Modeling and Recovering Non-Transitive Pairwise Comparison Matrices

Dehui Yang and Michael B. Wakin
Colorado School of Mines
Email: {dyang, mwakin}@mines.edu

Abstract—Pairwise comparison matrices arise in numerous applications including collaborative filtering, elections, economic exchanges, etc. In this paper, we propose a new low-rank model for pairwise comparison matrices that accommodates non-transitive pairwise comparisons. Based on this model, we consider the regime where one has limited observations of a pairwise comparison matrix and wants to reconstruct the whole matrix from these observations using matrix completion. To do this, we adopt a recently developed alternating minimization algorithm to this particular matrix completion problem and derive a theoretical guarantee for its performance. Numerical simulations using synthetic data support our proposed approach.

I. INTRODUCTION

Pairwise comparison matrices arise in a number of areas ranging from recommendation systems, economics, and psychology to competitions involving elections, sports, etc. For example, in recommendation systems, pairwise comparisons can be formed by using available user movie ratings. A pairwise comparison matrix $Y$ encodes all possible comparisons between pairs of items in a set: for a set of $n$ items, $Y$ will have size $n \times n$, and $Y(i,j)$ will denote the strength of the preference of item $i$ over item $j$. In many applications, due to the increasingly large volume of data sets, pairwise comparison matrices often have some number of missing entries. Given the available pairwise comparisons, inferring the missing entries of the matrix can be valuable for making better decisions and recommendations.

Pairwise comparison matrices are also relevant to the rank aggregation problem, where one seeks to rank a collection of items based on pairwise comparisons of the items. In [3], the authors assume that each item $i$, $1 \leq i \leq n$, has an intrinsic value $s(i)$ and describe a rank-two model for the resulting pairwise comparison matrix. In their model, $Y(i,j)$ takes the value $s(i) - s(j)$; this dictates that the full matrix $Y$ has rank two. Given partial observations of $Y$, they propose to fill in the missing entries of $Y$ using nuclear norm minimization, extract the intrinsic score vector $s = [s(1) \ s(2) \ \cdots \ s(n)]^T \in \mathbb{R}^n$, and then rank the items based on the values of the scores. However, there are inherent limitations to this approach. For example, the rank-two model above is only meaningful when the pairwise comparisons are transitive, i.e., when $Y(i,j) + Y(j,k) + Y(k,i) = 0$ for all $i, j, k$. In fact, this transitive property requires the underlying pairwise comparison matrix $Y$ to be rank-two, to be skew-symmetric (i.e., to satisfy $Y(i,j) = -Y(j,i)$ for all $i, j$), and to have the form of $Y(i,j) = s(i) - s(j)$ for some $s \in \mathbb{R}^n$ [3]. In other words, transitivity is necessary and sufficient for obtaining a consistent ranking.

However, social choice theory indicates that real-world data provided by human beings is far more complicated and often exhibits non-transitive behavior [6]. For instance, while people who like item A more than item B, and like item B more than item C, will generally like item A more than item C, this is not always the case. The comparison of two items may rely on multiple factors nonlinearly rather than simply on the difference of two individual intrinsic values. For such scenarios, it is desirable to have a new model that maintains a low complexity but has the ability to capture these complicated pairwise interactions and non-transitive behavior. An efficient algorithm that can infer the missing pairwise comparisons from available ones is also required. In these situations, where comparisons can rely on multiple intrinsic (and typically hidden) factors and many of the pairwise comparisons are non-transitive, one can also argue that it is more meaningful to focus on recovering the pairwise comparison matrix $Y$ itself rather than on recovering a scalar ranking of the items.

In this paper, we develop a new low-rank model for non-transitive pairwise comparison matrices. Our model is based on the fact that any skew-symmetric matrix $Y$ with rank at most $2r$ can be decomposed in a particular way as a sum of low-rank components:

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

for some vectors $s_1, s_2, \ldots, s_r \in \mathbb{R}^n$ and $a_1, a_2, \ldots, a_r \in \mathbb{R}^n$.

