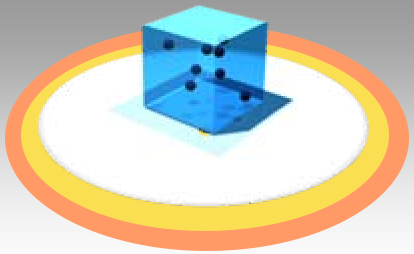


Fitting a Tensor Decomposition is a Nonlinear Optimization Problem

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

Sandia National Laboratories

LDRD
LABORATORY DIRECTED RESEARCH & DEVELOPMENT

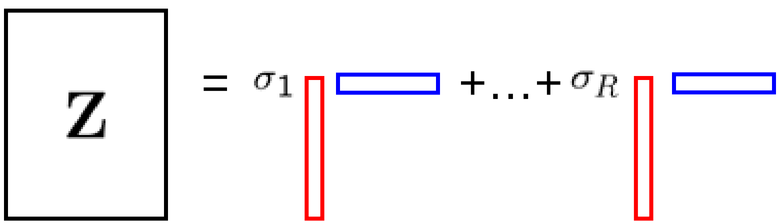
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

* = Speaker

Sandia National Laboratories

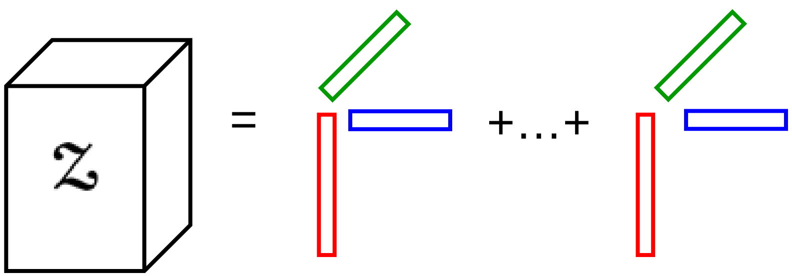
CANDECOMP/PARAFAC Decomposition (CPD)

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



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

CANDECOMP/PARAFAC (CP) expresses a tensor as the sum of rank-1 factors.



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

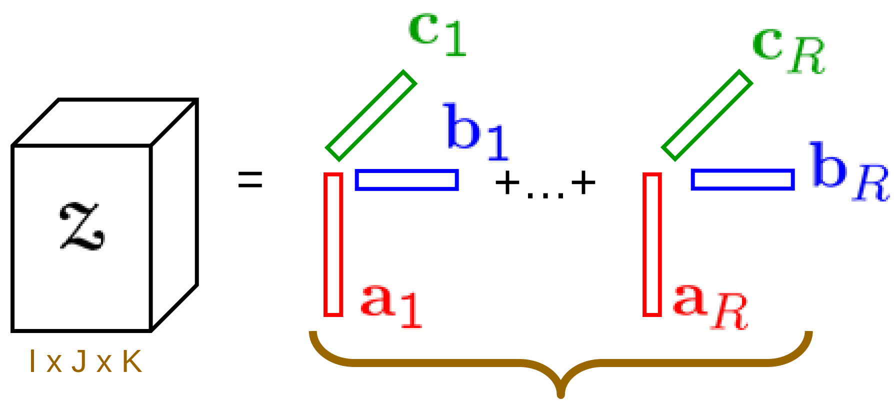
$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

CPD is a Nonlinear Optimization Problem

Given R (# of components), find A, B, C that solve the following problem:

Optimization Problem

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



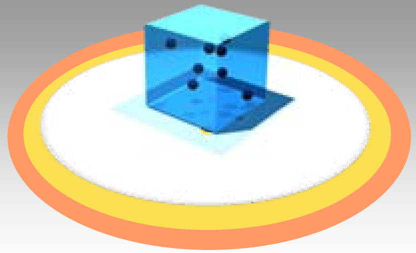
R rank-1 factors

R(I+J+K) variables

CONCLUSION:
We need to bring modern optimization methods to bear on tensor decomposition problems.

