SAND2014-19678C

Tensor Analysis for Sparse Data

Tamara G. Kolda
Sandia National Laboratories
Livermore, CA

Signature Discovery Workshop, Univ. Washington, Seattle
November 4, 2014

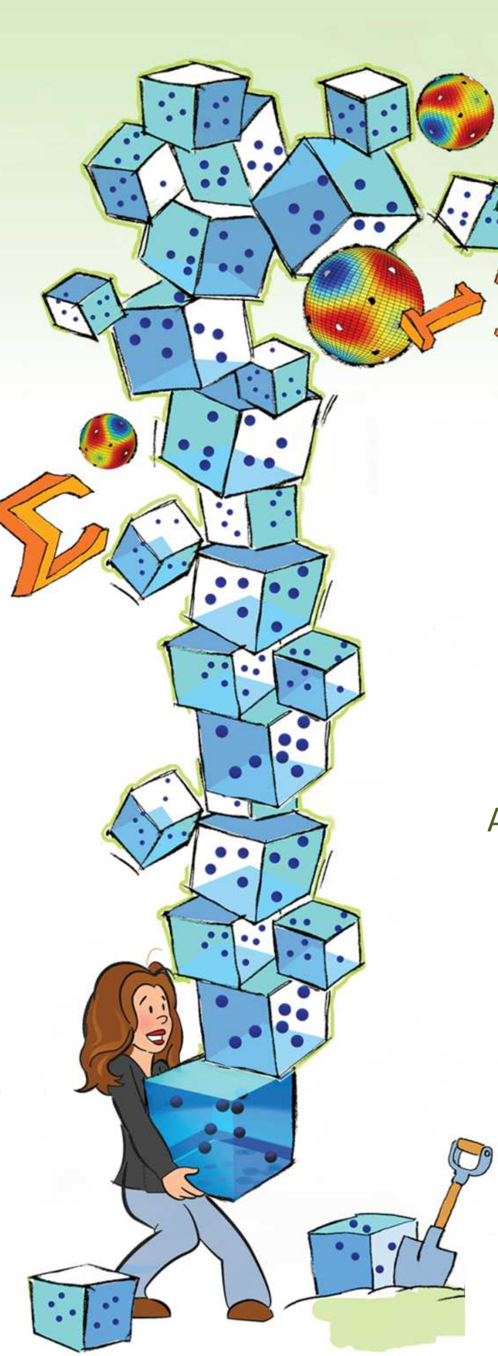
Illustration by Chris Brigman

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

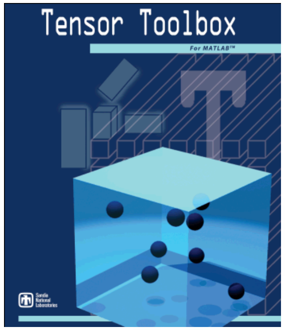
U.S. DEPARTMENT OF ENERGY

Sandia National Laboratories



Acknowledgements

Kolda and Bader, Tensor Decompositions and Applications, SIAM Review, 2009



Tensor Toolbox for MATLAB
Bader, Kolda, Acar, Dunlavy, and others

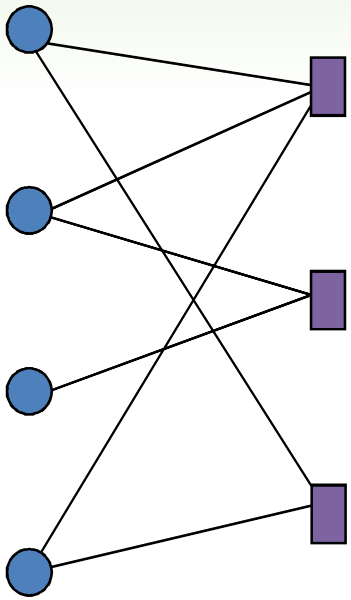
Illustration by Chris Brigman

Co-authors

- Evrim Acar (Univ. Copenhagen)
- Woody Austin (Univ. Texas Austin)
- Brett Bader (Digital Globe)
- Grey Ballard (Sandia)
- Eric Chi (Rice Univ.)
- Danny Dunlavy (Sandia)
- Sammy Hansen (Northwestern Univ.)
- Joe Kenny (Sandia)
- Jackson Mayo (Sandia)
- Morten Mørup (Denmark Tech. Univ.)
- Todd Plantenga (Sandia)
- Martin Schatz (Univ. Texas Austin)
- Teresa Selee (GA Tech Research Inst.)
- Jimeng Sun (GA Tech)

Plus many more collaborators for workshops, tutorials, etc.

Networks, Matrices, Factor Analysis

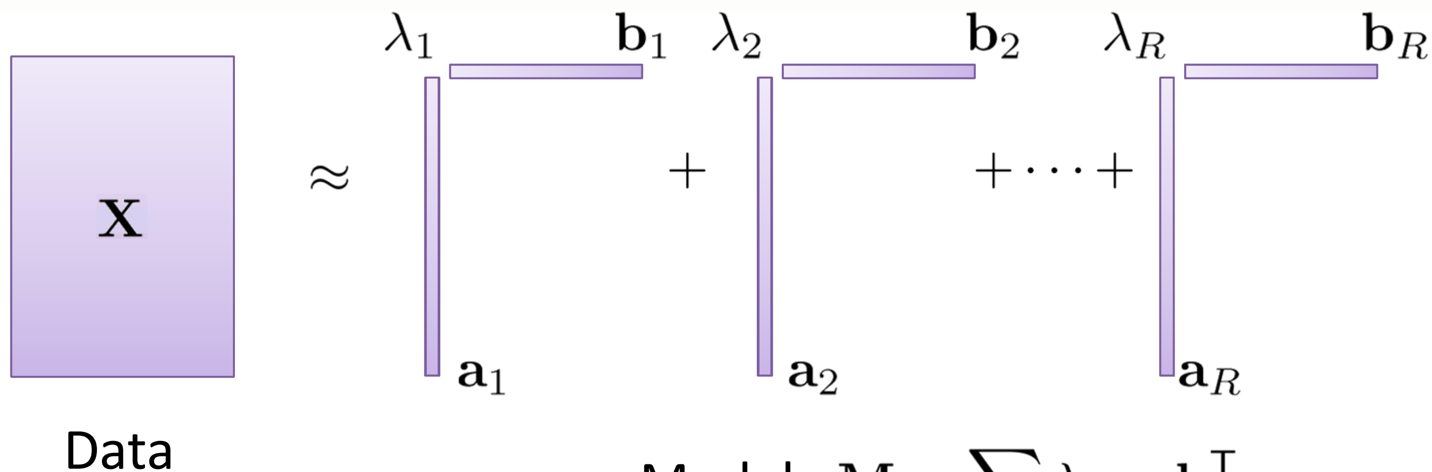


$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- Networks correspond to sparse matrices
  - Undirected Graph) Symmetric Matrix
  - Directed Graph) Asymmetric Matrix
  - Bipartite Graph) Rectangular Matrix
  - Unweighted Graph) Binary Matrix
- Matrix analysis yields insight
  - Matrix factorization
    - Singular Value Decomposition (SVD) and Principal Components Analysis (PCA)
    - Latent Semantic Indexing (LSI) (Dumais et al., 1988)
    - Independent Component Analysis (ICA) (Comon, 1994)
    - Nonnegative Matrix Factorization (Paatero, 1997; Bro & De Jong, 1997; Lee & Seung, 2001)
    - Compressive Sensing and related work (Candes, 2006)
  - Ranking methods
    - PageRank (Page et al., 1999)
    - Hubs & Authorities (Kleinberg, 1999)
  - Eigenvectors of Laplacian
    - Partitioning (Pothen, Simon, Liou, 1990)
    - Estimating commute time (Fouss et al., 2007)

Matrix Factorizations for Analysis

Singular Value Decomposition (SVD)

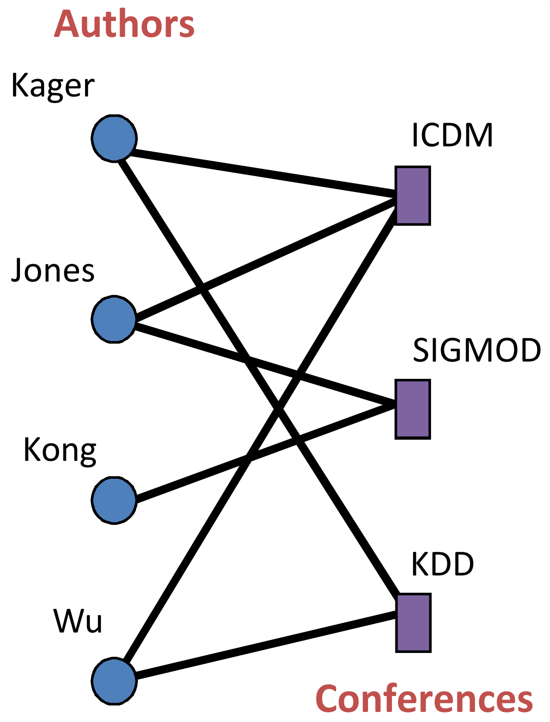


Model: $\mathbf{M} = \sum_r \lambda_r \, \mathbf{a}_r \mathbf{b}_r^{\mathsf{T}}$

$$\min \sum_{ij} (x_{ij} - m_{ij})^2 \quad \text{subject to} \quad m_{ij} = \sum_r \lambda_r \, a_{ir} \, b_{jr}$$

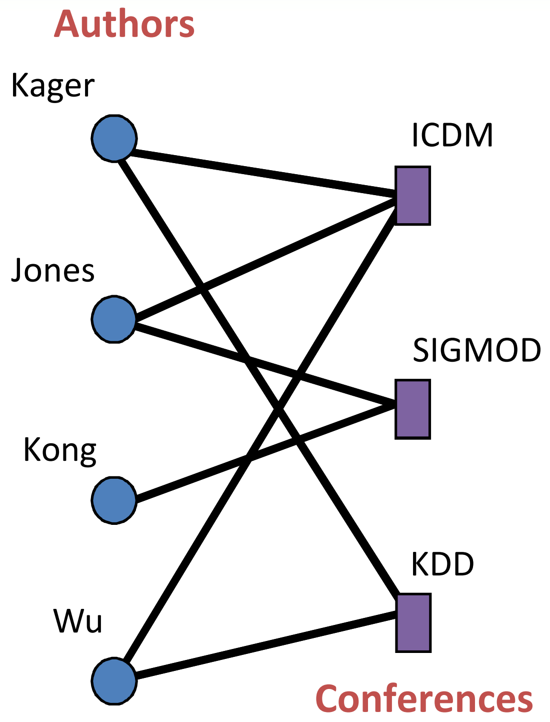
Key references: Beltrami (1873), Pearson (1901), Eckart & Young (1936)

Interpretation of 2-Way Factor Model



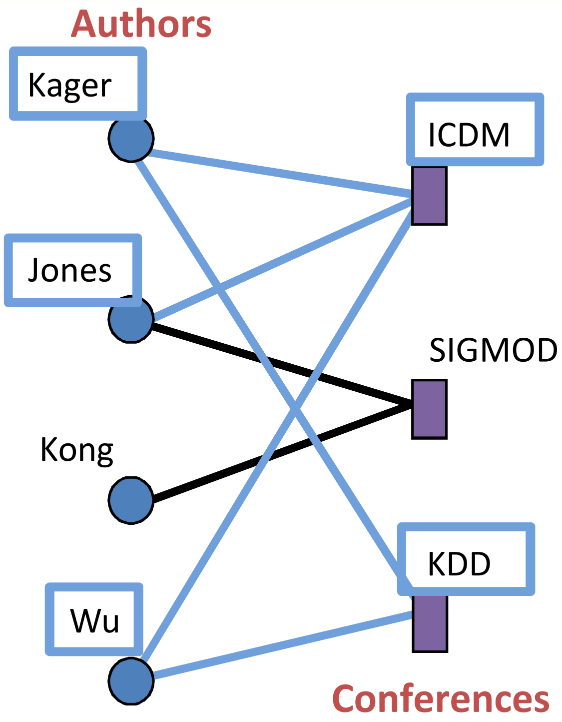
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Interpretation of 2-Way Factor Model



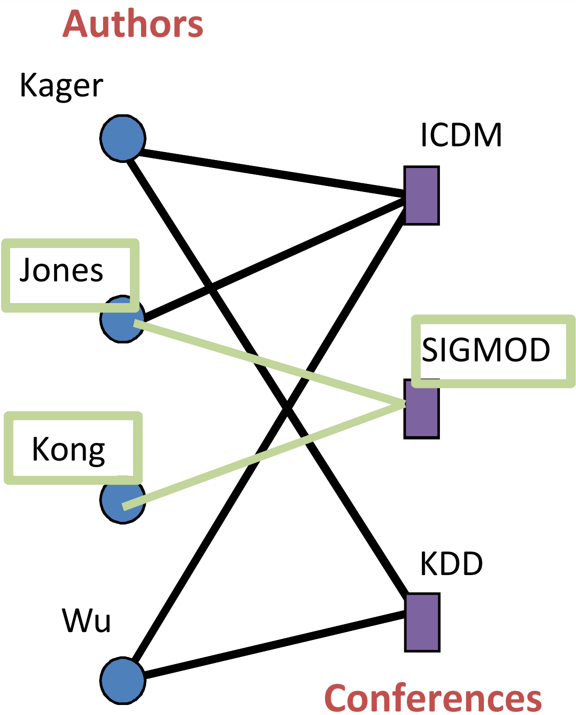
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model



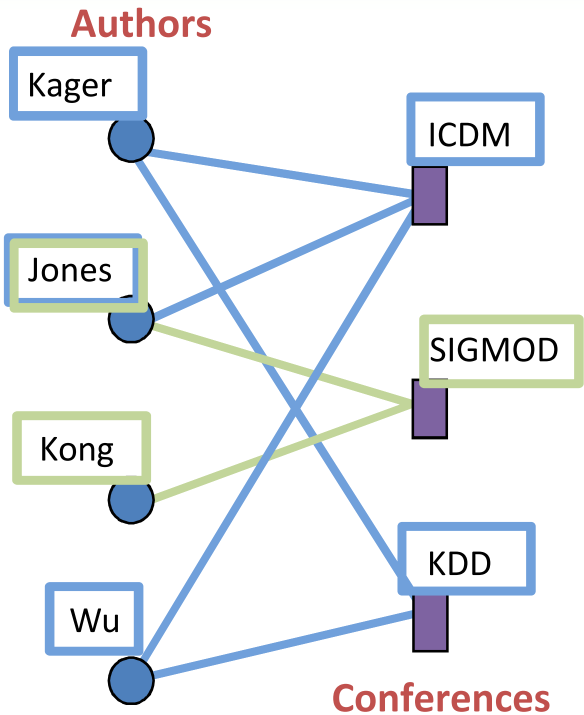
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model



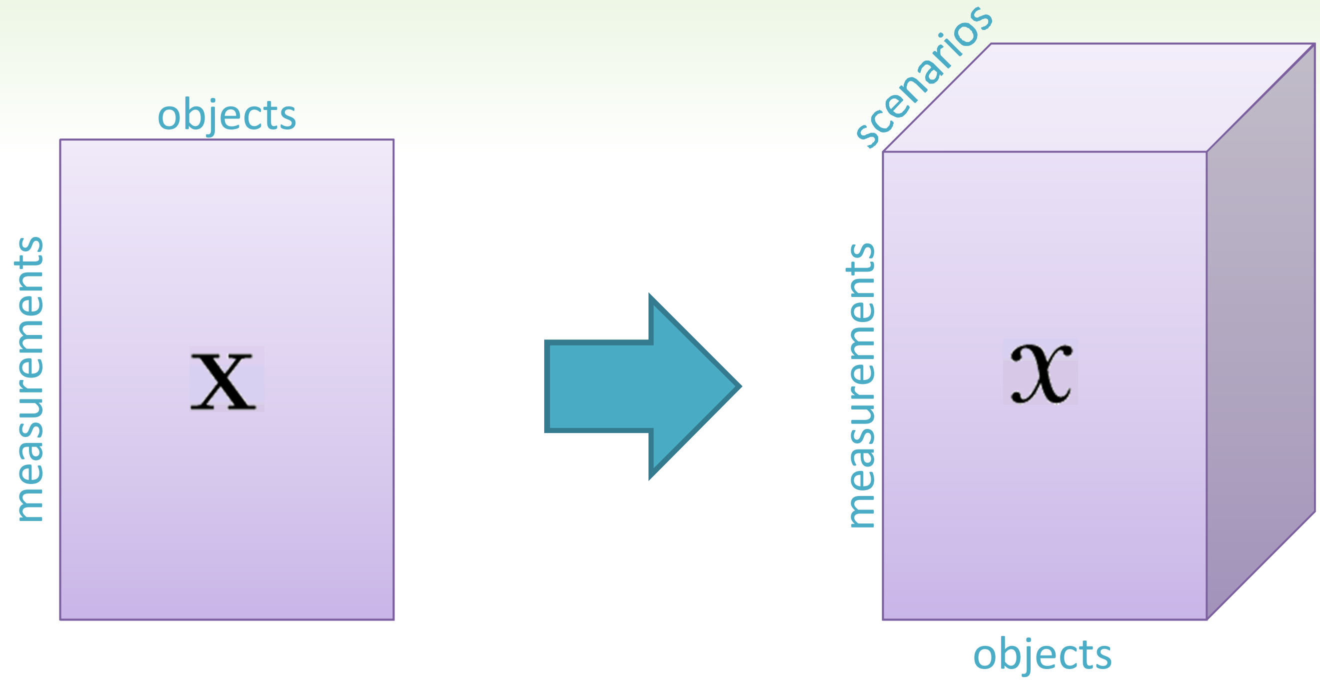
$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

Interpretation of 2-Way Factor Model



$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} .91 & -.38 \\ .72 & .75 \\ .19 & .75 \\ .91 & -.38 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1.15 & 0 \\ .41 & 1.06 \\ .83 & -.53 \end{bmatrix}^{\mathsf{T}}}_{\mathbf{B}^{\mathsf{T}}}$$

What about 3-way or N-way Data?



