Technical University of Crete
Department of Electronic and Computer Engineering

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

Contents

- Introduction & motivating list of applications
- 3-way arrays: similarities and differences with matrices
- Rank, and low-rank decomposition; 3-way notation
- Closer look at applications: Data modeling
- Uniqueness
- Algorithms
- Performance
- Web pointers
- What lies ahead & wrap-up

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. Gershman, S. Vorobyov, Y. Rong (Canada & Germany)
- Sponsors: NSF CCR 9733540, 0096165, 9979295, 0096164; ONR N/N00014-99-1-0693; DARPA/ATO MDA 972-01-0056; ARL C & N CTA Cooperative Agreement DADD19-01-2-0011

List of Applications - I

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

List of Applications - II

- Joint diagonalization problems (symmetric):
  - i) Blind spatial signature estimation from covariance matrices, using time-varying power loading, spectral color / multiple lags
  - ii) Blind source separation for multi-channel speech signals
  - iii) ACMA
- HOS-based parameter estimation and signal separation (“super-symmetric”)
- Analysis of individual differences (Psychology)
- Chromatography, spectroscopy, magnetic resonance, ...

Three-Way Arrays

- Two-way arrays, AKA matrices: $\mathbf{X} := [x_{i,j}] : (I \times J)$
- Three-way arrays: $[x_{i,j,k}] : (I \times J \times K)$
- CDMA w/ Rx Ant array @ baseband: chip $\times$ symbol $\times$ antenna
- MI SAP: subarray $\times$ element $\times$ snapshot
- Multiuser MIMO-OFDM: antenna $\times$ FFT bin $\times$ symbol
- Spectroscopy, NMR, Radar, analysis of food attributes (judge $\times$ attribute $\times$ sample), personality traits ...

Three-Way vs Two-Way Arrays - Similarities

- Rank := smallest number of rank-one “factors” (“terms” is probably better) for exact additive decomposition (same concept for both 2-way and 3-way)
- Two-way rank-one factor: rank-one MATRIX outer product of 2 vectors (containing all double products)
- Three-way rank-one factor: rank-one 3-WAY ARRAY outer product of 3 vectors (containing all triple products) - same concept

Three-Way vs Two-Way Arrays - Differences

- Two-way $(I \times J)$: row-rank = column-rank = rank $\leq \min(I, J)$;
- Three-way: row-rank $\neq$ column-rank $\neq$ "tube"-rank $\neq$ rank
- Two-way: rank(randn(I,J))=min(I,J) w.p. 1;
- Three-way: rank(randn(2,2,2)) is a RV (2 w.p. 0.3, 3 w.p. 0.7)
- 2-way: rank insensitive to whether or not underlying field is open or closed ($\mathbb{R}$ versus $\mathbb{C}$); 3-way: rank sensitive to $\mathbb{R}$ versus $\mathbb{C}$
- 3-way: Except for loose bounds and special cases [Kruskal; J.M.F. ten Berge], general results for maximal rank and typical rank sorely missing for decomposition over $\mathbb{R}$; theory more developed for decomposition over $\mathbb{C}$ [Burgisser, Clausen, Shokrollahi, *Algebraic complexity theory*, Springer, Berlin, 1997]

Khatri-Rao Product

☞ Column-wise Kronecker Product:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 5 & 10 \\ 15 & 20 \\ 25 & 30 \end{bmatrix}, \quad \mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 5 & 20 \\ 15 & 40 \\ 25 & 60 \\ 15 & 40 \\ 45 & 80 \\ 75 & 120 \end{bmatrix}$$

$$vec(\mathbf{A}\mathbf{D}\mathbf{B}^T) = (\mathbf{B} \odot \mathbf{A})\mathbf{d}(\mathbf{D})$$

$$\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}$$

LRD of Three-Way Arrays: Notation

- Scalar:

$$x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f}, \quad i = 1, \cdots, I, \; j = 1, \cdots, J, \; k = 1, \cdots, K$$

