Numerical multilinear algebra

From matrices to tensors

Lek-Heng Lim

University of California, Berkeley

March 1, 2008

Lesson 1: work over $\mathbb{R}$ or $\mathbb{C}$

Facts that every numerical linear algebraist takes for granted:

1. Every matrix has a singular value decomposition.
2. Normal equation $A^\top A\mathbf{x} = A^\top\mathbf{b}$ always consistent for any A and $\mathbf{b}$.
3. $\operatorname{rank}(A^\top A) = \operatorname{rank}(A)$.
4. If $A \in \mathbb{R}^{m\times n}$, then
$$\mathbb{R}^n = \operatorname{nullsp}(A) \oplus \operatorname{range}(A^\top),$$
$$\mathbb{R}^m = \operatorname{nullsp}(A^\top) \oplus \operatorname{range}(A).$$
5. $\langle\mathbf{x},\mathbf{y}\rangle = \sum_{i=1}^n x_i y_i$, $\langle A, B\rangle = \operatorname{tr}(A^\top B)$ define inner products.
6. $\|\mathbf{x}\|_p^p = \sum_{i=1}^n |x_i|^p$, $\|A\|_F^2 = \sum_{i,j=1}^{m,n} |a_{ij}|^2$ define norms.

All these statements are **false** in general over arbitrary fields.

Lesson 2: work with matrices

- Isomorphic doesn't mean identical.
- A matrix doesn't always come from an operator.
- Can be a list of column or row vectors:
  - gene-by-microarray matrix,
  - movies-by-viewers matrix,
  - list of codewords.
- Can be a convenient way to represent graph structures:
  - adjacency matrix,
  - graph Laplacian,
  - webpage-by-webpage matrix.
- Useful to regard them as matrices and apply matrix operations:
  - A gene-by-microarray matrix, $A = U\Sigma V^\top$ gives cellular states (eigengenes), biological phenotype (eigenarrays) [Alter, Golub; 2004],
  - A adjacency matrix, A^k counts number of paths of length $\leq k$ from node i to node j.

Lesson 3: look at other areas

Linear algebra is probably the topic least likely to yield interesting problems and tools for numerical linear algebra.

- **Algebraic geometry:** varieties of Segre (rank-1 matrices) and Veronese (rank-1 symmetric matrices)
- **Classical mechanics:** high frequency oscillations of membranes (pseudospectrum)
- **Lie groups:** Bruhat (LU), Cartan (SVD), Iwasawa (QR) decompositions
- **Machine learning:** collaborative filtering (maximum margin matrix factorization)
- **Psychology:** Eckart-Young theorem (optimal low rank approximation)
- **Representation theory:** cyclic representations (Krylov subspaces)
- **Statistics:** errors-in-variables model (total least squares)

Hypermatrices

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 $[\![b_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$, $\lambda \in \mathbb{R}$,

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

2. **Multilinear matrix multiplication:** for matrices $L = [\lambda_{\alpha i}] \in \mathbb{R}^{p \times l}$, $M = [\mu_{\beta j}] \in \mathbb{R}^{q \times m}$, $N = [\nu_{\gamma k}] \in \mathbb{R}^{r \times n}$,

$$(L, M, N) \cdot A := [\![c_{\alpha\beta\gamma}]\!] \in \mathbb{R}^{p \times q \times r}$$

where

$$c_{\alpha\beta\gamma} := \sum\nolimits_{i,j,k=1}^{l,m,n} \lambda_{\alpha i} \mu_{\beta j} \nu_{\gamma k} a_{ijk}.$$

