CS 598 EVS: Tensor Computations

Tensor Eigenvalues

Edgar Solomonik

University of Illinois, 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 $\boldsymbol{M}$ and associated linear function $\boldsymbol{f}^{(\boldsymbol{M})}(\boldsymbol{x}) = \boldsymbol{M}\boldsymbol{x}$
  - *Eigenvalues and eigenvectors can be used to characterize eigenfunctions of differential operators*
  - *Eigenvalues describe powers of the matrix and its limiting behavior*

$$\boldsymbol{M} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1} \quad \Rightarrow \quad \boldsymbol{M}^2 = \boldsymbol{X}\boldsymbol{D}^2\boldsymbol{X}^{-1}$$

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

$$\lim_{k\to\infty} \boldsymbol{M}^k / \|\boldsymbol{M}^{k-1}\| = \lambda \boldsymbol{x}\boldsymbol{y}$$

  - *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 $\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}$ by analogy from the role of matrix eigenvalues on $\boldsymbol{f}^{(\boldsymbol{M})}$
  - *Matrix eigenpairs* (λ, x) *satisfy* $f^{(\boldsymbol{M})}(\boldsymbol{x}) = \lambda\boldsymbol{x}$, *while for an order* d *symmetric tensor, we may define*[1,2]

$$\underbrace{\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x},\dots,\boldsymbol{x}) = \lambda\boldsymbol{x}}_{\text{Z-eigenpair}} \quad \underbrace{\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x},\dots,\boldsymbol{x}) = \lambda\boldsymbol{x}^{d-1}}_{\text{H-eigenpair}} \quad \underbrace{\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x},\dots,\boldsymbol{x}) = \lambda\boldsymbol{x}^{p-1}}_{l^p\text{-eigenpair}}$$

    where $\boldsymbol{x}^p = [x_1^p \dots x_n^p]^T$
  - *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* $(\sigma, \boldsymbol{x}_1, \dots, \boldsymbol{x}_d)$ *that satisfies* d *equations like* $\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}_2, \dots, \boldsymbol{x}_d) = \sigma\boldsymbol{x}_1^p$*, e.g., for* $d = 3, p = 1$*,*

$$\boldsymbol{T}_{(1)}(\boldsymbol{x}_2 \otimes \boldsymbol{x}_3) = \sigma\boldsymbol{x}_1, \quad \boldsymbol{T}_{(2)}(\boldsymbol{x}_1 \otimes \boldsymbol{x}_3) = \sigma\boldsymbol{x}_2, \quad \boldsymbol{T}_{(3)}(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2) = \sigma\boldsymbol{x}_3$$

---

[1] Liqun Qi, “Eigenvalues of a Real Supersymmetric Tensor”, 2005

[2] Lek-Heng Lim, “Singular Values and Eigenvalues of Tensors: A Variational Approach”, 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* $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ *subject to* $\|\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2\|\boldsymbol{x}\|_2 = 1$ *is*
  $$\mathcal{L}(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \lambda(\|\boldsymbol{x}\|_2^2 - 1)$$
  - *The first-order optimality conditions are* $\|\boldsymbol{x}\|_2 = 1$ *and*
  $$\frac{d\mathcal{L}}{d\boldsymbol{x}}(\boldsymbol{x}, \lambda) = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$
- Singular vectors and singular values of matrices may be derived analogously
  - *The Lagrangian function of* $f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}$ *subject to* $\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 = 1$ *is*
  $$\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \sigma) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y} - \sigma(\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 - 1)$$
  - *The first-order optimality conditions are* $\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 = 1$ *and*
  $$\frac{d\mathcal{L}}{d\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{y}, \sigma) = \mathbf{0} \quad \Rightarrow \quad \frac{\boldsymbol{A}\boldsymbol{y}}{\|\boldsymbol{y}\|} = \frac{\sigma\boldsymbol{x}}{\|\boldsymbol{x}\|}, \qquad \frac{d\mathcal{L}}{d\boldsymbol{y}}(\boldsymbol{x}, \boldsymbol{y}, \sigma) = \mathbf{0} \quad \Rightarrow \quad \frac{\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|} = \frac{\sigma\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

[3] Lek-Heng Lim, “Singular Values and Eigenvalues of Tensors: A Variational Approach”, 2005

