Multilinear Least Square, Eigenvalue, and Singular Value Problems

Lek-Heng Lim

Linear Algebra and Optimization Seminar
October 11, 2006

Thanks: Sou-Cheng Choi, Vin de Silva, Gene Golub, Morten Mørup, Liqun Qi, Michael Saunders, Berkant Savas

$$\left[a_{ij}\right]_{l\times m}\left[b_{jk}\right]_{m\times n}=\left[\sum_{j=1}^{m}a_{ij}b_{jk}\right]_{l\times n}$$

Tensors

A set of multiply indexed real numbers $A = [\![a_{ijk}]\!]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l \times m \times n}$ on which the following algebraic operations are defined:

1. Addition/Scalar Multiplication: for $[\![b_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$, $\lambda \in \mathbb{R}$,

   $[\![a_{ijk}]\!] + [\![b_{ijk}]\!] := [\![a_{ijk} + b_{ijk}]\!]$ and $\lambda [\![a_{ijk}]\!] := [\![\lambda a_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$

2. Multilinear Matrix Multiplication: for matrices $L = [\lambda_{i'i}] \in \mathbb{R}^{p \times l}, M = [\mu_{j'j}] \in \mathbb{R}^{q \times m}, N = [\nu_{k'k}] \in \mathbb{R}^{r \times n}$,

   $$(L, M, N) \cdot A := [\![c_{i'j'k'}]\!] \in \mathbb{R}^{p \times q \times r}$$

   where

   $$c_{i'j'k'} := \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \lambda_{i'i} \mu_{j'j} \nu_{k'k} a_{ijk}.$$

May think of A as a 3-dimensional array of numbers. $(L, M, N) \cdot A$ as multiplication on '3 sides' by matrices L, M, N.

Outer product rank

$\mathbf{u} \in \mathbb{R}^l$, $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{w} \in \mathbb{R}^n$, outer product defined by

$$\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} = [\![u_i v_j w_k]\!]_{i,j,k=1}^{l,m,n}.$$

A tensor $A \in \mathbb{R}^{l \times m \times n}$ is said to be decomposable if it can be written in the form

$$A = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}.$$

$A \in \mathbb{R}^{l \times m \times n}$, outer product rank is

$$\operatorname{rank}_{\otimes}(A) = \min\{r \mid A = \textstyle\sum_{i=1}^{r} \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i\}.$$

Tensor rank is difficult

Mystical Power of Twoness (Eugene L. Lawler). 2-SAT is easy, 3-SAT is hard; 2-dimensional matching is easy, 3-dimensional matching is hard; etc.

Matrix rank is easy, tensor rank is hard:

Theorem (Håstad). Computing $\operatorname{rank}_{\otimes}(A)$ for $A \in \mathbb{R}^{l \times m \times n}$ is an NP-hard problem.

Tensor rank depends on base field:

Theorem (Bergman). For $A \in \mathbb{R}^{l \times m \times n} \subset \mathbb{C}^{l \times m \times n}$, $\operatorname{rank}_{\otimes}(A)$ is base field dependent.

$\boxed{\text{Best rank-}r\text{ approximation of tensors}}$

Given $A \in \mathbb{R}^{l\times m\times n}$, solve

$$\operatorname{argmin}_{\operatorname{rank}_{\otimes}(B)\leq r}\|A - B\|_F.$$

No solution for all orders > 2, all norms, and many ranks:

Theorem 1 (de Silva, L). Let $k \geq 3$ and $d_1, \dots, d_k \geq 2$. For any s such that $2 \leq s \leq \min\{d_1, \dots, d_k\} - 1$, there exist $A \in \mathbb{R}^{d_1\times\cdots\times d_k}$ with $\operatorname{rank}_{\otimes}(A) = s$ such that A has no best rank-r approximation for some $r < s$. The result is independent of the choice of norms.

Tensor rank can jump over an arbitrarily large gap:

Theorem 2 (de Silva, L). Let $k \geq 3$. Given any $s \in \mathbb{N}$, there exists a sequence of order-k tensor A_n such that $\operatorname{rank}_{\otimes}(A_n) \leq r$ and $\lim_{n\to\infty} A_n = A$ with $\operatorname{rank}_{\otimes}(A) = r + s$.

