Sandia National Laboratories
SAND2011-5041C

Computing Tensor Eigenvalues: Theory and Practice

Tamara G. Kolda & Jackson Mayo
Sandia National Labs
Grey Ballard
UC Berkeley & Sandia National Labs

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

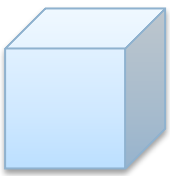
LABORATORY DIRECTED RESEARCH & DEVELOPMENT

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.

Real Tensor Eigenpairs

Qi (2005), Lim (2005)

$\mathcal{A}$ = symmetric m^{th} order n-dimensional real-valued tensor

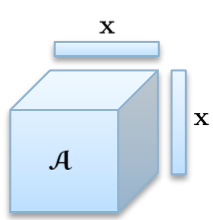


For m = 3, symmetry means $a_{ijk} = a_{jik} = a_{ikj} = a_{jki} = a_{kij} = a_{kji}$

We say that $\lambda \in \mathbb{R}$ is an eigenvalue if there exists $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathbf{x} \quad \text{and} \quad \mathbf{x}^{\mathrm{T}}\mathbf{x} = 1.$$

The vector $\mathbf{x}$ is called the eigenvector.



For m = 3, $\left(\mathcal{A}\mathbf{x}^{m-1}\right)_i = \sum_{jk} a_{ijk}\, x_j\, x_k$

Number of Eigenpairs

Cartwright & Sturmfels 2010

For a symmetric m^{th} order n-dimensional real-valued tensor, the # of distinct eigenpairs (real and complex) is $((m-1)^n-1)/(m-2)$

Note: For m odd, $(-\lambda,-\mathbf{x})$ is also an eigenpair.
For m even, $(\lambda,-\mathbf{x})$ is also an eigenpair.
These are not considered distinct.

Example: $\mathcal{A}$ is of size $2 \times 2 \times 2 \times 2 \quad (m = 4 \text{ and } n = 2)$
with $a_{ijkl} = 0$ except $a_{1111} = 1$ and $a_{2222} = -1$

$$\begin{aligned} \mathcal{A}\mathbf{x}^{m-1} &= \lambda \mathbf{x} \\ \mathbf{x}^{\mathsf{T}}\mathbf{x} &= 1 \end{aligned}$$

$$\begin{aligned} x_1^3 &= \lambda x_1 \\ -x_2^3 &= \lambda x_2 \\ x_1^2 + x_2^2 &= 1 \end{aligned}$$

[up to 4 solutions]

$$\lambda = 1, \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

or

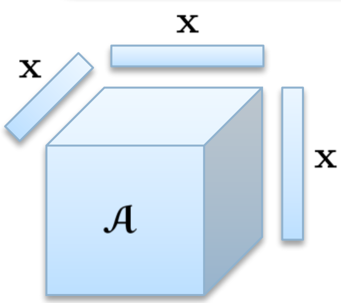
$$\lambda = -1, \mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

[2 real solutions]

Eigenpairs Correspond to Extrema of the Homogeneous Form

Lim (2005)

$$\max \;\; or \;\; \min \quad f(\mathbf{x}) \equiv \mathcal{A}\mathbf{x}^m$$
$$\text{s.t.} \quad \frac{1}{2}(\|\mathbf{x}\|^2 - 1) = 0$$



For m = 3,

$$\mathcal{A}\mathbf{x}^m = \sum_{ijk} a_{ijk}\, x_i\, x_j\, x_k$$

Lagrangian:

$$\mathcal{L}(\mathbf{x}, \mu) = \mathcal{A}\mathbf{x}^m + \mu \frac{1}{2}(\|\mathbf{x}\|^2 - 1)$$

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu) = m\mathcal{A}\mathbf{x}^{m-1} + \mu\mathbf{x}$$

A real eigenpair is any KKT point of the constrained homogeneous form. (Analogous to the matrix case.)

KKT Conditions:

$$m\mathcal{A}\mathbf{x}^{m-1} + \mu\mathbf{x} = 0 \text{ and } \|\mathbf{x}\| = 1$$

iff

Eigenpair:

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

$$(\text{with } \lambda = -\mu/m)$$

Eigenpairs Correspond to Best Symmetric Rank-1 Approximation

Eigenvalue Problem

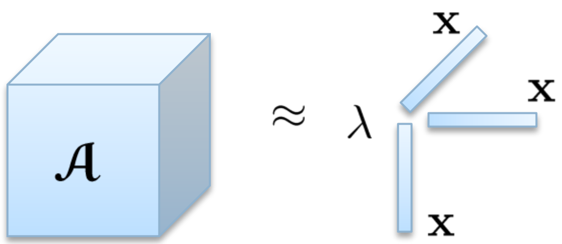
$$\mathcal{A}\mathbf{x}^{m-1} = \lambda\mathbf{x}$$
$$\mathbf{x}^{\mathsf{T}}\mathbf{x} = 1$$

Extrema of Constrained Homogenous Form

$$\max \quad f(\mathbf{x}) \equiv \mathcal{A}\mathbf{x}^m$$
$$\text{s.t.} \quad \frac{1}{2}(\|\mathbf{x}\|^2 - 1) = 0$$

Best Rank-1 Approximation

$$\min \quad \|\mathcal{A} - \lambda\mathbf{x} \circ \mathbf{x} \circ \cdots \circ \mathbf{x}\|^2$$
$$\text{s.t.} \quad \lambda = \mathcal{A}\mathbf{x}^m, \ \|\mathbf{x}\| = 1$$



Symmetric Higher-Order Power Method (S-HOPM)

De Lathauwer, De Moor, Vandewalle 2000

S-HOPM

For $k = 1, 2, \ldots$

$$\mathbf{x}_{k+1} = \mathcal{A}\mathbf{x}_k^{m-1} / \|\mathcal{A}\mathbf{x}_k^{m-1}\|$$

$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^{m}$$

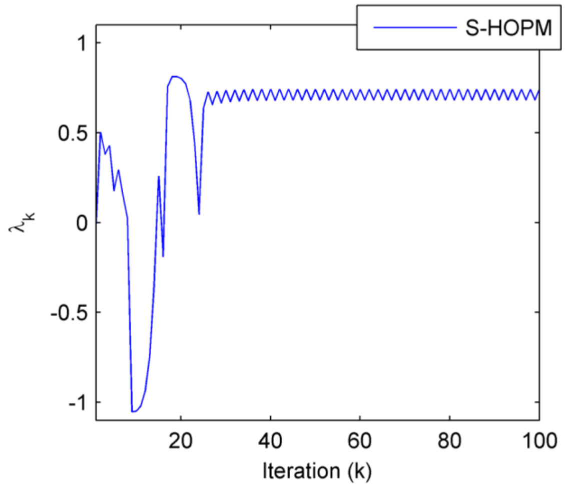
- Symmetric analogue of convergent Higher-Order Power Method
- Not guaranteed to converge
  - May diverge
  - May have chaotic behavior
  - *But sometimes works really well!*

Failure Example [Kofidis & Regalia 2002]

- 3 x 3 x 3 x 3 Symmetric Tensor

$$\begin{aligned}
a_{1111} &= 0.2883, & a_{1112} &= -0.0031, & a_{1113} &= 0.1973,\\
a_{1122} &= -0.2485, & a_{1123} &= -0.2939, & a_{1133} &= 0.3847,\\
a_{1222} &= 0.2972, & a_{1223} &= 0.1862, & a_{1233} &= 0.0919,\\
a_{1333} &= -0.3619, & a_{2222} &= 0.1241, & a_{2223} &= -0.3420,\\
a_{2233} &= 0.2127, & a_{2333} &= 0.2727, & a_{3333} &= -0.3054.
\end{aligned}$$

- Optimum: $|\lambda|$ = 1.09
- S-HOPM fails on this problem for every starting point we tried



Shifted S-HOPM (SS-HOPM) is Convergent

$$f(\mathbf{x}) = \mathcal{A}\mathbf{x}^m$$

$$\hat{f}(\mathbf{x}) \equiv f(\mathbf{x}) + \alpha(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{m/2}$$

S-HOPM

For $k = 1, 2, \ldots$

$$\mathbf{x}_{k+1} = \frac{\mathcal{A}\mathbf{x}_k^{m-1}}{\|\mathcal{A}\mathbf{x}_k^{m-1}\|}$$

$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^m$$

SS-HOPM

For $k = 1, 2, \ldots$

$$\mathbf{x}_{k+1} = \frac{\mathcal{A}\mathbf{x}_k^{m-1} + \alpha\mathbf{x}_k}{\|\mathcal{A}\mathbf{x}_k^{m-1} + \alpha\mathbf{x}_k\|}$$

$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^m$$

In the context of ICA, using a shift has previously been proposed by Regalia and Kofidis (2005) and Erdogen (2009).