Tensors Eigenvalues

- The Lagrangian approach to matrix eigenvalues generalizes naturally to symmetric tensors
  - *The symmetric tensor is associated with a multilinear scalar-valued function* $f^{(\mathcal{T})}(\boldsymbol{x}) = \sum_{i_1,\dots i_d} t_{i_1,\dots,i_d} x_{i_1} \cdots x_{i_d}$ *as well as the vector valued function* $\boldsymbol{f}^{(\mathcal{T})}(\boldsymbol{x}) = \sum_{i_1,\dots i_{d-1}} t_{i_1,\dots,i_{d-1}} x_{i_1} \cdots x_{i_{d-1}} = \frac{1}{d} \nabla f^{(\mathcal{T})}(\boldsymbol{x})$
  - *We consider its Lagrangian subject to a normalization condition* $\|\boldsymbol{x}\|_p^d = 1$ *(for matrices* $p = 2$*, so for order* d *tensors natural to pick either* $p = 2$ *or* $p = d$*),*

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

  - *The first order optimality conditions for* $p = 2$ *is* $\|\boldsymbol{x}\|_2 = 1$ *and*

$$\frac{d\mathcal{L}}{d\boldsymbol{x}}(\boldsymbol{x}, \lambda) = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{f}^{(\mathcal{T})}(\boldsymbol{x}) = \lambda \boldsymbol{x}$$

  - *The analogous first order optimality condition for* $p = d$ *and even* p *is*

$$\frac{d\mathcal{L}}{d\boldsymbol{x}}(\boldsymbol{x}, \lambda) = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{f}^{(\mathcal{T})}(\boldsymbol{x}) = \lambda \boldsymbol{x}^{d-1}$$

    is scale invariant (if $(\boldsymbol{x}*, \lambda)$ *minimizes* $\mathcal{L}$ *so does* $(\alpha \boldsymbol{x}^*, \lambda)$*)*

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
  - *An order d tensor is associated with a multilinear scalar-valued function*
    $$f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(d)}) = \sum_{i_1,\dots i_d} t_{i_1,\dots,i_d} x_{i_1}^{(d)} \cdots x_{i_d}^{(d)}$$
    as well as d vector valued functions
    $$\boldsymbol{f}_i^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}^{(1)}, \dots, \hat{\boldsymbol{x}}^{(i)}, \dots, \boldsymbol{x}^{(d)}) = \frac{df^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(d)})}{d\boldsymbol{x}^{(i)}}(\boldsymbol{x}^{(1)}, \dots, \hat{\boldsymbol{x}}^{(i)}, \dots, \boldsymbol{x}^{(d)})$$
    e.g., $\boldsymbol{f}_1^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}) = \boldsymbol{T}_{(1)}(\boldsymbol{x}^{(2)} \otimes \boldsymbol{x}^{(3)})$
  - *We consider its Lagrangian subject to a normalization condition* $\|\boldsymbol{x}_1\|_p \cdots \|\boldsymbol{x}_d\|_p = 1$
    $$\mathcal{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_d, \sigma) = f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_d) - \sigma(\|\boldsymbol{x}_1\|_p \cdots \|\boldsymbol{x}_d\|_p - 1)$$
  - *The first order optimality conditions for even p are, for all i in $\{1, \dots, d\}$,*
    $$\frac{d\mathcal{L}}{d\boldsymbol{x}_i}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_d, \sigma) = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{f}_i^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}_1, \dots, \hat{\boldsymbol{x}}_i, \dots, \boldsymbol{x}_d) = \sigma \boldsymbol{x}_i^p$$

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* $\phi_p(\boldsymbol{x}) = [\mathit{sgn}(x_1)|x_1|^p, \ldots, \mathit{sgn}(x_n)|x_n|^p]^T$ *then can generally write*
    $$\nabla\|\boldsymbol{x}\|_p = \phi_{p-1}(\boldsymbol{x})/\|\boldsymbol{x}\|_p^{p-1}$$
    when p *is even,* $\phi_{p-1}(\boldsymbol{x}) = \boldsymbol{x}^{p-1}$
  - *The eigenvalue equations can then be we written for general* p *as*
    $$\frac{d\mathcal{L}}{d\boldsymbol{x}}(\boldsymbol{x},\lambda) = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}) = \lambda\phi_{p-1}(\boldsymbol{x})$$
- The largest tensor singular value is the operator/spectral norm of the tensor
  - *Recall we defined the operator norm of the tensor as*
    $$\|\boldsymbol{\mathcal{T}}\| = \max_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d \in \mathbb{S}^{n-1}} |f^{\boldsymbol{\mathcal{T}}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d)|$$
    where $\mathbb{S}^{n-1}$ *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*

