Hyperdeterminants, secant varieties, and tensor approximations

Lek-Heng Lim

University of California, Berkeley

April 23, 2008

Joint work with Vin de Silva

Hypermatrices

Totally ordered finite sets: $[n] = \{1 < 2 < \cdots < n\}$, $n \in \mathbb{N}$.

- Vector or n-tuple
$$f : [n] \to \mathbb{R}.$$
If $f(i) = a_i$, then f is represented by $\mathbf{a} = [a_1, \dots, a_n]^\top \in \mathbb{R}^n$.
- Matrix
$$f : [m] \times [n] \to \mathbb{R}.$$
If $f(i,j) = a_{ij}$, then f is represented by $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{R}^{m \times n}$.
- Hypermatrix (order 3)
$$f : [l] \times [m] \times [n] \to \mathbb{R}.$$
If $f(i,j,k) = a_{ijk}$, then f is represented by $\mathcal{A} = [\![a_{ijk}]\!]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l \times m \times n}$.

Normally $\mathbb{R}^X = \{f : X \to \mathbb{R}\}$. Ought to be $\mathbb{R}^{[n]}, \mathbb{R}^{[m]\times[n]}, \mathbb{R}^{[l]\times[m]\times[n]}$.

Hypermatrices and tensors

Up to choice of bases

- $\mathbf{a} \in \mathbb{R}^n$ can represent a vector in V (contravariant) or a linear functional in V^* (covariant).
- $A \in \mathbb{R}^{m \times n}$ can represent a bilinear form $V^* \times W^* \to \mathbb{R}$ (contravariant), a bilinear form $V \times W \to \mathbb{R}$ (covariant), or a linear operator $V \to W$ (mixed).
- $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$ can represent trilinear form $U \times V \times W \to \mathbb{R}$ (covariant), bilinear operators $V \times W \to U$ (mixed), etc.

A hypermatrix is the same as a tensor if

1. we give it coordinates (represent with respect to some bases);
2. we ignore covariance and contravariance.

Basic operation on a hypermatrix

- A matrix can be multiplied on the left and right: $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{p \times m}$, $Y \in \mathbb{R}^{q \times n}$,

$$(X, Y) \cdot A = XAY^{\top} = [c_{\alpha\beta}] \in \mathbb{R}^{p \times q}$$

where

$$c_{\alpha\beta} = \sum\nolimits_{i,j=1}^{m,n} x_{\alpha i} y_{\beta j} a_{ij}.$$

- A hypermatrix can be multiplied on three sides: $\mathcal{A} = [\![a_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$, $X \in \mathbb{R}^{p \times l}$, $Y \in \mathbb{R}^{q \times m}$, $Z \in \mathbb{R}^{r \times n}$,

$$(X, Y, Z) \cdot \mathcal{A} = [\![c_{\alpha\beta\gamma}]\!] \in \mathbb{R}^{p \times q \times r}$$

where

$$c_{\alpha\beta\gamma} = \sum\nolimits_{i,j,k=1}^{l,m,n} x_{\alpha i} y_{\beta j} z_{\gamma k} a_{ijk}.$$

Basic operation on a hypermatrix

- Covariant version:

$$\mathcal{A} \cdot (X^\top, Y^\top, Z^\top) := (X, Y, Z) \cdot \mathcal{A}.$$

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

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

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

Symmetric hypermatrices

- Cubical hypermatrix $[\![a_{ijk}]\!] \in \mathbb{R}^{n\times n\times n}$ is **symmetric** if

$$a_{ijk} = a_{ikj} = a_{jik} = a_{jki} = a_{kij} = a_{kji}.$$

- Invariant under all permutations $\sigma \in \mathfrak{S}_k$ on indices.
- $\mathsf{S}^k(\mathbb{R}^n)$ denotes set of all order-k symmetric hypermatrices.

Example

Higher order derivatives of multivariate functions.

Example

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

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

Symmetric hypermatrices

Example

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

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

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

- Important in Independent Component Analysis (ICA).

