Shifted Power Method for Computing Tensor Eigenvalues

Tamara G. Kolda & Jackson Mayo
Sandia National Labs

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

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.

Maximizing a Homogeneous Form

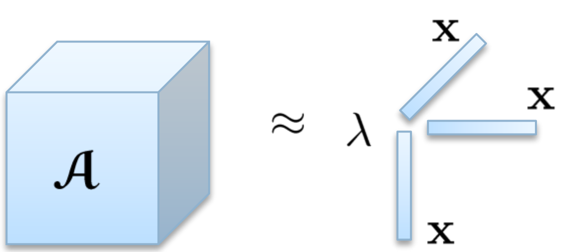
Let $\mathbf{A}$ be an n x n x $\cdots$ x n symmetric tensor of order m.

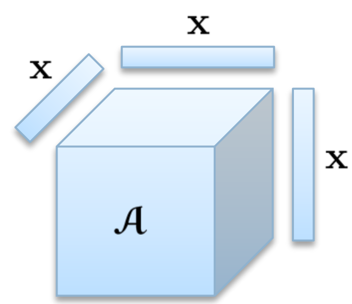
Every summand has degree m.

Homogeneous Form: $\mathcal{A}\mathbf{x}^m \equiv \sum_{i_1 i_2 \cdots i_m} a_{i_1 i_2 \ldots i_m} x_{i_1} x_{i_2} \cdots x_{i_m}$

Best Rank-1 Approximation: Equivalent to extreme point of homogenous form.

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





Homogeneous Form & Eigenpairs

Lim (2005)

Need to do both min and max.

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

Lagrangian:

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

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

We can define a real eigenpair as 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:

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

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

More General Definition of Tensor Eigenpairs

Qi (2005), Lim (2005)

Definition: 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.

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

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

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

Our Focus: Real Eigenpairs

$$m \text{ even} \ \Rightarrow (\lambda, -\mathbf{x}) \text{ is an eigenpair}$$

$$m \text{ odd} \ \Rightarrow (-\lambda, -\mathbf{x}) \text{ is an eigenpair}$$

These are eigenpairs in the same equivalence class.

Symmetric Higher-Order Power Method (S-HOPM)

De Lathauwer, De Moor, Vandewalle 2000

Symmetric Power Method

For $k = 1, 2, \ldots$

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

$$\lambda_{k+1} = \mathbf{x}_{k+1}^{\mathrm{T}} \mathbf{A} \mathbf{x}_{k+1}$$

S-HOPM

For $k = 1, 2, \ldots$

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

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

- Guaranteed to converge to the "leading" eigenpair
  - Leading eigenpair is the one with the largest magnitude eigenvalue
- Not guaranteed to converge in general
- In fact, may diverge or show chaotic behavior
- But sometimes works really well!

S-HOPM Analysis

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

Kofidis and Regalia (2002)

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

- 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 $l = m/2$.

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

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

S-HOPM Failure Example

Kofidis and 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

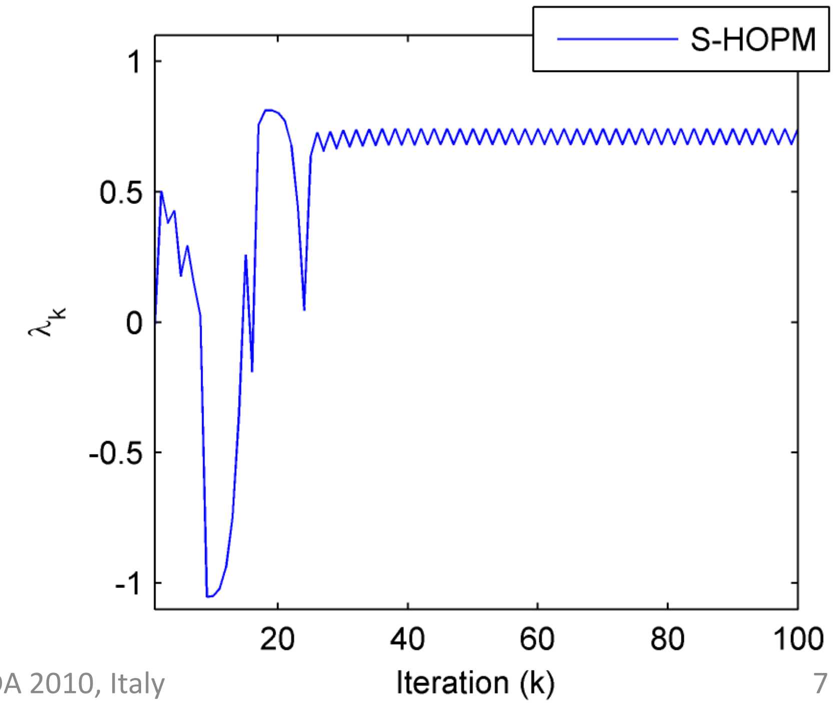
Why?
How can we fix it?

S-HOPM

For $k = 1, 2, \ldots$

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

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



Fixing & Analyzing S-HOPM

Forcing Convexity with a Shift

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

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

An analogue for even-order tensors:

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

Identity Tensor

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

A More General Shift for Convexity

Modify objective function:

$$f(\mathbf{x}) = \mathcal{A}\mathbf{x}^m \quad \Rightarrow \quad \hat{f}(\mathbf{x}) \equiv f(\mathbf{x}) + \alpha(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{m/2}$$

Max problem:

$$\begin{aligned} \max \quad & \mathcal{A}\mathbf{x}^m \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} \max \quad & \mathcal{A}\mathbf{x}^m + \alpha \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned}$$

$\hat{f}(\mathbf{x})$ convex for large positive α, λ_k inc.

Min problem:

$$\begin{aligned} \min \quad & \mathcal{A}\mathbf{x}^m \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min \quad & \mathcal{A}\mathbf{x}^m + \alpha \\ \text{s.t.} \quad & \|\mathbf{x}\| = 1 \end{aligned}$$

$\hat{f}(\mathbf{x})$ concave for large negative α, λ_k dec.