Eigenvalues of Nonsymmetric Tensors

- 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,*

    $$\boldsymbol{f}_i^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}, \ldots, \boldsymbol{x}) = \lambda \phi_{p-1}(\boldsymbol{x})$$

    so that $\lambda, \boldsymbol{x}$ are the mode-i an l^p-eigenpair
  - *For matrices, the mode-1 and mode-2 l^2-eigenvectors are the left/right eigenvectors*

Another View of Symmetric Tensor Eigenvalues

- We can characterize eigenvectors with $\lambda = 0$ by considering the polynomial associated with a symmetric tensor, and similar with others.[4]
  - *The zero eigenvalectors are defined by the singular points on the hypersurface* $\{\boldsymbol{x} : f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}) = 0\}$ *(singular points correspond to points where the gradient of* $f^{(\boldsymbol{\mathcal{T}})}$ *vanishes).*
  - *The polynomial* $f^{(T)}(\boldsymbol{x}) \geqslant 0$ *if all the eigenvalues of* $\boldsymbol{\mathcal{T}}$ *are nonnegative*
  - *For a general* $\lambda \in \mathbb{C}$*, Z-eigenvectors correspond to singular points on the hypersurface*
$$\{\boldsymbol{x} : f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}) - (\lambda/2)\boldsymbol{x}^T\boldsymbol{x} - (1/d - 1/2)\lambda = 0\}$$

[4] Cartwright, Dustin, and Bernd Sturmfels. "The number of eigenvalues of a tensor." Linear algebra and its applications 438.2 (2013): 942-952.

Number of eigenvalues in a tensor

- The number of eigenvalues over $\mathbb{C}$ can be derived from the view of polynomial equations
  - *For any tensor $\mathcal{T} \in \mathbb{C}^{n\times\cdots\times n}$ of order $d \geqslant 3$, the number of eigenvalues is either infinite or at most (with multiplicity)*

$$\frac{(d-1)^n - 1}{d-2}$$

  - *A generic (randomly chosen) tensor attains this bound and has all multiplicties equal to one*
  - *Symmetric tensors may not have an infinite number of eigenvalues, so long as eigenvectors are defined with the normalization $\boldsymbol{x}^T\boldsymbol{x} = 1$ (and not $\boldsymbol{x}^H\boldsymbol{x} = 1$), excluding eigenvectors that may not be normalized*
- In the real case, eigenvalues need not be real
  - *If $\mathcal{T}$ has real entries and either n or d is odd, it has a real eigenpair, otherwise it may not*

Example of Symmetric Tensor with Infinite Eigenvalues

- Concretely, the following symmetric tensor has infinite eigenvalues if they are normalized as $\boldsymbol{x}^H\boldsymbol{x}$[5]

$$a_{111} = 2, a_{122} = a_{212} = a_{221} = a_{133} = a_{313} = a_{331} = 1$$

and otherwise $a_{ijk} = 0$

  - *The eigenvalues of $\boldsymbol{\mathcal{A}}$ are solutions to the equations*

$$2x_1^2 + x_2^2 + x_3^2 = \lambda x_1, 2x_1x_2 = \lambda x_2, 2x_1x_3 = \lambda x_3,$$

  - *For any α, $\boldsymbol{x} = [1, i\alpha, \alpha]$ is an eigenvector of $\boldsymbol{\mathcal{A}}$*
  - *Then, $\boldsymbol{x}/\|\boldsymbol{x}\|_2$ gives an eigenvector with eigenvalue $2/\sqrt{1+2|\alpha|}$, while rescaling as $\boldsymbol{x}/(\boldsymbol{x}^T\boldsymbol{x})$ gives eigenvectors that have eigenvalue 2 for any α*
  - *On the other hand, the latter normalization is not always possible, since for $\boldsymbol{x} \in \mathbb{C}^n$, we can have $\boldsymbol{x} \neq 0$, but $\boldsymbol{x}^T\boldsymbol{x} = 0$*

[5] Cartwright, Dustin, and Bernd Sturmfels. "The number of eigenvalues of a tensor." Linear algebra and its applications 438.2 (2013): 942-952.