Key reference: Cattell , *Psychological Bulletin*, 1952

THE THREE BASIC FACTOR-ANALYTIC RESEARCH DESIGNS—THEIR INTERRELATIONS AND DERIVATIVES

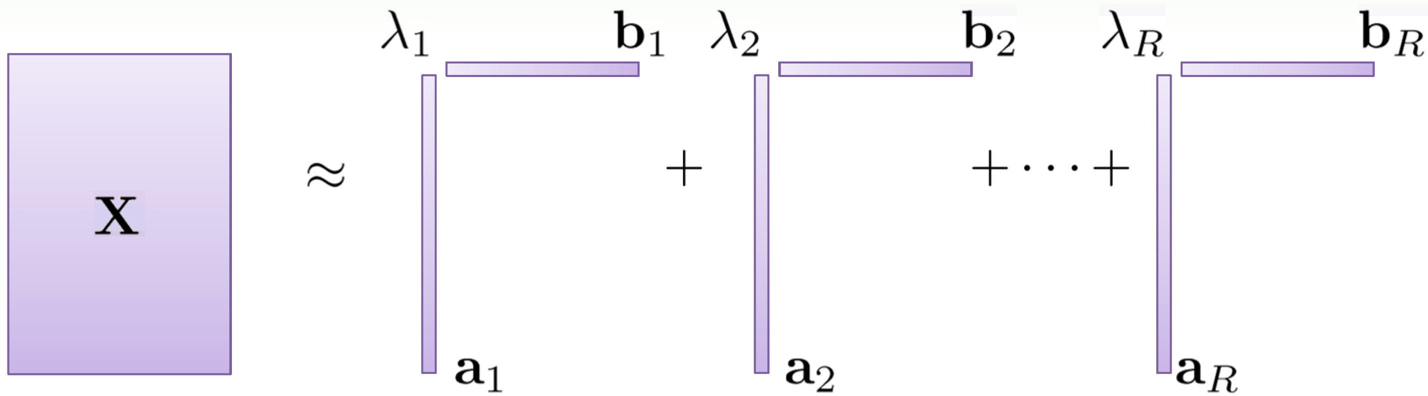
RAYMOND B. CATTELL

University of Illinois

Factor analysis began with the correlation of tests measured on populations of persons, but other arrangements have since been

Matrix Factorizations for Analysis

Think: SVD or NMF



$$\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_R]$$

$$\mathbf{B} = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_R]$$

Data

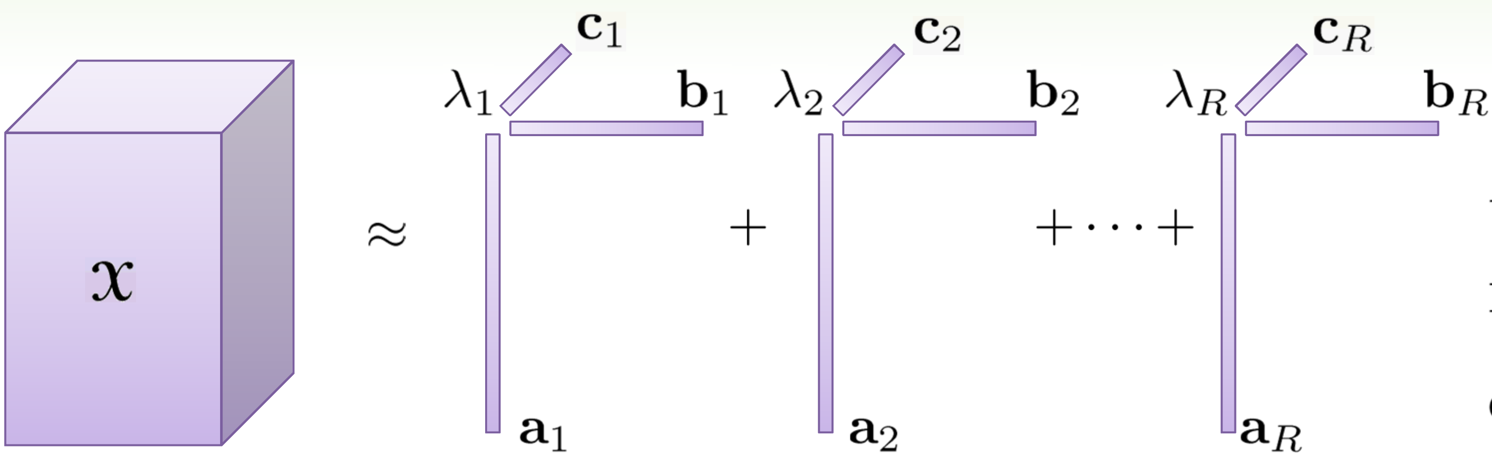
Model: $\mathbf{M} = \sum_r \lambda_r \, \mathbf{a}_r \mathbf{b}_r^{\mathsf{T}}$

$$\min \sum_{ij} (x_{ij} - m_{ij})^2 \quad \text{subject to} \quad m_{ij} = \sum_r \lambda_r \, a_{ir} \, b_{jr}$$

Key references: Beltrami (1873), Pearson (1901), Eckart & Young (1936)

Multi-way Factorizations for Analysis

CANDECOMP/PARAFAC (CP) Model



$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

Data

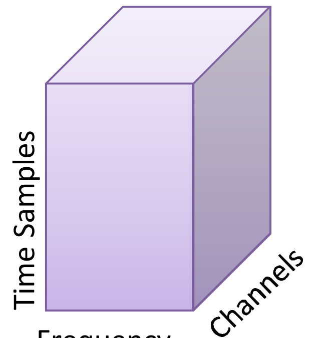
Model: $\mathcal{M} = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$

$$\min \sum_{ijk} (x_{ijk} - m_{ijk})^2 \quad \text{subject to} \quad m_{ijk} = \sum_r \lambda_r \, a_{ir} \, b_{jr} \, c_{kr}$$

Key references: Hitchcock (1927), Harshman (1970), Carroll and Chang (1970)

Tensor Factorization "Sorts Out" Comingled Data

Data measurements are recorded at multiple sites (channels) over time. The data is transformed via a continuous wavelet transform.

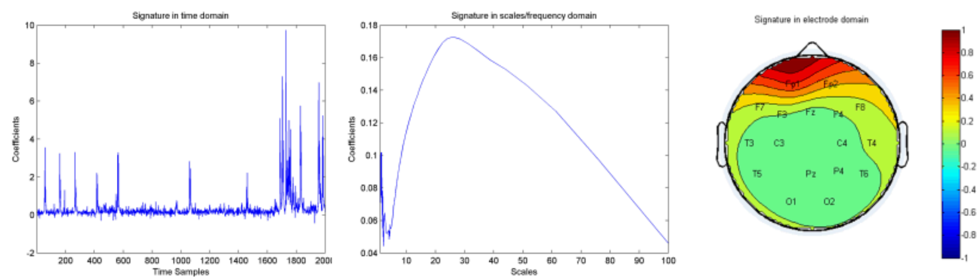


$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \mathcal{E}$$

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

Eye Artifact

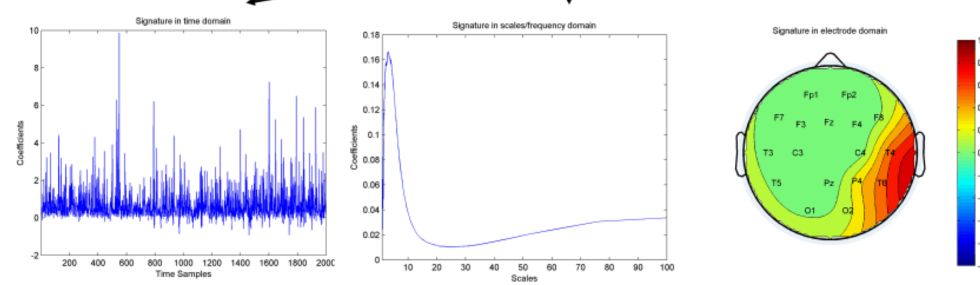
$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \mathcal{E}$$



Time Frequency Channel

Seizure

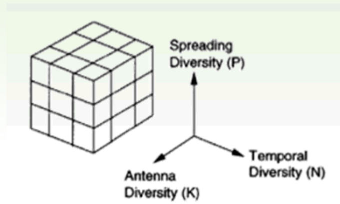
$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \mathcal{E}$$



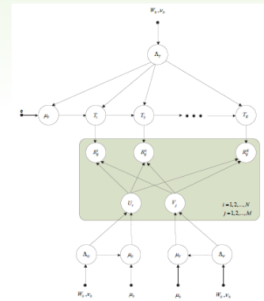
Time Frequency Channel

Tensor Factorizations have Numerous Applications

- Modeling fluorescence excitation-emission data (chemometrics)
- Signal processing
- Brain imaging (e.g., fMRI) data
- Network analysis and link prediction
- Image compression and classification; texture analysis
- Text analysis, e.g., multi-way LSI
- Approximating Newton potentials, stochastic PDEs, etc.
- Collaborative filtering
- Higher-order graph/image matching

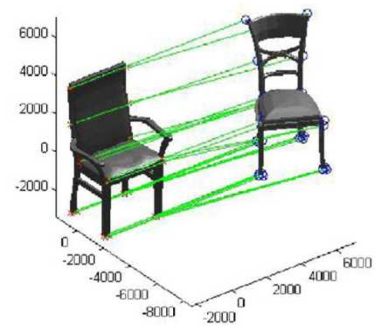


Sidiropoulos, Giannakis, and Bro, *IEEE Trans. Signal Processing*, 2000.

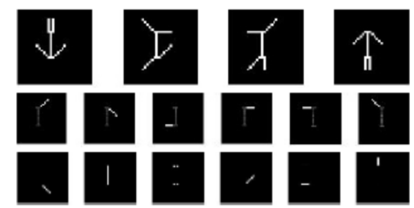


Xiong et al., *SDM'10*

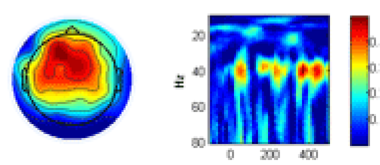
Furukawa, Kawasaki, Ikeuchi, and Sakauchi, *EGRW '02*



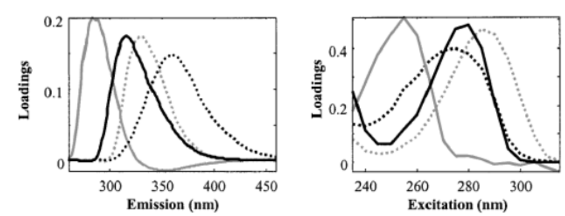
Duchenne, Bach, Kweon, Ponce, *TPAMI 2011*



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

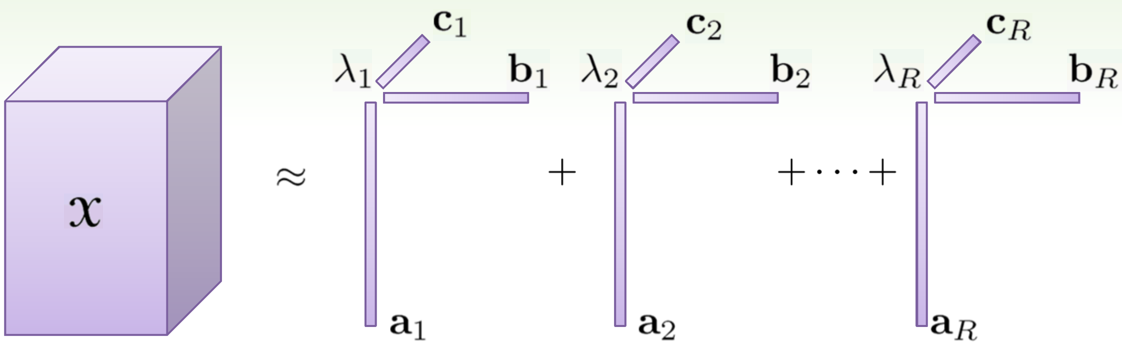


ERPWAVELAB by Morten Mørup.



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

Solving the Least Squares Problem



$$\mathcal{M} \approx \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix}$$

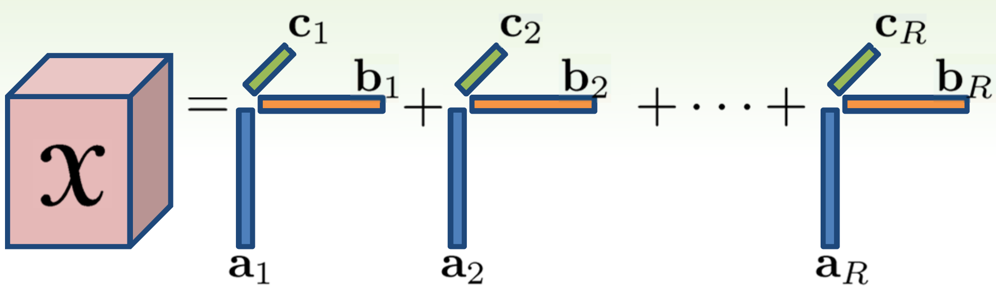
$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

$$\min_{\mathcal{M}} \sum_{ijk} (x_{ijk} - m_{ijk})^2$$

- Highly nonconvex problem!
  - Assume R is given
  - Need to find N factor matrices for N-way tensor
- Alternating least squares (ALS) (Harshman 1970; Phan et al. 2013)
  - Fix N-1 factor matrices and solve for the remaining one
  - Convex subproblem with easy solution (linear least squares)
- All-at-once optimization (Kolda, Dunlavy, Acar 2011; Phan et al. 2013)
  - Solves for all factor matrices simultaneously

Mathematical Facts on CP

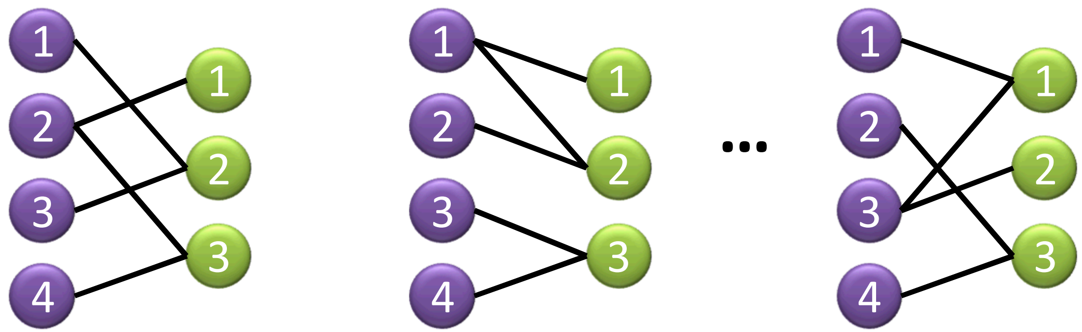


- Determining *exact* number of components is NP-Hard (Håstad 1990, Hillar & Lim 2009)
  - Example: Specific 9 x 9 x 9 tensor factorization problem corresponds to being able to do fast matrix multiplication of two 3x3 matrices – unknown what the rank is! (Bini et al. 1979)
  - More work needed on numerical techniques…
- Best low-rank factorization may not always exist (Silva & Lim 2006)
  - Sequence of low-rank factorizations may converge to a factorization of higher rank
- The best rank-(R-1) factorization is not necessarily part of the best rank-R factorization (Kolda 2001)
- Factorization is often *essentially unique* (unlike matrix factorization)
  - Up to permutation and scaling

Temporal Graphs & Tensors

Temporal Series of Bipartite Graphs

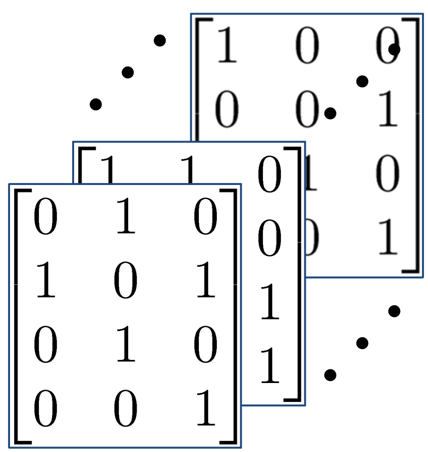
$G_1 = (V, W, E_1)$ $G_2 = (V, W, E_2)$ $G_T = (V, W, E_T)$



Tensor Representation

$$\mathcal{X} \in \mathbb{R}^{M \times N \times T}$$

$$x_{ijk} = \begin{cases} 1 & \text{if } (i,j) \in E_k \\ 0 & \text{if } (i,j) \notin E_k \end{cases}$$