SS-HOPM Finds Real Eigenpairs

100 Random Starting Points

S-HOPM

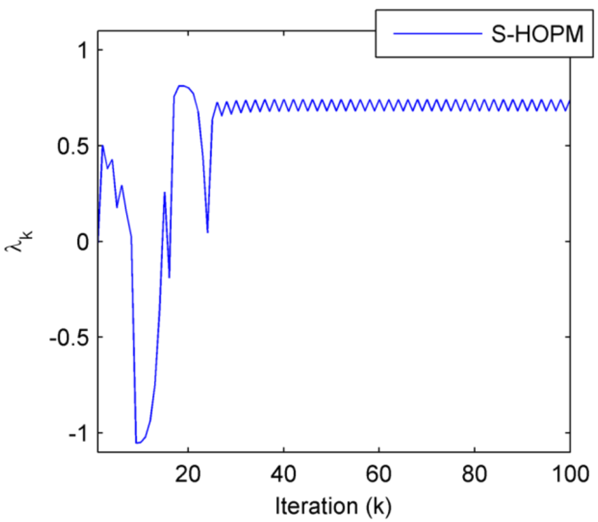
No Convergence

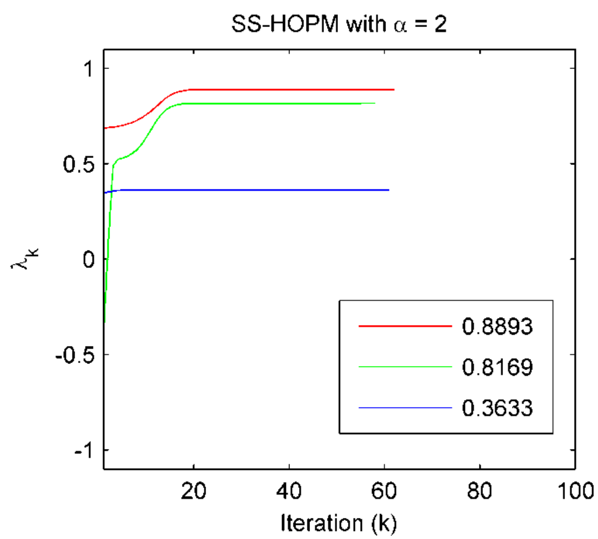
SS-HOPM with α = 2

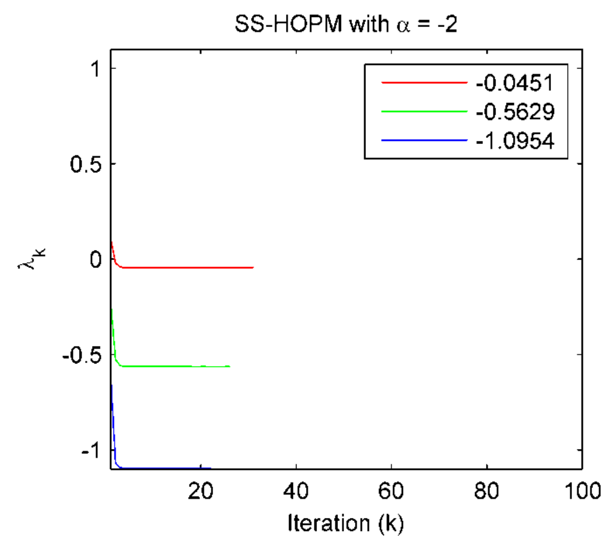
Occurrences	λ	Median Its.
46	0.8893	63
24	0.8169	52
30	0.3633	65

SS-HOPM with α = -2

Occurrences	λ	Median Its.
15	-0.0451	35
40	-0.5629	23
45	-1.0954	23







SS-HOPM as a Fixed Point Iteration

SS-HOPM

For $k = 1, 2, \ldots$

$$\mathbf{x}_{k+1} = \frac{\mathcal{A}\mathbf{x}_k^{m-1} + \alpha \mathbf{x}_k}{\|\mathcal{A}\mathbf{x}_k^{m-1} + \alpha \mathbf{x}_k\|}$$

$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^{m}$$

$$\phi(\mathbf{x};\alpha) = \frac{\mathcal{A}\mathbf{x}^{m-1} + \alpha \mathbf{x}}{\|\mathcal{A}\mathbf{x}^{m-1} + \alpha \mathbf{x}\|}$$

For our problem, any fixed point is an eigenvector and vice versa.

Fixed Point of ϕ: $\phi(\mathbf{x};\alpha) = \mathbf{x}$

Let $J(\mathbf{x};\alpha)$ denote the n x n Jacobian of $\phi(\mathbf{x};\alpha)$.

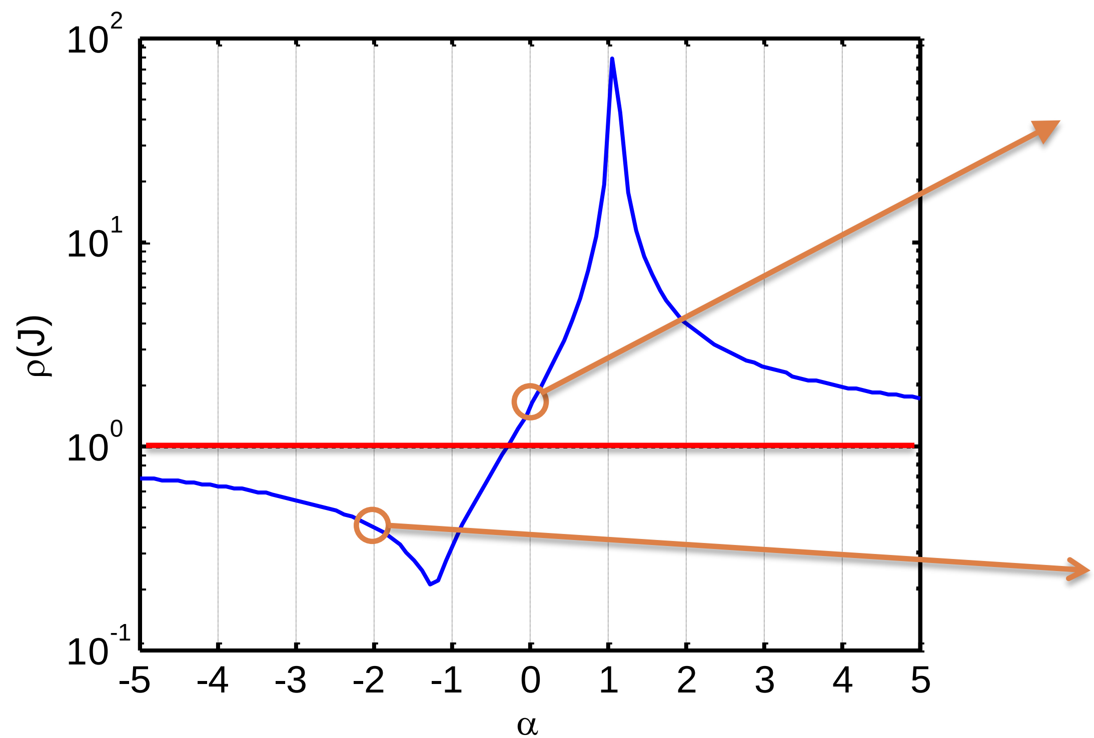
Fact 1: $\mathbf{x}$ is an attracting fixed point if $\sigma \equiv \rho(J(\mathbf{x};\alpha)) < 1$.