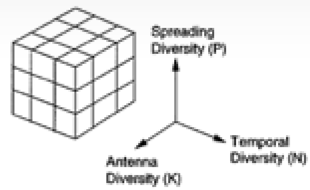
AIM Workshop on Computational Optimization for Tensor Decompositions,
Palo Alto, CA, March 29 - April 2, 2010.

Sandia National Laboratories



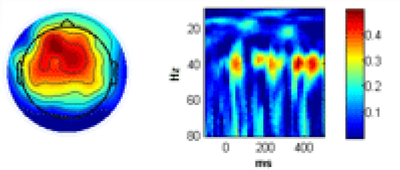
Applications of CPD

- 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.

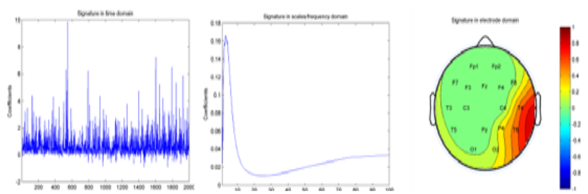


Sidiropoulos, Giannakis, and Bro, *IEEE Trans. Signal Processing*, 2000.

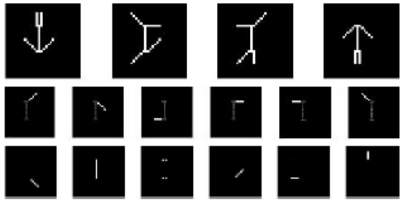
Furukawa, Kawasaki, Ikeuchi, and Sakauchi, *EGRW '02*



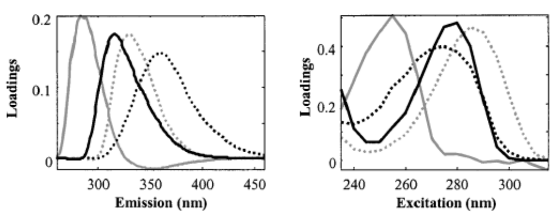
ERPWAVELAB by Morten Mørup.



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



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



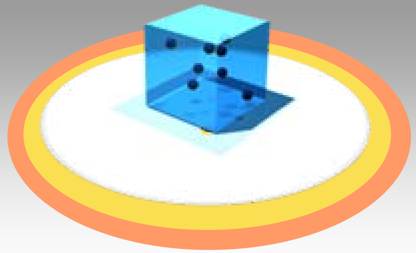
Andersen and Bro, *J. Chemometrics*, 2003.

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

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

Goals for Computing CPD

- Speed – Which method is fastest?
- Accuracy – Did we get the right answer?
- Scalability – Will the method scale to large problems? What about large and sparse?



Mathematical Background

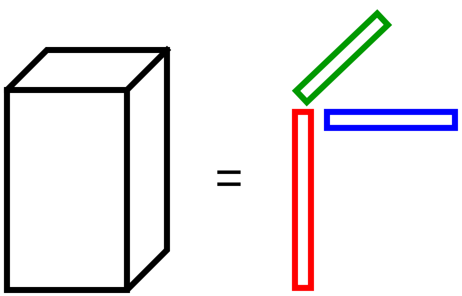
Tensor Order

The number of dimensions, modes, or ways in a tensor.

Vector Outer Product

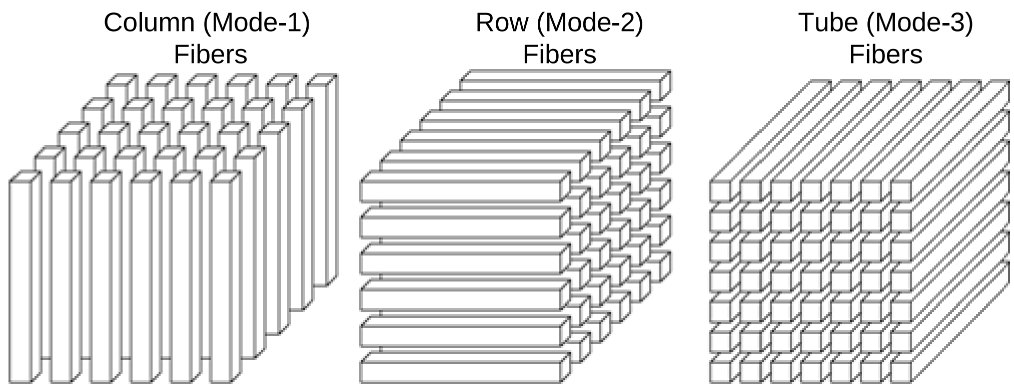
$$\mathcal{Z} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$$

