Convex Relaxations for Permutation Problems

Fajwel Fogel, *CNRS & ENS, Paris.*

with Alexandre d'Aspremont, Rodolphe Jenatton & Francis Bach

INRIA & ENS Paris

Seriation

The Seriation Problem.

Randomly ordered movie.

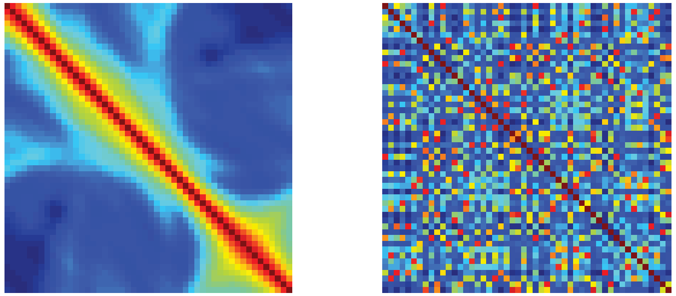


Image similarity matrix (true & observed)

Reconstructed movie.

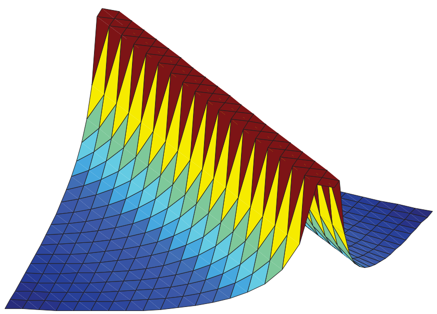
Seriation

The Seriation Problem.

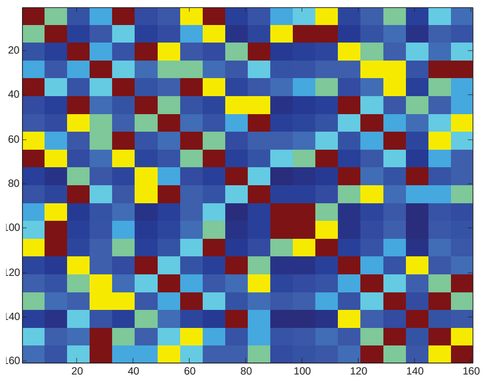
- Pairwise **similarity information** A_{ij} on n variables.
- Suppose the data has a **serial structure**, i.e. there is an order π such that

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

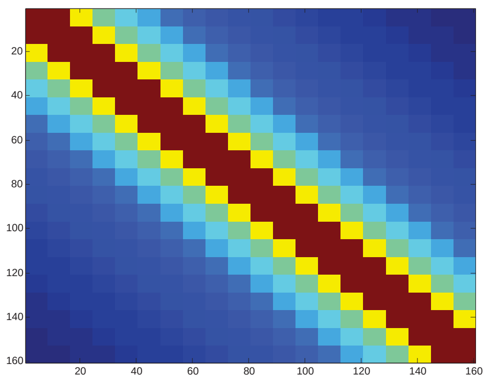
Recover π?



Similarity matrix



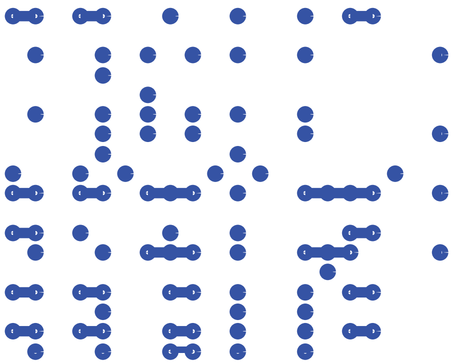
Input



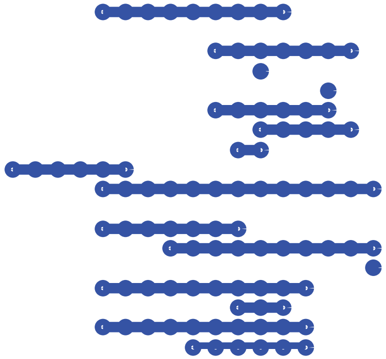
Reconstructed

The Continuous Ones Problem.

