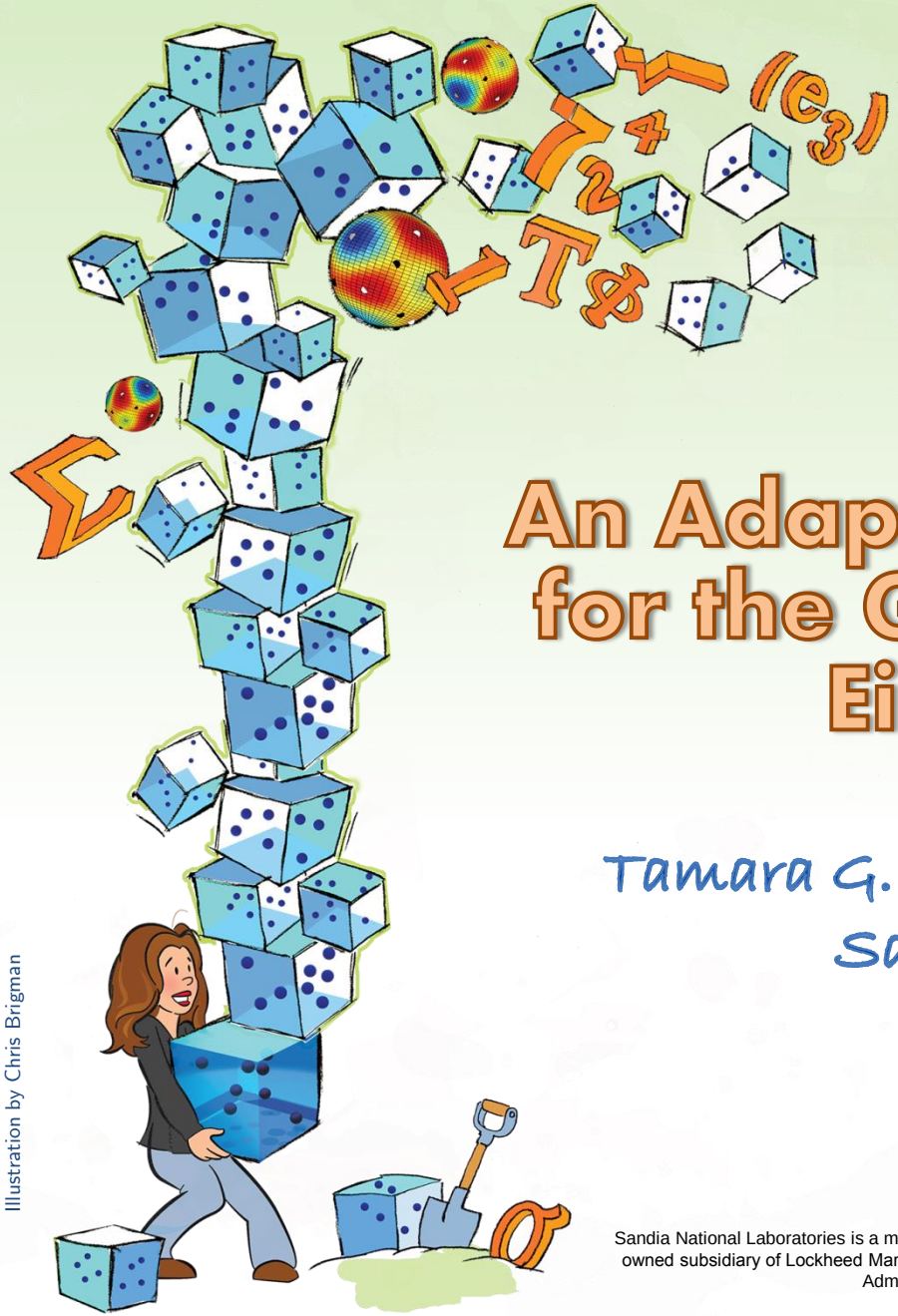


Illustration by Chris Brigman

Sandia National Laboratories

SAND2014-4327C

An Adaptive Power Method for the Generalized Tensor Eigenproblem

Tamara G. Kolda and Jackson C. Mayo
Sandia National Labs
Livermore, CA

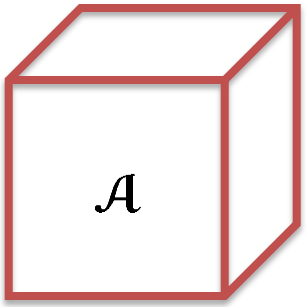
ASCR

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.

U.S. DEPARTMENT OF ENERGY

Symmetric Tensors

- $\mathbb{S}^{[m,n]}$ = set of m-way, n-dimensional symmetric tensors
- m = number of modes or ways
- n = size of each mode
- symmetric = entries invariant to permutation of indices



3-way tensor

Symmetry for 3-way tensor (m = 3)

$$a_{ijk} = a_{ikj} = a_{jik} = a_{kij} = a_{jki} = a_{kji}$$
$$\text{for all } i, j, k \in \{1, 2, \ldots, n\}$$

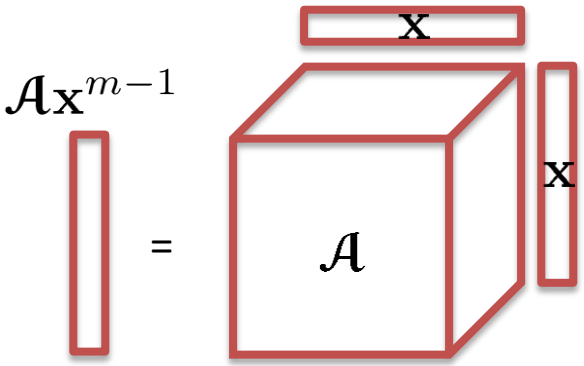
- Applications of symmetric tensors: diffusion tensor imaging (DTI/HARDI), higher-order statistics, higher-order derivatives, relativity, signal processing, etc.

