Convex Relaxations for Permutation Problems

Alex d'Aspremont, *CNRS & ENS, Paris.*

with Fajwel Fogel, Rodolphe Jenatton & Francis Bach, INRIA & ENS Paris.

Jobs

Recruiting **postdocs** (one or two years) at Ecole Normale Superieure in Paris.

Send your resume to aspremon@di.ens.fr.

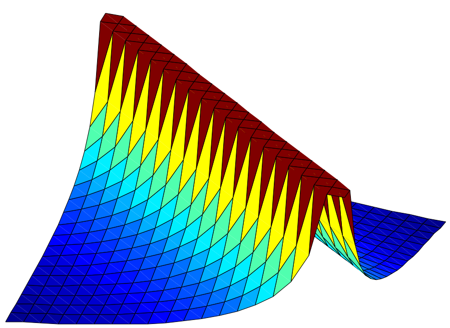
Seriation

The Seriation Problem.

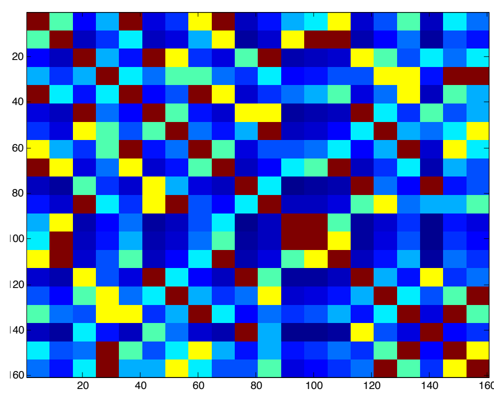
- Pairwise **similarity information** A_{ij} on n variables.
- Suppose the data has a **serial structure**, i.e. there is an order π such that

$$A_{\pi(i)\pi(j)} \text{ decreases with } |i-j| \quad \textbf{(R-matrix)}$$

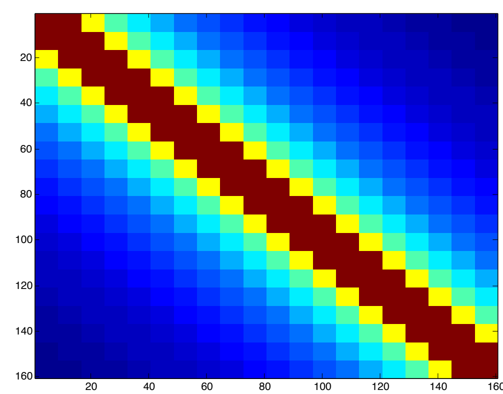
Recover π?



Similarity matrix



Input

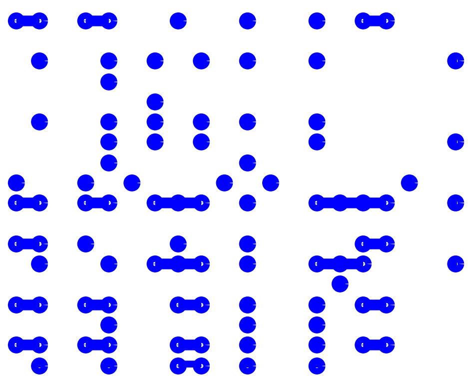


Reconstructed

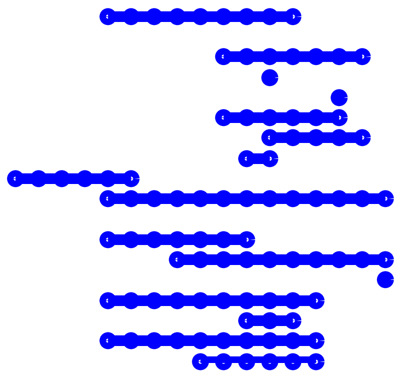
Seriation

The Continuous Ones Problem.

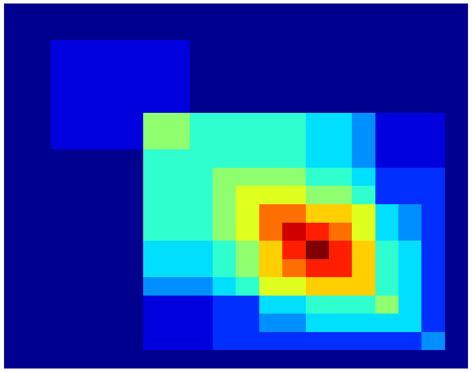
- We're given a rectangular binary $\{0,1\}$ matrix.
- Can we reorder its columns so that the ones in each row are contiguous (C1P)?