Shifted S-HOPM (SS-HOPM) Converges

S-HOPM

For $k = 1, 2, \ldots$

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

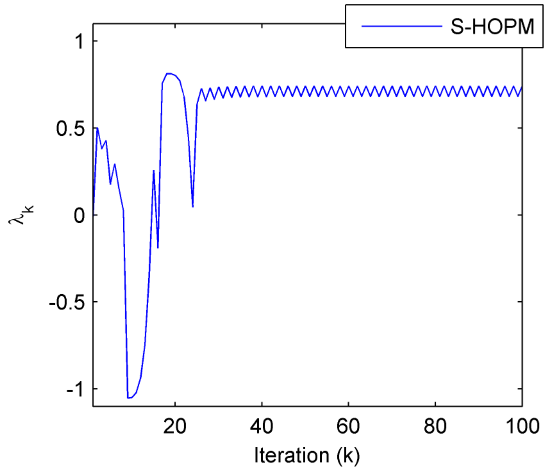
$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^m$$

SS-HOPM

For $k = 1, 2, \ldots$

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

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



For suitably large α…

- Nondecreasing λ_k
- $\lambda_k \to \lambda_*$
- $\mathbf{x}_k$ has a limit point $\mathbf{x}_*$
- $(\lambda_*, \mathbf{x}_*)$ is an eigenpair

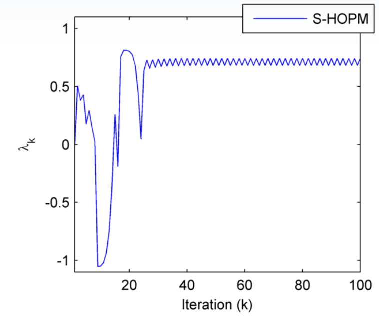
Example Convergence

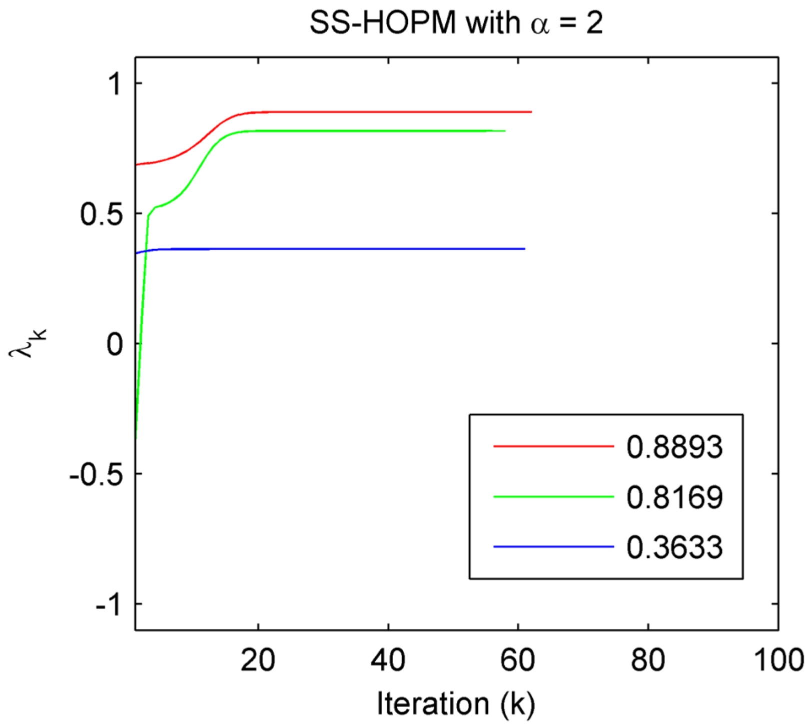
- 3 x 3 x 3 x 3 Symmetric Tensor

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

- Optimum: $|\lambda|$ = 1.09
- Experiment
  - 100 Random Starting Points
  - Use α = 2 (forces concavity)
- Results:

Occurrences	λ
46	0.8893
24	0.8169
30	0.3633





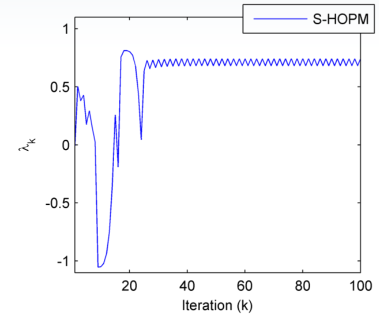
Different Eigenvalues with Negative Shift

- 3 x 3 x 3 x 3 Symmetric Tensor

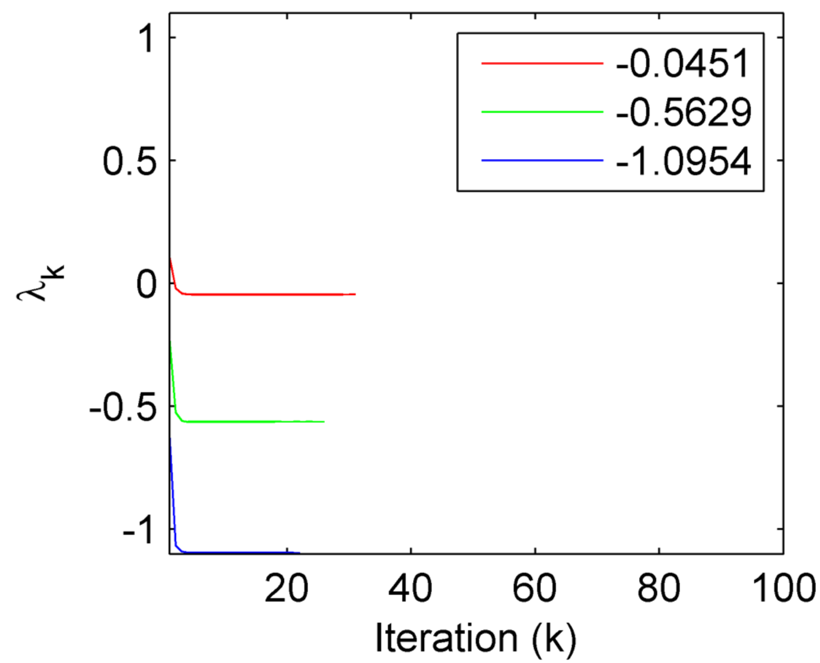
$$
\begin{array}{llll}
a_{1111} = 0.2883, & a_{1112} = -0.0031, & a_{1113} = 0.1973, \\
a_{1122} = -0.2485, & a_{1123} = -0.2939, & a_{1133} = 0.3847, \\
a_{1222} = 0.2972, & a_{1223} = 0.1862, & a_{1233} = 0.0919, \\
a_{1333} = -0.3619, & a_{2222} = 0.1241, & a_{2223} = -0.3420, \\
a_{2233} = 0.2127,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

- Optimum: $|\lambda|$ = 1.09
- Experiment
  - 100 Random Starting Points
  - Use α = -2 (forces convexity)
- Results:

Occurrences	λ
15	-0.0451
40	-0.5629
45	-1.0954



SS-HOPM with α = -2



SS-HOPM Convergence Theory (Part 1)

- Let $\mathbf{A}$ be an n x n x $\cdots$ x n symmetric tensor of order m
- For appropriate choice of α, SS-HOPM is guaranteed to converge to a tensor eigenpair for any starting point
  - Moreover, sequence of λ_k values is monotonic
- But…
  - How does the choice of α matter?
  - How fast does it converge?