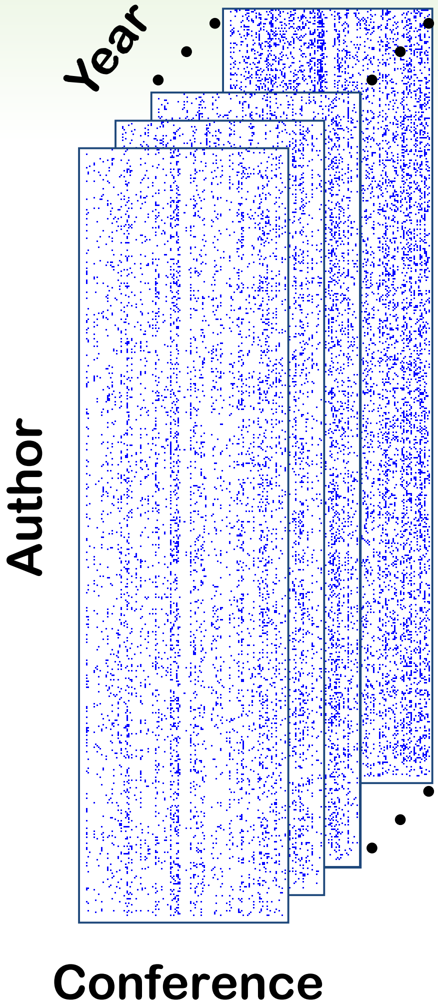
Tasks

- Principal Components
- Multidimensional Scaling
- Clustering
- Classification
- *Temporal Link Prediction*

Applications

- Obj. x Feature x Time
- Author x Conference x Year [Bibliometric]
- Person x Location x Time [GPS]

Temporal Analysis Example



DBLP has data from 1936-2007
(used only “inproceedings” from 1991-2000)

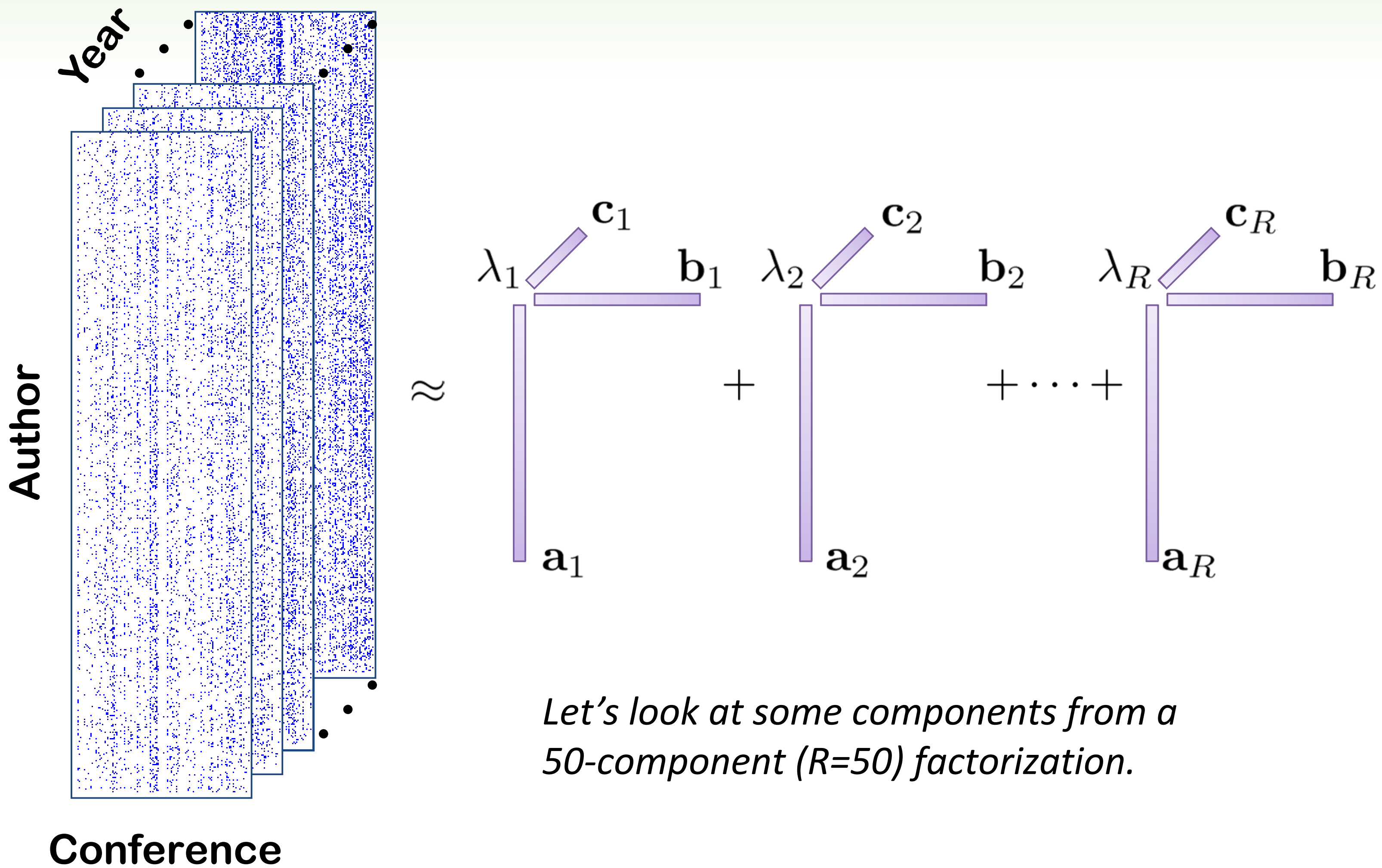
Training Data	10 Years: 1991-2000
# Authors (min 10 papers)	7108
# Conferences	1103
Links	113k (0.14% dense)

Nonzeros defined by:

$$x_{ijk} = \log(c_{ijk}) + 1 \text{ if } c_{ijk} > 0$$

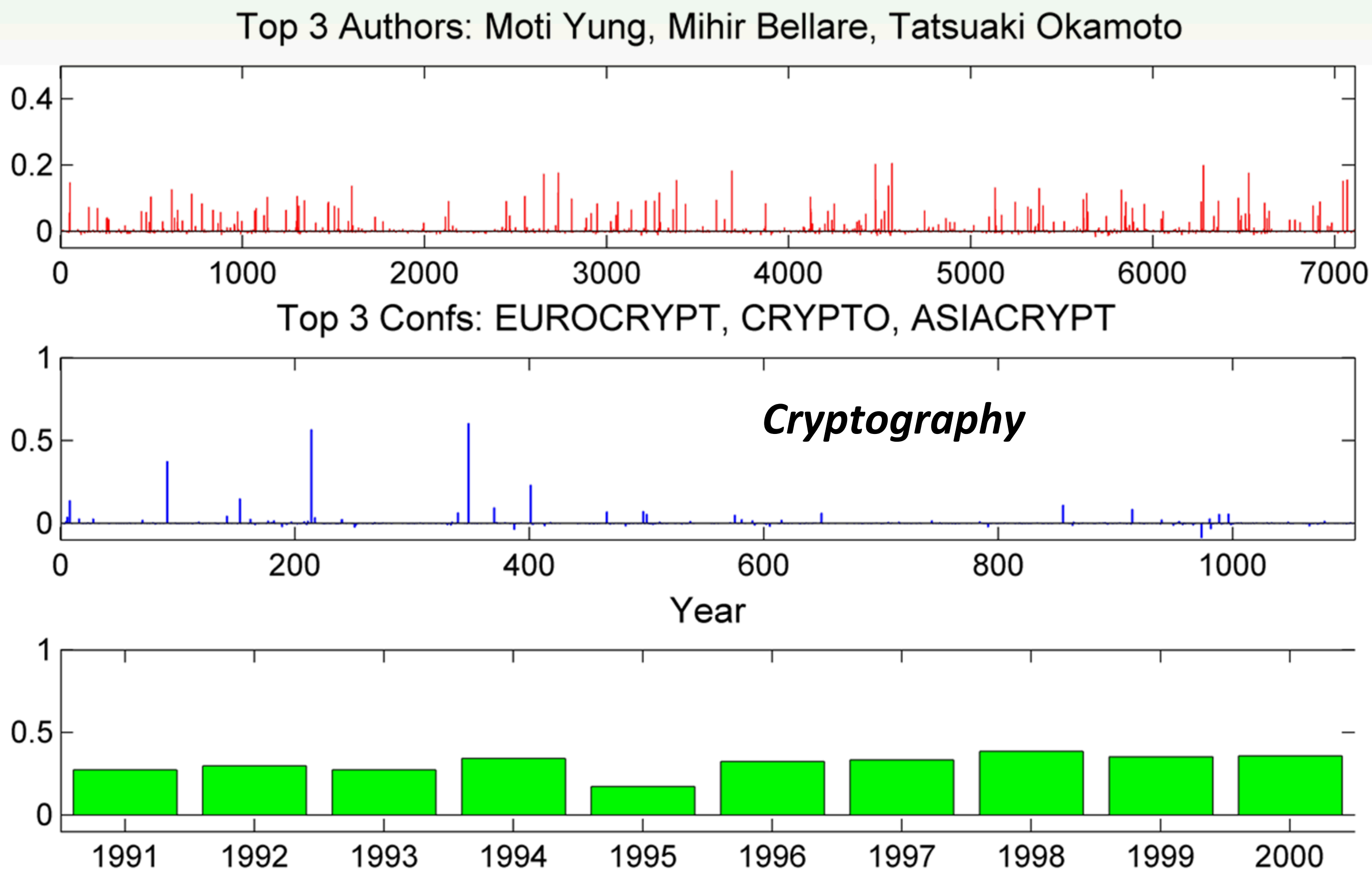
Acar, Dunlavy, & Kolda, Temporal Link Prediction using Matrix and Tensor Factorizations, ACM TKDD, 2010

Example: DBLP Data

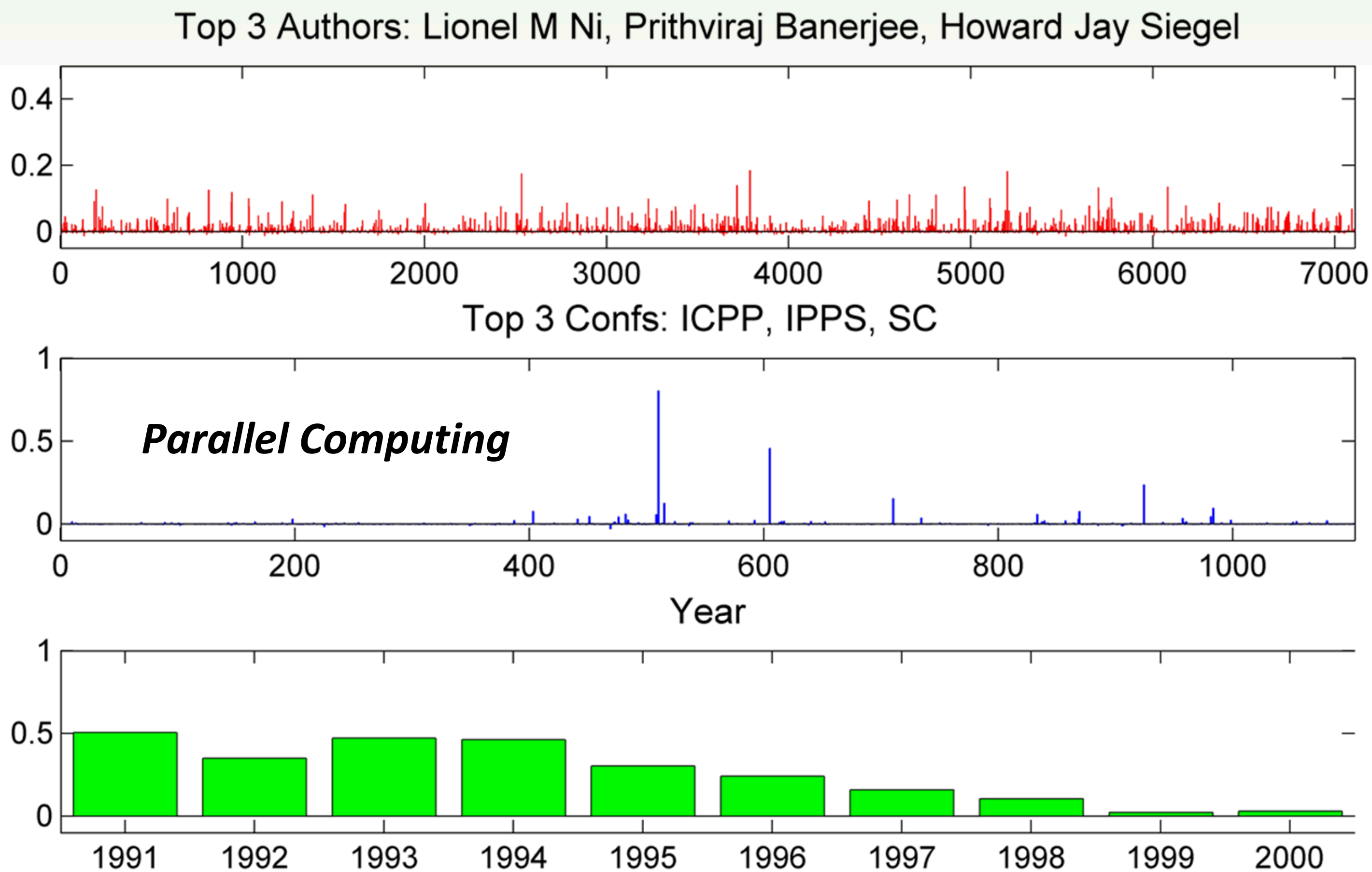


Let's look at some components from a 50-component (R=50) factorization.

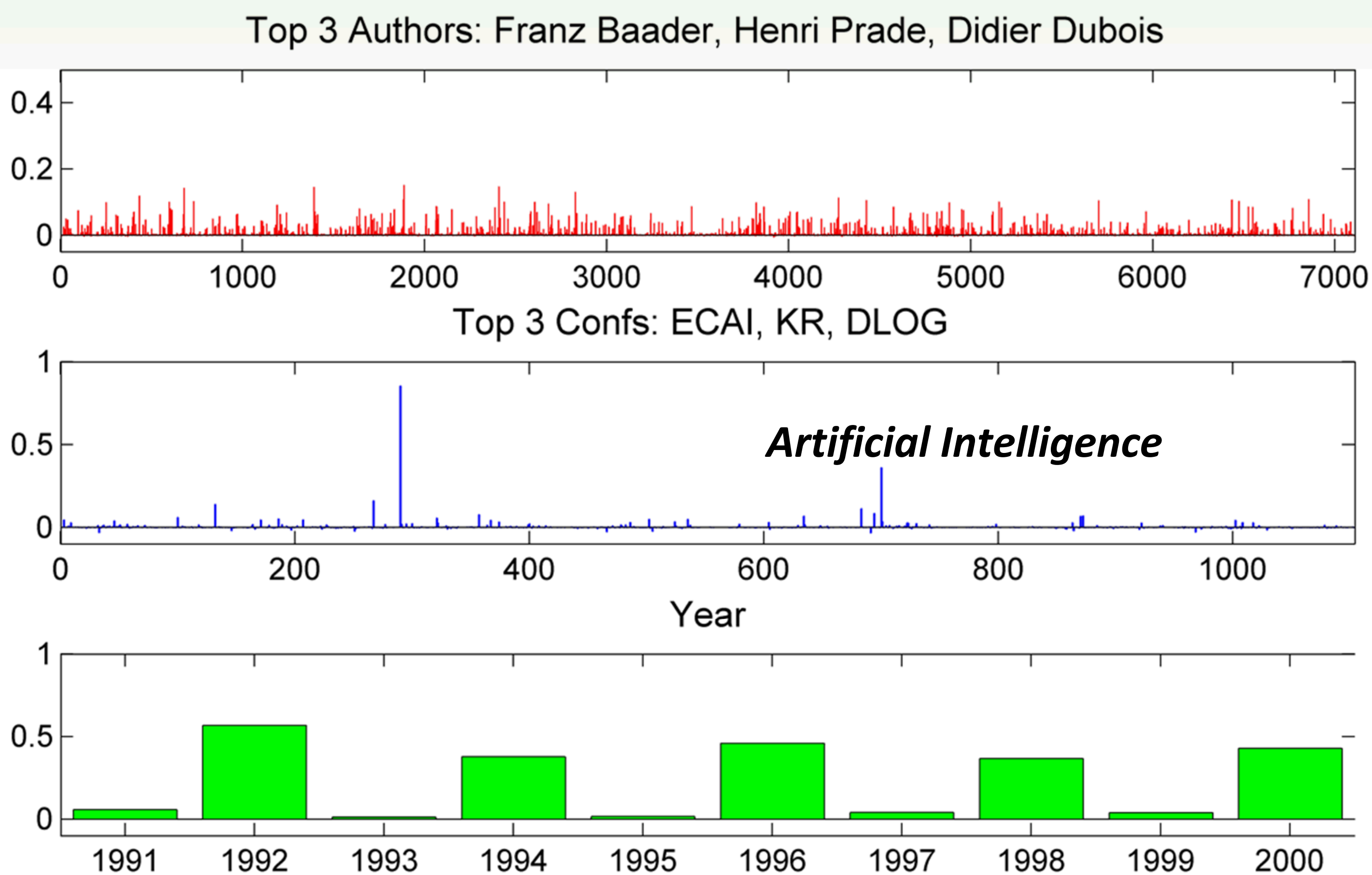
DBLP Component #30 (of 50)



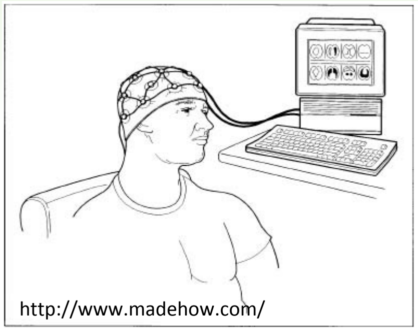
DBLP Component #19 (of 50)