- Slabs:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}^T, \; k = 1, \cdots, K$$

- Matrix:

$$\mathbf{X}^{(KJ \times I)} = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T$$

- Vector:

$$\mathbf{x}^{(KJI)} := vec\left(\mathbf{X}^{(KJ \times I)}\right) = (\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}))\,\mathbf{1}_{F \times 1} = (\mathbf{A} \odot \mathbf{B} \odot \mathbf{C})\,\mathbf{1}_{F \times 1}$$

LRD of N-Way Arrays: Notation

- Scalar:

$$x_{i_1,\cdots,i_N} = \sum_{f=1}^{F} \prod_{n=1}^{N} a_{i_n,f}^{(n)}$$

- Matrix:

$$\mathbf{X}^{(I_1 I_2 \cdots I_{N-1} \times I_N)} = \left(\mathbf{A}^{(N-1)} \odot \mathbf{A}^{(N-2)} \odot \cdots \odot \mathbf{A}^{(1)}\right) \left(\mathbf{A}^{(N)}\right)^T$$

- Vector:

$$\mathbf{x}^{(I_1 \cdots I_N)} := vec\left(\mathbf{X}^{(I_1 I_2 \cdots I_{N-1} \times I_N)}\right) =$$

$$\left(\mathbf{A}^{(N)} \odot \mathbf{A}^{(N-1)} \odot \mathbf{A}^{(N-2)} \odot \cdots \odot \mathbf{A}^{(1)}\right) \mathbf{1}_{F \times 1}$$

Closer look at applications: Data modeling

- CDMA: (i, j, k, f): (Rx antenna, symbol snapshot, chip, user)

$$x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f}, \quad i = 1, \cdots, I, \ j = 1, \cdots, J, \ k = 1, \cdots, K$$

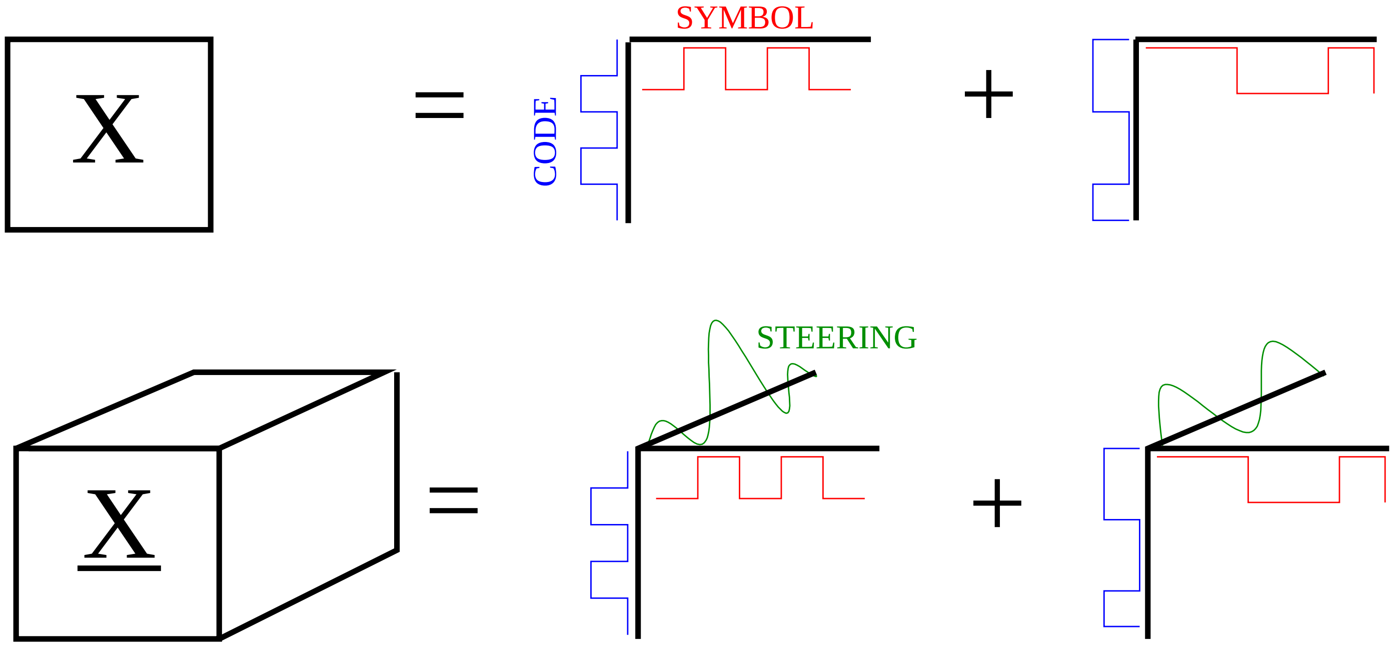
- MI-SAP: $\mathbf{A}$ is response of reference subarray, $\mathbf{B}^T$ is temporal signal matrix (usually denoted **S**), $\mathbf{D}_k(\mathbf{C})$ holds the phase shifts for the k-th displaced but otherwise identical subarray:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}^T, \ k = 1, \cdots, K$$