Tensor-vector product: $\mathcal{A}\mathbf{x}^{m-1}$

$$(\mathcal{A}\mathbf{x}^{m-1})_{i_1} = \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1 i_2 \ldots i_m} x_{i_2} \cdots x_{i_m} \text{ for } i_1 \in \{1, \ldots, n\}$$

Matrix Case (m = 2):

$$(\mathbf{A}\mathbf{x})_{i_1} = \sum_{i_2=1}^{n} a_{i_1 i_2} x_{i_2} \text{ for } i_1 \in \{1, \ldots, n\}$$



3-way Case (m = 3):

$$(\mathcal{A}\mathbf{x}^{2})_{i_1} = \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} a_{i_1 i_2 i_3} x_{i_2} x_{i_3} \text{ for } i_1 \in \{1, \ldots, n\}$$

$$\mathcal{A} \in \mathbb{S}^{[m,n]}$$

$$\mathbf{x} \in \mathbb{R}^{n}$$

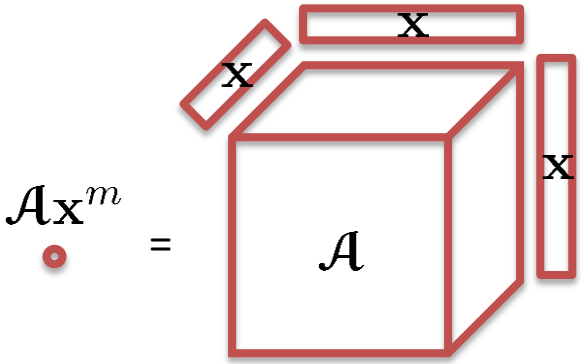
$$\mathcal{A}\mathbf{x}^{m-1} \in \mathbb{R}^{n}$$

Tensor-vector product: $\mathbf{A}\mathbf{x}^m$

$$\mathcal{A}\mathbf{x}^m = (\mathcal{A}\mathbf{x}^{m-1})^{\mathsf{T}}\mathbf{x} = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n} a_{i_1 i_2 \ldots i_m} x_{i_1} x_{i_2} \cdots x_{i_m}$$

Matrix Case (m = 2):

$$\mathbf{A}\mathbf{x}^2 = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} a_{i_1 i_2} x_{i_1} x_{i_2}$$



3-way Case (m = 3):

$$\mathcal{A}\mathbf{x}^3 = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} a_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3}$$

$$\mathcal{A} \in \mathbb{S}^{[m,n]}$$

$$\mathbf{x} \in \mathbb{R}^n$$

$$\mathcal{A}\mathbf{x}^m \in \mathbb{R}$$

Positive Definite Tensors

A symmetric tensor A is **positive definite** if

$$\mathcal{A}\mathbf{x}^m > 0 \text{ for all } \mathbf{x} \neq 0$$

Recall:

$$\mathcal{A}\mathbf{x}^m = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1 i_2 \ldots i_m} x_{i_1} x_{i_2} \cdots x_{i_m}$$

Observe that m must be even if A is positive definite:

$$\mathcal{A}(-\mathbf{x})^m = (-1)^m \cdot \mathcal{A}\mathbf{x}^m$$

Generalized Eigenproblem

Let A and B be m-way tensors of dimension n and assume B is positive definite. The generalized eigenproblem is defined as finding a pair (λ, x) that satisfies:

$$\mathcal{A} \in \mathbb{S}^{[m,n]}$$

$$\mathcal{B} \in \mathbb{S}_{+}^{[m,n]}$$

positive definite

Generalized Eigenpair

(Chang, Pearson, Zhang 2009)

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathcal{B}\mathbf{x}^{m-1}$$

$$\text{subject to } (\lambda, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^n$$

Note that

$$\lambda = \frac{\mathcal{A}\mathbf{x}^m}{\mathcal{B}\mathbf{x}^m}$$

Homogeneous system of n degree-(m-1) polynomials:

$$\sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} \left(a_{i_1 i_2 \ldots i_m} - \lambda\, b_{i_1 i_2 \ldots i_m} \right) x_{i_2} \cdots x_{i_m} = 0 \text{ for } i_1 \in \{1, \ldots, n\}$$

Optimization Formulation

Generalized Eigenpair

(Chang, Pearson, Zhang 2009)

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathcal{B}\mathbf{x}^{m-1}$$

$$\text{subject to } (\lambda, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^n$$

Nonlinear Program

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \equiv \frac{\mathcal{A}\mathbf{x}^m}{\mathcal{B}\mathbf{x}^m} \|\mathbf{x}\|^m$$

$$\text{subject to } \|\mathbf{x}\| = 1$$

Theorem: Any generalized eigenpair (λ,x) is a KKT point of the nonlinear program with λ as the Lagrange multiplier (and reverse is also true)