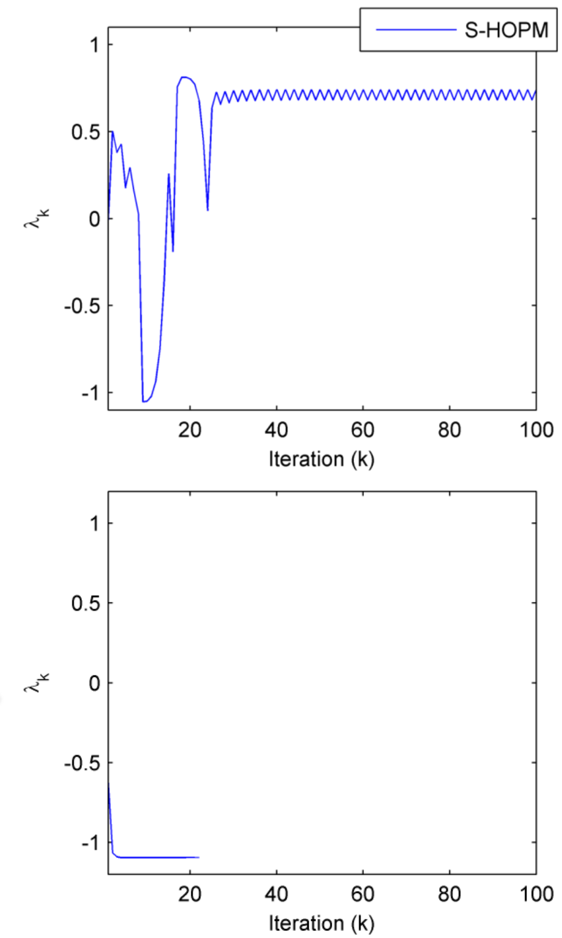
Fast 2: The convergence is linear with rate σ (smaller is faster).

Convergence Explained via Fixed Point Analysis

At eigenpair (λ,$\mathbf{x}$): $\mathbf{J}(\mathbf{x};\alpha) = \dfrac{(m-1)(\mathcal{A}\mathbf{x}^{m-2} - \lambda\mathbf{x}\mathbf{x}^{\mathrm{T}}) + \alpha(\mathbf{I} - \mathbf{x}\mathbf{x}^{\mathrm{T}})}{\lambda + \alpha}$

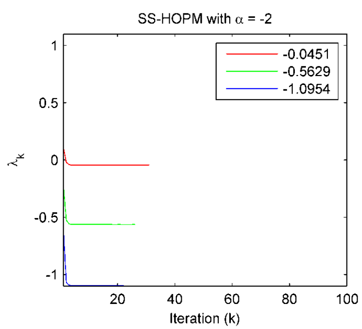
Spectral radius of Jacobian for eigenvector corresponding to λ = -1.09

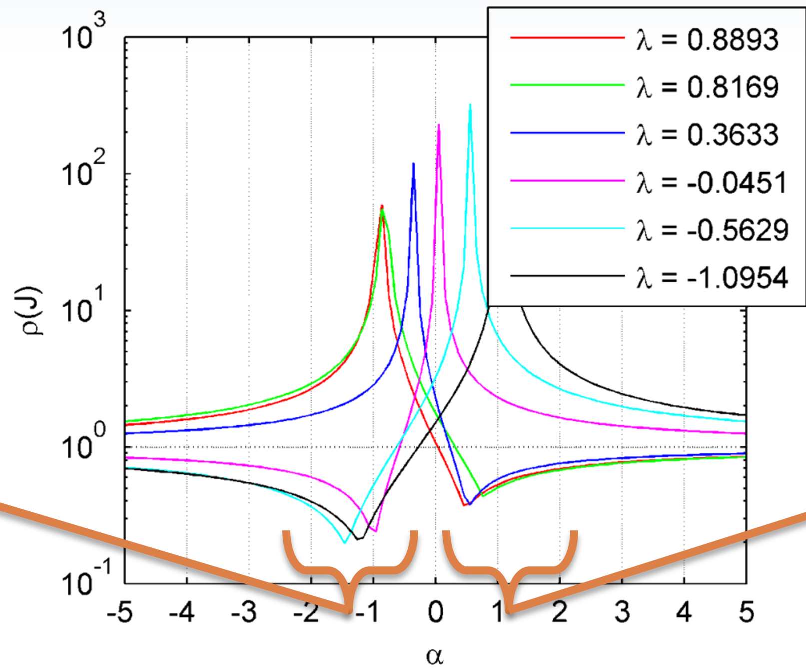




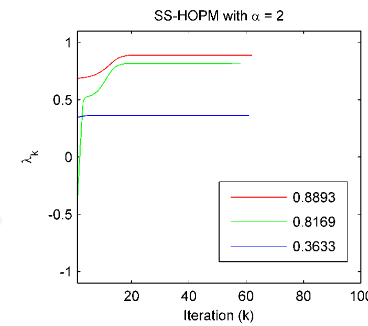
How Choice of α Impacts SS-HOPM

Negative values of α lead to local minima.





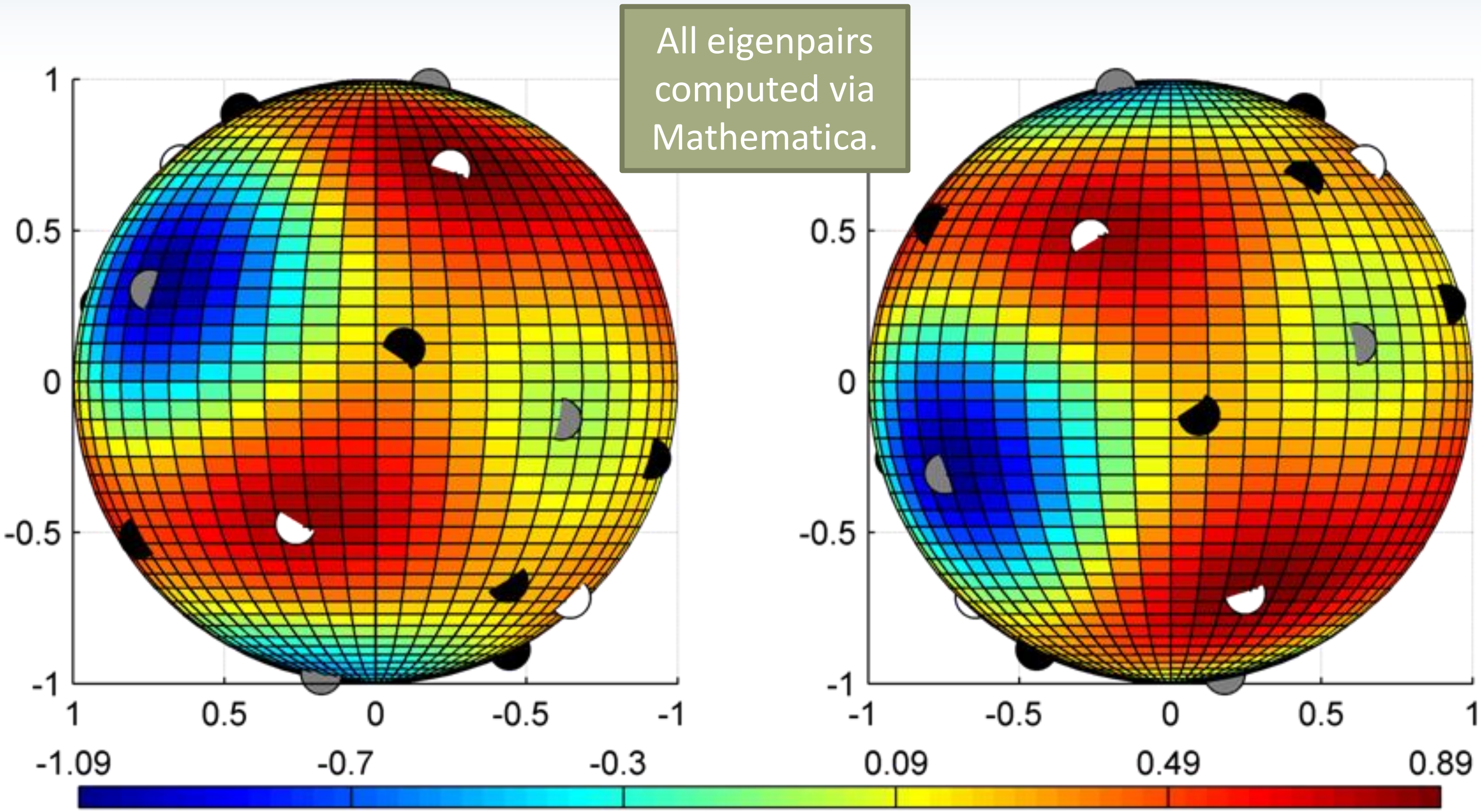
Positive values of α lead to local minima.



