & \left(\mathbf{vec}(\Pi)^T \mathcal{A}\, \mathbf{vec}(\Pi)\right)^{1/2} \\ \text{subject to} & (\mathbf{1}^T \otimes \mathbf{I})\, \mathbf{vec}(\Pi) = \mathbf{1},\ (\mathbf{I} \otimes \mathbf{1}^T)\, \mathbf{vec}(\Pi) = \mathbf{1},\ \Pi \geq 0. \end{array}$$

We use a square root in the objective here to maintain the same homogeneity properties as in the linear arrangement problems that follow. Because the objective is

constructed from a Laplacian matrix, we have $\mathbf{1}^T \mathcal{A}\mathbf{1} = 0$, so the objective is invariant by a shift in the variables. We now show that the equality constraints can be relaxed without loss of generality. We first recall a simple scaling algorithm due to [41] which shows how to normalize to one the row and column sums of a strictly positive matrix. Other algorithms based on geometric programming with explicit complexity bounds can be found in, e.g., [33].

Algorithm 2. Matrix scaling (Sinkhorn).

Input: A matrix $\Pi \in \mathbb{R}^{m \times n}$

1: **for** $k = 1$ to $N - 1$ **do**
2: Scale row sums to one: $\Pi_{k+1/2} = \mathrm{diag}(\Pi_k \mathbf{1})^{-1} \Pi_k$
3: Scale column sums to one: $\Pi_{k+1} = \Pi_{k+1/2} \, \mathrm{diag}(\mathbf{1}^T \Pi_{k+1/2})^{-1}$
4: **end for**

Output: A scaled matrix Π_N.

The next lemma shows that the only matrices satisfying both $\|\Pi\|_F = \sqrt{n}$ and $\Pi\mathbf{1} \leq \mathbf{1}$, $\Pi^T\mathbf{1} \leq \mathbf{1}$, with $\Pi \geq 0$ are doubly stochastic.

LEMMA 6.1. *Let $\Pi \in \mathbb{R}^{n \times n}$; if $\|\Pi\|_F = \sqrt{n}$ and $\Pi\mathbf{1} \leq \mathbf{1}$, $\Pi^T\mathbf{1} \leq \mathbf{1}$, with $\Pi \geq 0$, then Π is doubly stochastic.*

Proof. Suppose $\Pi\mathbf{1} \leq \mathbf{1}$, $\Pi^T\mathbf{1} \leq \mathbf{1}$, $\Pi > 0$; each iteration of Algorithm 2 multiplies $\mathbf{vec}(\Pi)$ by a diagonal matrix D with diagonal coefficients greater than one, with at least one coefficient strictly greater than one if Π is not doubly stochastic, and hence $\|\Pi\|_F$ is strictly increasing if Π is not doubly stochastic. This means that the only maximizers of $\|\Pi\|_F$ over the feasible set of (6.1) are doubly stochastic matrices. ∎

We let $z = \mathbf{vec}(\Pi)$; the above lemma means that problem (6.1) is equivalent to the following QP:

$$\text{(QP)} \qquad \begin{array}{ll} \text{maximize} & \|\mathcal{A}^{1/2} z\|_2 \\ \text{subject to} & (\mathbf{1}^T \otimes \mathbf{I}) z \leq \mathbf{1}, \, (\mathbf{I} \otimes \mathbf{1}^T) z \leq \mathbf{1}, \\ & z \geq 0 \end{array}$$

in the variable $z \in \mathbb{R}^{n^2}$. Furthermore, since permutation matrices are binary matrices, we can impose the redundant constraints that $z_i \in \{0, 1\}$ or equivalently $z_i^2 = z_i$ at the optimum. Lifting the quadratic objective and constraints as in [40, 30] yields the following relaxation:

$$\text{(SDP1)} \qquad \begin{array}{ll} \text{maximize} & \mathbf{Tr}(\mathcal{A} Z) \\ \text{subject to} & (\mathbf{1}^T \otimes \mathbf{I}) z \leq \mathbf{1}, \, (\mathbf{I} \otimes \mathbf{1}^T) z \leq \mathbf{1}, \\ & Z_{ii} = z_i, \, Z_{ij} \geq 0, \quad i, j = 1, \ldots, n, \\ & \begin{pmatrix} Z & z \\ z^T & 1 \end{pmatrix} \succeq 0, \end{array}$$

which is a semidefinite program in the matrix variable $Z \in \mathbf{S}_{n^2}$ and the vector $z \in \mathbb{R}^{n^2}$. By adapting a randomization argument used in the MaxCut relaxation bound in [34] and adapted to the k-dense-subgraph problem in [9], we can show the following $O(\sqrt{n})$ approximation bound on the quality of this relaxation.

PROPOSITION 6.2. *Let OPT be the optimal value of problem* (QP) *and SDP*1 *be that of* (SDP1)*; then*

$$0 \leq \frac{\mathbf{Tr}(\mathcal{A} G)}{4n} + \frac{SDP1}{2\pi n} \leq OPT^2 \leq SDP1$$

with $G_{ij} = \sqrt{Z_{ii}Z_{jj}}$, $i = 1, \ldots, n$, *and* $\mathbf{Tr}(\mathcal{A}G) \leq 0$.

Proof. The fact that $\mathcal{A} \succeq 0$ by construction shows that $0 \leq OPT^2 \leq SDP1$. Let $\xi \sim \mathcal{N}(0, Z)$, and define

$$y_i = \begin{cases} \sqrt{z_i} & \text{if } \xi_i \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

We write $C = \mathrm{diag}(Z)^{-1/2} Z \, \mathrm{diag}(Z)^{-1/2}$ as the correlation matrix associated with Z (under the convention that $C_{ij} = 0$ whenever $Z_{ii}Z_{jj} = 0$). A classical result from [39] (see also [23, p. 95]) shows that

$$\mathbf{E}[y_i y_j] = \sqrt{z_i z_j} \left(\tfrac{1}{4} + \tfrac{1}{2\pi} \arcsin(C) \right), \quad i = 1, \ldots, n,$$

and $\mathcal{A} \succeq 0$ together with $\arcsin(C) \succeq C$ (with the $\arcsin(\cdot)$ taken elementwise) and $z_i = Z_{ii}$ means that, writing $G_{ij} = \sqrt{z_i z_j} = \sqrt{Z_{ii}Z_{jj}}$, we get

$$\begin{aligned} \mathbf{E}[y^T \mathcal{A} y] &= \mathbf{E}[\mathbf{Tr}(\mathcal{A} y y^T)] \\ &= \mathbf{Tr}\left(\mathcal{A} \left(G \circ \left(\frac{1}{4} \mathbf{1}\mathbf{1}^T + \frac{1}{2\pi} \arcsin(C) \right) \right) \right) \\ &\leq \mathbf{Tr}\left(\mathcal{A} \left(\frac{1}{4} G + \frac{1}{2\pi} Z \right) \right) \\ &= \frac{1}{4} \mathbf{Tr}(\mathcal{A}G) + \frac{1}{2\pi} SDP1 \end{aligned}$$

because Schur's theorem shows that $A \circ B \succeq 0$ when $A, B \succeq 0$. It remains to notice that, because $(\mathbf{1}^T \otimes \mathbf{I}) z \leq \mathbf{1}$, and $(\mathbf{I} \otimes \mathbf{1}^T) z \leq \mathbf{1}$, with $z \geq 0$, then

$$(\mathbf{1}^T \otimes \mathbf{I}) \sqrt{z} \leq \sqrt{n}\mathbf{1} \quad \text{and} \quad (\mathbf{I} \otimes \mathbf{1}^T) \sqrt{z} \leq \sqrt{n}\mathbf{1},$$

so all the points y generated using this procedure are feasible for (QP) if we scale them by a factor $\sqrt{n}$. ∎

While the $O(\sqrt{n})$ bound grows relatively fast with problem dimension, remember that the problem has n^2 variables because it is written on permutation *matrices*. In what follows, we will see that better theoretical approximation bounds can be found if we write the seriation problem directly over permutation *vectors*, which is of course a much more restrictive formulation.

