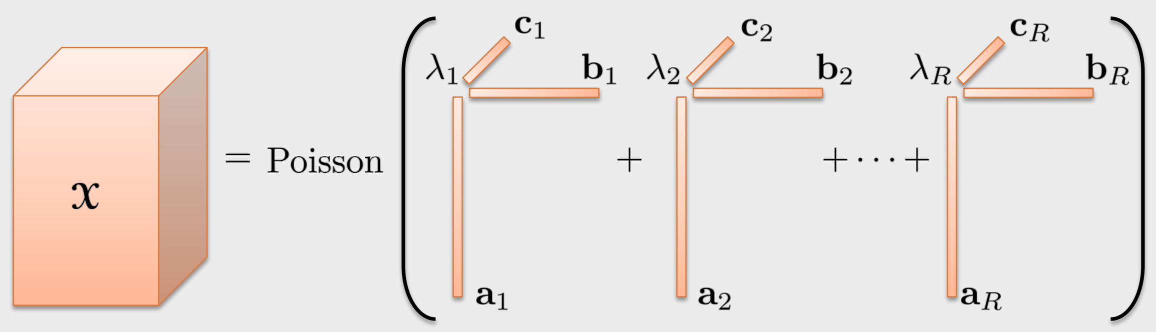
SAND2011-7599C

Nonnegative Tensor Factorizations for Sparse Count Data

Tamara G. Kolda
Sandia National Laboratories • California

Eric C. Chi
Rice University & UCLA

Poisson Tensor Factorization (PTF)



Model: Poisson/Multinomial distribution (nonnegative factorization)

$$\mathfrak{X}(i,j,k) \sim \text{Poisson}(\mathfrak{M}(i,j,k)) \quad \text{where} \quad \mathfrak{M}(i,j,k) = \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r(i) \mathbf{b}_r(j) \mathbf{c}_r(k)$$

Useful properties of Poisson distributed variables:

- Generally preferred for describing "count" data
- The expected value is equal to its parameter 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

Gaussian (typical)

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

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

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

$$\min_{\mathfrak{M}} \sum_{ijk} (\mathfrak{X}_{ijk} - \mathfrak{M}_{ijk})^2$$

Key references for Tensor Factorization with Gaussian fit: Harshman (1970), Carroll and Chang (1970)

Poisson

The random variable x is a discrete nonnegative integer.

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

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

$$\min_{\mathfrak{M}} \sum_{ijk} \mathfrak{M}_{ijk} - \mathfrak{X}_{ijk} \log \mathfrak{M}_{ijk}$$

Fitting a Poisson Factorization

$$\min_{\mathfrak{M}} \sum_{ijk} \mathfrak{M}_{ijk} - \mathfrak{X}_{ijk} \log \mathfrak{M}_{ijk}$$

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

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

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

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 λ into one of the factor matrices, e.g.,

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

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

Majorization-Minimization

- Definition: A function $g(\mathbf{y};\mathbf{x})$ **majorizes** a function $f(\mathbf{y})$ at $\mathbf{x}$ if $g(\mathbf{y},\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{y}$ and $g(\mathbf{x};\mathbf{x}) = f(\mathbf{x})$
- Algorithm: MM approach for minimizing $f(\mathbf{x})$:

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

- Idea: Choose a majorizing function that is *easy* to minimize

Alternating Poisson Regression (CP-APR) Algorithm

Repeat until converged…

1. $\bar{\mathbf{A}} \leftarrow \arg\min_{\bar{\mathbf{A}} \geq 0} \sum_{ijk} \mathfrak{M}_{ijk} - \mathfrak{X}_{ijk} \log \mathfrak{M}_{ijk}$ s.t. $\mathfrak{M} = [\![\bar{\mathbf{A}}, \mathbf{B}, \mathbf{C}]\!]$
2. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{A}} \mathbf{e}$; $\mathbf{A} \leftarrow \bar{\mathbf{A}} \cdot \text{diag}(1/\boldsymbol{\lambda})$

(Fix B,C; solve for A)

3. $\bar{\mathbf{B}} \leftarrow \arg\min_{\bar{\mathbf{B}} \geq 0} \sum_{ijk} \mathfrak{M}_{ijk} - \mathfrak{X}_{ijk} \log \mathfrak{M}_{ijk}$ s.t. $\mathfrak{M} = [\![\mathbf{A}, \bar{\mathbf{B}}, \mathbf{C}]\!]$
4. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{B}} \mathbf{e}$; $\mathbf{B} \leftarrow \bar{\mathbf{B}} \cdot \text{diag}(1/\boldsymbol{\lambda})$