Tensors that fail to have best low-rank approximations are not rare — they occur with non-zero probability:

Theorem 3 (de Silva, L). Let μ be a measure that is positive or infinite on Euclidean open sets in $\mathbb{R}^{d_1 \times \cdots \times d_k}$. There exists some $r \in \mathbb{N}$ such that

$$\mu(\{A \mid A \text{ does not have a best rank-}r\text{ approximation}\}) > 0.$$

Note 1. It is incorrect to think that if we just want an 'approximate solution', then this doesn't matter. If there is no solution in the first place, then what is it that are we trying to approximate? ie. what is the 'approximate solution' an approximate of?

Note 2. That the best rank-r approximation problem lacks a solution is not the same as the phenomenon commonly referred to as 'degeneracy' in psychometrics.

$\boxed{\textbf{Message}}$

Best rank-r approximation problem for tensors is difficult.

Let's study something else.

Symmetric tensors

$A = [\![a_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. For a permutation $\sigma \in \mathfrak{S}_k$, *$\sigma$-transpose of A* is

$$A^\sigma = [\![a_{i_{\sigma(1)} \cdots i_{\sigma(k)}}]\!] \in \mathbb{R}^{d_{\sigma(1)} \times \cdots \times d_{\sigma(k)}}.$$

Order-k generalization of 'taking transpose'.

For matrices (order-2), only one way to take transpose (ie. swapping row and column indices) since $\mathfrak{S}_2$ has only one non-trivial element. For an order-k tensor, there are $k! - 1$ different 'transposes' — one for each non-trivial element of $\mathfrak{S}_k$.

An order-k tensor $A = [\![a_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{n \times \cdots \times n}$ is called *symmetric* if $A = A^\sigma$ for all $\sigma \in \mathfrak{S}_k$, ie.

$$a_{i_{\sigma(1)} \cdots i_{\sigma(k)}} = a_{i_1 \cdots i_k}.$$

Rayleigh-Ritz approach to eigenpairs

$A \in \mathbb{R}^{n\times n}$ symmetric. Its eigenvalues and eigenvectors are critical values and critical points of Rayleigh quotient

$$\mathbb{R}^n\backslash\{\mathbf{0}\} \to \mathbb{R}, \qquad \mathbf{x} \mapsto \frac{\mathbf{x}^\top A\mathbf{x}}{\|\mathbf{x}\|_2^2}$$

or equivalently, critical values/points constrained to unit vectors, ie. $S^{n-1} = \{x \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 = 1\}$. Associated Lagrangian is

$$L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, \qquad L(\mathbf{x}, \lambda) = \mathbf{x}^\top A\mathbf{x} - \lambda(\|\mathbf{x}\|_2^2 - 1).$$

At a critical point $(\mathbf{x}_c, \lambda_c) \in \mathbb{R}^n\backslash\{\mathbf{0}\} \times \mathbb{R}$, we have

$$A\frac{\mathbf{x}_c}{\|\mathbf{x}_c\|_2} = \lambda_c\frac{\mathbf{x}_c}{\|\mathbf{x}_c\|_2} \qquad \text{and} \qquad \|\mathbf{x}_c\|_2^2 = 1.$$

Write $\mathbf{u}_c = \mathbf{x}_c/\|\mathbf{x}_c\|_2 \in S^{n-1}$. Get usual

$$A\mathbf{u}_c = \lambda_c\mathbf{u}_c.$$

Variational characterization of singular triples

Similar approach for singular triples of $A \in \mathbb{R}^{m\times n}$: singular values, left/right singular vectors are critical values and critical points of

$$\mathbb{R}^m \backslash \{\mathbf{0}\} \times \mathbb{R}^n \backslash \{\mathbf{0}\} \to \mathbb{R}, \qquad (\mathbf{x}, \mathbf{y}) \mapsto \frac{\mathbf{x}^\top A \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$$

