Rank aggregation via nuclear norm minimization

SAND2011-4281C

David F. Gleich

Sandia National Laboratories

Lek-Heng Lim

University of Chicago

Householder Symposium 18

Tahoe City, (Barely) CA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Which is a better list of good DVDs?

Lord of the Rings 3: The Return of …	Lord of the Rings 3: The Return of …
Lord of the Rings 1: The Fellowship	Lord of the Rings 1: The Fellowship
Lord of the Rings 2: The Two Towers	Lord of the Rings 2: The Two Towers
Lost: Season 1	Star Wars V: Empire Strikes Back
Battlestar Galactica: Season 1	Raiders of the Lost Ark
Fullmetal Alchemist	Star Wars IV: A New Hope
Trailer Park Boys: Season 4	Shawshank Redemption
Trailer Park Boys: Season 3	Star Wars VI: Return of the Jedi
Tenchi Muyo!	Lord of the Rings 3: Bonus DVD
Shawshank Redemption	The Godfather

Standard
rank aggregation
(the mean rating)

Nuclear Norm
based rank aggregation

Ranking is ***really*** *hard*

Ken Arrow

All rank aggregations involve some measure of compromise

John Kemeny

A good ranking is the "average" ranking under a permutation distance

Dwork, Kumar, Naor, Sivikumar

NP hard to compute Kemeny's ranking

Given a hard problem, what do you do?

Numerically relax!

It'll probably be easier.

Embody chair
John Cantrell (flickr)

Suppose we had scores

Let s_i be the score of the ith movie/song/paper/team to rank

Suppose we can compare the ith to jth:

$$Y_{i,j} = s_i - s_j$$

Then $\mathbf{Y} = \mathbf{s}\mathbf{e}^T - \mathbf{e}\mathbf{s}^T$ is skew-symmetric, rank 2.

Also works for $Y_{i,j} = s_i / s_j$ with an extra log.

Numerical ranking is intimately intertwined with skew-symmetric matrices

Kemeny and Snell, Mathematical Models in Social Sciences (1978)

Using ratings instead of comparisons

Ratings induce various skew-symmetric matrices.

$$Y_{i,j} = \frac{\sum_u R_{u,i} - R_{u,j}}{|\{u \mid R_{u,i} \text{ and } R_{u,j} \text{ exist}\}|}$$ *Arithmetic Mean*

$$Y_{i,j} = \log \frac{\Pr_u(R_{u,i} \geq R_{u,j})}{\Pr_u(R_{u,i} \leq R_{u,j})}$$ *Log-odds*

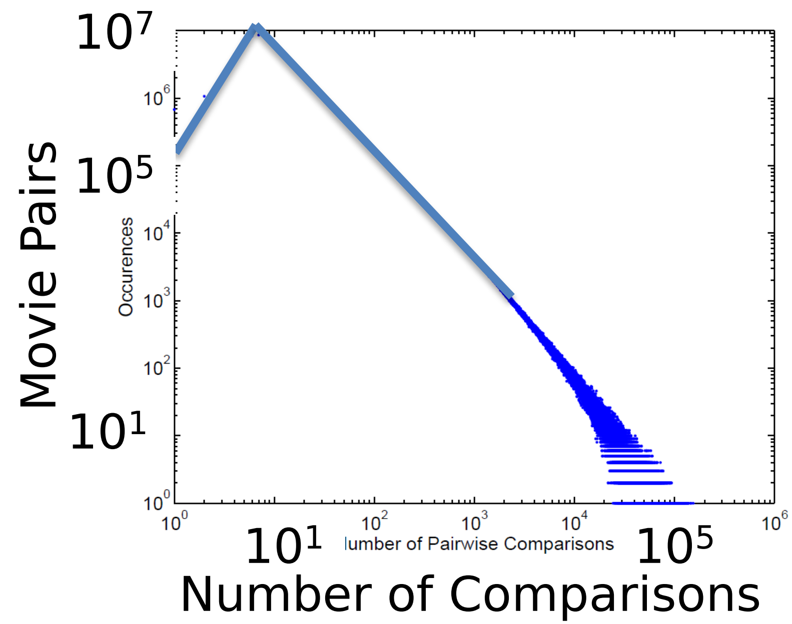
David 1988 – The Method of Paired Comparisons

Extracting the scores

Given $\boldsymbol{Y}$ with all entries, then
$\mathbf{s} = \frac{1}{n}\boldsymbol{Y}\mathbf{e}$ is the *Borda count*, the least-squares solution to $\mathbf{s}$

How many $Y_{i,j}$ do we have?
Most.