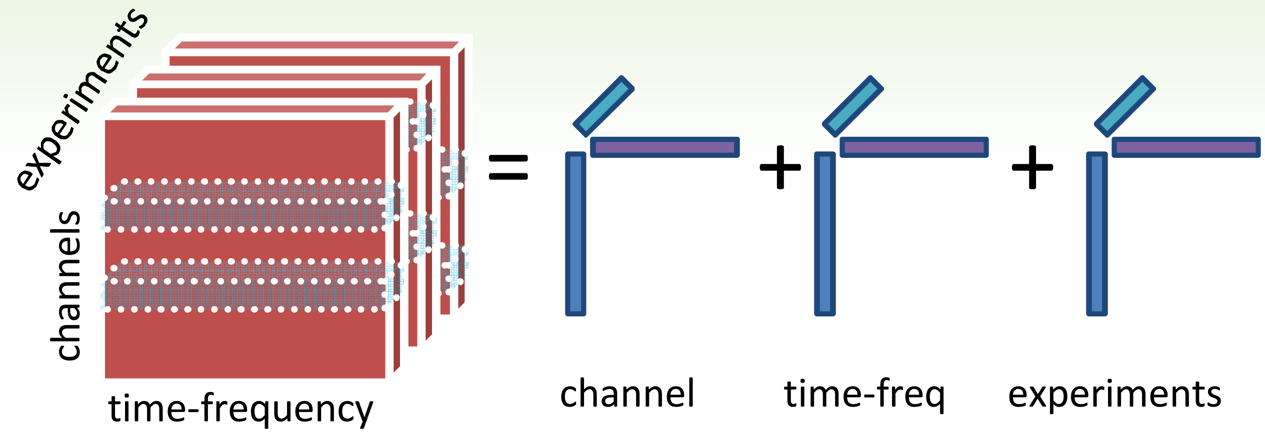


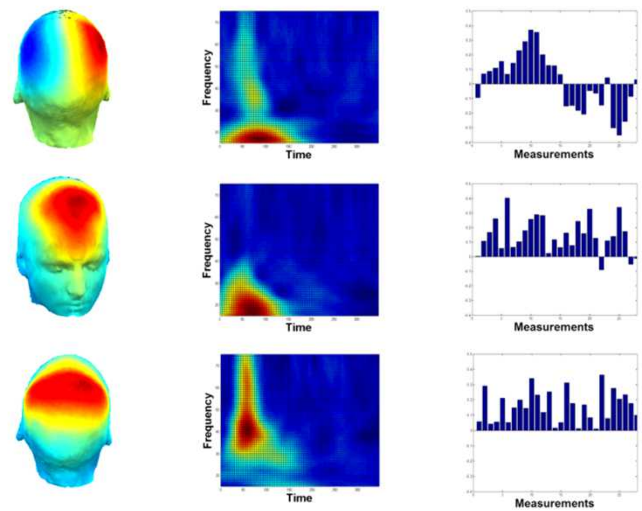
DBLP Component #43 (of 50)



Tensor Factorizations with Missing Data?







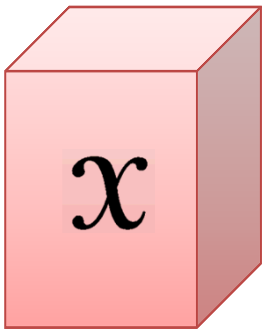
Biomedical signal processing

- EEG (electroencephalogram) signals can be recorded using electrodes placed on the scalp
- **Missing data problem** occurs when…
  - Electrodes get loose or disconnected, causing the signal to be unusable
  - Different experiments have over-lapping but not identical channels

Can we still do this calculation if data are missing?

The Missing Data Problem

Standard Problem:
Given tensor X, find A, B, and C such that…

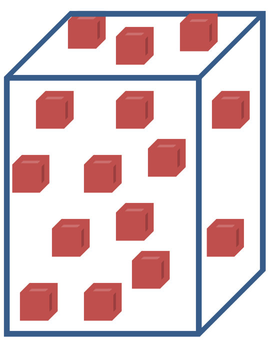


$$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

Typically formulated as a least squares problem.

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \frac{1}{2} \| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

Missing Data Problem:
Given a subset of the entries of X, find A, B, and C such that…

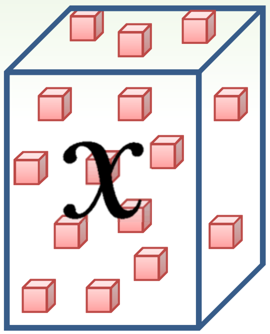


$$(\mathcal{X})_{ijk} = ([\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!])_{ijk}$$

for the known entries.

Mathematical Formulation

Define the "weight" tensor W such that



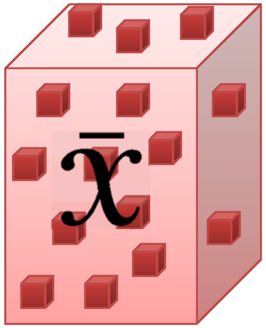
$$w_{ijk} = \begin{cases} 1 & \text{if entry } (i,j,k) \text{ of } \mathcal{X} \text{ is known} \\ 0 & \text{if entry } (i,j,k) \text{ of } \mathcal{X} \text{ is missing} \end{cases}$$

Then the least squares problem is…

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \frac{1}{2} \| \mathcal{W} * (\mathcal{X} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]) \|^2$$

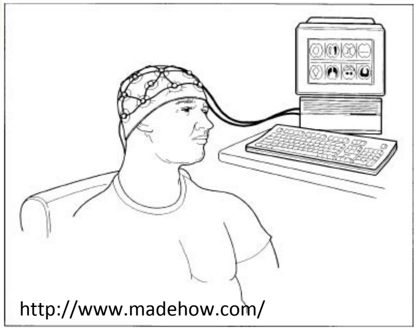
Elementwise product (.* in MATLAB)

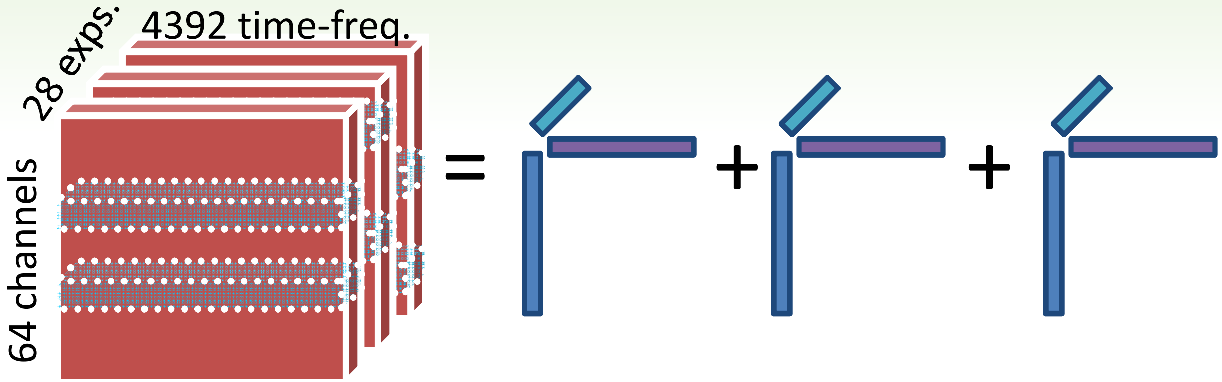
With a solution, the tensor can be "completed" via…



$$\bar{x}_{ijk} = \begin{cases} x_{ijk} & \text{if entry } (i,j,k) \text{ of } \mathcal{X} \text{ is known} \\ ([\![\mathbf{A},\mathbf{B},\mathbf{C}]\!])_{ijk} & \text{if entry } (i,j,k) \text{ of } \mathcal{X} \text{ is missing} \end{cases}$$

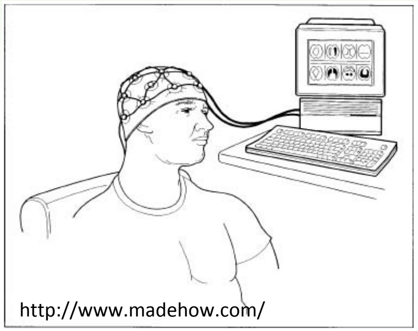
Brain dynamics can be captured even extensive missing channels

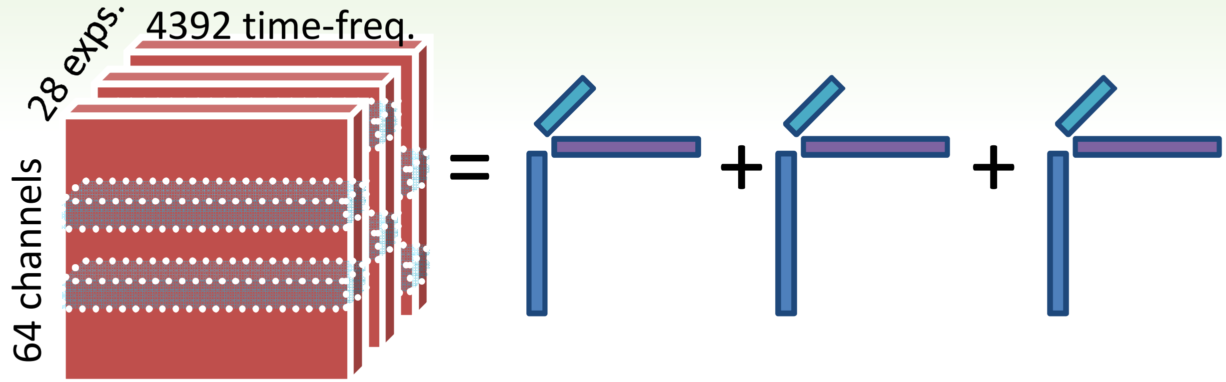




Number of Missing Channels	Replace Missing Entries with Zero
1	0.98
10	0.82
20	0.67
30	0.45
40	0.24

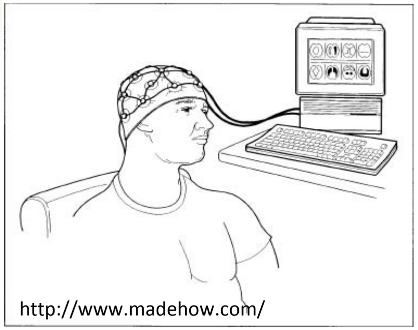
Brain dynamics can be captured even extensive missing channels

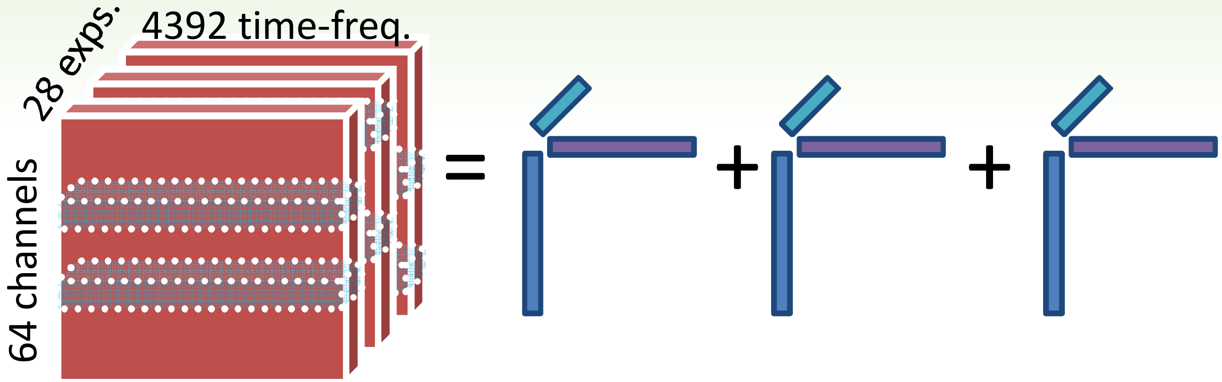




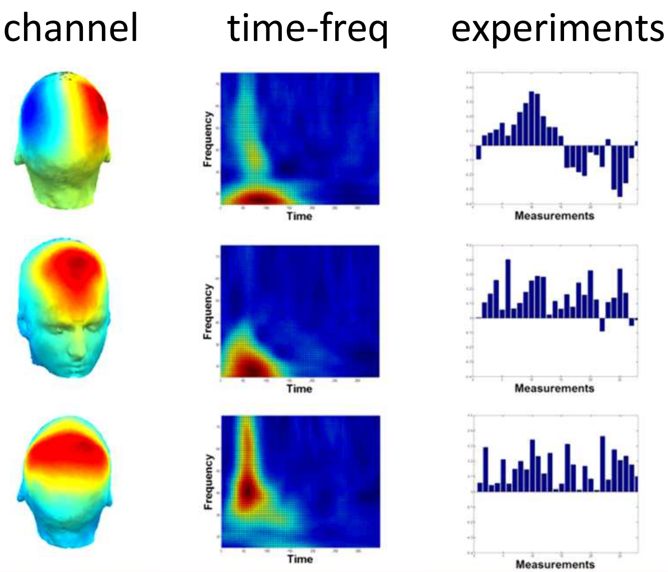
Number of Missing Channels	Replace Missing Entries with Zero	More Sensible Approach
1	0.98	1.00
10	0.82	0.98
20	0.67	0.95
30	0.45	0.89
40	0.24	0.65

Brain dynamics can be captured even extensive missing channels

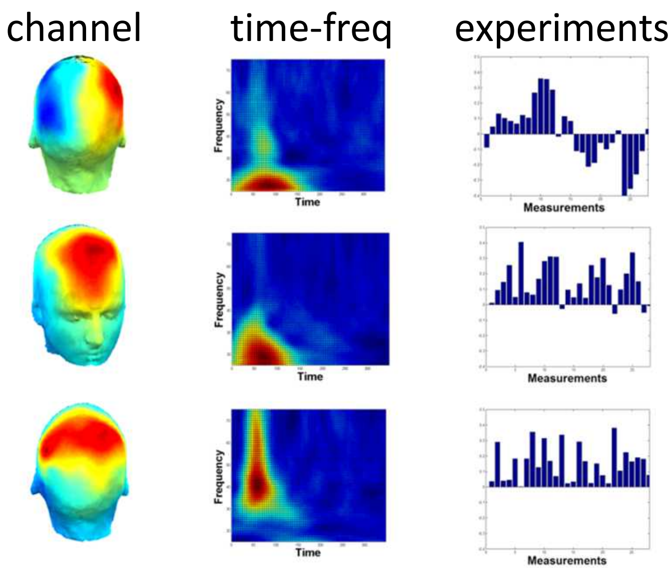




No Missing Data



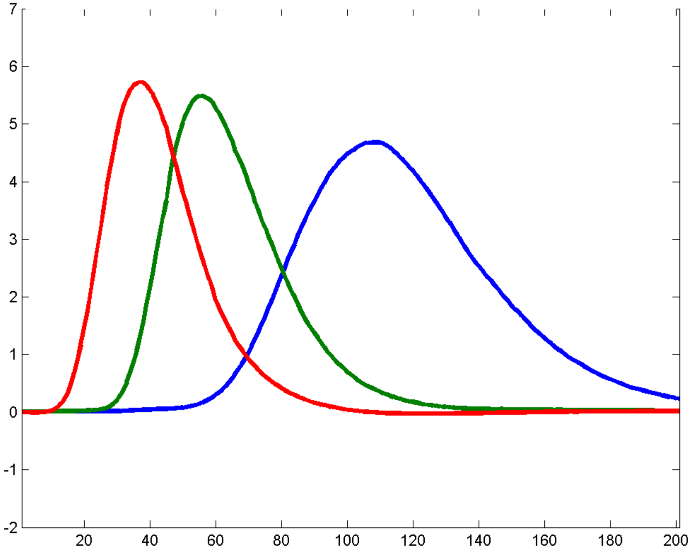
30 Chan./Exp. Missing



Chemometrics Example

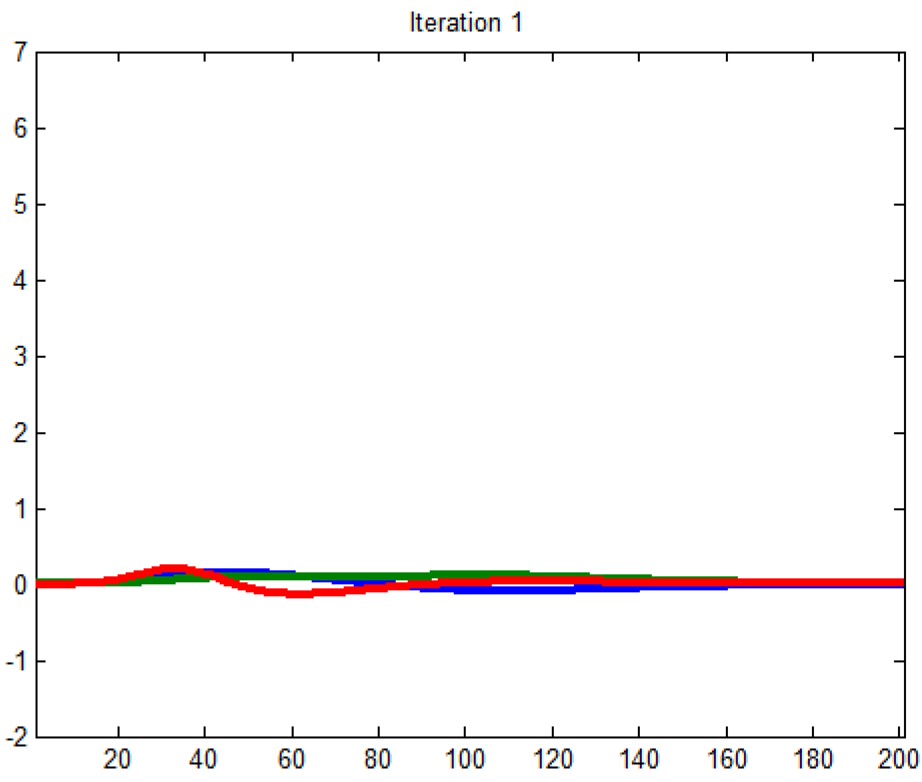
- Fluorescence measurements of 5 samples containing 3 amino acids
  - Tyrosine
  - Tryptophan
  - Phenylalanine
- Each amino acid corresponds to a rank-one components
- Tensor of size 5 x 51 x 201
  - 5 samples
  - 51 excitations
  - 201 emissions