- Blind signature estimation from covariance data: Symmetric PARAFAC/CANDECOMP (INDSCAL):

$$\mathbf{R}_k = \mathbf{A}\mathbf{D}_k(\mathbf{P})\mathbf{A}^H, \ k = 1, \cdots, K$$

Early Take-Home Point

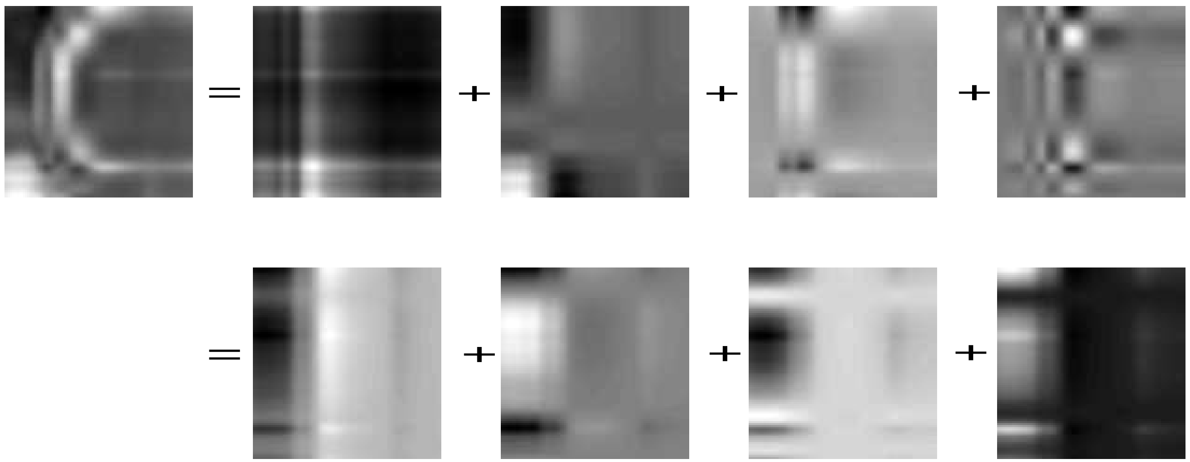


- ☞ Fact 1: Low-rank matrix (2-way array) decomposition not unique for rank > 1
- ☞ Fact 2: Low-rank 3- and higher-way array decomposition (PARAFAC) is unique under certain conditions

LRD of Matrices: Rotational Indeterminacy

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T = \mathbf{a}_1\mathbf{b}_1^T + \cdots + \mathbf{a}_{r_\mathbf{X}}\mathbf{b}_{r_\mathbf{X}}^T$$

$$x_{i,j} = \sum_{k=1}^{r_\mathbf{X}} a_{i,k} b_{j,k}$$



Reverse engineering of soup?

