Low-Rank Decomposition of Multi-Way Arrays: A Signal Processing Perspective

Nikos Sidiropoulos

Dept. ECE, TUC-Greece & UMN-U.S.A

nikos@telecom.tuc.gr
Acknowledgments

3-way Students: X. Liu (U. Louisville), T. Jiang (KTH)

3-way Collaborators: R. Bro (Denmark), J. ten Berge, A. Smilde (Netherlands), R. Rocci (Italy), A. Gerdsman, S. Vorobyov, Y. Rong (Canada & Germany)

Sponsors: NSF CCR 9733540, 0096165, 9979295, 0096164; ONR

N CT A Cooperative Agreement DADD19-01-2-0011
N/NO0014-99-1-0693; DARPA/ATO MDA 972-01-0056; ARL C &
List of Applications - I

- Blind multiuser detection-estimation in DS-CDMA, using Rx
- Multiple-invariance sensor array processing (MI-SAP)
- Joint detection-estimation in SIMO/MIMO OFDM systems subject to CFO, using receive diversity
- 3-D Radar clutter modeling and mitigation
- Multi-dimensional harmonic retrieval w/applications in DOA
- Blind decoding of a class of linear space-time codes
- Exploratory data analysis: clustering, scatter plots, multi-dimensional scaling
- Exploratory data analysis: clustering, scatter plots, multi-dimensional scaling
- Antenna array
- Blind multiuser detection-estimation in DS-CDMA, using Rx
List of Applications - II

- Chromatography, spectroscopy, magnetic resonance, ...
- Analysis of individual differences (Psychology)
  ("super-symmetric")
- HOS-based parameter estimation and signal separation
- ACMA
- Blind source separation for multi-channel speech signals
- Joint diagonalization problems (symmetric)
Three-Way Arrays

Two-way arrays, AKA matrices:

\[
X = [x_{ij}]
\]

Three-way arrays:

\[
X = [x_{ijk}]
\]

CDMA w/ Rx Ant array @ baseband: chip × symbol × antenna

MI SAP: Subarray × element × snapshot

MISAP: Subarray × element × snapshot

Multiuser MIMO-OFDM: antenna × FFT bin × symbol

Spectroscopy, NMR, Radar, analysis of food attributes (judge attribute × sample) × personality traits

Multiuser MIMO-OFDM: antenna × FFT bin × symbol

Three-way arrays:

\[
(f \times I) : [f \times x] = X
\]

Two-way arrays, AKA matrices:

\[
(f \times I)
\]
Three-Way vs Two-Way Arrays - Similarities

Three-way rank-one factor: rank-one 3-WAY ARRAY outer product of vectors (containing all triple products)

Two-way rank-one factor: rank-one MATRIX outer product of 2 vectors (containing all double products)

Rank := smallest number of rank-one "factors" ("terms" is probably better) for exact additive decomposition (same concept for both 2-way and 3-way)

- Three-way (containing all triple products) - same concept of 3 vectors (containing all triple products) - same concept

- Two-way rank-one factor: rank-one MATRIX outer product of 2 vectors (containing all double products)
Three-Way vs Two-Way Arrays - Differences

- **Two-way**:
  - \( \text{rank}(R) = \min(I, J) \text{ w.p. 1} \)
- **Three-way**:
  - \( \text{rank}(R) \) is a RV (2 \( \leq \) w.p. 0.3, 3 \( \leq \) w.p. 0.7)

- Three-way: row-rank = column-rank = tube-rank = rank
- Two-way: row-rank \neq \text{column-rank} \neq \text{tube-rank} rank


Decomposition over \( \mathbb{C} \) [Burgisser, Clausen, Shokrollahi, Algebraic]

For decomposition over \( \mathbb{R} \), theory more developed for ten Berge; general results for maximal rank and typical rank poorly

- Three-way: Except for loose bounds and special cases (Kruskal), J.M.F.
- Two-way: Closed (\( \mathbb{R} \) versus \( \mathbb{C} \)); three-way: rank sensitive to \( \mathbb{R} \) versus \( \mathbb{C} \)
- Two-way: Rank insensitive to whether or not underlying field is open or closed
- Three-way: \( \text{rank}(\text{randn}(2, 2, 2)) = \text{a RV (2 w.p. 0.3, 3 w.p. 0.7}) \)

Three-way: row-rank(\( f \times I \)) = rank(\( f(1, I) = \min(1, I) \)) = w.p. 1.