(Fix A,C; solve for B)

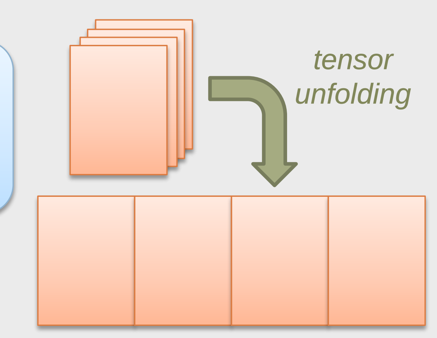
5. $\bar{\mathbf{C}} \leftarrow \arg\min_{\bar{\mathbf{C}} \geq 0} \sum_{ijk} \mathfrak{M}_{ijk} - \mathfrak{X}_{ijk} \log \mathfrak{M}_{ijk}$ s.t. $\mathfrak{M} = [\![\mathbf{A}, \mathbf{B}, \bar{\mathbf{C}}]\!]$
6. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{C}} \mathbf{e}$; $\mathbf{C} \leftarrow \bar{\mathbf{C}} \cdot \text{diag}(1/\boldsymbol{\lambda})$

(Fix A,B; solve for C)

Solving the Subproblem (Mode 1)

$$\min_{\bar{\mathbf{A}} \geq 0} \sum_{i\ell} \mathbf{M}_{i\ell} - \mathbf{X}_{i\ell} \log \mathbf{M}_{i\ell}$$

$$\text{subject to} \quad \mathbf{M} = \bar{\mathbf{A}}\boldsymbol{\Pi} \ \text{with} \ \boldsymbol{\Pi} = (\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$



- Suppose tensor is of size I x J x K. It can be "unfolded" into an I x L matrix where $L = JK$

$$\mathbf{X}(i,\ell) = \mathfrak{X}(i,j,k), \quad \mathbf{M}(i,\ell) = \mathfrak{M}(i,j,k), \quad \ell = j + (k-1) \cdot J$$

- We define the new matrix $\boldsymbol{\Pi}$ of size R x L as

$$\boldsymbol{\Pi}(r,\ell) = \mathbf{b}_r(j)\mathbf{c}_r(k)$$

- Then we can write the model as $\mathfrak{M} = \sum_r \bar{\mathbf{a}}_r \circ \mathbf{b}_r \mathbf{c}_r \Rightarrow \mathbf{M} = \bar{\mathbf{A}}\boldsymbol{\Pi}$
- Note we can rewrite objective as $\mathbf{e}^{\mathsf{T}} [\mathbf{M} - \mathbf{X} * \log \mathbf{M}] \, \mathbf{e}$

$$\min_{\bar{\mathbf{A}} \geq 0} \mathbf{e}^{\mathsf{T}} \left[\bar{\mathbf{A}}\boldsymbol{\Pi} - \mathbf{X} * \log (\bar{\mathbf{A}}\boldsymbol{\Pi}) \right] \mathbf{e}$$

Lemma: The subproblems are strictly convex if the data tensor has a sufficient number of nonzeros and they are reasonably distributed.

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

Subproblem objective function

$$f(\bar{\mathbf{A}}) = \sum_{i\ell} \left(\sum_r \bar{\mathbf{A}}_{ir} \boldsymbol{\Pi}_{r\ell} \right) - \mathbf{X}_{i\ell} \log \left(\sum_r \bar{\mathbf{A}}_{ir} \boldsymbol{\Pi}_{r\ell} \right)$$

Majorizer

$$g(\mathbf{A};\bar{\mathbf{A}}) = \sum_{ri\ell} \mathbf{A}_{ir}\boldsymbol{\Pi}_{r\ell} - \alpha_{ri\ell} \mathbf{X}_{i\ell} \log \left(\frac{\mathbf{A}_{ir}\boldsymbol{\Pi}_{r\ell}}{\alpha_{ri\ell}} \right)$$

where $\alpha_{ri\ell} = \bar{\mathbf{A}}_{ir}\boldsymbol{\Pi}_{r\ell} / \sum_{r'} \bar{\mathbf{A}}_{ir'}\boldsymbol{\Pi}_{r'\ell}$