Associated Lagrangian is

$$L : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, \qquad L(\mathbf{x}, \mathbf{y}, \sigma) = \mathbf{x}^\top A \mathbf{y} - \sigma(\|\mathbf{x}\|_2 \|\mathbf{y}\|_2 - 1).$$

The first order condition yields

$$A \frac{\mathbf{y}_c}{\|\mathbf{y}_c\|_2} = \sigma_c \frac{\mathbf{x}_c}{\|\mathbf{x}_c\|_2}, \qquad A^\top \frac{\mathbf{x}_c}{\|\mathbf{x}_c\|_2} = \sigma_c \frac{\mathbf{y}_c}{\|\mathbf{y}_c\|_2}, \qquad \|\mathbf{x}_c\|_2 \|\mathbf{y}_c\|_2 = 1$$

at a critical point $(\mathbf{x}_c, \mathbf{y}_c, \sigma_c) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}$. Write $\mathbf{u}_c = \mathbf{x}_c / \|\mathbf{x}_c\|_2 \in S^{m-1}$ and $\mathbf{v}_c = \mathbf{y}_c / \|\mathbf{y}_c\|_2 \in S^{n-1}$, get familiar

$$A\mathbf{v}_c = \sigma_c \mathbf{u}_c, \qquad A^\top \mathbf{u}_c = \sigma_c \mathbf{v}_c.$$

Multilinear functional

$A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$; multilinear functional defined by A is

$$f_A : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} \to \mathbb{R},$$
$$(\mathbf{x}^1, \ldots, \mathbf{x}^k) \mapsto A(\mathbf{x}^1, \ldots, \mathbf{x}^k).$$

Gradient of f_A with respect to $\mathbf{x}^i$,

$$\begin{aligned} \nabla_{\mathbf{x}^i} f_A(\mathbf{x}^1, \ldots, \mathbf{x}^k) &= \left(\frac{\partial f_A}{\partial x_1^i}, \ldots, \frac{\partial f_A}{\partial x_{d_i}^i} \right) \\ &= A(\mathbf{x}^1, \ldots, \mathbf{x}^{i-1}, I_{d_i}, \mathbf{x}^{i+1}, \ldots, \mathbf{x}^k) \end{aligned}$$

where I_{d_i} denotes $d_i \times d_i$ identity matrix.

Multilinear spectral theory

May extend the variational approach to tensors to obtain a theory of eigen/singular values/vectors for tensors (cf. [**L**] for details).

For $\mathbf{x} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$, write

$$\mathbf{x}^p := [x_1^p, \ldots, x_n^p]^\top.$$

We also define the 'ℓ^k-norm'

$$\|\mathbf{x}\|_k = (x_1^k + \cdots + x_n^k)^{1/k}.$$

Define ℓ^2- and ℓ^k-eigenvalues/vectors of $A \in \mathsf{S}^k(\mathbb{R}^n)$ as the critical values/points of the multilinear Rayleigh quotient $A(\mathbf{x}, \ldots, \mathbf{x})/\|\mathbf{x}\|_p^k$. Differentiating the Lagrangian

$$L(\mathbf{x}_1, \ldots, \mathbf{x}_k, \sigma) := A(\mathbf{x}_1, \ldots, \mathbf{x}_k) - \sigma(\|\mathbf{x}_1\|_{p_1} \cdots \|\mathbf{x}_k\|_{p_k} - 1).$$

yields

$$A(I_n, \mathbf{x}, \ldots, \mathbf{x}) = \lambda \mathbf{x}$$

and

$$A(I_n, \mathbf{x}, \ldots, \mathbf{x}) = \lambda \mathbf{x}^{k-1}$$

respectively. Note that for a symmetric tensor A,

$$A(I_n, \mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}) = A(\mathbf{x}, I_n, \mathbf{x}, \ldots, \mathbf{x}) = \cdots = A(\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}, I_n).$$

This doesn't hold for nonsymmetric cubical tensors $A \in \mathsf{S}^k(\mathbb{R}^n)$ and we get different eigenpair for different modes (this is to be expected: even for matrices, a nonsymmetric matrix will have different left/right eigenvectors).

These equations have also been obtained by L. Qi independently using a different approach.

