Modeling and Recovering Non-Transitive Pairwise Comparison Matrices

Dehui Yang and Michael B. Wakin

Electrical Engineering and Computer Science
Colorado School of Mines

Rank Aggregation

- Goal is to produce a single ranked list of n items (or candidates, teams, etc.) that best reflects the collective preferences of multiple voters.

	Voter 1	Voter 2	Voter 3
Best	A	D	C
	B	B	D
	C	A	B
Worst	D	C	A

- Classical problem well studied in social choice theory, computer science, etc.
  - Arrow's impossibility theorem

Rank Aggregation via Pairwise Comparisons

Two steps [Gleich and Lim, 2011]:

1. Distill voter preferences into **pairwise comparisons**
   - most voters prefer item A over item B
   - most voters prefer item D over item C
   - etc.

2. Form ranked list based on pairwise comparisons

Pairwise Comparison Matrices

- Let Y denote an $n \times n$ matrix where $Y(i,j)$ represents the strength of preference of item i over item j.

- Typically, $Y(i,j) =$ -$Y(j,i)$, making Y **skew-symmetric**:

$$Y = -Y^T.$$

- How to create a pairwise comparison matrix?
  - **implicitly:** aggregating voter rankings, ratings databases, etc.
  - **explicitly:** direct surveys, polling, competitions, etc.

- Data may be noisy, incomplete.

Special Case

- Suppose each item has an intrinsic value $s(i)$ and the comparison $Y(i,j)$ simply equals

$$Y(i,j) = s(i) - s(j).$$

Then the matrix Y will be **rank two**. In particular,

$$Y = se^T - es^T,$$

where $s = [s(1)\ s(2)\ \ldots\ s(n)]^T$ and $e = [1\ 1\ \ldots\ 1]^T$.

- This makes Y a natural candidate for recovery via Nuclear Norm Minimization [Gleich and Lim, 2011; see also Massimino and Davenport, 2013].

Transitivity

- Such pairwise comparisons are **transitive**:

$$Y(i,j) = Y(i,k) + Y(k,j) \text{ for all } i,j,k.$$

- Indeed, transitivity holds only in this special case where

$$Y(i,j) = s(i) - s(j)$$

for some score vector s.

Realistic Pairwise Comparisons

- Condorcet paradox: Collective preferences may be cyclic.

	Voter 1	Voter 2	Voter 3
Best	A	C	B
	B	A	C
Worst	C	B	A

- Moreover, an individual's own preferences may not even be transitive. Individual preferences are often determined using multiple factors.

	Cost	Appearance	Practicality
Best	A	C	B
	B	A	C
Worst	C	B	A

Non-transitive Pairwise Comparison Matrices

- Our interest: Modeling and recovering Y itself, rather than flattening to a one-dimensional ranking.
- Questions:
  - What structure can we anticipate in Y?
  - Can non-transitive matrices be low rank?
- Contributions:
  - New model for non-transitive pairwise comparisons.
  - Low-rank analysis of resulting pairwise comparison matrices.
  - Discussing the recovery of these matrices.

New Model for Pairwise Comparisons

- Recall: Transitive model

$$Y(i,j) = s(i) - s(j).$$

- New: Suppose

$$Y(i,j) = s(i)a(j) - s(j)a(i),$$

where $s(i)$ represents a latent "value" for item i as before, but $a(j)$ is a "weight" determined by item j that can inhibit this value.

Interactions and Competition

- In the model

$$Y(i,j) = s(i)a(j) - s(j)a(i),$$

item j affects how item i is evaluated, and vice versa.

