Coupled Matrix/Tensor Decompositions:

An Introduction

Laurent Sorber    Mikael Sørensen    Marc Van Barel
Lieven De Lathauwer

KU Leuven
Belgium

Lieven.DeLathauwer@kuleuven-kulak.be
Canonical Polyadic Decomposition

Rank: minimal number of rank-1 terms \([Hitchcock, 1927]\)

Canonical Polyadic Decomposition (CPD): decomposition in minimal number of rank-1 terms \([Harshman '70], [Carroll and Chang '70]\)

\[ T = a_1 b_1 c_1 + \cdots + a_R b_R c_R \]

- Unique under mild conditions on number of terms and differences between terms
- Orthogonality (triangularity, \ldots) not required (but may be imposed)
Factor Analysis and Blind Source Separation

- Decompose a data matrix in rank-1 terms that can be interpreted
  E.g. statistics, telecommunication, biomedical applications, chemometrics, data analysis, ...

\[ A = F \cdot G^T \]

\[ \begin{bmatrix} A \\ f_1 \\ f_2 \\ \vdots \\ f_R \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_R \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_R \end{bmatrix} \]

- \( F \): mixing matrix
- \( G \): source signals

- Decompose a data matrix in rank-1 terms that can be interpreted
What about SVD?

- SVD is unique
- ... thanks to orthogonality constraints

\[ A = U \cdot S \cdot V^T = \sum_{r=1}^{R} s_{rr} u_r v_r^T \]

\( U, V \) orthogonal, \( S \) diagonal
- Whether these constraints make sense, depends on the application
- SVD is great for dimensionality reduction
  best rank-\( R \) approximation \( \leftarrow \) truncated SVD
Uniqueness: \( C \) has full column rank

\[
\mathcal{T} = \sum_{r=1}^{R} a_r \odot b_r \odot c_r \in \mathbb{C}^{I \times J \times K}
\]
e.g. \( C \)-mode is sample mode

Khatri-Rao product second compound matrices:

\[
U = C_2(A) \odot C_2(B) \in \mathbb{C}^{\frac{I(I-1)}{2} \times \frac{J(J-1)}{2} \times \frac{R(R-1)}{2}}
\]

\[
u_{i_1i_2j_1j_2r_1r_2} = \begin{vmatrix}
a_{i_1r_1} & a_{i_2r_1} \\
a_{i_1r_2} & a_{i_2r_2}
\end{vmatrix} \cdot \begin{vmatrix}
b_{j_1r_1} & b_{j_2r_1} \\
b_{j_1r_2} & b_{j_2r_2}
\end{vmatrix}
\]

\[1 \leq i_1 < i_2 \leq I \quad 1 \leq j_1 < j_2 \leq J \quad 1 \leq r_1 < r_2 \leq R\]

Theorem: if \( U \) and \( C \) have full column rank, then CPD is unique

(proof is constructive)

[\text{Jiang and Sidiropoulos, '04}, \text{DL '06}]

15
Uniqueness: $C$ has full column rank (2)

Theorem: if $U \in \mathbb{C}^{\frac{I(I-1)J(J-1)}{2} \times \frac{R(R-1)}{2}}$ and $C \in \mathbb{C}^{K \times R}$ have full column rank, then CPD is unique.

Generic: CPD is unique for $R$ bounded by $I, J, K$ as in

$$\frac{I(I-1)J(J-1)}{2} \geq \frac{R(R-1)}{2} \quad \text{and} \quad K \geq R$$

Approximately: \(\frac{IJ}{\sqrt{2}} \geq R \quad K \geq R\)

Compare to Kruskal:

$$\min(I, R) + \min(J, R) \geq R + 2 \quad \text{and} \quad K \geq R$$
Recent results

Unifying theory

Constructive proof

Algorithm for Kruskal’s condition (and beyond)

[Domanov, DL, ’12], [Domanov, DL, ’13]
Canonical Polyadic Decomposition

Rank: minimal number of rank-1 terms \[ \text{[Hitchcock, 1927]} \]

Canonical Polyadic Decomposition (CPD): decomposition in minimal number of rank-1 terms \[ \text{[Harshman ’70], [Carroll and Chang ’70]} \]