Factors in Emission Mode



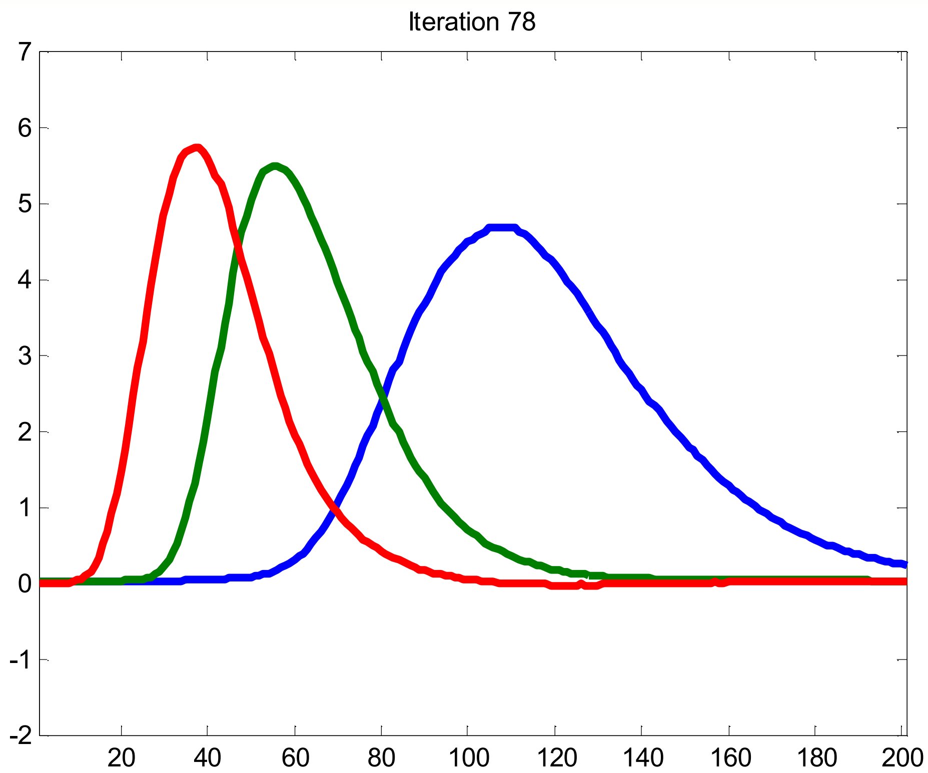
Bro (1997): http://www.models.kvl.dk/amino_acid_fluo

No Missing Data

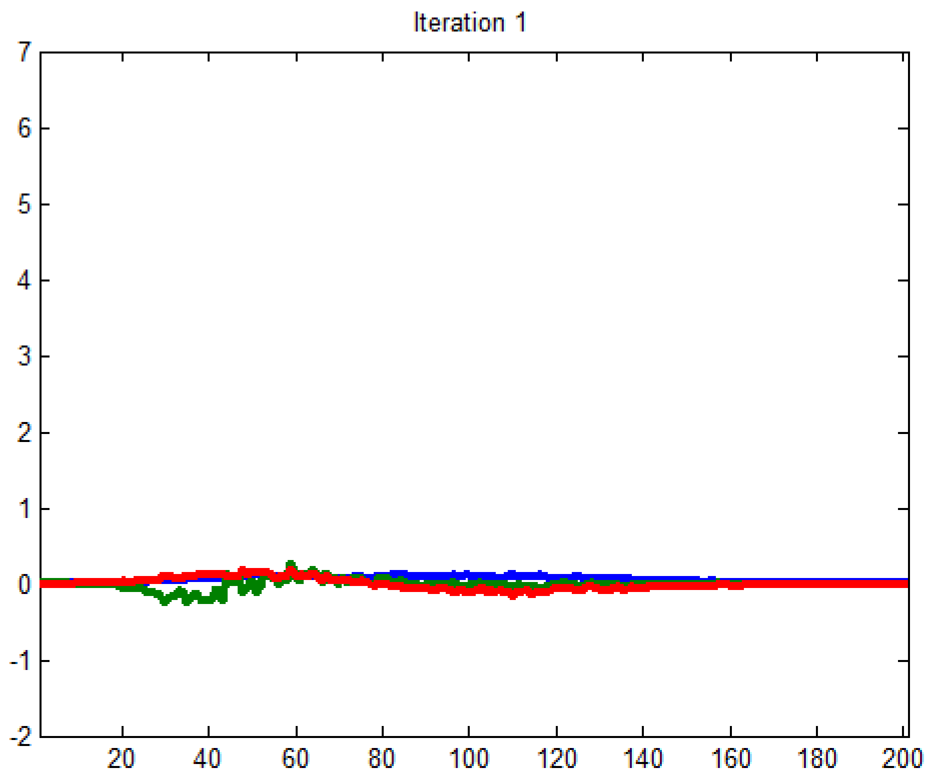


[Movie]

No Missing Data

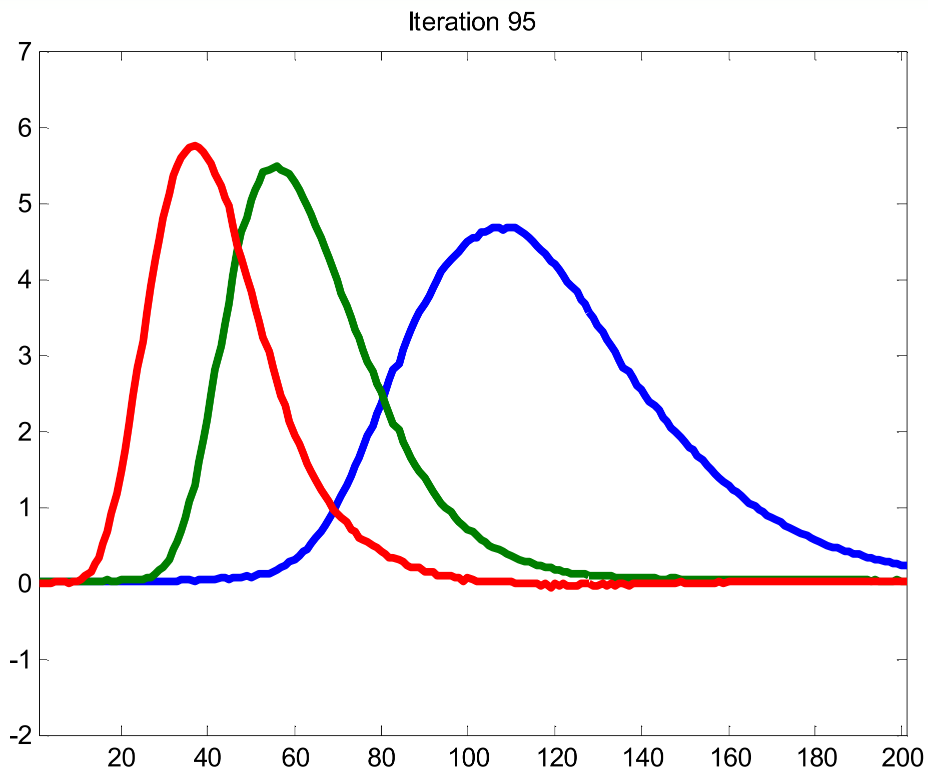


75% Missing Data using Sensible Approach



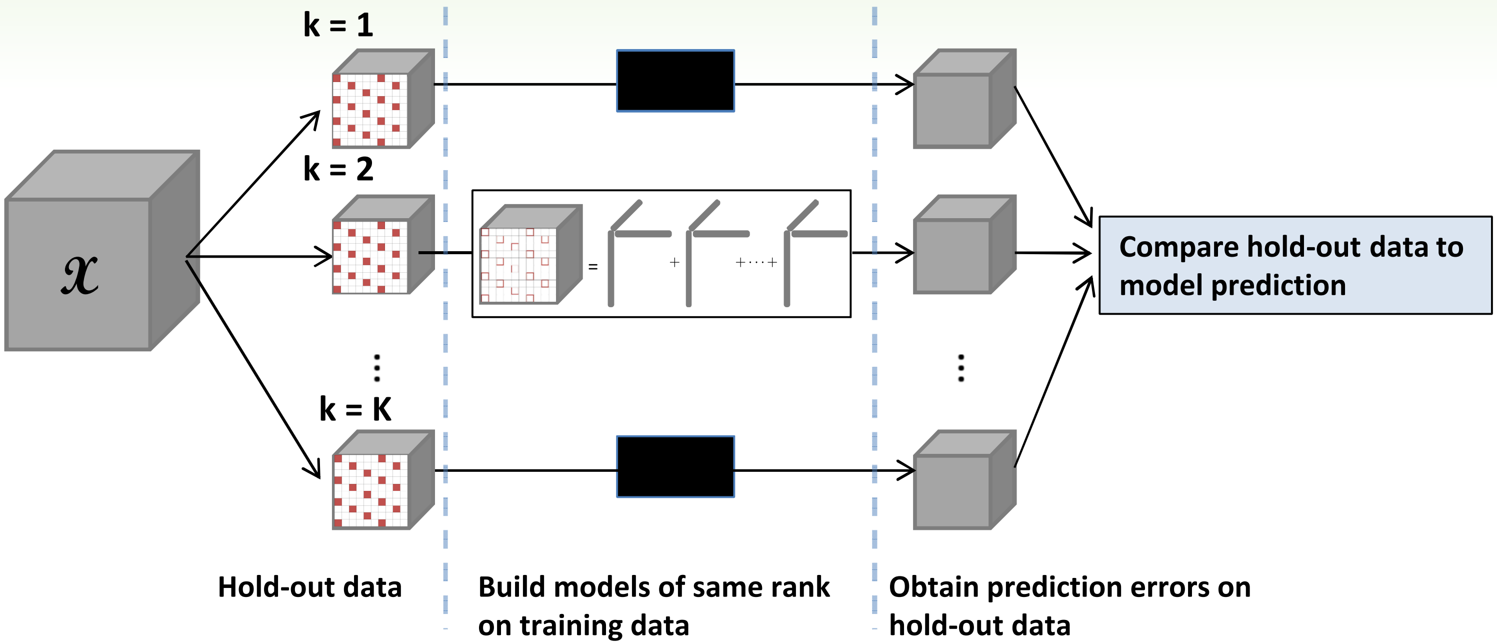
[Movie]

75% Missing Data using Sensible Approach



Statistical Rank Determination

For a given R use K-fold cross validation to calculate prediction error:

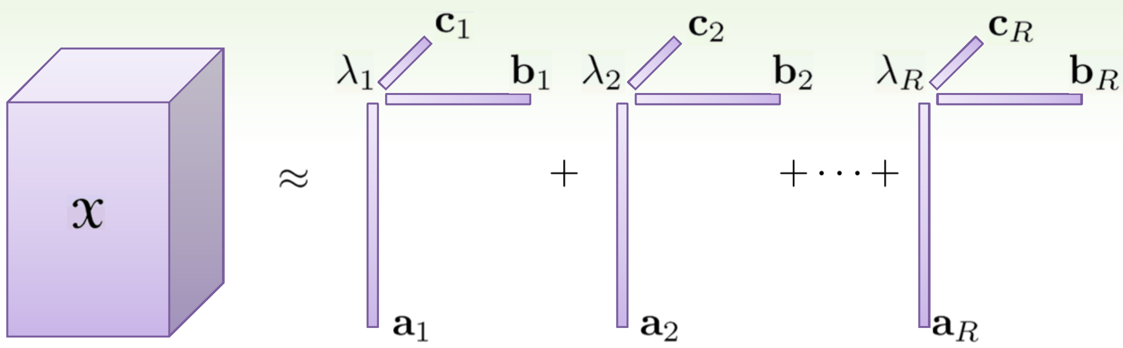


Ignore hold-out data in fitting model:

$$\arg\min_{\mathcal{M}} f(\mathcal{M}) = \sum_{\mathbf{i}\notin\mathbf{\Phi}}(m_{\mathbf{i}} - x_{\mathbf{i}})^2$$

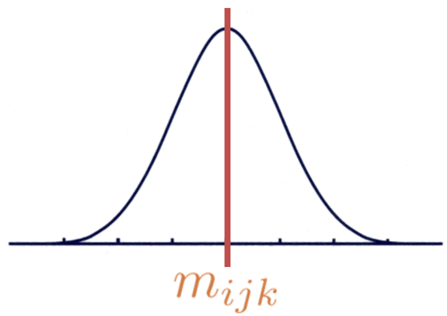
Austin & Kolda (in progress)

Gaussian Distributed Data



- Typically, we minimize the least-squares error
- This corresponds to maximizing the likelihood, assuming a **Gaussian distribution**

$$x_{ijk} \sim N(m_{ijk}, \sigma^2)$$



Maximize this:

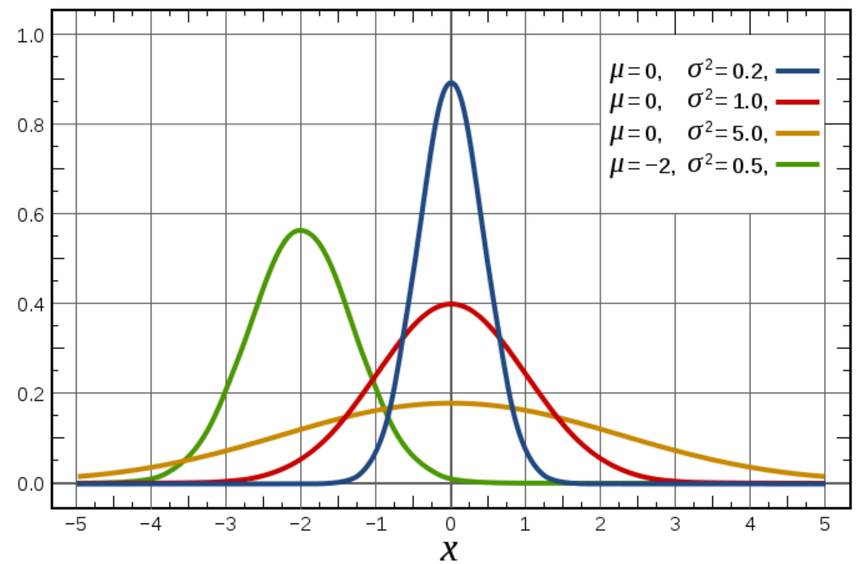
$$\text{likelihood}(\mathcal{M}) = \prod_{ijk} \frac{\exp(-(x_{ijk} - m_{ijk})^2 / 2\sigma^2)}{2\pi\sigma^2}$$

By monotonicity of log, same as maximizing this:

$$\text{log-likelihood}(\mathcal{M}) = c_1 - c_2 \sum_{ijk} (x_{ijk} - m_{ijk})^2$$

Gaussian is often Good, But…

- Gaussian (aka normal) distribution is prominent is statistics
  - Limiting distribution of the sum of a large number of random variables
  - Often a reasonable model for measurement/observational errors
- But, some data are better understood via alternative distributions
  - Data with outliers or multiple modes
  - Count data with many low counts



http://commons.wikimedia.org/wiki/File:Normal_Distribution_PDF.svg

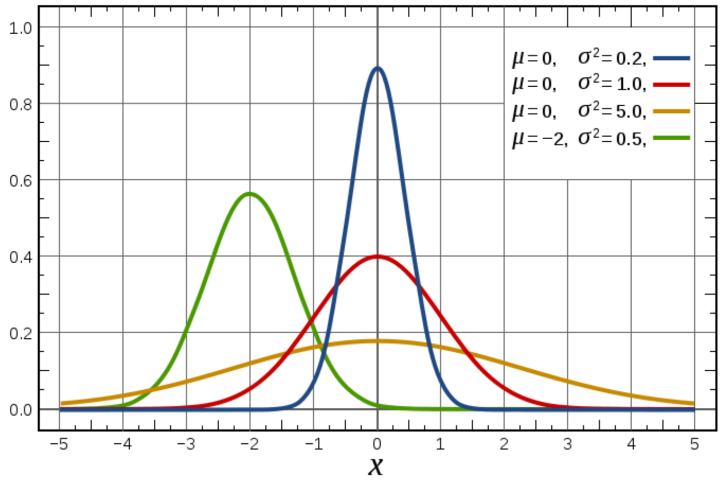
Poisson is Better for Count Data

Gaussian (typical)

The random variable x is a continuous real-valued number.

$$x \sim N(m, \sigma^2)$$

$$P(X = x) = \frac{\exp(-\frac{(x-m)^2}{2\sigma^2})}{\sqrt{2\pi\sigma^2}}$$



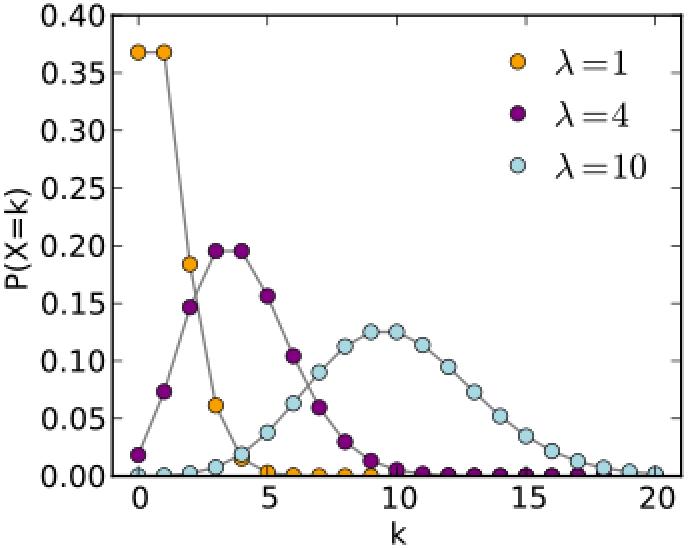
http://commons.wikimedia.org/wiki/File:Normal_Distribution_PDF.svg

Poisson

The random variable x is a discrete nonnegative integer.