☞ Can only guess recipe

Sample from two or more Cooks!

☞ Same ingredients, different proportions $\hookrightarrow$ recipe!

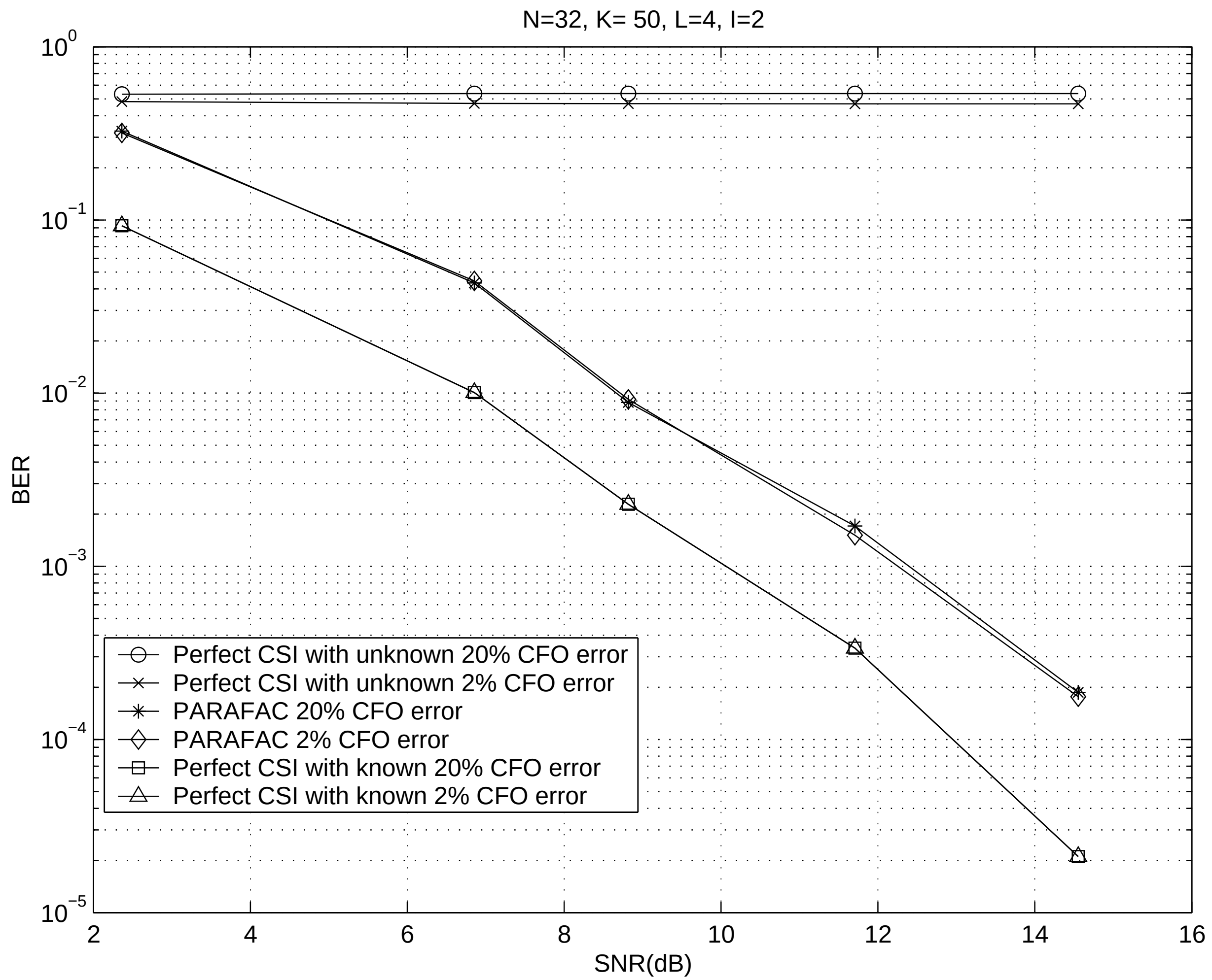
SIMO OFDM / CFO

- Collect K OFDM symbol snapshots

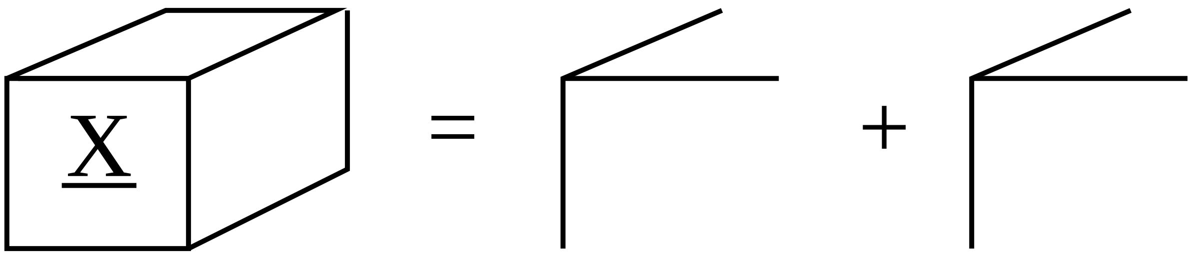
$$\mathbf{Y}_i = \mathbf{P}\mathbf{F}^H\mathbf{H}_i(\mathbf{Q}\mathbf{S})^T + \mathbf{W}_i =: \mathbf{A}\mathbf{D}_i\mathbf{B}^T + \mathbf{W}_i, i = 1, \cdots, I$$

- PARAFAC model (w/ special structure) $\Longrightarrow$ blindly identifiable [Jiang & Sidiropoulos, ’02]
- Deterministic approach, works with small sample sizes (channel coherence), relaxed ID conditions, performance within 2 dB from non-blind MMSE clairvoyant Rx

SIMO-OFDM / CFO - results



Uniqueness



☞ [Kruskal, 1977], $N = 3, \mathbb{R}$: $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2F + 2$
k-rank= maximum r such that *every* r columns are linearly independent ($\leq$ rank)

☞ [Sidiropoulos *et al*, IEEE TSP, 2000]: $N = 3$, $\mathbb{C}$

☞ [Sidiropoulos & Bro, J. Chem., 2000]: any N, $\mathbb{C}$:
$\sum_{n=1}^{N} k - ranks \geq 2F + (N - 1)$

Key-I

☞ Kruskal's Permutation Lemma [Kruskal, 1977]: Consider $\mathbf{A}$ $(I \times F)$ having no zero column, and $\bar{\mathbf{A}}$ $(I \times \bar{F})$. Let $w(\cdot)$ be the *weight* (# of nonzero elements) of its argument. If for any vector $\mathbf{x}$ such that

$$w(\mathbf{x}^H \bar{\mathbf{A}}) \leq \bar{F} - r_{\bar{\mathbf{A}}} + 1,$$

we have

$$w(\mathbf{x}^H \mathbf{A}) \leq w(\mathbf{x}^H \bar{\mathbf{A}}),$$

then $F \leq \bar{F}$; if also $F \geq \bar{F}$, then $F = \bar{F}$, and there exist a permutation matrix $\mathbf{P}$ and a non-singular diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \bar{\mathbf{A}}\mathbf{P}\mathbf{D}$.