ℓ^2-singular values of a tensor

Lagrangian is

$$L(\mathbf{x}^1, \dots, \mathbf{x}^k, \sigma) = A(\mathbf{x}^1, \dots, \mathbf{x}^k) - \sigma(\|\mathbf{x}^1\|_2 \cdots \|\mathbf{x}^k\|_2 - 1).$$

Then

$$\nabla L = (\nabla_{\mathbf{x}^1} L, \dots, \nabla_{\mathbf{x}^k} L, \nabla_\sigma L) = (\mathbf{0}, \dots, \mathbf{0}, 0).$$

yields

$$\begin{aligned} A\left(I_{d_1}, \frac{\mathbf{x}^2}{\|\mathbf{x}^2\|_2}, \frac{\mathbf{x}^3}{\|\mathbf{x}^3\|_2}, \dots, \frac{\mathbf{x}^k}{\|\mathbf{x}^k\|_2}\right) &= \sigma \frac{\mathbf{x}^1}{\|\mathbf{x}^1\|_2}, \\ &\vdots \\ A\left(\frac{\mathbf{x}^1}{\|\mathbf{x}^1\|_2}, \frac{\mathbf{x}^2}{\|\mathbf{x}^2\|_2}, \dots, \frac{\mathbf{x}^{k-1}}{\|\mathbf{x}^{k-1}\|_2}, I_{d_k}\right) &= \sigma \frac{\mathbf{x}^k}{\|\mathbf{x}^k\|_2}, \\ \|\mathbf{x}^1\|_2 \cdots \|\mathbf{x}^k\|_2 &= 1. \end{aligned}$$

Normalize to get $\mathbf{u}^i = \mathbf{x}^i / \|\mathbf{x}^i\|_2 \in S^{d_i - 1}$. We have

$$\begin{aligned} A(I_{d_1}, \mathbf{u}^2, \mathbf{u}^3, \dots, \mathbf{u}^k) &= \sigma \mathbf{u}^1, \\ &\vdots \\ A(\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^{k-1}, I_{d_k}) &= \sigma \mathbf{u}^k. \end{aligned}$$

Call $\mathbf{u}^i \in S^{d_i - 1}$ mode-i singular vector and σ singular value of A.

Same equations first appeared in the context of rank-1 tensor approximations. Our study differs in that we are interested in all critical values as opposed to only the maximum.

Norms of Multilinear Operators

Recall that the *norm* of a multilinear operator $f : V_\alpha \times \cdots \times V_\gamma \to V_\omega$ from a product of norm spaces $(V_\alpha, \|\cdot\|_\alpha), \ldots, (V_\gamma, \|\cdot\|_\gamma)$ to a norm space $(V_\omega, \|\cdot\|_\omega)$ is defined as

$$\sup \frac{\|f(\mathbf{x}^\alpha, \ldots, \mathbf{x}^\gamma)\|_\omega}{\|\mathbf{x}^\alpha\|_\alpha \cdots \|\mathbf{x}^\gamma\|_\gamma}$$

where the supremum is taken over all $\mathbf{x}^i \neq \mathbf{0}$.

Relation with spectral norm

Define *spectral norm* of a tensor $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ by

$$\|A\|_\sigma := \sup \frac{|A(\mathbf{x}^1, \dots, \mathbf{x}^k)|}{\|\mathbf{x}^1\|_2 \cdots \|\mathbf{x}^k\|_2}.$$

Note that this differs from the *Frobenius norm*,

$$\|A\|_F := \left(\sum_{i_1=1}^{d_1} \cdots \sum_{i_k=1}^{d_k} |a_{i_1 \cdots i_k}|^2\right)^{1/2}$$

for $A = [\![a_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$.

Proposition. Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. The largest singular value of A equals its spectral norm,

$$\sigma_{\max}(A) = \|A\|_\sigma.$$

Hyperdeterminant

Theorem (Gelfand, Kapranov, Zelevinsky, 1992).
$\mathbb{R}^{(d_1+1)\times\cdots\times(d_k+1)}$ has a non-trivial hyperdeterminant iff

$$d_j \leq \sum_{i\neq j} d_i$$

for all $j = 1, \ldots, k$.