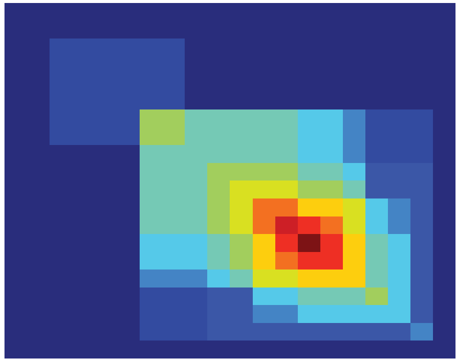
- We're given a rectangular binary $\{0, 1\}$ matrix.
- Can we reorder its columns so that the ones in each row are contiguous (C1P)?



Input matrix



Ordered C1P matrix



$C^T C$ (overlap)

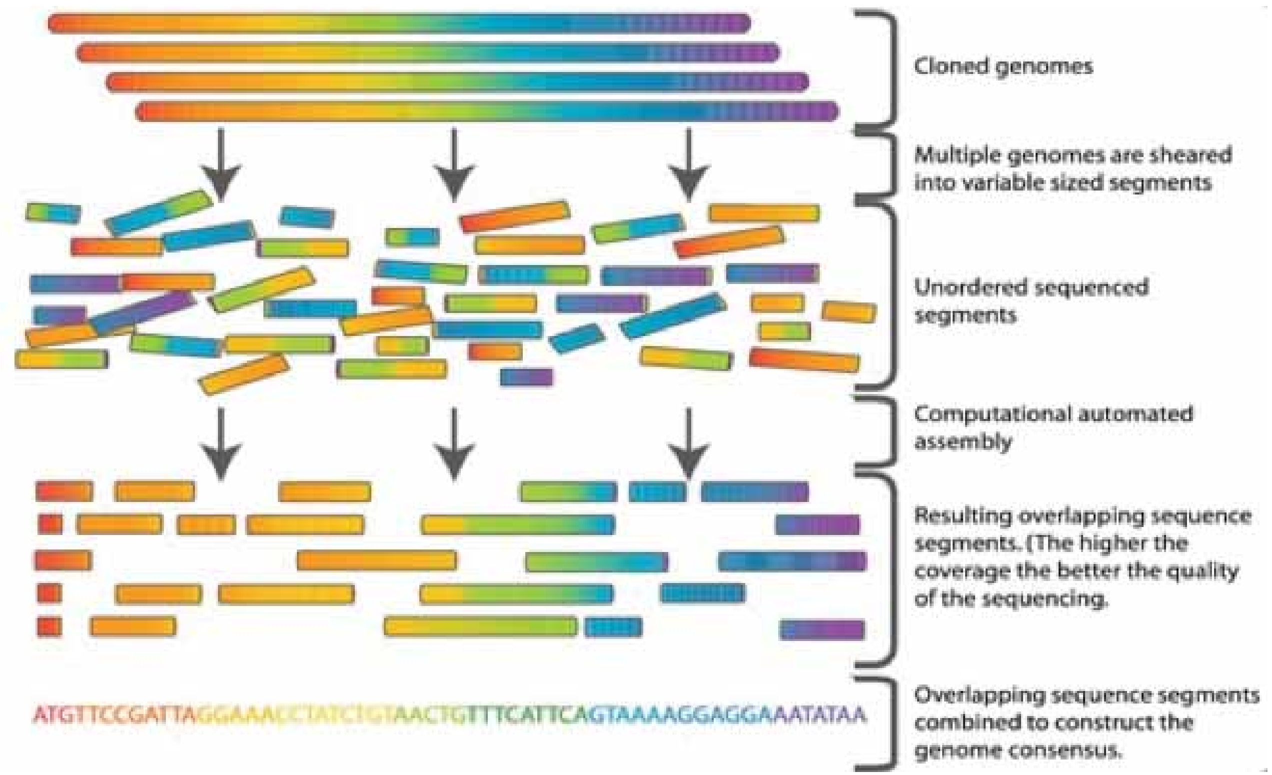
Lemma [Kendall, 1969]

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. . .)

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

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

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 spectral 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**?

Key questions here. . .

- Write seriation as an **optimization problem?**
- Define an **objective function?**

Seriation

Combinatorial problems.