Input matrix



Ordered C1P matrix



C^TC (overlap)

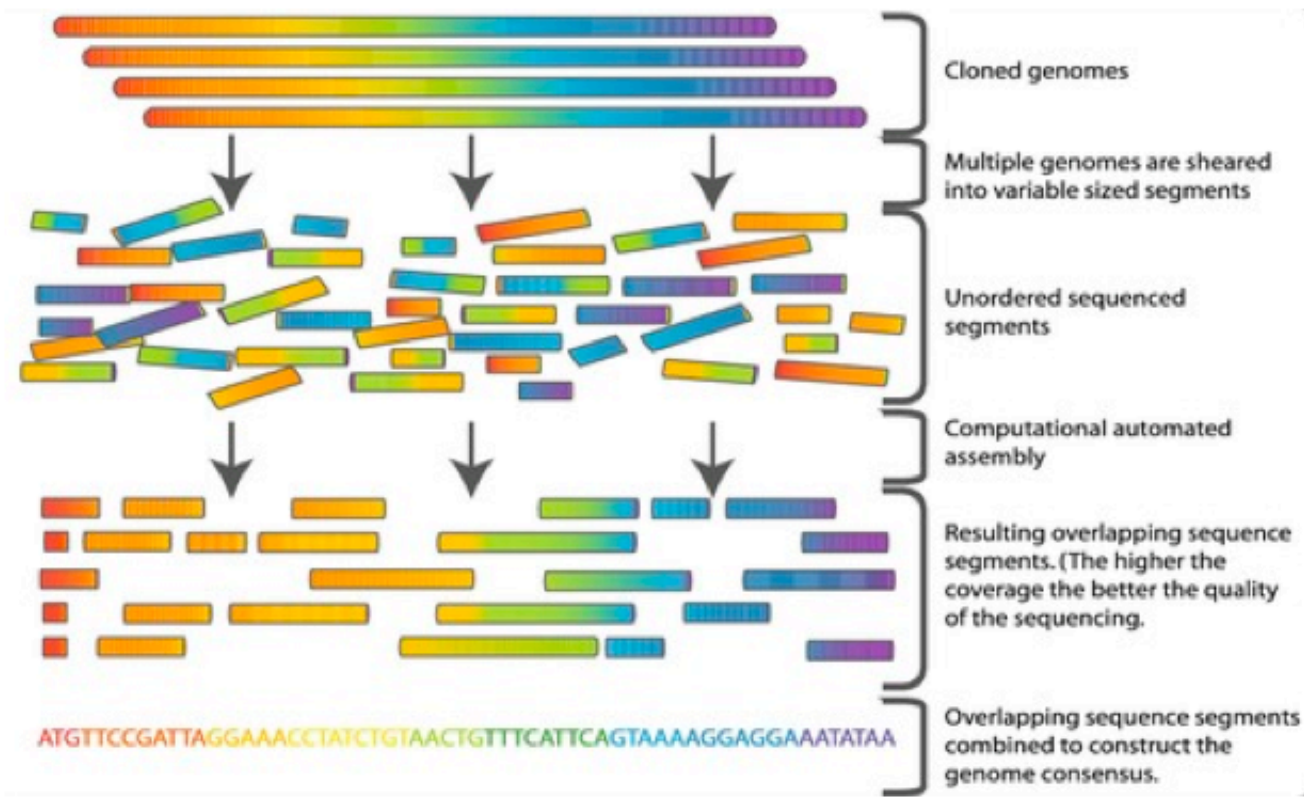
Lemma [Kendall, 1969]

Seriation and C1P. *Suppose there exists a permutation such that* C *is C1P, then* $C\Pi$ *is C1P if and only if* $\Pi^T C^T C \Pi$ *is an R-matrix.*

Shotgun Gene Sequencing

C1P has direct applications in shotgun gene sequencing.

- Genomes are cloned multiple times and randomly cut into shorter reads ($\sim$ 400bp), which are fully sequenced.
- Reorder the reads to recover the genome.