Larger values of α slow convergence.

Some eigenvectors <u>never</u> have a spectral radius less than one; SS-HOPM cannot find those eigenvectors.

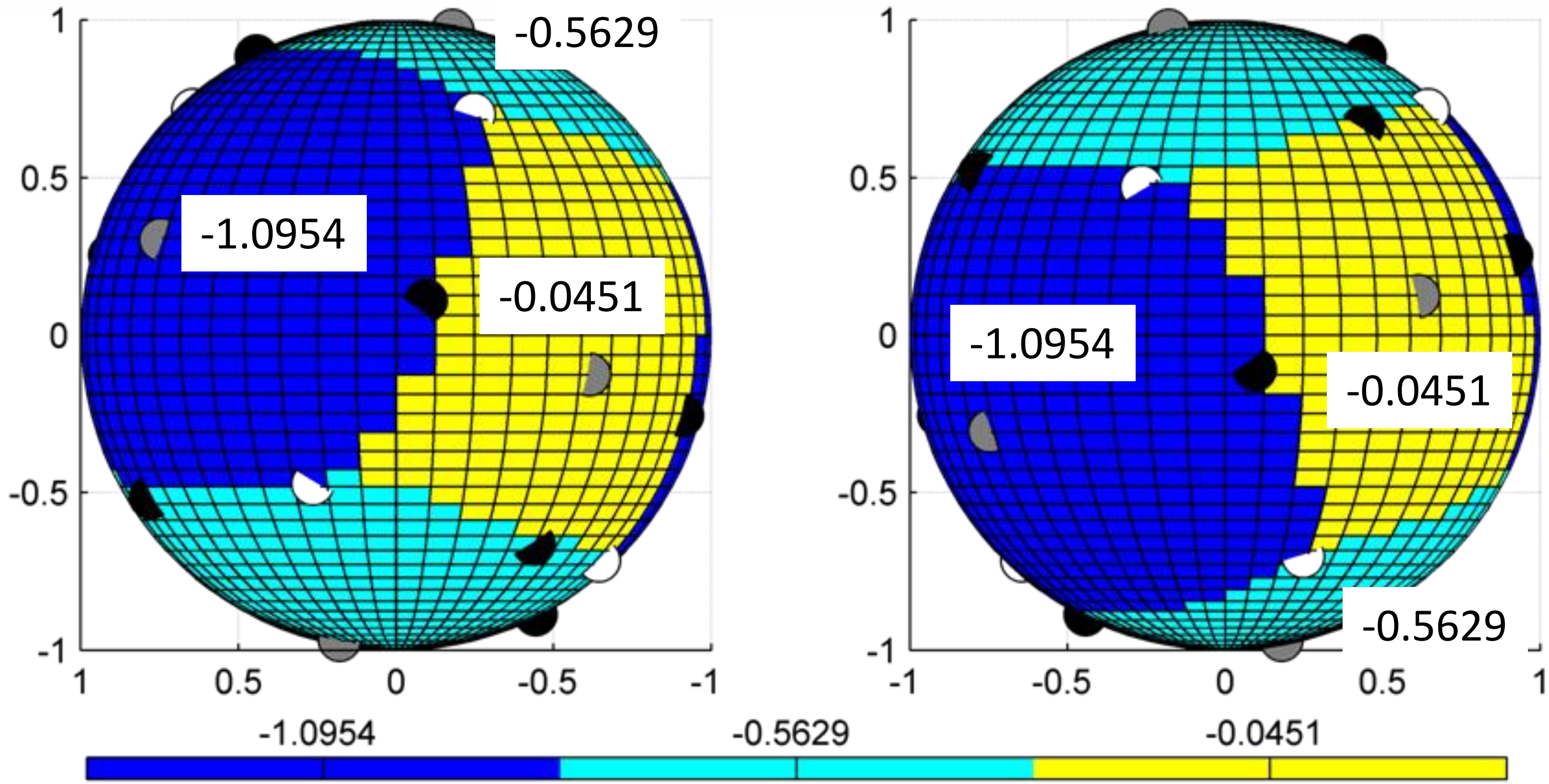
Visualization of Eigenvectors



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

Basins of Attraction for α = -2

Limit points correspond to local minima of function.

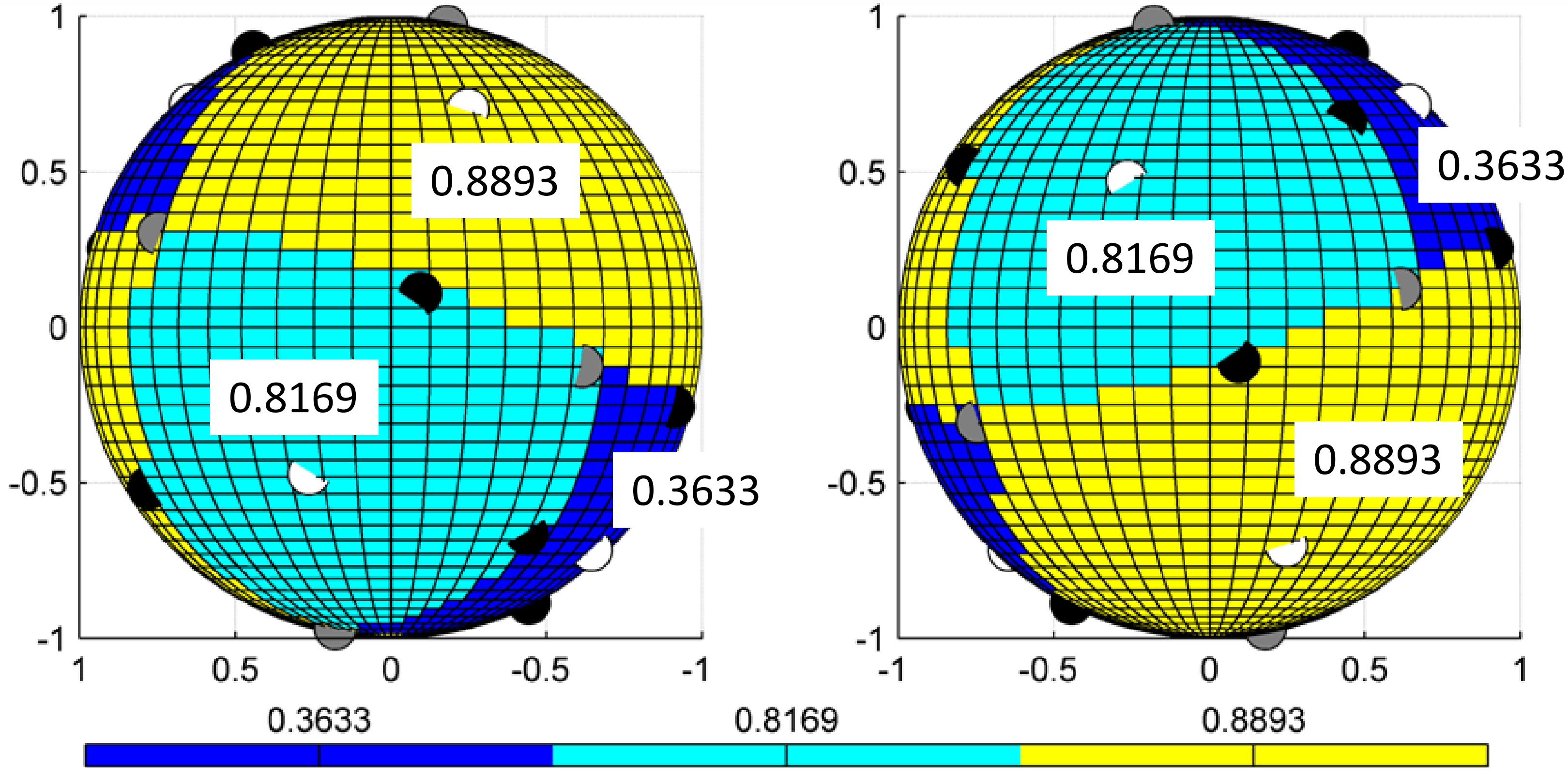


White = Local Max, Gray = Local Min,
Black = Saddle Point

Occurrences	λ
15	-0.0451
40	-0.5629
45	-1.0954

Basins of Attraction for $\alpha = 2$

Limit points correspond to local maxima of function.

