Optimization for Fitting CANDECOMP/PARAFAC Models

Evrim Acar, Daniel M. Dunlavy, and Tamara G. Kolda*
Sandia National Laboratories

LDRD
LABORATORY DIRECTED RESEARCH & DEVELOPMENT

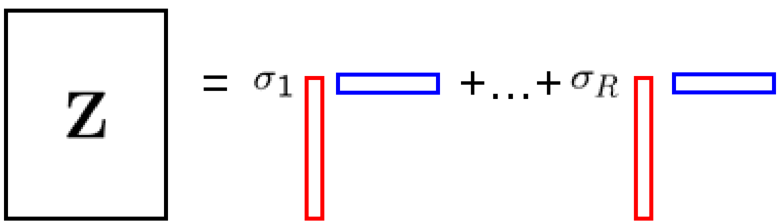
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

* = Speaker

Sandia National Laboratories

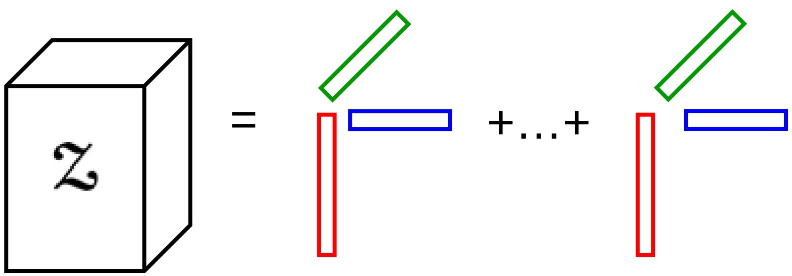
CANDECOMP/PARAFAC Decomposition (CPD)

Singular Value Decomposition (SVD) expresses a matrix as the sum of component rank-1 matrices.



$$\mathbf{Z} = \sum_{r=1}^{R} \sigma_r \, \mathbf{u}_r \circ \mathbf{v}_r$$

CANDECOMP/PARAFAC Decomposition (CPD) expresses a tensor as the sum of component rank-1 tensors.



$$\mathcal{Z} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

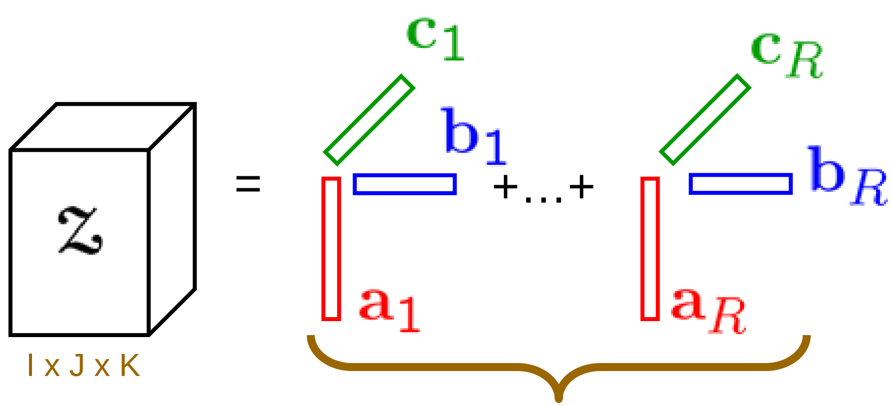
$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

CPD is a Nonlinear Optimization Problem

Given tensor $\mathcal{Z}$ and R (# of components), find matrices **A**, **B**, **C** that solve the following problem:

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|$$



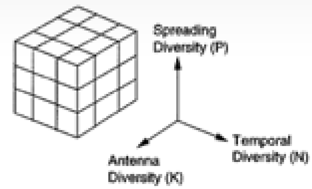
R(I+J+K) variables

R component rank-1 tensors

Sandia National Laboratories

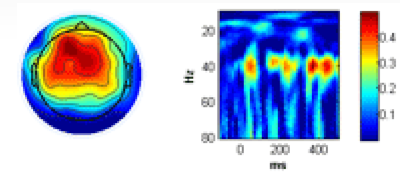
Applications of CPD

- Modeling fluorescence excitation-emission data
- Signal processing
- Brain imaging (e.g., fMRI) data
- Web graph plus anchor term analysis
- Image compression and classification
- Texture analysis
- Epilespy seizure detection
- Text analysis
- Approximating Newton potentials, stochastic PDEs, etc.