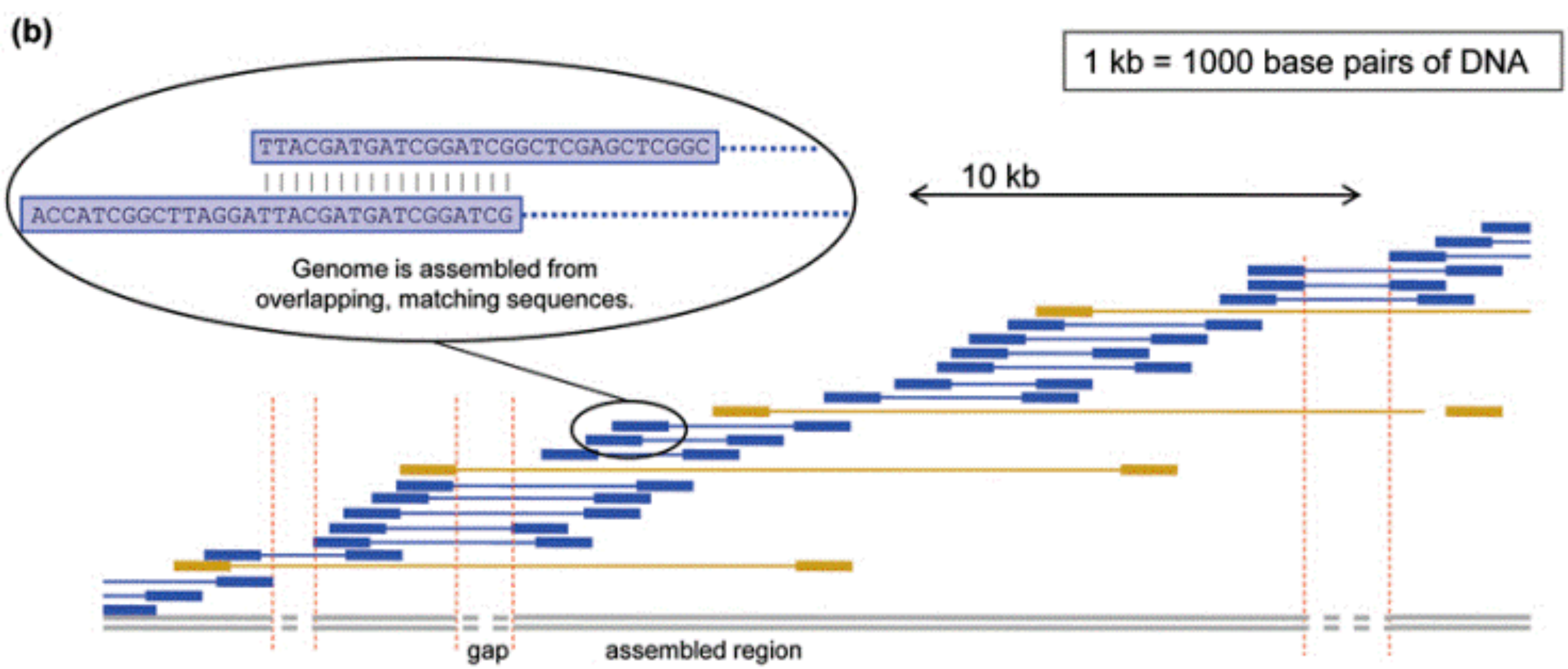


(from Wikipedia. . .)

Shotgun Gene Sequencing

C1P formulation.

- Scan the reads for k-mers (short patterns of bases).
- Form a *read* $\times$ *k-mer* matrix A, such that $A_{ij} = 1$ if k-mer j is in read i.
- Reorder the matrix A so that its **columns are C1P.**



(from [Gilchrist, 2010]). Only noiseless if the reads all have the same length.

Outline

- Introduction
- **Spectral solution**
- Combinatorial solution
- Convex relaxation
- Numerical experiments

A Spectral Solution

Spectral Seriation. Define the Laplacian of A as $L_A = \mathbf{diag}(A\mathbf{1}) - A$, the Fiedler vector of A is written

$$f = \operatorname*{argmin}_{\substack{\mathbf{1}^T x = 0, \\ \|x\|_2 = 1}} \; x^T L_A x.$$

and is the second smallest eigenvector of the Laplacian.

The Fiedler vector reorders a R-matrix in the noiseless case.

Theorem [Atkins, Boman, Hendrickson, et al., 1998]

Spectral seriation. *Suppose $A \in \mathbf{S}_n$ is a pre-R matrix, with a simple Fiedler value whose Fiedler vector f has no repeated values. Suppose that $\Pi \in \mathcal{P}$ is such that the permuted Fiedler vector Πv is monotonic, then $\Pi A \Pi^T$ is an R-matrix.*

Spectral Solution

A solution in search of a problem. . .

- What if the data is **noisy** and outside the perturbation regime? The spectral solution is only stable when the noise $\|\Delta L\|_2 \leq (\lambda_2 - \lambda_3)/2$.
- What if we have additional **structural information**?

Write seriation as an **optimization problem?**

Seriation

Combinatorial problems.

- **Ordering in 1D.** Given an increasing sequence $a_1 \leq \ldots \leq a_n$, solve

$$\min_{\pi \in \mathcal{P}} \sum_{i=1}^{n} a_i b_{\pi(i)}$$

Trivial solution: set π such that b_π is decreasing.

- **2D version.** The **2-SUM problem**, written

$$\min_{\pi \in \mathcal{P}} \sum_{i,j=1}^{n} A_{\pi(i)\pi(j)}(i-j)^2 \quad \text{or equivalently} \quad \min_{y \in \mathcal{P}} \sum_{i,j=1}^{n} A_{ij}(y_i - y_j)^2$$

where L_A is the Laplacian of A. The 2-SUM problem is **NP-Complete** for generic matrices A.

Seriation and 2-SUM

Combinatorial Solution. For certain matrices A, **2-SUM $\Longleftrightarrow$ seriation.**

Decompose the matrix A...