<u>SS-HOPM</u>

For $k = 1, 2, \ldots$

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

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

Fixed Point Analysis

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

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

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

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

SS-HOPM

For $k = 1, 2, \ldots$

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

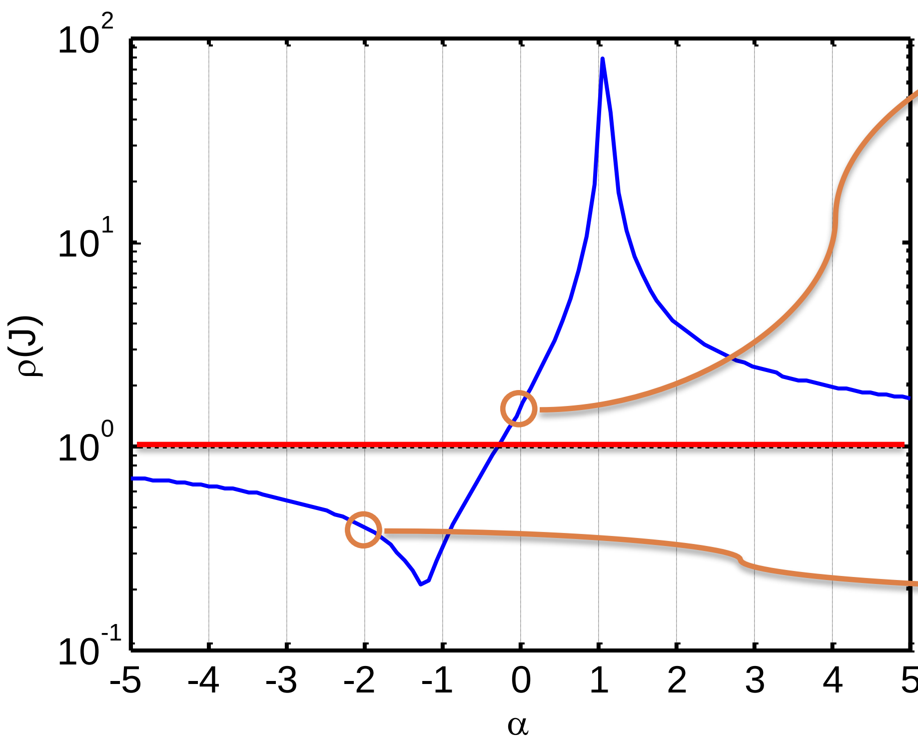
$$\lambda_{k+1} = \mathcal{A}\,\mathbf{x}_{k+1}^m$$

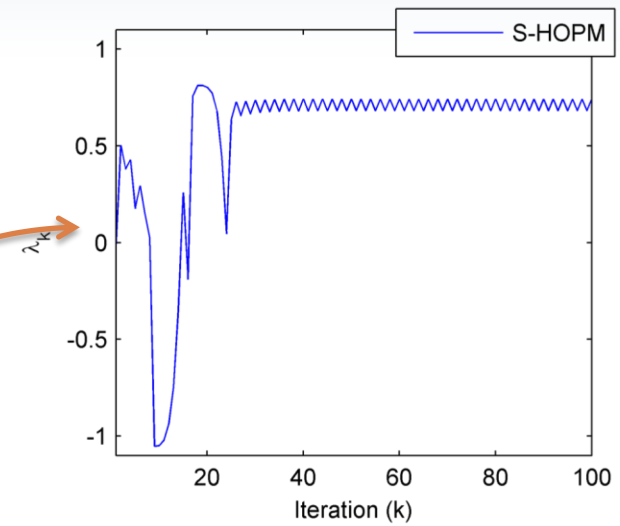
$$\phi(\mathbf{x}) = \frac{\mathcal{A}\mathbf{x}^{m-1} + \alpha\mathbf{x}}{\|\mathcal{A}\mathbf{x}^{m-1} + \alpha\mathbf{x}\|}$$

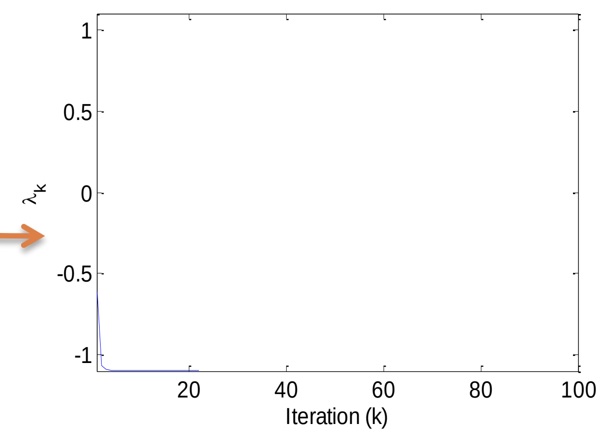
For our problem, any fixed point is an eigenpair and vice versa.

Understanding via Fixed Point Analysis

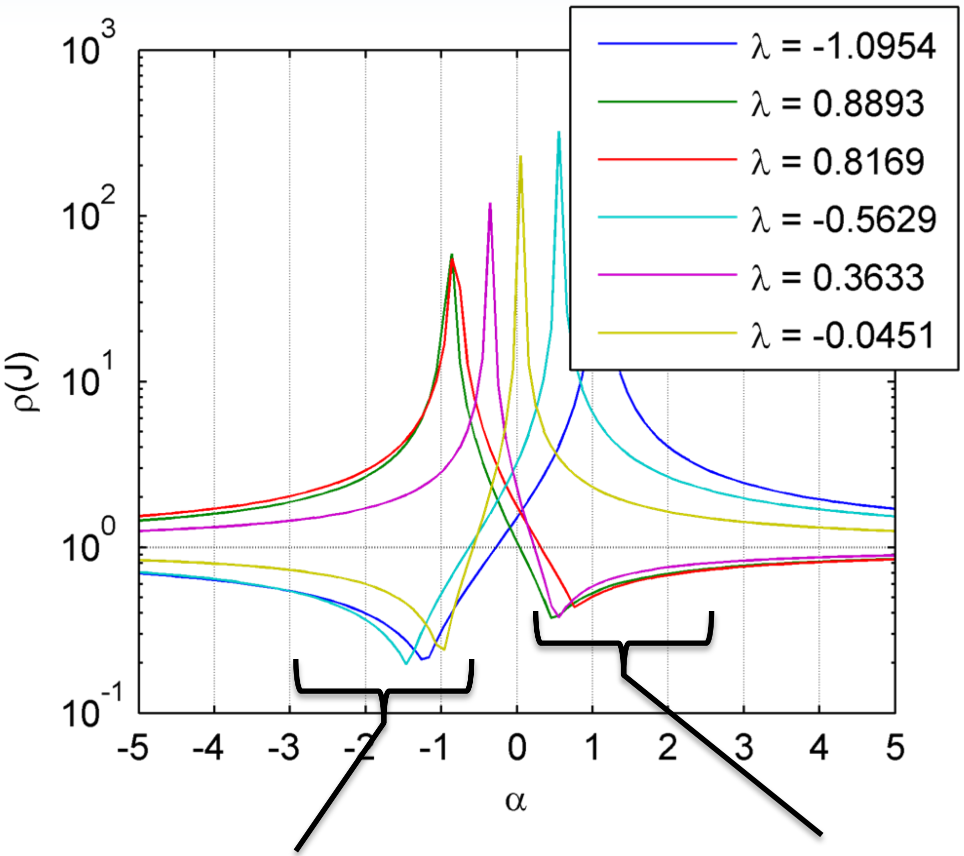
Spectral radius of Jacobian for
eigenvector corresponding to λ = -1.09







What choices of α create fixed points?