Proof: $\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda(\|\mathbf{x}\|^m - 1)$

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \lambda) = \frac{m}{\mathcal{B}\mathbf{x}^m}\left[\left(\mathcal{A}\mathbf{x}^m\right)\mathbf{x} + \mathcal{A}\mathbf{x}^{m-1} - \left(\frac{\mathcal{A}\mathbf{x}^m}{\mathcal{B}\mathbf{x}^m}\right)\mathcal{B}\mathbf{x}^{m-1}\right] - m\lambda\mathbf{x}$$

$$= \frac{m}{\mathcal{B}\mathbf{x}^m}\left[\left(\mathcal{A}\mathbf{x}^m\right)\mathbf{x}\right] - m\lambda\mathbf{x} = 0$$

Special Technique for Optimization on a Sphere

$$\max_{\mathbf{x}\in\mathbb{R}^n} \hat{f}(\mathbf{x}) \text{ s.t. } \mathbf{x} \in \Sigma = \{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1 \,\}$$

Assume $\mathbf{w} \in \Sigma$,
$\Omega =$ open nbhd of $\mathbf{w}$,
$\hat{f}$ convex and C^1 on Ω

&

Define
$$\mathbf{v} = \frac{\nabla \hat{f}(\mathbf{w})}{\|\nabla \hat{f}(\mathbf{w})\|}$$

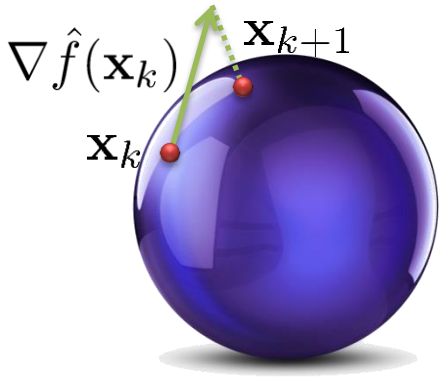
$\Rightarrow$

If $\mathbf{v} \in \Omega$ and $\mathbf{v} \neq \mathbf{w}$,
then $\hat{f}(\mathbf{v}) > \hat{f}(\mathbf{w})$

(Regalia & Kofidis , 2003)

Simple Algorithm:

$$\mathbf{x}_{k+1} \leftarrow \frac{\nabla \hat{f}(\mathbf{x}_k)}{\|\nabla \hat{f}(\mathbf{x}_k)\|}$$



Yields monotonically *increasing* function values and guaranteed to converge since operating on compact set.

Special Technique for Optimization on a Sphere

$$\max_{\mathbf{x}\in\mathbb{R}^n} \hat{f}(\mathbf{x}) \text{ s.t. } \mathbf{x} \in \Sigma = \{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1 \,\}$$

Assume $\mathbf{w} \in \Sigma$,
$\Omega =$ open nbhd of $\mathbf{w}$,
$\hat{f}$ convex and C^1 on Ω

&

Define
$$\mathbf{v} = \frac{\nabla \hat{f}(\mathbf{w})}{\|\nabla \hat{f}(\mathbf{w})\|}$$

$\Rightarrow$

If $\mathbf{v} \in \Omega$ and $\mathbf{v} \neq \mathbf{w}$,
then $\hat{f}(\mathbf{v}) > \hat{f}(\mathbf{w})$

(Regalia & Kofidis , 2003)

Simple Algorithm: $$\mathbf{x}_{k+1} \leftarrow \frac{\nabla \hat{f}(\mathbf{x}_k)}{\|\nabla \hat{f}(\mathbf{x}_k)\|}$$

Yields monotonically *increasing* function values and guaranteed to converge since operating on compact set.

Creating Local Convexity

Original function: $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}, \quad \mathbf{g} \equiv \nabla f(\mathbf{x}), \quad \mathbf{H} \equiv \nabla^2 f(\mathbf{x})$

Shifted function: $\hat{f}(\mathbf{x}) = f(\mathbf{x}) + \alpha \|\mathbf{x}\|^m$

For $\mathbf{x} \in \Sigma$:

$$\hat{f} = f + \alpha, \quad \hat{\mathbf{g}} \equiv \nabla \hat{f}(\mathbf{x}) = \mathbf{g} + \alpha m \mathbf{x},$$
$$\hat{\mathbf{H}} \equiv \nabla^2 \hat{f}(\mathbf{x}) = \mathbf{H} + \alpha m \mathbf{I} + \alpha m(m-2)\mathbf{x}\mathbf{x}^\mathsf{T}$$

$$\hat{f} \text{ convex} \Leftrightarrow \lambda_{\min}(\hat{\mathbf{H}}) > 0 \text{ for } i = 1, \ldots, n$$

Choose α to make all eigenvalues of shifted Hessian positive.