- Define **CUT(u,v) matrices** [Frieze and Kannan, 1999] as elementary $\{0,1\}$ R-matrices (one constant symmetric square block), with

$$CUT(u,v) = \begin{cases} 1 & \text{if } u \leq i,j \leq v \\ 0 & \text{otherwise,} \end{cases}$$

- The combinatorial objective for $A = CUT(u,v)$, is

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

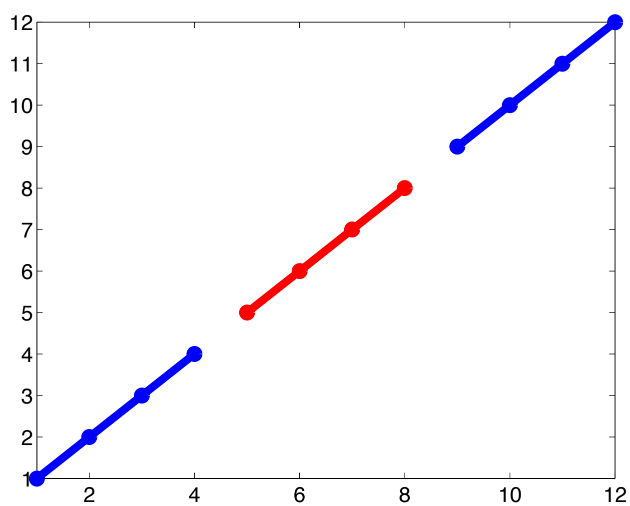
  it measures the **variance** of $y_{[u,v]}$.

Seriation and 2-SUM

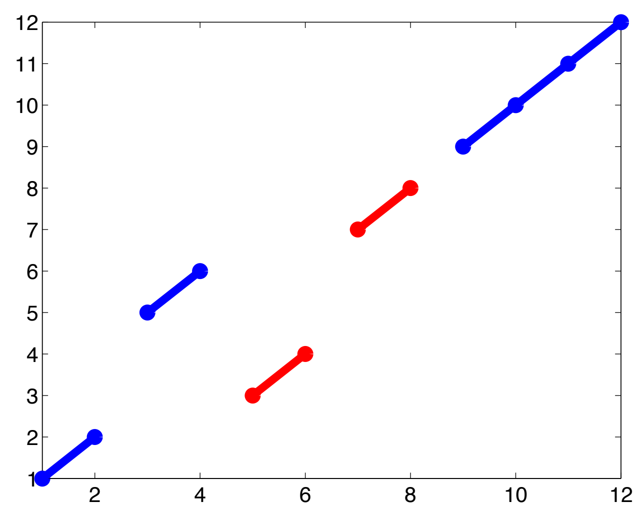
Combinatorial Solution. Solve

$$\min_{\pi \in \mathcal{P}} \sum_{i,j=1}^{n} A_{ij}(y_i - y_j)^2 = y^T L_A y$$

- For CUT matrices, **contiguous sequences** have **low variance**.
- All contiguous solutions have the **same variance** here.
- Simple graphical example with $A = CUT(5,8)\ldots$



$y^T L_A y = \mathbf{var}(y_{[5,8]}) = 1.6$



$y^T L_A y = \mathbf{var}(y_{[5,8]}) = 5.6$

Combinatorial Solution.

Lemma [Fogel, Jenatton, Bach, and d'Aspremont, 2013]

CUT decomposition. *If A is pre-R (or pre-P), then $A^T A = \sum_i A_i^T A_i$ is a sum of CUT matrices.*

Lemma [Fogel et al., 2013]

Contiguous 2-SUM solutions. *Suppose $A = CUT(u, v)$, and write $z = y_\pi$ the optimal solution to $\min_\pi y_\pi L_A y_\pi$. If we call $I = [u, v]$ and I^c its complement in $[1, n]$, then*

$$z_j \notin [\min(z_I), \max(z_I)], \quad \textit{for all } j \in I^c,$$

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

Seriation and 2-SUM

Combinatorial Solution. Solving 2-SUM

$$\min_{\pi \in \mathcal{P}} \sum_{i,j=1}^{n} A_{ij}(\pi(i) - \pi(j))^2 = \pi^T L_A \pi \tag{1}$$

when $y_i = i$, $i = 1, \ldots, n$ and A is a conic combination of CUT matrices.

Laplacian operator is linear, y_π monotonic **optimal for all CUT components.**

Proposition [Fogel et al., 2013]

Seriation and 2-SUM. *Suppose* $C \in \mathbf{S}_n$ *is a* $\{0, 1\}$ *pre-R matrix and* $y_i = i$ *for* $i = 1, \ldots, n$. *If* Π *is such that* $\Pi C \Pi^T$ *(hence* $\Pi A \Pi^T$*) is an R-matrix, then the permutation* π *solves the combinatorial minimization problem* (1) *for* $A = C^2$.

Outline

- Introduction
- Spectral solution
- Combinatorial solution
- **Convex relaxation**
- Numerical experiments

Convex Relaxation

What's the point?

