Shifted Power Method for Computing Tensor Eigenvalues

Tamara G. Kolda & Jackson Mayo
Sandia National Labs
Maximizing a Homogeneous Form

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

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

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

\[
\begin{align*}
\min \quad & \| \mathbf{A} - \lambda \mathbf{x} \circ \mathbf{x} \circ \cdots \circ \mathbf{x} \|^2 \\
\text{s.t.} \quad & \lambda = \mathbf{A} x^m, \quad \| \mathbf{x} \| = 1
\end{align*}
\]

\[
\begin{align*}
\max \quad & | \mathbf{A} x^m | \\
\text{s.t.} \quad & \| \mathbf{x} \| = 1
\end{align*}
\]
Homogeneous Form & Eigenpairs

Lagrangean:
\[ \mathcal{L}(x, \mu) = Ax^m + \mu \frac{1}{2} (\|x\|^2 - 1) \]

We can define a real eigenpair as any KKT point of the constrained homogeneous form. (Analogous to the matrix case.)

KKT Conditions:
\[ mAx^{m-1} + \mu x = 0 \text{ and } \|x\| = 1 \]

Eigenpair:
\[ Ax^{m-1} = \lambda x \text{ and } \|x\| = 1 \]
\[ \text{(with } \lambda = -\mu/m) \]

Need to do both min and max.

\[ \max \; f(x) \equiv Ax^m \]
\[ \text{s.t. } \frac{1}{2} (\|x\|^2 - 1) = 0 \]
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 $x \in \mathbb{C}^n$ such that

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

The vector $x$ is called the eigenvector.

**Theorem**: The number of distinct complex eigenvalues is

$$\frac{(m-1)n-1}{m-2}$$

Cartwright/Sturmfels 2010

*Eigenpairs are not “unique” but define an equivalence class:*

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

Our Focus: Real Eigenpairs

$m$ even $\Rightarrow (\lambda, -x)$ is an eigenpair

$m$ odd $\Rightarrow (-\lambda, -x)$ 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 \( k = 1, 2, \ldots \)
\[
x_{k+1} = \frac{Ax_k}{\|Ax_k\|}
\]
\[
\lambda_{k+1} = x_{k+1}^T A x_{k+1}
\]

**S-HOPM**
For \( k = 1, 2, \ldots \)
\[
x_{k+1} = \mathcal{A} x_{k}^{m-1} / \| \mathcal{A} x_{k}^{m-1} \|
\]
\[
\lambda_{k+1} = \mathcal{A} x_{k+1}^{m}
\]

- 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

Kofidis and Regalia (2002)

- **Theorem:** S-HOPM $\lambda_k$ converges to eigenvalue if $f(x)$ is convex or concave on unit ball

- **Key Lemma:** Assume $f(x)$ convex on unit ball and let $v$ be such that $\|v\|=1$.
  - If $w = \nabla f(v)/\|\nabla f(v)\|
  - Then $f(w) \geq f(v)$

- **Importance:** If $f(x)$ is convex, then S-HOPM has $\lambda_{k+1} \geq \lambda_k$ for all $k$

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

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

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

- **S-HOPM**
  For $k = 1, 2, \ldots$
  - $x_{k+1} = \mathcal{A}x_{k}^{m-1}/\|\mathcal{A}x_{k}^{m-1}\|
  - $\lambda_{k+1} = \mathcal{A}x_{k+1}^m$
S-HOPM Failure Example

Kofidis and Regalia (2002)

- $3 \times 3 \times 3 \times 3$ Symmetric Tensor

$$
\begin{align*}
     a_{111} &= 0.2883, & a_{112} &= -0.0031, & a_{113} &= 0.1973, \\
     a_{112} &= -0.2485, & a_{113} &= -0.2939, & a_{113} &= 0.3847, \\
     a_{122} &= 0.2972, & a_{123} &= 0.1862, & a_{123} &= 0.0919, \\
     a_{133} &= -0.3619, & a_{222} &= 0.1241, & a_{223} &= -0.3420, \\
     a_{223} &= 0.2127, & a_{233} &= 0.2727, & a_{333} &= -0.3054.
\end{align*}
$$

- Optimum: $|\lambda| = 1.09$

- S-HOPM fails on this problem for every starting point we tried

**S-HOPM**

For $k = 1, 2, \ldots$

$$
x_{k+1} = \mathcal{A}x_k^{m-1} / \|\mathcal{A}x_k^{m-1}\|$$

$$
\lambda_{k+1} = \mathcal{A}x_k^m
$$

September 14, 2010
Fixing & Analyzing S-HOPM
A quadratic function is convex if all the eigenvalues of \( A \) are positive (and concave if all are negatives).

\[
\begin{align*}
\max \quad & f(x) \equiv x^T Ax \\
\text{s.t.} \quad & \|x\| = 1
\end{align*}
\]

\[
\begin{align*}
\max \quad & \hat{f}(x) \equiv x^T (A + \alpha I)x \\
\text{s.t.} \quad & \|x\| = 1
\end{align*}
\]