Weyl's Inequality $\lambda_i(\mathbf{H}) + m\alpha \leq \lambda_i(\hat{\mathbf{H}}) \leq \lambda_i(\mathbf{H}) + m\alpha + m(m-2)\alpha \text{ for } \alpha > 0$

Choose $\alpha = \max\{\, 0, \frac{1}{m}(\tau - \lambda_{\min}(\mathbf{H}))\,\}$ $\tau > 0$: Convexity threshold

Gradient and Hessian

$$\max_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \equiv \frac{\mathcal{A}\mathbf{x}^m}{\mathcal{B}\mathbf{x}^m}\|\mathbf{x}\|^m$$

$$\text{subject to } \|\mathbf{x}\| = 1$$

$$\mathbf{g}(\mathbf{x}) \equiv \nabla f(\mathbf{x}) = \frac{m}{\mathcal{B}\mathbf{x}^m}\left[\left(\mathcal{A}\mathbf{x}^m\right)\mathbf{x} + \mathcal{A}\mathbf{x}^{m-1} - \left(\frac{\mathcal{A}\mathbf{x}^m}{\mathcal{B}\mathbf{x}^m}\right)\mathcal{B}\mathbf{x}^{m-1}\right]$$

$$\mathbf{H}(\mathbf{x}) \equiv \nabla^2 f(\mathbf{x}) = \frac{m^2\mathcal{A}\mathbf{x}^m}{(\mathcal{B}\mathbf{x}^m)^3}\left(\mathcal{B}\mathbf{x}^{m-1} \circledcirc \mathcal{B}\mathbf{x}^{m-1}\right)$$

$$+ \frac{m}{\mathcal{B}\mathbf{x}^m}\left[(m-1)\mathcal{A}\mathbf{x}^{m-2} + \mathcal{A}\mathbf{x}^m\left(\mathbf{I} + (m-2)\mathbf{x}\mathbf{x}^\mathsf{T}\right) + m\left(\mathcal{A}\mathbf{x}^{m-1} \circledcirc \mathbf{x}\right)\right]$$

$$- \frac{m}{(\mathcal{B}\mathbf{x}^m)^2}\left[(m-1)\mathcal{A}\mathbf{x}^m\mathcal{B}\mathbf{x}^{m-2} + m\left(\mathcal{A}\mathbf{x}^{m-1} \circledcirc \mathcal{B}\mathbf{x}^{m-1}\right)\right.$$

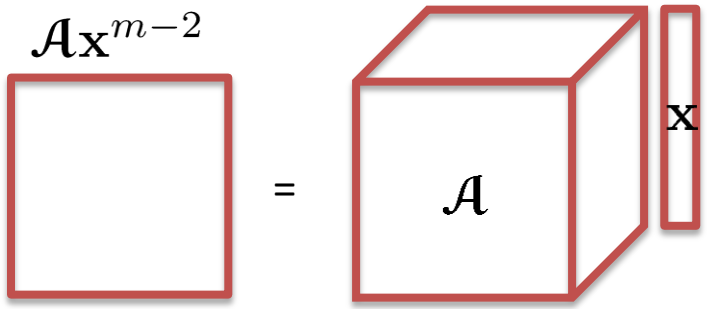
$$\left. + m\mathcal{A}\mathbf{x}^m\left(\mathbf{x} \circledcirc \mathcal{B}\mathbf{x}^{m-1}\right)\right]$$

Tensor-vector product: $\mathcal{A}\mathbf{x}^{m-2}$

$$(\mathcal{A}\mathbf{x}^{m-2})_{i_1 i_2} = \sum_{i_3=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1 i_2 \ldots i_m} x_{i_3} \cdots x_{i_m} \text{ for } i_1, i_2 \in \{1, \ldots, n\}$$

Matrix Case (m = 2):

$$\mathbf{A}\mathbf{x}^{0} = \mathbf{I}$$



3-way Case (m = 3):

$$(\mathcal{A}\mathbf{x}^{1})_{i_1 i_2} = \sum_{i_3=1}^{n} a_{i_1 i_2 i_3} x_{i_3} \text{ for } i_1, i_2 \in \{1, \ldots, n\}$$

$$\mathcal{A} \in \mathbb{S}^{[m,n]}$$

$$\mathbf{x} \in \mathbb{R}^{n}$$

$$\mathcal{A}\mathbf{x}^{m-2} \in \mathbb{S}^{[2,n]}$$

Computing Products

$$\mathcal{A}\mathbf{x}^{m-2}$$

$$\big((m-2)\cdot n^{m-2} \text{ multplies } + n^{m-2} \text{ additions}\big)\cdot n^2 \text{ entries} = (m-1)n^m$$

exploit symmetry to reduce cost to $O(n^m/m!)$

$$\mathcal{A}\mathbf{x}^{m-1} = (\mathcal{A}\mathbf{x}^{m-2})\mathbf{x}$$

$2n^2$ operations

$$\mathcal{A}\mathbf{x}^{m} = (\mathcal{A}\mathbf{x}^{m-1})^{\mathsf{T}}\mathbf{x}$$

$2n$ operations

Generalized Eigenproblem Adaptive Power (GEAP) Method

Sandia National Laboratories

$\mathbf{x}_0 \leftarrow \hat{\mathbf{x}}_0 / \|\hat{\mathbf{x}}_0\|$

for $k = 0, 1, \ldots$ **do**

Precompute $\mathcal{A}\mathbf{x}_k^{m-2}$, $\mathcal{B}\mathbf{x}_k^{m-2}$, $\mathcal{A}\mathbf{x}_k^{m-1}$, $\mathcal{B}\mathbf{x}_k^{m-1}$, $\mathcal{A}\mathbf{x}_k^{m}$, $\mathcal{B}\mathbf{x}_k^{m}$ ← O($n^m/m!$) work