- Write seriation as an optimization problem.
- Also gives a spectral (hence polynomial) solution for 2-SUM on some R-matrices ([Atkins et al., 1998] mention both problems, but don't show the connection).
- Write a **convex relaxation** for 2-SUM and seriation.
  - Spectral solution scales very well (cf. Pagerank, spectral clustering, etc.)
  - Not very robust. . .
  - Not flexible. . . Hard to include additional structural constraints.

Convex Relaxation

- Let $\mathcal{D}_n$ the set of doubly stochastic matrices, where

$$\mathcal{D}_n = \{X \in \mathbb{R}^{n \times n} : X \geqslant 0, X\mathbf{1} = \mathbf{1}, X^T\mathbf{1} = \mathbf{1}\}$$

  is the **convex hull of the set of permutation matrices.**

- Notice that $\mathcal{P} = \mathcal{D} \cap \mathcal{O}$, i.e. Π permutation matrix if and only Π is both **doubly stochastic** and **orthogonal.**

Convex Relaxation

We solve

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

in the variable $\Pi \in \mathbb{R}^{n \times n}$, where $P = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ and $Y \in \mathbb{R}^{n \times p}$ is a matrix whose columns are small perturbations of $g = (1, \ldots, n)^T$.

Convex Relaxation

Objective. $\mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) - \mu\|P\Pi\|_F^2$

- **2-SUM** term $\mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) = \sum_{i=1}^p y_i^T \Pi^T L_A \Pi y_i$ where y_i are small perturbations of the vector $g = (1, \ldots, n)^T$.
- **Orthogonalization penalty** $-\mu\|P\Pi\|_F^2$, where $P = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$.
  - Among all DS matrices, rotations (hence permutations) have the highest Frobenius norm.
  - Setting $\mu \leq \lambda_2(L_A)\lambda_1(YY^T)$, keeps the problem **a convex QP.**

Constraints.

- $e_1^T \Pi g + 1 \leq e_n^T \Pi g$ breaks degeneracies by imposing $\pi(1) \leq \pi(n)$. Without it, both monotonic solutions are optimal and this degeneracy can significantly deteriorate relaxation performance.
- $\Pi\mathbf{1} = \mathbf{1}$, $\Pi^T\mathbf{1} = \mathbf{1}$ and $\Pi \geq 0$, keep Π doubly stochastic.

Convex Relaxation

Other relaxations.

- A lot of work on relaxations for orthogonality constraints, e.g. SDPs in [Nemirovski, 2007, Coifman et al., 2008, So, 2011].
- Simple idea: $Q^TQ = \mathbf{I}$ is a quadratic constraint on Q, **lift it.** This yields a $O(\sqrt{n})$ approximation ratio.
- We could also use $O(\sqrt{\log n})$ approximation bounds for MLA [Even et al., 2000, Feige, 2000, Blum et al., 2000, Rao and Richa, 2005, Feige and Lee, 2007, Charikar et al., 2010].
- All these relaxations form extremely large SDPs.

Our simplest relaxation is a QP. No approximation bounds at this point however.

Semi-Supervised Seriation

Convex Relaxation.

- **Semi-Supervised Seriation.** We can add structural constraints to the relaxation, where

$$a \leq \pi(i) - \pi(j) \leq b \quad \text{is written} \quad a \leq e_i^T \Pi g - e_j^T \Pi g \leq b.$$

  which are linear constraints in Π.

- **Sampling permutations.** We can generate permutations from a doubly stochastic matrix D
  - Sample monotonic random vectors u.
  - Recover a permutation by reordering Du.

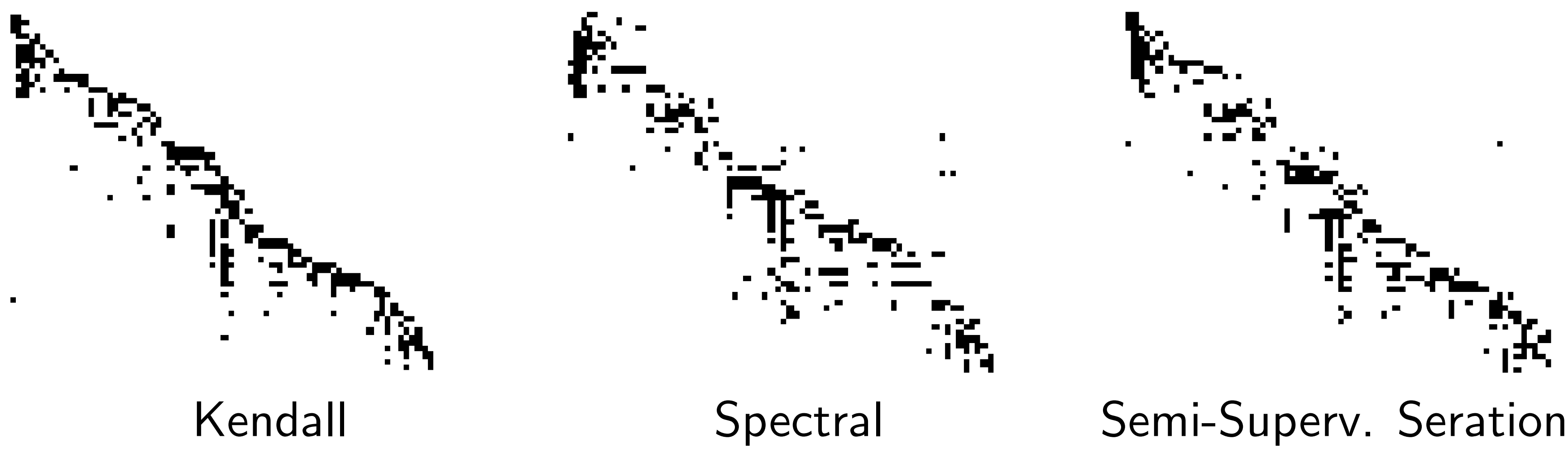
- **Algorithms.** Large QP, projecting on doubly stochastic matrices can be done very efficiently, using block coordinate descent on the dual. We use accelerated first-order methods.