Hypermatrices

- Think of A as 3-dimensional array of numbers. $(L, M, N) \cdot A$ as multiplication on '3 sides' by matrices L, M, N.
- Generalizes to arbitrary order k. If $k = 2$, ie. matrix, then $(M, N) \cdot A = MAN^\top$.
- Covariant version:

$$A \cdot (L^\top, M^\top, N^\top) := (L, M, N) \cdot A.$$

- Gives convenient notations for multilinear functionals and multilinear operators. For $\mathbf{x} \in \mathbb{R}^l, \mathbf{y} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^n$,

$$A(\mathbf{x}, \mathbf{y}, \mathbf{z}) := A \cdot (\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum\nolimits_{i,j,k=1}^{l,m,n} a_{ijk} x_i y_j z_k,$$

$$A(I, \mathbf{y}, \mathbf{z}) := A \cdot (I, \mathbf{y}, \mathbf{z}) = \sum\nolimits_{j,k=1}^{m,n} a_{ijk} y_j z_k.$$

Hypermatrices and tensors

Up to a choice of bases,

- a matrix $A \in \mathbb{R}^{m \times n}$ may represent
  - an order-2 tensor in $V_1 \otimes V_2$,
  - a bilinear functional $V_1 \times V_2 \to \mathbb{R}$,
  - a linear operator $V_2 \to V_1$,

  where $\dim(V_1) = m$ and $\dim(V_2) = n$;
- a hypermatrix $A \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_k}$ may represent
  - an order-k tensor in $V_1 \otimes V_2 \otimes \cdots \otimes V_k$,
  - a multilinear functional $V_1 \times V_2 \times \cdots \times V_k \to \mathbb{R}$,
  - a multilinear operator $V_2 \times \cdots \times V_k \to V_1$,

  where $\dim(V_i) = d_i$, $i = 1, \dots, k$.

Numerical multilinear algebra

Bold claim: every topic discussed in Golub-Van Loan has a multilinear generalization.

- Numerical tensor rank (GV Chapter 2)
- Conditioning of multilinear systems (GV Chapter 3)
- Unsymmetric eigenvalue problem for hypermatrices (GV Chapter 7)
- Symmetric eigenvalue problem for hypermatrices (GV Chapter 8)
- Regularization of tensor approximation problems (GV Chapter 12)

DARPA mathematical challenge eight

One of the twenty three mathematical challenges announced at DARPA Tech 2007.

Problem

Beyond convex optimization: *can linear algebra be replaced by algebraic geometry in a systematic way?*

- **Algebraic geometry in a slogan:** polynomials are to algebraic geometry what matrices are to linear algebra.
- Polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ of degree d can be expressed as

$$f(\mathbf{x}) = a_0 + \mathbf{a}_1^\top \mathbf{x} + \mathbf{x}^\top A_2 \mathbf{x} + A_3(\mathbf{x}, \mathbf{x}, \mathbf{x}) + \cdots + A_d(\mathbf{x}, \dots, \mathbf{x}).$$

  $a_0 \in \mathbb{R}, \mathbf{a}_1 \in \mathbb{R}^n, A_2 \in \mathbb{R}^{n \times n}, A_3 \in \mathbb{R}^{n \times n \times n}, \dots, A_d \in \mathbb{R}^{n \times \cdots \times n}.$
- Numerical linear algebra: $d = 2$.
- Numerical multilinear algebra: $d > 2$.

Tensor ranks

- **Matrix rank.** $A \in \mathbb{R}^{m\times n}$.

$$\begin{aligned}
\operatorname{rank}(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet 1}, \dots, A_{\bullet n}\}) && \text{(column rank)}\\
&= \dim(\operatorname{span}_{\mathbb{R}}\{A_{1\bullet}, \dots, A_{m\bullet}\}) && \text{(row rank)}\\
&= \min\{r \mid A = \textstyle\sum_{i=1}^{r} \mathbf{u}_i \mathbf{v}_i^{\mathsf{T}}\} && \text{(outer product rank).}
\end{aligned}$$

- **Multilinear rank.** $A \in \mathbb{R}^{l\times m\times n}$. $\operatorname{rank}_{\boxplus}(A) = (r_1(A), r_2(A), r_3(A))$,

$$\begin{aligned}
r_1(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{1\bullet\bullet}, \dots, A_{l\bullet\bullet}\})\\
r_2(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet 1\bullet}, \dots, A_{\bullet m\bullet}\})\\
r_3(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet\bullet 1}, \dots, A_{\bullet\bullet n}\})
\end{aligned}$$

- **Outer product rank.** $A \in \mathbb{R}^{l\times m\times n}$.