- The **2-SUM problem**, written

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

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 $\iff$ 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 $\pi^T L_A \pi$ for $A = CUT(u,v)$, is

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

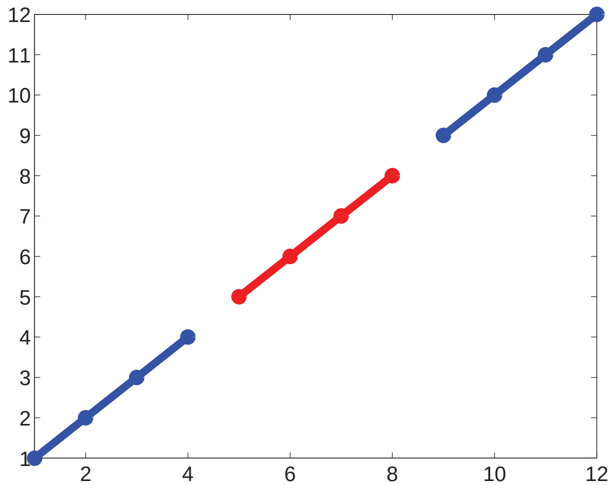
it measures the **variance** of $y_{[u,v]}$.

Seriation and 2-SUM

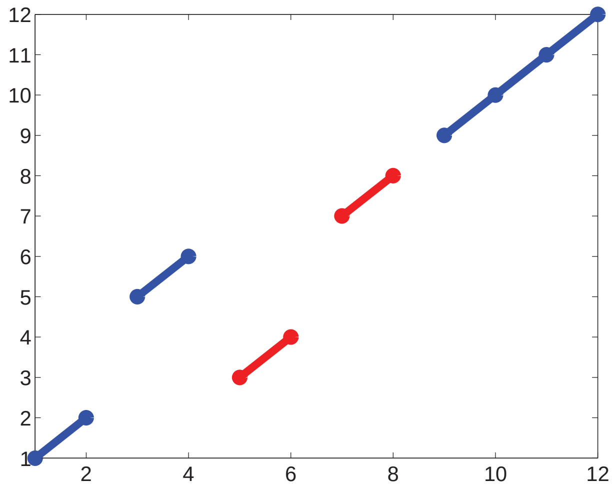
Combinatorial Solution. Solve

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

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



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



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

Seriation and 2-SUM

Combinatorial Solution.

- 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.
- 2-SUM optimization problem:

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

  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 [F., Jenatton, Bach, d'Aspremont, 2013]

Seriation and 2-SUM. *If A can be written as a conic combination of cut matrices, then the identity permutation is optimal for the 2-SUM problem (1). More generally if, for some permutation $\pi \in \mathcal{P}$, A_π can be written as a conic combination of cut matrices, then π is optimal for the 2-SUM problem (1).*

Combinatorial Solution.

Generalization: equivalence between seriation and 2-SUM for any R-matrix A.

Proposition [Laurent, Seminaroti, 2014]

Seriation and 2-SUM: generalization. *Let $A, B \in \mathbf{S}_n$ and assume that A is a Robinson similarity matrix, B is a Robinson dissimilarity matrix and moreover A or B is a Toeplitz matrix. Then the identity permutation is an optimal solution to the problem QAP(A,B).*

Outline

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

Convex Relaxation

2-SUM $\iff$ Seriation: What's the point?

- Spectral (hence polynomial) solution for 2-SUM on for most R-matrices.
- Write seriation as an optimization problem.
- 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

- Write $\mathcal{D}_n$ 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.**

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

- Form a **convex relaxation**

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

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

- **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

Approximation bounds.

- A lot of work on relaxations for orthogonality constraints, e.g. SDPs in [Nemirovski, 2007, Coifman et al., 2008, So, 2011]. All of this could be used here.
- Forms SDP of dimension $O(n^4)$, e.g. $O(n^9)$ for naive IPM implementations
- Simple idea: $Q^TQ = \mathbf{I}$ is a quadratic constraint on Q, **lift it.**
- $O(\sqrt{\log n})$ approximation bounds for some instances of Minimum Linear Arrangement. [Even et al., 2000, Feige, 2000, Blum et al., 2000, Rao and Richa, 2005, Feige and Lee, 2007, Charikar et al., 2010].
- Usual tradeoff with SDP relaxations: **higher complexity** but easier to **quantify approximation quality**.

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

Semi-Supervised Seriation

- **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.

- Recent work by Cong Han Lim and Stephen J. Wright (2014): optimize over permutahedron using sorting networks representation of Goemans. Seems faster to solve.

Outline

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

Comparing orderings

Compare permutations $x = \{3, 5, 7, \ldots, 1\}$ and $y = \{4, 5, 2, \ldots, 3\}$?