Outline

- Introduction
- Spectral solution
- Combinatorial solution
- Convex relaxation
- **Numerical experiments**

Numerical results

Dead people. Row ordering **59 graves** $\times$ **70 artifacts** matrix [Kendall, 1971]. Find the chronology of the 59 graves by making artifact occurrences **contiguous in columns.**



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

Numerical results

Dead people.

	Kendall [1971]	**Spectral**	**QP Reg**	**QP Reg + 0.1%**	**QP Reg + 47.5%**
Kendall τ	1.00±0.00	0.75±0.00	0.73±0.22	0.76±0.16	0.97±0.01
Spearman ρ	1.00±0.00	0.90±0.00	0.88±0.19	0.91±0.16	1.00±0.00
Comb. Obj.	38520±0	38903±0	41810±13960	43457±23004	**37602±775**
# R-constr.	1556±0	1802±0	2021±484	2050±747	**1545±43**

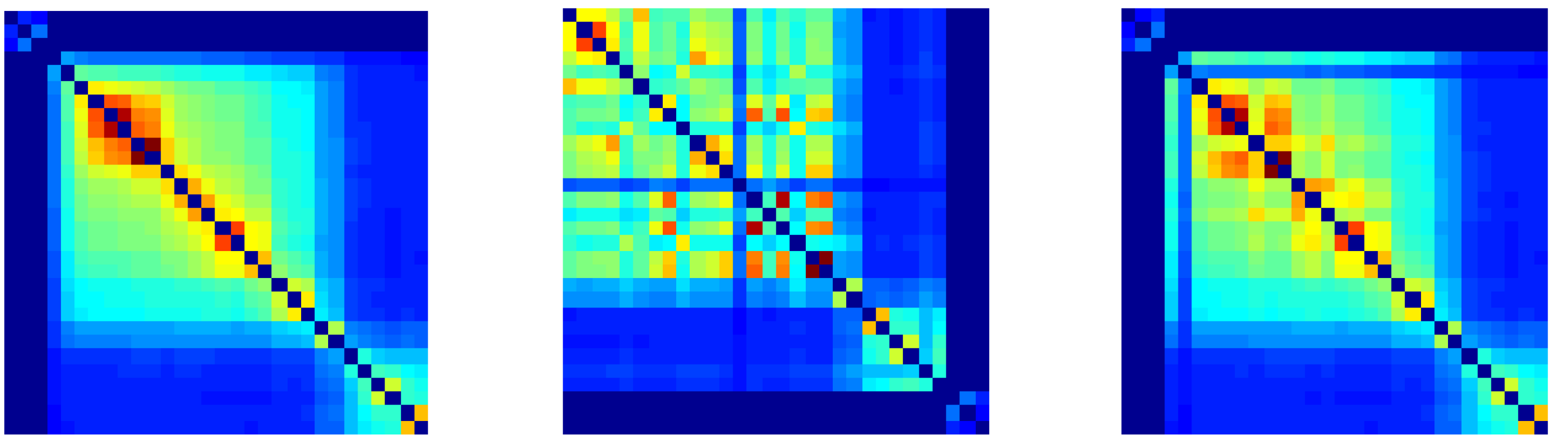
Performance metrics (median and stdev over 100 runs of the QP relaxation). We compare Kendall's original solution with that of the Fiedler vector, the seriation QP in (2) and the semi-supervised seriation QP with 0.1% and 24% pairwise ordering constraints specified.

Note that the **semi-supervised solution** actually improves on both Kendall's manual solution and on the spectral ordering.

Numerical results

Markov chain. Observe random permutations from a Markov chain.

- Gaussian Markov chain written $X_{i+1} = b_i X_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma_i^2)$.
- **Mutual information matrix** decreasing with $|i - j|$ when ordered according to the true Markov chain [Cover and Thomas, 2012], it is a **pre-R matrix.**



True Spectral Unsupervised QP

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

Numerical results

Markov chain.