Do we *trust* all $Y_{i,j}$?
Not really.



Netflix data 17k movies, 500k users, 100M ratings– 99.17% filled

Only partial info? ***Complete it!***

Let $\hat{Y}_{i,j}$ be known for $(i,j) \in \Omega$ *We trust these scores.*

Goal Find the simplest skew-symmetric matrix that matches the data $\hat{Y}_{i,j}$

NP hard

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(\boldsymbol{Y}) \\ \text{subject to} & \boldsymbol{Y} = -\boldsymbol{Y}^T \\ & Y_{i,j} = \hat{Y}_{i,j} \text{ for all } (i,j) \in \Omega \end{array}$$

Convex

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{Y}\|_* \\ \text{subject to} & \boldsymbol{Y} = -\boldsymbol{Y}^T \\ & Y_{i,j} = \hat{Y}_{i,j} \text{ for all } (i,j) \in \Omega \end{array}$$

Heuristic

$\|\boldsymbol{Y}\|_* = \sum \sigma_i(\boldsymbol{Y})$ best convex underestimator of rank on unit ball.

Solving the nuclear norm problem

Use a LASSO formulation

$\mathbf{b} = \text{vec}(\hat{Y}_{i,j})$

$$\begin{array}{ll} \text{minimize} & \|\Omega(\boldsymbol{Y}) - \mathbf{b}\| \\ \text{subject to} & \|\boldsymbol{Y}\|_* \leq 2 \\ & \boldsymbol{Y} = -\boldsymbol{Y}^T \end{array}$$

Jain et al. propose SVP for this problem without $\boldsymbol{Y} = -\boldsymbol{Y}^T$

1. $\boldsymbol{Y}_0 = 0, t = 0$
2. REPEAT
3. $\boldsymbol{U}_t \boldsymbol{\Sigma}_t \boldsymbol{V}_t^T$ = rank-k SVD of $\Omega(\boldsymbol{Y}_t) - \eta(\Omega(\boldsymbol{Y}_t) - \mathbf{b})$
4. $\boldsymbol{Y}_{t+1} = \boldsymbol{U}_t \boldsymbol{\Sigma}_t \boldsymbol{V}_t^T$
5. $t = t + 1$
6. UNTIL $\|(\Omega(\boldsymbol{Y}_t) - \mathbf{b})\| < \varepsilon$

Jain et al. NIPS 2010

Skew-symmetric SVDs

Let $\boldsymbol{A} = -\boldsymbol{A}^T$ be an $n \times n$ skew-symmetric matrix with eigenvalues $i\lambda_1, -i\lambda_1, i\lambda_2, -i\lambda_2, \ldots, i\lambda_j, -i\lambda_j$, where $\lambda_i > 0, \lambda_i \geq \lambda_{i+1}$ and $j = \lfloor n/2 \rfloor$. Then the SVD of $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \boldsymbol{U} \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_1 & & & & & \\ & & \lambda_2 & & & & \\ & & & \lambda_2 & & & \\ & & & & \ddots & & \\ & & & & & \lambda_j & \\ & & & & & & \lambda_j \end{bmatrix} \boldsymbol{V}^T$$

for $\boldsymbol{U}$ and $\boldsymbol{V}$ given in the proof.

Proof Use the Murnaghan-Wintner form and the SVD of a 2x2 skew-symmetric block

This means that SVP will give us the skew-symmetric constraint "for free"

Exact recovery results

David Gross showed how to recover Hermitian matrices.
i.e. the conditions under which we get the exact **s**

Note that $\imath \boldsymbol{Y}$ is Hermitian. Thus our new result!