An analogue for even-order tensors:

\[
\begin{align*}
\max \quad & f(x) \equiv A x^m \\
\text{s.t.} \quad & \|x\| = 1
\end{align*}
\]

\[
\begin{align*}
\max \quad & \hat{f}(x) \equiv (A + \alpha E) x^m \\
\text{s.t.} \quad & \|x\| = 1
\end{align*}
\]

Identity Tensor

\[E x^{m-1} = x \quad \forall x\]
A More General Shift for Convexity

Modify objective function:

$$f(x) = Ax^m$$

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

**Max problem:**

$$\max \ A x^m$$

subject to $$\|x\| = 1$$

$$\Rightarrow$$

$$\max \ A x^m + \alpha$$

subject to $$\|x\| = 1$$

$$\hat{f}(x)$$ convex for large positive $$\alpha$$, $$\lambda_k$$ inc.

**Min problem:**

$$\min \ A x^m$$

subject to $$\|x\| = 1$$

$$\Rightarrow$$

$$\min \ A x^m + \alpha$$

subject to $$\|x\| = 1$$

$$\hat{f}(x)$$ concave for large negative $$\alpha$$, $$\lambda_k$$ dec.
Shifted S-HOPM (SS-HOPM) Converges

**S-HOPM**
For \( k = 1, 2, \ldots \)
\[
    x_{k+1} = \frac{Ax_k^{m-1}}{\|Ax_k^{m-1}\|} \\
    \lambda_{k+1} = A x_{k+1}^m
\]

**SS-HOPM**
For \( k = 1, 2, \ldots \)
\[
    x_{k+1} = \frac{Ax_k^{m-1} + \alpha x_k}{\|Ax_k^{m-1} + \alpha x_k\|} \\
    \lambda_{k+1} = A x_{k+1}^m
\]

For suitably large \( \alpha \ldots \)
- Nondecreasing \( \lambda_k \)
- \( \lambda_k \to \lambda_* \)
- \( x_k \) has a limit point \( x_* \)
- \( (\lambda_*, x_*) \) is an eigenpair
Example Convergence

- **3 x 3 x 3 x 3 Symmetric Tensor**

  \[
  a_{111} = 0.2883, \quad a_{112} = -0.0031, \quad a_{113} = 0.1973, \\
  a_{122} = -0.2485, \quad a_{123} = -0.2939, \quad a_{133} = 0.3847, \\
  a_{222} = 0.2972, \quad a_{223} = 0.1862, \quad a_{233} = 0.0919, \\
  a_{333} = -0.3619, \quad a_{222} = 0.1241, \quad a_{223} = -0.3420, \\
  a_{233} = 0.2127, \quad a_{233} = 0.2727, \quad a_{333} = -0.3054.
  \]

- **Optimum:** \(|\lambda| = 1.09\)

- **Experiment**
  - 100 Random Starting Points
  - Use \(\alpha = 2\) (forces concavity)

- **Results:**

<table>
<thead>
<tr>
<th>Occurrences</th>
<th>(\lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>0.8893</td>
</tr>
<tr>
<td>24</td>
<td>0.8169</td>
</tr>
<tr>
<td>30</td>
<td>0.3633</td>
</tr>
</tbody>
</table>

[Graph showing convergence of \(\lambda\) over iterations with different colors for different \(\lambda\) values.]
Different Eigenvalues with Negative Shift

- **3 x 3 x 3 x 3 Symmetric Tensor**
  
  \[
  a_{111} = 0.2883, \quad a_{112} = -0.0031, \quad a_{113} = 0.1973, \\
  a_{122} = -0.2485, \quad a_{123} = -0.2939, \quad a_{133} = 0.3847, \\
  a_{222} = 0.2972, \quad a_{223} = 0.1862, \quad a_{233} = 0.0919, \\
  a_{333} = -0.3619, \quad a_{222} = 0.1241, \quad a_{223} = -0.3420, \\
  a_{233} = 0.2127, \quad a_{233} = 0.2727, \quad a_{333} = -0.3054.
  \]

- **Optimum:** \(|\lambda| = 1.09\)

- **Experiment**
  - 100 Random Starting Points
  - Use \(\alpha = -2\) (forces convexity)

- **Results:**

<table>
<thead>
<tr>
<th>Occurrences</th>
<th>(\lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>-0.0451</td>
</tr>
<tr>
<td>40</td>
<td>-0.5629</td>
</tr>
<tr>
<td>45</td>
<td>-1.0954</td>
</tr>
</tbody>
</table>

SS-HOPM with \(\alpha = -2\)

\[
\begin{align*}
\lambda_k & = -0.0451 \\
\lambda_k & = -0.5629 \\
\lambda_k & = -1.0954
\end{align*}
\]
Let $\mathbf{A}$ be an $n \times n \times \cdots \times n$ symmetric tensor of order $m$.

For appropriate choice of $\alpha$, SS-HOPM is guaranteed to converge to a tensor eigenpair for any starting point.

Moreover, sequence of $\lambda_k$ values is monotonic.

But...

How does the choice of $\alpha$ matter?