Three-way: row-rank(\( f(1, I) \)) = \( f(I, I) \subseteq \text{rank} \neq \text{column-rank} 

- Three-way vs Two-Way Arrays - Differences

\[
\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}
\]
\[
\text{vec}(\mathbf{AB}) = (\mathbf{B} \odot \mathbf{A})(D)
\]

\[
\begin{bmatrix}
75 \\
45 \\
15 \\
25 \\
15 \\
20 \\
5 \\
20
\end{bmatrix}
\]

\[
\mathbf{A} \odot \mathbf{B} =
\begin{bmatrix}
0 & 30 & 25 \\
15 & 20 & 15 \\
10 & 5 & 10
\end{bmatrix}
\]

\[
\mathbf{B} =
\begin{bmatrix}
4 & 3 \\
2 & 1 \\
1 & 2
\end{bmatrix}
\]

\[
\mathbf{A} =
\begin{bmatrix}
9
\end{bmatrix}
\]

**Khatri-Rao Product:**

\[
\text{Column-wise Kronecker Product}
\]

**TKC & UMN**

LRD of Three-Way Arrays: Notation

Scalar:
\[ x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \]

Slabs:
\[ X_k = A D_k(B^T) \]

Matrix:
\[ X (KJ) = (B C) A^T \]

Vector:
\[ x_{(KJI)} = \text{vec}(X) \]

Scalar:
\[ \sum_{f=1}^{F} \sum_{i,j,k} x_{i,j,k} = \sum_{i,j,k} x_{i,j,k} \]
Scalar: $x_{i_1 \ldots i_N} = F \sum_{f=1}^{N} \prod_{n=1}^{N} a(n) x_{i_n};$

Matrix: $X(I_1 I_2 \ldots I_N) = A(N_1) A(N_2) \ldots A(1) T$

Vector: $\mathbf{x}(I_1 I_2 \ldots I_N) = \text{vec} X(I_1 I_2 \ldots I_N) = A(N) A(N_1) A(N_2) \ldots A(1) 1^{1 \times N_1 \ldots N_2 \ldots 1}$

Matrix of N-Way Arrays: Notation
Closer look at applications: Data modeling

**CDMA**: (i, j, k, f, \( p_A \)) \((R\), antenna, symbol snapshot, chiph, user)

**MI-SAP**: \( A \) is response of reference subarray, \( B \) is temporal signal

Blind signature estimation from covariance data: Symmetric

\[ R \xi = A D \gamma H (p_A) \]

**PARAFAC/CANDECOMP (INDSCAL)**:

Displaced but otherwise identical subarray matrix (usually denoted \( s \)), \( D \) holds the phase shifts for the \( k \)-th

\[ X \xi = A D \gamma (c) B \]

\( i, j, k = 1, \ldots, K \)
Fact 2: Low-rank 3- and higher-way array decomposition (PARAFAC) is unique under certain conditions.

Fact 1: Low-rank matrix (2-way array) decomposition is not unique for rank $\leq 1$.

### Early Take-Home Point

- **Code**: Represents the outcome or result of the decomposition.
- **Symbol**: Indicative of the input or data being processed.

$$
\begin{align*}
\text{Code} + \text{Symbol} &= \overline{X} \\
\text{Code} + \text{Symbol} &= X
\end{align*}
$$
\[
\begin{align*}
X &= a_1 b^T + \cdots + a_r X b^T + \cdots + a_i X b^T + \cdots + a_i X b^T \\
&= \mathbf{A}^T \mathbf{B} \mathbf{X}
\end{align*}
\]
Reverse engineering of soup?

Can only guess recipe.
Same ingredients, different proportions recipe!
Collect $K$ OFDM symbol snapshots

\[ Y \left( P_{HF} \right) = W \]

Deteriministic approach, works with small sample sizes (channel coherence), relaxed ID conditions, performance within 2 dB from non-blind MMSE clairvoyant RX

PARAFAC model (w/ special structure) blindly identifiable [Jiang & Sidiropoulos, 02]

$F \bullet \cdots \bullet F = \begin{bmatrix} 1 & \ddots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} W + B T + W_i = A D_i B T + W_i, i = 1, \ldots, I$
SIMO-OFDM/CFO - Results

Uniqueness

$\text{Uniqueness} = \sum_{i=1}^{n} (k - \text{ranks}) i \geq 2F$

[Kruskal, 1977], $N = 3$, $C$

[Sidiropoulos et al., IEEEE TSP, 2000]: $N = 3$, $C$


$\sum_{n} k = 1 \geq F + 2$

$\text{rank} \geq \text{maximum } k \text{ such that every } r \text{ columns are linearly independent}$