Based on this model, we adopt the alternating minimization algorithm [5] for low-rank matrix completion [2], allowing reconstruction of missing pairwise comparisons.

The rest of this paper is structured as follows. Section II presents our model and discusses its connection with the transitive pairwise comparison matrices from [3]. In Section III, we describe the alternating minimization algorithm for pairwise comparison matrix completion and give a theoretical guarantee on its performance under our model. Section IV provides numerical simulations on synthetic data. Section V discusses future work.
II. MODELS FOR PAIRWISE COMPARISON MATRICES

A. New model for pairwise comparison matrices

Our model for pairwise comparison matrices is based on the following observation. Suppose that there are $r$ latent factors on which pairwise comparisons among $n$ items are based. For each $k$, $1 \leq k \leq r$, the comparison between item $i$ and item $j$ based on the $k^{th}$ factor is determined by the difference

$$s_k(i) a_k(j) - s_k(j) a_k(i),$$

where $s_k = [s_k(1) \ s_k(2) \ \cdots \ s_k(n)]^T \in \mathbb{R}^n$ is a vector of values for the $n$ items associated with feature $k$, and $a_k = [a_k(1) \ a_k(2) \ \cdots \ a_k(n)]^T \in \mathbb{R}^n$ is a vector of weights for the $n$ items associated with feature $k$. The interactions of the "cross terms" in the expression $s_k(i) a_k(j) - s_k(j) a_k(i)$ allow $a_k(j)$ to inhibit $s_k(i)$ or $a_k(i)$ to inhibit $s_k(j)$. For the final pairwise comparison between item $i$ and item $j$, we sum the comparisons based on the $r$ factors:

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

This model has a simple expression in matrix form. Defining $Y := s_k a_k^T - a_k s_k^T$, we have

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

This model has a number of possible applications in practical scenarios. Consider an example involving competitions between football teams. Suppose $s_1 \in \mathbb{R}^n$ represents the strength of each team’s offense (with high numbers being better), and suppose $a_1 \in \mathbb{R}^n$ represents the strength of each team’s defense (with low numbers being better). Then $s_1(i) a_1(j)$ could be a reasonable model for the number of points that team $i$ is expected to score against team $j$, $s_1(j) a_1(i)$ could model the number of points that team $j$ is expected to score against team $i$, and the difference $s_1(i) a_1(j) - s_1(j) a_1(i)$ could model the expected margin of victory (or loss, if negative) of team $i$ over team $j$. The incorporation of additional offensive factors $s_2, \ldots, s_r$ and defensive factors $a_2, \ldots, a_r$ could model the use of different player configurations, special teams (field goals and kickoffs), etc. Such models could be useful for modeling other scenarios where comparisons involve multiple features and interactions where features are inhibited or accentuated.

By inspection, one can deduce that the matrix $Y$ defined in (1) is skew-symmetric and has rank at most $2r$. In fact, the property of skew-symmetry is intimately related to decompositions of the form appearing in (1). This is formalized in Lemma II.1.

Lemma II.1. [1] Any skew-symmetric matrix $A \in \mathbb{R}^{n \times n}$ with rank at most $2r$ can be decomposed as $A = \sum_{k=1}^{r} x_k y_k^T - y_k x_k^T$ for some vectors $x_1, x_2, \ldots, x_r \in \mathbb{R}^n$ and $y_1, y_2, \ldots, y_r \in \mathbb{R}^n$.

The following results will also be useful for our analysis.

Lemma II.2. [3] Suppose that $A = -AT$ is an $n \times n$ real skew-symmetric matrix with rank $2r$. Then $A$ can be factorized into the form

$$A = X \begin{bmatrix} 0 & \lambda_1 & 0 & \cdots & 0 \\ \lambda_1 & 0 & & & \\ 0 & -\lambda_2 & 0 & & \\ & \ddots & \ddots & \ddots & \\ 0 & & \ddots & \ddots & 0 \\ \end{bmatrix} X^T,$$

where $X$ is an orthonormal matrix that depends on $A$, and $T$ is a block diagonal matrix determined by $\lambda_1, \lambda_2, \ldots, \lambda_r$. Since $A$ is a real skew-symmetric matrix, the eigenvalues of $A$ are purely imaginary and are given by $\pm i \lambda_1, \pm i \lambda_2, \ldots, \pm i \lambda_r$, where $i$ is the imaginary unit.