Inner products and norms

- $\ell^2([n])$: $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b} = \sum_{i=1}^{n} a_i b_i$.
- $\ell^2([m] \times [n])$: $A, B \in \mathbb{R}^{m \times n}$, $\langle A, B \rangle = \operatorname{tr}(A^\top B) = \sum_{i,j=1}^{m,n} a_{ij} b_{ij}$.
- $\ell^2([l] \times [m] \times [n])$: $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{l \times m \times n}$, $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i,j,k=1}^{l,m,n} a_{ijk} b_{ijk}$.
- In general,

$$\ell^2([m] \times [n]) = \ell^2([m]) \otimes \ell^2([n]),$$
$$\ell^2([l] \times [m] \times [n]) = \ell^2([l]) \otimes \ell^2([m]) \otimes \ell^2([n]).$$

- Frobenius norm

$$\|\mathcal{A}\|_F^2 = \sum_{i,j,k=1}^{l,m,n} a_{ijk}^2.$$

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

$$f(\mathbf{x}) = a_0 + \mathbf{a}_1^\top \mathbf{x} + \mathbf{x}^\top A_2 \mathbf{x} + \mathcal{A}_3(\mathbf{x}, \mathbf{x}, \mathbf{x}) + \dots + \mathcal{A}_d(\mathbf{x}, \dots, \mathbf{x}).$$

  $a_0 \in \mathbb{R}, \mathbf{a}_1 \in \mathbb{R}^n, A_2 \in \mathbb{R}^{n \times n}, \mathcal{A}_3 \in \mathbb{R}^{n \times n \times n}, \dots, \mathcal{A}_d \in \mathbb{R}^{n \times \dots \times n}$.
- Numerical linear algebra: $d = 2$.
- Numerical multilinear algebra: $d > 2$.

Multilinear spectral theory

Let $\mathcal{A} \in \mathbb{R}^{n \times n \times n}$ (easier if $\mathcal{A}$ symmetric).

1. How should one define its eigenvalues and eigenvectors?
2. What is a decomposition that generalizes the eigenvalue decomposition of a matrix?

Let $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$

1. How should one define its singular values and singular vectors?
2. What is a decomposition that generalizes the singular value decomposition of a matrix?

Somewhat surprising: (1) and (2) have different answers.

Multilinear spectral theory

- May define eigenvalues/vectors of $\mathcal{A} \in \mathsf{S}^k(\mathbb{R}^n)$ as critical values/points of the multilinear Raleigh quotient

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

- Lagrangian

$$L(\mathbf{x}, \lambda) := \mathcal{A}(\mathbf{x}, \dots, \mathbf{x}) - \lambda(\|\mathbf{x}\|_k^k - 1).$$

- At a critical point

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

- Ditto for singular values/vectors of $\mathcal{A} \in \mathbb{R}^{d_1 \times \dots \times d_k}$.
- Perron-Frobenius theorem for irreducible non-negative hypermatrices, spectral hypergraph theory:
  - L, "Singular values and eigenvalues of tensors: a variational approach," *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, **1** (2005).

Tensor ranks (Hitchcock, 1927)

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

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

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

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

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

Eigenvalue and singular value decompositions

- Rank revealing decompositions associated with outer product rank.
- **Symmetric eigenvalue decomposition** of $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$,

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

  where $\operatorname{rank}_{\mathsf{S}}(A) = \min\left\{r \;\middle|\; \mathcal{A} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i\right\} = r$.
  - P. Comon, G. Golub, L, B. Mourrain, "Symmetric tensor and symmetric tensor rank," *SIAM J. Matrix Anal. Appl.*
- **Singular value decomposition** of $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$,

$$\mathcal{A} = \sum\nolimits_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i \tag{2}$$

  where $\operatorname{rank}_{\otimes}(\mathcal{A}) = r$.
  - V. de Silva, L, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*
- (1) used in applications of ICA to signal processing; (2) used in applications of the PARAFAC model to analytical chemistry.