6.2. SDP relaxations and MLA. Several other semidefinite relaxations have been derived for the 2-SUM problem and the directly related 1-SUM, or MLA, problem. While these relaxations have very high computational complexity, to the point of being impractical, they come with excellent approximation bounds. We briefly recall these results in what follows. The 2-SUM minimization problem (1) is written (after taking square roots) as

$$\begin{array}{ll} \text{minimize} & \left(\sum_{i,j=1}^n A_{ij} (\pi(i) - \pi(j))^2 \right)^{\frac{1}{2}} \\ \text{subject to} & \pi \in \mathcal{P} \end{array} \tag{12}$$

in the variable $\pi \in \mathcal{P}$ which is a permutation of the vector $(1, \ldots, n)^T$. The authors of [12, 13, 6] form the following semidefinite relaxation:

(SDP2)

$$\begin{array}{ll} \text{minimize} & \sum_{i,j=1}^n A_{ij} X_{ij} \\ \text{subject to} & \frac{1}{|S|} \sum_{j \in S} (X_{ii} - 2X_{ij} + X_{jj}) \geq \frac{1}{6}(|S|/2 + 1)(|S| + 1) \quad \text{for all } S \subset [1, n],\ i = 1, \ldots, n, \\ & \frac{1}{|S|} \sum_{k \in S} \Delta^2(i, j, k) \geq \epsilon (X_{ii} - 2X_{ij} + X_{jj}) |S|^2 \quad \text{for all } S \subset [1, n],\ i, j = 1, \ldots, n, \\ & X \succeq 0,\ X_{ij} \geq 0, \quad i, j = 1, \ldots, n, \end{array}$$

in the variable $X \in \mathbf{S}_n$, where $\epsilon > 0$ and $\Delta(i,j,k)$ is given by the determinant

$$\Delta(i,j,k) = \begin{vmatrix} X_{jj} - 2X_{ij} + X_{ii} & X_{jk} - X_{ij} - X_{jk} + X_{ii} \\ X_{jk} - X_{ij} - X_{jk} + X_{ii} & X_{kk} - 2X_{ik} + X_{ii} \end{vmatrix}.$$

In [6, Thm. 2] it is shown that if OPT is the optimal value of the 2-SUM problem (12) and SDP2 the optimal value of the relaxation in (SDP2), then

$$SDP2\ (\log n)^{-1/2} \leq OPT \leq SDP2\ (\log n)^{3/2}.$$

While problem (SDP2) has an exponential number of constraints, efficient linear separation oracles can be constructed for the last two spreading constraints; hence the problem can be solved in polynomial time [21].

Tighter bounds can be obtained by exploiting approximation results on the MLA problem, noting that, after taking the square root of the objective, the 2-SUM problem is equivalent to

$$\min_{\pi \in \mathcal{P}} \max_{\{\|V\|_F \leq 1,\, V \geq 0\}} \sum_{i,j=1}^{n} V_{ij} A_{ij}^{1/2} |\pi(i) - \pi(j)| \tag{13}$$

in the variables $\pi \in \mathcal{P}$ and $V \in \mathbb{R}^{n \times n}$ (note that this is true for the support function of any set contained in the nonnegative orthant). Using results in [37, 14, 7], the MLA problem, written as

$$\min_{\pi \in \mathcal{P}} \sum_{i,j=1}^{n} W_{ij} |\pi(i) - \pi(j)| \tag{MLA}$$