Positive Stable Fixed Points

Negative Stable Fixed Points

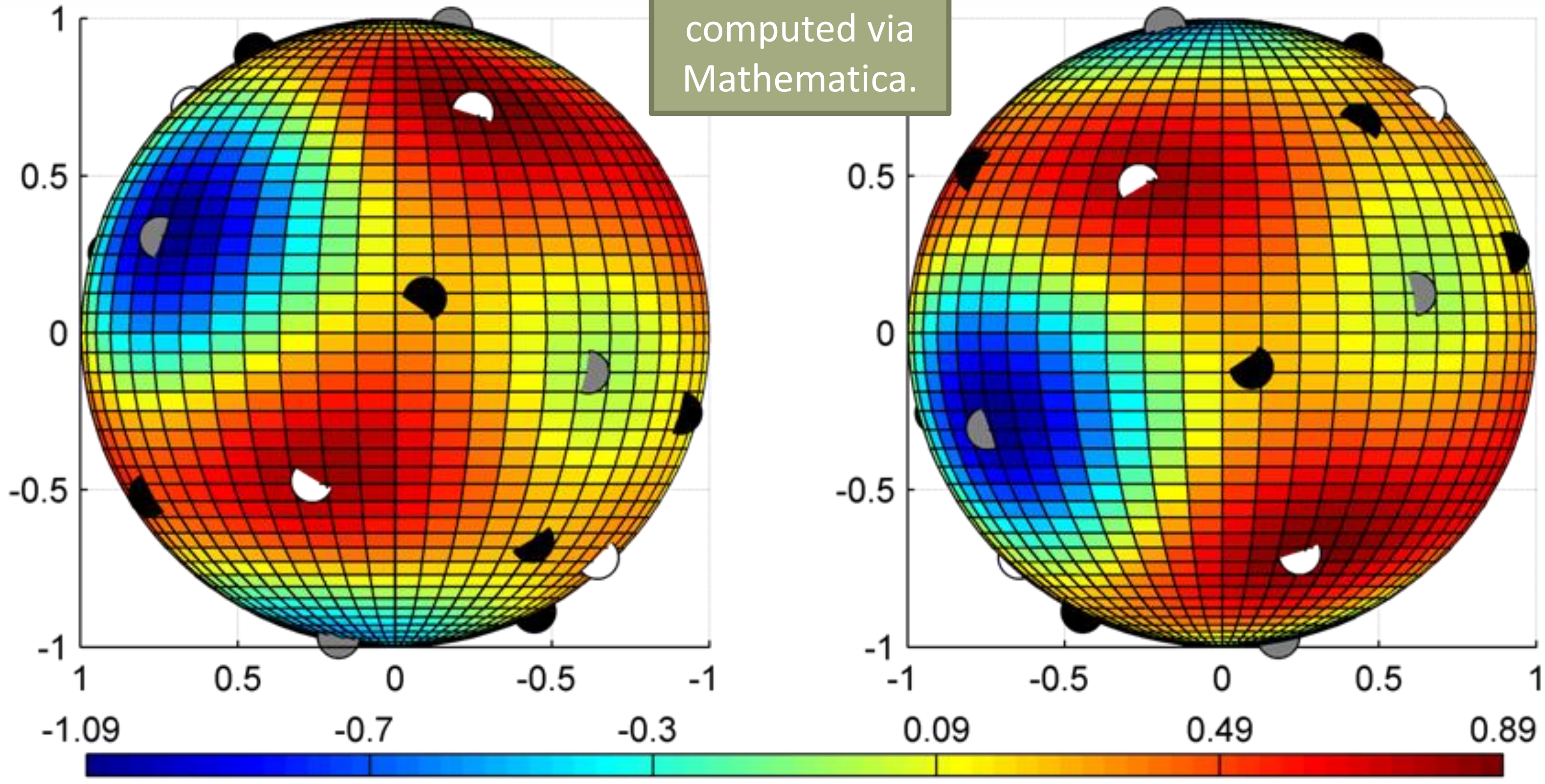
Not shown: Unstable Fixed Points (never attracting for any value of α)

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

Connections:
- Positive Stable – Local Minimum
- Negative Stable – Local Maximum
- Unstable – Saddle Point

Function Values for Example

All eigenpairs computed via Mathematica.



White = Negative Stable, Gray = Positive Stable, Black = Unstable

SS-HOPM Convergence Theory (Part 2)

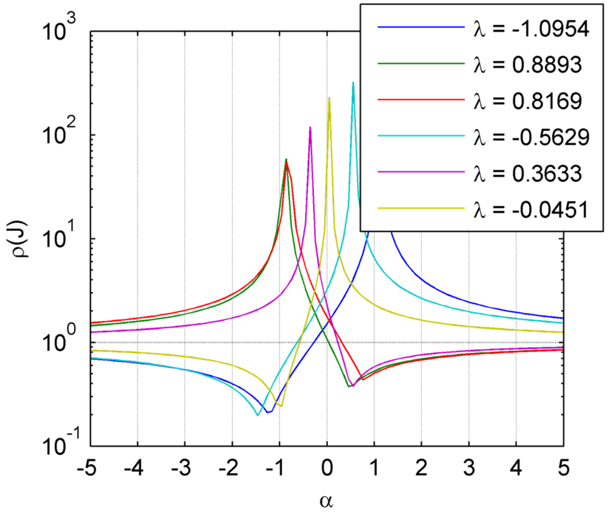
- Let $\mathcal{A}$ be an n x n x $\cdots$ x n symmetric tensor of order m
- For appropriate choice of α, SS-HOPM is guaranteed to converge to a tensor eigenpair for any starting point
  - Moreover, sequence of λ_k values is monotonic
- We can classify all eigenpairs as…
  - Positive stable
  - Negative stable
  - Unstable
- For appropriate choice of α, SS-HOPM can find all the positive and negative stable eigenpairs
  - Rate of convergence is determined by α