- Possible examples:
  - “Anchoring” in human judgment [Tversky and Kahneman, 1974]
  - Competitions and sporting events
    - $s(i)$ = offensive strength of team i (higher is better)
    - $a(j)$ = defensive strength of team j (lower is better)
    - $Y(i,j)$ = anticipated margin of victory for team i over team j
    - similar models have been proposed/discovered in linear regression of sporting outcomes [Pfitzner et al., 2009; Guo et al., 2012]

Example

- Vectors and resulting matrix:

$$s = \begin{bmatrix} -0.5749 \\ 0.7154 \\ 1.8577 \\ 0.0780 \end{bmatrix} \quad a = \begin{bmatrix} 0.1 \\ 0.5 \\ 0.5 \\ 1 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -0.359 & -0.473 & -0.583 \\ 0.359 & 0 & -0.571 & 0.676 \\ 0.473 & 0.571 & 0 & 1.819 \\ 0.583 & -0.676 & -1.819 & 0 \end{bmatrix}$$

- Non-transitive sign changes:

$$Y(1,2) + Y(2,4) = 0.317 > -0.583 = Y(1,4)$$

$$Y(1,3) + Y(3,4) = 1.346 > -0.583 = Y(1,4)$$

Non-transitivity

- The degree of non-transitivity in a pairwise comparison matrix can be measured [Jiang et al., 2010].

- For a skew-symmetric Y, define

$$R(Y) = \min_{\tilde{s}} \|Y - (\tilde{s}e^T - e\tilde{s}^T)\|_F$$

to be the distance between Y and the closest transitive matrix. The closest transitive matrix is generated using the score vector

$$\tilde{s} = \frac{1}{n} Y e.$$

Non-transitivity

- Under our model, where

$$Y(i,j) = s(i)a(j) - s(j)a(i),$$

we can show that

$$R(Y) \leq 2\, \|s\|_2 \, \|a\|_2 \sin \angle(\{a\}, \{s, e\})$$

- So the degree of non-transitivity is low if a is close to span$\{s$,$e\}$.

Extension to Multiple Factors

- Suppose there are r **latent factors** on which pairwise comparisons are based:

$$Y(i,j) = \sum_{k=1}^{r} s_k(i)a_k(j) - s_k(j)a_k(i).$$

- We can write

$$Y = \sum_{k=1}^{r} s_k a_k^T - a_k s_k^T,$$

showing that Y is skew-symmetric and has **rank at most** $2r$.

Non-transitivity

- Under our multi-factor model, where

$$Y = \sum_{k=1}^{r} s_k a_k^T - a_k s_k^T ,$$

we can show that

$$R(Y) \leq 2 \sum_{k=1}^{r} \|s_k\|_2 \, \|a_k\|_2 \sin \angle(\{a_k\}, \{s_k, e\})$$

- So the degree of non-transitivity is low if all a_k are close to span$\{s_k,e\}$.

Low-rank Structure

- In fact, any skew-symmetric matrix Y with rank at most $2r$ can be decomposed as

$$Y = \sum_{k=1}^{r} s_k a_k^T - a_k s_k^T ,$$

for some s_k and a_k [Brualdi et al., 2010].

- Therefore, any low-rank, skew-symmetric pairwise comparison matrix must fit our model, although the factors are not uniquely recoverable from the matrix itself.

Singular Value Decomposition

- The SVD of a skew-symmetric matrix Y with rank at most $2r$ is given by

$$Y = X \underbrace{\begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & & & \\ & & 0 & 1 & \\ & & 1 & 0 & \\ & & & & \ddots \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \lambda_1 & & & & \\ & \lambda_1 & & & \\ & & \ddots & & \\ & & & \lambda_r & \\ & & & & \lambda_r \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} -1 & 0 & & & \\ 0 & 1 & & & \\ & & -1 & 0 & \\ & & 0 & 1 & \\ & & & & \ddots \end{bmatrix}}_{V^T} X^T$$

where X is a matrix with orthonormal columns [Gleich and Lim, 2011].

Analysis

- Key to analysis:

$$\begin{aligned}\text{colspan}(Y) &= \text{rowspan}(Y) \\ &= \text{colspan}(X) \\ &= \text{span}\{s_1, s_2, \ldots, s_r, a_1, a_2, \ldots, a_r\}\end{aligned}$$

- Coherence of Y can be determined from any orthobasis for span$\{s_1, s_2, \ldots, s_r, a_1, a_2, \ldots, a_r\}$.