☞ Easy to show for a pair of square nonsingular matrices (use rows of pinv); but the result is very deep and difficult for fat matrices - see [Jiang & Sidiropoulos, TSP:04]

Key-II

☞ Property: [Sidiropoulos & Liu, 1999; Sidiropoulos & Bro, 2000]

If $k_{\mathbf{A}} \geq 1$ and $k_{\mathbf{B}} \geq 1$, then it holds that

$$k_{\mathbf{B}\odot\mathbf{A}} \geq \min(k_{\mathbf{A}} + k_{\mathbf{B}} - 1, F),$$

whereas if $k_{\mathbf{A}} = 0$ or $k_{\mathbf{B}} = 0$

$$k_{\mathbf{B}\odot\mathbf{A}} = 0$$

Stepping stone

☞ A proof of Kruskal's result is beyond our scope. The following is more palatable & conveys flavor (see SAM2004 paper for compact proof):

Theorem: Given $\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C})$, with $\mathbf{A}: I \times F$, $\mathbf{B}: J \times F$, and $\mathbf{C}: K \times F$, it is *necessary* for uniqueness of **A**, **B**, **C** that $\min(r_{\mathbf{A}\odot\mathbf{B}}, r_{\mathbf{C}\odot\mathbf{A}}, r_{\mathbf{B}\odot\mathbf{C}}) = F$. If $F > 1$, then it is also necessary that $\min(k_{\mathbf{A}}, k_{\mathbf{B}}, k_{\mathbf{C}}) \geq 2$.

If, in addition $r_{\mathbf{C}} = F$, and $k_{\mathbf{A}} + k_{\mathbf{B}} \geq F + 2$, then **A**, **B**, and **C** are unique up to permutation and scaling of columns, meaning that if $\underline{\mathbf{X}} = (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$, for some $\bar{\mathbf{A}}: I \times F$, $\bar{\mathbf{B}}: J \times F$, and $\bar{\mathbf{C}}: K \times F$, then there exists a permutation matrix Π and diagonal scaling matrices $\Lambda_1, \Lambda_2, \Lambda_3$ such that

$$\bar{\mathbf{A}} = \mathbf{A}\Pi\Lambda_1,\ \bar{\mathbf{B}} = \mathbf{B}\Pi\Lambda_2,\ \bar{\mathbf{C}} = \mathbf{C}\Pi\Lambda_3,\ \Lambda_1\Lambda_2\Lambda_3 = \mathbf{I}.$$

Is Kruskal's Condition Necessary?

- Long-held conjecture (Kruskal'89): Yes
- ten Berge & Sidiropoulos, *Psychometrika*, 2002: Yes for $F \in \{2,3\}$, no for $F > 3$
- Jiang & Sidiropoulos '03: new insights that explain part of the puzzle: E.g., for $r_{\mathbf{C}} = F$, the following condition has been proven to be *necessary and sufficient*:

  No linear combination of two or more columns of $\mathbf{A} \odot \mathbf{B}$ can be written as KRP of two vectors

Why Care?

☞ 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 $\longrightarrow$ monotone convergence, usually to global minimum (uniqueness), close to CRB for $F << IJK$

Algorithms

- ALS is based on matrix view:

$$\mathbf{X}^{(KJ\times I)} = (\mathbf{B}\odot\mathbf{C})\mathbf{A}^T$$

- Given interim estimates of **B**, **C**, solve for conditional LS update of **A**:

$$\mathbf{A}_{CLS} = \left((\mathbf{B}\odot\mathbf{C})^{\dagger}\mathbf{X}^{(KJ\times I)}\right)^T$$

- Similarly for the CLS updates of **B**, **C** (symmetry); repeat in a circular fashion until convergence in fit (guaranteed)

Algorithms

- ALS initialization matters, not crucial for heavily over-determined problems
- Alt: rank-1 updates possible [Kroonenberg], but inferior
- COMFAC (Tucker3 compression), G-N, Levenberg, ATLD, DTLD, ESPRIT-like,...
- G-N converges faster than ALS, but it may fail
- In general, no ”algebraic” solution like SVD
- Possible if e.g., a subset of columns in A is known [Jiang & Sidiropoulos, JASP/SMART 2003]; or under very restrictive rank conditions