THEOREM 5. *Let* $\mathbf{s}$ *be centered, i.e.,* $\mathbf{s}^T\mathbf{e} = 0$. *Let* $\boldsymbol{Y} = \mathbf{s}\mathbf{e}^T - \mathbf{e}\mathbf{s}^T$ *where* $\theta = \max_i s_i^2/(\mathbf{s}^T\mathbf{s})$ *and* $\rho = ((\max_i s_i) - (\min_i s_i))/\|\mathbf{s}\|$. *Also, let* $\Omega \subset \mathcal{H}$ *be a random set of elements with size* $|\Omega| \geq O(2n\nu(1+\beta)(\log n)^2)$ *where* $\nu = \max((n\theta + 1)/4, n\rho^2)$. *Then the solution of*

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{X}\|_* \\ \text{subject to} & \operatorname{trace}(\boldsymbol{X}^*\boldsymbol{W}_i) = \operatorname{trace}((\imath\boldsymbol{Y})^*\boldsymbol{W}_i), \quad \boldsymbol{W}_i \in \Omega \end{array}$$

is equal to $\imath\boldsymbol{Y}$ *with probability at least* $1 - n^{-\beta}$.

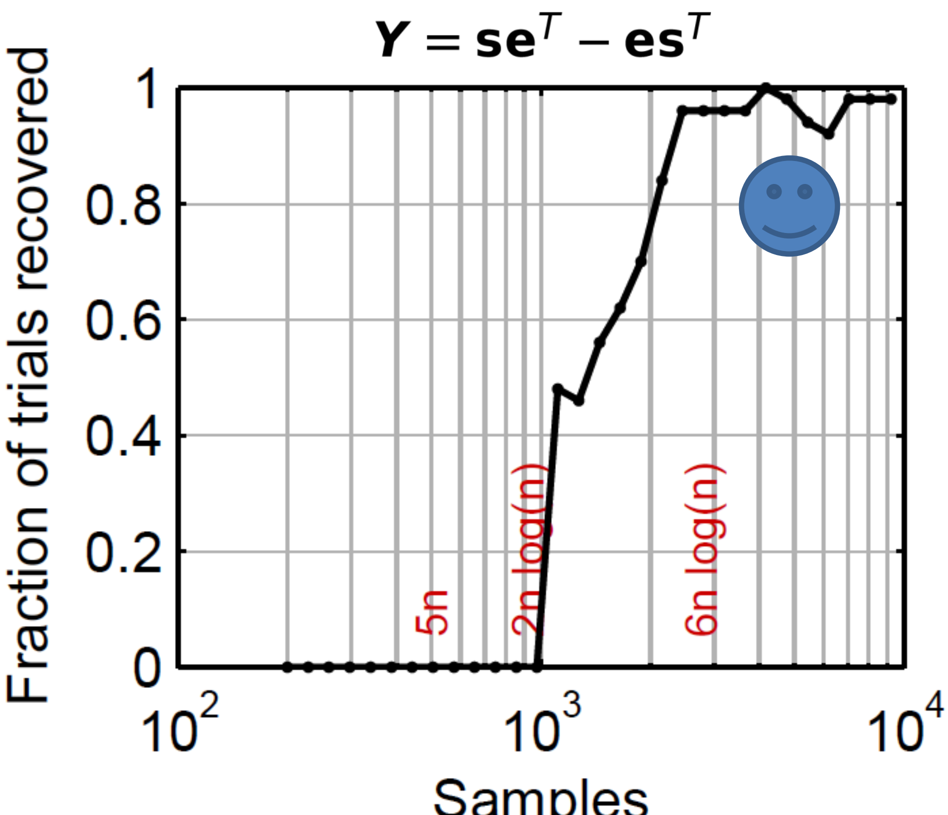
“$n \log(n)$”