$\lambda_k \leftarrow \mathcal{A}\mathbf{x}_k^m / \mathcal{B}\mathbf{x}_k^m$

$\mathbf{H}_k \leftarrow \mathbf{H}(\mathbf{x}_k)$

β = 1 ⇒ local maxima
β = -1 ⇒ local minima

Adaptive Shift

$\alpha_k \leftarrow \beta \max\{0, (\tau - \lambda_{\min}(\beta \mathbf{H}_k))/m\}$ ← O(n^3) work

$\hat{\mathbf{x}}_{k+1} \leftarrow \beta\left(\mathcal{A}\mathbf{x}_k^{m-1} - \lambda_k \mathcal{B}\mathbf{x}_k^{m-1} + (\alpha_k + \lambda_k)\mathcal{B}\mathbf{x}_k^m \mathbf{x}_k\right)$

$\mathbf{x}_{k+1} = \hat{\mathbf{x}}_{k+1} / \|\hat{\mathbf{x}}_{k+1}\|$

end for

Pros: Extremely simple and fast
Cons: Requires multiple starting points

Comparison to Other Work

- Polynomial Solver
  - E.g., Nsolve in Mathematica
  - Does not scale
- SS-HOPM: Shifted Symmetric Higher-Order Power Method (Kolda & Mayo, 2011)
  - Only solves one special case of generalized eigenpair problem
  - Does not have adaptive shift
  - Otherwise method is identical
- Han's Method (Han, 2012)
  - Unconstrained optimization: $f(\mathbf{x}) = (\mathcal{B}\mathbf{x}^m)^2/2m - \beta\mathcal{A}\mathbf{x}^m/m$
  - Uses any optimization method (e.g., L-BFGS)
  - Did not always converge to generalized eigenpair in our experiments

Z-eigenpairs: B = Identity Tensor

The identity tensor (denoted E) is defined as

$$\mathcal{E}\mathbf{x}^{m-1} = \|\mathbf{x}\|^{m-1}\mathbf{x}$$

Only defined for m even.

Z-Eigenpair

(Qi 2005, Lim 2005)

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda\mathbf{x}$$

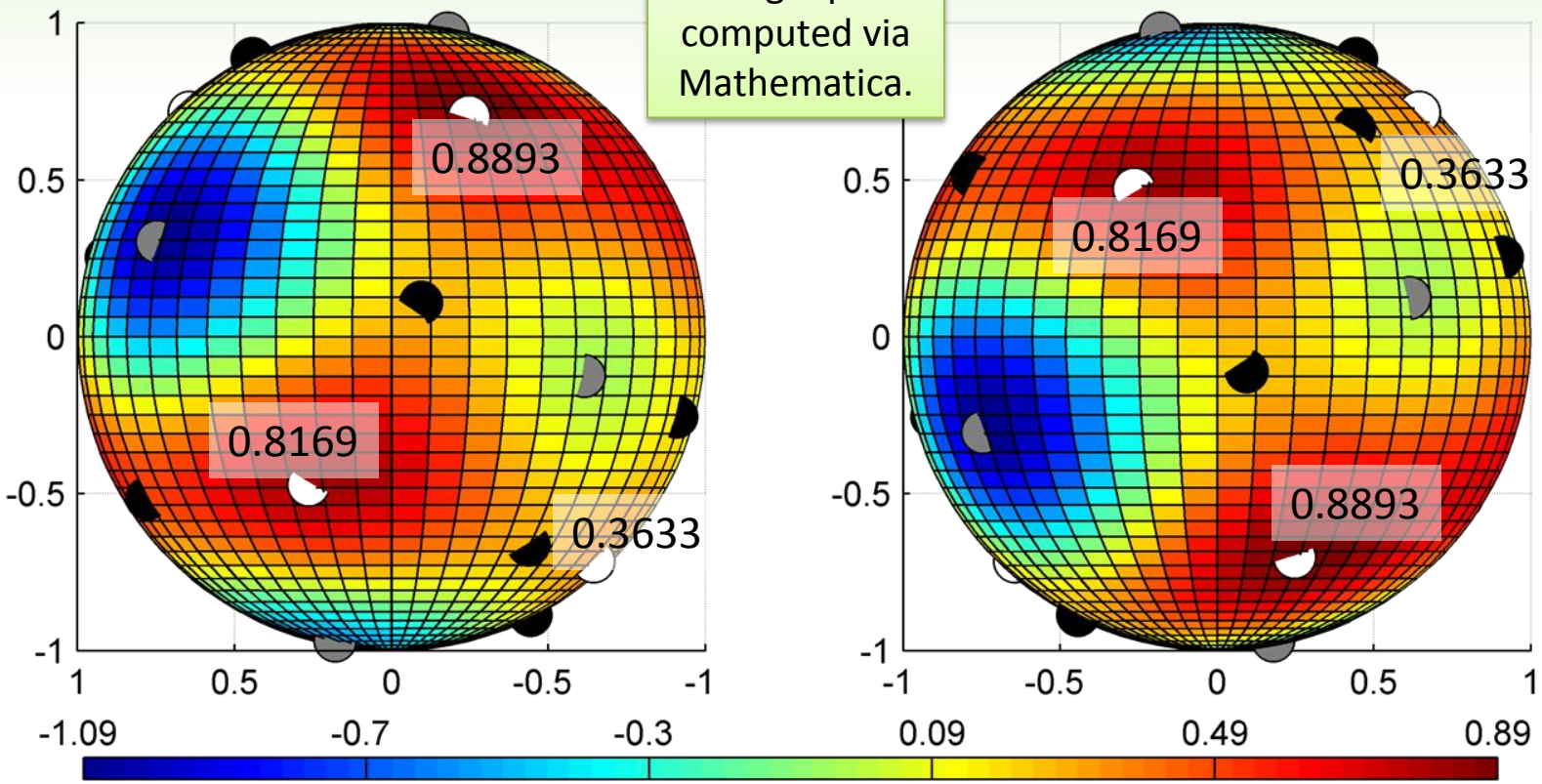
$$\text{subject to } \|\mathbf{x}\| = 1 \text{ and } (\lambda, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^n$$