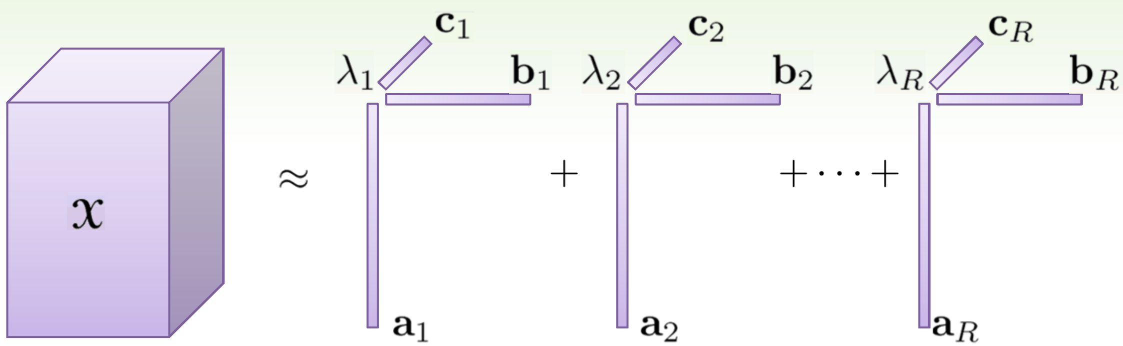
$$x \sim \text{Poisson}(m)$$

$$P(X = x) = \frac{\exp(-m)m^x}{x!}$$



http://en.wikipedia.org/wiki/File:Poisson_pmf.svg

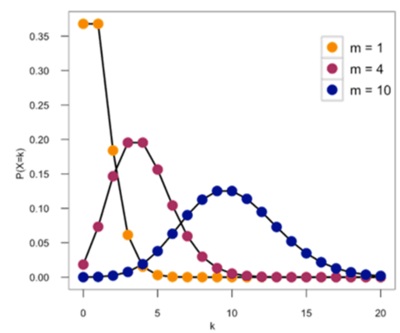
Poisson Tensor Factorization (PTF)



- Poisson preferred for sparse count data
- Automatically nonnegative
- More difficult objective function than least squares
- Note that this objective is also called Kullback-Liebler (KL) divergence

$$x_{\mathbf{i}} \sim \text{Poisson}(m_{\mathbf{i}})$$

$$P(X=x) = \frac{\exp(-m)m^x}{x!}$$



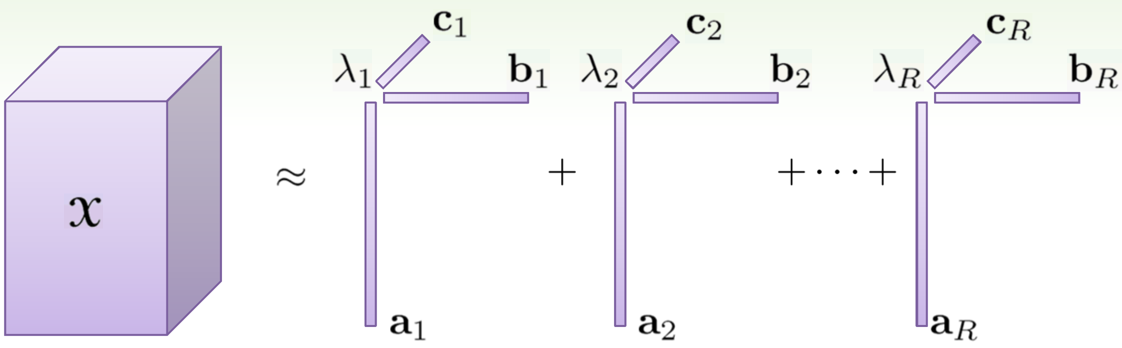
Maximize this:

$$\text{likelihood}(\mathcal{M}) = \prod_{\mathbf{i}} \frac{\exp(-m_{\mathbf{i}})\, m_{\mathbf{i}}^{x_{\mathbf{i}}}}{x_{\mathbf{i}}!}$$

By monotonicity of log, same as maximizing this:

$$\text{log-likelihood}(\mathcal{M}) = c - \sum_{\mathbf{i}} m_{\mathbf{i}} - x_{\mathbf{i}} \log(m_{\mathbf{i}})$$

Solving the Poisson Regression Problem



$$\mathcal{M} \approx \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix}$$

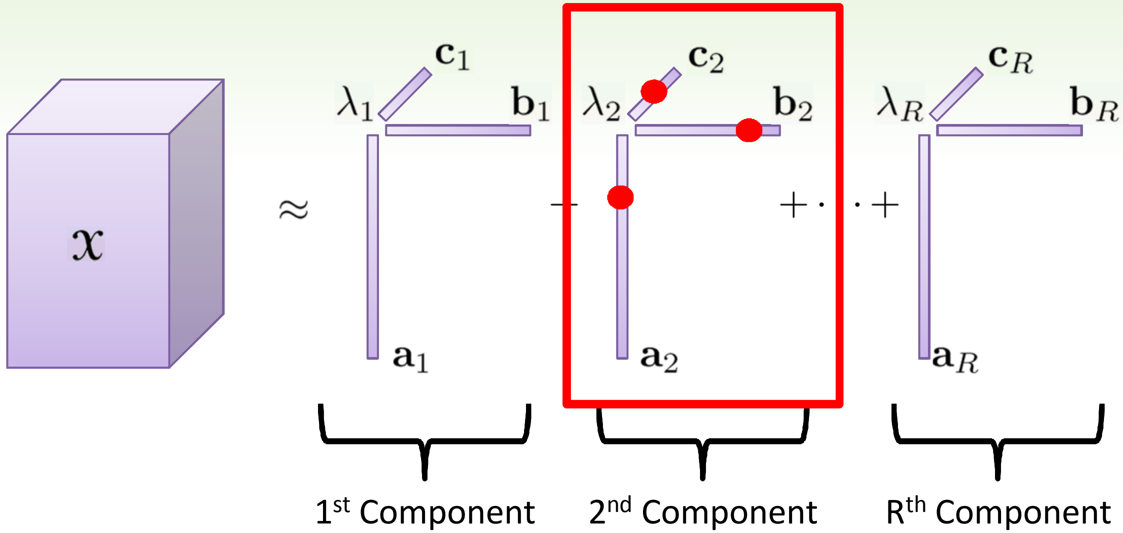
$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

$$\min_{\mathcal{M}} \sum_{ijk} m_{ijk} - x_{ijk} \log m_{ijk}$$

- Highly nonconvex problem!
  - Assume R is given
  - Need to find N factor matrices for N-way tensor
- Alternating Poisson regression
  - Assume (N-1) factor matrices are known and solve for the remaining one
  - Multiplicative updates (Chi & Kolda 2013)
  - Newton or Quasi-Newton method (Hansen, Plantenga, Kolda 2014)
  - Can adapt statistical rank test for Poisson too (Austin & Kolda TBD)

Interpreting PTF



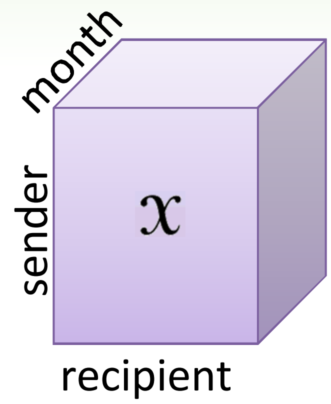
$$\boldsymbol{\lambda} \geq 0,\ \mathbf{A} \geq 0,$$
$$\mathbf{B} \geq 0,\ \mathbf{C} \geq 0,$$
$$\|\mathbf{a}_r\| = \|\mathbf{b}_r\| = \|\mathbf{c}_r\| = 1$$
$$\text{for } r = 1, \ldots, R$$

¸$_r$ / Probability of choosing component r
a_{ir} = Probability of choosing object i (given component r)
b_{jr} = Probability of choosing object j (given component r)
c_{kr} = Probability of choosing object k (given component r)

To generate data according to this model:
Choose r according to ¸. Choose (i,j,k) according to $\mathbf{a}_r$, $\mathbf{b}_r$, $\mathbf{c}_r$. Add 1 to entry (i,j,k). Repeat.

PTF for Enron Email

- Emails from Enron FERC investigation
  - 8540 Messages
  - 28 Months (from Dec 1999 to Mar 2002)
  - 105 People (sent and received at least one email every month)
  - x_{ijk} = # emails from sender i to recipient j in month k
  - 105 x 105 x 28 = 308,700 possible entries
  - 8,500 nonzero counts
  - **3% dense**



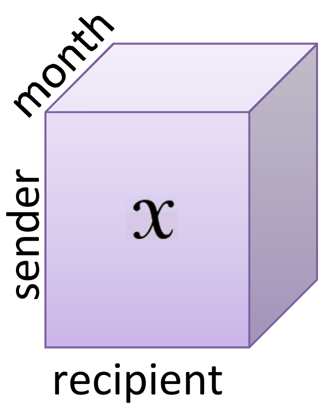
- Questions: What can we learn about this data?
  - Each person labeled by Zhou et al. (2007); see also Owen and Perry (2010)
    - Seniority: 57% senior, 43% junior
    - Gender: 67% male, 33% female
    - Department: 24% legal, 31% trading, 45% other

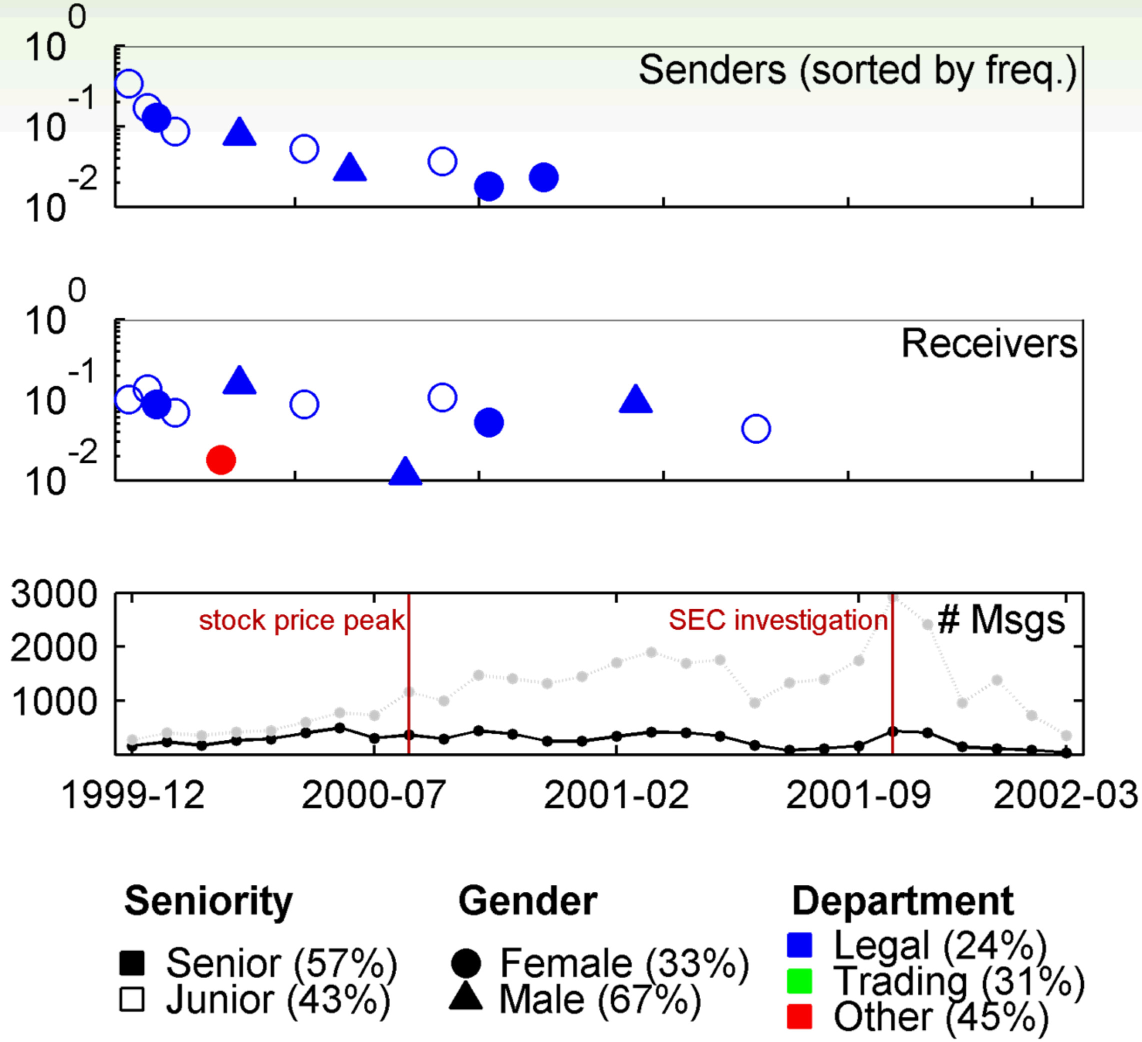
This information is not part of the tensor factorization

Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 1)

Legal Dept;
Mostly Female

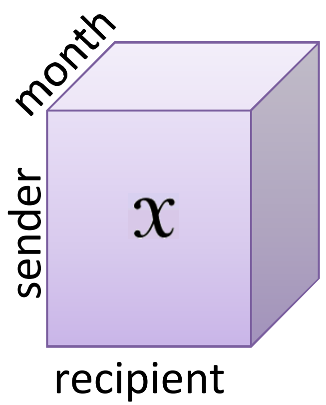


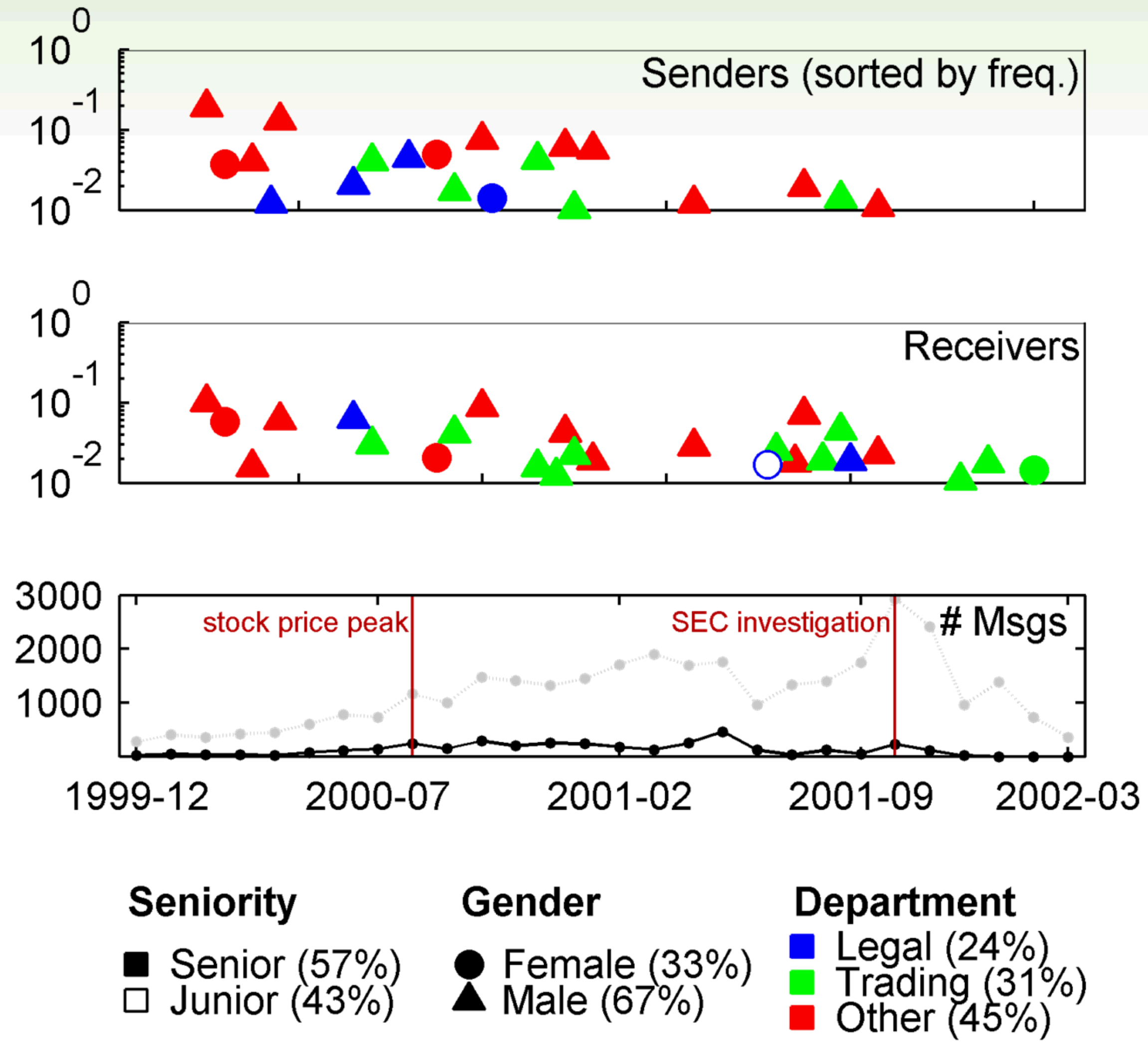


Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 3)