- **Spearman's ρ.** Pearson correlation between permutation vectors x and y.

- **Kendall's τ.** Pairs concordant if $(x_i, y_i) \leq (x_j, y_j)$ or $(x_i, y_i) \geq (x_j, y_j)$, then

$$\tau = \frac{\#\text{concordant pairs} - \#\text{non concordant pairs}}{n(n-1)/2}$$

- **2-SUM objective.** Compute

$$\sum_{i,j=1}^{n} A_{x_i x_j} (y_i - y_j)^2$$

- **# R constraints violated.** Number of pairwise R-constraints violated by permuted similarity matrix.

Numerical results

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

Kendall Spectral Semi-Superv. Seration

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.

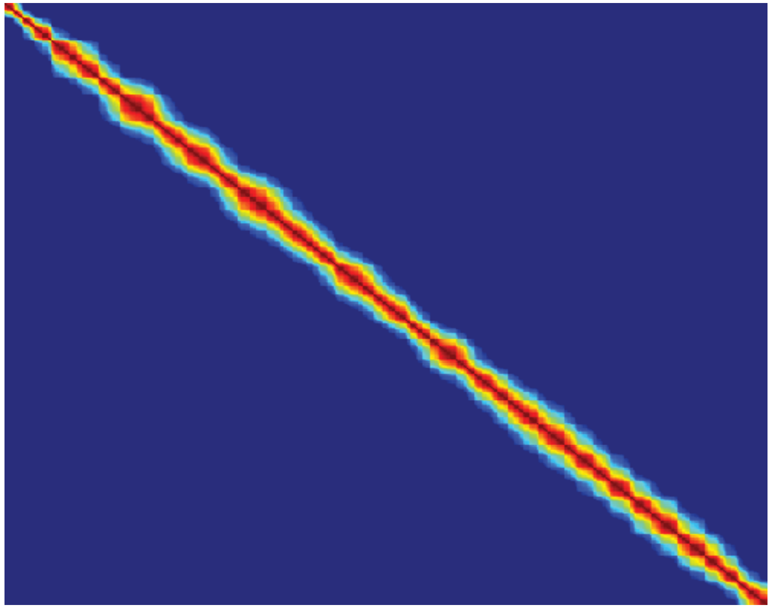
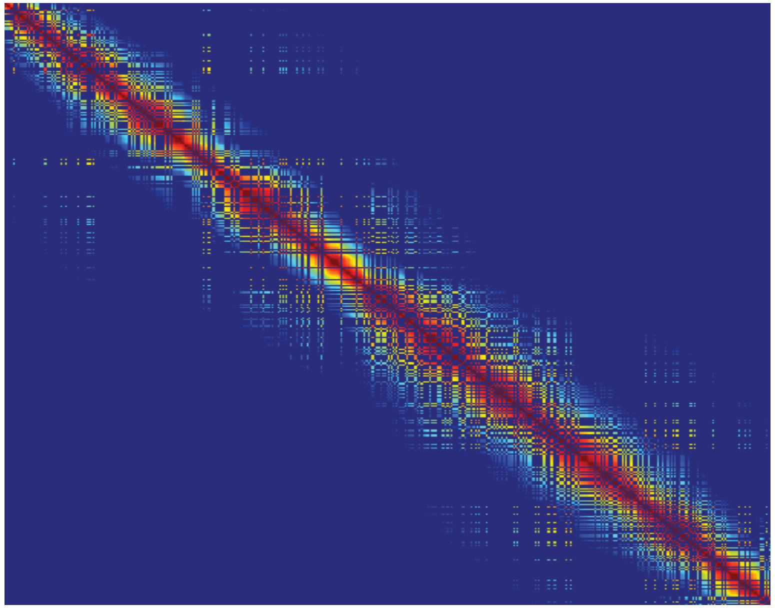
	Kendall [1971]	**Spectral**	**QP Reg**	**QP Reg + 0.1%**	**QP Reg + 47.5%**
Kendall τ	1.00±0.00	0.75±0.00	0.73±0.22	0.76±0.16	0.97±0.01
Spearman ρ	1.00±0.00	0.90±0.00	0.88±0.19	0.91±0.16	1.00±0.00
Comb. Obj.	38520±0	38903±0	41810±13960	43457±23004	**37602±775**
# R-constr.	1556±0	1802±0	2021±484	2050±747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

