Sandia National Laboratories

SAND2011-8480P

Nonnegative Tensor Factorizations for Sparse Count Data

Tamara G. Kolda
Sandia National Labs

Eric C. Chi
Rice University/UCLA
DOE CSGF

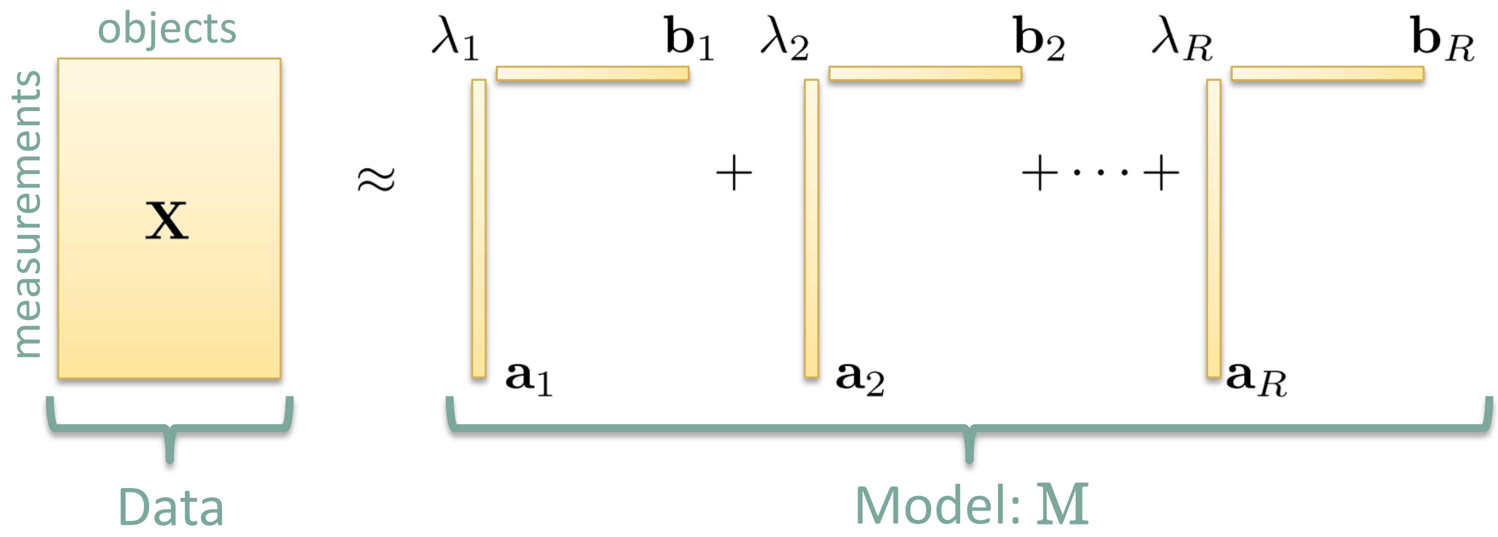
ASCR

U.S. Department of Energy
Office of Advanced Scientific Computing Research

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.

Factorizations for Data Analysis

$$\mathbf{X} \approx \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r$$



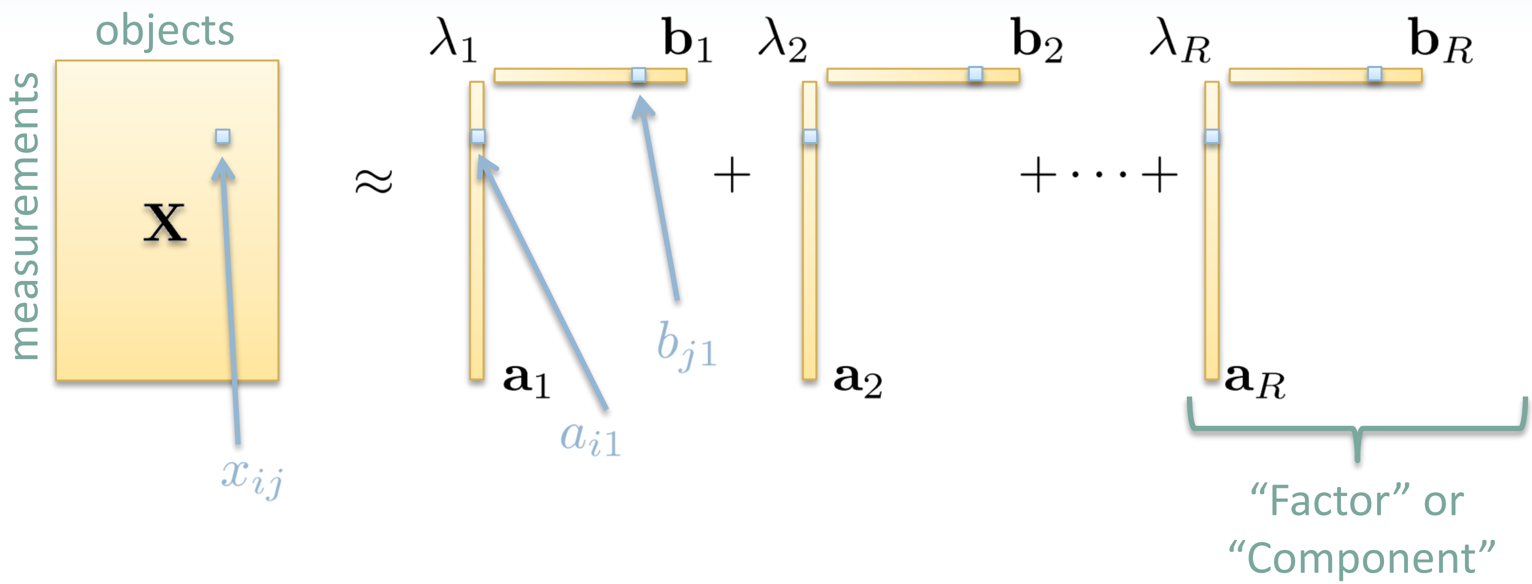
λ_r = weight of rth component

$\mathbf{a}_r$ = mode-1 factor ("principal component"), assumed to be scaled to norm 1

$\mathbf{b}_r$ = mode-2 factor ("loading"), assumed to be scaled to norm 1

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

Weighted Combination of Factors



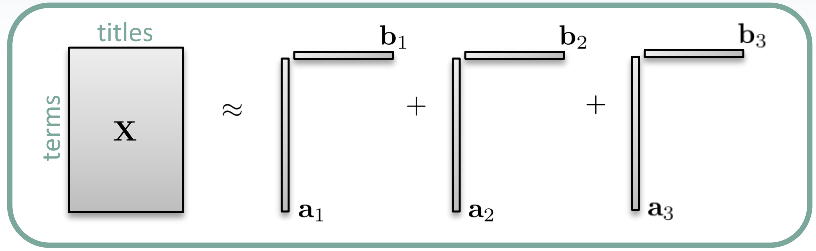
Description of jth object (i.e., jth column): $\mathbf{x}_j \approx \sum_r \gamma_r \, \mathbf{a}_r, \quad \gamma_r \equiv \lambda_r \, b_{jr}$

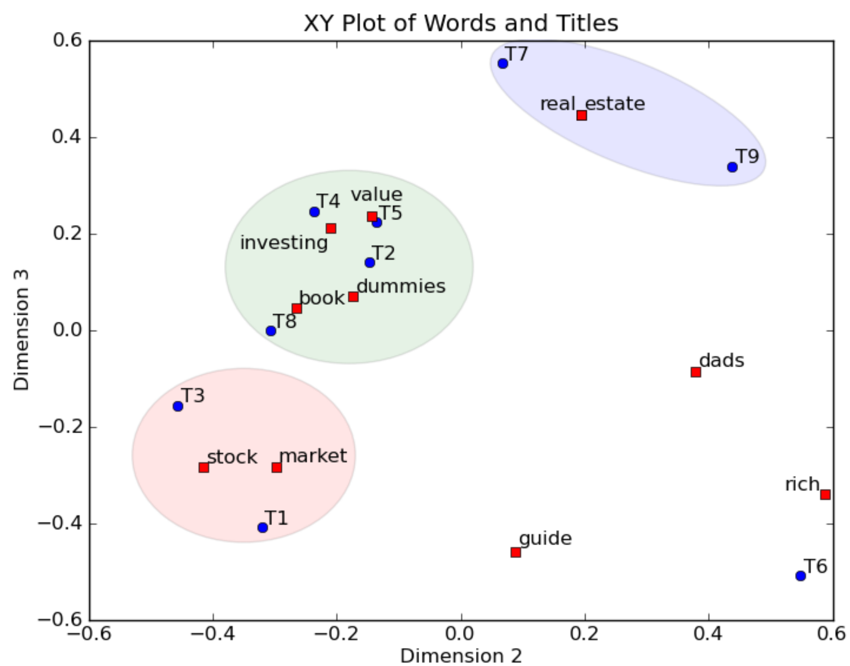
Description of a single data element: $m_{ij} \approx \sum_r \lambda_r \, a_{ir} \, b_{jr}$