Senior;
Mostly Male

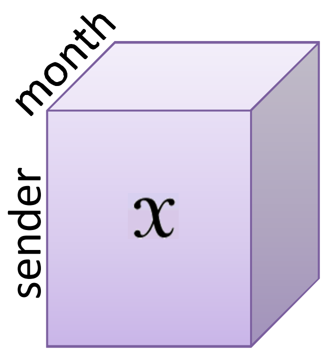


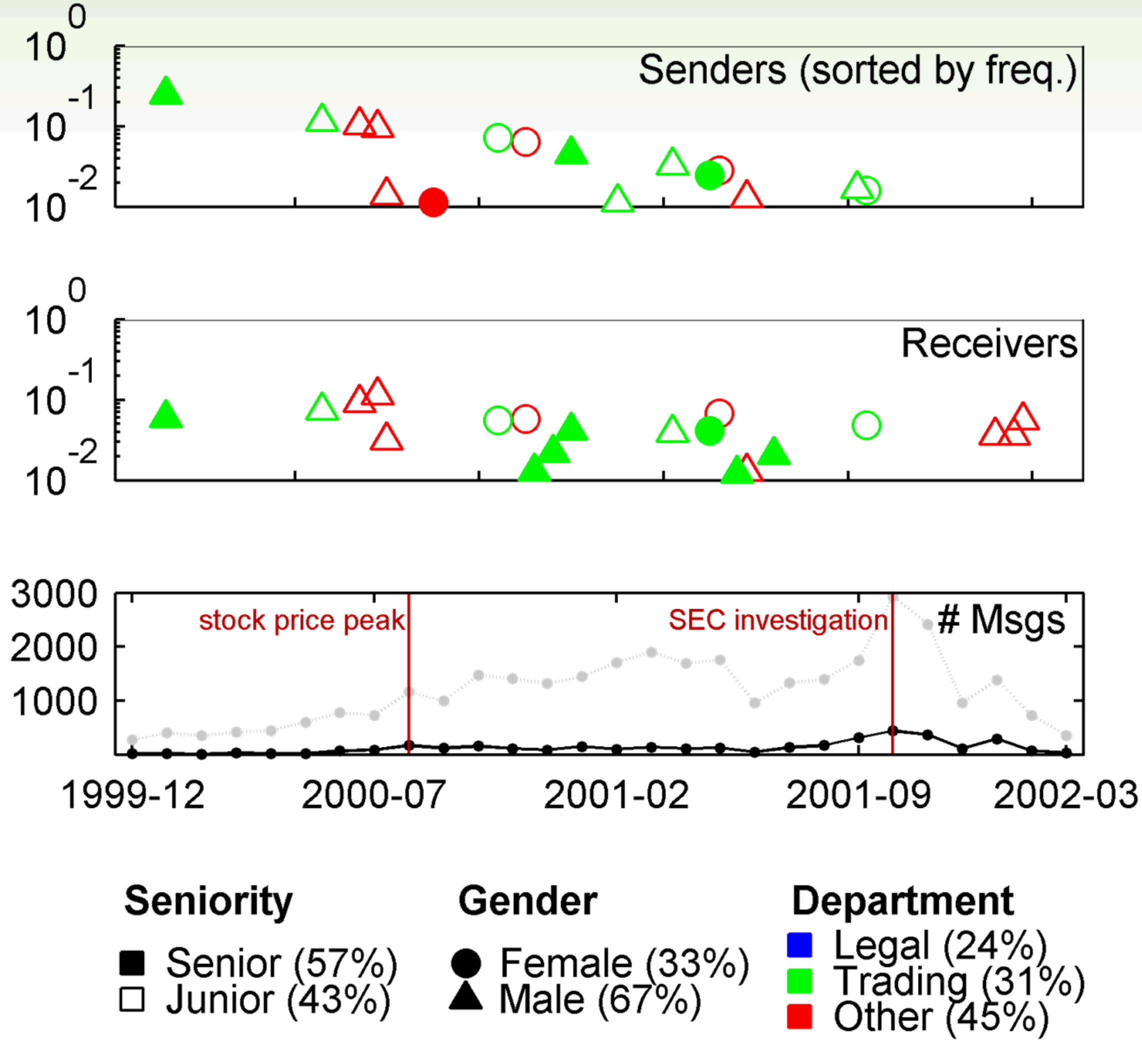


Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

Enron Email Data (Component 4)

Not Legal

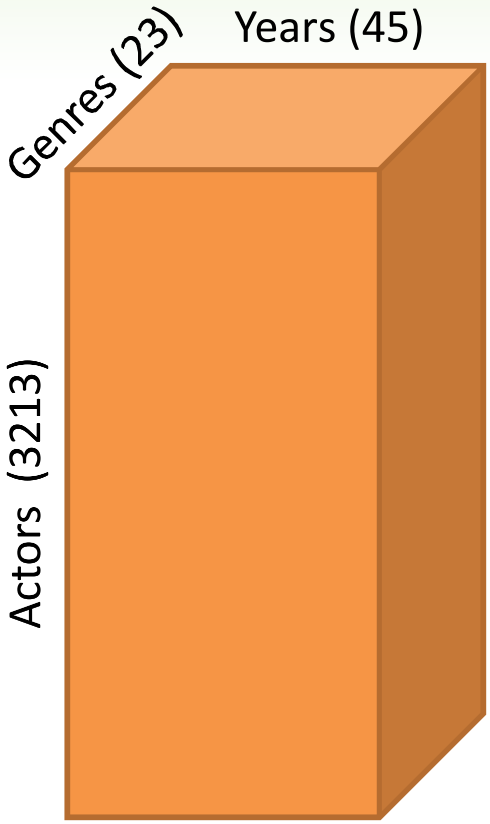




Chi and Kolda. ***On Tensors, Sparsity, and Nonnegative Factorizations***, 2012, http://arxiv.org/abs/1112.2414

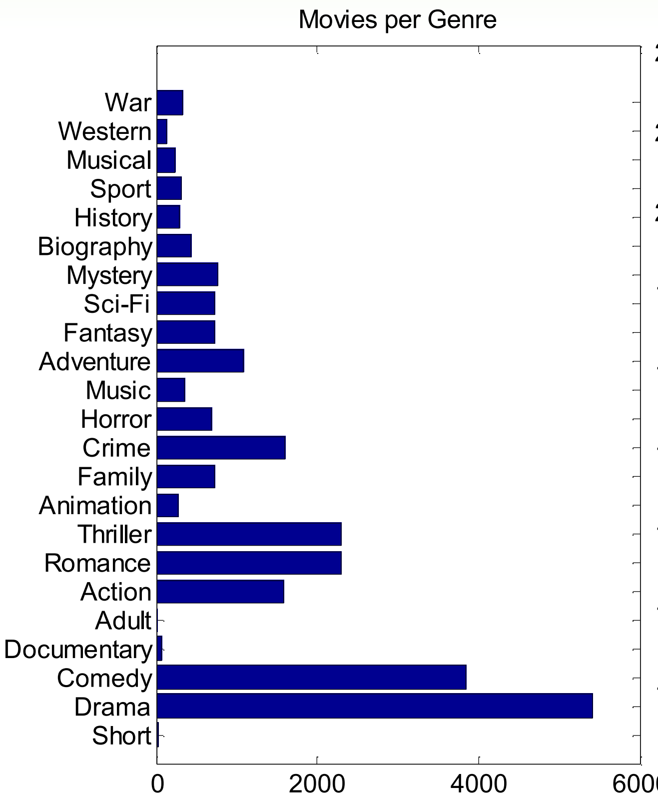
Running Example: Actor x Genre x Year Tensor

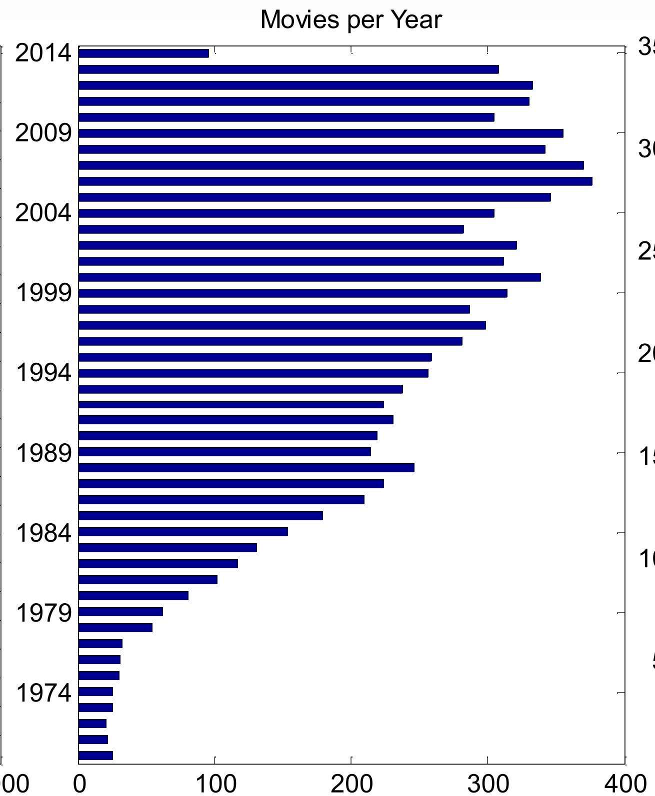
- Entry (i,j,k) = # movies for **actor** i in a movie of **genre** j in **year** k
  - 3213 actors (incl. actresses)
  - 23 genres
  - 45 years (1970-2014)
  - 99361 entries

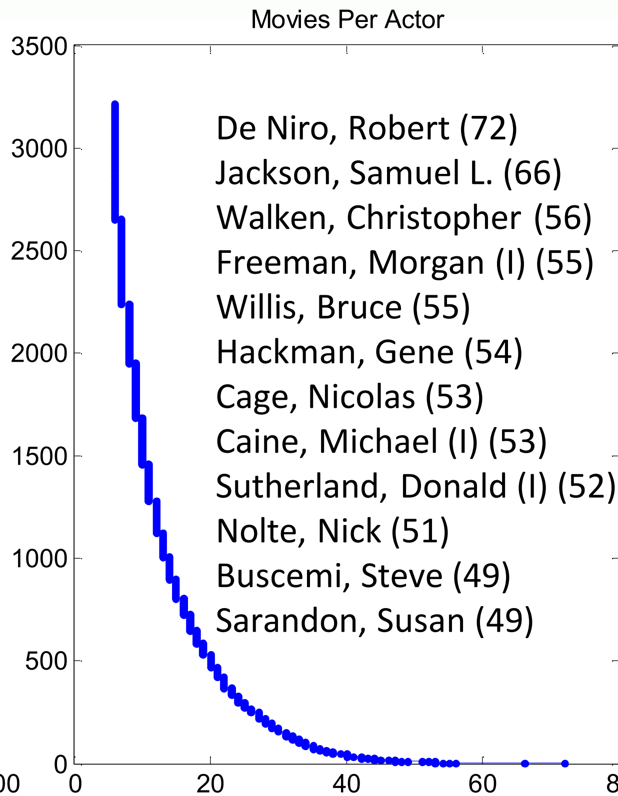


- Data details
  - IMDB data (http://www.imdb.com/interfaces)
  - Data comes from 9273 movies
    - Used movies with *reported* gross revenue in the USA
    - Skipped TV series, TV or video-only movies, and a few other filters
    - Each movie has one or more specified genres (avg. genres per movie = 2.6)
    - Each actor/actress appeared in at least 5 movies during this period (movies per actor = 12.8)

Movie Distributions



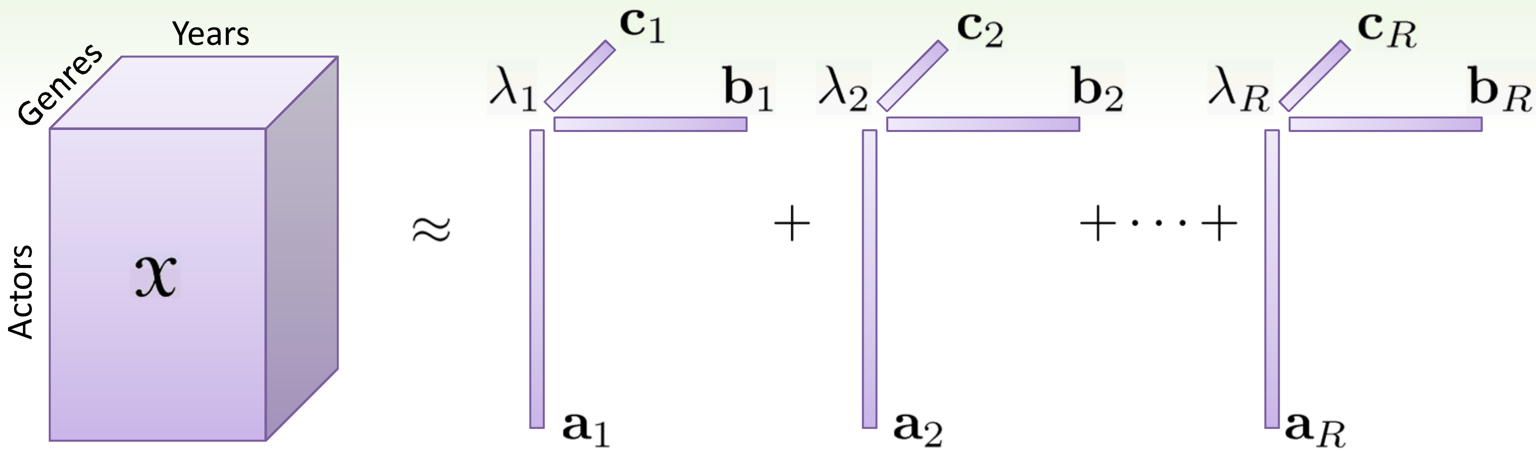




2.6 genres per movie

12.8 movies per actor

Tensor Factorization Interpretation



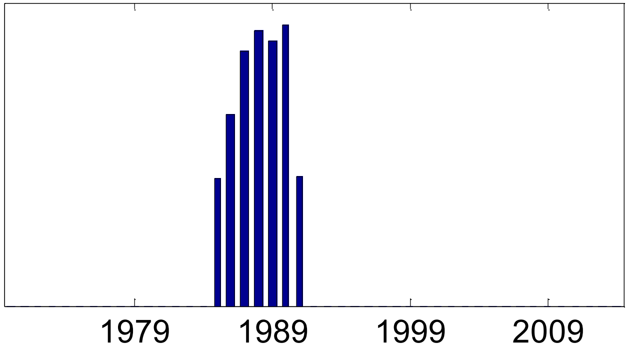
- Each component = **related group** of actors, genres, and years
- Each vector entry is a score between 0 and 1
- We show highest scoring actors, genres, and years for each

Component 1 (weight = 14730)

Top Genres:
- Drama (20.9)
- Comedy (17.4)
- Thriller (10.4)

Top Years:
- 1985 (8.3)
- 1986 (12.5)
- 1987 (16.7)
- 1988 (18.0)
- 1989 (17.5)
- 1990 (18.5)
- 1991 (8.5)

Year Weights



Top Actors:
- **Hackman, Gene (0.3)**
- **Sheen, Charlie (0.3)**
- **Sutherland, Kiefer (0.3)**
- **Candy, John (0.3)**
- **Goodman, John (I) (0.2)**
- **Costner, Kevin (0.2)**
- **Lloyd, Christopher (I) (0.2**)
- Heard, John (I) (0.2)
- Walsh, M. Emmet (0.2)
- Julia, Raul (0.2)
- Aiello, Danny (0.2)
- Walsh, J.T. (0.2)
- **Belushi, James (0.2)**
- Turturro, John (0.2)
- Berenger, Tom (0.2)
- **Neeson, Liam (0.2)**
- **Schwarzenegger, Arnold (0.2)**
- Loggia, Robert (0.2)
- Mantegna, Joe (0.2)
- Spader, James (0.2)

Component 2 (weight = 14503)

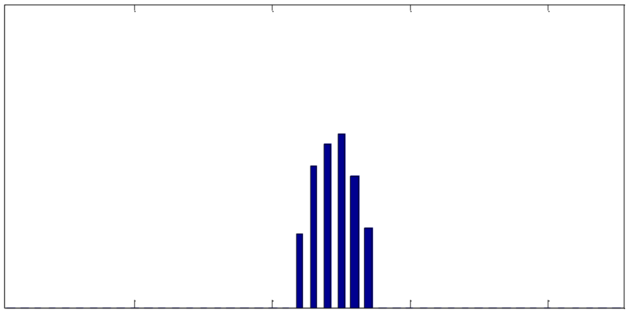
Top Genres:

- Drama (22.4)
- Comedy (17.0)

Top Years:

- 1991 (9.7)
- 1992 (18.5)
- 1993 (21.3)
- 1994 (22.7)
- 1995 (17.3)
- 1996 (10.5)