Eigenvalue and singular value decompositions

- Rank revealing decompositions associated with the multilinear rank.
- **Symmetric eigenvalue decomposition** of $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$,

$$\mathcal{A} = (U, U, U) \cdot \mathcal{C} \tag{3}$$

  where $\operatorname{rank}_{\boxplus}(A) = (r, r, r)$, $U \in \mathbb{R}^{n \times r}$ has orthonormal columns and $\mathcal{C} \in \mathsf{S}^3(\mathbb{R}^r)$.
- **Singular value decomposition** of $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$,

$$\mathcal{A} = (U, V, W) \cdot \mathcal{C} \tag{4}$$

  where $\operatorname{rank}_{\boxplus}(A) = (r_1, r_2, r_3)$, $U \in \mathbb{R}^{l \times r_1}$, $V \in \mathbb{R}^{m \times r_2}$, $W \in \mathbb{R}^{n \times r_3}$ have orthonormal columns and $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$.
  - L. De Lathauwer, B. De Moor, J. Vandewalle "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, **21** (2000), no. 4.
  - B. Savas, L, "Best multilinear rank approximation with quasi-Newton method on Grassmannians," *preprint.*

Segre variety and its secant varieties

- The set of all rank-1 hypermatrices is known as the Segre variety in algebraic geometry.
- It is a closed set (in both the Euclidean and Zariski sense) as it can be described algebraically:

$$\operatorname{Seg}(\mathbb{R}^l, \mathbb{R}^m, \mathbb{R}^n) = \{\mathcal{A} \in \mathbb{R}^{l \times m \times n} \mid \mathcal{A} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}\} = \\ \{A \in \mathbb{R}^{l \times m \times n} \mid a_{i_1 i_2 i_3} a_{j_1 j_2 j_3} = a_{k_1 k_2 k_3} a_{l_1 l_2 l_3}, \{i_\alpha, j_\alpha\} = \{k_\alpha, l_\alpha\}\}$$

- Hypermatrices that have rank > 1 are elements on the higher secant varieties of $\mathscr{S} = \operatorname{Seg}(\mathbb{R}^l, \mathbb{R}^m, \mathbb{R}^n)$.
- E.g. a hypermatrix has rank 2 if it sits on a secant line through two points in $\mathscr{S}$ but not on $\mathscr{S}$, rank 3 if it sits on a secant plane through three points in $\mathscr{S}$ but not on any secant lines, etc.

Decomposition approach to data analysis

- More generally, $\mathbb{F} = \mathbb{C}$, $\mathbb{R}$, $\mathbb{R}_+, \mathbb{R}_{\max}$ (max-plus algebra), $\mathbb{R}[x_1, \dots, x_n]$ (polynomial rings), etc.
- Dictionary, $\mathscr{D} \subset \mathbb{F}^N$, not contained in any hyperplane.
- Let $\mathscr{D}_2 =$ union of bisecants to $\mathscr{D}$, $\mathscr{D}_3 =$ union of trisecants to $\mathscr{D}$, $\dots$, $\mathscr{D}_r =$ union of r-secants to $\mathscr{D}$.
- Define $\mathscr{D}$-rank of $\mathcal{A} \in \mathbb{F}^N$ to be $\min\{r \mid \mathcal{A} \in \mathscr{D}_r\}$.
- If $\varphi : \mathbb{F}^N \times \mathbb{F}^N \to \mathbb{R}$ is some measure of 'nearness' between pairs of points (e.g. norms, Bregman divergences, etc), we want to find a best low-rank approximation to $\mathcal{A}$:

$$\operatorname{argmin}\{\varphi(\mathcal{A}, \mathcal{B}) \mid \mathscr{D}\text{-rank}(B) \leq r\}.$$

Decomposition approach to data analysis

- In the presence of noise, approximation instead of decomposition

$$\mathcal{A} \approx \alpha_1 \cdot \mathcal{B}_1 + \cdots + \alpha_r \cdot \mathcal{B}_r \in \mathscr{D}_r.$$

  $\mathcal{B}_i \in \mathscr{D}$ reveal features of the dataset $\mathcal{A}$.