Latent Semantic Analysis of Term-Document Matrices

Book Titles

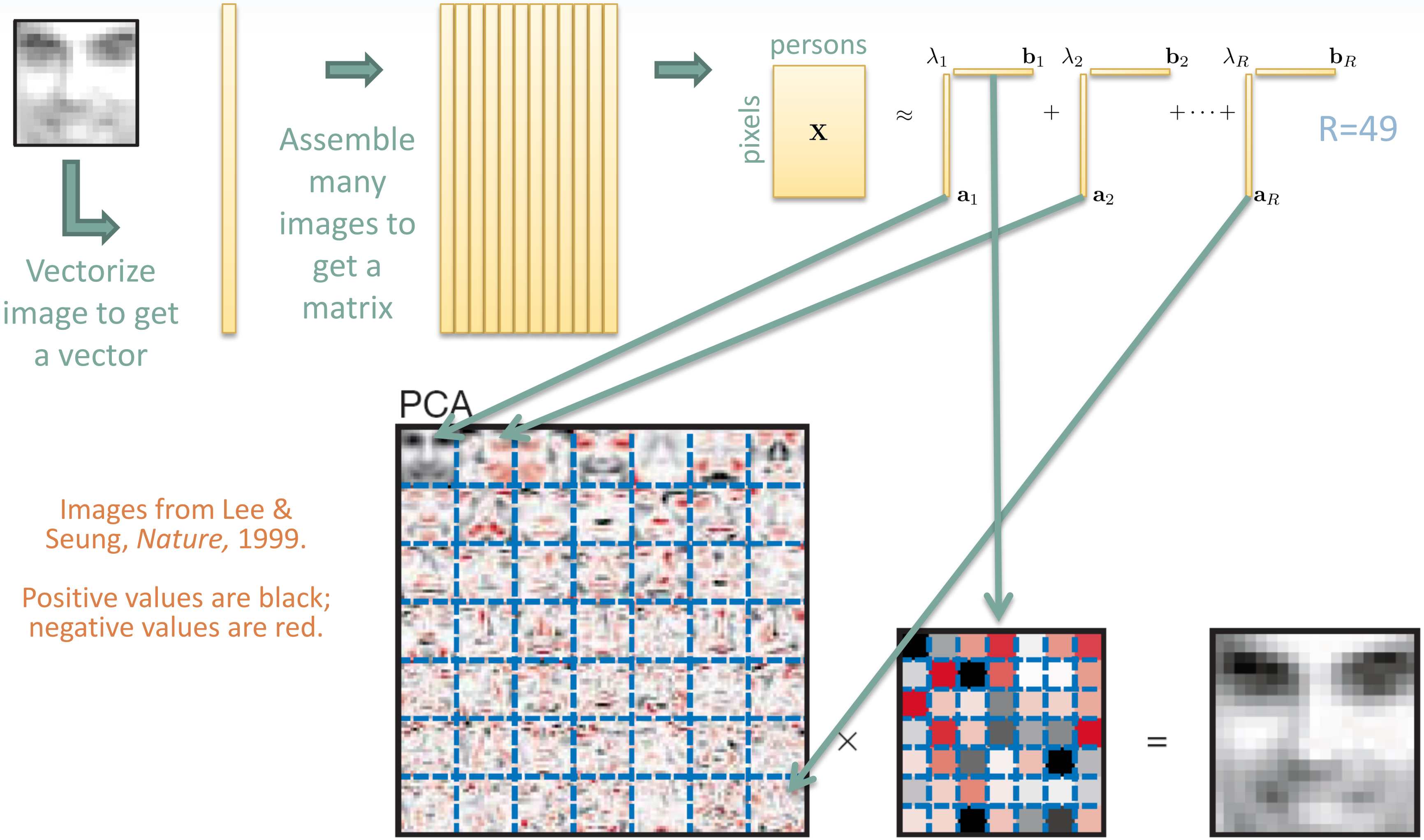
1. The Neatest Little Guide to Stock Market Investing
2. Investing For Dummies, 4th Edition
3. The Little Book of Common Sense Investing: The Only Way to Guarantee Your Fair Share of Stock Market Returns
4. The Little Book of Value Investing
5. Value Investing: From Graham to Buffett and Beyond
6. Rich Dad's Guide to Investing: What the Rich Invest in, That the Poor and the Middle Class Do Not!
7. Investing in Real Estate, 5th Edition
8. Stock Investing For Dummies
9. Rich Dad's Advisors: The ABC's of Real Estate Investing: The Secrets of Finding Hidden Profits Most Investors Miss





See full details at Latent Semantic Analysis (LSA) Tutorial at http://www.puffinwarellc.com/index.php/news-and-articles/articles/33.html

Facial Image Decomposition for Compression and Analysis

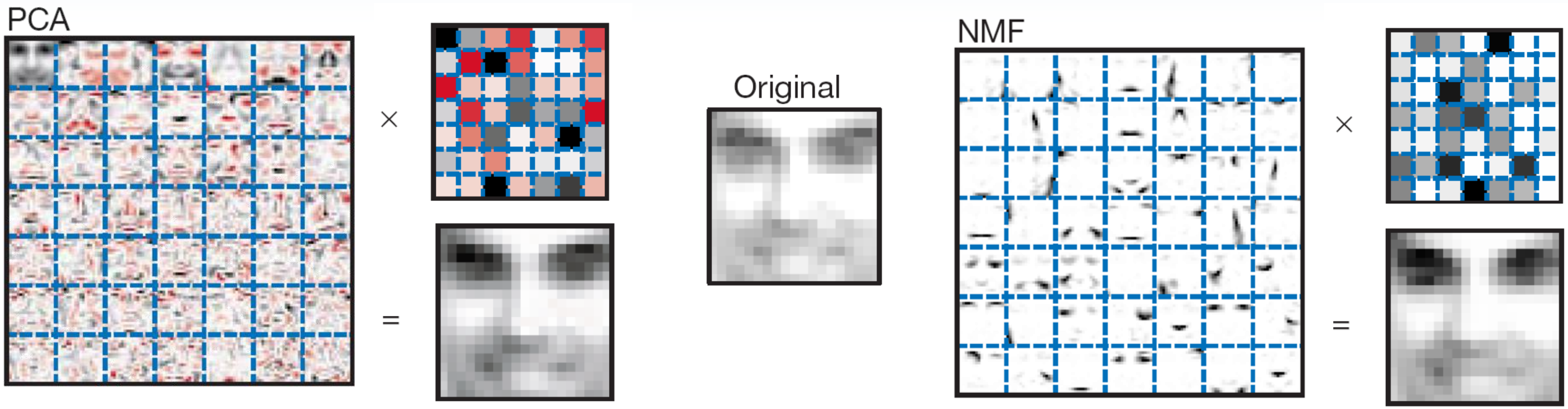


Images from Lee & Seung, *Nature,* 1999.

Positive values are black; negative values are red.

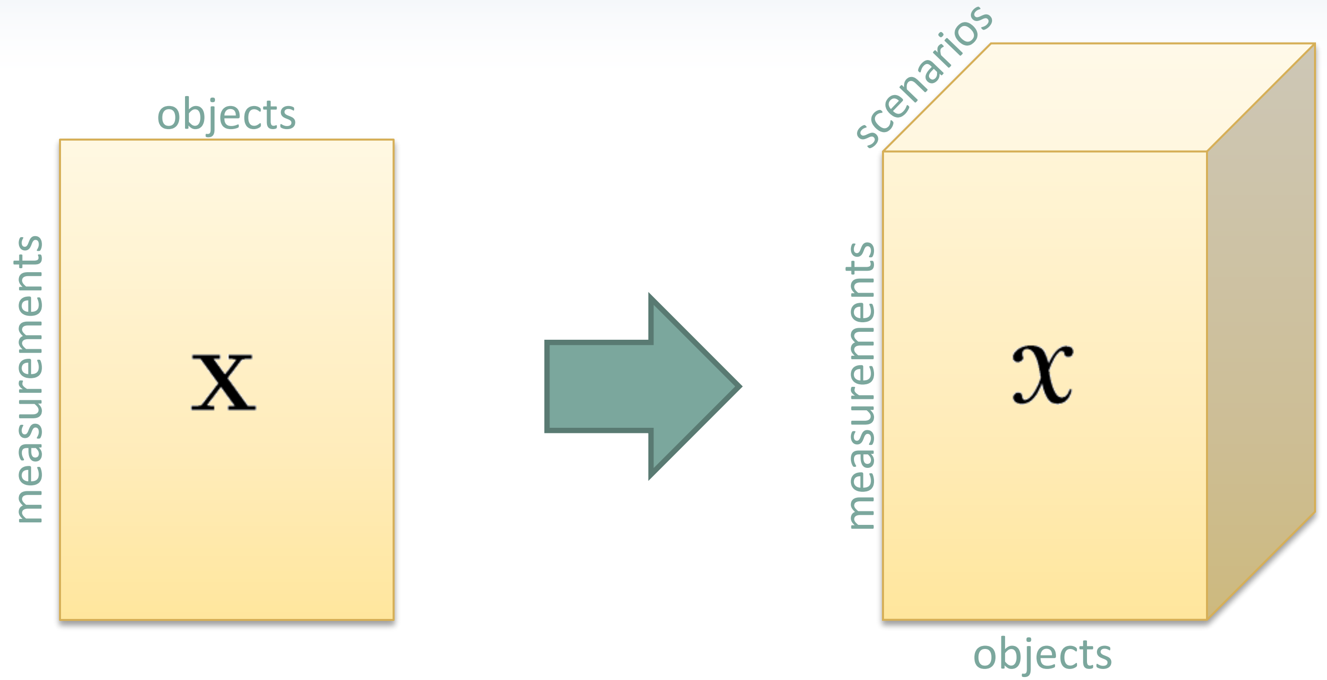
Many Different Two-Way Models

Images from Lee & Seung, *Nature,* 1999



- Singular Value Decomposition (SVD) and Principal Components Analysis (PCA)
  - *Factors are required to be orthogonal*
- Independent Component Analysis (ICA) [e.g., Comon, 1994]
  - *Factors are required to be maximally independent*
- Compressive Sensing and related work [Candes, 2006]
  - *Sparse factors*
- Nonnegative Matrix Factorization [Paatero, 1997; Bro & De Jong, 1997; Lee & Seung, 2001]
  - *Nonnegative factors*
  - *Alternative assumptions on distribution*

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

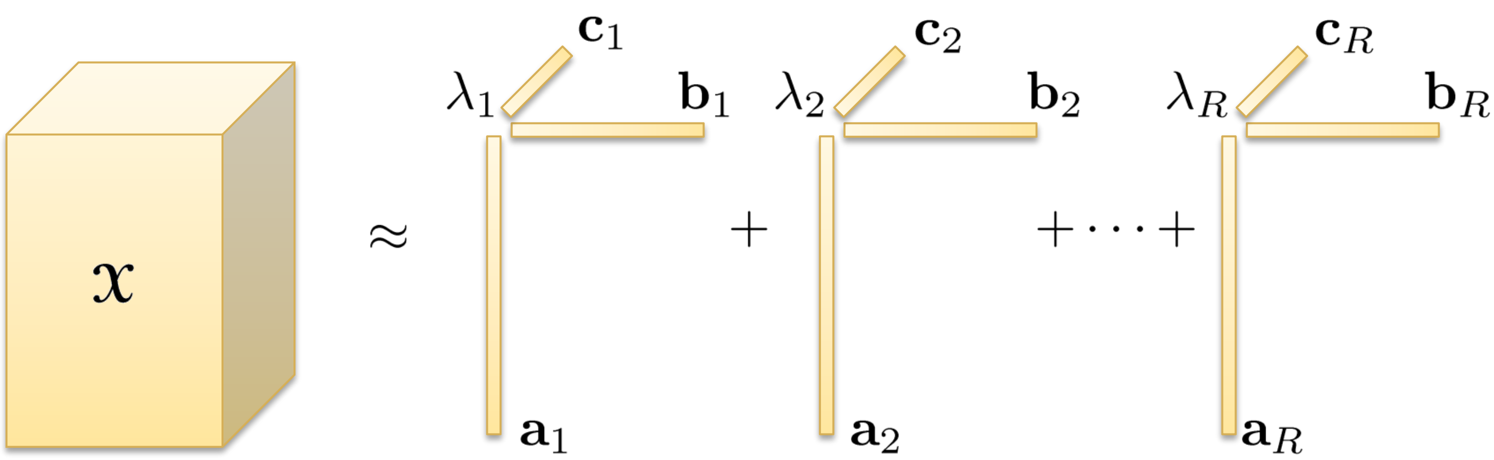
RAYMOND B. CATTELL

University of Illinois

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

Multi-way Factorizations for Analysis

CANDECOMP/PARAFAC (CP) Model



Data

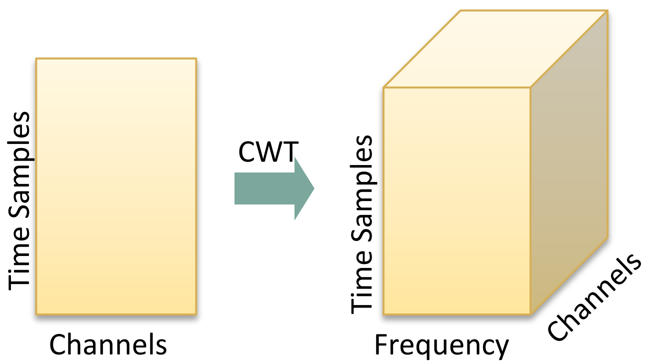
Model: $\mathcal{M} = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$