Update in element form:

$$\bar{\mathbf{A}}_{ir} \leftarrow \bar{\mathbf{A}}_{ir} \sum_{\ell} \frac{\mathbf{X}_{i\ell}}{\sum_{r'} \bar{\mathbf{A}}_{ir'}\boldsymbol{\Pi}_{r'\ell}} \boldsymbol{\Pi}_{r\ell}$$

Observe this can never be negative

Update in matrix form:

$$\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * \boldsymbol{\Phi} \quad \text{where} \quad \boldsymbol{\Phi} = [\mathbf{X} \oslash (\bar{\mathbf{A}}\boldsymbol{\Pi})] \, \boldsymbol{\Pi}^{\mathsf{T}}$$

Constrained Optimality (KKT) Conditions

$$\bar{\mathbf{A}} \geq 0 \quad \longleftarrow \text{automatically guaranteed}$$

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

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

These conditions enable us to check for "inadmissible" zeros

Convergence criterion:

$$\min(\hat{\mathbf{A}}, |\mathbf{E} - \boldsymbol{\Phi}|) \leq \text{tol}$$

CP-APR with MM Subproblem Solver

Repeat until converged…

1. Repeat until converged: $\bar{\mathbf{A}} \leftarrow \bar{\mathbf{A}} * [\mathbf{X}_{(1)} \oslash \bar{\mathbf{A}}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}] \, (\mathbf{C} \odot \mathbf{B})$
2. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{A}} \mathbf{e}$; $\mathbf{A} \leftarrow \bar{\mathbf{A}} \cdot \text{diag}(1/\boldsymbol{\lambda})$

(Fix B,C; solve for A)

3. Repeat until converged: $\bar{\mathbf{A}} \leftarrow \bar{\mathbf{B}} * [\mathbf{X}_{(2)} \oslash \bar{\mathbf{B}}(\mathbf{C} \odot \mathbf{A})^{\mathsf{T}}] \, (\mathbf{C} \odot \mathbf{A})$
4. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{B}} \mathbf{e}$; $\mathbf{B} \leftarrow \bar{\mathbf{B}} \cdot \text{diag}(1/\boldsymbol{\lambda})$

(Fix A,C; solve for B)

5. Repeat until converged: $\bar{\mathbf{C}} \leftarrow \bar{\mathbf{C}} * [\mathbf{X}_{(3)} \oslash \bar{\mathbf{C}}(\mathbf{B} \odot \mathbf{A})^{\mathsf{T}}] \, (\mathbf{B} \odot \mathbf{A})$
6. $\boldsymbol{\lambda} \leftarrow \mathbf{e}^{\mathsf{T}} \bar{\mathbf{C}} \mathbf{e}$; $\mathbf{C} \leftarrow \bar{\mathbf{C}} \cdot \text{diag}(1/\boldsymbol{\lambda})$

(Fix A,B; solve for C)

$\mathbf{X}_{(n)}$ = tensor unfolded into a matrix w.r.t. mode n

The Lee & Seung version of Nonnegative Tensor Factorization (Welling & Weber, 2001) is a special case of CP-APR with a single step for each subproblem solve.

Simulated Data

- Idea: Generate data from an existing model and see if we can recover that model with our algorithm
- Create generative model with R factors:
  - Each factor (i.e., $\mathbf{a}_r$) has $1/R$ entries from U[0,10R] and the remainder from U[0,1]

$$\mathfrak{M} = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

- For n = 1,…,N (# entries to insert)
  - Choose factor r proportional to λ_r
  - Choose index i proportional to $\mathbf{a}_r$
  - Choose index j proportional to $\mathbf{b}_r$
  - Choose index k proportional to $\mathbf{c}_r$

$$\mathfrak{X}(i,j,k) = \mathfrak{X}(i,j,k) + 1$$

- 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

Nonzeros	R Components	R+1 Components
480,000 (.1%)	0.99	0.98
240,000 (.05%)	0.99	0.99
48,000 (.01%)	0.95	0.80
24,000 (.005%)	0.77	0.87

$$\text{score} = \frac{1}{R} \sum_r \left(1 - \frac{|\lambda_r - \hat{\lambda}_r|}{\max(\lambda_r, \hat{\lambda}_r)} \right) \cos(\mathbf{a}_r, \hat{\mathbf{a}}_r) \cos(\mathbf{b}_r, \hat{\mathbf{b}}_r) \cos(\mathbf{c}_r, \hat{\mathbf{c}}_r)$$

Fixing Undesirable Zeros

Problem: Zeros never change with multiplicative updates!