Recovery Algorithms

1. SVP [Jain et al., 2010]
   – Advantages: Output matrix guaranteed to be skew-symmetric [Gleich and Lim, 2011].
   – Disadvantages: Speed, lack of theoretical guarantees.

2. Alternating minimization [Jain et al., 2013]

$$\min_{U,V \in \mathbf{R}^{n \times 2r}} \|P_\Omega(Y - UV^T)\|_F^2$$

   – Advantages: Speed, theoretical guarantees.
   – Disadvantages: Not guaranteed to preserve skew-symmetry.

Example Recovery Result

- Suppose $s_1, s_2, \ldots, s_r, a_1, a_2, \ldots, a_r$ are orthonormal with coherence μ, and that

$$Y = \sum_{k=1}^{r} \lambda_k (s_k a_k^T - a_k s_k^T).$$

Then with

$$m = O\left(\mu^2 \left(\frac{\lambda_1}{\lambda_r}\right)^6 r^7 n \log n \log \frac{r\|Y\|_F}{\epsilon}\right)$$

random samples, with high probability Altmin returns an estimate $\widehat{Y}$ after $\log(1/\epsilon)$ iterations that satisfies

$$\|Y - \widehat{Y}\|_F \leq \epsilon.$$

Recovery Algorithms [ctd.]

3. Skew-symmetric alternating minimization

$$\min_{P,Q \in \mathbf{R}^{n\times r}} \|P_\Omega(Y - (PQ^T - QP^T))\|_F^2$$

- Implementation: Fix $\widehat{P}$ and solve the least-squares problem

$$\min_{Q \in \mathbf{R}^{n\times r}} \|\mathrm{vec}(P_\Omega Y) - M_{\widehat{P}}\,\mathrm{vec}(Q)\|_2^2$$

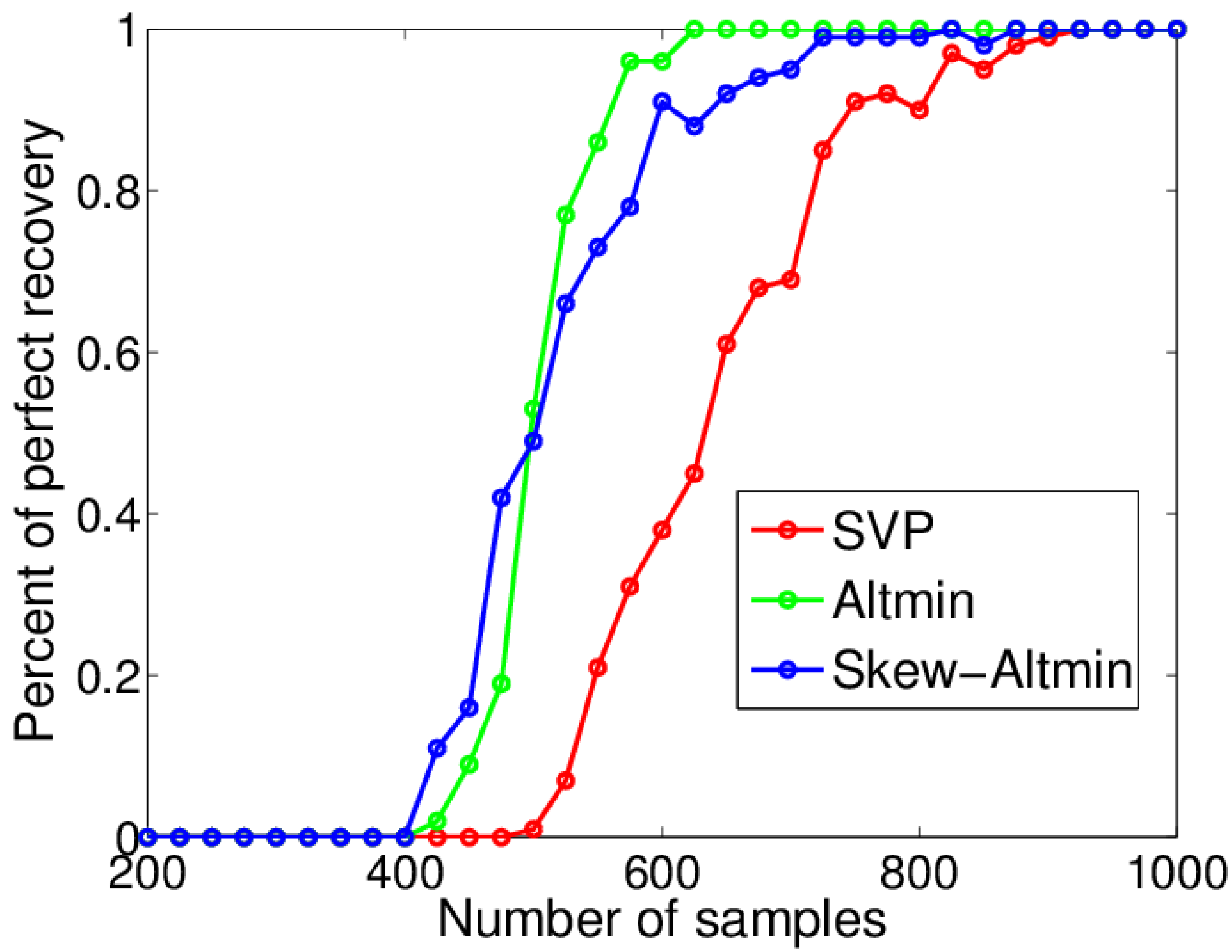
  Then fix $\widehat{Q}$ and solve for P.