Corollary II.3. [3] The reduced singular value decomposition (SVD) of an $n \times n$, rank $2r$, real skew-symmetric matrix $A$ is given by the following:

$$A = X \begin{bmatrix} \lambda_1 & & & \\ 0 & \lambda_2 & & \\ & \ddots & \ddots & \\ 0 & & \lambda_r & \\ \end{bmatrix} \Sigma_A V_A^T,$$

where $X$ and $\lambda_1, \lambda_2, \ldots, \lambda_r$ are as in Lemma II.2. In this decomposition, $\lambda_1, \lambda_1, \lambda_2, \ldots, \lambda_r, \lambda_r$ are the singular values of $A$, and $U_A$ and $V_A$ contain the left and right singular vectors of $A$, respectively.

Lemma II.4. For a rank-two, real skew-symmetric matrix $A = x_1 y_1^T - y_1 x_1^T$, the Frobenius norm $\|A\|_F$ is given by

$$\|A\|_F = \sqrt{2} \|y_1\|_2 \|x_1^c\|_2,$$ (2)

where $x_1^c = x_1 - \frac{\langle x_1, y_1 \rangle}{\|y_1\|^2} y_1$ is the component of $x_1$ that is orthogonal to $y_1$.

Proof: Denoting $x_1^o = \frac{\langle x_1, y_1 \rangle}{\|y_1\|^2} y_1$, we have $x_1 = x_1^o + x_1^c$.

Thus,

$$A = x_1 y_1^T - y_1 x_1^T$$

$$= (x_1^o + x_1^c) y_1^T - y_1 (x_1^o + x_1^c)^T$$

$$= (x_1^o y_1^T - y_1 x_1^o^T) + (x_1^c y_1^T - y_1 x_1^c^T)$$

$$= x_1^c y_1^T - y_1 x_1^c^T.$$ Considering

$$\frac{1}{w} A = \frac{x_1^c}{\|x_1^c\|_2} \frac{y_1^T}{\|y_1\|_2} - \frac{y_1}{\|y_1\|_2} \frac{x_1^c^T}{\|x_1^c\|_2}$$
with \( w = \| x_1^e \|_2 \| y_1 \|_2 \). According to Corollary II.3, we have the following SVD for \( \frac{1}{w} A \)

\[
\frac{1}{w} A = \begin{bmatrix}
\frac{y_1}{\| y_1 \|_2} & \frac{x^e}{\| x^e \|_2} \\
\frac{\| y_1 \|_2}{\| x^e \|_2} & \frac{y_1}{\| y_1 \|_2}
\end{bmatrix} \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix} \begin{bmatrix}
-\frac{x^e}{\| x^e \|_2} & \frac{y_1}{\| y_1 \|_2}
\end{bmatrix}^T.
\]

Therefore, we have \( \| A \|_F = \sqrt{2} \). Consequently, \( \| A \|_F = \sqrt{2} \| y_1 \|_2 \| x^e \|_2 \).

B. Connection to the transitive rank-two skew-symmetric model

Let \( e \in \mathbb{R}^n \) denote the vector with all entries equal to 1. We note that when all \( a_k = e \) for all \( k \), the pairwise comparison matrix in (1) takes the form \( Y = se^T - es^T \), where \( s = \sum_{k=1}^n s_k \). In this case, our model reduces to the rank-two skew-symmetric case studied in [3]. By introducing more degrees of freedom (via the \( a_k \)) into the model, inhibitions and non-transitive behavior can be captured. In fact, this model allows for non-transitive matrices \( Y \) even in the case where \( r = 1 \) (by choosing \( a_1 \neq e \)). Therefore, our model generalizes the rank-two skew-symmetric case [3] both by allowing for non-transitivity and by allowing for higher rank.