$$x_{ijk} \approx m_{ijk} = \sum_r \lambda_r \, a_{ir} \, b_{jr} \, c_{kr}$$

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

Factor Example: Epilepsy

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

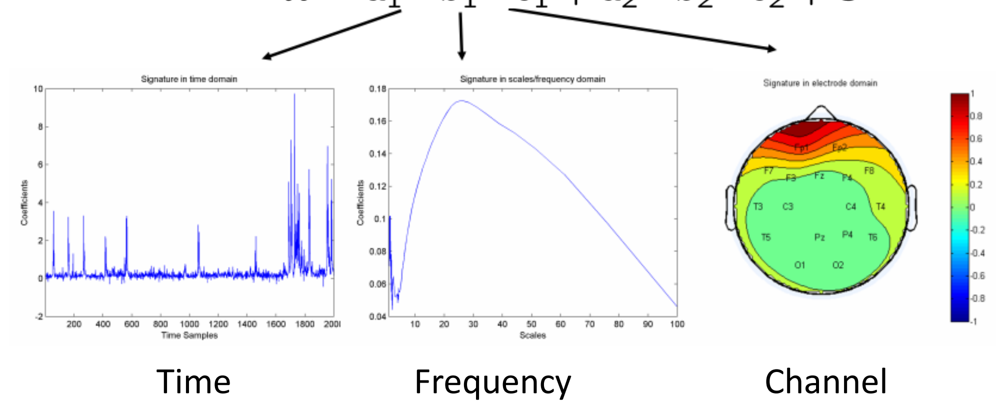


$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \mathcal{E}$$

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

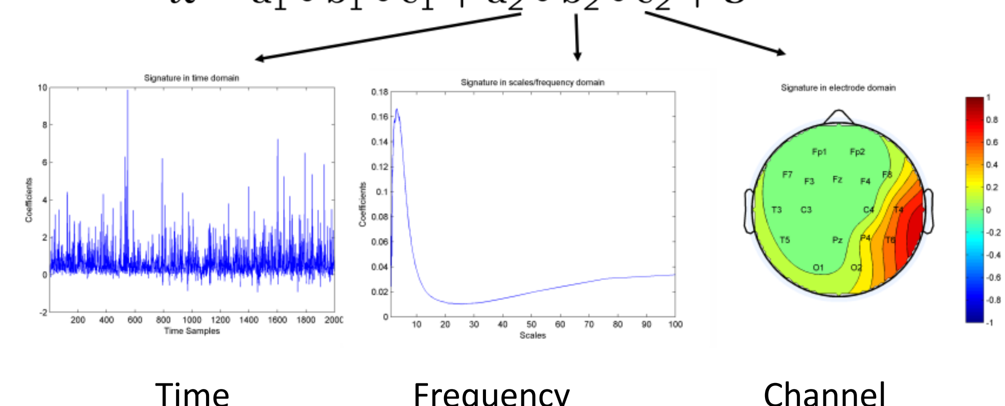
Eye Artifact

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



Seizure

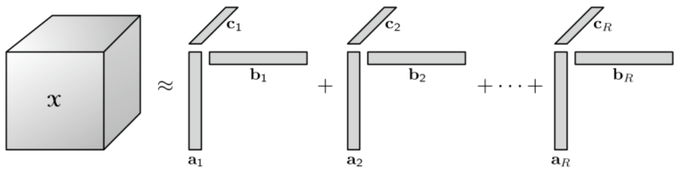
$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \mathcal{E}$$



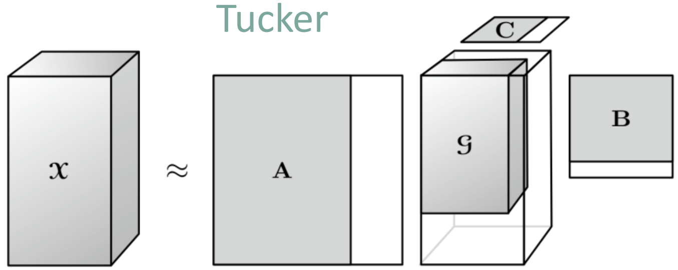
Many Tensor Decompositions, Methods, Software, etc.

K. & Bader, Tensor Decompositions and Applications, SIAM Review, 2009

CANDECOMP/PARAFAC (CP)



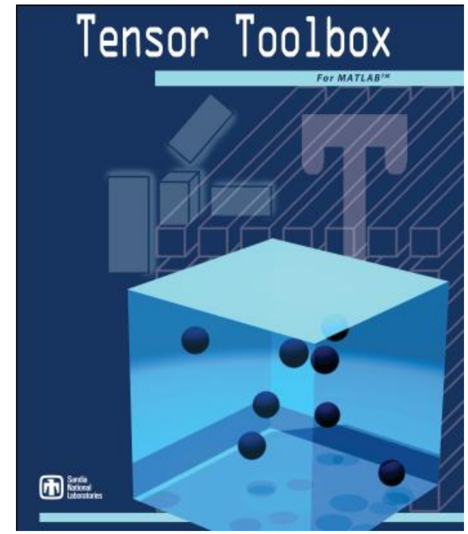
Tucker



See also past work in

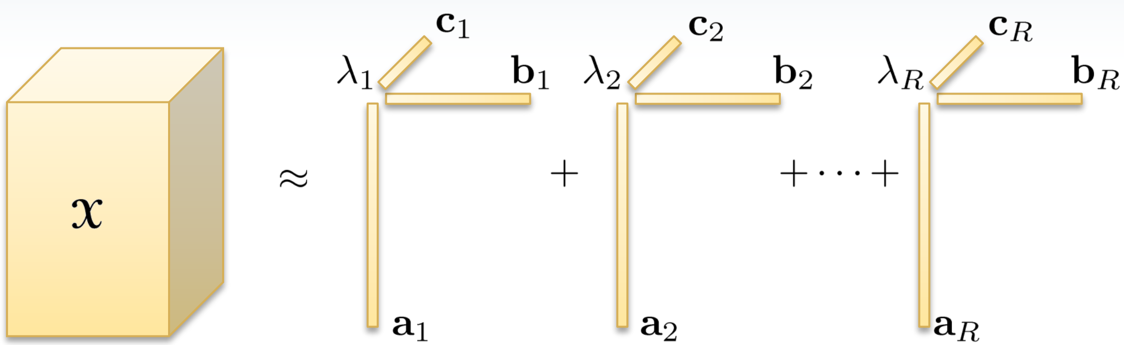
- Sparse computations
- Model fitting
- Missing data
- Applications to graphs

http://www.sandia.gov/~tgkolda/



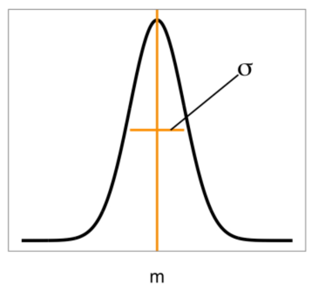
Tensor Toolbox for MATLAB
Bader & Kolda
plus
Acar, Dunlavy, Sun, et al.

But what does "≈" mean?



- 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:

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

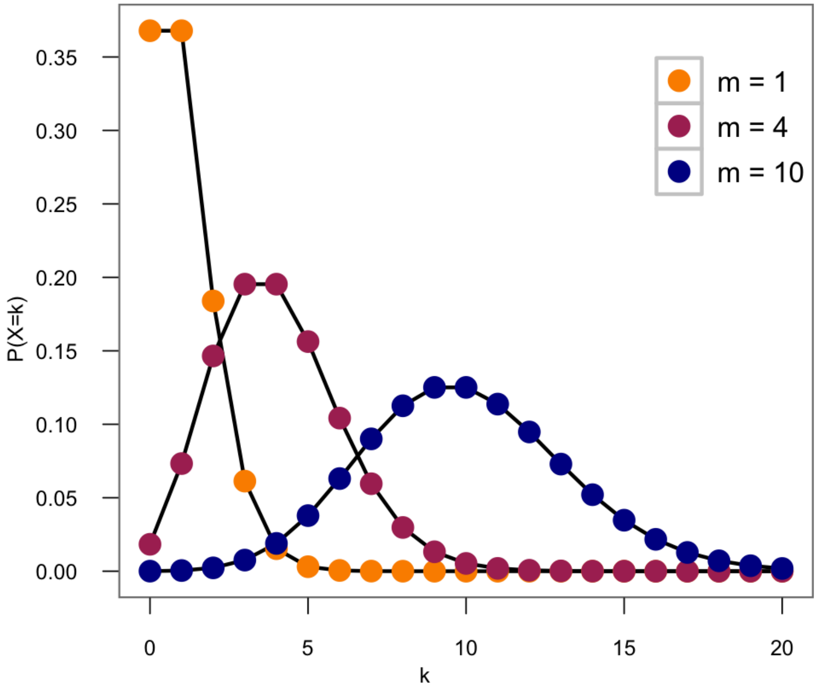
By monotonicity of log, same as maximizing this:

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

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
  - Non-symmetric errors (e.g., data that grows exponentially)
  - Data with outliers or multiple modes
  - Count data with many low counts
    - High counts can be reasonably approximated by a Gaussian!