- Unique under mild conditions on number of terms and differences between terms
- Orthogonality (triangularity, . . .) not required (but may be imposed)
Decomposition in rank-\((L, L, 1)\) terms

\[ T = A_1 B_1 c_1 + \cdots + A_R B_R c_R \]

Unique under mild conditions

[DL ’08]
Decomposition in rank-$(R_1, R_2, R_3)$ terms

$T = A_1 B_1 C_1 + \cdots + A_R B_R C_R$

Unique under mild conditions

Rank-1 term $\sim$ data atom
Block term $\sim$ data molecule

[DL ’08]
Constraints

Examples: orthogonality [Sørensen and DL '12]
          nonnegativity [Cichocki et al. '09]
          Vandermonde [Sørensen and DL '12]
          independence [De Vos et al. '12]
          ...

Not needed for uniqueness in tensor case

Pro: relaxed uniqueness conditions
     easier interpretation
     no degeneracy (NN, orthogonality)
     higher accuracy

Depending on type of constraints, lower or higher computational cost
Tensorlab — a MATLAB toolbox for tensor decompositions
esat.kuleuven.be/sista/tensorlab

- Elementary operations on tensors
  Multicore-aware and profiler tuned

- Tensor decompositions with structure and/or symmetry
  CPD, LMLRA, MLSVD, block term decompositions

- Global minimization of bivariate polynomials
  Exact line and plane search for tensor optimization

- Cumulants, tensor visualization, estimating a tensor’s rank or multilinear rank, ...
Coupled matrix/tensor decompositions

One or more matrices and/or one or more tensors

Symmetric and/or nonsymmetric

One or more factors shared (or parts of factors, or generators)

Constraints (orthogonal, nonnegative, exponential, constant modulus, polynomial, rational, Toeplitz, Hankel, . . . )

Large, possibly incomplete data

Multi-view data / data fusion

E.g. coupled EEG-fMRI; person-location-activity and person-person and location-location . . . ; . . .
Uniqueness: $\mathbf{C}$ has full column rank

Coupled CPD: $\mathbf{T}^{(n)} = \sum_{r=1}^{R} \mathbf{a}_r^{(n)} \odot \mathbf{b}_r^{(n)} \odot \mathbf{c}_r \in \mathbb{C}^{I_n \times J_n \times K}$

Khatri-Rao product second compound matrices:

$$\mathbf{U}^{(n)} = C_2(\mathbf{A}^{(n)}) \odot C_2(\mathbf{B}^{(n)}) \in \mathbb{C}^{\frac{I_n(I_n-1)}{2} \frac{J_n(J_n-1)}{2} \frac{R(R-1)}{2}}$$

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}^{(1)} \\ \mathbf{U}^{(2)} \\ \vdots \\ \mathbf{U}^{(N)} \end{pmatrix}$$

Theorem: if $\mathbf{U}$ and $\mathbf{C}$ have full column rank, then coupled CPD is unique

(proof is constructive)
Increase spatial diversity: Widely Separated Antenna Arrays

\[ y^{(1)}_{1jk} = \sum_{r=1}^{R} a^{(r,1)}_{i} h^{(r,1)}_{j} s^{(r)}_{k} \]

\[ y^{(N)}_{1jk} = \sum_{r=1}^{R} a^{(r,N)}_{i} h^{(r,N)}_{j} s^{(r)}_{k} \]

(A biotensor example is multimodal data fusion: EEG × fMRI × MEG × ...)
Signal Separation from a Coupled Tensorial Perspective

\[
\mathbf{y}^{(1)} = \mathbf{s}^{(1)} \mathbf{a}^{(1,1)} + \cdots + \mathbf{s}^{(R)} \mathbf{a}^{(R,1)}
\]

\[
\mathbf{y}^{(N)} = \mathbf{s}^{(1)} \mathbf{a}^{(1,N)} + \cdots + \mathbf{s}^{(R)} \mathbf{a}^{(R,N)}
\]
Extension to Incoherent Multipath with Small Delay Spread

\[ y_{ijk}^{(N)} = \sum_{r=1}^{R} \sum_{p=1}^{P_{r,N}} a_{i}^{(p,r,N)} h_{j}^{(p,r,N)} s_{k}^{r} \]

\[ y_{ijk}^{(1)} = \sum_{r=1}^{R} \sum_{p=1}^{P_{1,r,1}} a_{i}^{(p,r,1)} h_{j}^{(p,r,1)} s_{k}^{r} \]

\[ y_{ijk} = \sum_{r=1}^{R} \sum_{p=1}^{P_{r}} a_{i}^{(p,r)} h_{j}^{(p,r)} s_{k}^{r} \]