$$z_{ijk} = a_i b_j c_k$$



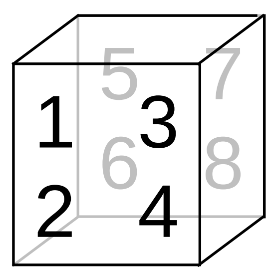
Rank-1 Tensor

Tensor Fibers (Higher-Order Analogue of Rows and Columns)



Unfolding or Matricization

Aligning the mode-n fibers as the columns of a matrix.

$\mathcal{X} =$ 

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Sandia National Laboratories

CPALS – Solves for One Block of Variables at a Time

OLD WAY

Optimization Problem

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

Alternating Algorithm

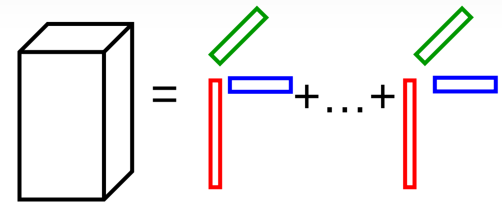
For k = 1,…

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

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

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

End



This can be converted to a matrix least squares problem:

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

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

I x JK   JK x R

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

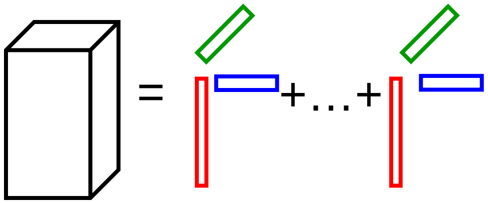
I x R   I x JK   JK x R   R x R matrix

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

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

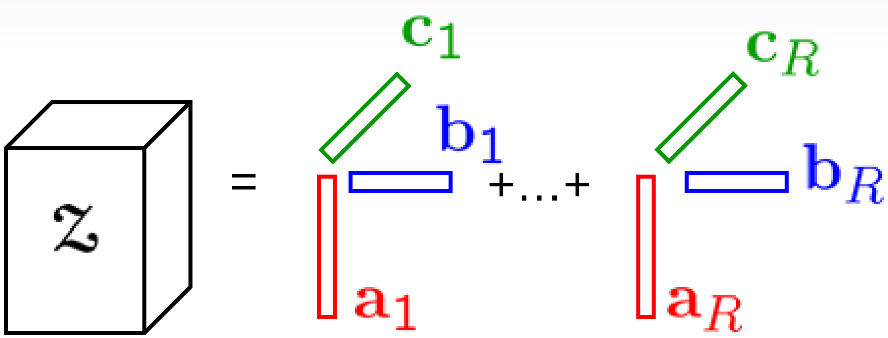
$$\frac{\partial f}{\partial \mathbf{B}} = -\mathbf{Z}_{(2)}(\mathbf{C} \odot \mathbf{A}) + \mathbf{B}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{A}^{\mathsf{T}}\mathbf{A})$$

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

NEW WAY

Sandia National Laboratories

Indeterminacies of CP



- CP has two fundamental indeterminacies
  - **Permutation** – The factors can be reordered
    - Swap $\mathbf{a}_1$, $\mathbf{b}_1$, $\mathbf{c}_1$ with $\mathbf{a}_3$, $\mathbf{b}_3$, $\mathbf{c}_3$
  - **Scaling** – The vectors comprising a single rank-one factor can be scaled
    - Replace $\mathbf{a}_1$ and $\mathbf{b}_1$ with 2 $\mathbf{a}_1$ and ½ $\mathbf{b}_1$

Does this matter? We don't think so but may be an open question…

This leads to a continuous space of equivalent solutions. Therefore singular Hessian matrix.