The combinatorial Hodge theory developed in [6] allows us to quantify the degree of non-transitivity expressed in our model.

Lemma II.5. [6] Given an arbitrary skew-symmetric matrix \( Y \), denote

\[
R(s) = Y - (se^T - es^T)
\]

as the difference between \( Y \) and a transitive rank-two skew-symmetric matrix induced by a score vector \( s \in \mathbb{R}^n \). Then, the minimum norm solution of the following optimization problem

\[
\min_{s} \| R(s) \|_F
\]

is given by \( \hat{s} = \frac{1}{n} Ye \).

The proof of the lemma above is quite straightforward and can be found in [7].

To appreciate the implications of Lemma II.5, consider a scenario where many triples \((i, j, k)\) satisfy \( |Y(i, j) + Y(j, k) + Y(k, i)| \geq \gamma \) with \( \gamma \) being a positive number. In this case, \( Y \) is a pairwise comparison matrix with many non-transitive comparisons. In such a case, we might suspect that it is not appropriate to look for a score vector \( s \) that provides a ranking consistent with the pairwise comparisons. In [6], the degree of non-transitivity in \( Y \) is characterized by the residual \( \| R(\hat{s}) \|_F \) of the least squares problem (3). Hence, examining \( \| R(\hat{s}) \|_F \) can reveal whether a score vector exists that will provide a consistent ranking. If \( \| R(\hat{s}) \|_F \) is small, the pairwise comparison matrix \( Y \) is less non-transitive and the extracted score vector \( \hat{s} \) is plausible. In the extreme case where \( \| R(\hat{s}) \|_F = 0 \), the extracted score vector \( \hat{s} \) perfectly accounts for the pairwise comparisons, in that \( Y(i, j) = \hat{s}(i) - \hat{s}(j) \). On the other hand, if \( \| R(\hat{s}) \|_F \) is large, \( Y \) exhibits a high degree of non-transitivity, no score vector \( s \) exists that accurately explains the pairwise comparisons, and the transitive rank-two skew-symmetric model is not appropriate. In the discussion that follows below, we give a sufficient condition under which our proposed model is close to a transitive rank-two skew-symmetric model.

To do this, we apply the result in Lemma II.5 to our model. For the sake of simplicity in this paper, we consider the case when \( r = 1 \) and \( Y = s_1 a_1^T - a_1 s_1^T \). Then, the residual is given by

\[
R(s) = \|(s_1 a_1^T - a_1 s_1^T) - (se^T - es^T)\|_F.
\]

According to Lemma II.5, the optimal solution \( \hat{s} \) that minimizes \( R(s) \) is given by

\[
\hat{s} = \frac{1}{n} (s_1 a_1^T - a_1 s_1^T) e.
\]

Here we are interested in characterizing the size of the residual \( R(\hat{s}) \) in terms of properties of the feature and weight vectors \( s_1 \) and \( a_1 \) which were used to generate \( Y \). To bound this quantity, let us consider \( R(s_1) \), which measures the distance from \( Y \) to a transitive matrix generated using oracle knowledge of \( s_1 \). Because \( \hat{s} \) minimizes (4), it follows that \( R(\hat{s}) \leq R(s_1) \). Therefore,

\[
R(\hat{s}) \leq R(s_1)
\]

\[
= \|(s_1 a_1^T - a_1 s_1^T) - (s_1 e^T - es^T)\|_F
\]

\[
= \|s_1(a_1 - e)^T - (a_1 - e)s_1^T\|_F
\]

\[
= \sqrt{2}\|s_1\|_2\|a_1 - e\|_2,
\]

where \( (a_1 - e)^c = (a_1 - e) - \frac{(a_1 - e)^c}{\|a_1 - e\|_2} s_1 \) and the third equality above follows from Lemma II.4.