Undesirable Zero: $\mathbf{A}_{ij} = 0$ and $\boldsymbol{\Phi}_{ij} > 1$

Fix: If $\mathbf{A}_{ij}$ is close to zero and $\boldsymbol{\Phi}_{ij} > 1$, then bump $\mathbf{A}_{ij}$ to away from zero, e.g., reset $\mathbf{A}_{ij} = 0.2$

See example of problem in Gonzalez & Zhang, 2005

This fixes Lee-Seung updates too!

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?

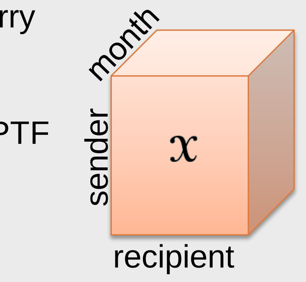
Can we find patterns of behavior?

Can we spot anomalies?

Can we predict future behavior?

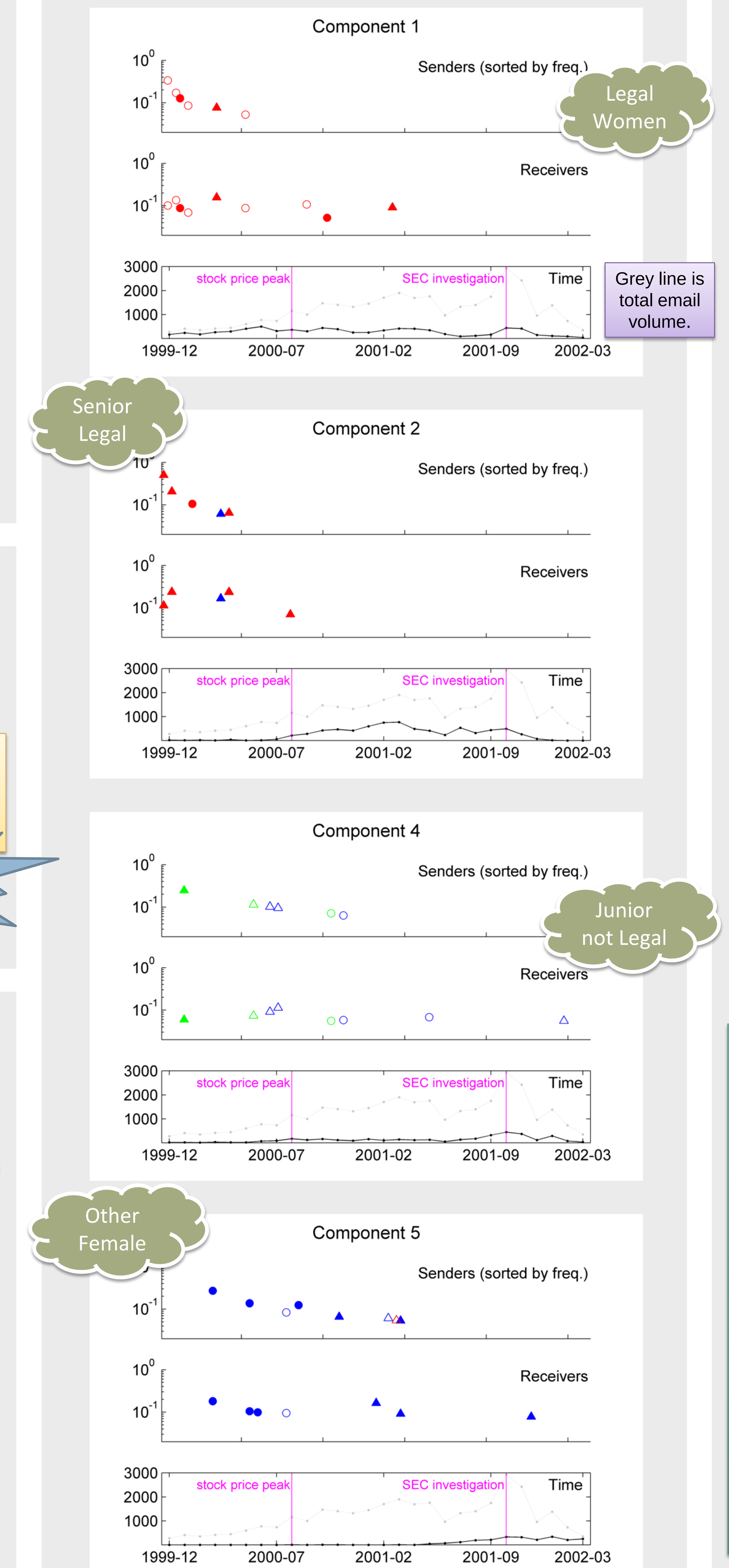
Email Analysis Clusters by Gender, Seniority, Dept.