Adding Regularization

Objective Function

$$f(\mathbf{x}) = \frac{1}{2} \| \mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \frac{\lambda}{2} \left(\| \mathbf{A} \|_F^2 + \| \mathbf{B} \|_F^2 + \| \mathbf{C} \|_F^2 \right)$$

Gradient (for r = 1,…,R)

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

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

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

Resolves issue with scaling ambiguity and resulting singular Hessian.

Our methods: CPOPT & CPOPTR

CPOPT: Apply derivative-based optimization method to the following objective function:

$$f = \frac{1}{2} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

CPOPTR: Apply derivative-based optimization method to the following regularized objective function:

$$f = \frac{1}{2} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \frac{\lambda}{2} \left(\| \mathbf{A} \|_F^2 + \| \mathbf{B} \|_F^2 + \| \mathbf{C} \|_F^2 \right)$$

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 $(I+J+K)R \times IJK$, 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 factors 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	CPOPTR
$20 \times 20 \times 20$	0.5 ± 1.0	1.0 ± 1.1	0.3 ± 0.2	0.2 ± 0.1
$50 \times 50 \times 50$	0.3 ± 0.3	16.0 ± 17.7	0.7 ± 0.5	0.5 ± 0.1
$100 \times 100 \times 100$	1.7 ± 1.1	153.2 ± 142.3	5.6 ± 3.6	4.3 ± 1.3
$250 \times 250 \times 250$	26.6 ± 9.1	—	83.5 ± 35.2	81.9 ± 22.8

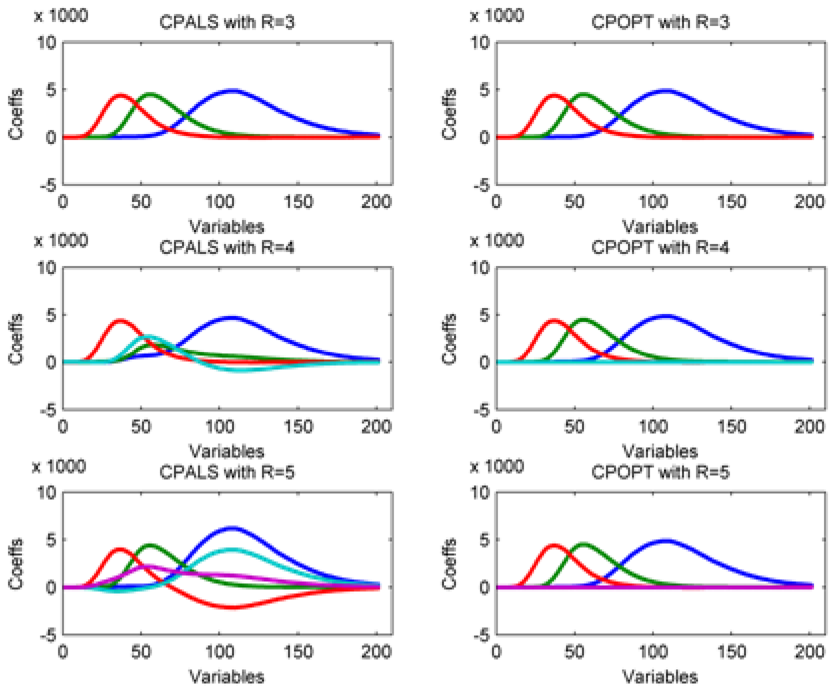
	Accuracy (%)			
Size	CPALS	CPNLS	CPOPT	CPOPTR
$20 \times 20 \times 20$	78.8	99.7	99.9	100.0
$50 \times 50 \times 50$	65.7	99.9	100.0	100.0
$100 \times 100 \times 100$	63.5	99.9	100.0	100.0
$250 \times 250 \times 250$	62.2	—	100.0	100.0

Further, CPOPT is scalable (see Evrim's talk)…

Sandia National Laboratories

Many Open Questions around Nonlinear Optimization Formulation

- CPD is a nonlinear optimization problem – great results with gradient approach, but we still need to consider…
  - Sensitivity to starting point
  - How to regularize
  - Issues of rank
  - Many more tests and methods…
- Other tensor decompositions can also be posed as optimization problems
  - See Elden and Savas for Tucker
- Consider imposing constraints
  - Symmetry
  - Sparsity in solution
  - Nonnegativity
  - Etc.