For $\mathbb{R}^{m\times n}$, the condition becomes $m \leq n$ and $n \leq m$ — that's why matrix determinants are only defined for square matrices.

Relation with hyperdeterminant

Assume

$$d_i - 1 \leq \sum_{j \neq i} (d_j - 1)$$

for all $i = 1, \ldots, k$. Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. Easy to see that

$$\begin{aligned} A(I_{d_1}, \mathbf{u}^2, \mathbf{u}^3, \ldots, \mathbf{u}^k) &= \mathbf{0}, \\ A(\mathbf{u}^1, I_{d_2}, \mathbf{u}^3, \ldots, \mathbf{u}^k) &= \mathbf{0}, \\ &\vdots \\ A(\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^{k-1}, I_{d_k}) &= \mathbf{0}. \end{aligned}$$

has a solution $(\mathbf{u}^1, \ldots, \mathbf{u}^k) \in S^{d_1 - 1} \times \cdots \times S^{d_k - 1}$ iff

$$\Delta(A) = 0$$

where Δ is the hyperdeterminant in $\mathbb{R}^{d_1 \times \cdots \times d_k}$.

In other words, $\Delta(A) = 0$ iff 0 is a singular value of A.

Homogeneous system of multilinear equations

The hyperdeterminant of $A = [\![a_{ijk}]\!] \in \mathbb{R}^{2\times 2\times 2}$ is

$$\begin{aligned}\Delta(A) := (&a_{000}^2 a_{111}^2 + a_{001}^2 a_{110}^2 + a_{010}^2 a_{101}^2 + a_{011}^2 a_{100}^2) \\ &- 2(a_{000}a_{001}a_{110}a_{111} + a_{000}a_{010}a_{101}a_{111} + a_{000}a_{011}a_{100}a_{111} \\ &+ a_{001}a_{010}a_{101}a_{110} + a_{001}a_{011}a_{110}a_{100} + a_{010}a_{011}a_{101}a_{100}) \\ &+ 4(a_{000}a_{011}a_{101}a_{110} + a_{001}a_{010}a_{100}a_{111}).\end{aligned}$$

Result that parallels matrix case: the system of bilinear equations

$$\begin{aligned} a_{000}x_0y_0 + a_{010}x_0y_1 + a_{100}x_1y_0 + a_{110}x_1y_1 &= 0, \\ a_{001}x_0y_0 + a_{011}x_0y_1 + a_{101}x_1y_0 + a_{111}x_1y_1 &= 0, \\ a_{000}x_0z_0 + a_{001}x_0z_1 + a_{100}x_1z_0 + a_{101}x_1z_1 &= 0, \\ a_{010}x_0z_0 + a_{011}x_0z_1 + a_{110}x_1z_0 + a_{111}x_1z_1 &= 0, \\ a_{000}y_0z_0 + a_{001}y_0z_1 + a_{010}y_1z_0 + a_{011}y_1z_1 &= 0, \\ a_{100}y_0z_0 + a_{101}y_0z_1 + a_{110}y_1z_0 + a_{111}y_1z_1 &= 0. \end{aligned}$$

has a non-trivial solution iff $\Delta(A) = 0$.

Multilinear forms

$A = [\![a_{j_1\cdots j_k}]\!] \in \mathbb{R}^{n\times\cdots\times n}$ symmetric tensor; multilinear form defined by A is homogeneous polynomial

$$
\begin{aligned}
g_A &: \mathbb{R}^n \to \mathbb{R}, \\
&\mathbf{x} \mapsto A(\mathbf{x},\ldots,\mathbf{x}) = \sum\nolimits_{j_1=1}^{n} \cdots \sum\nolimits_{j_k=1}^{n} a_{j_1\cdots j_k} x_{j_1} \cdots x_{j_k}.
\end{aligned}
$$