Year Weights



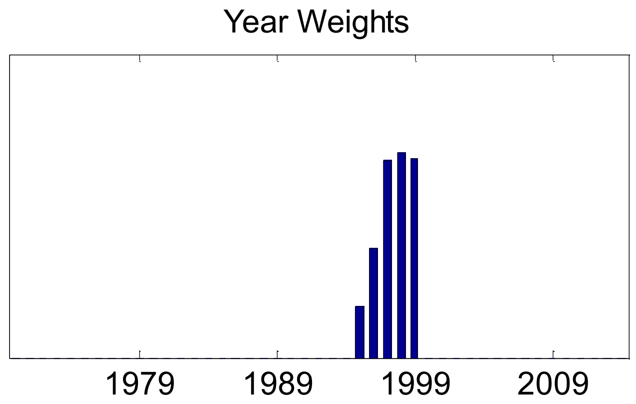
Top Actors:

- **Jackson, Samuel L. (0.3)**
- Keitel, Harvey (0.3)
- Madsen, Michael (I) (0.3)
- Walsh, J.T. (0.2)
- LaPaglia, Anthony (0.2)
- **Williams, Robin (I) (0.2)**
- **Snipes, Wesley (0.2)**
- Henriksen, Lance (0.2)
- **Bullock, Sandra (0.2)**
- **Kilmer, Val (0.2)**
- Curry, Tim (I) (0.2)
- Whaley, Frank (I) (0.2)
- McGinley, John C. (0.2)
- **Willis, Bruce (0.2)**
- Wincott, Michael (0.2)
- Fonda, Bridget (0.2)
- Whitaker, Forest (0.2)
- Mulroney, Dermot (0.2)
- **Estevez, Emilio (0.2)**
- Pollak, Kevin (I) (0.2)

Component 3 (weight = 12732)

Top Genres:
- Drama (26.4)
- Comedy (15.5)
- Romance (12.1)
- Thriller (11.9)

Top Years:
- 1995 (6.7)
- 1996 (14.4)
- 1997 (25.9)
- 1998 (26.9)
- 1999 (26.1)



Top Actors:
- Paymer, David (0.3)
- Cox, Brian (I) (0.3)
- Cromwell, James (I) (0.3)
- Ulrich, Skeet (0.3)
- **Paltrow, Gwyneth (0.3)**
- Voight, Jon (0.2)
- Woods, James (I) (0.2)
- **Willis, Bruce (0.2)**
- Walken, Christopher (0.2)
- Leary, Denis (I) (0.2)
- **Heche, Anne (0.2)**
- **Washington, Denzel (0.2)**
- Wilkinson, Tom (I) (0.2)
- Macy, William H. (0.2)
- Rapaport, Michael (I) (0.2)
- Schreiber, Liev (0.2)
- Morse, David (I) (0.2)
- Keener, Catherine (0.2)
- **Nolte, Nick (0.2)**
- **Hopkins, Anthony (I) (0.2)**

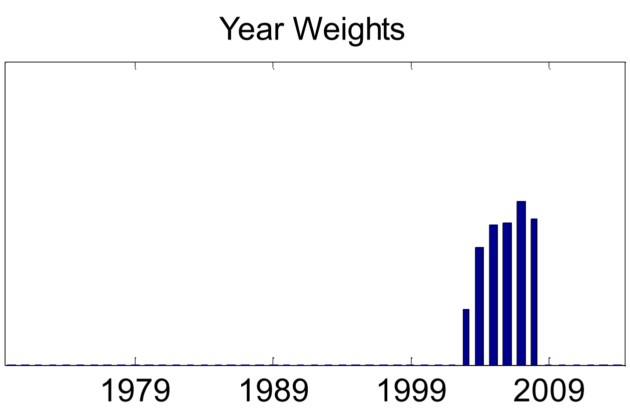
Component 6 (weight = 11206)

Top Genres:

- Drama (21.5)
- Thriller (21.0)
- Crime (12.5)
- Action (11.0)

Top Years:

- 2003 (7.2)
- 2004 (15.5)
- 2005 (18.3)
- 2006 (18.6)
- 2007 (21.3)
- 2008 (19.1)



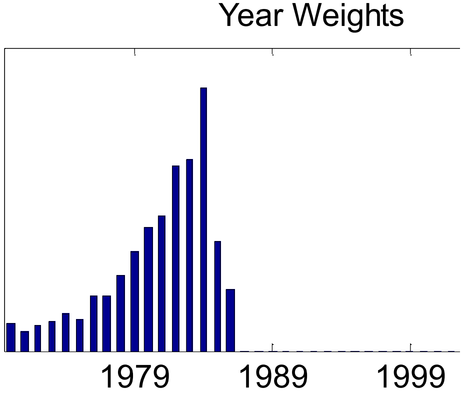
Top Actors:

- **Jackson, Samuel L. (0.4)**
- Craig, Daniel (I) (0.4)
- **Freeman, Morgan (I) (0.3)**
- **Jolie, Angelina (0.3)**
- **Bale, Christian (0.3)**
- **Farrell, Colin (I) (0.3)**
- Howard, Terrence (I) (0.3)
- **Knightley, Keira (0.3)**
- Statham, Jason (0.3)
- Kretschmann, Thomas (0.2)
- **Cage, Nicolas (0.2)**
- **Willis, Bruce (0.2)**
- **Kilmer, Val (0.2)**
- Cox, Brian (I) (0.2)
- Wilkinson, Tom (I) (0.2)
- Wahlberg, Mark (I) (0.2)
- Ribisi, Giovanni (0.2)
- Ejiofor, Chiwetel (0.2)
- Caine, Michael (I) (0.2)
- Franco, James (0.2)

Component 8 (weight = 10538)

Top Genres:
- Drama (20.2)
- Comedy (15.7)

Top Years:
- 1970 (1.8)
- 1971 (1.2)
- 1972 (1.7)
- 1973 (1.9)
- 1974 (2.5)
- 1975 (2.0)
- 1976 (3.7)
- 1977 (3.6)
- 1978 (5.0)
- 1979 (6.5)
- 1980 (8.1)
- 1981 (8.9)
- 1982 (12.2)
- 1983 (12.5)
- 1984 (17.4)
- 1985 (7.1)
- 1986 (4.0)

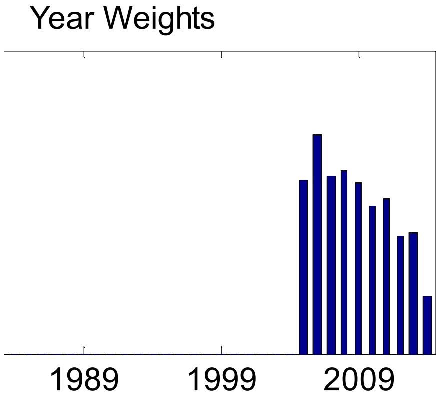


Top Actors:
- **Reynolds, Burt (I) (0.6)**
- **Eastwood, Clint (0.5)**
- Durning, Charles (0.4)
- Bridges, Jeff (I) (0.4)
- **Ford, Harrison (I) (0.4)**
- Walsh, M. Emmet (0.4)
- **Hackman, Gene (0.4)**
- **Norris, Chuck (0.4)**
- Beatty, Ned (0.3)
- **Connery, Sean (0.3)**
- Caine, Michael (I) (0.3)
- von Sydow, Max (I) (0.3)
- Hurt, John (0.3)
- Garfield, Allen (0.3)
- Lauter, Ed (0.3)
- Nicholson, Jack (I) (0.3)
- McMillan, Kenneth (I) (0.3)
- Lewis, Geoffrey (I) (0.3)
- DeLuise, Dom (0.3)
- Oz, Frank (0.3)

Component 9 (weight = 6826)

Top Genres:
- Comedy (25.2)
- Family (20.5)
- Adventure (16.6)
- Fantasy (13.1)
- Animation (11.6)

Top Years:
- 2005 (11.4)
- 2006 (14.4)
- 2007 (11.6)
- 2008 (12.1)
- 2009 (11.3)
- 2010 (9.7)
- 2011 (10.2)
- 2012 (7.6)
- 2013 (7.9)
- 2014 (3.7)



Top Actors:
- Arnett, Will (0.8)
- Nighy, Bill (0.7)
- Cleese, John (0.6)
- **Rogen, Seth (0.6)**
- **Stiller, Ben (0.6)**
- Tatasciore, Fred (I) (0.5)
- Cross, David (II) (0.5)
- Hutcherson, Josh (0.5)
- Long, Justin (I) (0.5)
- Warburton, Patrick (0.5)
- **Wilson, Owen (I) (0.4)**
- Taylor, James Arnold (0.4)
- Hill, Jonah (0.4)
- **Poehler, Amy (0.4)**
- **Carell, Steve (0.4)**
- Faris, Anna (0.4)
- **Cusack, Joan (0.4)**
- Miller, T.J. (0.4)
- Alazraqui, Carlos (0.4)
- McKellen, Ian (0.4)

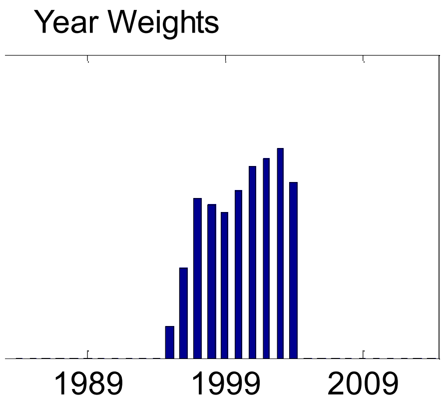
Component 10 (weight = 5724)

Top Genres:

- Comedy (21.9)
- Adventure (21.4)
- Family (18.0)
- Fantasy (13.3)

Top Years:

- 1995 (2.0)
- 1996 (5.9)
- 1997 (10.5)
- 1998 (10.1)
- 1999 (9.5)
- 2000 (11.0)
- 2001 (12.6)
- 2002 (13.1)
- 2003 (13.8)
- 2004 (11.6)

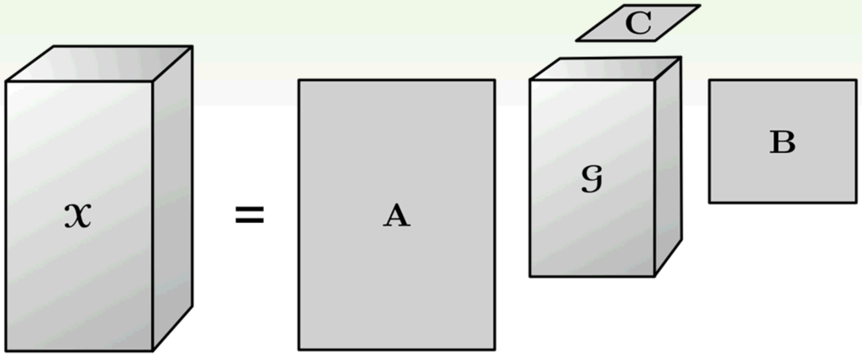


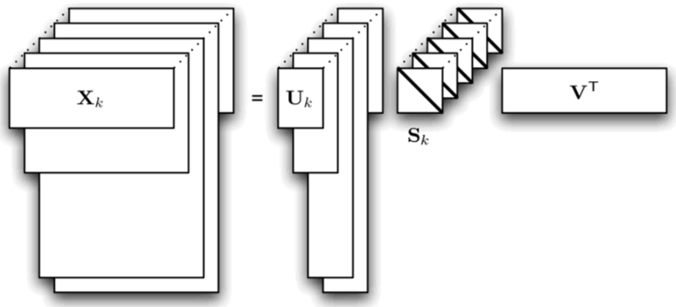
Top Actors:

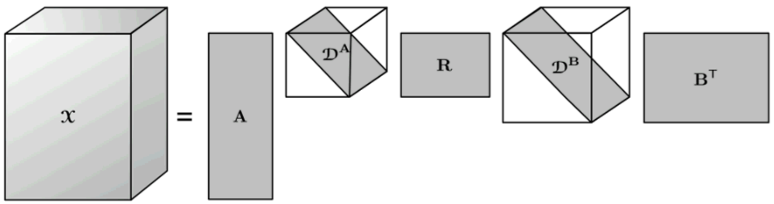
- Cummings, Jim (I) (0.6)
- **Murphy, Eddie (I) (0.6)**
- **Goodman, John (I) (0.5)**
- Daily, Elizabeth (0.5)
- Dunst, Kirsten (0.4)
- Soucie, Kath (0.4)
- Buscemi, Steve (0.4)
- Lane, Nathan (I) (0.4)
- Curry, Tim (I) (0.4)
- **Fox, Michael J. (I) (0.4)**
- Woods, James (I) (0.4)
- Myers, Mike (I) (0.4)
- **Allen, Tim (I) (0.4)**
- Fraser, Brendan (0.4)
- Burton, Corey (I) (0.4)
- Tambor, Jeffrey (0.4)
- Torn, Rip (0.3)
- Idle, Eric (0.3)
- **Carrey, Jim (0.3)**
- **Banderas, Antonio (0.3)**

More on Computing Tensor Factorizations

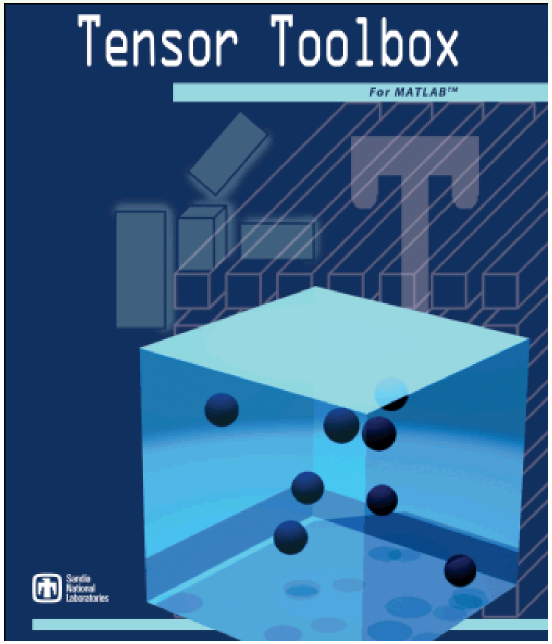
- Other objective functions and constraints
  - Nonnegative least squares (Bro and Jong 1997, Paatero 1997, Welling & Weber 2001)
  - Orthogonal constraints (generally fails)
  - Bayesian tensor factorization
  - Binary tensor factorization
- Computational issues to consider
  - Compression (Bro and Andersson 1998)
  - Sparse tensors (Bader and Kolda 2007)
  - Symmetry (Comon et al. 2008)
  - Missing data (Acar et al. 2011)
- Other types of factorizations
  - Tucker (Tucker 1966) and Higher-order SVD (De Lathauwer 1997)
  - INDSCAL (Carroll & Chang 1972)
  - PARAFAC2 (Harshman 1978)
  - DEDICOM (Harshman & Lundy 1996)
  - Hierarchical SVD (Grasedyck 2010)
  - Tensor Train (Oseledets 2011)







Tensor Software



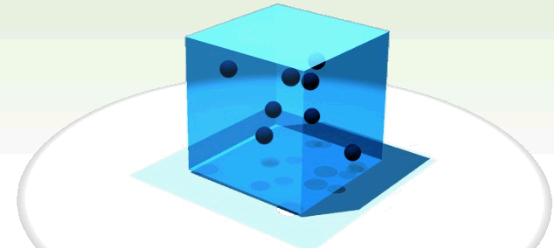
Tensor Toolbox for MATLAB
Bader, Kolda, Acar, Dunlavy, and others

- MATLAB
  - N-way Toolbox (Andersen and Bro, Univ. Copenhagen)
    - The forerunner of all today's software
  - Tensor Toolbox for MATLAB (Bader and Kolda, Sandia)
    - Key unique capability: handles sparse tensors
  - TensorLab (Sorber et al., KU Leuven)
    - Key unique capability: handles complex data

Sparse Tensors

- Sparse if majority of entries (x_{ijk}) are zero
- Some storage options
  - Each two-dimensional slice stored as sparse matrix
  - Unfold and store as sparse matrix
  - Coordinate format
- Storage for sptensor
  - P = # nonzeros
  - subs = P x 3 matrix of subscripts
  - vals = P x 1 vector of values
- Optimized calculations
  - Sparse tensor times vector(s) keeps being reinvented!

Bader & Kolda, SISC, 2007



2 x 2 x 2 Tensor with P = 4 Nonzeros

$$x_{111} = 1.5$$
$$x_{121} = 2.7$$
$$x_{212} = 3.3$$
$$x_{222} = 8.5$$

$$\texttt{subs} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 1 & 2 \\ 2 & 2 & 2 \end{bmatrix} \quad \texttt{vals} = \begin{bmatrix} 1.5 \\ 2.7 \\ 3.3 \\ 8.5 \end{bmatrix}$$

Conclusions & Future Work

- CANDECOMP/PARAFAC
  - Decomposes tensor into sum of rank-1 tensors (i.e., outer products)
  - Typically computed via alternating least squares
  - Poisson Tensor Factorization (PTF) instead uses KL divergence objective function
- Applications include
  - Missing data for EEG brain analysis
  - Enron email analysis
  - Actor-genre-time correlations
- Developing new methods for statistical rank prediction
  - Based on cross-validation
- Computations with sparse tensors
  - Fast and efficient methods have been developed

Other Topics

- Need BLAS for tensor computations
  - Plus parallel methods
  - Plus more methods for sparse data
- Better computational algorithms
  - Extremely difficult non-convex optimization problem
- Tensor Eigenpairs
  - Polynomial optimization methods
- Tucker decomposition is a useful method for compression
  - Related methods include Tucker Train and Tensor Quantization
  - Need fast and space efficient computational methods

Tamara G. Kolda
tgkolda@sandia.gov
http://www.sandia.gov/~tgkolda