- Advantages: Speed, preserves skew-symmetry.
- Disadvantages: Lack of theoretical guarantees.

Performance

- $n = 100;\ r = 1$ (rank $= 2$)
- s_1, a_1 random with entries $\mathrm{U}[0,1]$
- coherence: low
- non-transitivity:

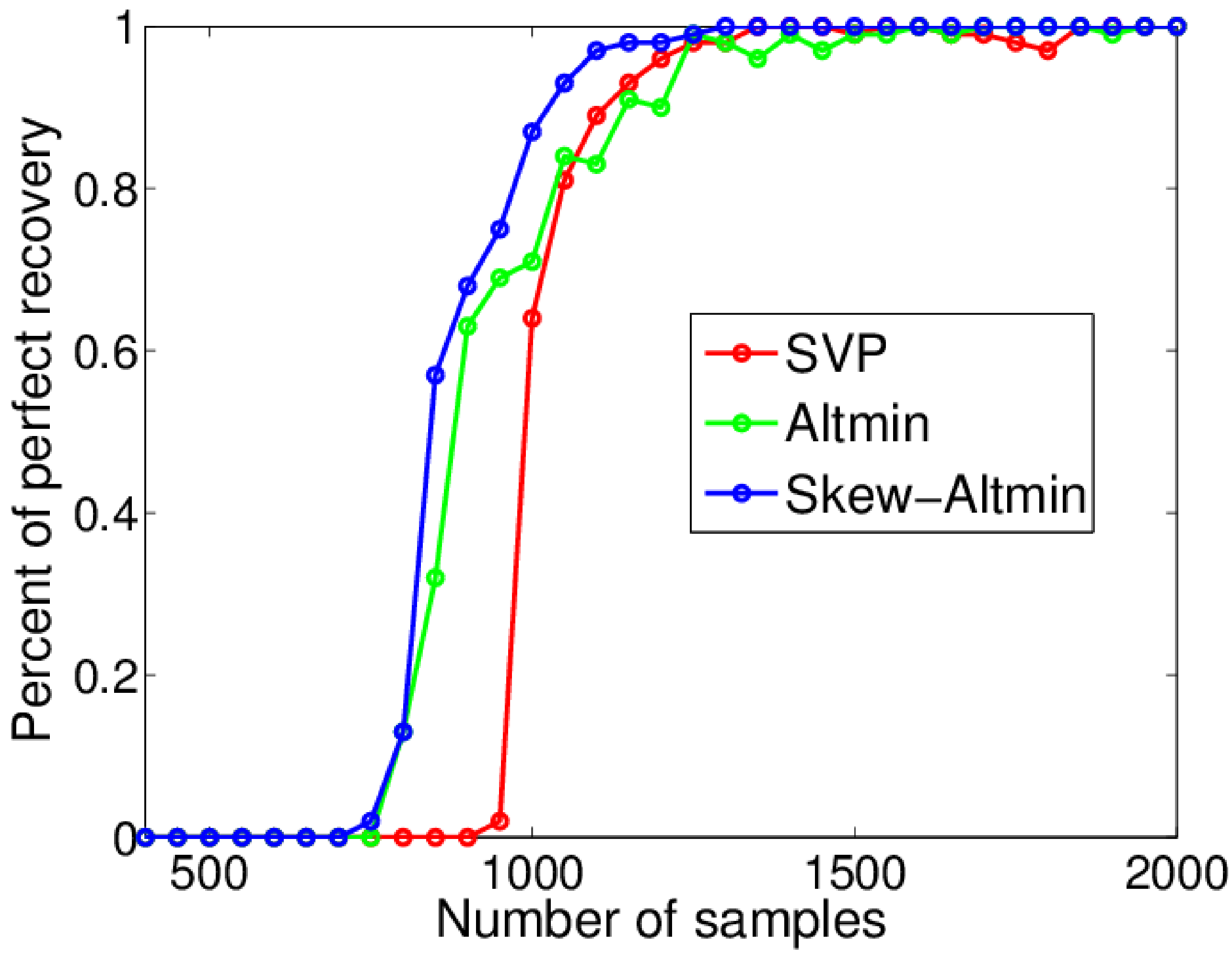