Gradient of g_A,

$$\nabla g_A(\mathbf{x}) = \left(\frac{\partial g_A}{\partial x_1},\ldots,\frac{\partial g_A}{\partial x_n}\right) = kA(I_n,\mathbf{x},\ldots,\mathbf{x})$$

where $\mathbf{x} = (x_1,\ldots,x_n)^\top$ occurs $k-1$ times in the argument. This is a multilinear generalization of

$$\frac{d}{dx} ax^k = kax^{k-1}.$$

Note that for a symmetric tensor,

$$A(I_n,\mathbf{u},\mathbf{u},\ldots,\mathbf{u}) = A(\mathbf{u},I_n,\mathbf{u},\ldots,\mathbf{u}) = \cdots = A(\mathbf{u},\mathbf{u},\ldots,\mathbf{u},I_n).$$

ℓ^2-eigenvalues of a symmetric tensor

In this case, the Lagrangian is

$$L(\mathbf{x}, \lambda) = A(\mathbf{x}, \ldots, \mathbf{x}) - \lambda(\|\mathbf{x}\|_2^k - 1)$$

Then $\nabla_{\mathbf{x}} L = \mathbf{0}$ yields

$$kA(I_n, \mathbf{x}, \ldots, \mathbf{x}) = k\lambda\|\mathbf{x}\|_2^{k-2}\mathbf{x},$$

or, equivalently

$$A\left(I_n, \frac{\mathbf{x}}{\|\mathbf{x}\|_2}, \ldots, \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\right) = \lambda\frac{\mathbf{x}}{\|\mathbf{x}\|_2}.$$

$\nabla_\lambda L = 0$ yields $\|\mathbf{x}\|_2 = 1$. Normalize to get $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_2 \in S^{n-1}$, giving

$$A(I_n, \mathbf{u}, \mathbf{u}, \ldots, \mathbf{u}) = \lambda\mathbf{u}.$$

$\mathbf{u} \in S^{n-1}$ will be called an ℓ^2-*eigenvector* and λ will be called an ℓ^2-*eigenvalue* of A.

ℓ^2-eigenvalues of a nonsymmetric tensor

How about eigenvalues and eigenvectors for $A \in \mathbb{R}^{n\times\cdots\times n}$ that may not be symmetric? Even in the order-2 case, the critical values/points of the Rayleigh quotient no longer gives the eigenpairs.

However, as in the order-2 case, eigenvalues and eigenvectors can still be defined via

$$A(I_n, \mathbf{v}^1, \mathbf{v}^1, \dots, \mathbf{v}^1) = \mu\mathbf{v}^1.$$

Except that now, the equations

$$\begin{aligned} A(I_n, \mathbf{v}^1, \mathbf{v}^1, \dots, \mathbf{v}^1) &= \mu_1\mathbf{v}^1, \\ A(\mathbf{v}^2, I_n, \mathbf{v}^2, \dots, \mathbf{v}^2) &= \mu_2\mathbf{v}^2, \\ &\vdots \\ A(\mathbf{v}^k, \mathbf{v}^k, \dots, \mathbf{v}^k, I_n) &= \mu_k\mathbf{v}^k, \end{aligned}$$

are distinct.

We will call $\mathbf{v}^i \in \mathbb{R}^n$ an mode-i *eigenvector* and μ_i an mode-i *eigenvalue*. This is just the order-k generalization of left- and right-eigenvectors for nonsymmetric matrices.

Note that the unit-norm constraint on ℓ^2-eigenvectors cannot be omitted for order 3 or higher because of the lack of scale invariance.

Characteristic polynomial

Let $A \in \mathbb{R}^{n\times n}$. One way to get the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$ is as follows.

$$\begin{cases} \sum_{j=1}^{n} a_{ij}x_j = \lambda x_i, \quad i = 1,\dots,n, \\ x_1^2 + \cdots + x_n^2 = 1. \end{cases}$$

System of $n+1$ polynomial equations in $n+1$ variables, $x_1,\dots,x_n,\lambda$.

Use Elimination Theory to eliminate all variables $x_1,\dots,x_n$, leaving a one-variable polynomial in λ — a simple case of the multivariate resultant.

The $\det(A - \lambda I)$ definition does not generalize to higher order but the elimination theoretic approach does.