over the variable $\pi \in \mathcal{P}$, with nonnegative weights $W \in \mathbb{R}^{n \times n}$, can be relaxed as (SDP3)

$$\begin{array}{ll} \text{minimize} & \sum_{i,j=1}^{n} W_{ij}(X_{ii} - 2X_{ij} + X_{jj}) \\ \text{subject to} & \frac{1}{|S|} \sum_{j \in S} (X_{ii} - 2X_{ij} + X_{jj}) \geq \frac{|S|^2}{5} \quad \text{for all } S \subset [1,n],\ i = 1, \ldots, n, \\ & (X_{ii} - 2X_{ij} + X_{jj}) \leq (X_{ii} - 2X_{ik} + X_{kk}) + (X_{kk} - 2X_{kj} + X_{jj}), \quad i,j,k = 1, \ldots, n, \\ & (X_{ii} - 2X_{ij} + X_{jj}) \geq 1, \quad i,j = 1, \ldots, n, \\ & X \succeq 0 \end{array}$$

in the variable $X \in \mathbf{S}_n$. The constraints above ensure that $d_{ij} = (X_{ii} - 2X_{ij} + X_{jj})$ is a squared Euclidean metric (hence a metric of negative type). If MLA is the optimal value of the minimum linear arrangement problem (MLA) and SDP3 the optimum of the relaxation in (SDP3), then in [14, Thm. 2.1] and [7] it is shown that

$$SDP3 \leq MLA \leq SDP3\ O(\sqrt{\log n}\, \log\log n),$$

which immediately yields a convex relaxation with $O(\sqrt{\log n}\, \log\log n)$ approximation ratio for the minmax formulation of the 2-SUM problem in (13).

6.3. Procedure for gene sequencing. We first order all the reads using the spectral algorithm. Then, in order to handle repeats in the DNA sequence, we adopt a divide and conquer approach and reorder smaller groups of reads partitioned using the spectral order. Finally we use the information given by mate pairs to reorder the resulting clusters of reads, using the QP relaxation. Outside of spectral computations which take less than a minute in our experiments, most computations can be naively parallelized. The details of the procedure are given below.

- Extract uniformly reads of length a few hundred bp (base pairs) from the DNA sequence. In our experiments, we artificially extract reads of length 200 bp from the true sequence of a million bp of the human chromosome 22. We perform a high coverage (each bp is contained in approximately 50 reads) uniform sampling. To replicate the setting of real sequencing data, we extract pairs of reads, with a distance of 5000 bp between each "mate" pair.
- Extract all possible k-mers from reads; i.e., for each read, record all subsequences of size k. We use $\mathrm{k} = 100$ in our experiments. The size of k-mers may be tuned to deal with noise in sequencing data (use small k) or repeats (use large k).
- Solve the C1P problem on the $\{0, 1\}$-matrix whose rows correspond to k-mers hits for each read; i.e., the element (i, j) of the matrix is equal to one if k-mer j is included in read i. Note that solving this C1P problem corresponds to reordering the similarity matrix between reads whose element (r, s) is the number of shared k-mers between reads r and s. In the presence of noise in sequencing data, this similarity matrix can be made more robust by recomputing, for instance, an edit distance between reads sharing k-mers. Moreover, if there are no repeated k-mers in the original sequence, i.e., a k-mer appears in two reads only if they overlap in the original sequence, then the C1P problem is solved exactly by the spectral relaxation and the original DNA sequence is retrieved by concatenating the overlapping reordered reads. Unfortunately, for large sequences, repeats are frequent and the spectral solution "mixes" together different areas of the original sequence. We deal with repeats in what follows.
- We extract contigs from the reordered reads: extract with high coverage (e.g., 10) sequences of a few thousand reads from the reordered sequence of reads (250000 reads in our experiments). Although there were repeats in the whole sequence, a good proportion of the contigs does not contain reads with repeats. By reordering each contig (using the spectral relaxation) and looking at the corresponding similarity (R-like) matrix, we can discriminate between "good" contigs (with no repeats and therefore a perfectly reordered similarity matrix which is an R-matrix) and "bad" contigs (with repeats and a badly reordered similarity matrix).
- Reorder the "good" contigs from the previous step using the spectral relaxation and agglomerate overlapping contigs. The aggregation can be done using again the spectral algorithm on the submatrix of the original similarity matrix corresponding to the two clusters of reads. Now there should be only a few (long) contigs left (usually less than a few hundred in our experiments).
- Use the mate pairs to refine the order of the contigs with the QP relaxation to solve the semisupervised seriation problem. Gaps are filled by incorporating the reads from the "bad" contigs (contigs with repeats).