$$\frac{R(Y)}{\|Y\|_F} \approx 0.37$$



Performance

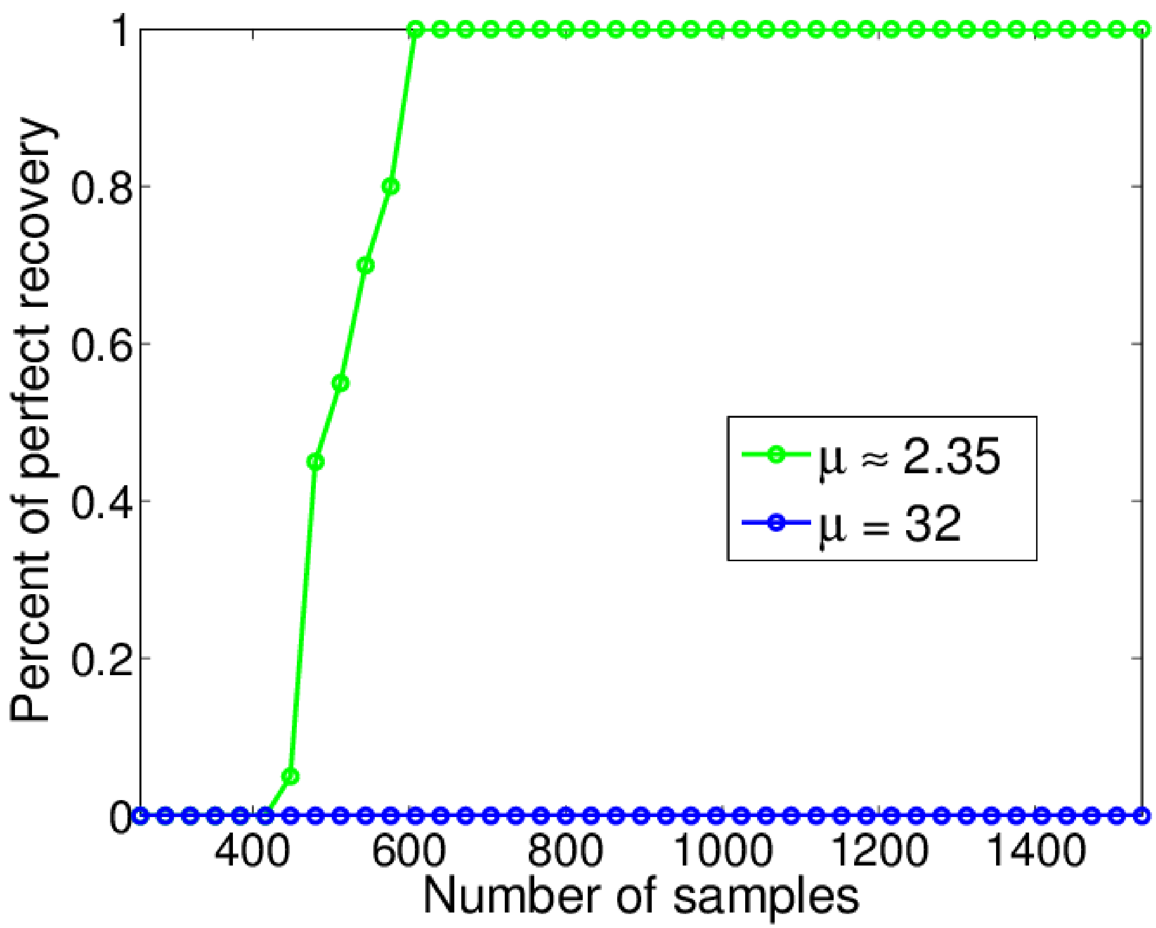
- $n = 100;\ r = 2$ (rank $= 4$)
- $s_1,\ s_2,\ a_1,\ a_2$ random with entries $\mathrm{U}[0,1]$
- coherence: low
- non-transitivity:

$$\frac{R(Y)}{\|Y\|_F} \approx 0.37$$



Performance

- $n = 64$; $r = 2$ (rank $= 4$)
- low coherence: s_k, a_k random with entries $\mathrm{U}[0,1]$
- high coherence: s_1 from identity matrix; s_2, a_k ~ iid $\mathrm{U}[0,1]$



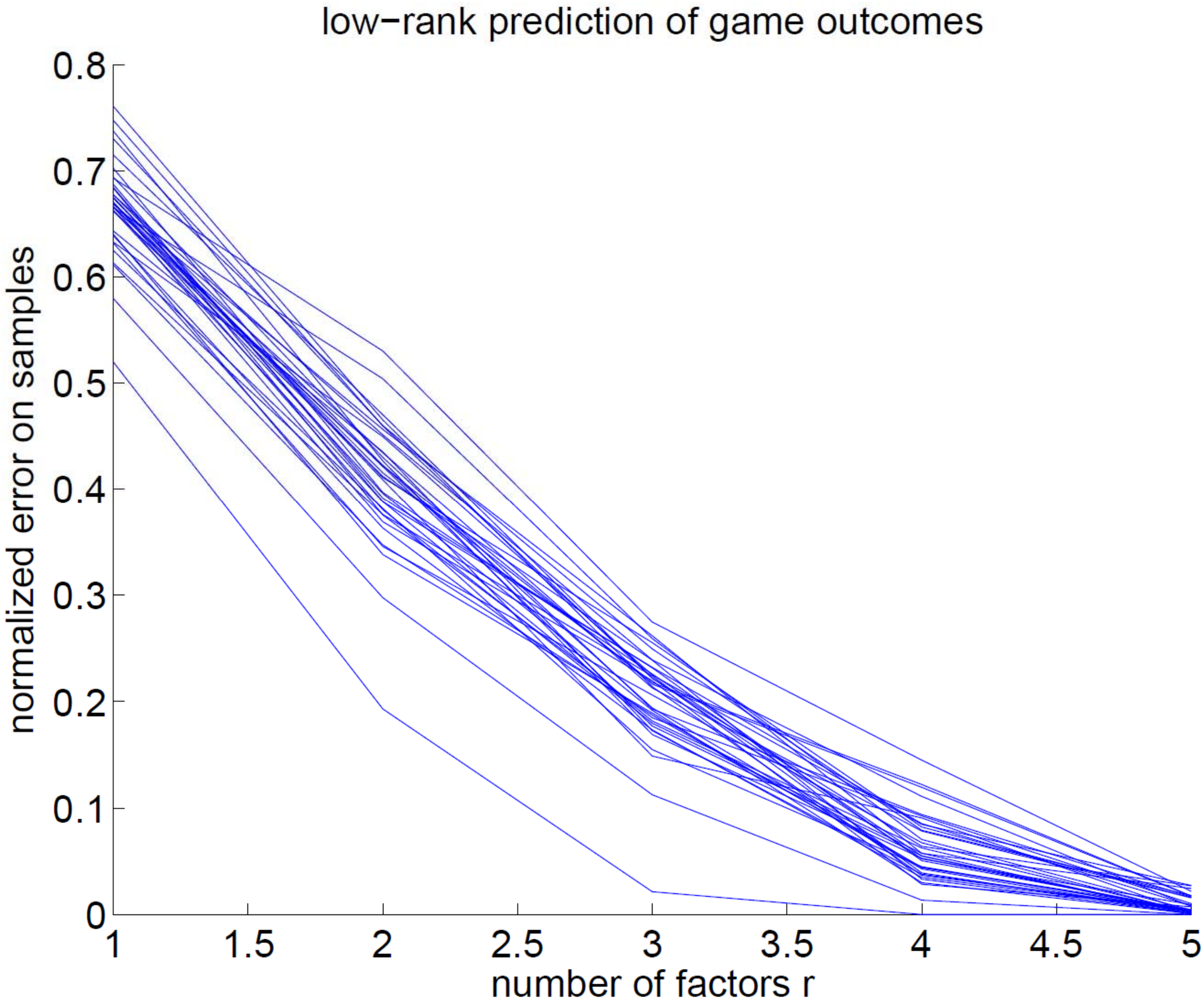
NFL Game Outcomes (1978-2013)