Multilinear characteristic polynomial

Let $A \in \mathbb{R}^{n\times\cdots\times n}$, not necessarily symmetric. Use mode-1 for illustration.

$$A(I_n, \mathbf{x}^1, \mathbf{x}^1, \dots, \mathbf{x}^1) = \mu\mathbf{x}^1.$$

and the unit-norm condition gives a system of $n+1$ equations in $n+1$ variables $x_1, \dots, x_n, \lambda$:

$$\begin{cases} \sum_{j_2=1}^{n} \cdots \sum_{j_k=1}^{n} a_{ij_2\cdots j_k} x_{j_2} \cdots x_{j_k} = \lambda x_i, \quad i = 1, \dots, n, \\ x_1^2 + \cdots + x_n^2 = 1. \end{cases}$$

Apply elimination theory to obtain the *multipolynomial resultant* or *multivariate resultant* — a one-variable polynomial $p_A(\lambda)$. Efficient algorithms exist:

D. Manocha and J.F. Canny, "Multipolynomial resultant algorithms," *J. Symbolic Comput.*, **15** (1993), no. 2, pp. 99–122.

If the $a_{ij_2\cdots j_k}$'s assume numerical values, $p_A(\lambda)$ may be obtained by applying Gröbner bases techniques to system of equations directly.

Roots of $p_A(\lambda)$ are precisely the eigenvalues of the tensor A. Adopt matrix terminology and call it *characteristic polynomial* of A, which has an expression

$$p_A(\lambda) = \begin{cases} \det M(\lambda)/\det L & \text{if } \det L \neq 0, \\ \det m(\lambda) & \text{if } \det L = 0. \end{cases}$$

$M(\lambda)$ is a square matrix whose entries are polynomials in λ (for order-2, $M(\lambda) = A - \lambda I$). In the $\det(L) = 0$ case, $\det m(\lambda)$ denotes the largest non-vanishing minor of $M(\lambda)$.

Polynomial matrix eigenvalue problem

The matrix $M(\lambda)$ (or $m(\lambda)$ in the $\det(L) = 0$ case) allows numerical linear algebra to be used in the computations of eigenvectors as

$$\begin{cases} \sum_{j_2=1}^{n} \cdots \sum_{j_k=1}^{n} a_{ij_2\cdots j_k} x_{j_2} \cdots x_{j_k} = \lambda x_i, \quad i = 1, \ldots, n, \\ x_1^2 + \cdots + x_n^2 = 1. \end{cases}$$

may be reexpressed in the form

$$M(\lambda)(1, x_1, \ldots x_n, \ldots, x_n^n)^\top = (0, \ldots, 0)^\top.$$

So if $(\mathbf{x}, \lambda)$ is an eigenpair of A. Then $M(\lambda)$ must have a non-trivial kernel.

Observe that $M(\lambda)$ may be expressed as

$$M(\lambda) = M_0 + M_1\lambda + \cdots + M_d\lambda^d$$

where M_i's are matrices with numerical entries.

Perron-Frobenius theorem for nonnegative tensors

An order-k cubical tensor $A \in \mathsf{T}^k(\mathbb{R}^n)$ is *reducible* if there exist a permutation $\sigma \in \mathfrak{S}_n$ such that the permuted tensor

$$[\![b_{i_1 \cdots i_k}]\!] = [\![a_{\sigma(j_1) \cdots \sigma(j_k)}]\!]$$

has the property that for some $m \in \{1, \dots, n-1\}$, $b_{i_1 \cdots i_k} = 0$ for all $i_1 \in \{1, \dots, n-m\}$ and all $i_2, \dots, i_k \in \{1, \dots, m\}$. We say that A is *irreducible* if it is not reducible. In particular, if $A > 0$, then it is irreducible.

Theorem (L). Let $0 \le A = [\![a_{j_1 \cdots j_k}]\!] \in \mathsf{T}^k(\mathbb{R}^n)$ be irreducible. Then A has a positive real l^k-eigenvalue μ with an l^k-eigenvector $\mathbf{x}$ that may be chosen to have all entries non-negative. Furthermore, μ is simple, ie. $\mathbf{x}$ is unique modulo scalar multiplication.

Hypergraphs

For notational simplicity, the following is stated for a 3-hypergraph but it generalizes to k-hypergraphs for any k.