$\text{rank} \geq \text{maximum } k \text{ such that every } r \text{ columns are linearly independent}$

$N_{19} = \sum k^r + k^r + k^r \geq 2F + 2$

$\text{rank} \geq \text{maximum } k \text{ such that every } r \text{ columns are linearly independent}$

$\text{rank} \geq \text{maximum } k \text{ such that every } r \text{ columns are linearly independent}$
Kruskal's Permutation Lemma \cite{Kruskal,1977}:

Consider a matrix $$A$$ having no zero column, and let $$\bar{F} = \{ \not\exists x \mid \text{weight of non-zero elements of } x \text{ in } \bar{F} \}$$.

If also $$\bar{F} \preceq \bar{F}'$$, then there exist a permutation matrix and a non-singular diagonal matrix such that $$A = \mathbf{V}\mathbf{P}\mathbf{D}$$.

To show for a pair of square nonsingular matrices (use rows of $$\mathbf{V}^{-1}$$), but the result is very deep and difficult for fat matrices - see \cite{Jiang \& Sidiropoulos, TSP:04}.

The formula is:

$$\left( \mathbf{V}_H \mathbf{x} \right)^\top \preceq \left( \mathbf{V}_H \mathbf{x} \right)^\top$$

This implies:

$$\mathbf{I} + \mathbf{V}_d - \mathbf{D} \succeq \left( \mathbf{V}_H \mathbf{x} \right)^\top$$

Consider $$\mathbf{V}$$ Kruskal's Permutation Lemma \cite{Kruskal, 1977}. The condition is the weight of columns in $$\mathbf{V}$$ is the minimum weight of columns in $$\mathbf{H}$$.
$0 = \varphi_{\mathbf{b} \odot \mathbf{a}}$

whereas if $\mathbf{a} \vartriangleright 0$ or $\mathbf{b} \vartriangleright 0$, then it holds that

$$0 = \varphi_{\mathbf{b} \odot \mathbf{a}} \leq \min\{\mathbf{a} \vartriangleright 1, \mathbf{b} - I, \mathbf{b} + \mathbf{a} \vartriangleright 1, I, \mathbf{a}, \mathbf{b} \}$$

If $\mathbf{a} \vartriangleright 1$ and $\mathbf{b} \vartriangleright 1$, then it holds that

\[ \mathbf{a} \odot \mathbf{b} = \mathbf{0} \]

\[ \mathbf{0} = \mathbf{a} \odot \mathbf{b} \]

Property: [Sidiropoulos & Liu, 1999; Sidiropoulos & Bro, 2000]
A = AV \cdot I, B = BV \cdot I, C = CV \cdot I, V^2 V^\dagger = I.

Stepping Stone

Theorem: Given \( X = (A, B, C) \) with \( A \neq B \neq C \), it is necessary for uniqueness of \( A, B, C \) that \( \min_k (k_A, k, k_B) \geq 2 \).

Theorem: Given \( X = (A, B, C) \), with \( A \neq B \neq C \), it is necessary for uniqueness of \( A, B, C \) that \( \min_k (r_A B^\dagger, r_B^\dagger, r_C^\dagger) = 2 \).

For some \( A : F \times I \), and \( C : F \), and \( B : F \times I \), then there exists a permutation matrix \( P \) and diagonal scaling matrices \( \Lambda_1, \Lambda_2, \Lambda_3 \) such that for permutation and scaling of columns, meaning that if \( \bar{X} = (A, B, C) \), then \( \min \in \{\bar{X}^T A, \bar{X}^T B, \bar{X}^T C\} \).

If, in addition, \( r_B \neq F \), then it is also necessary that \( \min (r_B^\dagger + k_A^\dagger, k_B, k_C) \).
Is Kruskal’s Condition Necessary?

No linear combination of two or more columns of $A \odot B$

No for $F > 3$

Jiang & Sidiropoulos, 03: new insights that explain part of the puzzle: E.g., for $F = r \subset F$, the following condition has been proven to be necessary and sufficient:

A nonlinear combination of two or more columns of $A$ can be written as KRP of twovectors

Ten Berge & Sidiropoulos, Psychometrika, 2002: Yes for $F \in \{2, 3\}$

Long-held conjecture (Kruskal, 89): Yes
So, LRD for 3- or higher-way arrays unique, provided rank is "low enough", often works for rank $>1$

- In CDMA application, each user contributes a rank-1 factor.
- In MI-SAP application, each source contributes a rank-1 factor.
- In multiuser MIMO-OFDM, each Tx antenna contributes rank-1 factor.