Poisson Distribution



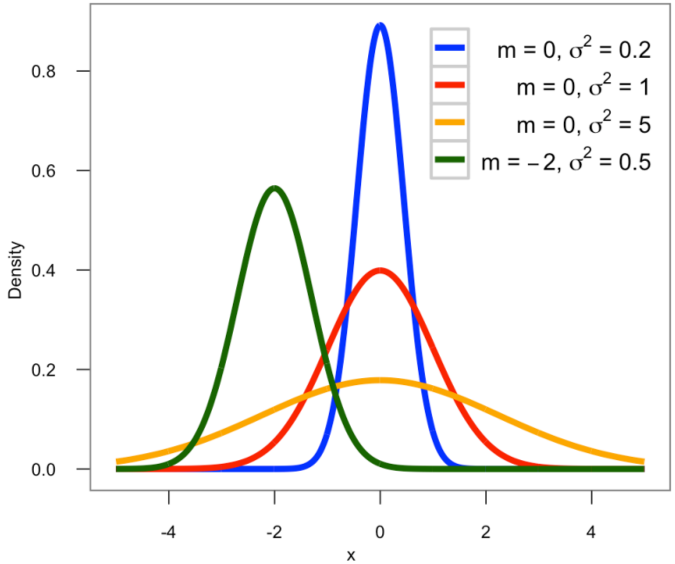
Poisson for Sparse Count Data

Gaussian (typical)

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

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

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

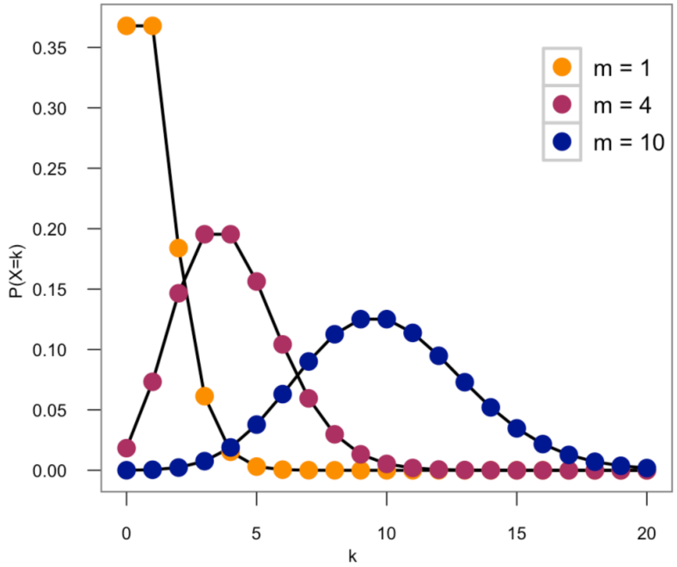


Poisson

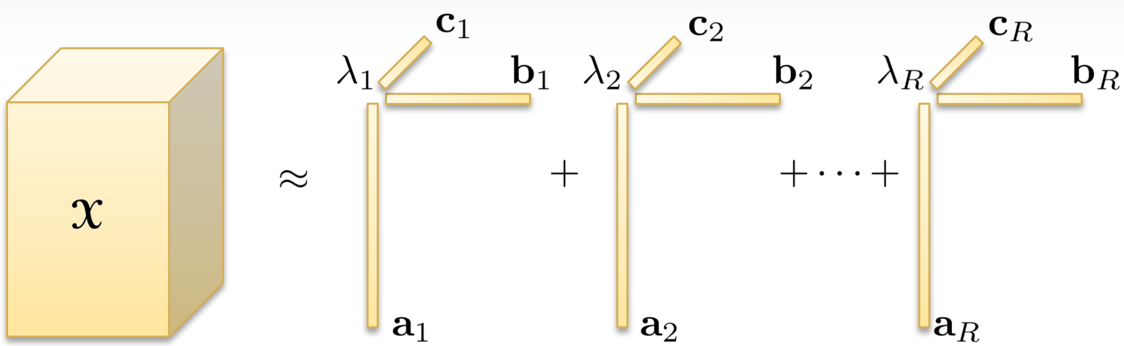
The random variable x is a discrete nonnegative integer.

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

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



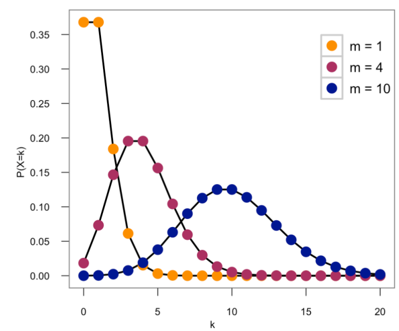
Poisson Tensor Factorization



- Poisson preferred for sparse count data
- Automatically nonnegative
- More difficult objective function than least squares

$$x_{ijk} \sim \text{Poisson}(m_{ijk})$$

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



Maximize this:

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

By monotonicity of log, same as maximizing this:

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

Sparse Count Data Abounds

- Computer network traffic
  - User visits to websites
  - IP x IP x Port communications
  - Packet routing
  - Computer logins
- Communications
  - Email traffic
  - Social network interactions
- Financial
  - Purchase records
  - Bank transfers
  - Credit card transactions
- Bibliometric data
  - Co-authorship
  - Author x Term

- *Any of the above binned into time intervals*

How do we make sense of this data?

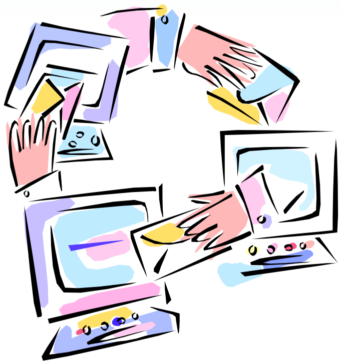
Can we find patterns of behavior?

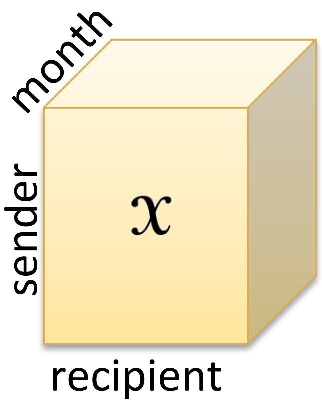
Can we spot anomalies?

Can we predict future behavior?

Motivating Example: Enron Email

- Emails from Enron FERC investigation
  - Zhou et al., 2007 version
  - 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
  - 0.03% dense





Fitting a Poisson Factorization

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

$$\text{subject to } \mathcal{M} = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$$\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C} \geq 0$$

$$\|\mathbf{a}_r\|_1 = 1, \|\mathbf{b}_r\|_1 = 1, \|\mathbf{c}_r\|_1 = 1 \quad \forall r$$

Assumption
x log m = 0
if x=0 & m=0

- We will solve for each factor matrix in turn, using a **Gauss-Seidel** (or Alternating Optimization) approach
- We can rewrite the model by absorbing the weights $\boldsymbol{\lambda}$ into one of the factor matrices, e.g.,

$$\mathcal{M} = \sum_r \bar{\mathbf{a}}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \text{ with } \bar{\mathbf{A}} = \mathbf{A} \cdot \text{diag}(\boldsymbol{\lambda})$$

- Matrix $\bar{\mathbf{A}}$ is only constrained by be nonnegative
- This can be done for any of the three factor matrices

New Method: Alternating Poisson Regression (CP-APR)

Sandia National Laboratories

Repeat until converged…

1. $\bar{\mathbf{A}} \leftarrow \arg\min_{\bar{\mathbf{A}} \geq 0} \sum_{ijk} m_{ijk} - x_{ijk} \log m_{ijk}$ s.t. $\mathcal{M} = \sum_r \bar{\mathbf{a}}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$
2. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{A}}$; $\mathbf{A} \leftarrow \bar{\mathbf{A}} \cdot \operatorname{diag}(1/\boldsymbol{\lambda})$

Fix $\mathrm{B,C}$; solve for A

3. $\bar{\mathbf{B}} \leftarrow \arg\min_{\bar{\mathbf{B}} \geq 0} \sum_{ijk} m_{ijk} - x_{ijk} \log m_{ijk}$ s.t. $\mathcal{M} = \sum_r \mathbf{a}_r \circ \bar{\mathbf{b}}_r \circ \mathbf{c}_r$
4. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{B}}$; $\mathbf{B} \leftarrow \bar{\mathbf{B}} \cdot \operatorname{diag}(1/\boldsymbol{\lambda})$

Fix $\mathrm{A,C}$; solve for B

5. $\bar{\mathbf{C}} \leftarrow \arg\min_{\bar{\mathbf{C}} \geq 0} \sum_{ijk} m_{ijk} - x_{ijk} \log m_{ijk}$ s.t. $\mathcal{M} = \sum_r \mathbf{a}_r \circ \mathbf{b}_r \circ \bar{\mathbf{c}}_r$
6. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{C}}$; $\mathbf{C} \leftarrow \bar{\mathbf{C}} \cdot \operatorname{diag}(1/\boldsymbol{\lambda})$

Fix $\mathrm{A,B}$; solve for C

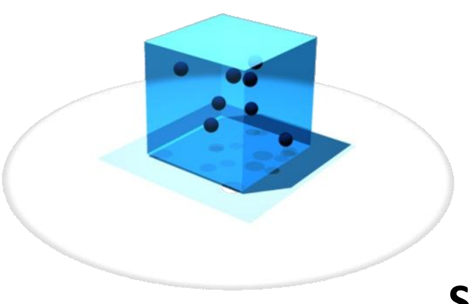
Convergence Theory

Theorem: The CP-APR algorithm will **converge to a constrained stationary point** if the subproblems are strictly convex and solved exactly at each iteration.

Solving the Subproblem

$$\min_{\bar{\mathbf{A}} \geq 0} \sum_{ijk} m_{ijk} - x_{ijk} \log m_{ijk} \text{ s.t. } \mathcal{M} = \sum_r \bar{\mathbf{a}}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$$\min_{\bar{\mathbf{A}} \geq 0} \sum_{ijk} \left(\sum_r \bar{a}_{ir} b_{jr} c_{kr} \right) - x_{ijk} \log \left(\sum_r \bar{a}_{ir} b_{jr} c_{kr} \right)$$

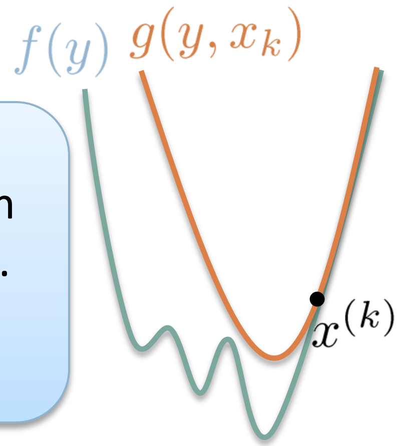


Lemma: The subproblems are **strictly convex** if no columns of the factor matrices go to zero and the data tensor has a sufficient number of reasonably distributed nonzeros.

Sufficient number: Must have an average of at least R nonzeros per slice to compute an R-component factorization.