Special case of generalized eigenpair with

$$\mathcal{B} = \mathcal{E}$$

Example: Z-eigenpairs

All eigenpairs computed via Mathematica.



White = Local Max, Gray = Local Min, Black = Saddle Point

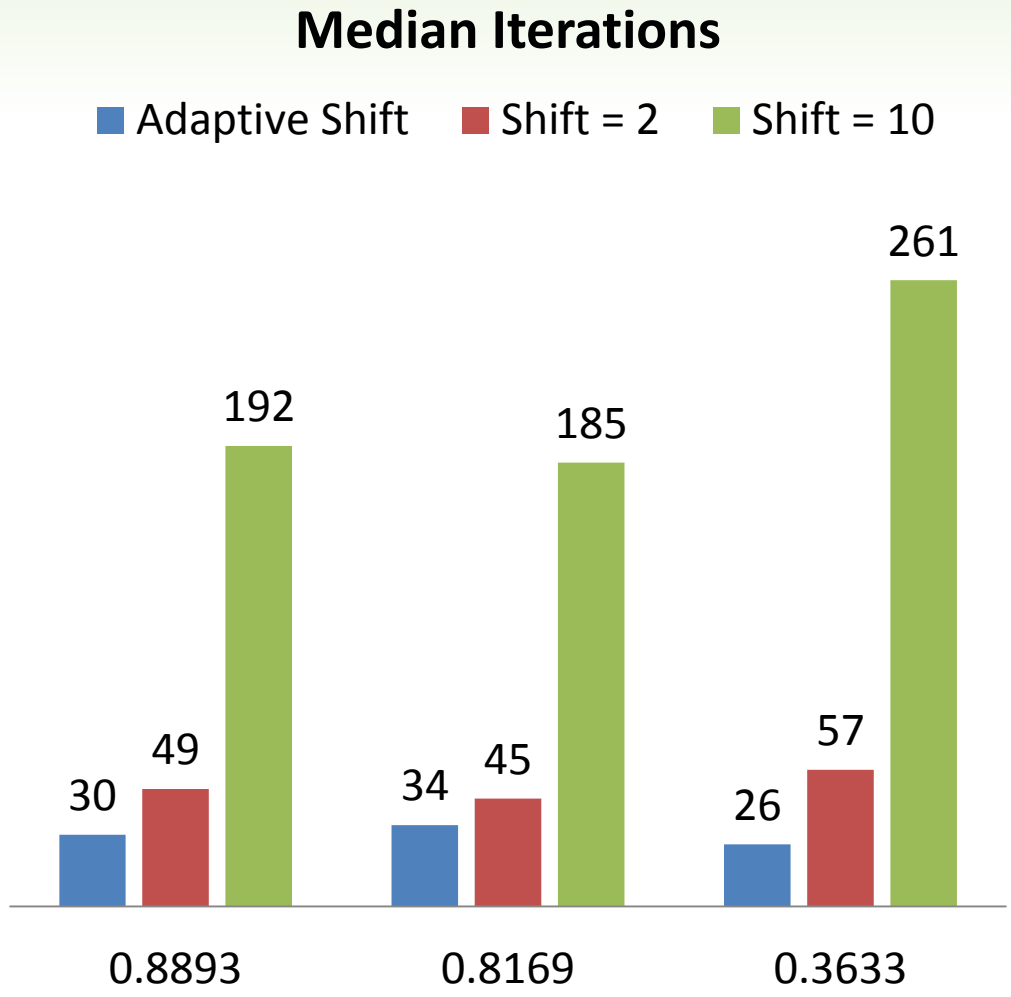
$$\mathcal{A} \in \mathbb{S}^{[4,3]}, \mathcal{B} = \mathcal{E} \in \mathbb{S}^{[4,3]}$$

Z-Eigenpairs: Adaptive Shift is Faster than Fixed Shift

Sandia National Laboratories

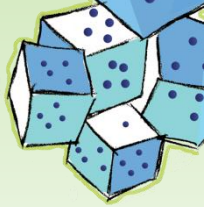
$\mathcal{A} \in \mathbb{S}^{[4,3]}$, $\mathcal{B} = \mathcal{E} \in \mathbb{S}^{[4,3]}$

- Comparison to SS-HOPM (fixed shift)
- 100 random starts
- Three local maxima
- Convergence: $|\lambda_{k+1} - \lambda_k| < 10^{-15}$
- Choosing shift too small (e.g., zero) means the iterations will *never* converge
- Choosing shift too large leads to slow convergence
- Adaptive shift is much faster than fixed shift!