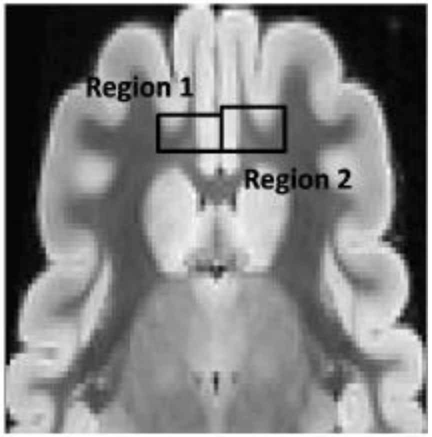


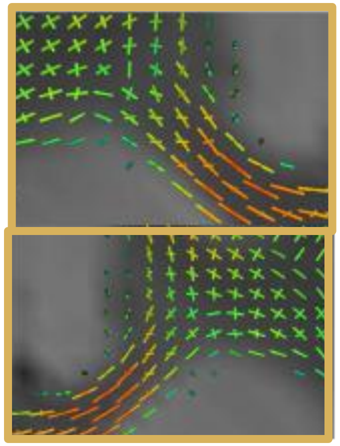
White = Local Max, Gray = Local Min,
Black = Saddle Point

Occurrences	λ
46	0.8893
24	0.8169
30	0.3633

Application to Brain Imaging

- Diffusion Tensor MRI (DT-MRI) is generally used to infer white matter connectivity in the brain
  - Limited resolution
  - Reduces to an eigenproblem
- High Angular Resolution Diffusion Imaging (HARDI)
  - Higher resolution
  - Reduces to a tensor problem
- U. Utah: F. Jiao, Y. Gur, C. Johnson, S. Joshi, *Detection of Crossing White Matter Fibers with High-Order Tensors and Rank-k Decompositions*, In Proc. Information Processing in Medical Imaging (IPMI), 2011
  - Focus on challenge of small crossing angle, a.k.a., high congruence
- Thanks to Yaniv Gur and Fiang Jiao for providing us sample data

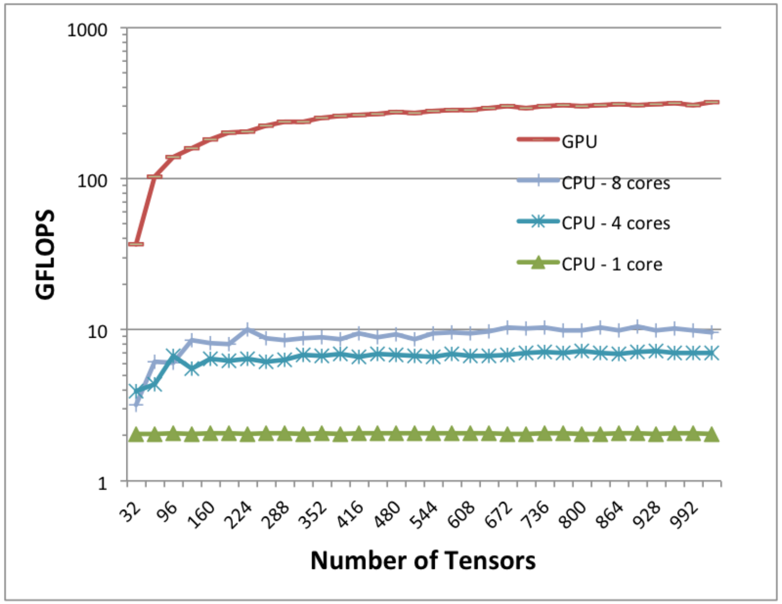




SS-HOPM on a GPU gets 317 Gflops/s

- Motivating application
  - Diffusion-weighted MRI
  - Need to solve millions of 3x3x3x3 tensor eigen-problems
  - Use 128 starting vectors per tensor
- New storage format for symmetric tensors
  - Storage ~ (n^m) / m!
  - Cost of $\mathbf{A}\mathbf{x}^m$ ~ (n^m) / $(m$-1)!
  - Cost of $\mathbf{A}\mathbf{x}^{(m\text{-}1)}$ ~ (mn^m) / $(m$-1)!
- GPU implementation
  - One "thread block" per tensor
  - One "thread" per starting point
  - Loop unrolling gives up to 20x speed-up

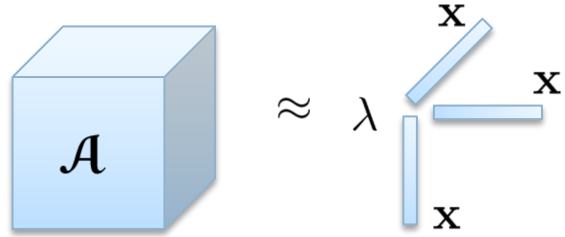
Compute Engine	Gflops/s
Intel Nahelem (1 core)	2.05 (9% peak)
Intel Nahelem (4 cores)	7.07 (8% peak)
nVidia Tesla 2050 (Fermi) 16 streaming multiprocessors (SMPs) 32 cores per SMP	317.83 (31% peak)



But… Brain Imaging Application Actually Needs Best Sym. Rank-K Approximation

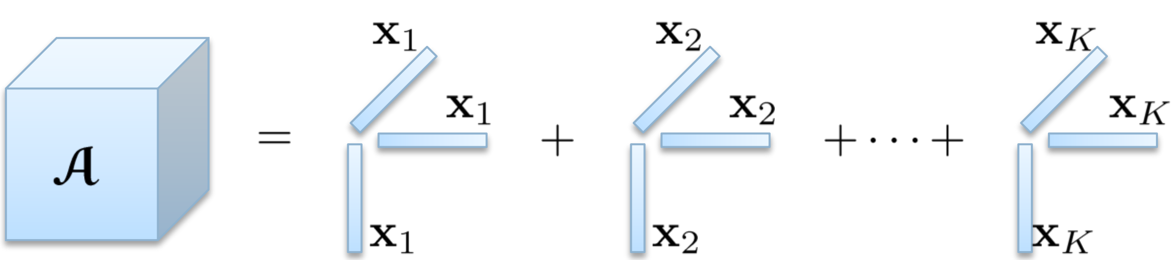
Best Symmetric Rank-1 Approximation

$$\min \quad \|\mathcal{A} - \lambda \mathbf{x} \circ \mathbf{x} \circ \cdots \circ \mathbf{x}\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\| = 1$$



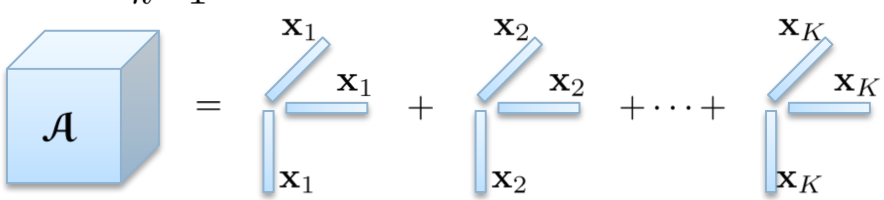
Best Symmetric Rank-K Approximation

$$\min \quad \|\mathcal{A} - \sum_k \lambda_k \mathbf{x}_k \circ \mathbf{x}_k \circ \cdots \circ \mathbf{x}_k\|^2 \quad \text{s.t.} \quad \|\mathbf{x}_k\| = 1 \forall k$$



Optimization Formulation

$$\mathcal{A} = \sum_{k=1}^{K} \mathbf{x}_k \circ \mathbf{x}_k \circ \mathbf{x}_k = [\![\mathbf{X}, \mathbf{X}, \mathbf{X}]\!]$$