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[6]
  - *Given a nonsymmetric tensor $\boldsymbol{\mathcal{T}}$ the rank-1 tensor decomposition objective is*

$$\min_{\boldsymbol{u}^{(1)},\ldots,\boldsymbol{u}^{(d)}\in\mathbb{S}^{n-1}} \|\boldsymbol{\mathcal{T}} - \sigma \boldsymbol{u}^{(1)} \otimes \cdots \otimes \boldsymbol{u}^{(d)}\|_F^2$$

  - *The problem is equivalent to the maximum l^2-singular value problem for $\boldsymbol{\mathcal{T}}$*

$$\max_{\boldsymbol{u}^{(1)},\ldots,\boldsymbol{u}^{(d)}\in\mathbb{S}^{n-1}} \sigma \quad \textit{s.t.} \quad \forall_i \ \boldsymbol{f}_i^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}^{(1)},\ldots,\hat{\boldsymbol{u}}^{(i)},\ldots,\boldsymbol{u}^{(d)}) = \sigma \boldsymbol{u}^{(i)},$$

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(R_1, R_2,..., 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*

$$\mathcal{L}(\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(d)}, \sigma, \boldsymbol{\lambda}) = \|\boldsymbol{\mathcal{T}} - \sigma \boldsymbol{u}^{(1)} \otimes \cdots \otimes \boldsymbol{u}^{(d)}\|_F^2 + \sum_i \lambda_i \Big(\sum_j (\boldsymbol{u}_j^{(i)T} \boldsymbol{u}_j^{(i)} - 1)\Big)$$

  - *For order 3, we have*

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

  - *The optimality conditions give*

$$\begin{aligned}
\frac{d\mathcal{L}}{d\boldsymbol{\lambda}} = \mathbf{0} \quad &\Rightarrow \quad \boldsymbol{u}^T \boldsymbol{u} = 1, \quad \boldsymbol{v}^T \boldsymbol{v} = 1, \quad \boldsymbol{w}^T \boldsymbol{w} = 1 \\
\frac{d\mathcal{L}}{d\sigma} = \mathbf{0} \quad &\Rightarrow \quad f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \sigma \\
\frac{d\mathcal{L}}{d\boldsymbol{u}} = \mathbf{0} \quad &\Rightarrow \quad \sigma \boldsymbol{f}_1^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{v}, \boldsymbol{w}) = (\sigma^2 + \lambda_1)\boldsymbol{u}
\end{aligned}$$

  and similar for $\frac{d\mathcal{L}}{d\boldsymbol{v}}$, $\frac{d\mathcal{L}}{d\boldsymbol{w}}$. Premultiplying the last condition by $\boldsymbol{u}^T$, gives the second modulo λ_1, so $\lambda_1 = 0$, giving the singular value equation $\boldsymbol{f}_1^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{v}, \boldsymbol{w}) = \sigma \boldsymbol{u}$.

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[7]
  - *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*[8]

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

  - *The tensor bilinear feasibility problem associated with an order 3 tensor $\boldsymbol{\mathcal{T}}$ is defined by the set of equations*

$$\boldsymbol{f}_1^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{v},\boldsymbol{w}) = \boldsymbol{0}, \quad \boldsymbol{f}_2^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u},\boldsymbol{w}) = \boldsymbol{0}, \quad \boldsymbol{f}_3^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{0}$$

    where we seek solutions $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w} \neq \boldsymbol{0}$
  - *This problem is a special case of the l^p singular value problem for any choice of p with $\sigma = 0$, similar ideas (with a bit more technology) have been used to show hardness of maximization / rank-1 approximation*

---

[7] C.J. Hillar and L.-H. Lim, “Most tensor problems are NP-hard”, 2013

[8] S. Banach, “On homogeneous polynomials in L^2”, 1938

Hardness of Eigenvalue Computation

- 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 set of equations over variables* $\boldsymbol{x} \in \mathbb{C}^n$ *that describe the color (each will take on a power of the third root of unity), as well as auxiliary variables* $\boldsymbol{y} \in \mathbb{C}^n, z \in \mathbb{C}$*,*

$$\forall i \in \{1, \dots, n\}, \quad x_i y_i - z^2 = 0, \quad y_i z - x^2 = 0, \quad x_i z - y_i^2 = 0$$
$$\forall i \in \{1, \dots, n\}, \quad \sum_{(i,j) \in E} \underbrace{x_i^2 + x_i x_j + x_j^2}_{\frac{x_i^3 - x_j^3}{x_i - x_j}} = 0$$

  - *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* $(i, j) \in E$

Power Method for Singular Value Computation

- The *high-order power method (HOPM)* can be used to compute the largest singular value[9]
  - *The algorithm updates factors in an alternating manner until convergence, with the ith factor matrix updated as*
    1. $\boldsymbol{v}^{(i)} = \boldsymbol{f}_i^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}^{(1)}, \dots, \hat{\boldsymbol{u}}^{(i)}, \dots, \boldsymbol{u}^{(d)})$,
    2. $\sigma = \|\boldsymbol{v}^{(i)}\|_2$
    3. $\boldsymbol{u}_{new}^{(i)} = \boldsymbol{v}^{(i)}/\sigma$
  - *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*