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

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

- Both notions of tensor ranks first appeared in [Hitchcock; 1927].

Recall: conditioning for linear systems

- Let $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Suppose we want to solve system of linear equations $A\mathbf{x} = \mathbf{b}$.
- $\mathcal{M} = \{A \in \mathbb{R}^{n\times n} \mid \det(A) = 0\}$ is the manifold of ill-posed problems.
- $A \in \mathcal{M}$ iff $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions.
- Note that $\det(A)$ is a poor measure of conditioning.
- Conditioning is the inverse distance to ill-posedness [Demmel; 1987] (also Dedieu, Shub, Smale), ie.

$$\frac{1}{\|A^{-1}\|_2}.$$

- Normalizing by $\|A\|_2$ yields

$$\frac{1}{\|A\|_2\|A^{-1}\|_2} = \frac{1}{\kappa_2(A)}.$$

- Note that

$$\|A^{-1}\|_2^{-1} = \sigma_n = \min_{\mathbf{x}_i,\mathbf{y}_i} \|A - \mathbf{x}_1 \otimes \mathbf{y}_1 - \cdots - \mathbf{x}_{n-1} \otimes \mathbf{y}_{n-1}\|_2.$$

Recall: conditioning for linear systems

Important for error analysis [Wilkinson, 1961]. Let $A = U\Sigma V^\top$ and

$$\begin{aligned} S_{\text{forward}}(\varepsilon) &= \{\mathbf{x}' \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}, \quad \|\mathbf{x}' - \mathbf{x}\|_2 \le \varepsilon\} \\ &= \{\mathbf{x}' \in \mathbb{R}^n \mid \textstyle\sum_{i=1}^n |x_i' - x_i|^2 \le \varepsilon^2\}, \\ S_{\text{backward}}(\varepsilon) &= \{\mathbf{x}' \in \mathbb{R}^n \mid A\mathbf{x}' = \mathbf{b}', \quad \|\mathbf{b}' - \mathbf{b}\|_2 \le \varepsilon\} \\ &= \{\mathbf{x}' \in \mathbb{R}^n \mid \mathbf{x}' - \mathbf{x} = V(\mathbf{y}' - \mathbf{y}), \quad \textstyle\sum_{i=1}^n \sigma_i^2 |y_i' - y_i|^2 \le \varepsilon^2\}. \end{aligned}$$

Then

$$S_{\text{backward}}(\varepsilon) \subseteq S_{\text{forward}}(\sigma_n^{-1}\varepsilon), \quad S_{\text{forward}}(\varepsilon) \subseteq S_{\text{backward}}(\sigma_1\varepsilon).$$

Determined by $\sigma_1 = \|A\|_2$ and $\sigma_n^{-1} = \|A^{-1}\|_2$.

What about multilinear systems?

Look at the simplest case. Take $A = [\![a_{ijk}]\!] \in \mathbb{R}^{2\times 2\times 2}$ and $\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^2$.

$$
\begin{aligned}
a_{000}x_0y_0 + a_{010}x_0y_1 + a_{100}x_1y_0 + a_{110}x_1y_1 &= b_{00},\\
a_{001}x_0y_0 + a_{011}x_0y_1 + a_{101}x_1y_0 + a_{111}x_1y_1 &= b_{01},\\
a_{000}x_0z_0 + a_{001}x_0z_1 + a_{100}x_1z_0 + a_{101}x_1z_1 &= b_{10},\\
a_{010}x_0z_0 + a_{011}x_0z_1 + a_{110}x_1z_0 + a_{111}x_1z_1 &= b_{11},\\
a_{000}y_0z_0 + a_{001}y_0z_1 + a_{010}y_1z_0 + a_{011}y_1z_1 &= b_{20},\\
a_{100}y_0z_0 + a_{101}y_0z_1 + a_{110}y_1z_0 + a_{111}y_1z_1 &= b_{21}.
\end{aligned}
$$