	No noise	Noise within spectral gap	Large noise
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**

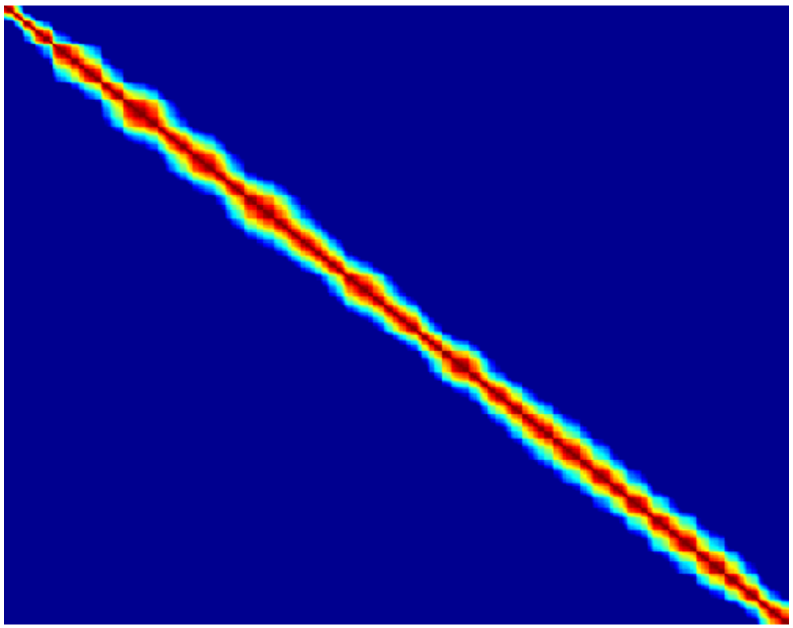
Kendall's τ between true Markov chain ordering, Fiedler vector, seriation QP and semi-supervised seriation QP with some pairwise orders specified.

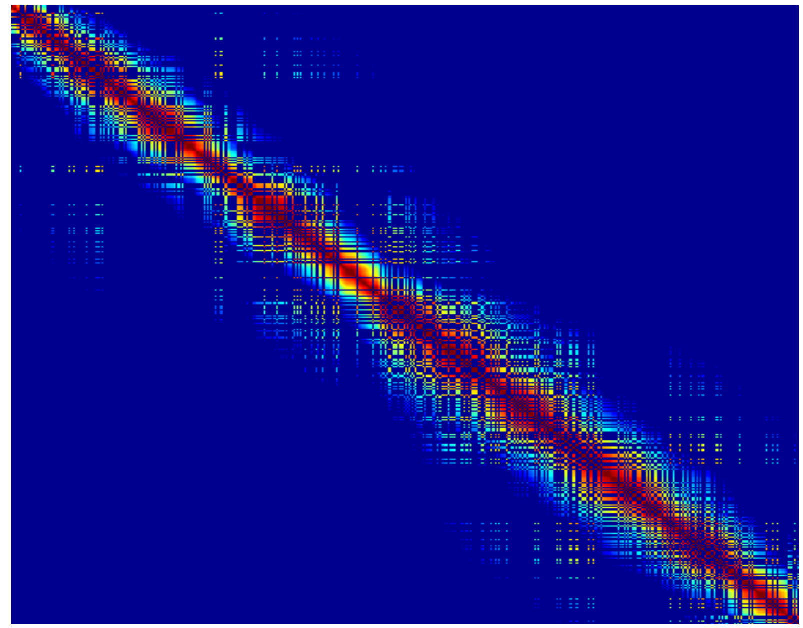
We observe:

- 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 *(noise within spectral gap)*.
- A sample covariance computed using much fewer samples so the spectral perturbation condition fails *(large noise)*.

Numerical results

DNA. Reorder the *read* similarity matrix to solve C1P on 250 000 reads from human chromosome 22.



