CS 598 EVS: Tensor Computations
Tensor Eigenvales

Edgar Solomonik
University of Illinois at Urbana-Champaign
Matrix Eigenvalues

- The eigenvalue and singular value decompositions of matrices enable not only low-rank approximation (which we can get for tensors via decomposition), but also describe important properties of the matrix $M$ and associated linear function $f(M)(x) = Mx$
  - Eigenvalues and eigenvectors can be used to characterize eigenfunctions of differential operators
  - Eigenvalues describe powers of the matrix and its limiting behavior

$$M = XDX^{-1} \Rightarrow M^2 = XD^2X^{-1}$$

if there is a unique largest eigenvalue $\lambda$ with associated left/right eigenvectors are $x$, $y$ then

$$\lim_{k \to \infty} \frac{M^k}{\|M^{k-1}\|} = \lambda xy$$

- They can be used to find stationary states of statistical processes and to find low-cut partitions in graphs
Tensor Eigenvalues

- Tensor eigenvalues and singular values can be defined based on the function $f^{(T)}$ by analogy from the role of matrix eigenvalues on $f^{(M)}$

  - Matrix eigenpairs $(\lambda, x)$ satisfy $f^{(M)}(x) = \lambda x$, while for an order $d$ symmetric tensor, we may define\(^1,\)\(^2\)
    
    $f^{(T)}(x, \ldots, x) = \lambda x$ \quad $f^{(T)}(x, \ldots, x) = \lambda x^{d-1}$ \quad $f^{(T)}(x, \ldots, x) = \lambda x^{p-1}$

    where $x^p = [x_1^p \ldots x_n^p]^T$

  - For matrices, Z-eigenpairs ($l^p$-eigenpairs with $p = 1$) and H-eigenpairs ($l^p$-eigenpairs with $p = d - 1$) are the same

  - Singular value/vector pairs can be defined by a tuple $(\sigma, x_1, \ldots, x_d)$ that satisfies $d$ equations like $f^{(T)}(x_2, \ldots, x_d) = \sigma x_1^p$, e.g., for $d = 3, p = 1$, 
    \[ T_{(1)}(x_2 \otimes x_3) = \sigma x_1, \quad T_{(2)}(x_1 \otimes x_3) = \sigma x_2, \quad T_{(3)}(x_1 \otimes x_2) = \sigma x_3 \]

---

\(^1\) Liqun Qi, “Eigenvalues of a Real Supersymmetric Tensor”, 2005

Matrix Eigenvalues and Critical Points

- The eigenvalues/eigenvectors of a matrix are the critical values/points of its Rayleigh quotient\(^3\)
  - The Lagrangian function of \( f(x) = x^T A x \) subject to \( \|x\|_2^2 = \|x\|_2 \|x\|_2 = 1 \) is
    \[
    \mathcal{L}(x, \lambda) = x^T A x - \lambda(\|x\|_2^2 - 1)
    \]
  - The first-order optimality condition are \( \|x\|_2 = 1 \) and
    \[
    \frac{d\mathcal{L}}{dx}(x, \lambda) = 0 \Rightarrow A x = \lambda x
    \]

- Singular vectors and singular values of matrices may be derived analogously
  - The Lagrangian function of \( f(x, y) = x^T A y \) subject to \( \|x\|_2 \|y\|_2 = 1 \) is
    \[
    \mathcal{L}(x, y, \sigma) = x^T A y - \sigma(\|x\|_2 \|y\|_2 - 1)
    \]
  - The first-order optimality conditions are \( \|x\|_2 \|y\|_2 = 1 \) and
    \[
    \frac{d\mathcal{L}}{dx}(x, y, \sigma) = 0 \Rightarrow \frac{A y}{\|y\|} = \frac{\sigma x}{\|x\|}, \quad \frac{d\mathcal{L}}{dy}(x, y, \sigma) = 0 \Rightarrow \frac{A x}{\|x\|} = \frac{\sigma y}{\|y\|}
    \]

\(^3\)Lek-Heng Lim, “Singular Values and Eigenvalues of Tensors: A Variational Approach”, 2005
The Lagrangian approach to matrix eigenvalues generalizes naturally to symmetric tensors. The symmetric tensor is associated with a multilinear scalar-valued function 
\[ f^{(T)}(x) = \sum_{i_1,\ldots,i_d} t_{i_1,\ldots,i_d} x_{i_1} \cdots x_{i_d} \] as well as the vector valued function 
\[ f^{(T)}(x) = \sum_{i_1,\ldots,i_{d-1}} t_{i_1,\ldots,i_{d-1}} x_{i_1} \cdots x_{i_{d-1}} = \frac{1}{d} \nabla f^{(T)}(x) \]