Robust Regression Algorithms

- Laplacian, Cauchy-distributed errors, outliers
- Least Absolute Error (LAE) criterion: optimal (ML) for Laplacian, robust across α-stable
- Similar to ALS, each conditional matrix update can be shown equivalent to a LP problem $\longrightarrow$ alternating LP [Vorobyov, Rong, Sidiropoulos, Gershman, 2003]
- Alternatively, very simple element-wise updating using *weighted median filtering* [Vorobyov, Rong, Sidiropoulos, Gershman, 2003]
- Robust algorithms perform well for Laplacian, Cauchy, and not far from optimal in the Gaussian case

CRBs for the PARAFAC model

- Dependent on how scale-permutation ambiguity is resolved
- Real i.i.d. Gaussian, 3-way, Complex circularly symmetric i.i.d. Gaussian, 3-way & 4-way [Liu & Sidiropoulos, TSP 2001]
- Compact expressions for complex 3-way case & asymptotic CRB when one mode length goes to infinity [Jiang & Sidiropoulos, JASP/SMART:04]
- Laplacian, Cauchy [Vorobyov, Rong, Sidiropoulos, Gershman, TSP:04] - scaled versions of the Gaussian CRB; scaling parameter only dependent on noise pdf

Performance

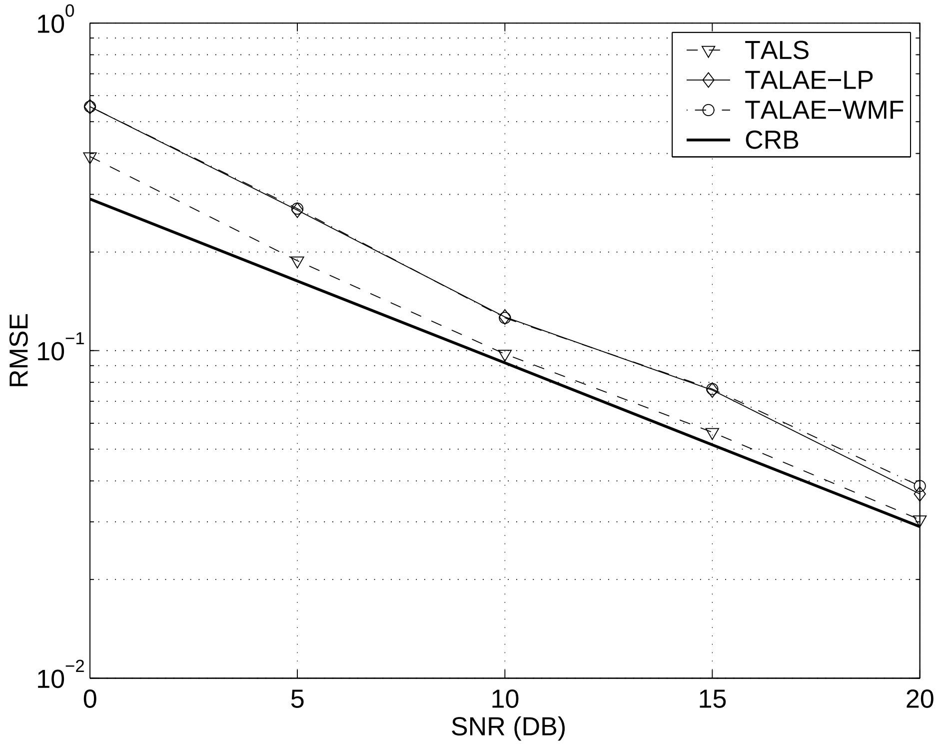


Figure 1: RMSEs versus SNR: Gaussian noise, $8 \times 8 \times 20$, $F = 2$

Performance

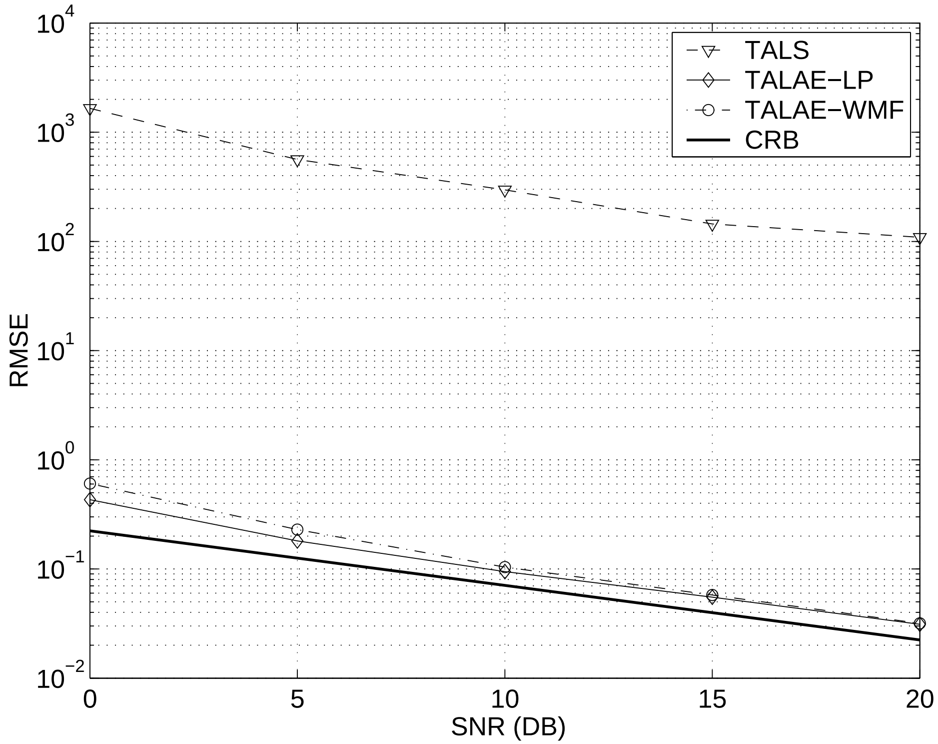


Figure 2: RMSEs versus SNR: Cauchy noise, $8 \times 8 \times 20$, $F = 2$

Performance

- ☞ ALS works well in AWGN because it is ML-driven, and with 3-way data it is easy to get to the large-samples regime: e.g., $10 \times 10 \times 10 = 1000$