Randomly Generated Generalized Eigenproblems

Sandia National Laboratories



Question: How do we generate a random *positive definite* tensor?

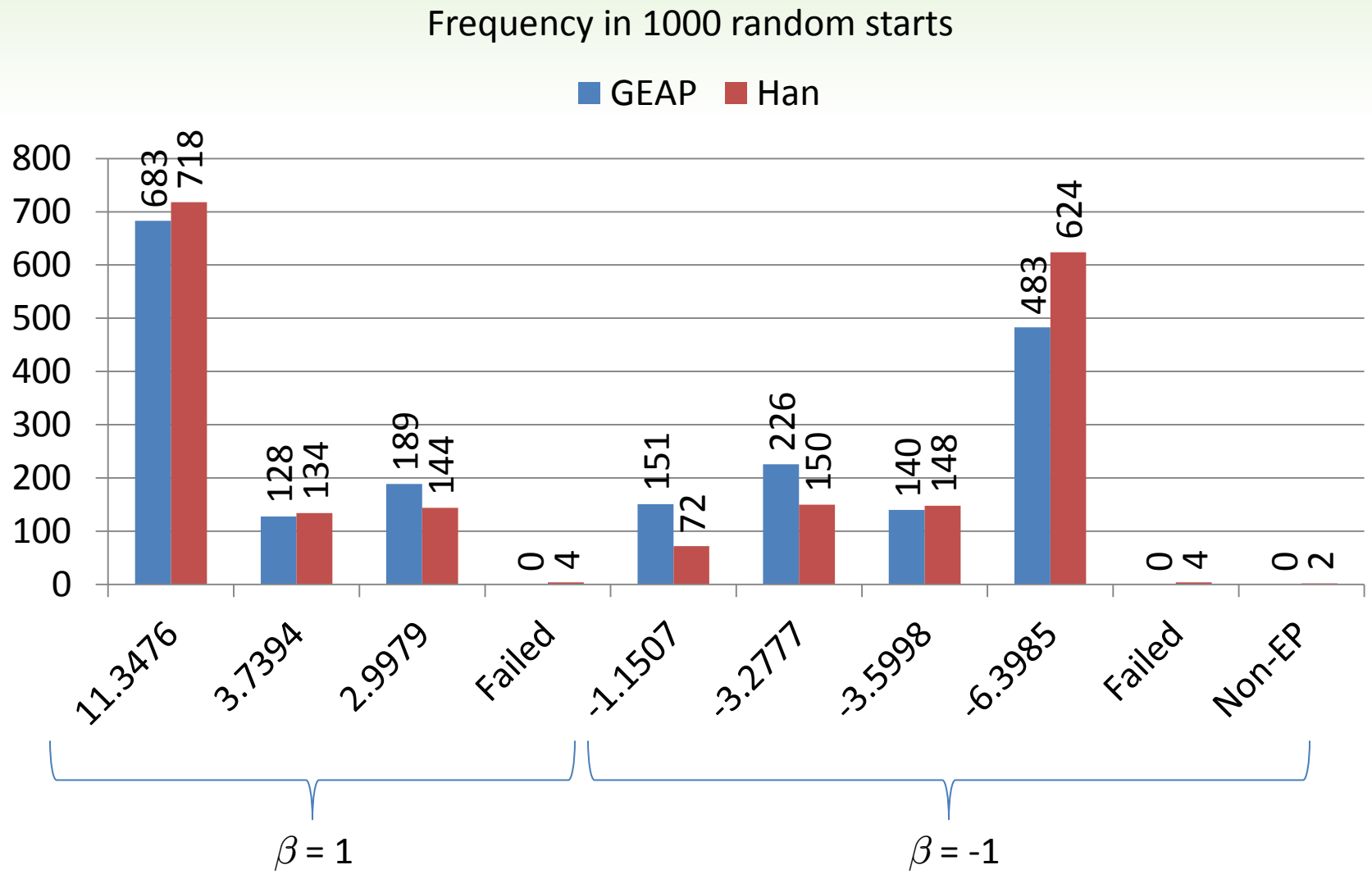
Theorem: Let S be a symmetric matrix. Let (μ, x) be an eigenpair of S (with ||x||=1). Define an m-way tensor B such that

$$b_{i_1 \cdots i_m} = \sum_{j_1=1}^{n} \cdots \sum_{j_m=1}^{n} e_{j_1 \cdots j_m} s_{i_1 j_1} \cdots s_{i_m j_m}$$