Majorization-Minimization for Subproblem

$$f(\bar{\mathbf{A}}) = \sum_{ijk} \left(\sum_r \bar{a}_{ir} b_{jr} c_{kr} \right) - x_{ijk} \log \left(\sum_r \bar{a}_{ir} b_{jr} c_{kr} \right)$$



A function $g(y,x)$ **majorizes** $f(x)$ if $g(y,x) \geq f(y)$ for all y and $g(x,x) = f(x)$. Majorization-minimization (MM) algorithms minimize a majorizing function at the current iterate, set that minimizer to be the next iterate, and repeat.

$$x^{(k+1)} = \arg\min_y g(y, x^{(k)})$$

Insight: Easy to minimize majorizer moves sum outside

$$g(\mathbf{A}; \bar{\mathbf{A}}) = \sum_{rijk} a_{ir} b_{jr} c_{kr} - \alpha_{rijk} \, x_{ijk} \log \left(\frac{a_{ir} b_{jr} c_{kr}}{\alpha_{rijk}} \right), \quad \alpha_{rijk} = \frac{\bar{a}_{ir} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}}$$

$$\bar{a}_{ir} \leftarrow \bar{a}_{ir} \phi_{ir}, \quad \phi_{ir} = \sum_{jk} \frac{x_{ijk} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}} \geq 0$$

$$\bar{\mathbf{A}} = \bar{\mathbf{A}} * \mathbf{\Phi}$$

Elementwise Multiplication

MM Subproblem Algorithm

$$f(\bar{\mathbf{A}}) = \sum_{ijk} \left(\sum_r \bar{a}_{ir} b_{jr} c_{kr} \right) - x_{ijk} \log \left(\sum_r \bar{a}_{ir} b_{jr} c_{kr} \right)$$

Repeat until convergence:

$$\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * \mathbf{\Phi} \quad \text{where} \quad \phi_{ir} = \sum_{jk} \frac{x_{ijk} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}}$$

Constrained Optimality (KKT) Conditions

$$\bar{\mathbf{A}} \geq 0$$

← automatically guaranteed

$$\nabla f(\bar{\mathbf{A}}) = \mathbf{E} - \mathbf{\Phi} \geq 0$$

$$\bar{\mathbf{A}} * (\mathbf{E} - \mathbf{\Phi}) = 0$$

Convergence criterion: $|\min(\bar{\mathbf{A}}, \mathbf{E} - \mathbf{\Phi})| \leq \text{tol}$

Elementwise ↗

Novel Algorithm: CP-APR with MM Subproblem Solver

Repeat until converged…

1. Repeat until converged: $\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * \mathbf{\Phi}$, where $\phi_{ir} = \sum_{jk} \frac{x_{ijk} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}}$ — Fix B,C; solve for A
2. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{A}}$; $\mathbf{A} \leftarrow \bar{\mathbf{A}} \cdot \mathrm{diag}(1/\boldsymbol{\lambda})$
3. Repeat until converged: $\bar{\mathbf{B}} \leftarrow \bar{\mathbf{B}} * \mathbf{\Phi}$, where $\phi_{jr} = \sum_{ik} \frac{x_{ijk} a_{ir} c_{kr}}{\sum_{r'} a_{ir'} \bar{b}_{jr'} c_{kr'}}$ — Fix A,C; solve for B
4. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{B}}$; $\mathbf{B} \leftarrow \bar{\mathbf{B}} \cdot \mathrm{diag}(1/\boldsymbol{\lambda})$
5. Repeat until converged: $\bar{\mathbf{C}} \leftarrow \bar{\mathbf{C}} * \mathbf{\Phi}$, where $\phi_{kr} = \sum_{ij} \frac{x_{ijk} a_{ir} b_{jr}}{\sum_{r'} a_{ir'} b_{jr'} \bar{c}_{kr'}}$ — Fix A,B; solve for C
6. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{C}}$; $\mathbf{C} \leftarrow \bar{\mathbf{C}} \cdot \mathrm{diag}(1/\boldsymbol{\lambda})$

Lee-Seung is a Special Case of CP-APR

Lee & Seung, 1999 [matrix version]; Welling & Weber, 2001 [tensor extension]

Repeat until converged…

1. Update matrix as: $\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * \mathbf{\Phi}$, where $\phi_{ir} = \sum_{jk} \frac{x_{ijk} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}}$ — Fix B,C; update A
2. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{A}}$; $\mathbf{A} \leftarrow \bar{\mathbf{A}} \cdot \mathrm{diag}(1/\boldsymbol{\lambda})$
3. Update matrix as: $\bar{\mathbf{B}} \leftarrow \bar{\mathbf{B}} * \mathbf{\Phi}$, where $\phi_{jr} = \sum_{ik} \frac{x_{ijk} a_{ir} c_{kr}}{\sum_{r'} a_{ir'} \bar{b}_{jr'} c_{kr'}}$ — Fix A,C; update B
4. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{B}}$; $\mathbf{B} \leftarrow \bar{\mathbf{B}} \cdot \mathrm{diag}(1/\boldsymbol{\lambda})$
5. Update matrix as: $\bar{\mathbf{C}} \leftarrow \bar{\mathbf{C}} * \mathbf{\Phi}$, where $\phi_{kr} = \sum_{ij} \frac{x_{ijk} a_{ir} b_{jr}}{\sum_{r'} a_{ir'} b_{jr'} \bar{c}_{kr'}}$ — Fix A,B; update C
6. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{C}}$; $\mathbf{C} \leftarrow \bar{\mathbf{C}} \cdot \mathrm{diag}(1/\boldsymbol{\lambda})$

New Insight: How to Fix "Undesirable" Zeros

Zeros never change with multiplicative updates!

$$\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * \mathbf{\Phi} \quad \text{where} \quad \phi_{ir} = \sum_{jk} \frac{x_{ijk} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}}$$

Recall the Constrained Optimality (KKT) Conditions:

$$\bar{\mathbf{A}} \geq 0$$

← automatically guaranteed

$$\nabla f(\bar{\mathbf{A}}) = \mathbf{E} - \mathbf{\Phi} \geq 0$$

$$\bar{\mathbf{A}} * (\mathbf{E} - \mathbf{\Phi}) = 0$$

These conditions enable us to check for "inadmissible" zeros

Undesirable Zero: $\bar{a}_{ij} = 0$ and $\phi_{ij} > 1$

Fix: If $\bar{a}_{ij}$ is close to zero and $\phi_{ij} > 1$, then bump $\bar{a}_{ij}$ from zero, i,e, set $\bar{a}_{ij} = 0.2$,

Fixes Lee-Seung updates too!

See example of problem in Gonzalez & Zhang, 2005

Our Code Supports Sparse Computations

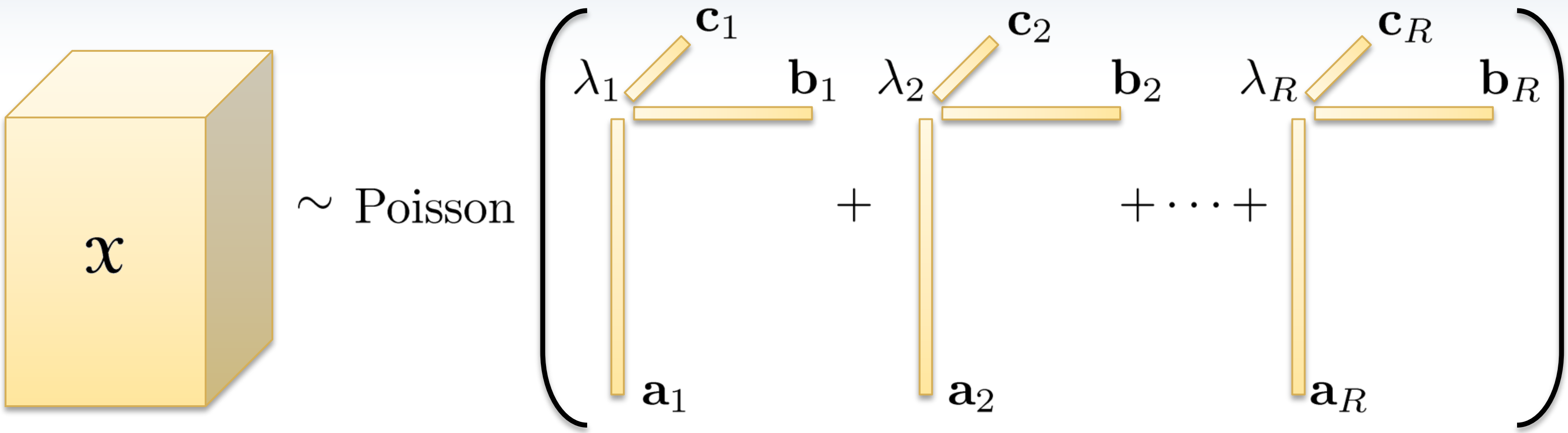
Too big to form product.
Only form entries that correspond to existing data.

Very sparse

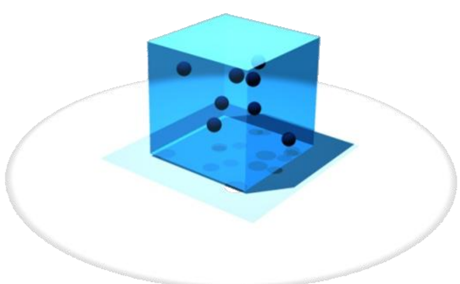
$$\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * \mathbf{\Phi} \text{ where } \phi_{ir} = \sum_{jk} \frac{x_{ijk} b_{jr} c_{kr}}{\sum_{r'} \bar{a}_{ir'} b_{jr'} c_{kr'}}$$

Current estimate of model (dense).
Only form entries that correspond to existing data.

Generating Test Data

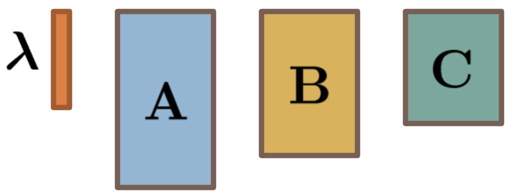


- Each "occurrence" generated as follows
- Choose factor r proportional to $\boldsymbol{\lambda}$
- Given factor r:
  - Choose index i proportional to $\mathbf{a}_r$
  - Choose index j proportional to $\mathbf{b}_r$
  - Choose index k proportional to $\mathbf{c}_r$
- Increment x_{ijk} by one

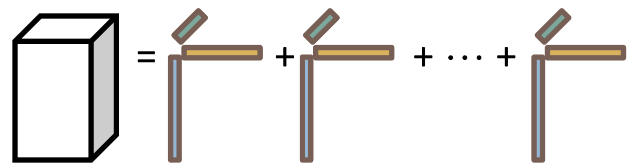


Numerical Experiment Set-up for Simulated Data

Step 1: Generate factor matrices with R=10 columns. Choose 10% entries from U(0,100) and remainder from U(0,1). Renormalize so that each column sums to one. Choose λ entries from U(0,1).