SS-HOPM

For $k = 1, 2, \ldots$

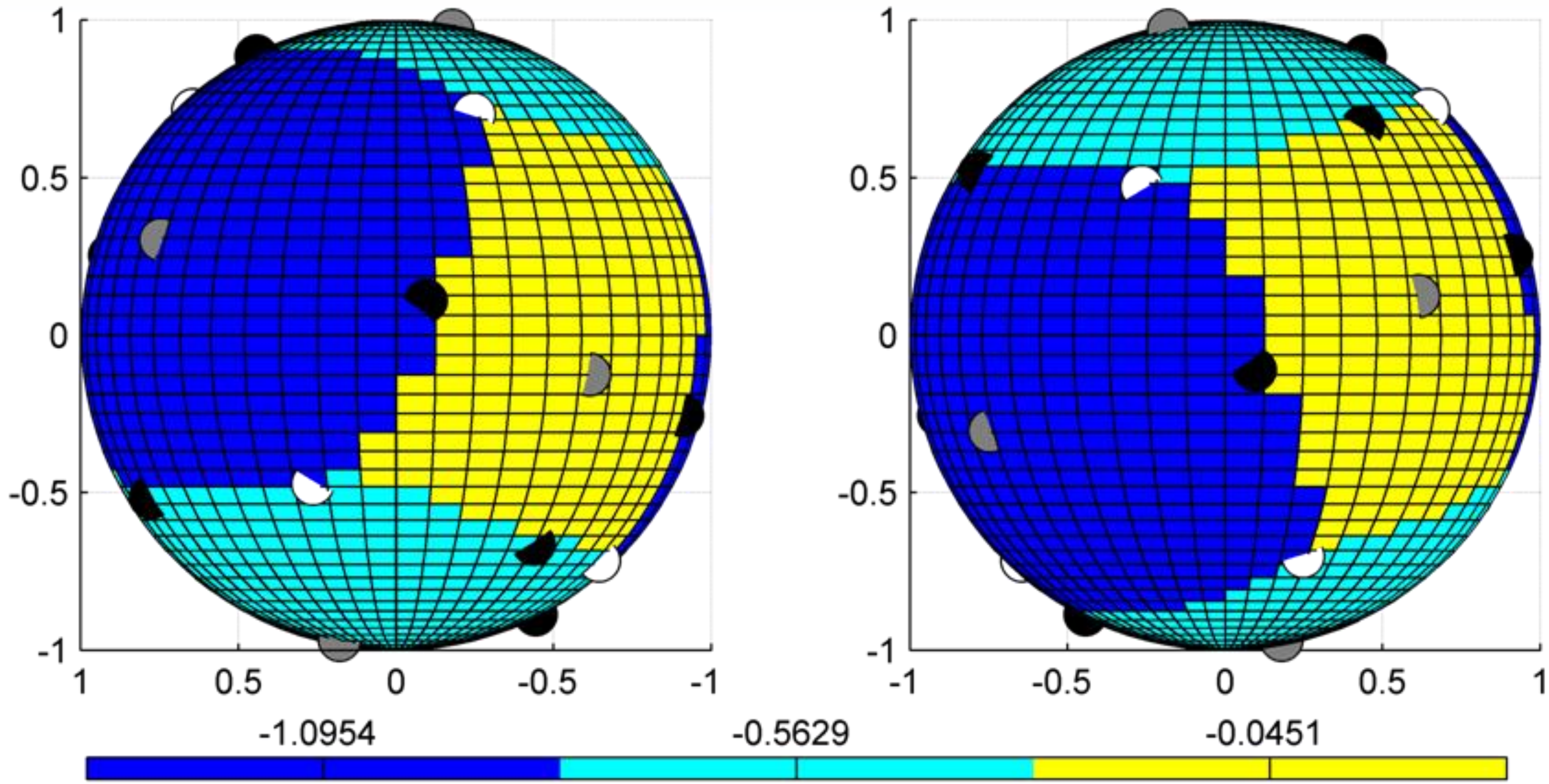
$$\mathbf{x}_{k+1} = \frac{\mathcal{A}\mathbf{x}_k^{m-1} + \alpha \mathbf{x}_k}{\|\mathcal{A}\mathbf{x}_k^{m-1} + \alpha \mathbf{x}_k\|}$$

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



Basins of Attraction for α = -2

Limit points correspond to local minima of function.



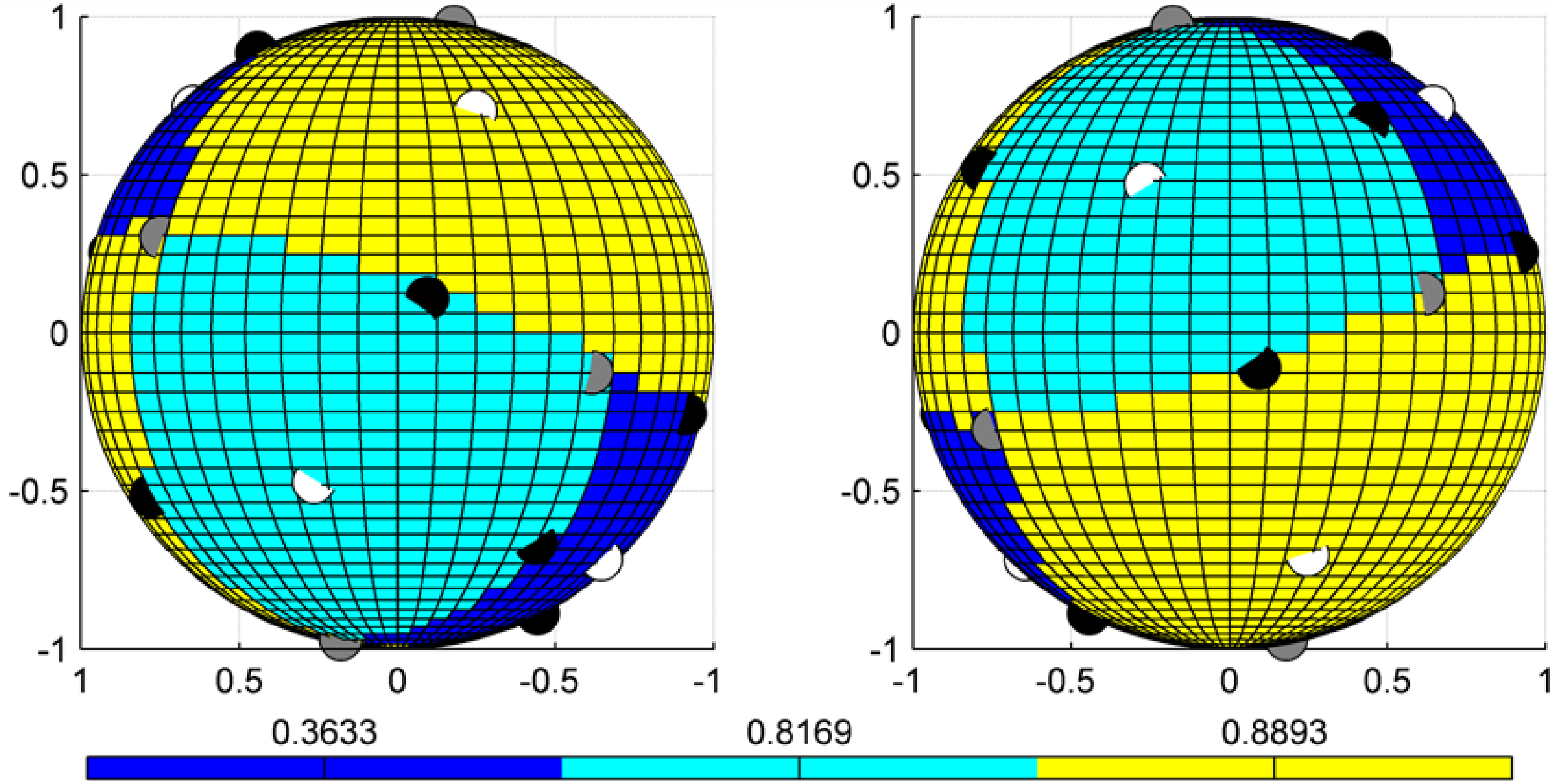
White = Negative Stable, Gray = Positive Stable,
Black = Unstable

Occurrences	λ
15	-0.0451
40	-0.5629
45	-1.0954

Sandia National Laboratories

Basins of Attraction for α = 2

Limit points correspond to local maxima of function.