$G = (V, E)$ be a 3-hypergraph. V is the finite set of vertices and E is the subset of hyperedges, ie. 3-element subsets of V. We write the elements of E as $[x, y, z]$ $(x, y, z \in V)$.

G is undirected, so $[x, y, z] = [y, z, x] = \cdots = [z, y, x]$. A hyperedge is said to degenerate if it is of the form $[x, x, y]$ or $[x, x, x]$ (hyperloop at x). We do not exclude degenerate hyperedges.

G is m-regular if every $v \in V$ is adjacent to exactly m hyperedges. We can 'regularize' a non-regular hypergraph by adding hyperloops.

Adjacency tensor of a hypergraph

Define the order-3 adjacency tensor A by

$$A_{xyz} = \begin{cases} 1 & \text{if } [x, y, z] \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Note that A is $|V|$-by-$|V|$-by-$|V|$ nonnegative symmetric tensor.

Consider cubic form $A(f, f, f) = \sum_{x,y,z} A_{xyz} f(x) f(y) f(z)$ (note that f is a vector of dimension $|V|$).

Call critical values and critical points of $A(f, f, f)$ constrained to the set $\sum_x f(x)^3 = 1$ (like the ℓ^3-norm except we do not take absolute value) the ℓ^3-eigenvalues and ℓ^3-eigenvectors of A respectively.

Very basic spectral hypergraph theory I

As in the case of spectral graph theory, combinatorial/topological properties of a k-hypergraph may be deduced from ℓ^k-eigenvalues of its adjacency tensor (henceforth, in the context of a k-hypergraph, an eigenvalue will always mean an ℓ^k-eigenvalue).

Straightforward generalization of a basic result in spectral graph theory:

Theorem (Drineas, L). Let G be an m-regular 3-hypergraph and A be its adjacency tensor. Then

(a) m is an eigenvalue of A;

(b) if μ is an eigenvalue of A, then $|\mu| \leq m$;

(c) μ has multiplicity 1 if and only if G is connected.

Very basic spectral hypergraph theory II

A hypergraph $G = (V, E)$ is said to be k-partite or k-colorable if there exists a partition of the vertices $V = V_1 \cup \cdots \cup V_k$ such that for any k vertices $u, v, \ldots, z$ with $A_{uv\cdots z} \neq 0$, $u, v, \ldots, z$ must each lie in a distinct V_i ($i = 1, \ldots, k$).

Lemma (Drineas, L). Let G be a connected m-regular k-partite k-hypergraph on n vertices. Then

(a) If k is odd, then every eigenvalue of G occurs with multiplicity a multiple of k.

(b) If k is even, then the spectrum of G is symmetric (ie. if μ is an eigenvalue, then so is $-\mu$). Furthermore, every eigenvalue of G occurs with multiplicity a multiple of $k/2$. If μ is an eigenvalue of G, then μ and $-\mu$ occurs with the same multiplicity.

Liqun Qi's work

L. Qi, "Eigenvalues of a real supersymmetric tensor," *J. Symbolic Comput.*, **40** (2005), no. 6, pp. 1302–1324.

(a) Gershgorin circle theorem for ℓ^k-eigenvalues;

(b) characterizing positive definiteness of even-ordered forms (e.g. quartic forms) using ℓ^k-eigenvalues;

(c) generalization of trace-sum equality for ℓ^2-eigenvalues;

(d) six open conjectures.

See also work by Qi's postdocs and students: Yiju Wang, Guyan Ni, Fei Wang.

References

`http://www-sccm.stanford.edu/nf-publications-tech.html`

[**dSL**] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, to appear.

[**CGLM2**] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," *SCCM Tech. Rep.*, **06-02**, 2006.

[**CGLM1**] P. Comon, B. Mourrain, L.-H. Lim, and G.H. Golub, "Genericity and rank deficiency of high order symmetric tensors," *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, **31** (2006), no. 3, pp. 125–128.

[**L**] L.-H. Lim, "Singular values and eigenvalues of tensors: a variational approach," *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, **1** (2005), pp. 129–132.