How fast does it converge?

SS-HOPM

For $k = 1, 2, \ldots$

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

$$
\lambda_{k+1} = \mathbf{A} \mathbf{x}_k^m
$$
Fixed Point Analysis

**Fixed Point of** \( \phi \): \( \phi(x) = x \)

Let \( J(x) \) denote the \( n \times n \) Jacobian of \( \phi(x) \).

**Fact 1:** \( x \) is an **attracting** fixed point if \( \sigma \equiv \rho(J(x)) < 1 \).

**Fast 2:** The convergence is linear with rate \( \sigma \) (smaller is faster).

**SS-HOPM**

For \( k = 1, 2, \ldots \)

\[
    x_{k+1} = \frac{Ax_{k}^{m-1} + \alpha x_{k}}{\|Ax_{k}^{m-1} + \alpha x_{k}\|}
\]

\[
    \lambda_{k+1} = Ax_{k}^{m+1}
\]

\[
    \phi(x) = \frac{Ax^{m-1} + \alpha x}{\|Ax^{m-1} + \alpha x\|}
\]

For our problem, any fixed point is an eigenpair and vice versa.
Spectral radius of Jacobian for eigenvector corresponding to $\lambda = -1.09$
What choices of $\alpha$ create fixed points?

Not shown: Unstable Fixed Points (never attracting for any value of $\alpha$)

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

Connections:
- Positive Stable – Local Minimum
- Negative Stable – Local Maximum
- Unstable – Saddle Point
White = Negative Stable, Gray = Positive Stable, Black = Unstable

All eigenpairs computed via Mathematica.
Let $A$ be an $n \times n \times \cdots \times n$ symmetric tensor of order $m$.

For appropriate choice of $\alpha$, SS-HOPM is guaranteed to converge to a tensor eigenpair for any starting point.

Moreover, sequence of $\lambda_k$ values is monotonic.

We can classify all eigenpairs as...

- Positive stable
- Negative stable
- Unstable

For appropriate choice of $\alpha$, SS-HOPM can find all the positive and negative stable eigenpairs.

Rate of convergence is determined by $\alpha$.

SS-HOPM

For $k = 1, 2, \ldots$

$$x_{k+1} = \frac{Ax_k^{m-1} + \alpha x_k}{\|Ax_k^{m-1} + \alpha x_k\|}$$

$$\lambda_{k+1} = A x_k^m$$
Basins of Attraction for $\alpha = -2$

Limit points correspond to local minima of function.

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

<table>
<thead>
<tr>
<th>Occurrences</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>-0.0451</td>
</tr>
<tr>
<td>40</td>
<td>-0.5629</td>
</tr>
<tr>
<td>45</td>
<td>-1.0954</td>
</tr>
</tbody>
</table>
Basins of Attraction for $\alpha = 2$

Limit points correspond to local maxima of function.

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

<table>
<thead>
<tr>
<th>Occurrences</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>0.8893</td>
</tr>
<tr>
<td>24</td>
<td>0.8169</td>
</tr>
<tr>
<td>30</td>
<td>0.3633</td>
</tr>
</tbody>
</table>
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 \( \sim (n^m)/m! \)
  – Cost of \( Ax^m \sim (n^m)/(m-1)! \)
  – Cost of \( Ax^{(m-1)} \sim (mn^m)/(m-1)! \)

• GPU implementation
  – One “thread block” per tensor
  – One “thread” per starting point
  – Loop unrolling gives up to 20x speed-up

<table>
<thead>
<tr>
<th>Compute Engine</th>
<th>Gflops/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel Nahelem (1 core)</td>
<td>1.79</td>
</tr>
<tr>
<td>Intel Nahelem (4 cores)</td>
<td>10.03</td>
</tr>
<tr>
<td>nVidia Tesla 2050 (Fermi)</td>
<td>339.96</td>
</tr>
<tr>
<td>16 streaming multiprocessors (SMPs)</td>
<td></td>
</tr>
<tr>
<td>32 cores per SMP</td>
<td></td>
</tr>
</tbody>
</table>

September 14, 2010  T. G. Kolda – TDA 2010, Italy  22
Complex SS-HOPM

For $k = 1, 2, \ldots$

$$\hat{x}_{k+1} = \frac{Ax_k^{m-1} + \alpha x_k}{\lambda_k + \alpha}$$

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

$$\lambda_{k+1} = x_{k+1}^\dagger A x_k^{m-1}$$

| $|\lambda|$ | $\alpha = 2$ | $\alpha = 2^{1/2}(1+i)$ |
|-----------|--------------|--------------------------|
| 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                       |

Eigenrings
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

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

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

\[ \rho(J) \]

-5 \quad -4 \quad -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5

-2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5

-5 \quad -4 \quad -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5

-2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5

\[ \lambda = 0.8730 \]
\[ \lambda = 0.4306 \]
\[ \lambda = 0.0180 \]
\[ \lambda = 0.0006 \]
Jacobian explains Convergence

White = Negative Stable, Gray = Positive Stable, Black = Unstable
Basins of Attraction

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