White = Negative Stable, Gray = Positive Stable,
Black = Unstable

Occurrences	λ
46	0.8893
24	0.8169
30	0.3633

SS-HOPM on a GPU gets 340 Gflops/s

Ballard, K., Plantenga (2010)

- 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 $\mathbf{Ax}^m$ ~ (n^m) / $(m$-1)!
  - Cost of $\mathbf{Ax}^{(m\text{-}1)}$ ~ (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)	1.79
Intel Nahelem (4 cores)	10.03
nVidia Tesla 2050 (Fermi) 16 streaming multiprocessors (SMPs) 32 cores per SMP	339.96

Complex SS-HOPM

Complex SS-HOPM

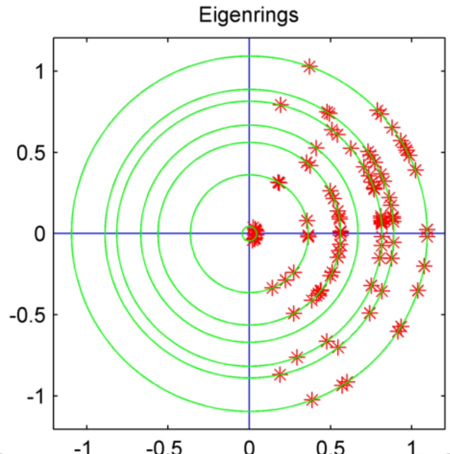
For $k = 1, 2, \ldots$

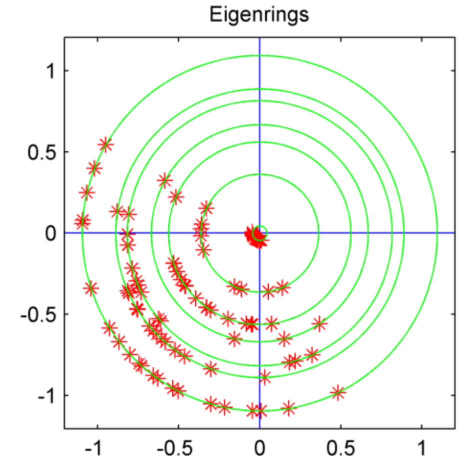
$$\hat{\mathbf{x}}_{k+1} = \frac{\mathcal{A}\mathbf{x}_k^{m-1} + \alpha \mathbf{x}_k}{\lambda_k + \alpha}$$

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

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

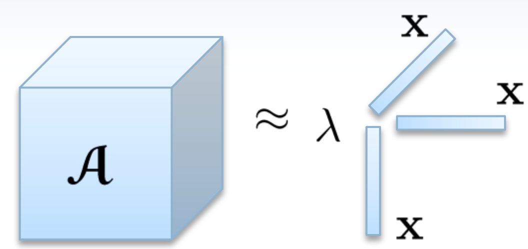
$\lvert\lambda\rvert$	$\alpha = 2$	$\alpha = 2^{1/2}(1+i)$
1.0954	18	22
0.8893	18	15
0.8169	21	12
0.6694	1	4
0.5629	22	16
0.3633	8	9
0.0451	12	20

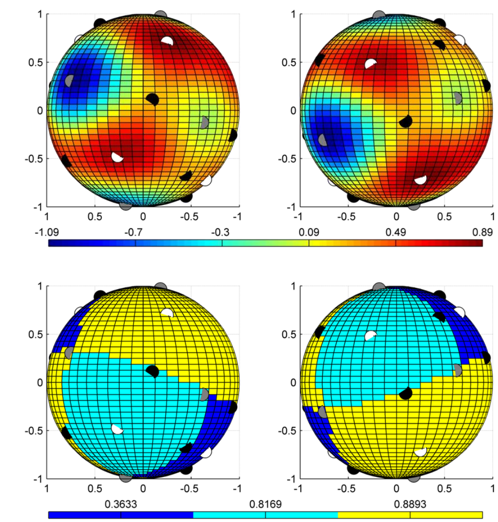




Conclusions & Future Work

- SS-HOPM is a convergent method for finding positive or negative stable real tensor eigenpairs
  - Convexity/concavity of (shifted) function sufficient
  - Even if function is not convex, fixed point analysis provides an alternative theoretical explanation
- Easily parallelizable
  - GPU implementation of SS-HOPM by Grey Ballard
- Applications
  - Signal Processing [Kofidis and Regalia 2002]
  - Diffusion tensor imaging [Schultz and Seidel 2008]
  - Molecular conformation [Rogers, unpublished]
- A few open problems
  - Perturbation analysis
  - Computing unstable eigenpairs
  - Eigendecomposition of a tensor?
  - Storage for symmetric tensors
  - Analysis of complex algorithm





For more info: Tammy Kolda
tgkolda@sandia.gov

Kolda and Mayo, ***Shifted Power Method for Computing Tensor Eigenpairs.*** arXiv:1007.1267.