We consider its Lagrangian subject to a normalization condition \( \|x\|_p^d = 1 \) (for matrices \( p = 2 \), so for order \( d \) tensors natural to pick either \( p = 2 \) or \( p = d \)),

\[ \mathcal{L}(x, \lambda) = f(x) - \lambda(\|x\|_p^d - 1) \]

The first order optimality conditions for \( p = 2 \) is \( \|x\|_2 = 1 \) and

\[ \frac{d\mathcal{L}}{dx}(x, \lambda) = 0 \quad \Rightarrow \quad f^{(T)}(x) = \lambda x \]

The analogous first order optimality condition for \( p = d \) and even \( p \) is

\[ \frac{d\mathcal{L}}{dx}(x, \lambda) = 0 \quad \Rightarrow \quad f^{(T)}(x) = \lambda x^{d-1} \]

is scale invariant (if \( (x^*, \lambda) \) minimizes \( \mathcal{L} \) so does \( (\alpha x^*, \lambda) \))
Tensor Singular Values and Singular Vectors

- Tensor singular values again can be viewed as critical points of the Lagrangian function of the multilinear map given by a tensor $f^{(\mathcal{T})}(x^{(1)}, \ldots, x^{(d)}) = \sum_{i_1, \ldots, i_d} t_{i_1, \ldots, i_d} x_{i_1}^{(d)} \cdots x_{i_d}^{(d)}$

as well as $d$ vector valued functions

$$f_i^{(\mathcal{T})}(x^{(1)}, \ldots, \hat{x}^{(i)}, \ldots, x^{(d)}) = \frac{df^{(\mathcal{T})}(x^{(1)}, \ldots, x^{(d)})}{dx^{(i)}} (x^{(1)}, \ldots, \hat{x}^{(i)}, \ldots, x^{(d)})$$

e.g., $f_1^{(\mathcal{T})}(x^{(2)}, x^{(3)}) = T_1(x^{(2)} \otimes x^{(3)})$

- We consider its Lagrangian subject to a normalization condition

$$\|x_1\|_p \cdots \|x_d\|_p = 1$$

$$L(x_1, \ldots, x_d, \sigma) = f(x_1, \ldots, x_d) - \sigma(\|x_1\|_p \cdots \|x_d\|_p - 1)$$

- The first order optimality conditions for even $p$ are, for all $i$ in $\{1, \ldots, d\}$,

$$\frac{dL}{d\hat{x}_i}(x_1, \ldots, x_d, \sigma) = 0 \implies f_i^{(\mathcal{T})}(x_1, \ldots, \hat{x}_i, \ldots, x_d) = \sigma x_i^p$$
Immediate Properties of Tensor Eigenvectors and Singular Vectors

- When the tensor order $d$ is odd, $H$-eigenvectors ($l^d$-eigenvectors) and singular vectors must be defined with additional care
  - Let $\phi_p(x) = [\text{sgn}(x_1)x_1^p, \ldots, \text{sgn}(x_n)x_n^p]^T$ then can generally write
    $$\nabla \|x\|_p = \frac{\phi_{p-1}(x)}{\|x\|_p^{p-1}}$$
    when $p$ is even, $\phi_{p-1}(x) = x^{p-1}$

- The eigenvalue equations can then be written for general $p$ as
  $$\frac{d\mathcal{L}}{dx}(x, \lambda) = 0 \implies f(\mathcal{T})(x) = \lambda \phi_d(x^{d-1})$$

- The largest tensor singular value is the operator/spectral norm of the tensor
  - Recall we defined the operator norm of the tensor as
    $$\| \mathcal{T} \| = \max_{x_1, \ldots, x_d \in S^{n-1}} |f(T)(x_1, \ldots, x_d)|$$
    where $S^{n-1}$ is the unit sphere (norm-1 vectors)
  - This value corresponds to the largest $l^2$ tensor singular value, or in the symmetric case, the largest magnitude of any of the tensor $Z$-eigenvalues
For nonsymmetric matrices case, the Lagrangian approach used above cannot be used to describe the eigenvalues

- The eigenvalues of a real nonsymmetric matrix may be complex
- For tensors, we can still define the eigenvalue equations in a consistent way with respect to matrices,

\[ f_i^{(\mathbf{T})}(x, \ldots, x) = \lambda \phi_p(x) \]

so that \( \lambda, x \) are the mode-1 and mode-2 \( l^2 \)-eigenpair

- For matrices, the mode-1 and mode-2 \( l^2 \)-eigenvectors are the left/right eigenvectors
Connection Between Decomposition and Eigenvalues