Not surprisingly, this result indicates that if \( a_1 \) is close to \( e \), meaning that \( \|a_1 - e\|_2 \) is small, then \( \| R(\hat{s}) \|_F \) is small and \( Y \) exhibits a low degree of non-transitivity. In addition, though, this result indicates it is sufficient for \( \|a_1 - e\|_2 \) to be small (even if \( \|a_1 - e\|_2 \) is large). That is, if there is a small angle between \( s_1 \) and \( a_1 - e \), then again \( \| R(\hat{s}) \|_F \) will be small and \( Y \) will exhibit a low degree of non-transitivity.

III. ALTERNATING MINIMIZATION FOR PAIRWISE COMPARISON MATRIX COMPLETION

In many applications, pairwise comparison matrices may be known only partially. In order to infer the missing entries, we consider an approach involving alternating minimization [5]. Alternating minimization emerged as a useful tool for matrix completion during the Netflix Prize competition. However, only recently has it been shown that alternating minimization can have global convergence as long as certain conditions (e.g., a satisfactory initialization, a sufficient number of measurements, an incoherence condition) are met.

Our aim is to solve

\[
\min_{U, V} \| \mathcal{P}_\Omega (Y - UV^T) \|_2^2,
\]

where \( \Omega \) is the observed index set, \( \mathcal{P}_\Omega \) zeros out all entries of a matrix outside \( \Omega \), and \( Y \) is the true pairwise comparison matrix. Under the assumption that \( Y \) obeys the model (1) (and thus has rank at most \( 2r \)), we consider \( U \in \mathbb{R}^{n \times 2r} \) and \( V \in \mathbb{R}^{n \times 2r} \). The alternating minimization algorithm for solving (7)
Algorithm 1 Alternating Minimization for Pairwise Comparison Matrix Completion

**input:** partial pairwise comparison matrix $\mathcal{P}_{12}(Y)$, number of latent factors $r$

**initialize:** $U^0$: the top $2r$ left singular vectors of $\frac{1}{p}\mathcal{P}_{12}(Y)$, where $p$ is the ratio between the number of measurements and $n^2$

**while:** stopping criterion not met do

**solve:** $V^{k+1} = \arg\min_{V \in \mathbb{R}^{n \times 2r}} ||\mathcal{P}_{12}(Y - U^kV^T)||_F^2$

and then $U^{k+1} = \arg\min_{U \in \mathbb{R}^{n \times 2r}} ||\mathcal{P}_{12}(Y - UV^{k+1T})||_F^2$.

**output:** $\hat{U} = U^k$, $\hat{V} = V^k$

---

**Theorem III.2.** Suppose that $Y = \sum_{i=1}^r \left( \lambda_i s_i a_i^T - \lambda_i a_i s_i^T \right)$ and that the matrix $[s_1, s_2, \ldots, s_r, a_1, a_2, \ldots, a_r]$ has orthonormal columns and coherence $\mu$. Assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$. Suppose that we sample $m = O \left( \mu^2 \left( \frac{1}{\lambda_1} \right)^6 r^7 n \log n \log \frac{r||Y||_F}{\epsilon} \right)$ observations of $Y$ uniformly at random. Then with $K = O(\log \frac{1}{\epsilon})$ iterations, Algorithm 1 converges geometrically to the true $Y$ and the recovered $\hat{Y} = \hat{U}\hat{V}^T$ satisfies $||Y - \hat{Y}||_F \leq \epsilon$.

**Proof:** From Lemma 2.2 and Corollary 2.3, we conclude that the condition number of $Y$ is $\frac{r\lambda_1}{\lambda_r}$. Meanwhile, from our discussion above, we know that $U_Y$ and $V_Y$ have the same coherence as $[s_1, s_2, \ldots, s_r, a_1, a_2, \ldots, a_r]$. Therefore, $Y$ has coherence $\mu$ as well. By applying Theorem 2.5 of [5], we conclude that observing $m = O \left( \mu^2 \left( \frac{1}{\lambda_1} \right)^6 r^7 n \log n \log \frac{r||Y||_F}{\epsilon} \right)$ random entries is sufficient for recovery with the predefined accuracy $\epsilon$.