Sandia National Laboratories

NIPS Workshop on Tensors, Kernels, and Machine Learning

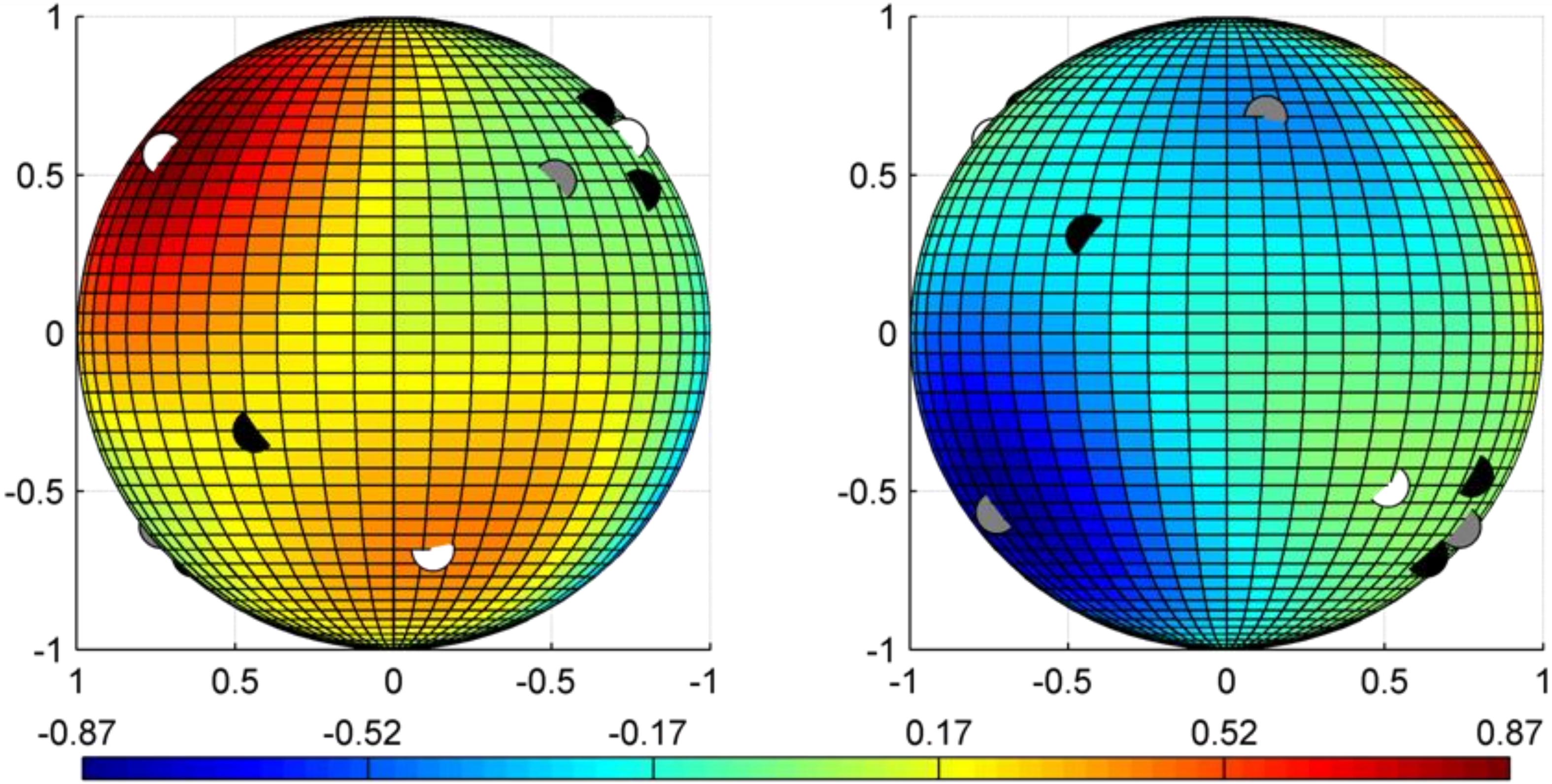
- Time & Place
  - Whistler, BC
  - December 10th or 11th 2010
- Organizers
  - Andreas Argyriou, *Toyota Institute of Tech. at Chicago*
  - David F. Gleich, *Sandia National Labs*
  - Tamara G. Kolda, *Sandia National Labs*
  - Vicente Malave, *University of California - San Diego*
  - Marco Signoretto, *K. U. Leuven*
  - Johan Suykens,*K. U. Leuven*
- Contributions
  - 4 pages
  - Deadline Sept 27, 2010