- In the matrix-case, the largest magnitude eigenvalue and singular value may be associated with a rank-1 term that gives the best rank-1 decomposition of a matrix
  - *For symmetric matrices, it suffices to consider the dominant eigenpair*
  - *For nonsymmetric matrices, a rank-1 truncated SVD gives the largest singular vector/value pair and associated rank-1 approximation*
- In the tensor case, the rank-1 approximation problem corresponds to a maximization problem\(^4\)
  - *Given a nonsymmetric tensor \(\mathcal{T}\) the rank-1 tensor decomposition objective is*
    \[
    \min_{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d)} \in \mathbb{S}^{n-1}} \| \mathcal{T} - \sigma \mathbf{u}^{(1)} \otimes \cdots \otimes \mathbf{u}^{(d)} \|_F^2
    \]
  - *The problem is equivalent to the maximum \(l^2\)-singular value problem for \(\mathcal{T}\)*
    \[
    \max_{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d)} \in \mathbb{S}^{n-1}} \sigma \quad s.t. \quad \forall i \ f_i(\mathcal{T}) (\mathbf{u}^{(1)}, \ldots, \hat{\mathbf{u}}^{(i)}, \ldots, \mathbf{u}^{(d)}) = \sigma \mathbf{u}^{(i)}
    \]

---

\(^4\) L. De Lathauwer, B. De Moor, and J. Vandewalle, “On the best rank-1 and rank-(\(R_1, R_2, \ldots, R_n\)) approximation of higher-order tensors”, 2000
Derivation of Equivalence

- The singular value problem can be derived from decomposition via the method of Lagrange multipliers

  - In general, consider the Lagrangian function

    \[ \mathcal{L}(u^{(1)}, \ldots, u^{(d)}, \sigma, \lambda) = \|\mathcal{T} - \sigma u^{(1)} \otimes \cdots \otimes u^{(d)}\|_F^2 + \sum_i \lambda_i (\sum_j (\|u_j^{(i)}\|_2^2 - 1)) \]

  - For order 3, we have

    \[ \mathcal{L}(u, v, w, \sigma, \lambda) = \|\mathcal{T} - \sigma u \otimes v \otimes w\|_F^2 + \lambda_1 (u^T u - 1) + \lambda_2 (v^T v - 1) + \lambda_3 (w^T w - 1) \]

  - The optimality conditions give

    \[
    \begin{align*}
    \frac{d\mathcal{L}}{d\lambda} &= 0 \quad \Rightarrow \quad u^T u = 1, \quad v^T v = 1, \quad w^T w = 1 \\
    \frac{d\mathcal{L}}{d\sigma} &= 0 \quad \Rightarrow \quad f^{(\mathcal{T})}(u, v, w) = \sigma \\
    \frac{d\mathcal{L}}{du} &= 0 \quad \Rightarrow \quad \sigma f^{(\mathcal{T})}_1(v, w) = (\sigma^2 + \lambda_1)u
    \end{align*}
    \]

    and similar for \( \frac{d\mathcal{L}}{dv}, \frac{d\mathcal{L}}{dw} \). Premultiplying the last condition by \( u^T \), gives the second modulo \( \lambda_1 \), so \( \lambda_1 = 0 \), giving the singular value equation \( f^{(\mathcal{T})}_1(v, w) = \sigma u \).
Hardness of Eigenvalue Computation

- Like rank-1 approximation, computing eigenvalues of singular values of a tensor is NP-hard, which can be demonstrated by considering the tensor bilinear feasibility problem\(^5\)
  - Restricting the tensor to be symmetric still leads to NP-hard problems, the largest singular vector will be the largest eigenvector a result of Banach\(^6\)

$$\max_{x, y, z \in S^{n-1}} f(T)(x, y, z) = \max_{x \in S^{n-1}} f(T)(x, x, x)$$

- The tensor bilinear feasibility problem associated with an order 3 tensor \(T\) is defined by the set of equations

\[
\begin{align*}
  f_1(T)(v, w) &= 0, \\
  f_2(T)(u, w) &= 0, \\
  f_3(T)(u, v) &= 0
\end{align*}
\]

where we seek solutions \(u, v, w \neq 0\)

- This problem is a special case of the \(l^p\) singular value problem for any choice of \(p\) with \(\sigma = 0\)

---

\(^{5}\)C.J. Hillar and L.-H. Lim, “Most tensor problems are NP-hard”, 2013

\(^{6}\)S. Banach, “On homogeneous polynomials in \(L^2\)”, 1938
NP-hardness of the tensor bilinear feasibility problem can be demonstrated by reduction from 3-colorability

*The 3-coloring problem seeks to find (if possible) an assignment of one of 3 colors to each vertex of a graph that is different from the color of any of its neighbors*