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

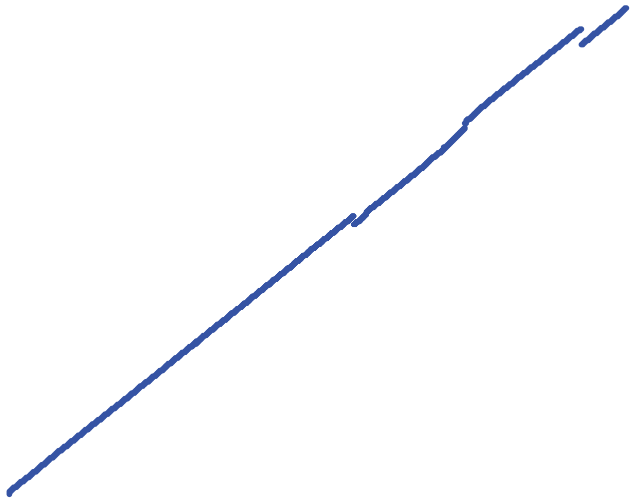
$\# \, reads \times \# \, reads$ matrix measuring the number of common k-mers between read pairs, reordered according to the spectral ordering.

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.

Advertisement: SerialRank

New method for ranking based on pairwise comparisons

- Comparison matrix C ($c_{ij} \in [-1, 1]$).
- Define similarity $S = nI + CC^T$: "number of similar outcomes against other opponents".
- Apply spectral and/or convex relaxation.

See preprint on Arxiv for more details and nice experiments!!

www.di.ens.fr/~fogel

Conclusion

Results.

- Equivalence **2-SUM** $\iff$ **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 for DNA sequencing and ranking.

Merci!

SerialRank

Fajwel Fogel (DI ENS), Alexandre d'Aspremont (DI ENS), Milan Vojnovic (Microsoft Research Cambridge)

Ranking from pairwise comparisons

Goal

Given pairwise comparisons between a set of elements, find the most consistent global order of these elements

Classical methods

- Ranking by score (e.g. *#wins - #losses*)
- Ranking by “skills” under a probabilistic model (e.g. Bradley Terry model)
- Ranking according to principal eigenvector of a transition matrix (e.g. PageRank, Rank centrality…)

Two main issues

- Missing comparisons
- Non transitive comparisons (i.e. $a<b$ and $b<c$ but $a>c$)

Applications

- Sports competitions (e.g. chess, football…)
- Crowdsourcing services (e.g. TopCoder…)
- Online computer games…

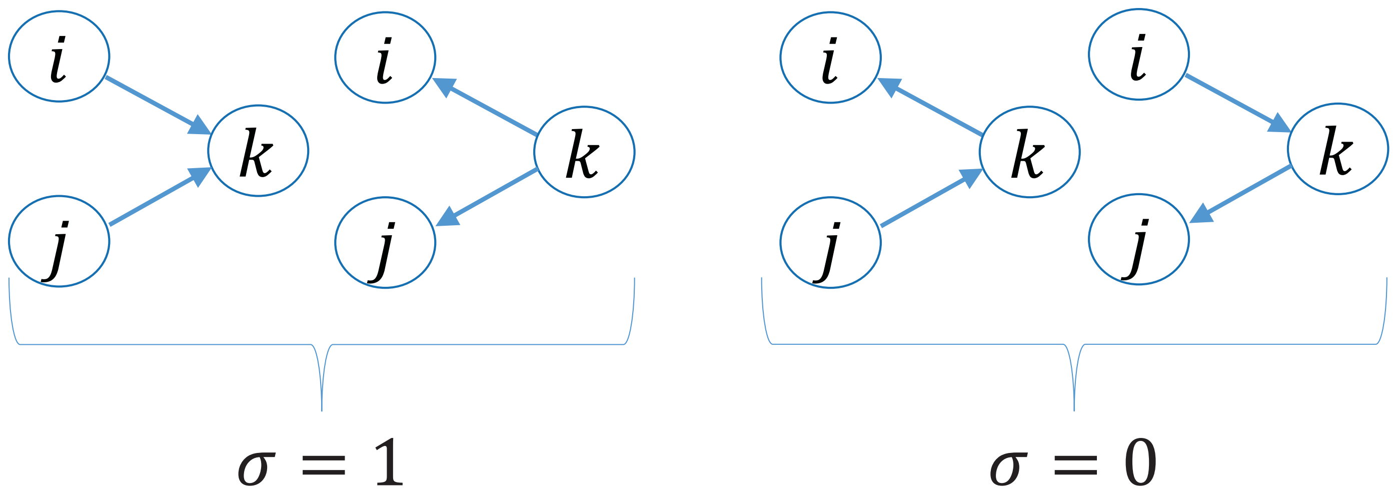
Constructing similarities: triplets outcomes