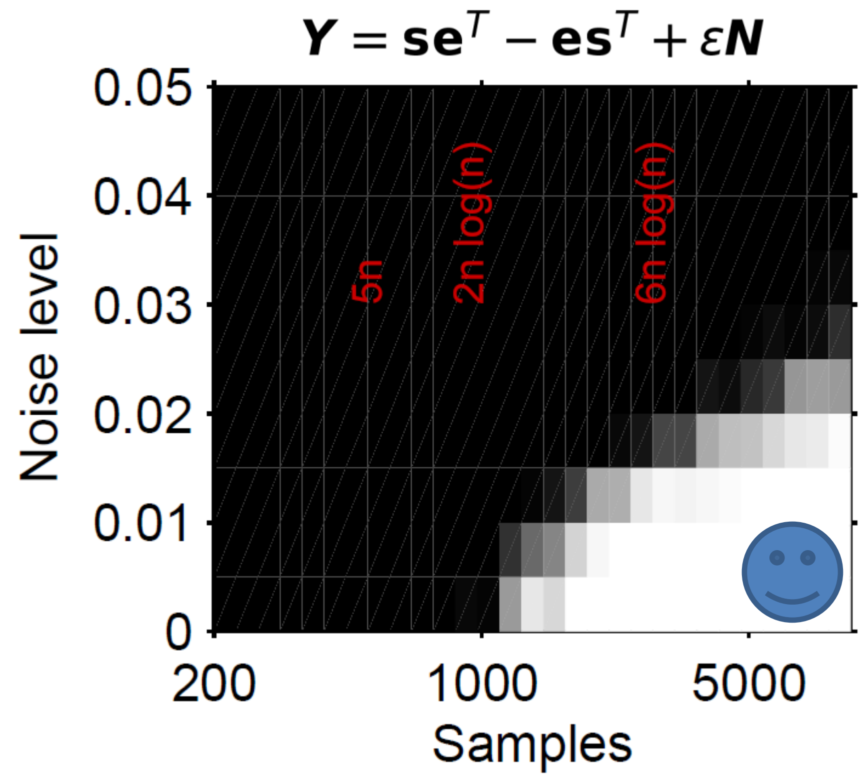
Gross arXiv 2010.

Recovery Discussion and Experiments

If $\boldsymbol{Y} = \mathbf{s}\mathbf{e}^T - \mathbf{e}\mathbf{s}^T$, then just look at differences from a connected set. Constants? Not very good.