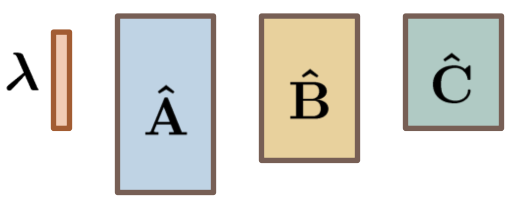


Step 2: Generate sparse tensor from Poisson distribution using model



$$\mathcal{X} \sim \text{Poisson}\left(\sum_r \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right)$$

Step 3: Factorize spares tensor using CP-APR



Step 4: Compute **FMS** = factor match score of computed factors against truth. Assume columns are two-norm normalized and ξ_r is the product of the norms

$$\frac{1}{R} \sum_r \left(1 - \frac{|\xi_r - \hat{\xi}_r|}{\max(\xi_r, \hat{\xi}_r)} \right) \mathbf{a}_r^{\mathsf{T}} \hat{\mathbf{a}}_r \, \mathbf{b}_r^{\mathsf{T}} \hat{\mathbf{b}}_r \, \mathbf{c}_r^{\mathsf{T}} \hat{\mathbf{c}}_r$$

Best FMS is 1

Sandia National Laboratories

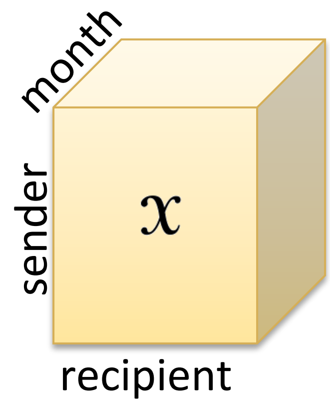
Accuracy is High For Very Sparse Data

Data: 1000 x 800 x 600 Tensor with R=10 Components
CP-APR: Max Iterations = 200, Max Inner Iterations = 30 (10 per mode), Tol = 1e-4 (KKT)
CP-ALS: Max Iterations = 200, Tol = 1e-8 (change in fit)

Nonzeros	Poisson Regression FMS	Gaussian Regression FMS
480,000 (.100%)	0.99	0.57
240,000 (.050%)	0.81	0.49
48,000 (.010%)	0.77	0.47
24,000 (.005%)	0.74	0.46

Motivating Example: Enron Email

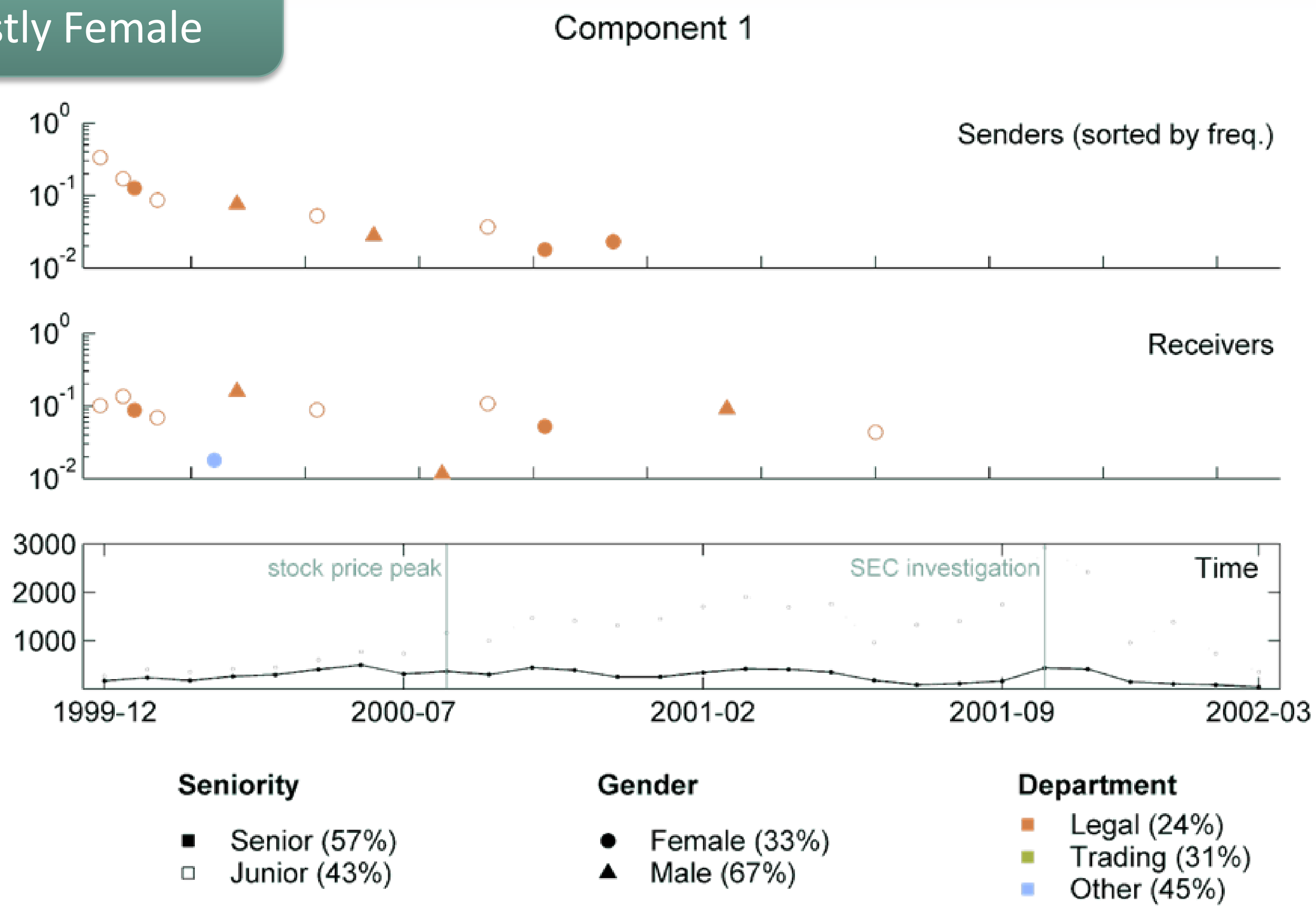
- Emails from Enron FERC investigation
  - 8540 Messages
  - 28 Months (from Dec 1999 to Mar 2002)
  - 105 People (sent and received at least one email every month)
  - x_{ijk} = # emails from sender i to recipient j in month k
  - 105 x 105 x 28 = 308,700 possible entries
  - 8,500 nonzero counts
  - **0.03% 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