- Given a matrix of pairwise comparisons $C = [c_{i,j}]$ where $c_{i,j} \in [-1,1]$, e.g. for a tournament $c_{i,j} \in \{-1,0,1\}$ (loss, tie, win)
- Construct a similarity matrix $S = [s_{i,j}]$

$$s_{i,j} = \sum_{i,j\ compared\ with\ k} \sigma(c_{i,k}, c_{j,k})$$

where σ is a similarity measure

Idea: count matching comparisons of *i* and *j* against other items *k*



SerialRank: ranking as a seriation problem

Combinatorial optimization problem

- Assign similar candidates to nearby positions in ranking i.e. find ranking π of candidates 1 to n that minimizes

$$\sum_{i,j=1}^{n} s_{i,j}\left(\pi(i) - \pi(j)\right)^2$$

Spectral approach

- Compute Laplacian matrix $L_S = diag(S\mathbf{1}) - S$
- Compute the eigenvector of the second smallest eigenvalue of the Laplacian (Fiedler vector)
- Rank candidates in decreasing order or the Fiedler vector

High scalability when few comparisons available

Computing the Fiedler vector of *S* amounts to computing one eigenvector of a sparse matrix

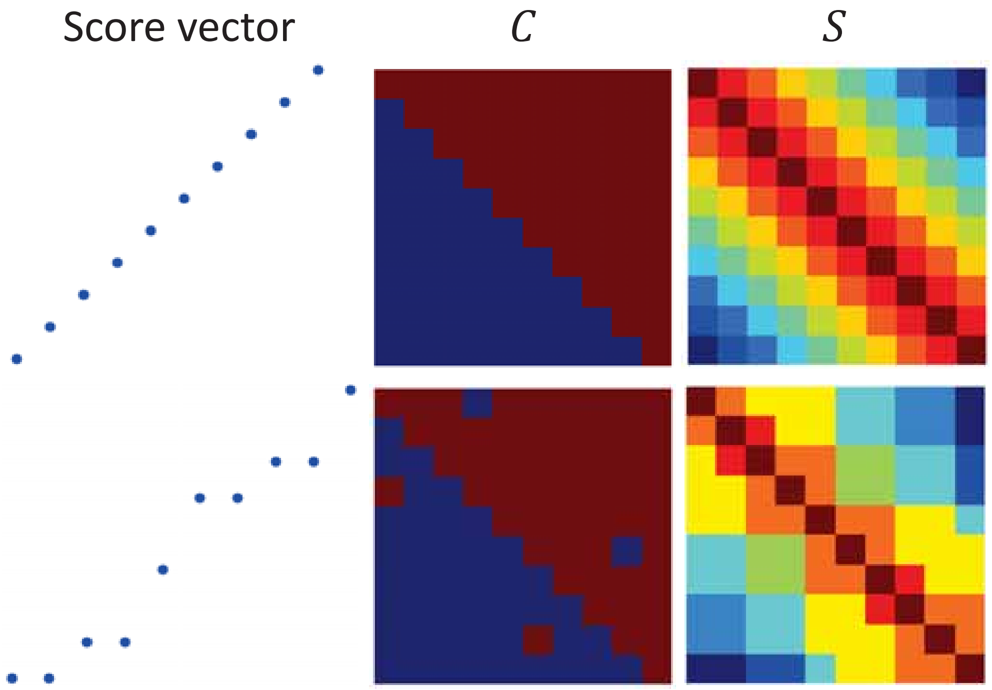
Performance guarantees

Robustness to missing/corrupted comparisons

- Similarity based ranking is more robust than typical score based rankings (i.e. *#wins - #losses)*

All comparisons given, no corrupted entries

All comparisons given, corrupted entries induce ties in score based ranking but not in similarity based ranking



Exact recovery regime

- Exact recovery of underlying ranking with probability $1 - o(1)$ for $\mathrm{o}(\sqrt{n})$ random missing/corrupted comparisons