We define an optimization problem over a set of variables $x \in \mathbb{C}^n$ that describe the color (each will take on a power of the third root of unity), as well as auxiliary variables $y \in \mathbb{C}^n$, $z \in \mathbb{C}$, then define the bilinear equations

\begin{align*}
\forall i \in \{1, \ldots, n\}, & \quad x_i y_i - z^2 = 0, \quad y_i z - x^2 = 0, \quad x_i z - y_i^2 = 0 \\
\forall i \in \{1, \ldots, n\}, & \quad \sum_{(i, j) \in E} \frac{x_i^2 + x_i x_j + x_j^2}{x_i^3 - x_j^3}(\frac{x_i^3 - x_j^3}{x_i - x_j})
\end{align*}

Assume (normalize) so that $z = 1$, then the first set of equations implies $y_i = 1/x_i$ and further $x_i^3 = 1$, so labels are cubic roots of unity

For the second set of equations, we then must have $x_i \neq x_j$ if $(i, j) \in E$
The **high-order power method (HOPM)** can be used to compute the largest singular value\(^7\)

- The algorithm updates factors in an alternating manner until convergence, with the \(i\)th factor matrix updated as
  1. \(v^{(i)} = f_i^T(u^{(1)}, \ldots, \hat{u}^{(i)}, \ldots, u^{(d)})\),
  2. \(\sigma = \|v^{(i)}\|_2\)
  3. \(u_{\text{new}}^{(i)} = v^{(i)}/\sigma\)

- The algorithm can be derived from the Lagrangian and converges to a local minimum

- Effective initialization can be achieved by HOSVD and the algorithm is equivalent to the rank-1 version of the HOOI procedure

---

\(^7\)L. De Lathauwer, B. De Moor, and J. Vandewalle, “On the best rank-1 and rank-\((R_1, R_2, \ldots, R_n)\) approximation of higher-order tensors”, 2000
The HOPM algorithm can be adapted to symmetric tensors

- The aforementioned Banach’s polynomial maximization theorem implies HOPM will converge to symmetric solution even if intermediate results are nonsymmetric
- If symmetry is enforced on the iterates, so that

\[ v = f(\mathcal{T})(u) = f_i(\mathcal{T})(u, \ldots, u), \quad u^{(\text{new})} = v/\|v\|, \]

the algorithm is no longer guaranteed to converge (it does if the tensor order is even and the underlying function is convex)

- The shifted symmetric HOPM method\(^8\) alleviates this problem and enables convergence to other eigenvalues by adding a shift so as to minimize

\[ f(\mathcal{T})(u) + \alpha(u^T u)^{d/2} \]

for order \(d\) tensor \(\mathcal{T}\), yielding to updates such as

\[ v = f(\mathcal{T})(u) + \alpha u, \quad u^{(\text{new})} = v/\|v\|, \]

The Perron-Frobenius theorem states that positive matrices have a unique real eigenvalue and the associated eigenvector is positive. Can be extended to nonnegative matrices so long as matrix is not reducible, i.e., cannot be put into the form

\[ PAP^{-1} = \begin{bmatrix} E & F \\ 0 & G \end{bmatrix} \]

where \( P \) is a permutation matrix and \( G \) has at least 1 row.

This theorem is prominent in the study of nonsymmetric matrices. Its applications include analysis of stochastic processes and algebraic graph theory.

Tensor eigenvalues satisfy a generalized Perron-Frobenius theorem. If tensor is positive, the eigenvector with the largest eigenvalue is positive. A nonnegative order \( d \) tensor is irreducible if there is no \( d \)-dimensional blocking into \( 2^d \) blocks that yields an off-diagonal zero block.

Tensor Eigenvalues and Hypergraphs

- Matrix eigenvalues are prominent in algebraic graph theory
  - For an unweighted graph we typically consider a binary adjacency matrix $A$ or the Laplacian matrix $D - A$ where $D$ is a diagonal degree matrix
  - The eigenvector with the second smallest eigenvalue can be used to find a partitioning of vertices with a provably small cut value
  - Clustering can be done via constrained low-rank approximations methods

- Tensor eigenvalues can be used to understand partitioning/clustering properties of uniform hypergraphs\(^9\)
  - A uniform hypergraph $H = (V, E)$ is described by a set of vertices $V$ and a set of hyperedges $E$, each of which is a subset of $r$ vertices in $E$
  - Each hyperedge $(v_i, v_j, v_k) \in E$ may be associated with a tensor entry $t_{ijk}$
  - Laplacian-like choice of $t_{ijk}$ yields symmetric and semidefinite tensor
  - The tensor must have a zero eigenvalue and the multiplicity of the zero eigenvalue is the number of components in the hypergraph
  - The second smallest eigenvalue lower bounds the minimum cut of $H$

---
\(^9\) J. Chang, Y. Chen, L. Qi, H. Yan, ”Hypergraph Clustering Using a New Laplacian Tensor with Applications in Image Processing”, 2019