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

Sandia National Laboratories

Another Nonlinear Optimization Problem: Tensor Eigenpairs

$\mathcal{Z} \in \mathbb{R}^{K \times K \times K}$ supersymmetric

Qi, J. Symbolic Computation (2005);
Lim, IEEE Workshop (2005).

Definition 1

$$\sum_{j=1}^{K} \sum_{k=1}^{K} z_{ijk} x_j x_k = \lambda x_i$$

for i =1,…,K

Definition 2

$$\sum_{j=1}^{K} \sum_{k=1}^{K} z_{ijk} x_j x_k = \lambda x_i^2$$

for i =1,…,K

- Computational methods?
- How to construct test problems?
- What are the properties of tensor eigenvalues and eigenvectors?
- What are the applications?

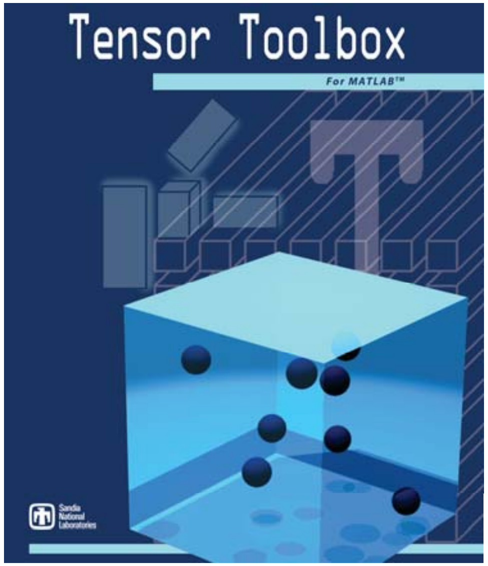
Sandia National Laboratories

Comments on Computing with Tensors

- Propose as model: Interface in Matlab Tensor Toolbox
  - Useful for writing new algorithms
  - If you aren't using it, tell us why!
  - Is there a need/demand for C++ or another language?
- Memory-efficient Tucker (MET)

$$g_{rst} = \sum_i \sum_j \sum_k x_{ijk} a_{ir} b_{js} c_{kt}$$

  - Avoids "intermediate blow-up" problem
  - May be of interest in terms of its simple optimization for "index fusion"



Bader & Kolda

Over 1900 Downloads since 9/2006 release.

Sandia National Laboratories

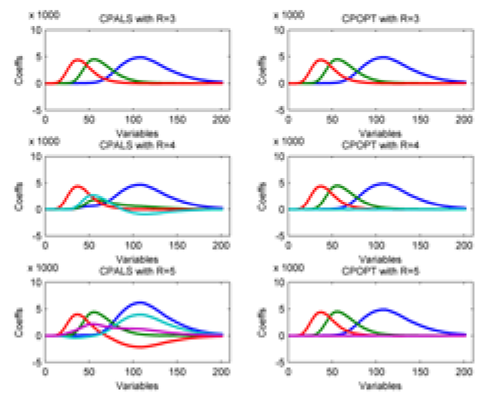
References & Contact Info

All papers available at: http://csmr.ca.sandia.gov/~tgkolda/

- **OPT:** Acar, Kolda and Dunlavy. **An Optimization Approach for Fitting Canonical Tensor Decompositions**, Technical Report SAND2009-0857, Feb 2009
- **MET:** Kolda and Sun. **Scalable Tensor Decompositions for Multi-aspect Data Mining**. In: ICDM 2008, pp. 363-372, Dec 2008 (paper prize winner)
- **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

Contacts

- Tammy Kolda, tgkolda@sandia.gov
- Evrim Acar, eacarat@sandia.gov
- Danny Dunlavy, dmdunla@sandia.gov



Sandia National Laboratories