Objective Function

$$f(\mathbf{X}) = \frac{1}{2} \|\mathcal{A} - [\![\mathbf{X}, \cdots, \mathbf{X}]\!]\|^2$$

$$= \frac{1}{2}\|\mathcal{A}\|^2 - \sum_{k=1}^{K} \mathcal{A}\mathbf{x}_k^m + \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \left(\mathbf{x}_i^{\mathsf{T}} \mathbf{x}_j\right)^m$$

Gradient

$$\frac{\partial f}{\partial \mathbf{x}_k} = -m \mathcal{A} \mathbf{x}_k^{m-1} + m \sum_{i=1}^{K} (\mathbf{x}_i^{\mathsf{T}} \mathbf{x}_k)^{(m-1)} \mathbf{x}_i$$

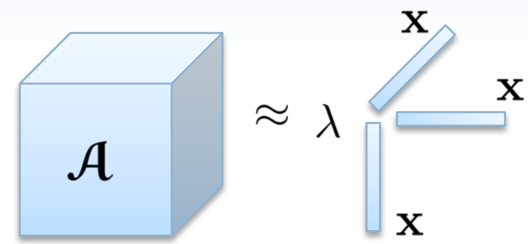
Gradient in Matrix Format

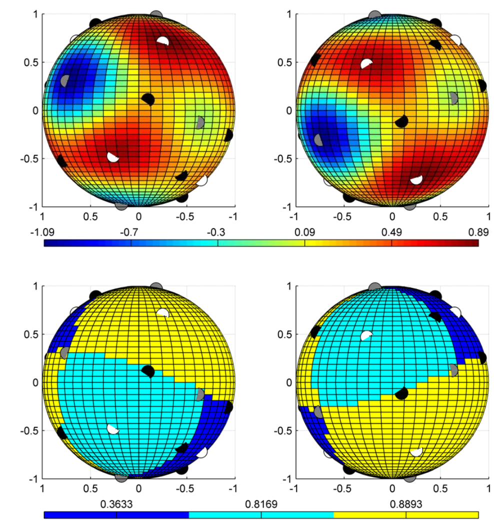
$$\nabla f(\mathbf{X}) = -m \mathbf{A}_{(1)} \underbrace{(\mathbf{X} \odot \mathbf{X} \odot \cdots \odot \mathbf{X})}_{(m-1) \text{ times}} + m \mathbf{X} (\mathbf{X}^{\mathsf{T}} \mathbf{X})^{[m-1]}$$

- Direct optimization
  - Motivated by CP-OPT and similar approaches
- Benefits
  - Can use any optimization method (we use NCG)
  - Extensible to higher-order methods
- Disadvantages
  - Can require extensive parameter tuning
  - May converge to local minimum

Conclusions and Future Work

- SS-HOPM is a convergent method for finding tensor eigenvalues
  - Corresponds to best symmetric rank-1 approximation problem
  - There is also a version for finding complex eigenpairs
  - PROS: Easily implemented, parallelized
  - CONS: Cannot find *all* real eigenpairs
- More generally interested in best symmetric rank-K approximation
  - See talk from Householder 2011





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

Kolda and Mayo, ***Shifted Power Method for Computing Tensor Eigenpairs***, SIMAX (to appear)
Ballard, Kolda, and Plantenga, ***Efficiently Computing Tensor Eigenvalues on a GPU***, PDSEC-11

Backup Slides

Interesting result because operating on unit sphere which is not convex.

S-HOPM Analysis

Kofidis and Regalia (2002)

$$\max \quad f(\mathbf{x}) \equiv \mathcal{A}\mathbf{x}^m$$
$$\text{s.t.} \quad \|\mathbf{x}\| = 1$$

- Theorem: S-HOPM λ_k converges to eigenvalue if $f(\mathbf{x})$ is convex or concave on unit ball
- Key Lemma: Assume $f(\mathbf{x})$ convex on unit ball and let $\mathbf{v}$ be such that $\|\mathbf{v}\|$=1.
  - If $\mathbf{w} = \nabla f(\mathbf{v})/\|\nabla f(\mathbf{v})\|$
  - Then $f(\mathbf{w}) \geq f(\mathbf{v})$
- Importance: If $f(\mathbf{x})$ is convex, then S-HOPM has $\lambda_{k+1} \geq \lambda_k$ for all k

S-HOPM

For $k = 1, 2, \ldots$

$$\mathbf{x}_{k+1} = \mathcal{A}\mathbf{x}_k^{m-1} / \|\mathcal{A}\mathbf{x}_k^{m-1}\|$$
$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^m$$

Assumes m even.
Let $l = m/2$.

$$f(\mathbf{x}) = \mathcal{A}\mathbf{x}^m = (\underbrace{\mathbf{x} \otimes \cdots \otimes \mathbf{x}}_{l \text{ times}})^{\mathrm{T}} \mathbf{A} (\underbrace{\mathbf{x} \otimes \cdots \otimes \mathbf{x}}_{l \text{ times}})$$

$$\nabla^2 f(\mathbf{x}) = (\mathbf{I} \otimes \underbrace{\mathbf{x} \otimes \cdots \otimes \mathbf{x}}_{l-1 \text{ times}})^{\mathrm{T}} \mathbf{A} (\mathbf{I} \otimes \underbrace{\mathbf{x} \otimes \cdots \otimes \mathbf{x}}_{l-1 \text{ times}})$$

Forcing Convexity with a Shift

A quadratic function is convex if all the eigenvalues of A are positive (and concave if all are negative).

$$\begin{aligned} \max \quad & f(\mathbf{x}) \equiv \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} \max \quad & \hat{f}(\mathbf{x}) \equiv \mathbf{x}^{\mathrm{T}}(\mathbf{A} + \alpha\mathbf{I})\mathbf{x} \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned}$$

An analogue for even-order tensors:

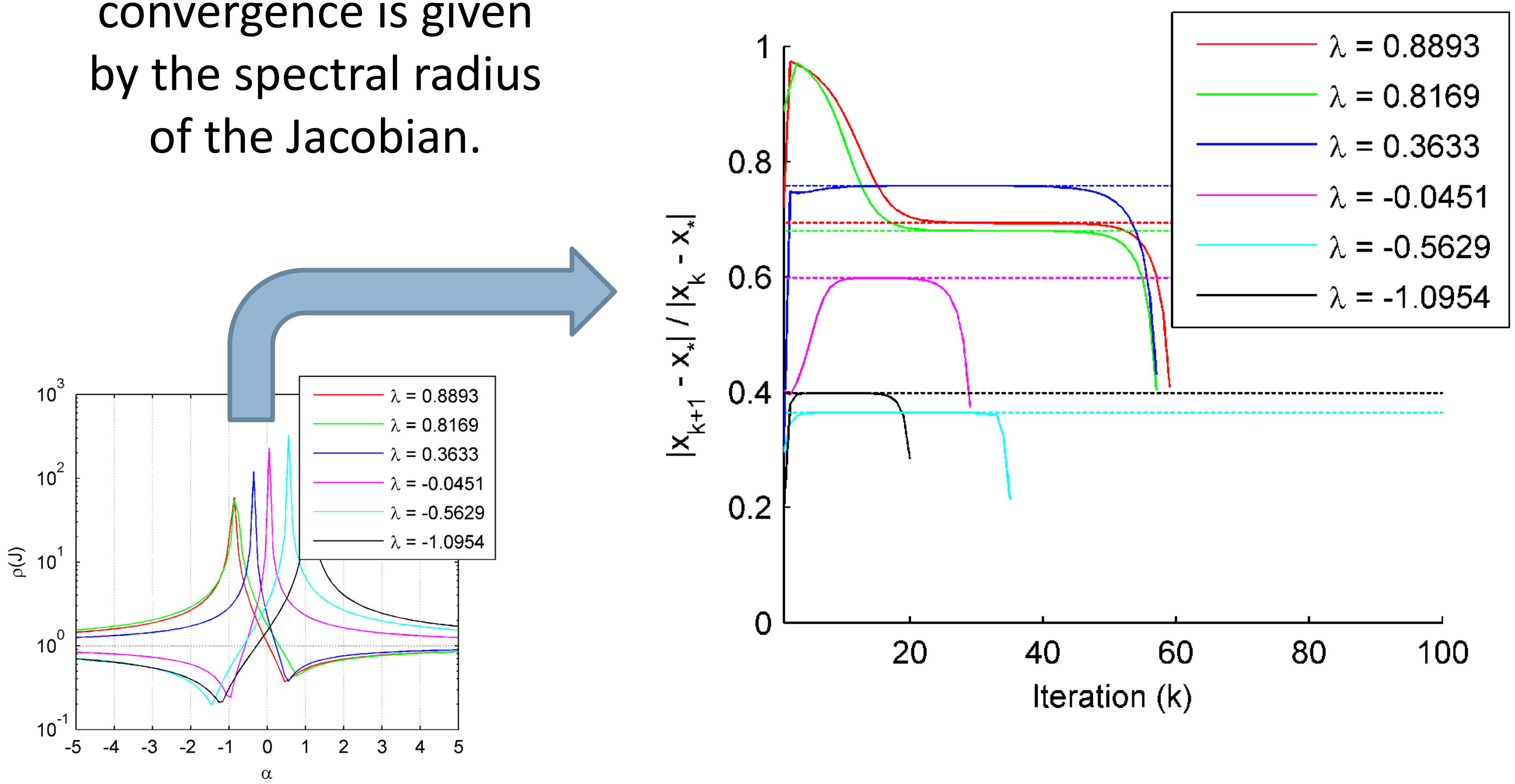
$$\begin{aligned} \max \quad & f(\mathbf{x}) \equiv \mathcal{A}\mathbf{x}^{m} \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} \max \quad & \hat{f}(\mathbf{x}) \equiv (\mathcal{A} + \alpha\mathcal{E})\,\mathbf{x}^{m} \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned}$$