- When does this have a solution?
- What is the corresponding manifold of ill-posed problems?
- When does the homogeneous system, ie. $\mathbf{b}_0 = \mathbf{b}_1 = \mathbf{b}_2 = \mathbf{0}$, have a non-trivial solution, ie. $\mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}, \mathbf{z} \neq \mathbf{0}$?

Hyperdeterminant

- Work in $\mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)}$ for the time being ($d_i \geq 1$). Consider

$$\mathcal{M} := \{A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \nabla A(\mathbf{x}_1, \dots, \mathbf{x}_k) = \mathbf{0} \quad \text{for non-zero } \mathbf{x}_1, \dots, \mathbf{x}_k\}.$$

Theorem (Gelfand, Kapranov, Zelevinsky)

$\mathcal{M}$ is a hypersurface iff for all $j = 1, \dots, k$,

$$d_j \leq \sum\nolimits_{i \neq j} d_i.$$

- The **hyperdeterminant** $\operatorname{Det}(A)$ is the equation of the hypersurface, ie. a multivariate polynomial in the entries of A such that

$$\mathcal{M} = \{A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \operatorname{Det}(A) = 0\}.$$

- $\operatorname{Det}(A)$ may be chosen to have integer coefficients.
- For $\mathbb{C}^{m\times n}$, condition becomes $m \leq n$ and $n \leq m$, ie. square matrices.

$2 \times 2 \times 2$ hyperdeterminant

Hyperdeterminant of $A = [\![a_{ijk}]\!] \in \mathbb{R}^{2\times2\times2}$ [Cayley; 1845] is

$$\operatorname{Det}(A) = \frac{1}{4}\left[\det\left(\begin{bmatrix} a_{000} & a_{010} \\ a_{001} & a_{011} \end{bmatrix} + \begin{bmatrix} a_{100} & a_{110} \\ a_{101} & a_{111} \end{bmatrix}\right) - \det\left(\begin{bmatrix} a_{000} & a_{010} \\ a_{001} & a_{011} \end{bmatrix} - \begin{bmatrix} a_{100} & a_{110} \\ a_{101} & a_{111} \end{bmatrix}\right)\right]^2 - 4\det\begin{bmatrix} a_{000} & a_{010} \\ a_{001} & a_{011} \end{bmatrix}\det\begin{bmatrix} a_{100} & a_{110} \\ a_{101} & a_{111} \end{bmatrix}.$$

A result that parallels the matrix case is the following: 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 $\operatorname{Det}(A) = 0$.

$2 \times 2 \times 3$ hyperdeterminant

Hyperdeterminant of $A = [\![a_{ijk}]\!] \in \mathbb{R}^{2\times 2\times 3}$ is

$$\operatorname{Det}(A) = \det \begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{100} & a_{101} & a_{102} \\ a_{010} & a_{011} & a_{012} \end{bmatrix} \det \begin{bmatrix} a_{100} & a_{101} & a_{102} \\ a_{010} & a_{011} & a_{012} \\ a_{110} & a_{111} & a_{112} \end{bmatrix} - \det \begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{100} & a_{101} & a_{102} \\ a_{110} & a_{111} & a_{112} \end{bmatrix} \det \begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{010} & a_{011} & a_{012} \\ a_{110} & a_{111} & a_{112} \end{bmatrix}$$