- Note that another way to say 'best low-rank' is 'sparsest possible'.

Examples

1. CANDECOMP/PARAFAC: $\mathscr{D} = \{\mathcal{A} \mid \operatorname{rank}_{\otimes}(\mathcal{A}) \leq 1\}$, $\varphi(\mathcal{A}, \mathcal{B}) = \|\mathcal{A} - \mathcal{B}\|_F$.
2. De Lathauwer model: $\mathscr{D} = \{\mathcal{A} \mid \operatorname{rank}_{\boxplus}(\mathcal{A}) \leq (r_1, r_2, r_3)\}$, $\varphi(\mathcal{A}, \mathcal{B}) = \|\mathcal{A} - \mathcal{B}\|_F$.

Fundamental problem of multiway data analysis

- $\mathcal{A}$ hypermatrix, symmetric hypermatrix, or nonnegative hypermatrix.
- Solve
$$\operatorname{argmin}_{\operatorname{rank}(\mathcal{B})\leq r}\|\mathcal{A}-\mathcal{B}\|.$$
- rank may be outer product rank, multilinear rank, symmetric rank (for symmetric hypermatrix), or nonnegative rank (nonnegative hypermatrix).

Example

Given $\mathcal{A}\in\mathbb{R}^{d_1\times d_2\times d_3}$, find $\mathbf{u}_i,\mathbf{v}_i,\mathbf{w}_i$, $i=1,\dots,r$, that minimizes
$$\|\mathcal{A}-\mathbf{u}_1\otimes\mathbf{v}_1\otimes\mathbf{w}_1-\mathbf{u}_2\otimes\mathbf{v}_2\otimes\mathbf{w}_2-\cdots-\mathbf{u}_r\otimes\mathbf{v}_r\otimes\mathbf{z}_r\|$$
or $\mathcal{C}\in\mathbb{R}^{r_1\times r_2\times r_3}$ and $U\in\mathbb{R}^{d_1\times r_1}, V\in\mathbb{R}^{d_2\times r_2}, W\in\mathbb{R}^{d_3\times r_3}$, that minimizes
$$\|\mathcal{A}-(U,V,W)\cdot\mathcal{C}\|.$$

Fundamental problem of multiway data analysis

Example

Given $\mathcal{A} \in \mathsf{S}^k(\mathbb{C}^n)$, find $\mathbf{u}_i$, $i = 1, \dots, r$, that minimizes

$$\|\mathcal{A} - \mathbf{u}_1^{\otimes k} - \mathbf{u}_2^{\otimes k} - \cdots - \mathbf{u}_r^{\otimes k}\|$$

or $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and $U \in \mathbb{R}^{n \times r_i}$ that minimizes

$$\|\mathcal{A} - (U, U, U) \cdot \mathcal{C}\|.$$

Separation of variables

Approximation by sum or integral of separable functions

- Continuous

$$f(x, y, z) = \int \theta(x, t)\varphi(y, t)\psi(z, t)\, dt.$$

- Semi-discrete

$$f(x, y, z) = \sum\nolimits_{p=1}^{r} \theta_p(x)\varphi_p(y)\psi_p(z)$$

$\theta_p(x) = \theta(x, t_p)$, $\varphi_p(y) = \varphi(y, t_p)$, $\psi_p(z) = \psi(z, t_p)$, r possibly ∞.

- Discrete

$$a_{ijk} = \sum\nolimits_{p=1}^{r} u_{ip}v_{jp}w_{kp}$$

$a_{ijk} = f(x_i, y_j, z_k)$, $u_{ip} = \theta_p(x_i)$, $v_{jp} = \varphi_p(y_j)$, $w_{kp} = \psi_p(z_k)$.