Identity Tensor

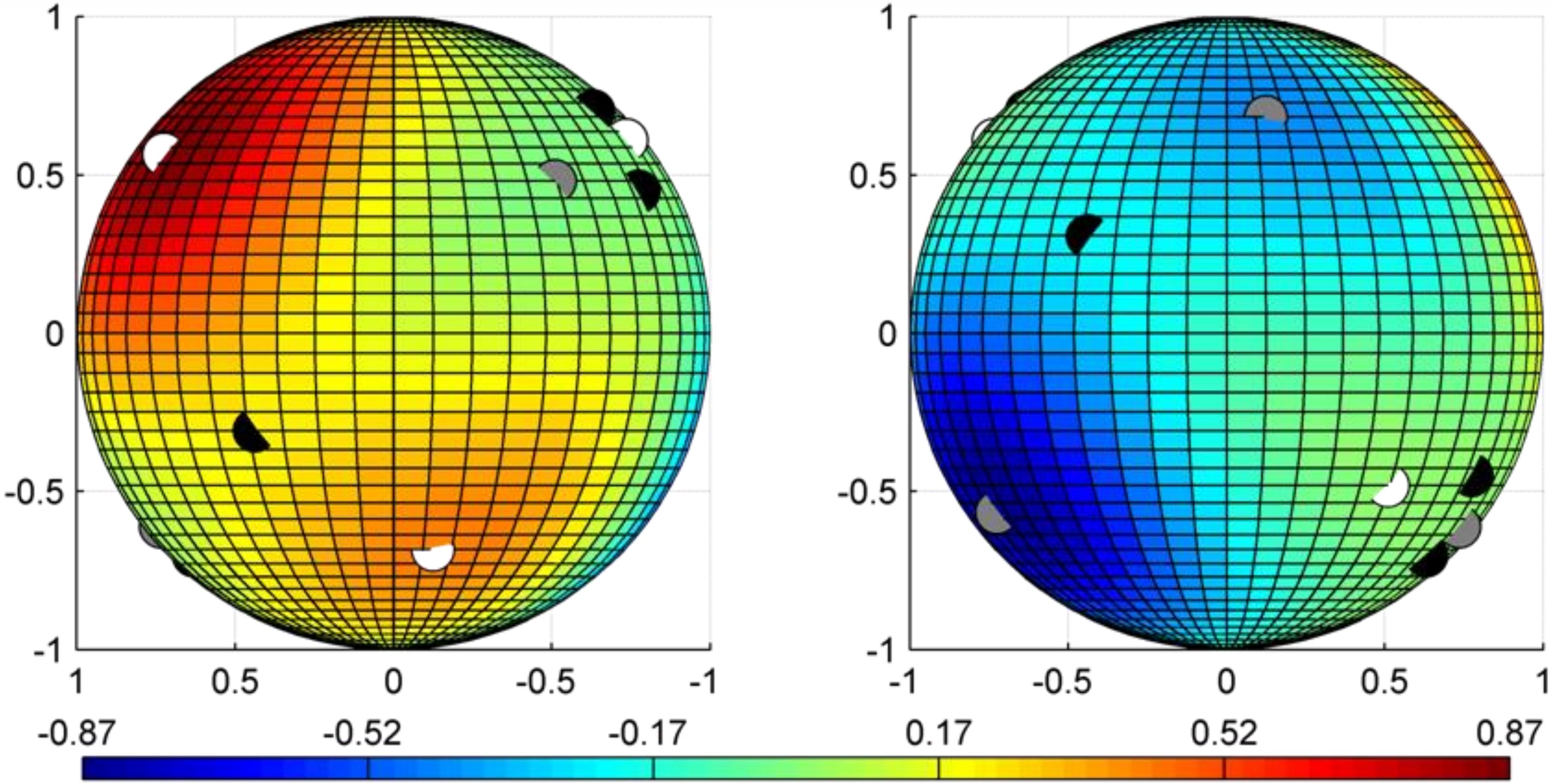
$$\mathcal{E}\mathbf{x}^{m-1} = \mathbf{x} \quad \forall \mathbf{x}$$

Rate of Convergence

The rate of convergence is given by the spectral radius of the Jacobian.