Again, the following is true:

$$\begin{aligned}
a_{000}x_0y_0 + a_{010}x_0y_1 + a_{100}x_1y_0 + a_{110}x_1y_1 &= 0,\\
a_{001}x_0y_0 + a_{011}x_0y_1 + a_{101}x_1y_0 + a_{111}x_1y_1 &= 0,\\
a_{002}x_0y_0 + a_{012}x_0y_1 + a_{102}x_1y_0 + a_{112}x_1y_1 &= 0,\\
a_{000}x_0z_0 + a_{001}x_0z_1 + a_{002}x_0z_2 + a_{100}x_1z_0 + a_{101}x_1z_1 + a_{102}x_1z_2 &= 0,\\
a_{010}x_0z_0 + a_{011}x_0z_1 + a_{012}x_0z_2 + a_{110}x_1z_0 + a_{111}x_1z_1 + a_{112}x_1z_2 &= 0,\\
a_{000}y_0z_0 + a_{001}y_0z_1 + a_{002}y_0z_2 + a_{010}y_1z_0 + a_{011}y_1z_1 + a_{012}y_1z_2 &= 0,\\
a_{100}y_0z_0 + a_{101}y_0z_1 + a_{102}y_0z_2 + a_{110}y_1z_0 + a_{111}y_1z_1 + a_{112}y_1z_2 &= 0,
\end{aligned}$$

has a non-trivial solution iff $\operatorname{Det}(A) = 0$.

Condition number of a multilinear system

- Like the matrix determinant, the value of the hyperdeterminant is a poor measure of conditioning. Need to compute distance to $\mathcal{M}$.

Theorem (de Silva, L)

Let $A \in \mathbb{R}^{2\times 2\times 2}$. $\operatorname{Det}(A) = 0$ *iff*

$$A = \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{y} + \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{x} + \mathbf{y} \otimes \mathbf{x} \otimes \mathbf{x}$$

for some $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^2$, $i = 1, 2, 3$.

- Conditioning of the problem can be obtained from

$$\min_{\mathbf{x},\mathbf{y}} \|A - \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{y} - \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{x} - \mathbf{y} \otimes \mathbf{x} \otimes \mathbf{x}\|.$$

- $\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{y} + \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{x} + \mathbf{y} \otimes \mathbf{x} \otimes \mathbf{x}$ has outer product rank 3 generically (in fact, iff $\mathbf{x}, \mathbf{y}$ are linearly independent).
- Surprising: the manifold of ill-posed problem has full rank almost everywhere!

Generalization

Use Gaussian elimination to get 'LDU-decomposition' for hypermatrices.

Lemma (de Silva, L)

Let $A \in \mathbb{R}^{l \times m \times n}$. If $\operatorname{rank}_{\boxplus}(A) \leq (r, r, r)$ or if $\operatorname{rank}_{\otimes}(A) \leq r$, then there exists unit lower-triangular matrices L_1, L_2, L_3 such that

$$A = (L_1, L_2, L_3) \cdot C$$

where C is everywhere zero except for an $r \times r \times r$ block.

Corollary

If $r = 2$ above, then $A \in \mathbb{R}^{l \times m \times n}$ is ill-posed iff $\operatorname{Det}(C) = 0$ iff

$$A = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{y}_3 + \mathbf{x}_1 \otimes \mathbf{y}_2 \otimes \mathbf{x}_3 + \mathbf{y}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$$

for some $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^l$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^m$, $\mathbf{x}_3, \mathbf{y}_3 \in \mathbb{R}^n$.

Symmetric hypermatrices

- An order-k cubical hypermatrix $[\![a_{i_1\cdots i_k}]\!] \in \mathbb{R}^{n\times\cdots\times n}$ is **symmetric** if

$$a_{i_{\sigma(1)}\cdots i_{\sigma(k)}} = a_{i_1\cdots i_k}, \qquad i_1,\dots,i_k \in \{1,\dots,n\},$$

for all permutations $\sigma \in \mathfrak{S}_k$. $\mathsf{S}^k(\mathbb{R}^n)$ is the set of all order-k symmetric hypermatrices.

Example

Higher order derivatives of multivariate functions.