Separation of variables

- Useful for data analysis, machine learning, pattern recognition.
- Gaussians are separable

$$\exp(x^2 + y^2 + z^2) = \exp(x^2)\exp(y^2)\exp(z^2).$$

- More generally for symmetric positive-definite $A \in \mathbb{R}^{n\times n}$,

$$\exp(\mathbf{x}^\top A\mathbf{x}) = \exp(\mathbf{z}^\top \Lambda\mathbf{z}) = \prod_{i=1}^{n} \exp(\lambda_i z_i^2).$$

- Gaussian mixture models

$$f(\mathbf{x}) = \sum_{j=1}^{m} \alpha_j \exp[(\mathbf{x} - \boldsymbol{\mu}_j)^\top A_j(\mathbf{x} - \boldsymbol{\mu}_j)],$$

f is a sum of separable functions.

Integral kernels

Approximation by sum or integral kernels

- Continuous

$$f(x,y,z) = \iiint K(x',y',z')\theta(x,x')\varphi(y,y')\psi(z,z')\,dx'dy'dz'.$$

- Semi-discrete

$$f(x,y,z) = \sum_{i',j',k'=1}^{p,q,r} c_{i'j'k'}\theta_{i'}(x)\varphi_{j'}(y)\psi_{k'}(z)$$

$c_{i'j'k'} = K(x'_{i'}, y'_{j'}, z'_{k'})$, $\theta_{i'}(x) = \theta(x, x'_{i'})$, $\varphi_{j'}(y) = \varphi(y, y'_{j'})$, $\psi_{k'}(z) = \psi(z, z'_{k'})$, p, q, r possibly ∞.

- Discrete

$$a_{ijk} = \sum_{i',j',k'=1}^{p,q,r} c_{i'j'k'}u_{ii'}v_{jj'}w_{kk'}$$

$a_{ijk} = f(x_i, y_j, z_k)$, $u_{ii'} = \theta_{i'}(x_i)$, $v_{jj'} = \varphi_{j'}(y_j)$, $w_{kk'} = \psi_{k'}(z_k)$.

Lemma (de Silva, L)

Let $r \geq 2$ and $k \geq 3$. Given the norm-topology on $\mathbb{R}^{d_1 \times \cdots \times d_k}$, the following statements are equivalent:

1. *The set* $\mathscr{S}_r(d_1, \dots, d_k) := \{\mathcal{A} \mid \operatorname{rank}_\otimes(\mathcal{A}) \leq r\}$ *is not closed.*
2. *There exists a sequence* $\mathcal{A}_n$, $\operatorname{rank}_\otimes(\mathcal{A}_n) \leq r$, $n \in \mathbb{N}$, *converging to* $\mathcal{B}$ *with* $\operatorname{rank}_\otimes(\mathcal{B}) > r$.
3. *There exists* $\mathcal{B}$, $\operatorname{rank}_\otimes(\mathcal{B}) > r$, *that may be approximated arbitrarily closely by hypermatrices of strictly lower rank, i.e.*
$$\inf\{\|\mathcal{B} - \mathcal{A}\| \mid \operatorname{rank}_\otimes(\mathcal{A}) \leq r\} = 0.$$
4. *There exists* $\mathcal{C}$, $\operatorname{rank}_\otimes(\mathcal{C}) > r$, *that does not have a best rank-r approximation, i.e.*
$$\inf\{\|\mathcal{C} - \mathcal{A}\| \mid \operatorname{rank}_\otimes(\mathcal{A}) \leq r\}$$
is not attained (by any $\mathcal{A}$ *with* $\operatorname{rank}_\otimes(\mathcal{A}) \leq r$*).*

Non-existence of best low-rank approximation

- For $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^{d_i}$, $i = 1, 2, 3$,

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

- For $n \in \mathbb{N}$,

$$\mathcal{A}_n := n\left(\mathbf{x}_1 + \frac{1}{n}\mathbf{y}_1\right) \otimes \left(\mathbf{x}_2 + \frac{1}{n}\mathbf{y}_2\right) \otimes \left(\mathbf{x}_3 + \frac{1}{n}\mathbf{y}_3\right) - n\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3.$$

Lemma (de Silva, L)

$\operatorname{rank}_\otimes(\mathcal{A}) = 3$ *iff* $\mathbf{x}_i, \mathbf{y}_i$ *linearly independent,* $i = 1, 2, 3$. *Furthermore, it is clear that* $\operatorname{rank}_\otimes(\mathcal{A}_n) \leq 2$ *and*

$$\lim_{n\to\infty} \mathcal{A}_n = \mathcal{A}.$$

- Original result, in a different form, due to:
  - D. Bini, G. Lotti, F. Romani, "Approximate solutions for the bilinear form computational problem," *SIAM J. Comput.*, **9** (1980), no. 4.