Hence if the number of users/sources/Tx is not too big, completely blind identification is possible.

Resulting ID conditions beat anything published to date.
**Algorithms**

- SVD/EVD or TLS 2-slab solution (similar to ESPRIT) in some cases (but conditions for this to work are restrictive).
- Workhorse: ALS [Harshman, 1970]: LS-driven (ML for AWGN), iterative, initialized using 2-slab solution or multiple random cold starts.
- ALS → monotone convergence, usually to global minimum (uniqueness), close to CRB for $F << IJK$. 

---

circular fashion until convergence is fit (guaranteed).

Similarly for the CLS updates of $B$, $C$ (symmetry); repeat in a
given intermediate estimates of $B$, $C$, solve for conditional LS update of

$$
\mathbf{A}_{CLS} = \left( (I \times I) \mathbf{X}_{t} \right) \mathbf{C} \odot \mathbf{B}
$$

ALS is based on matrix view:
Algorithms

ALS initialization matters, not crucial for heavily over-determined conditions

Possible if e.g., a subset of columns in A is known [Jiang & Sidiropoulos, JASP/SMART 2003]

In general, no "algebraic" solution like SVD

G-N converges faster than ALS, but it may fail

"ESPRIIT-like..."

COMFACT (Tucker3 compression), G-N, Levenberg, ATLD, DTLD

All rank-1 updates possible [Kroonenberg], but interior problems

ALS initialization matters, not crucial for heavily over-determined
Robust algorithms perform well for Laplacian, Cauchy, and not far from optimal in the Gaussian case.

Alternatingly, very simple element-wise updating using weighted median filtering [Vorobyov, Rong, Sidiropoulos, Gershman, 2003] is equivalent to a LP problem — alternating LP [Vorobyov, Rong, 2003] — similar to ALS, each conditional matrix update can be shown robust across α-stable.

Least Absolute Error (LAE) criterion: optimal (ML) for Laplacian, Cauchy-distributed errors, outliers.

Alternative, very simple element-wise updating using weighted median filtering [Vorobyov, Rong, Sidiropoulos, Gershman, 2003].
CRBs for the PARAFAC model

- only dependent on noise pdf
- scaled versions of the Gaussian CRB; scaling parameter
- Laplacian, Cauchy [Yurovsky, Ron, Sidiropoulos, Gerstenmaier]
- compact expressions for complex 3-way case & asymptotic CRB
- Laplacian, Cauchy [Yurovsky, Ron, Sidiropoulos, Gerstenmaier]

- Gaussian, 3-way & 4-way [Liu & Sidiropoulos, TSP 2001]
- Real i.i.d., Gaussian, 3-way, Complex, Circularly Symmetric i.i.d.

Dependent on how scale-permutation ambiguity is resolved
Performance

Figure 1: RMSEs versus SNR: Gaussian noise, 8 x 8 x 20, $F = 2$.
Figure 2: RMSE versus SNR: Cauchy noise, 8 × 8 × 20, $F = 2$.
Performance

Still workhorse, after all these years...

Convergence becomes extremely slow for difficult datasets, so-called swamps are possible: progress towards computational cost.

Main shortcoming of ALS and related algorithms is the high parameter estimates are still accurate. Identifiability conditions only, which means that at high SNR the performance is worse (and further from the CRB) when operating close to the identifiability boundary; but ALS works under model.

Performance is worse (and further from the CRB) when operating.

\[10 \times 10 = 100\]

Data is easy to get to the large-samples regime: e.g.

ATS works well in AWGN because it is ML-driven, and with 3-way...
Learn more - Tutorials, bibliography, papers, software...
What lies ahead & wrap-up

Take home point: 

- New exciting applications: Yours!

- Incorporation of application-specific constraints

Major challenges: Algorithms: Faster at small performance loss;

Super-symmetric models (INDSCAL, Jd. HOS)
- Application-specific constraints (e.g., Toeplitz; Jd. symmetric x)
- Uniqueness under sufficient conditions; ii) Higher-way models; iii) Uniqueness under symmetric x (INDSCAL, JD, HOS)

Major challenges: Uniqueness: i) Easy to check necessary x models unique; have many applications
- Take home point: \( N < 3 \)-way arrays are different; low-rank
Thank you!