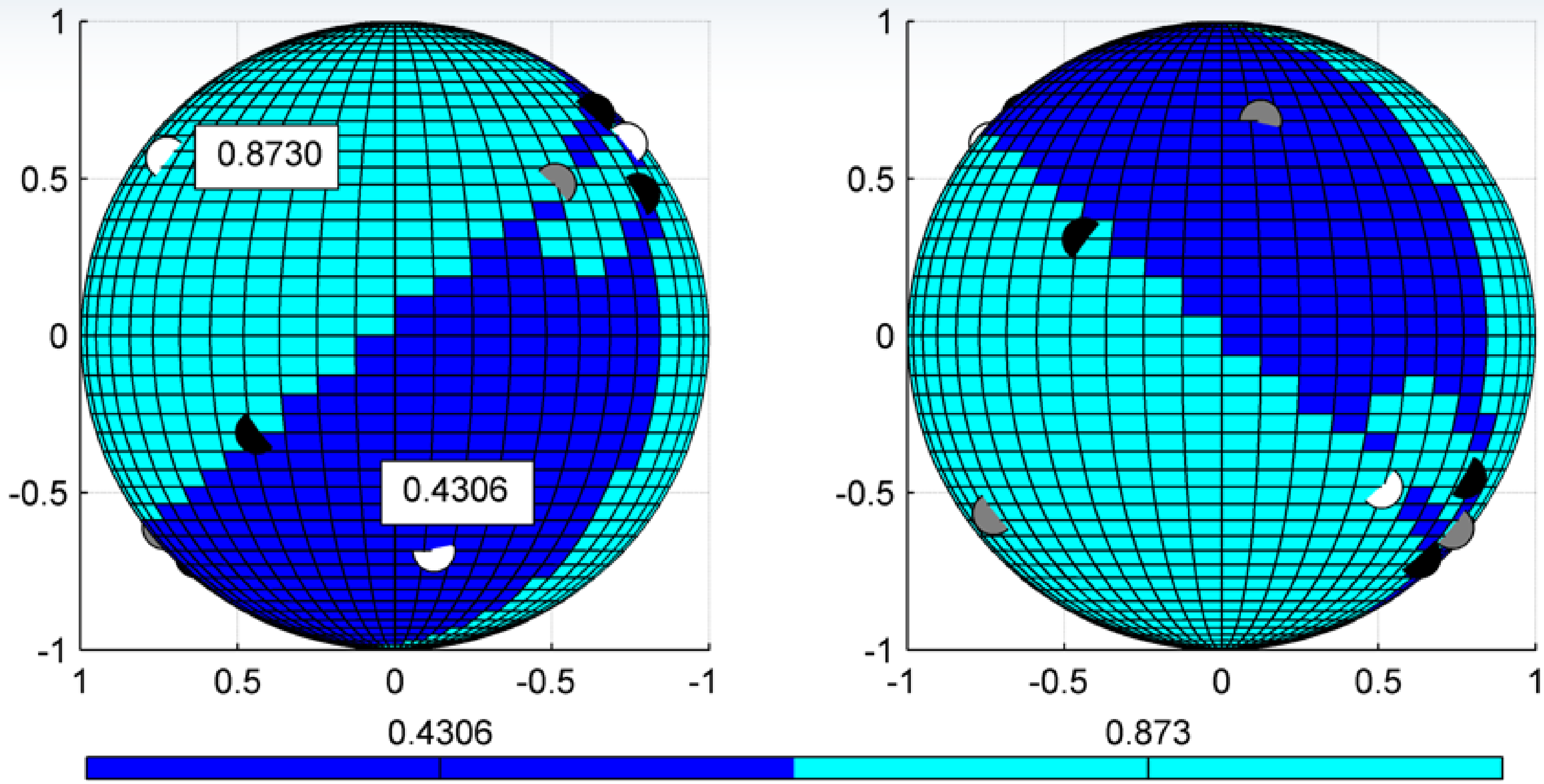
Third-Order Example



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

Sandia National Laboratories

Jacobian explains Convergence

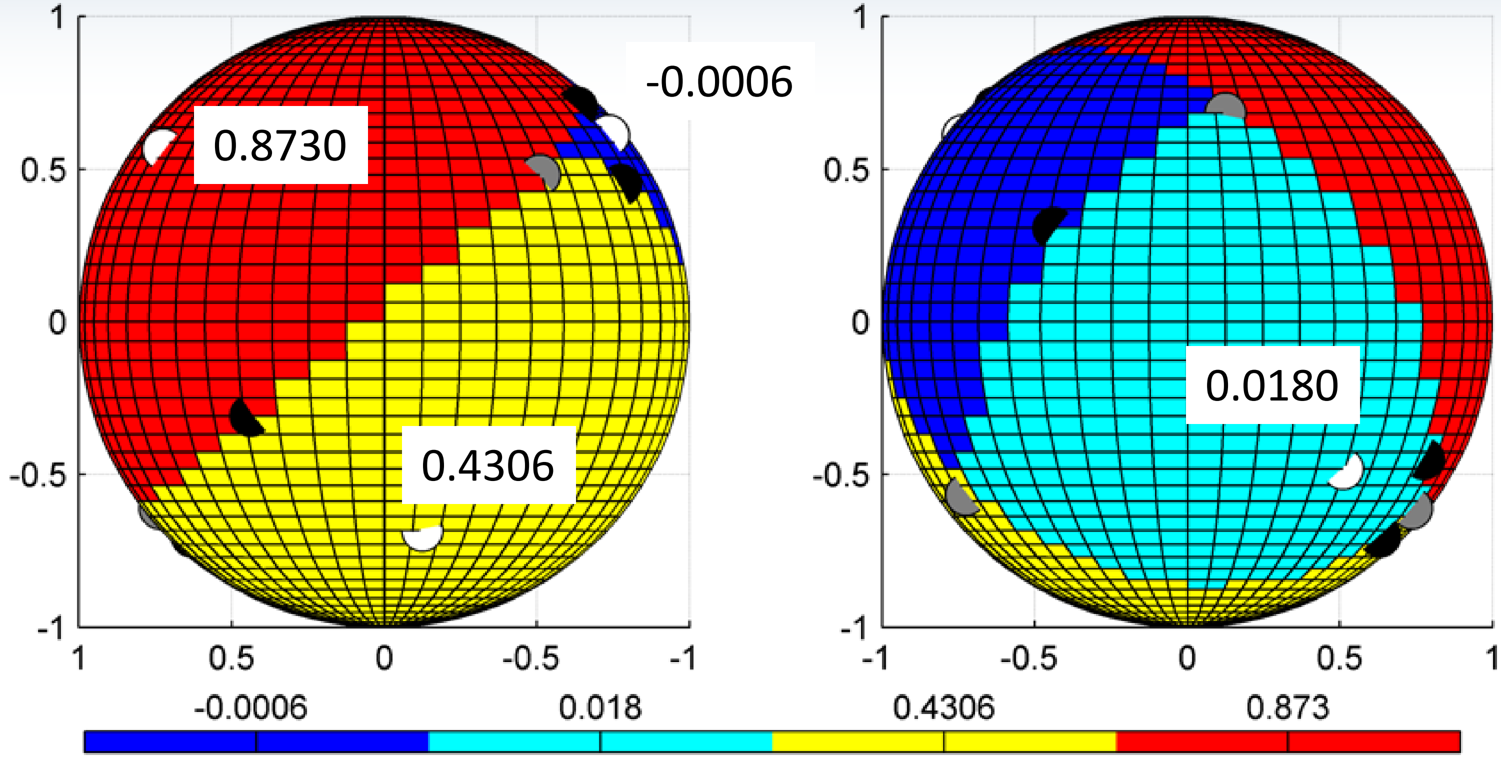


White = Local Max,
Gray = Local Min,
Black = Saddle Point

100 Random Starts

Occurrences	Lambda	Median Its.
62	0.8730	19
38	0.4306	184

Basins of Attraction for α = 1



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

100 Random Starts

Occurrences	Lambda	Median Its
40	0.8730	32
29	0.4306	48
18	0.0180	116
13	-0.0006	145

Relationship to Matrix Power Method

Sandia National Laboratories

<u>Symmetric Power Method</u>

$$\begin{aligned}&\text{For } k = 1, 2, \ldots\\ &\quad \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k / \|\mathbf{A}\mathbf{x}_k\|\\ &\quad \lambda_{k+1} = \mathbf{x}_{k+1}^{\mathrm{T}} \mathbf{A} \mathbf{x}_{k+1}\end{aligned}$$

Adding a shift moves the eigenvalues, potentially altering which eigenvalue is largest in magnitude.

$$\mathbf{A} \leftarrow \mathbf{A} + \alpha \mathbf{I}$$

$$\lambda_j \leftarrow \lambda_j + \alpha$$

Jacobian of fixed point operator at $(\lambda_j, \mathbf{x}_j)$ has eigenvalues:

$$\{0\} \cup \left\{ \tfrac{\lambda_i + \alpha}{\lambda_j + \alpha} : 1 \leq i \leq n \text{ with } i \neq j \right\}$$

Can only possibly have spectral radius less than one for largest or smallest eigenvalue.

Complex Tensor Eigenpairs

Qi (2005), Lim (2005)

<u>Definition</u>: Assume $\mathcal{A}$ is a symmetric m^{th} order n-dimensional real-valued tensor. We say that $\lambda \in \mathbb{C}$ is an eigenvalue if there exists $\mathbf{x} \in \mathbb{C}^n$ such that

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathbf{x} \quad \text{and} \quad \mathbf{x}^\dagger \mathbf{x} = 1.$$

The vector $\mathbf{x}$ is called the eigenvector.

Eigenpairs are not "unique" but define an equivalence class:

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

<u>Theorem</u>: # of distinct eigenvalues (real and complex) is exactly $((m\text{-}1)^n\text{-}1)/(m\text{-}2)$
Cartwright/Sturmfels 2010

For m = 3 and n = 4, we should have 7 distinct eigenvalues.

Complex SS-HOPM

Complex SS-HOPM

For $k = 1, 2, \ldots$

$$\hat{\mathbf{x}}_{k+1} = \frac{\mathcal{A}\mathbf{x}_k^{m-1} + \alpha \mathbf{x}_k}{\lambda_k + \alpha}$$

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

$$\lambda_{k+1} = \mathbf{x}_{k+1}^{\dagger} \mathcal{A} \mathbf{x}_{k+1}^{m-1}$$

$\lvert\lambda\rvert$	α = 2	$\alpha = 2^{1/2}(1+i)$
1.0954	18	22
0.8893	18	15
0.8169	21	12
0.6694	1	4
0.5629	22	16
0.3633	8	9
0.0451	12	20