[9]L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(R_1, R_2,..., R_n) approximation of higher-order tensors", 2000

Power Method for Symmetric Eigenvalue Problems

- 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*

$$\boldsymbol{v} = \boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}) = \boldsymbol{f}_i^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}, \dots, \boldsymbol{u}), \quad \boldsymbol{u}^{(\textit{new})} = \boldsymbol{v}/\|\boldsymbol{v}\|,$$

    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*[10] *alleviates this problem and enables convergence to other eigenvalues by adding a shift so as to minimize* $\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}) + \alpha(\boldsymbol{u}^T\boldsymbol{u})^{d/2}$ *for order* d *tensor* $\boldsymbol{\mathcal{T}}$*, yielding to updates such as*

$$\boldsymbol{v} = \boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{u}) + \alpha\boldsymbol{u}, \quad \boldsymbol{u}^{(\textit{new})} = \boldsymbol{v}/\|\boldsymbol{v}\|,$$

[10] T.G. Kolda and J.R. Mayo, "Shifted Power Method for Computing Tensor Eigenpairs", 2011

Newton-based Methods for Eigenvalue Computation

- A state-of-the-art method of Newton-type method Newton Correction Method (NCM) for computing real eigenvectors of a symmetric tensor[11]
  - *The gradient and Hessian of the Lagrangian function* $\mathcal{L}(\boldsymbol{x},\lambda) = f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}) - \frac{d\lambda}{2}(\boldsymbol{x}^T\boldsymbol{x}-1)$*, at a critical point (at which* $\lambda = f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x})$*) satisfy*

$$\frac{1}{d}\nabla\mathcal{L}(\boldsymbol{x},\lambda) = \boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}) - f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x})\boldsymbol{x}$$
$$\frac{1}{d}\boldsymbol{H}_{\mathcal{L}}(\boldsymbol{x},\lambda) = (d-1)\boldsymbol{T}_{(1,2)}(\boldsymbol{x}\otimes\cdots\otimes\boldsymbol{x})\boldsymbol{f}^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x}) - f^{(\boldsymbol{\mathcal{T}})}(\boldsymbol{x})\boldsymbol{I}$$

    where $\boldsymbol{T}_{1,2} \in \mathbb{R}^{n^2\times n^{d-2}}$ *is a matricization with the first two modes of the tensor enumerating matrix rows*
  - *To first-order in* $\boldsymbol{y} = \boldsymbol{x}^* - \boldsymbol{x}$*, for eigenvector* $\boldsymbol{x}^*$*, the eigenvector equations reduce to*

$$(\boldsymbol{H}_{\mathcal{L}}(\boldsymbol{x},\lambda) - d\cdot\boldsymbol{x}\boldsymbol{f}^{(T)}(\boldsymbol{x})^T)\boldsymbol{y} = -\nabla\mathcal{L}(\boldsymbol{x},\lambda)$$

  - *The NCM method achieves quadratic convergence provided the Hessian at the eigenvector is positive definite*

[11] Jaffe, Ariel, Roi Weiss, and Boaz Nadler. ”Newton correction methods for computing real eigenpairs of symmetric tensors.” SIAM Journal on Matrix Analysis and Applications 39.3 (2018).

Perron-Frobenius Theorem for Tensor Eigenvalues

- 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 in not reducible, i.e., cannot be put into the form*

$$\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1} = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \mathbf{0} & \boldsymbol{G} \end{bmatrix}$$

    where P is a permutation matrix and $\boldsymbol{G}$ 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*
  - *For further properties, see LH Lim, "Singular Values and Eigenvalues of Tensors: A Variational Approach", 2005 and Q Yang, Y Yang, "Further results for Perron–Frobenius theorem for nonnegative tensors II", 2011*

Tensor Eigenvalues and Hypergraphs

- Matrix eigenvalues are prominent in algebraic graph theory
  - *For an unweighted graph we typically consider a binary adjacency matrix* $\boldsymbol{A}$ *or the Laplacian matrix* $\boldsymbol{D} - \boldsymbol{A}$ *where* $\boldsymbol{D}$ *is a diagonal degree matrix*
  - *The eigenvector with the second smallest eigenvalue can be used to find a partitioning of verticies 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[12]
  - *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

---

[12] J. Chang, Y. Chen, L. Qi, H. Yan, "Hypergraph Clustering Using a New Laplacian Tensor with Applications in Image Processing", 2019