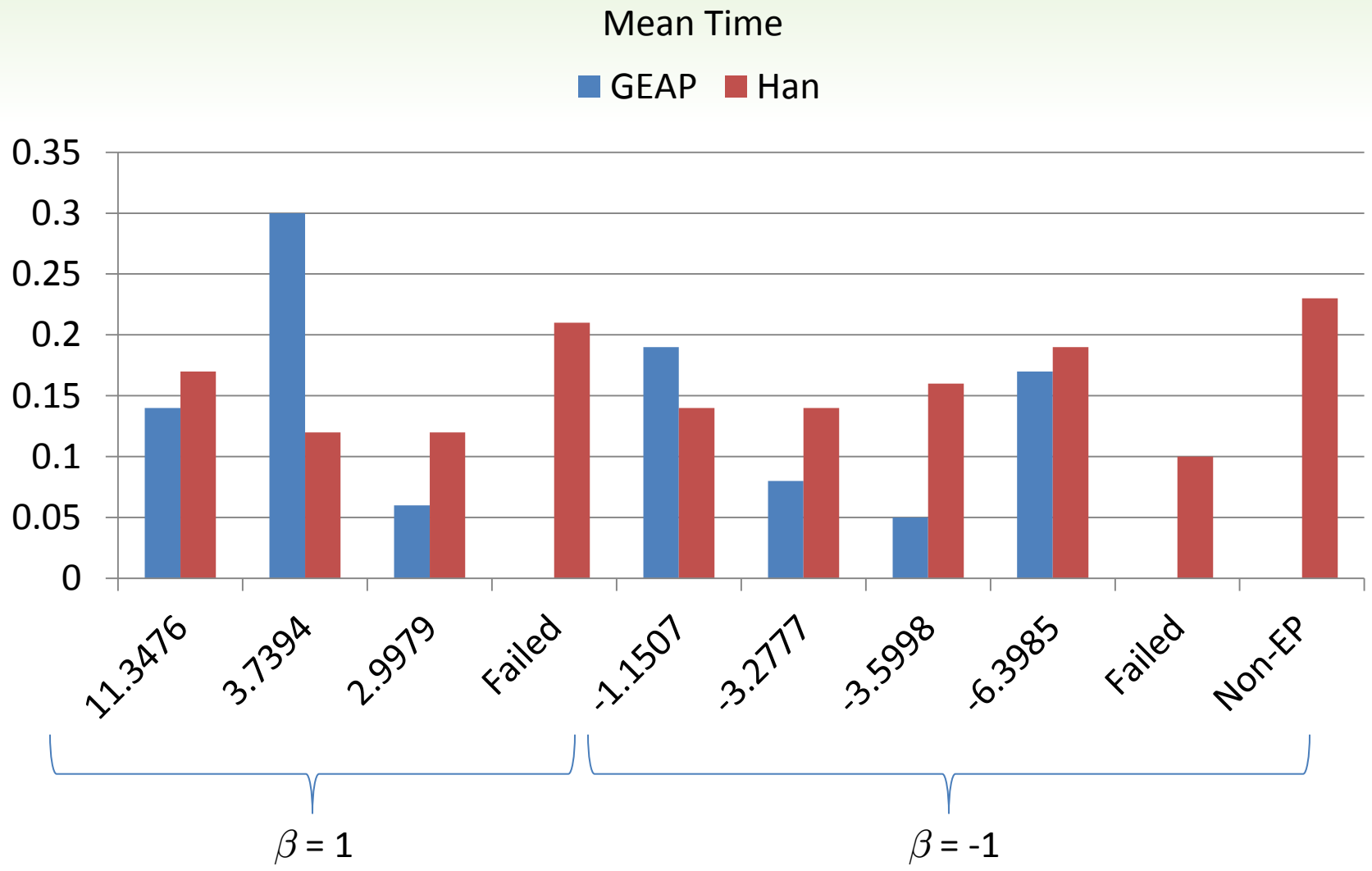
Then (λ,x) is an eigenpair of B with $\lambda = \mu^m$. Moreover, By$^m \geq \min_i (\mu_i)^m$.

Note: To keep B from being nearly indefinite, the smallest magnitude eigenvalue of S must be relatively large since it is raised to the mth power

Random Generalized Eigenpairs: GEAP versus Han's Method



Random Generalized Eigenpairs: GEAP versus Han's Method



Conclusions and Other Work

- Generalized tensor eigenproblem

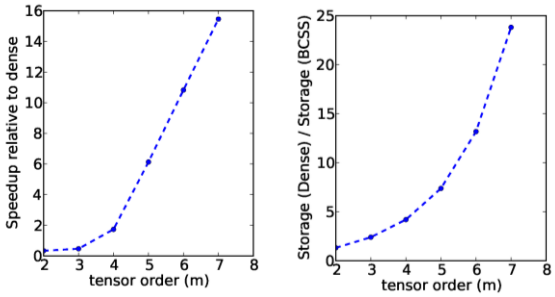
$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathcal{B}\mathbf{x}^{m-1}$$
$$\text{subject to } (\lambda, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^n$$

  - We reformulated GTEP as nonlinear program
- Monotonic algorithm for optimizing on the unit sphere
  - Any C2 function, not just GTEP
  - We propose adaptively choosing shift to ensure convexity
- Open questions
  - Scalability
  - Particular min/max eigenpair?
  - Saddle points?

For more info: Tammy Kolda, tgkolda@sandia.gov

- Kolda and Mayo, ***Shifted Power Method for Computing Tensor Eigenpairs***, SIMAX 32(4), 2011
- Kolda and Mayo, ***An Adaptive Shifted Power Method for Computing Generalized Tensor Eigenpairs***, arXiv:1401.1183

- Related Work
  - Symmetric computations (SISC, 2014)
    - With Schatz, van de Geijn, Low @ UT
    - Reduce storage of symmetric tensor by O(m!) without sacrificing performance



  - Poisson tensor decomposition
    - "Count" tensors with Chi (SIMAX, 2012)
    - Speed improvements with Hansen, Plantenga (arXiv:1304.4964, 2013)