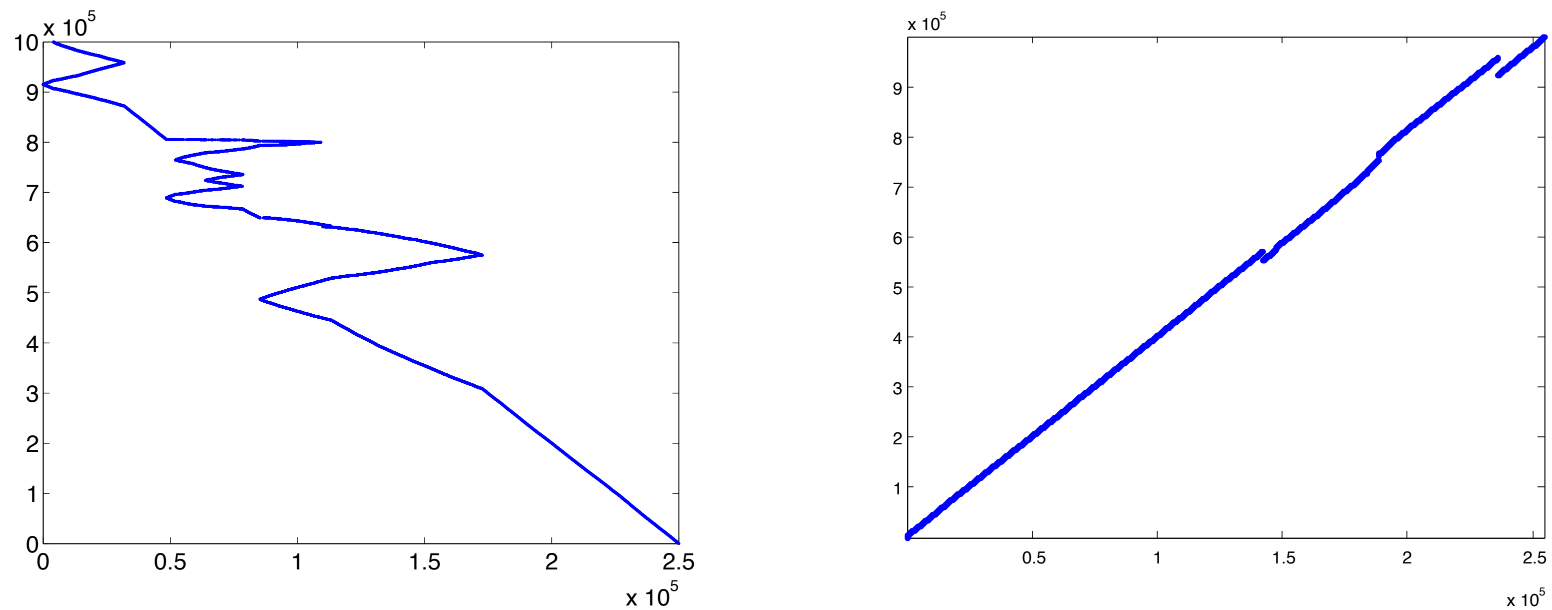


$\# \, reads \times \# \, reads$ matrix measuring the number of common k-mers between read pairs, reordered according to the spectral ordering.

The matrix is 250 000 $\times$ 250 000, we zoom in on two regions.

Numerical results

DNA. 250 000 reads from **human chromosome 22.**



Spectral   Spectral + QP

Recovered read position versus true read position for the **spectral solution** and the **spectral solution followed by semi-supervised seriation**.

We see that the number of misplaced reads significantly decreases in the semi-supervised seriation solution.

Conclusion

Results.

- Equivalence **2-SUM** $\Longleftrightarrow$ **seriation.**
- QP relaxation for **semi supervised seriation.**
- Good performance on shotgun gene sequencing.

Open problems.

- Approximation bounds.
- Large-scale QPs (without spectral preprocessing).
- Impact of similarity measures.

*

References

J.E. Atkins, E.G. Boman, B. Hendrickson, et al. A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.*, 28 (1):297–310, 1998.

Avrim Blum, Goran Konjevod, R Ravi, and Santosh Vempala. Semidefinite relaxations for minimum bandwidth and other vertex ordering problems. *Theoretical Computer Science*, 235(1):25–42, 2000.

Moses Charikar, Mohammad Taghi Hajiaghayi, Howard Karloff, and Satish Rao. l_2^2 spreading metrics for vertex ordering problems. *Algorithmica*, 56(4):577–604, 2010.

R. Coifman, Y. Shkolnisky, F.J. Sigworth, and A. Singer. Cryo-EM structure determination through eigenvectors of sparse matrices. *working paper*, 2008.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. Wiley-interscience, 2012.

Guy Even, Joseph Seffi Naor, Satish Rao, and Baruch Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *Journal of the ACM (JACM)*, 47(4):585–616, 2000.

Uriel Feige. Approximating the bandwidth via volume respecting embeddings. *Journal of Computer and System Sciences*, 60(3):510–539, 2000.

Uriel Feige and James R Lee. An improved approximation ratio for the minimum linear arrangement problem. *Information Processing Letters*, 101(1):26–29, 2007.

F. Fogel, R. Jenatton, F. Bach, and A. d'Aspremont. Convex relaxations for permutation problems. *NIPS 2013*, 2013.

A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

M. Gilchrist. Bringing it all back home: Next-generation sequencing technology and you. In *Mill Hill Essays 2010*. 2010.

David G Kendall. Incidence matrices, interval graphs and seriation in archeology. *Pacific Journal of mathematics*, 28(3):565–570, 1969.

David G Kendall. Abundance matrices and seriation in archaeology. *Probability Theory and Related Fields*, 17(2):104–112, 1971.

A. Nemirovski. Sums of random symmetric matrices and quadratic optimization under orthogonality constraints. *Mathematical programming*, 109(2):283–317, 2007.

Satish Rao and Andréa W Richa. New approximation techniques for some linear ordering problems. *SIAM Journal on Computing*, 34(2): 388–404, 2005.

Anthony Man-Cho So. Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical programming*, 130(1):125–151, 2011.