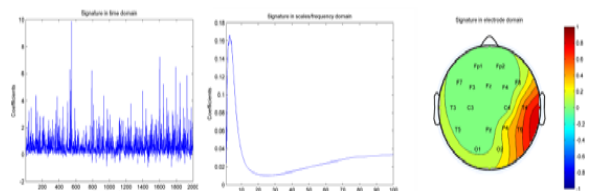


Sidiropoulos, Giannakis, and Bro, *IEEE Trans. Signal Processing*, 2000.

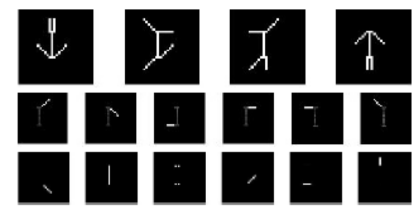
Furukawa, Kawasaki, Ikeuchi, and Sakauchi, *EGRW '02*



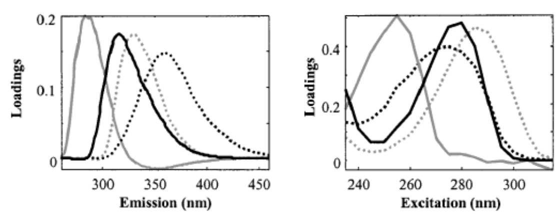
ERPWAVELAB by Morten Mørup.



Acar, Bingol, Bingol, Bro and Yener, *Bioinformatics*, 2007.



Hazan, Polak, and Shashua, *ICCV 2005*.



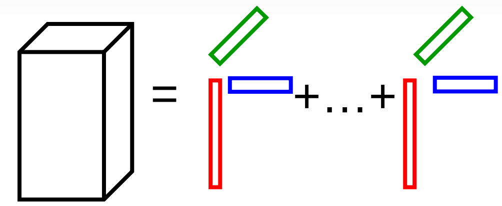
Andersen and Bro, *J. Chemometrics*, 2003.

$$\mathcal{L}(x,t,\omega;u) = f(x,t,\omega) \quad (x,t) \in \mathcal{D} \times [0,T]$$
$$\mathcal{B}(x,t,\omega;u) = g(x,t) \quad (x,t) \in \partial\mathcal{D} \times [0,T]$$
$$\mathcal{I}(x,0,\omega;u) = h(x,\omega) \quad x \in \mathcal{D},$$

Doostan, Iaccarino, and Etemadi, Stanford University TR, 2007

CPALS – Solves for One Block of Variables at a Time

Old Way



Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|$$

Alternating Algorithm

For k = 1,…

$$\min_{\mathbf{A}} \| \mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|$$

$$\min_{\mathbf{B}} \| \mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|$$

$$\min_{\mathbf{C}} \| \mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|$$

End

This can be converted to a matrix least squares problem:

$$\min_{\mathbf{A}} \left\| \mathbf{Z}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}} \right\|$$

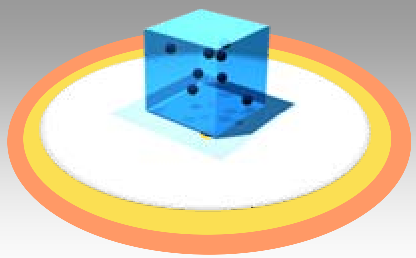
$$\mathbf{A} = \mathbf{Z}_{(1)} \left((\mathbf{C} \odot \mathbf{B})^{\mathsf{T}} \right)^{\dagger}$$

