Multilinear Spectral Theory (and its applications)

Lek-Heng Lim

Workshop on Tensor Decompositions and Applications
CIRM, Luminy, France
August 29–September 2, 2005

Thanks: Pierre Comon, Gene Golub, NSF DMS 01-01364

Acknowledgement

Pierre Comon
Laboratoire I3S
Université de Nice Sophia Antipolis

Lieven de Lathauwer
Equipes Traitement des Images et du Signal
Ecole Nationale Supérieure d'Electronique et de ses Applications

Multilinear Matrix Multiplication

Multilinear map $g : V_1 \times \cdots \times V_k \to \mathbb{R}$, $g(\mathbf{y}_1, \ldots, \mathbf{y}_k)$.

Linear maps $f_\alpha : U_\alpha \to V_\alpha$, $\mathbf{y}_\alpha = f_\alpha(\mathbf{x}_i)$, $\alpha = 1, \ldots, k$.

Compose g by $f_1, \ldots, f_k$ to get $h : U_1 \times \cdots \times U_k \to \mathbb{R}$,

$$h(\mathbf{x}_1, \ldots, \mathbf{x}_k) = g(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_k)).$$

$A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ represents g;

$M_\alpha = [m^\alpha_{j_1 i_1}] \in \mathbb{R}^{d_\alpha \times s_\alpha}$ represents f_α, $\alpha = 1, \ldots, k$;

Then h represented by

$$A(M_1, \ldots, M_k) = [\![c_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{s_1 \times \cdots \times s_k}$$
$$c_{i_1 \cdots i_k} := \sum\nolimits_{j_1=1}^{d_1} \cdots \sum\nolimits_{j_k=1}^{d_k} a_{j_1 \cdots j_k} m^1_{j_1 i_1} \cdots m^k_{j_k i_k}.$$

Call the above *covariant multilinear matrix multiplication*.

Contravariant version: compose multilinear map

$$g : V_1^* \times \cdots \times V_k^* \to \mathbb{R}$$

with the adjoint of linear maps $f_\alpha : V_\alpha \to U_\alpha$, $\alpha = 1, \ldots, k$,

$$\begin{aligned}(L_1, \ldots, L_k)A = \llbracket b_{i_1 \cdots i_k} \rrbracket \in \mathbb{R}^{r_1 \times \cdots \times r_k}, \\ b_{i_1 \cdots i_k} := \textstyle\sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} \ell^1_{i_1 j_1} \cdots \ell^k_{i_k j_k} a_{j_1 \cdots j_k}.\end{aligned}$$

Symmetric Tensors

$A = [\![a_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. For a permutation $\sigma \in \Sigma_k$, *σ-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 Σ_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 Σ_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 $\sigma \in \Sigma_k$, 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

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

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

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

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

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

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

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

Variational Characterization of Singular Pairs

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

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

Associated Lagrangian is

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

The first order condition yields

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

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

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

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.

Singular Values and Singular Vectors of a Tensor

Take a variational approach as in the case of matrices. Lagrangian is

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

where $\sigma \in \mathbb{R}$ is the Lagrange multiplier. Then

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

yields

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

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

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

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

P. Comon, "Tensor decompostions: state of the art and applications," in *Mathematics in signal processing*, **V** (Coventry, UK, 2000), pp. 1–24, *Inst. Math. Appl. Conf. Ser.*, **71**, Oxford University Press, Oxford, UK, 2002.

L. de Lathauwer, B. de Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, \dots, R_N)$ approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, **21** (4), 2000, pp. 1324–1342.

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 $f : V_1 \times \cdots \times V_k \to V_0$ from a product of norm spaces $(V_1, \|\cdot\|_1), \ldots, (V_k, \|\cdot\|_k)$ to a norm space $(V_0, \|\cdot\|_0)$ is defined as

$$\sup \frac{\|f(\mathbf{x}^1, \ldots, \mathbf{x}^k)\|_0}{\|\mathbf{x}^1\|_1 \cdots \|\mathbf{x}^k\|_k}$$

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

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

where $\|\cdot\|$ in the denominator denotes the usual Euclidean 2-norm. 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,

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

Relation with Hyperdeterminant

Assume

$$d_i - 1 \le \sum_{j \ne i} (d_j - 1)$$

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.

Multilinear Homogeneous Polynomial

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

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

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,

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

Eigenvalues and Eigenvectors of a Symmetric Tensor

In this case, the Lagrangian is

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

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

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

or, equivalently

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

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

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

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

Eigenvalues and Eigenvectors of a 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

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

Except that now, the equations

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

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 unsymmetric matrices.

Note that the unit-norm constraint on the 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.

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

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

Use Elimination Theory to eliminate all variables $x_1, \ldots, x_n$, 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.

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

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

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

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.

This reduces the multilinear eigenvalue problem to a *polynomial eigenvalue problem*. Efficient algorithms for solving such problems will be discussed in the next talk.

Note that the preceding discussions also apply in the context of singular pairs, where we solve a system of $d_1 + \cdots + d_k + 1$ equations in $d_1 + \cdots + d_k + 1$ variables.

Applications

Singular values/vectors — Nash equilibria for n-person games.

Symmetric eigenvalues/vectors — spectral hypergraph theory.

Unsymmetric eigenvalues/vectors — multilinear Perron-Frobenius theory.

R.D. McKelvey and A. McLennan, "The maximal number of regular totally mixed Nash equilibria," *J. Econom. Theory*, **72** (1997), no. 2, pp. 411–425.

P. Drineas and L.-H. Lim, "A multilinear spectral theory of hypergraphs and expander hypergraphs," work in progress.

L.-H. Lim, "Multilinear PageRank: measuring higher order connectivity in linked objects," poster, *The Internet: Today & Tomorrow*, 2005 School of Engineering Summer Research Forum, July 28, 2005, Stanford University, Stanford, CA, 2005.