Tensors Kernels and Machine Learning 2010

http://csmr.ca.sandia.gov/~dfgleic/tkml2010/

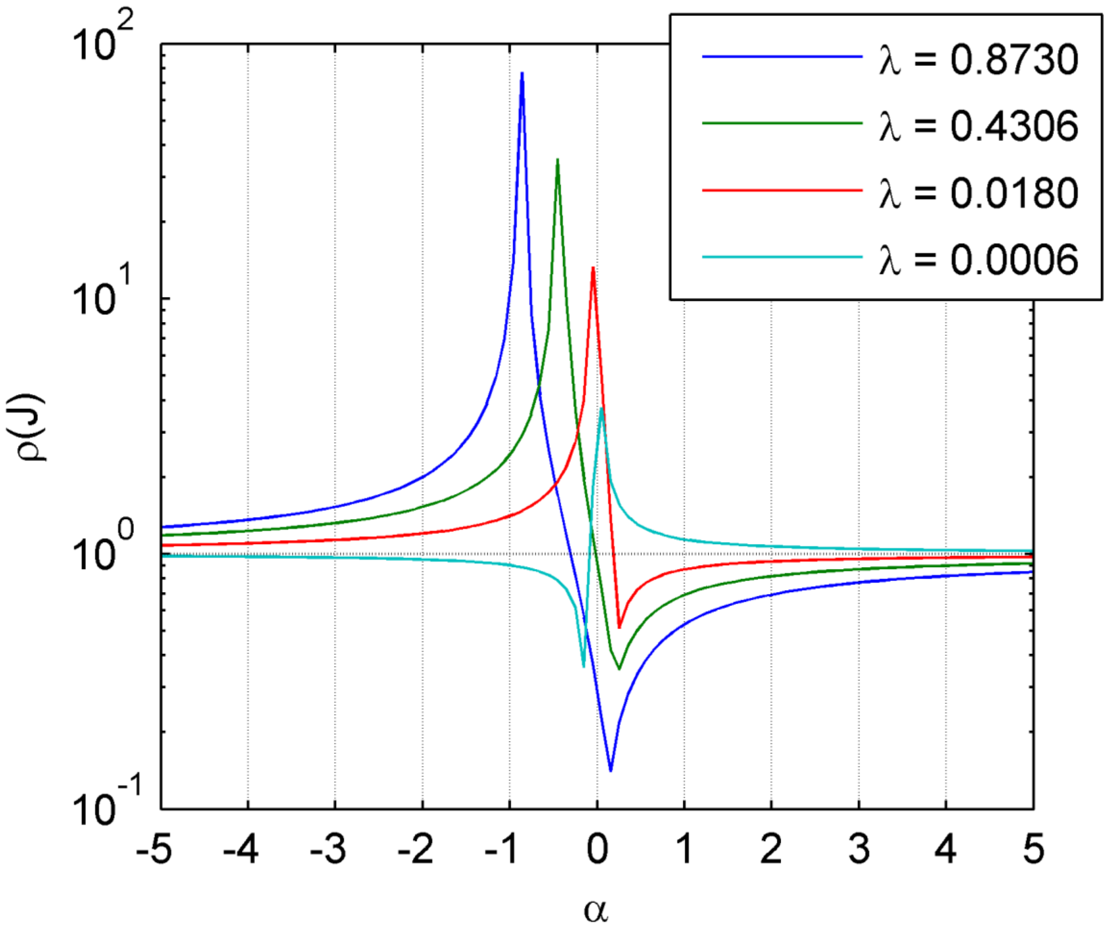
Another Example

Third-Order Example

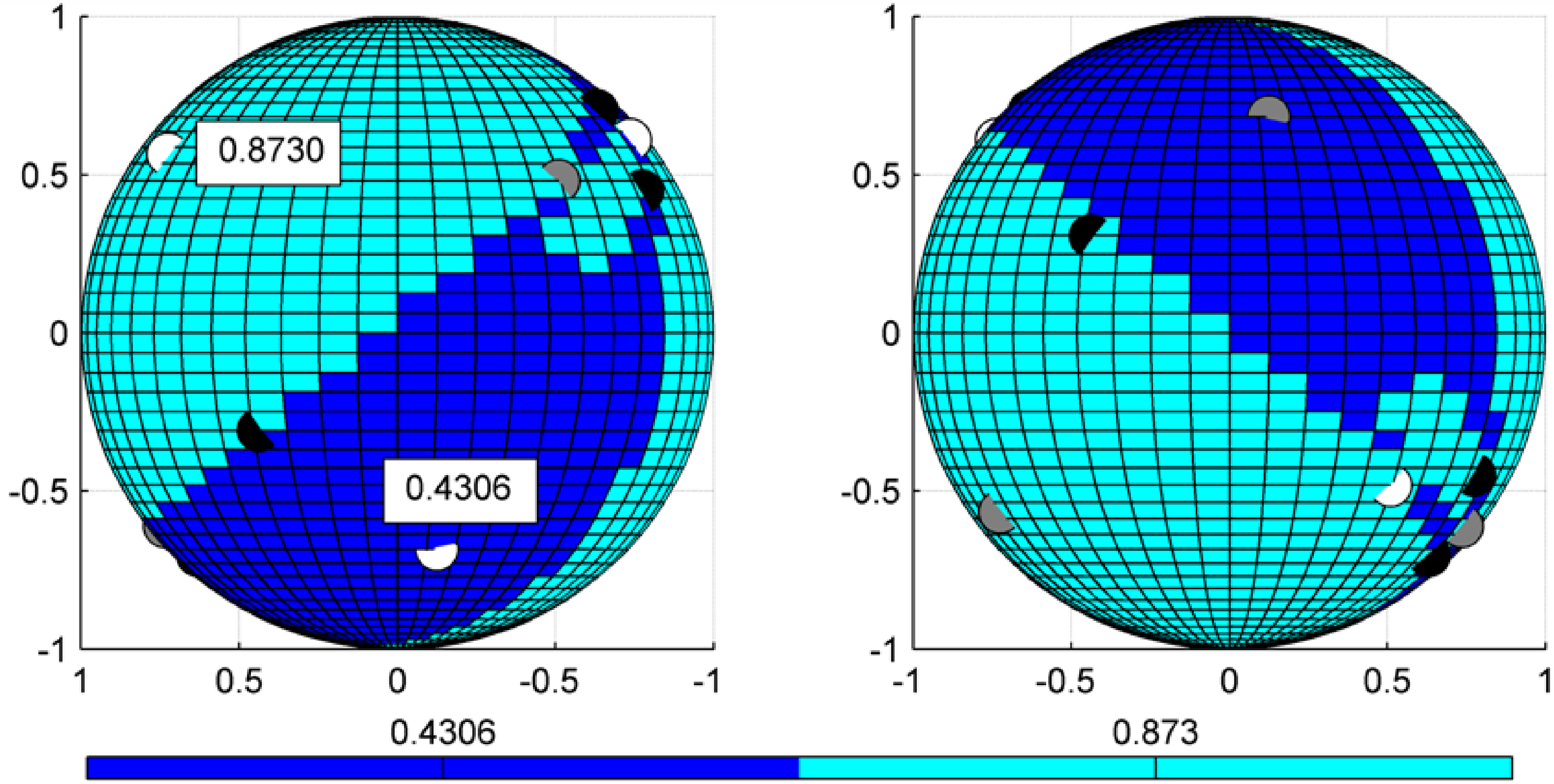


White = Negative Stable, Gray = Positive Stable, Black = Unstable

Stability of Third-Order Example

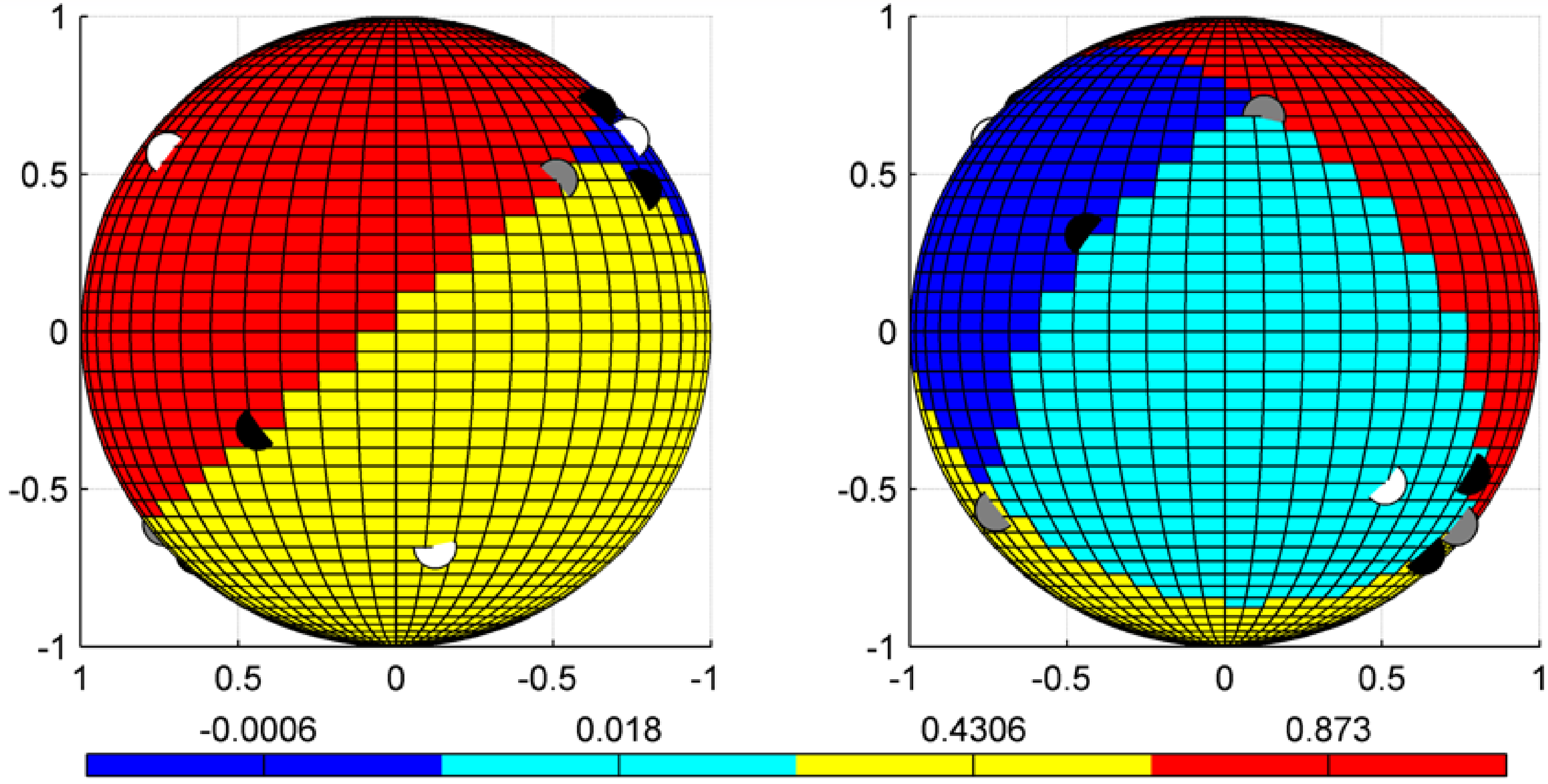


Jacobian explains Convergence



White = Negative Stable, Gray = Positive Stable, Black = Unstable

Basins of Attraction



White = Negative Stable, Gray = Positive Stable, Black = Unstable