"$n\log(n)$" **Intuition for *the truth*.**



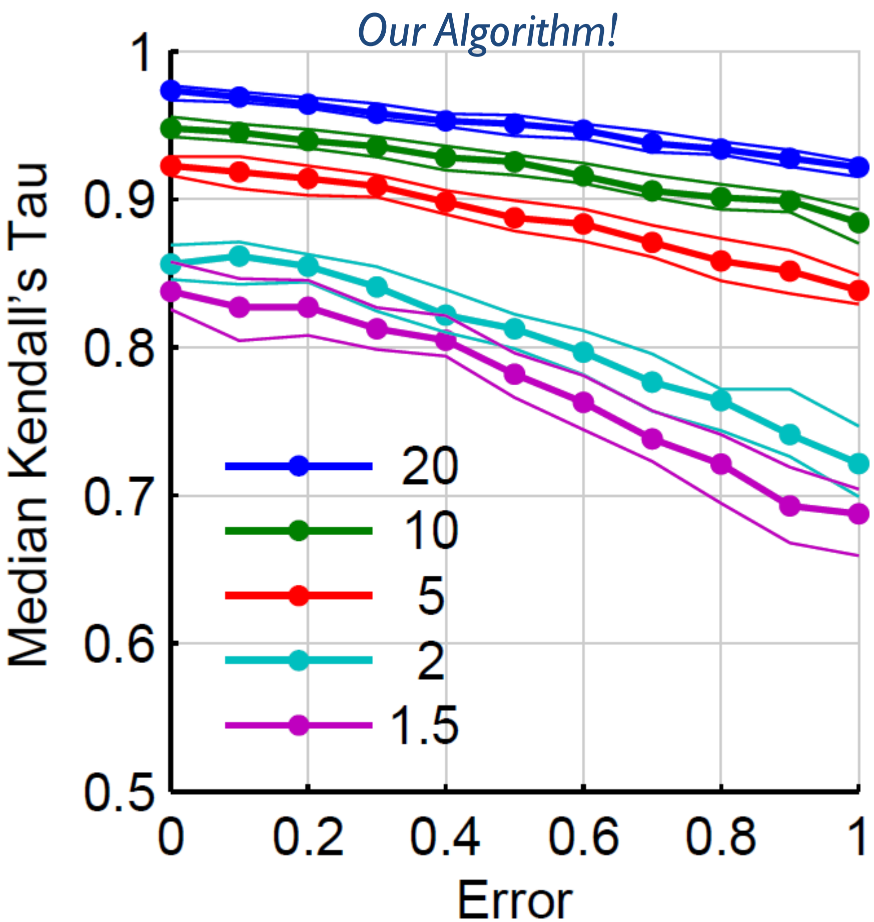


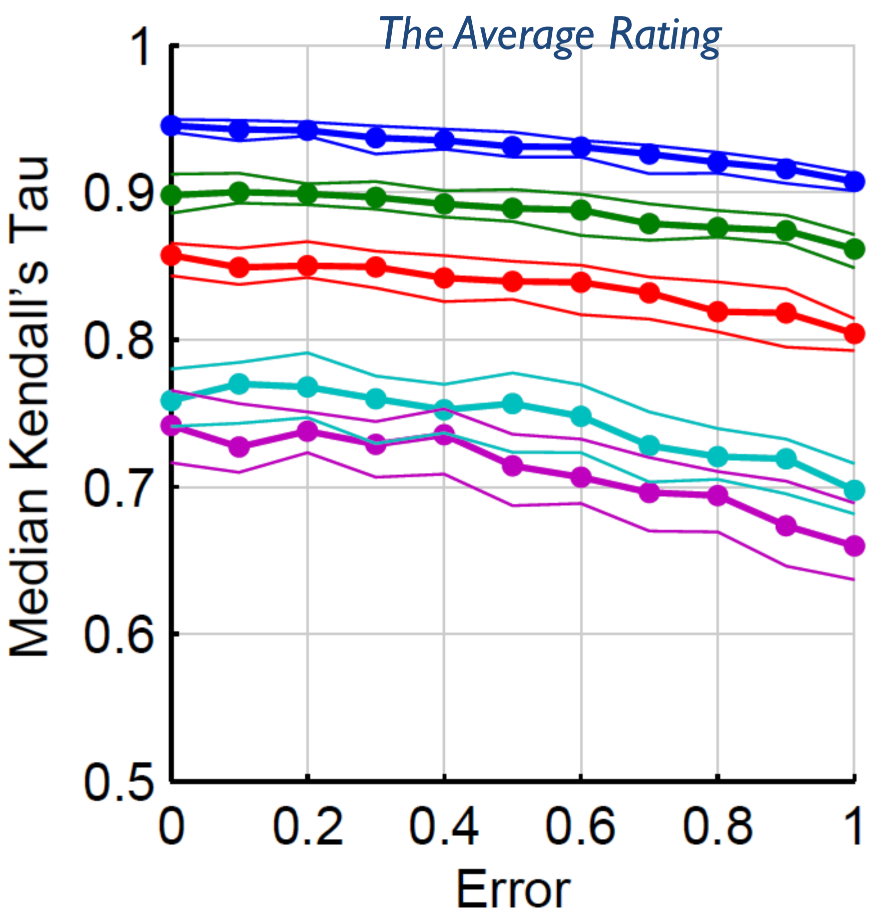
The Ranking Algorithm

0. INPUT $\boldsymbol{R}$ (ratings data) and c (for trust on comparisons)
1. Compute $\boldsymbol{Y}$ from $\boldsymbol{R}$
2. Discard entries with fewer than c comparisons
3. Set $\Omega, \mathbf{b}$ to be indices and values of what's left
4. $\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}^T = \text{SVP}(\Omega, \mathbf{b}, 2)$
5. OUTPUT $\mathbf{s} = (1/n)\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T\mathbf{e}$

Synthetic Results

Construct an Item Response Theory model. Vary number of ratings per user and a noise/error level





Conclusions and Future Work

Rank aggregation with the nuclear norm is:

principled

easy to compute

The results are much better than simple approaches.

1. Compare against others van Dooren (fitting) Massey (direct least squares for $\mathbf{s}$)

2. Noisy recovery! More realistic sampling.

3. Skew-symmetric Lanczos based SVD?

Google nuclear ranking gleich

https://dgleich.com/projects/skew-nuclear

To appear KDD2011

Approximation Residuals