- ☞ Performance is worse (and further from the CRB) when operating close to the identifiability boundary; but ALS *works* under model identifiability conditions only, which means that at high SNR the parameter estimates are still accurate
- ☞ Main shortcoming of ALS and related algorithms is the high computational cost
- ☞ For difficult datasets, so-called *swamps* are possible: progress towards convergence becomes extremely slow
- ☞ Still workhorse, after all these years ...

Learn more - tutorials, bibliography, papers, software,...

- Group homepage (Nikos Sidiropoulos):
  `www.telecom.tuc.gr/~nikos` and
  `www.ece.umn.edu/users/nikos`
- 3-way group at KVL/DK (Rasmus Bro):
  `http://www.models.kvl.dk/users/rasmus/` and
  `http://www.models.kvl.dk/courses/`
- 3-Mode Company (Peter Kroonenburg):
  `http://www.leidenuniv.nl/fsw/three-mode/3modecy.htm`
- Hard-to-find original papers (Richard Harshman):
  `http://publish.uwo.ca/~harshman/`
- 3-way workshop: TRICAP 2000: Faaborg, DK; 2003, Kentucky, USA; 2006, Chania-Crete Greece.

What lies ahead & wrap-up

- Take home point: $(N > 3)$-way arrays *are* different; low-rank models unique, have many applications
- Major challenges: Uniqueness: i) Easy to check necessary & sufficient conditions; ii) Higher-way models; iii) Uniqueness under application-specific constraints (e.g., Toeplitz); iv) symmetric & super-symmetric models (INDSCAL, JD, HOS)
- Major challenges: Algorithms: Faster at small performance loss; incorporation of application-specific constraints
- New exciting applications: Yours!

Preaching the Gospel of 3-Way Analysis

☞ Thank you!