Approximate recovery regime

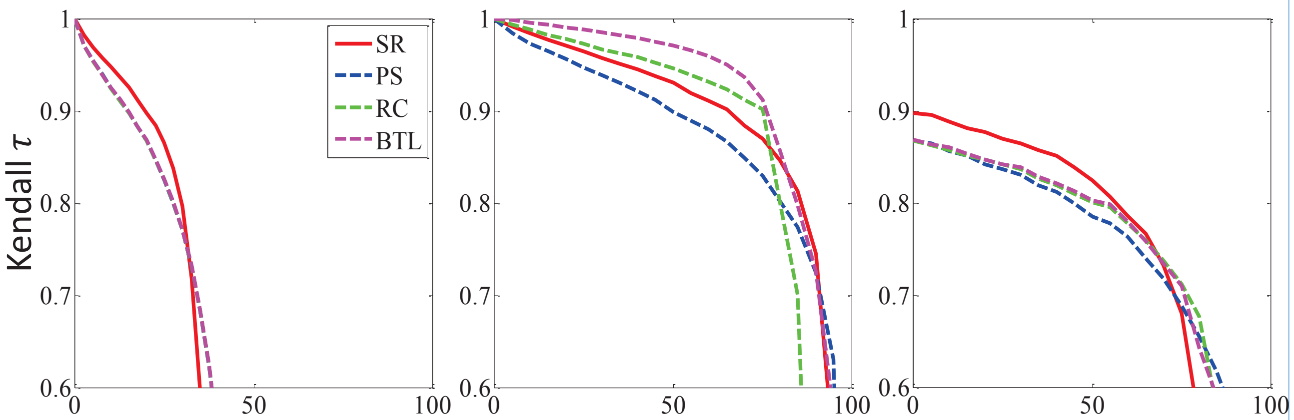
- Competitive to other approaches for partial observations and corrupted comparisons (cf. numerical experiments)

Numerical experiments

Synthetic datasets with random missing/corrupted comp.

Evaluate Kendall rank correlation coefficient τ between recovered ranking and “true” ranking

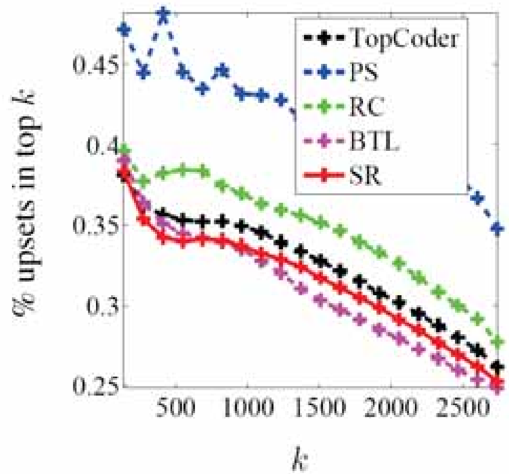
($\tau \in [-1,1]$, $\tau = 1$ means identical rankings)



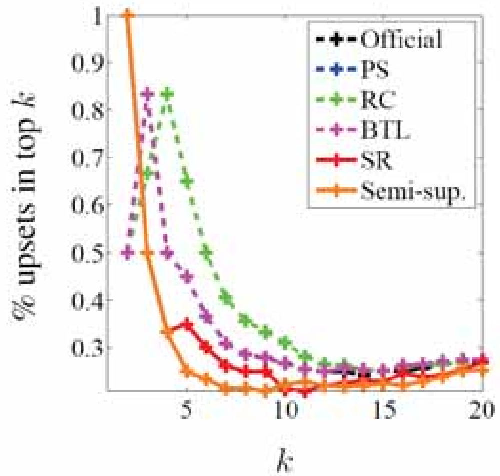
100 items, SR: SerialRank, PS: point-score, RC: rank centrality, BTL: Bradley-Terry

Real datasets (cf. paper for more details…)

TopCoder



England Premier League



Current work

- Theoretical guarantees for *SerialRank* in settings with few/corrupted comparisons: *perturbation results* for Fiedler vector
- *Semi-supervised* spectral ranking

*

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.

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.

Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre D Aspremont. Convex relaxations for permutation problems. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1016–1024. 2013. URL `http://papers.nips.cc/paper/4986-convex-relaxations-for-permutation-problems.pdf`.

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

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.