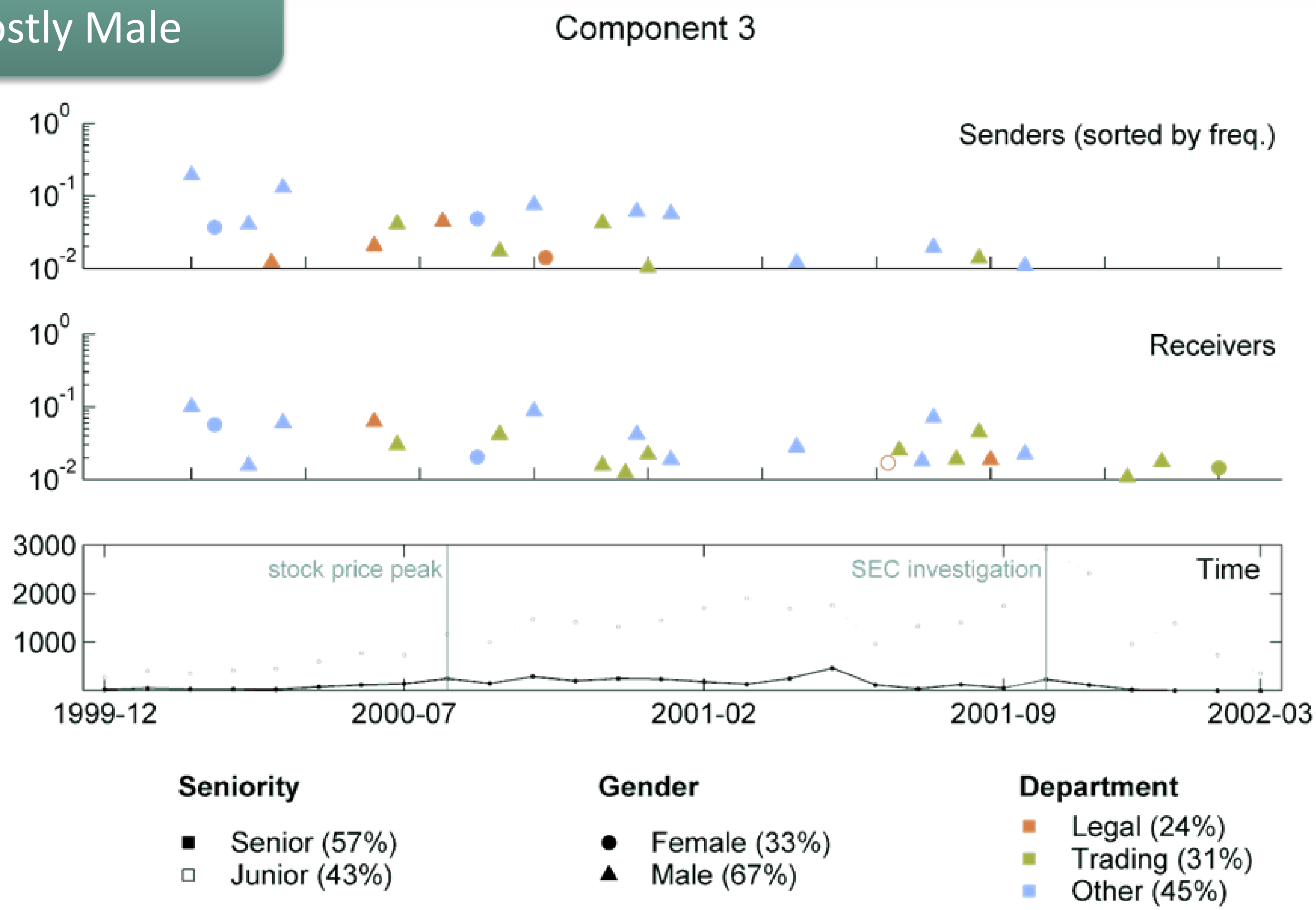
Enron Email Data

Legal Dept;
Mostly Female



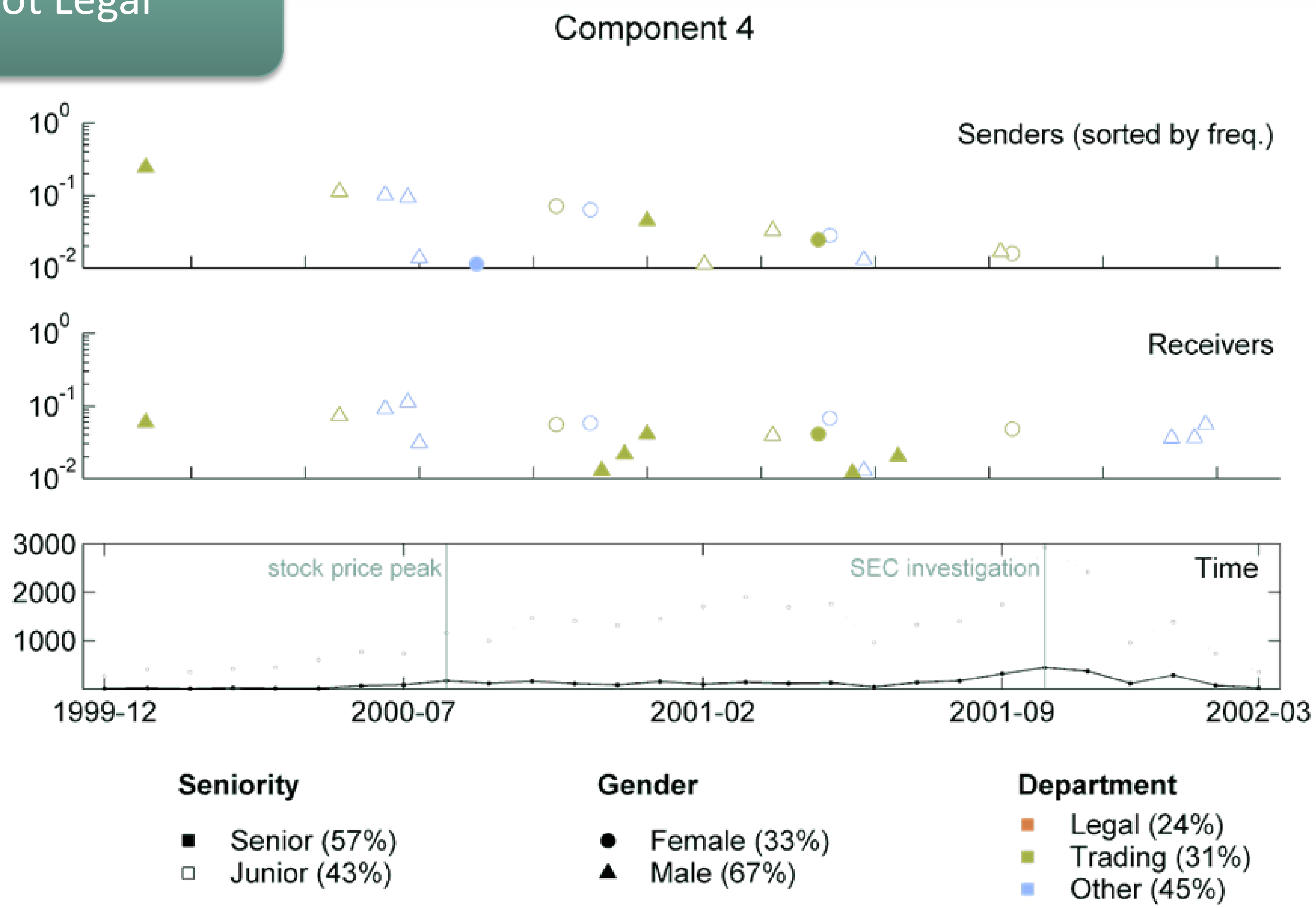
Enron Email Data

Senior;
Mostly Male



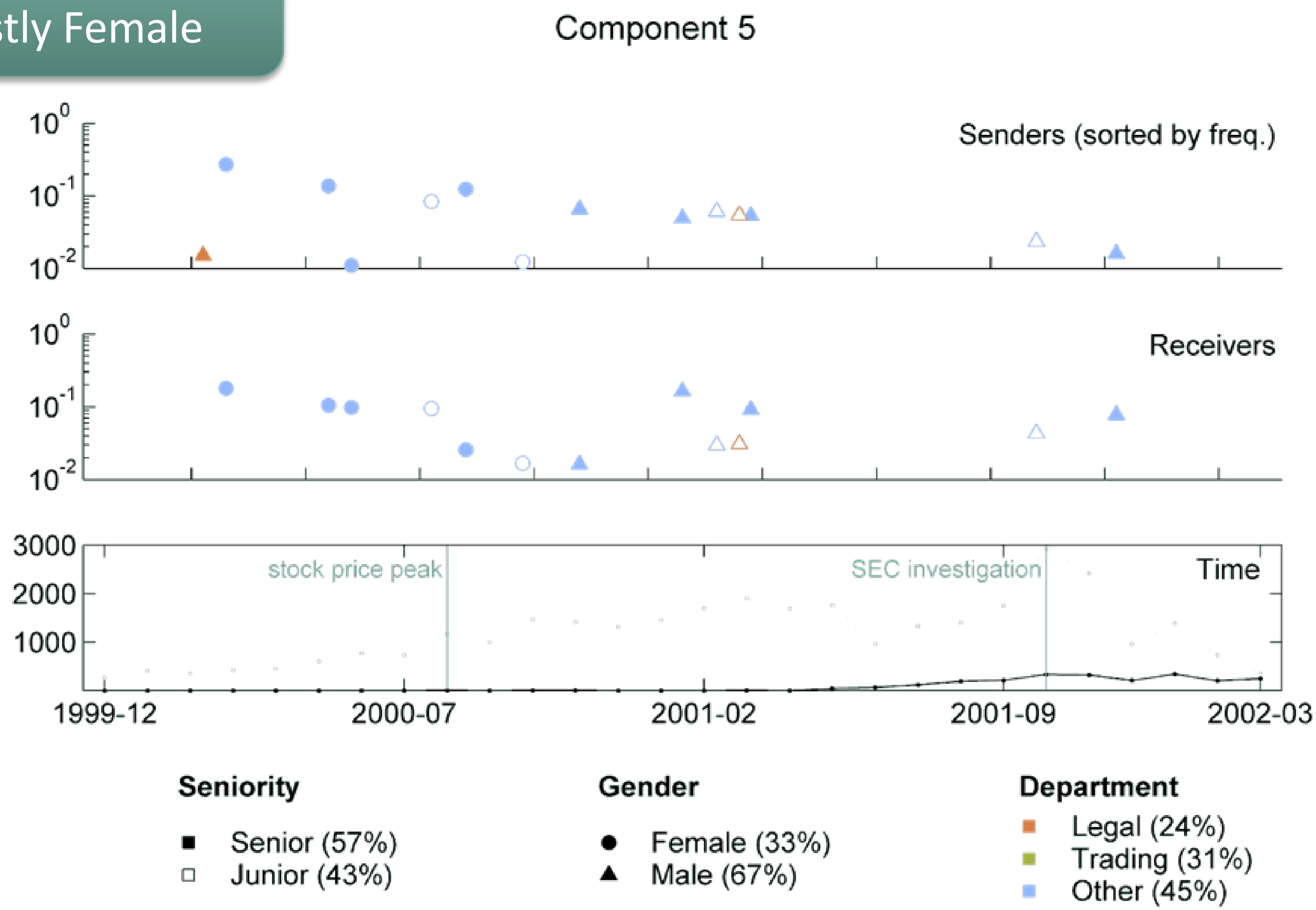
Enron Email Data

Not Legal



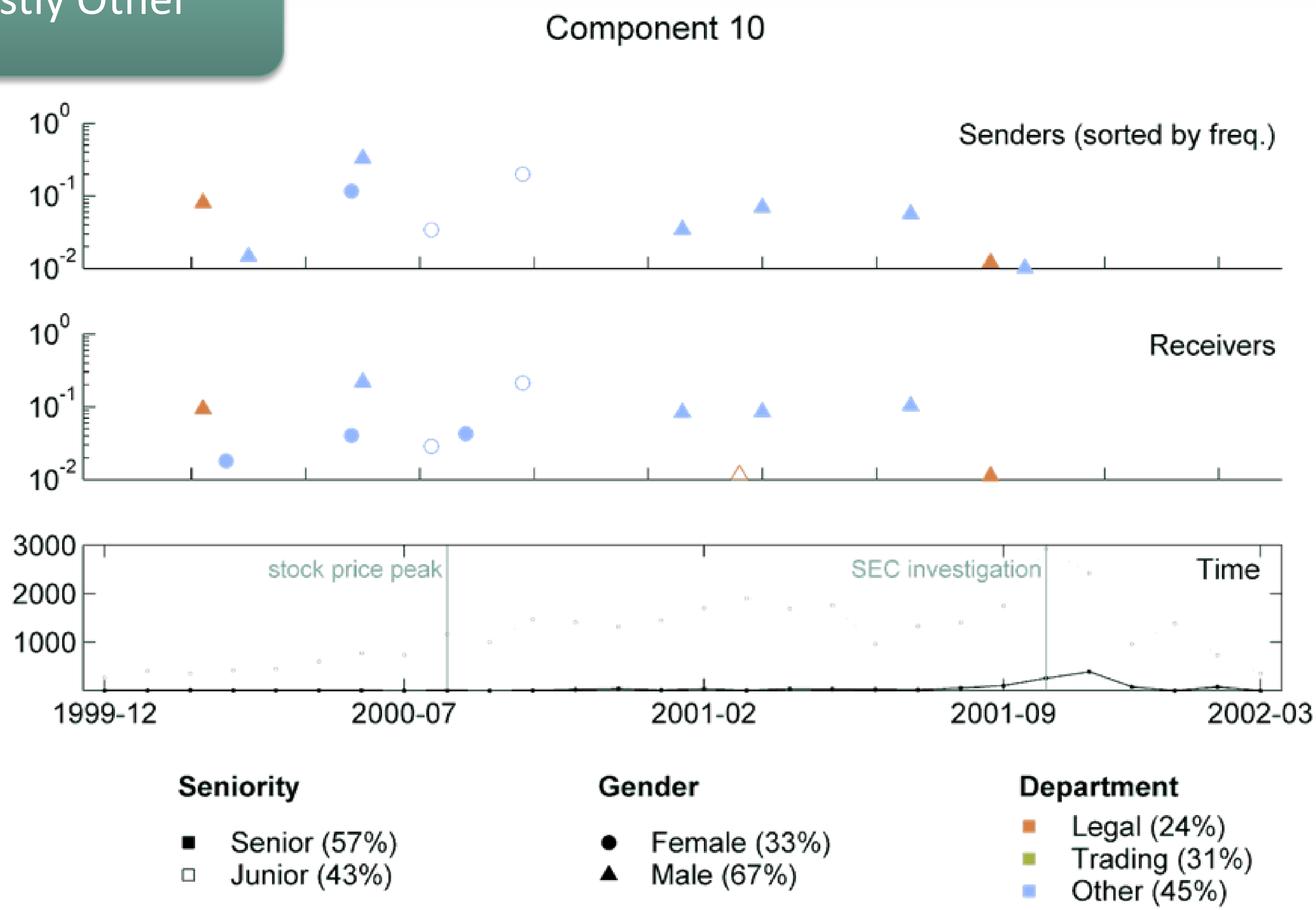
Enron Email Data

Other;
Mostly Female



Enron Email Data

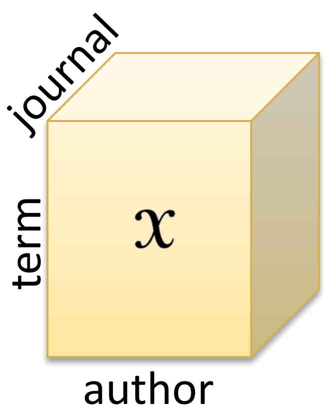
Mostly Other



Example: Publication Data

SIAM publications 1999-2004

- 4676 articles
- 11 journals
- 6955 authors
- 4952 title terms
  (after stop-word removal)