$$n \approx 30$$

teams

$$m \approx 200$$

unique matchups



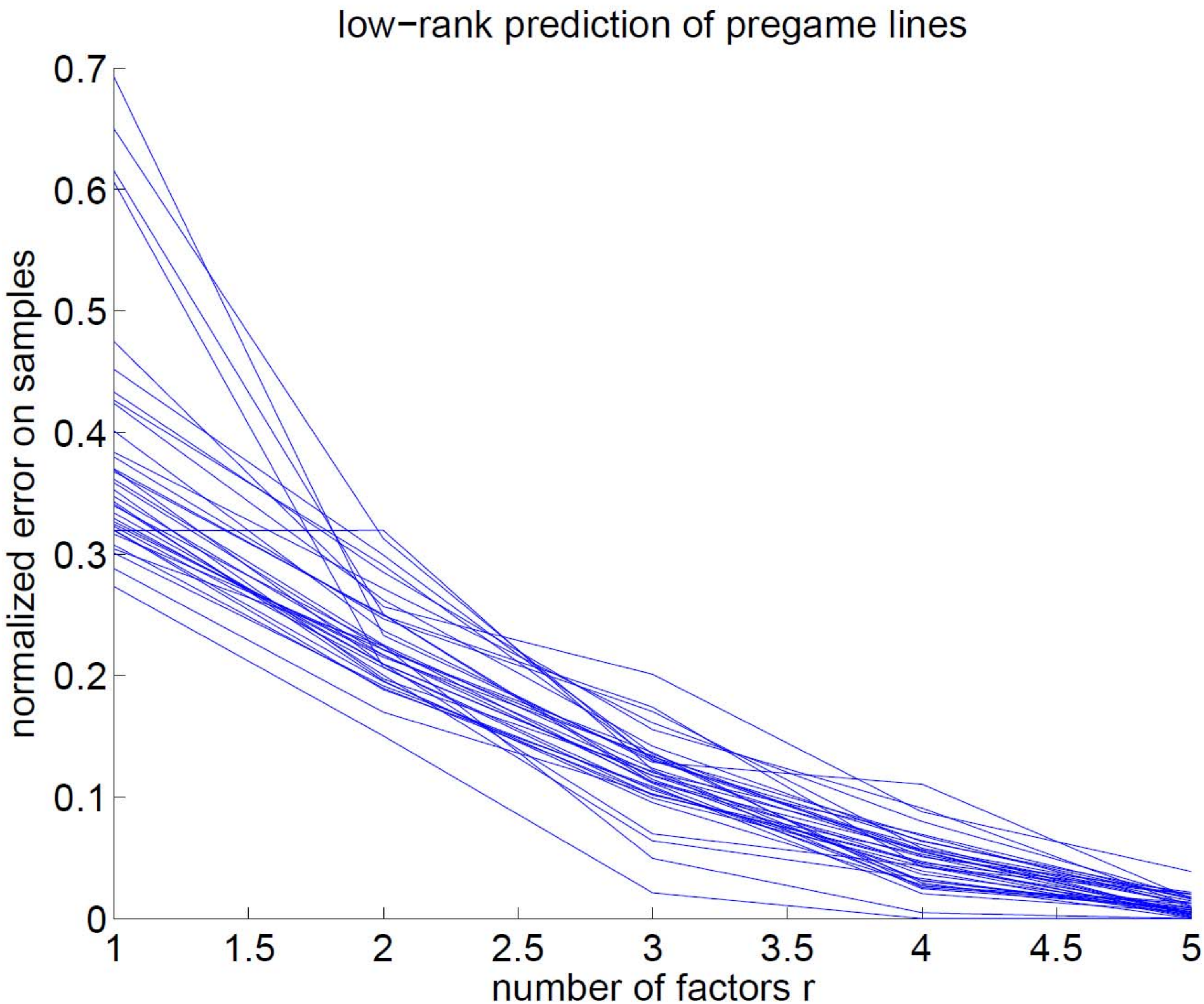
NFL Pregame Lines (1978-2013)

$$n \approx 30$$

teams

$$m \approx 200$$

unique matchups



Conclusions

- Low-rank models can support non-transitivity
- Matrix structure determined by feature vectors
  - could also give insight into leverage score sampling
- Skew-symmetric Altmin preserves structure, performs well
- Ongoing work
  - algorithm analysis
  - evaluating model for real data sets

IEEE Journal of Selected Topics in Signal Processing (J-STSP)

Special Issue on
Structured Matrices in Signal and Data Processing

- Low-rank matrix recovery
- Blind deconvolution and phase retrieval
- Matrix-based recommendation systems and collaborative filtering
- Non-negative matrix factorization
- Blind source separation
- Computer vision
- Matrix structures in radar and sensor array signal processing
- Subspace identification and tracking
- Dictionary learning and sparse coding

Manuscript submission due: ~~July 15, 2015~~ **July 30, 2015**

mines.edu/~mwakin

Proof by Induction

- Suppose that Y is transitive and for some $s(1)$, $s(2)$, $s(3)$,

$$Y = \begin{bmatrix} 0 & s(1)-s(2) & s(1)-s(3) & Y(1,4) \\ s(2)-s(1) & 0 & s(2)-s(3) & Y(2,4) \\ s(3)-s(1) & s(3)-s(2) & 0 & Y(3,4) \\ -Y(1,4) & -Y(2,4) & -Y(3,4) & 0 \end{bmatrix}$$

- Define $s(4) := s(1) - Y(1,4)$. Then for any $i = 1,\ 2,\ 3$,

$$\begin{aligned} Y(i,4) &= Y(i,1) + Y(1,4) \quad \text{(by transitivity)} \\ &= (s(i) - s(1)) + Y(1,4) \\ &= s(i) - (s(1) - Y(1,4)) \\ &= s(i) - s(4). \end{aligned}$$