Overall, the spectral preprocessing usually shrinks the ordering problem down to dimension $n \sim 100$, which is then solvable using the convex relaxations detailed in section 3.

REFERENCES

[1] E. D. Andersen and K. D. Andersen, *The Mosek interior point optimizer for linear programming: An implementation of the homogeneous algorithm*, in High Performance Optimization, Appl. Optim. 33, Kluwer Academic Publishers, Dordrecht, The Netherlands,

2000, pp. 197–232.
[2] J. E. Atkins, E. G. Boman, and B. Hendrickson, *A spectral algorithm for seriation and the consecutive ones problem*, SIAM J. Comput., 28 (1998), pp. 297–310.
[3] S. T. Barnard, A. Pothen, and H. Simon, *A spectral algorithm for envelope reduction of sparse matrices*, Numer. Linear Algebra Appl., 2 (1995), pp. 317–334.
[4] A. Barvinok, *Approximating orthogonal matrices by permutation matrices*, Pure Appl. Math. Quart., 2 (2006), pp. 943–961.
[5] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1998.
[6] A. Blum, G. Konjevod, R. Ravi, and S. Vempala, *Semidefinite relaxations for minimum bandwidth and other vertex ordering problems*, Theoret. Comput. Sci., 235 (2000), pp. 25–42.
[7] M. Charikar, M. T. Hajiaghayi, H. Karloff, and S. Rao, *l_2^2 spreading metrics for vertex ordering problems*, Algorithmica, 56 (2010), pp. 577–604.
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, 2012.
[9] A. d'Aspremont and N. El Karoui, *Weak recovery conditions from graph partitioning bounds and order statistics*, Math. Oper. Res., 38 (2013), pp. 228–247.
[10] C. Ding and X. He, *Linearized cluster assignment via spectral ordering*, in Proceedings of the Twenty-First ACM International Conference on Machine Learning, 2004, pp. 233–240.
[11] P. Erdös and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci, 5 (1960), pp. 17–61.
[12] G. Even, J. S. Naor, S. Rao, and B. Schieber, *Divide-and-conquer approximation algorithms via spreading metrics*, J. ACM, 47 (2000), pp. 585–616.
[13] U. Feige, *Approximating the bandwidth via volume respecting embeddings*, J. Comput. System Sci., 60 (2000), pp. 510–539.
[14] U. Feige and J. R. Lee, *An improved approximation ratio for the minimum linear arrangement problem*, Inform. Process. Lett., 101 (2007), pp. 26–29.
[15] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., 3 (1956), pp. 95–110.
[16] A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica, 19 (1999), pp. 175–220.
[17] D. R. Fulkerson and O. A. Gross, *Incidence matrices and interval graphs*, Pacific J. Math., 15 (1965), pp. 835–855.
[18] G. C. Garriga, E. Junttila, and H. Mannila, *Banded structure in binary matrices*, Knowl. Inf. Syst., 28 (2011), pp. 197–226.
[19] A. George and A. Pothen, *An analysis of spectral envelope reduction via quadratic assignment problems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 706–732.
[20] M. X. Goemans, *Smallest compact formulation for the permutahedron*, Math. Program., 153 (2015), pp. 5–11.
[21] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer, Berlin, 1988.
[22] F. R. Hodson, *The La Tène Cemetery at Münsingen-Rain: Catalogue and Relative Chronology*, Vol. 5, Stämpfli, Bern, Switzerland, 1968.
[23] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York, London, Sydney, 1972.
[24] D. G. Kendall, *Abundance matrices and seriation in archaeology*, Probab. Theory Related Fields, 17 (1971), pp. 104–112.
[25] J. Kuczyński and H. Woźniakowski, *Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start*, SIAM J. Matrix Anal. Appl, 13 (1992), pp. 1094–1122.
[26] M. Laurent and M. Seminaroti, *The quadratic assignment problem is easy for Robinsonian matrices with Toeplitz structure*, Oper. Res. Lett., 43 (2015), pp. 103–109.
[27] E. L. Lawler, *The quadratic assignment problem*, Management Sci., 9 (1963), pp. 586–599.
[28] I. Liiv, *Seriation and matrix reordering methods: An historical overview*, Stat. Anal. Data Min., 3 (2010), pp. 70–91.
[29] C. H. Lim and S. Wright, *Beyond the Birkhoff polytope: Convex relaxations for vector permutation problems*, in Advances in Neural Information Processing Systems, Vol. 27, MIT Press, Cambridge, MA, 2014, pp. 2168–2176.
[30] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and* 0–1 *optimization*, SIAM J. Optim., 1 (1991), pp. 166–190.
[31] J. Meidanis, O. Porto, and G. P. Telles, *On the consecutive ones property*, Discrete Appl. Math., 88 (1998), pp. 325–354.