"Coupled Tensor Decompositions"
Signal Separation from a Coupled Tensorial Perspective

\[ \mathbf{y}^{(1)} = \mathbf{s}^{(1)} \mathbf{H}^{(1,1)} \mathbf{A}^{(1,1)^T} + \cdots + \mathbf{s}^{(R)} \mathbf{H}^{(R,1)} \mathbf{A}^{(R,1)^T} \]

\[ \mathbf{y}^{(N)} = \mathbf{s}^{(1)} \mathbf{H}^{(1,N)} \mathbf{A}^{(1,N)^T} + \cdots + \mathbf{s}^{(R)} \mathbf{H}^{(R,N)} \mathbf{A}^{(R,N)^T} \]
Tensorlab v2.0

Major upgrade which brings:

- Full support for sparse and incomplete tensors
- Major improvements in computational and memory efficiency
- Structured data fusion

*Structured*: choose from a large library of constraints to impose on factors (nonnegative, orthogonal, Toeplitz, ...)

*Data fusion*: jointly factorize multiple data sets
Example 4: GPS data set

Five coupled data sets: user-location-activity, user-user, location-feature, activity-activity and user-location

Challenge: predict user participation in activities

Solution with SDF: compute coupled tensor factorization

\[
\begin{align*}
\text{minimize} & \quad \frac{\omega_1}{2} \left\| \mathcal{M}^{(1)}(U, L, A) - \mathcal{T}^{(1)} \right\|_{\mathcal{W}^{(1)}}^2 \\
& + \frac{\omega_2}{2} \left\| \mathcal{M}^{(2)}(U, U, \lambda) - \mathcal{T}^{(2)} \right\|_F^2 \\
& + \frac{\omega_3}{2} \left\| \mathcal{M}^{(3)}(L, F) - \mathcal{T}^{(3)} \right\|_F^2 \\
& + \frac{\omega_4}{2} \left\| \mathcal{M}^{(4)}(A, A, \mu) - \mathcal{T}^{(4)} \right\|_F^2 \\
& + \frac{\omega_5}{2} \left\| \mathcal{M}^{(5)}(U, L, \nu) - \mathcal{T}^{(5)} \right\|_F^2 \\
& + \frac{\omega_6}{2} \left( \left\| U \right\|_F^2 + \left\| L \right\|_F^2 + \left\| A \right\|_F^2 + \left\| F \right\|_F^2 + \left\| \lambda \right\|_F^2 + \left\| \mu \right\|_F^2 + \left\| \nu \right\|_F^2 \right)
\end{align*}
\]
Example 4: 80% missing entries in user-location-activity tensor
Example 4: 50 users missing in user-location-activity tensor
9th-order \((100 \times 100 \times \ldots \times 100)\) data set

10-component thermodynamic phase diagram

incomplete tensor: \(130,000\) known samples (out of \(10^{18}\))

sum of structured rank-1 terms
Example 3: InsPyro materials data set

**Data set:** an incomplete tensor in which each dimension represents the concentration of a metal in an alloy and the entries are the alloy’s melting temperature

**Challenge:** predict melting temperatures of different alloys

**Solution with SDF:** use structured CPD where each factor vector $u_{(n)}^{r}$ is a sum of RBF kernels

$$u_{b,r}^{(n)} = \sum_{i=1}^{8} a \exp \left( -\frac{(t - b)^2}{2c^2} \right)$$

where $a$, $b$, and $c$ are the free parameters in $u_{r}^{(n)}$
Example 3: InsPyro materials data set
Example 3: InsPyro materials data set
Example 3: InsPyro materials data set
Example 3: InsPyro materials data set
Conclusion

- Coupled matrix/tensor decompositions = important new concept
- Many applications
- Gold mine for research topics:
  - algebra
  - (numerical) linear algebra
  - randomized NLA
  - large-scale numerical optimization
  - ...