In the special case when $r = 1$, we can eliminate the condition number term in Theorem III.2 since rank-two skew-symmetric matrices have two repeated singular values. In particular, the condition number is 1. Therefore, when $r = 1$, $m = O(\mu^2 n \log n \log \frac{r||Y||_F}{\epsilon})$ uniformly random observations of $Y$ are sufficient for recovery up to an accuracy of $\epsilon$.

We note that, in practical scenarios, pairwise comparison matrices may not be perfectly low rank. Algorithm 1 can in fact be used for noisy pairwise comparison matrix completion although we do not provide a recovery guarantee for this case. Extending the low-rank model may be another way to deal with noise in pairwise comparison matrices. For example, by combining the low-rank model proposed here with a sparse model for the noise, it may be possible to recover the low-rank matrix in the presence of outliers.

**IV. NUMERICAL SIMULATIONS**

We provide simulations on synthetic data to demonstrate the effectiveness of Algorithm 1. We set $n = 100$ and consider the cases $r = 1$ and $r = 2$. In these simulations, the coherence in [3].

---

**Fig. 1.** Recovery performance for $r = 1$.

**Fig. 2.** Recovery performance for $r = 2$. 

1Technically, for this theorem to hold one must partition the observation set $\Omega$ and use a unique set of measurements during each iteration of Algorithm 1. See [5] for further details.

---

**Definition III.1.** The coherence of a subspace $U$ with dimension $r$ is defined by

$$\mu(U) = \frac{n}{r} \max_{1 \leq i \leq n} ||P_U e_i||^2,$$

where $P_U$ denotes the orthogonal projection onto $U$ and $e_i$ is a vector of all zeros with a single 1 in position $i$.
parameter $\mu$ is restricted to the range $[1, 50]$ when $r = 1$ and $[1, 25]$ when $r = 2$. We generate elements of $s_k$ and $a_k$ uniformly at random between 0 and 1. Then, we construct the pairwise comparison matrix according to (1). We reconstruct $Y$ from various numbers of random observations. For each trial, we declare success if the reconstruction relative error is less than $10^{-3}$. We note that on average, for $r = 1$, the value of $\mu$ is typically around 3, and for $r = 2$, the value of $\mu$ is typically around 2.35. Moreover, the normalized residual $\|R(\hat{S})\|_F$ is typically around 0.37 for $r = 1$ and $r = 2$; thus, the pairwise comparison matrices $Y$ in both experiments have a certain degree of non-transitivity. Figures 1 and 2 show the reconstruction performance of the alternating minimization algorithm and the singular value projection (SVP) algorithm for $r = 1$ (rank 2) and $r = 2$ (rank 4), respectively. SVP is a matrix completion algorithm first proposed in [4] and adopted in [3] for pairwise comparison matrix completion. Although SVP performs comparably to alternating minimization in both experiments, it has a drawback of Algorithm 1 is that it cannot preserve the skew-symmetric property during the iteration process. An alternative is to solve the following alternating minimization problem which can preserve the skew-symmetric property during the iterations:

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

We plan to derive theoretical guarantees for this skew-symmetric recovery algorithm. Additional testing and validation of our model and algorithm on real-world data sets are topics of current work.

**ACKNOWLEDGMENT**

This work was supported by NSF CAREER grant CCF-1149225.

**REFERENCES**


**V. QUESTIONS FOR FUTURE WORK**

In the case where all latent factors and weights are positive and bounded, we would like to explore the problem of identifying these latent factors. This may be possible thanks to the success of non-negative matrix factorization. Also, one drawback of Algorithm 1 is that it cannot preserve the skew-symmetry of $UV^T$ during the iteration process. An alternative is to solve the following alternating minimization problem which can preserve the skew-symmetric property during the iterations:

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