[32] A. Nemirovski, *Sums of random symmetric matrices and quadratic optimization under orthogonality constraints*, Math. Program., 109 (2007), pp. 283–317.
[33] A. Nemirovski and U. Rothblum, *On complexity of matrix scaling*, Linear Algebra Appl., 302 (1999), pp. 435–460.
[34] Y. Nesterov, *Global Quadratic Optimization via Conic Relaxation*, CORE Discussion Paper 9860, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 1998.
[35] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, Boston, MA, 2004.
[36] L. Portugal, F. Bastos, J. Júdice, J. Paixão, and T. Terlaky, *An investigation of interior-point algorithms for the linear transportation problem*, SIAM J. Sci. Comput., 17 (1996), pp. 1202–1223.
[37] S. Rao and A. W. Richa, *New approximation techniques for some linear ordering problems*, SIAM J. Comput., 34 (2004), pp. 388–404.
[38] W. S. Robinson, *A method for chronologically ordering archaeological deposits*, Am. Antiquity, 16 (1951), pp. 293–301.
[39] W. F. Sheppard, *On the calculation of the double integral expressing normal correlation*, Trans. Cambridge Philos. Soc., 19 (1900), pp. 23–66.
[40] N. Z. Shor, *Quadratic optimization problems*, Soviet J. Comput. Systems Sci., 25 (1987), pp. 1–11.
[41] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist., 35 (1964), pp. 876–879.
[42] A. M.-C. So, *Moment inequalities for sums of random matrices and their applications in optimization*, Math. Program., 130 (2011), pp. 125–151.
[43] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, MA, 1990.
[44] N. Vuokko, *Consecutive ones property and spectral ordering*, in Proceedings of the 10th SIAM International Conference on Data Mining (SDM'10), 2010, pp. 350–360.
[45] K. Q. Weinberger and L. K. Saul, *Unsupervised learning of image manifolds by semidefinite programming*, Int. J. Comput. Vis., 70 (2006), pp. 77–90.
[46] Q. Zhao, S. E. Karisch, F. Rendl, and H. Wolkowicz, *Semidefinite programming relaxations for the quadratic assignment problem*, J. Combin. Optim., 2 (1998), pp. 71–109.