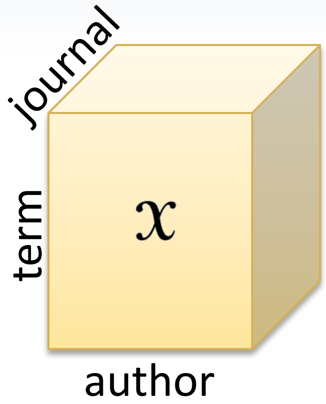
x_{ijk} = occurrences of term i in titles of articles by author j in journal k

64,133 nonzeros
< .01% dense

Sample Publications

- EIBECK A, WAGNER W, An efficient stochastic algorithm for studying coagulation dynamics and gelation phenomena, SIAM J SCI COMPUT, 2000
- KIM S, KWON O, SEO JK, Location search techniques for a grounded conductor, SIAM J APPL MATH, 2002
- CROWDY D, MARSHALL J, Constructing multiply connected quadrature domains, SIAM J APPL MATH, 2004
- LIPTON R, Optimal inequalities for gradients of solutions of elliptic equations occurring in two-phase heat conductors, SIAM J MATH ANAL, 2001
- LAFITTE OD, Diffraction in the high frequency regime by a thin layer of dielectric material I: The equivalent impedance boundary condition, SIAM J APPL MATH, 1999

Publication Data Yields Topics, with Authors and Journals



SIAM Data
1999-2004

- 4952 terms
- 6955 authors
- 11 journals
- 64k nonzeros

10 Component
Poisson Tensor
Factorization

Component 1
graphs
problem
algorithms
approximation
algorithm
complexity
optimal
trees
problems
bounds

Kao MY
Peleg D
Motwani R
Cole R
Devroye L

SIAM J Comput
SIAM J Discrete Math
SIAM Rev

Component 2
method
equations
methods
problems
numerical
multigrid
finite
element
solution
systems

Chan TF
Saad Y
Golub GH

SIAM J Sci Comput

Component 3
finite
methods
equations
method
element
problems
numerical
error
analysis

Du Q
Shen J
Ainsworth M
Mccormick SF
Wang JP
Manteuffel TA
Schwab C
Ewing RE
Widlund OB
Babuska I

SIAM J Numer Anal
SIAM J Comput

Component 4
control
systems
optimal
problems
stochastic
linear
nonlinear
stabilization
equations
equation

Zhou XY
Kushner HJ
Kunisch K
Ito K
Tang SJ
Raymond JP
Ulbrich S
Borkar VS
Altman E
Budhiraja A

SIAM J Control Optim

Publication Data Results, Cont'd.

Component 5	Component 6	Component 7	Component 8	Component 9	Component 10
equations	matrices	optimization	model	equations	education
solutions	matrix	problems	nonlinear	flow	introduction
problem	problems	programming	equations	model	health
equation	systems	methods	solutions	problem	analysis
boundary	algorithm	method	dynamics	theory	problems
nonlinear	linear	algorithm	waves	asymptotic	matrix
system	method	nonlinear	diffusion	models	method
stability	symmetric	point	system	method	methods
model	problem	semidefinite	analysis	analysis	control
systems	sparse	convergence	phase	singular	programming
Wei JC	Higham NJ	Qi LQ	Venakides S	Klar A	Flaherty J
Chen XF	Guo CH	Tseng P	Knessl C	Ammari H	Trefethen N
Frid H	Tisseur F	Roos C	Sherratt JA	Wegener R	Schnabel B
Yang T	Zhang ZY	Sun DF	Ermentrout GB	Schuss Z	[None]
Krauskopf B	Johnson CR	Kunisch K	Scherzer O	Stevens A	Moon G
Hohage T	Lin WW	Ng KF	Haider MA	Velazquez JJL	Shor PW
Seo JK	Mehrmann V	Jeyakumar V	Kaper TJ	Miura RM	Babuska IM
Krylov NV	Gu M	Qi HD	Ward MJ	Movchan AB	Sauter SA
Nishihara K	Zha HY	Fukushima M	Tier C	Fannjiang A	Van Dooren P
Friedman A	Golub GH	Kojima M	Warne DP	Ryzhik L	Adjei S
SIAM J Math Anal	SIAM J Matrix Anal A	SIAM J Optimiz	SIAM J Appl Math	SIAM J Appl Math	SIAM Rev
SIAM J Appl Dyn Syst	SIAM J Sci Comput			SIAM J Optimiz	

Similar Solutions found with Different Starting Points

Component 1

graphs
problem
algorithms
approximation
algorithm
complexity
optimal
trees
problems
bounds

Kao MY
Peleg D
Motwani R
Cole R
Devroye L

SIAM J Comput
SIAM J Discrete Math

Component 1

graphs
problem
algorithms
approximation
algorithm
complexity
optimal
trees
problems
bounds

Kao MY
Peleg D
Motwani R
Cole R
Devroye L

SIAM J Comput
AM J Discrete Math
SIAM Rev

Component 5

equations
solutions
problem
boundary
equation
nonlinear
stability
model
systems
system

Wei JC
Chen XF
Frid H
Yang T
Seo JK
Hohage T
Krylov NV
Nishihara K
Wu JH
Friedman A

SIAM J Math Anal

Component 5

equations
solutions
problem
equation
boundary
nonlinear
system
stability
model
systems

Wei JC
Chen XF
Frid H
Yang T
Krauskopf B
Hohage T
Seo JK
Krylov NV
Nishihara K
Friedman A

AM J Math Anal
SIAM J Appl Dyn Syst

Component 10

analysis
education
health
introduction
method
problems
methods
matrix
control
survey

Flaherty J
Trefethen N
Krauskopf B
Schnabel B
[None]
Hoffman K
Guckenheimer J
Moon G
Osinga HM
Shor PW

SIAM Rev
SIAM J Appl Dyn Syst

Component 10

education
introduction
health
analysis
problems
matrix
method
methods
control
programming

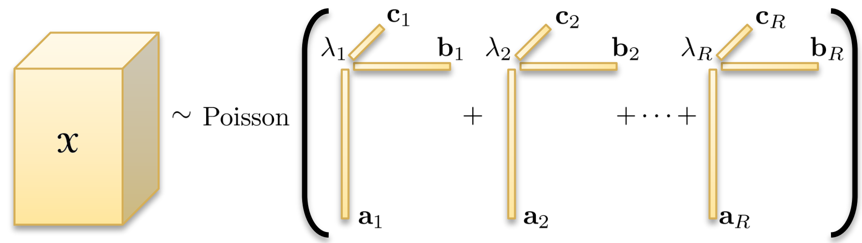
Flaherty J
Trefethen N
Schnabel B
[None]
Moon G
Shor PW
Babuska IM
Sauter SA
Van Dooren P
Adjei S

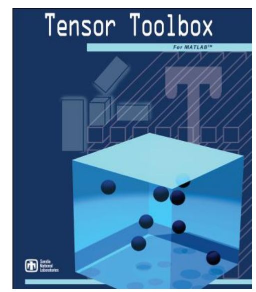
SIAM Rev

Sandia National Laboratories

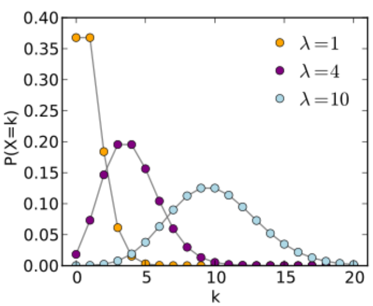
Concluding Remarks

- Data distribution matters!
  - Least squares fitting implies Gaussian
  - Poisson distribution better for sparse count data
- Model fitting via CP-APR
  - Alternating algorithm with multiplicative updates
    - Lee-Seung method is a special case
  - Can directly check convergence conditions
    - Fix for “undesirable zero” problem
- Future work
  - Modified version of Anderson acceleration for fixed point iterations
  - Alternate optimization methods
- Other on-going tensor work
  - Generalized tensor eigenproblem
  - Symmetric tensor decompositions





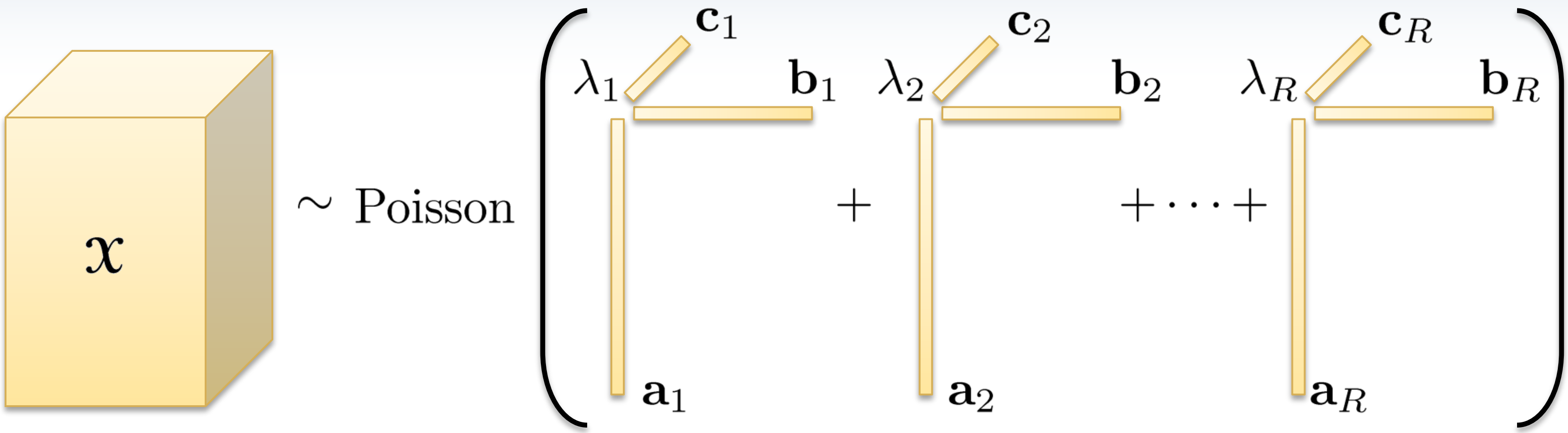
CP-APR will be in the next release of the Tensor Toolbox for MATLAB.



For more information:
Tammy Kolda
tgkolda@sandia.gov

Back-up/Old Slides

Poisson Tensor Factorization (PTF)



<u>Model</u>: Poisson/Multinomial distribution (nonnegative factorization)

$$x_{ijk} \sim \text{Poisson}(m_{ijk}) \text{ where } m_{ijk} = \sum_r \lambda_r \, a_{ir} \, b_{jr} \, c_{kr}$$

<u>Useful properties of Poisson distributed variables</u>:

- Generally preferred for describing “count” data
- Model is “naturally” nonnegative
- The expected value is equal to its parameters and so is its variance
- Sums of Poisson-distributed random variables also follow a Poisson distribution whose parameter is the sum of the component parameters