Example

Moments of a random vector $\mathbf{x} = (X_1,\dots,X_n)$:

$$m_k(\mathbf{x}) = \bigl[E(x_{i_1}x_{i_2}\cdots x_{i_k})\bigr]_{i_1,\dots,i_k=1}^{n} = \left[\int\cdots\int x_{i_1}x_{i_2}\cdots x_{i_k}\; d\mu(x_{i_1})\cdots d\mu(x_{i_k})\right]_{i_1,\dots,i_k=1}^{n}.$$

Symmetric hypermatrices

Example

Cumulants of a random vector $\mathbf{x} = (X_1, \dots, X_n)$:

$$\kappa_k(\mathbf{x}) = \left[\sum_{A_1 \sqcup \dots \sqcup A_p = \{i_1, \dots, i_k\}} (-1)^{p-1}(p-1)! E\left(\prod_{i \in A_1} x_i\right) \cdots E\left(\prod_{i \in A_p} x_i\right) \right]_{i_1, \dots, i_k = 1}^{n}.$$

For $n = 1$, $\kappa_k(x)$ for $k = 1, 2, 3, 4$ are the expectation, variance, skewness, and kurtosis.

- Symmetric hypermatrices, in the form of cumulants, are of particular importance in Independent Component Analysis (ICA).

Symmetric eigenvalue decomposition

- Want to understand properties of symmetric rank, defined for $A \in \mathsf{S}^k(\mathbb{R}^n)$ as

$$\operatorname{rank}_{\mathsf{S}}(A) = \min\left\{ r \;\middle|\; A = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \cdots \otimes \mathbf{v}_i \right\}.$$

Lemma (Comon, Golub, L, Mourrain)

Let $A \in \mathsf{S}^k(\mathbb{R}^n)$. Then there exist $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{R}^n$ such that

$$A = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \cdots \otimes \mathbf{v}_i.$$

- If $A \in \mathsf{S}^k(\mathbb{R}^n)$, is $\operatorname{rank}_{\mathsf{S}}(A) = \operatorname{rank}_{\otimes}(A)$? Yes in many cases:
- P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensor and symmetric tensor rank," *SIAM J. Matrix Anal. Appl.*, to appear.

Multilinear spectral theory

- Eigenvalues/vectors of symmetric A are critical values/points of Rayleigh quotient, $\mathbf{x}^\top A\mathbf{x}/\|\mathbf{x}\|_2^2$.
- Similar characterization exists for singular values/vectors
- For $\mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$, write $\mathbf{x}^p := [x_1^p, \dots, x_n^p]^\top$. Define the '$\ell^k$-norm' $\|\mathbf{x}\|_k = (x_1^k + \cdots + x_n^k)^{1/k}$.
- Define eigenvalues/vectors of $A \in \mathsf{S}^k(\mathbb{R}^n)$ as critical values/points of the multilinear Rayleigh quotient

$$A(\mathbf{x}, \dots, \mathbf{x})/\|\mathbf{x}\|_k^k.$$

- $\mathbf{x}$ eigenvector iff $A(I, \mathbf{x}, \dots, \mathbf{x}) = \lambda \mathbf{x}^{k-1}$.
- Note that for a symmetric hypermatrix A,

$$A(I, \mathbf{x}, \mathbf{x}, \dots, \mathbf{x}) = A(\mathbf{x}, I, \mathbf{x}, \dots, \mathbf{x}) = \cdots = A(\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}, I).$$

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

Perron-Frobenius theorem for hypermatrices

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

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

  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*

1. *a positive real eigenvalue* λ *with an eigenvector* $\mathbf{x}$*;*
2. λ *may be chosen to have all entries non-negative;*
3. λ *is simple, ie.* $\mathbf{x}$ *is unique up to scaling.*

Spectral hypergraph theory

- Define the order-3 **adjacency hypermatrix** A by

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

- A is $|V|$-by-$|V|$-by-$|V|$ nonnegative symmetric hypermatrix.
- Consider cubic form $A(f, f, f) = \sum_{x,y,z} A_{xyz} f(x) f(y) f(z)$ (f is a vector of dimension $|V|$).
- Look at eigenvalues/vectors of A, ie. critical values/points of $A(f, f, f)$ constrained to $\sum_x f(x)^3 = 1$.