I x JK JK x R

$$\mathbf{A} = \mathbf{Z}_{(1)} (\mathbf{C} \odot \mathbf{B}) (\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}$$

I x R I x JK JK x R R x R matrix

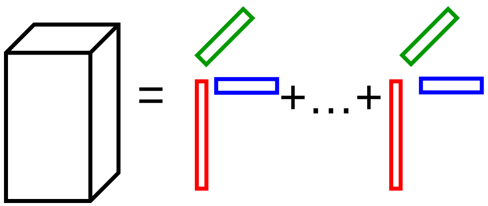
ALS procedure dates back to early work by Harshman (1970) and Carroll and Chang (1970)



CPOPT - Instead, Solve for All Variables Simultaneously

Objective Function

$$f(\mathbf{A},\mathbf{B},\mathbf{C}) = \min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|$$



Gradient

$$\frac{\partial f}{\partial \mathbf{A}} = -\mathbf{Z}_{(1)}(\mathbf{C} \odot \mathbf{B}) + \mathbf{A}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})$$

$$\frac{\partial f}{\partial \mathbf{B}} = -\mathbf{Z}_{(2)}(\mathbf{C} \odot \mathbf{A}) + \mathbf{B}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{A}^{\mathsf{T}}\mathbf{A})$$

$$\frac{\partial f}{\partial \mathbf{C}} = -\mathbf{Z}_{(3)}(\mathbf{B} \odot \mathbf{A}) + \mathbf{C}(\mathbf{B}^{\mathsf{T}}\mathbf{B} * \mathbf{A}^{\mathsf{T}}\mathbf{A})$$

NEW Way

Our implementation uses **nonlinear CG** with line search for optimization.

CPNLS – Tackle CPD as a nonlinear equation

CPNLS: Apply nonlinear least squares solver to the following equations:

$$F(\mathbf{x}) = \text{vec}\left(\mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\right)$$

⬇

$$F : \mathbb{R}^{(I+J+K)R} \mapsto \mathbb{R}^{IJK}$$

⬇

Jacobian is of size $(I+J+K)R$ £ IJK, which can be quite large.

This approach has been proposed by **Paatero**, *Chemometrics and Intelligent Laboratory Systems*, 1997 and also **Tomasi and Bro**, *Chemometrics and Intelligent Laboratory Systems*, 2005.

Sandia National Laboratories

Optimization-Based Approach is Fast and Accurate

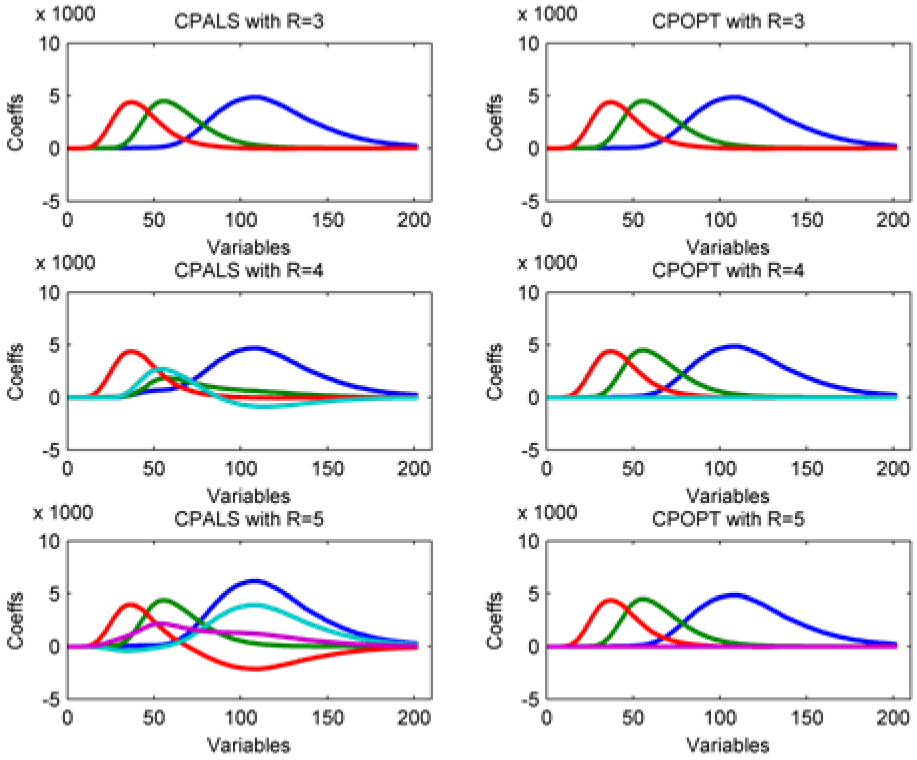
Generated 360 dense test problems (with ranks 3 and 5) and factorized with R as the correct number of components and one more than that. Total of 720 tests for each entry below.