Analysis of emails from Enron FERC investigation. Dataset from Zhou et al. (2007). Owen and Perry (2010) hypothesize that the data groups by gender, seniority, and department. Our 10-component PTF analysis shows this type of grouping.



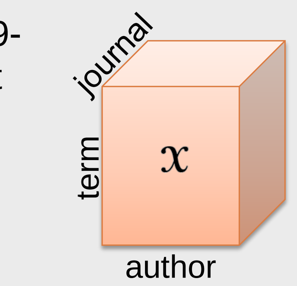
8540 Email Messages
105 Senders/Recipients
28 Months
8.5k nonzeros

■ Senior (57%) □ Junior (43%)
● Female (33%) ▲ Male (67%)
■ Legal (24%) ■ Trading (31%) ■ Other (45%)



Publication Data Yields Topics, with Authors and Journals

We analyze SIAM journal publication data from 1999-2004 using 10-component PTF analysis.



4952 terms
6955 authors
11 journals
64k nonzeros

Component 1	Component 2	Component 3
graphs, problem, algorithms, approximation, algorithm, complexity, optimal, trees, problems, bounds	method, equations, methods, problems, numerical, multigrid, finite, element, solution, systems	finite, methods, equations, method, element, problems, numerical, error, analysis
Kao MY, Peleg D, Motwani R, Cole R, Devroye L	Chan TF, Saad Y, Golub GH	Du Q, Shen J, Ainsworth M, Mccormick SF, Wang JP, Manteuffel TA, Schwab C, Ewing RE, Widlund OB, Babuska I
SIAM J Comput, SIAM J Discrete Math, SIAM Rev	SIAM J Sci Comput	SIAM J Numer Anal, SIAM J Comput

Component 4	Component 5	Component 6
control, systems, optimal, problems, stochastic, linear, nonlinear, stabilization, equations, equation	equations, solutions, problem, equation, boundary, nonlinear, system, stability, model, systems	matrices, matrix, problems, systems, algorithm, linear, method, symmetric, problem, sparse
Zhou XY, Kushner HJ, Kunisch K, Ito K, Tang SJ, Raymond JP, Ulbrich S, Borkar VS, Altman E, Budhiraja A	Wei JC, Chen XF, Frid H, Yang T, Krauskopf B, Hohage T, Seo JK, Krylov NV, Nishihara K, Friedman A	Higham NJ, Guo CH, Tisseur F, Zhang ZY, Johnson CR, Lin WW, Mehrmann V, Gu M, Zha HY, Golub GH
SIAM J Control Optim	SIAM J Math Anal, SIAM J Appl Dyn Syst	SIAM J Matrix Anal A, SIAM J Sci Comput

Component 7	Component 8	Component 9	Component 10
optimization, problems, programming, methods, method, algorithm, nonlinear, point, semidefinite, convergence	model, nonlinear, equations, solutions, dynamics, waves, diffusion, system, analysis, phase	equations, flow, model, problem, theory, asymptotic, models, method, analysis, singular	education, introduction, health, analysis, problems, matrix, method, methods, control, programming
Qi LQ, Tseng P, Roos C, Sun DF, Kunisch K, Ng KF, Jeyakumar V, Qi HD, Fukushima M, Kojima M	Venakides S, Knessl C, Sherratt JA, Ermentrout GB, Scherzer O, Haider MA, Kaper TJ, Ward MJ, Tier C, Warne DP	Klar A, Ammari H, Wegener R, Schuss Z, Stevens A, Velazquez JJL, Miura RM, Movchan AB, Fannjiang A, Ryzhik L	Flaherty J, Trefethen N, Schnabel B, [None], Moon G, Shor PW, Babuska IM, Sauter SA, Van Dooren P, Adjei S
SIAM J Optimiz	SIAM J Appl Math	SIAM J Appl Math, SIAM J Optimiz	SIAM Rev

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.

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