Lemma (L)

G m-regular 3-hypergraph. A its adjacency hypermatrix. Then

1. *m is an eigenvalue of A;*
2. *if λ is an eigenvalue of A, then $|\lambda| \leq m$;*
3. *λ has multiplicity 1 if and only if G is connected.*

Spectral hypergraph theory

- 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 (L)

Let G be a connected m-regular k-partite k-hypergraph on n vertices. Then

1. *If $k \equiv 1 \bmod 4$, then every eigenvalue of G occurs with multiplicity a multiple of k.*
2. *If $k \equiv 3 \bmod 4$, then the spectrum of G is symmetric, ie. if λ is an eigenvalue, then so is $-\lambda$.*
3. *Furthermore, every eigenvalue of G occurs with multiplicity a multiple of $k/2$, ie. if λ is an eigenvalue of G, then λ and $-\lambda$ occurs with the same multiplicity.*

Regularization

```
Date: Fri, 19 Oct 2007 06:50:56 -0700 (PDT)
From: Gene H Golub <golub@stanford.edu>
To: Lek-Heng Lim <lekheng@math.berkeley.edu>
Subject: Return

Dear Lek-Heng,

I am returning from Luminy tomorrow morning; my flight leaves at 7:15!
It's not been an easy meeting for me. I have not felt well --- my stomach
is bothering me. And depression is looming greater tan ever. If you recall,
I needed to see a doctor the last time I was here.

It occurred to me that it might be fun for the two of us to apply for a
grant together. Volker Mehrmann said he feels tensor decompositions are one
of the three most important problems this next decade! Have you any
knowledge of tensor decompositions and regularization. It could be a very
interesting topic.

Let me hear your thoughts.

Best,
Gene
```

Regularization

- The best low-rank approximation problem for hypermatrices is ill-posed in general.
- Over a wide range of dimensions, orders, ranks, one can construct hypermatrices A for which

$$\inf\{\|A - A_r\| \mid \operatorname{rank}_{\otimes}(A_r) \le r\}$$

  is not attained by any A_r with $\operatorname{rank}_{\otimes}(A_r) \le r$ [de Silva-L; 2006].
- This ill-posedness can be overcomed with appropriate regularization.
- If $\mathbf{u}_i, \mathbf{v}_i, \dots, \mathbf{z}_i$ are restricted to compact sets, then clearly a solution must exist. What about more general non-compact constraints?

Conic regularization

Theorem (L)

Let $A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. *Then*

$$\inf\{\|A - \sum_{i=1}^{r} \mathbf{u}_i \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{z}_i\| \mid \mathbf{u}_i, \ldots, \mathbf{z}_i \geq 0\}$$

is always attained.

Note that the optimization is over a product of nonnegative orthants. The result extends to more general cones.

Corollary

Nonnegative tensor approximation always has an optimal solution.

Final words

> *We take familiar things for granted. In particular, it is obvious to us that numerical practice is underpinned by solid, honest-to-god mathematical theory and this informs much of our professional life. This paradigm, which transcends any single theorem or result, we owe mainly to three individuals: Germund Dahlquist, Peter Lax, and Jim Wilkinson. In the early fifties they demonstrated that numerical algorithms do not just 'happen.' They can be understood and must be justified by rigourous mathematical analysis. If, as numerical analysts, we can see so far today, it is because we are standing on the shoulders of these giants and their generation.*

Arieh Iserles and Syvert Nørsett, "Colleagues remember Germund Dahlquist," *SIAM News*, **38** (2005), no. 4, pp. 3.