	Time (sec)		
Size	CPALS	CPNLS	CPOPT
$20 \times 20 \times 20$	0.5 ± 1.0	1.0 ± 1.1	0.3 ± 0.2
$50 \times 50 \times 50$	0.3 ± 0.3	16.0 ± 17.7	0.7 ± 0.5
$100 \times 100 \times 100$	1.7 ± 1.1	153.2 ± 142.3	5.6 ± 3.6
$250 \times 250 \times 250$	26.6 ± 9.1	—	83.5 ± 35.2

	Accuracy (%)		
Size	CPALS	CPNLS	CPOPT
$20 \times 20 \times 20$	78.8	99.7	99.9
$50 \times 50 \times 50$	65.7	99.9	100.0
$100 \times 100 \times 100$	63.5	99.9	100.0
$250 \times 250 \times 250$	62.2	—	100.0

Further, CPOPT is scalable (see MS45 – Acar et al.)

Nonlinear optimization is an important tool for tensor decompositions



Comparison of ALS and OPT when the rank is higher than is physically meaningful

- CPD is a nonlinear optimization problem
  - OPT = compute gradients and apply nonlinear CG
  - OPT is more scalable than NLS
  - OPT is more accurate than ALS
- Future work
  - Choice of starting point
  - Regularization
  - Constraints (nonnegativity, sparsity)
  - Selecting number of components
  - Exploiting symmetry

Sandia National Laboratories

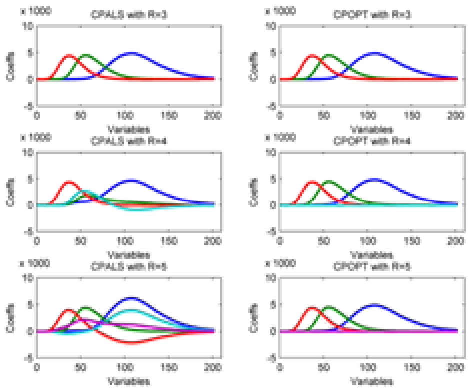
References & Contact Info

- **OPT:** Acar, Kolda and Dunlavy. **An Optimization Approach for Fitting Canonical Tensor Decompositions**, Technical Report SAND2009-0857, Feb 2009
- **Survey:** Kolda and Bader, **Tensor Decompositions and Applications**, *SIAM Review*, Sep 2009 (to appear)
- **Tensor Toolbox:** Bader and Kolda, **Efficient MATLAB computations with sparse and factored tensors**. SISC 30(1):205-231, 2007

All papers available at: http://csmr.ca.sandia.gov/~tgkolda/
(or just google “Tamara Kolda”)

Related Talk: MS45 Data Mining using Tensors (Symphony II)

3:30pm – **Link Prediction on Evolving Data using Tensor Factorization** by Evrim Acar, Danny Dunlavy, and Tamara Kolda.



Sandia National Laboratories