Outer product approximations are ill-behaved

- Such phenomenon can and will happen for all orders > 2, all norms, and many ranks:

Theorem (de Silva, L)

Let $k \geq 3$ and $d_1, \dots, d_k \geq 2$. For any s such that

$$2 \leq s \leq \min\{d_1, \dots, d_k\},$$

there exists $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ with $\operatorname{rank}_{\otimes}(\mathcal{A}) = s$ such that $\mathcal{A}$ has no best rank-r approximation for some $r < s$. The result is independent of the choice of norms.

- For matrices, the quantity $\min\{d_1, d_2\}$ will be the maximal possible rank in $\mathbb{R}^{d_1 \times d_2}$. In general, a hypermatrix in $\mathbb{R}^{d_1 \times \cdots \times d_k}$ can have rank exceeding $\min\{d_1, \dots, d_k\}$.

Outer product approximations are ill-behaved

- Tensor rank can jump over an arbitrarily large gap:

Theorem (de Silva, L)

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

- Hypermatrices that fail to have best low-rank approximations are not rare. May occur with non-zero probability; sometimes with certainty.

Theorem (de Silva, L)

Let μ be a measure that is positive or infinite on Euclidean open sets in $\mathbb{R}^{l\times m\times n}$. There exists some $r \in \mathbb{N}$ such that

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

In $\mathbb{R}^{2\times 2\times 2}$, all rank-3 hypermatrices fail to have best rank-2 approximation.

Message

- That the best rank-r approximation problem for hypermatrices has no solution poses serious difficulties.
- 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? i.e. what is the 'approximate solution' an approximate of?

Weak solutions

- For a hypermatrix $\mathcal{A}$ that has no best rank-r approximation, we will call a $\mathcal{C} \in \overline{\{\mathcal{A} \mid \operatorname{rank}_{\otimes}(\mathcal{A}) \leq r\}}$ attaining

$$\inf\{\|\mathcal{C} - \mathcal{A}\| \mid \operatorname{rank}_{\otimes}(\mathcal{A}) \leq r\}$$

  a **weak solution**. In particular, we must have $\operatorname{rank}_{\otimes}(\mathcal{C}) > r$.
- It is perhaps surprising that one may completely parameterize all limit points of order-3 rank-2 hypermatrices.

Weak solutions

Theorem (de Silva, L)

Let $d_1, d_2, d_3 \geq 2$. *Let* $\mathcal{A}_n \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ *be a sequence of hypermatrices with* $\operatorname{rank}_{\otimes}(\mathcal{A}_n) \leq 2$ *and*

$$\lim_{n\to\infty} \mathcal{A}_n = \mathcal{A},$$

where the limit is taken in any norm topology. If the limiting hypermatrix $\mathcal{A}$ *has rank higher than* 2*, then* $\operatorname{rank}_{\otimes}(\mathcal{A})$ *must be exactly* 3 *and there exist pairs of linearly independent vectors* $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^{d_1}$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^{d_2}$, $\mathbf{x}_3, \mathbf{y}_3 \in \mathbb{R}^{d_3}$ *such that*

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

- In particular, a sequence of order-3 rank-2 hypermatrices cannot ‘jump rank’ by more than 1.

Hyperdeterminant

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

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

Theorem (Gelfand, Kapranov, Zelevinsky)

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

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

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

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

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

$2 \times 2 \times 2$ hyperdeterminant

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

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

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 $\operatorname{Det}(\mathcal{A}) = 0$.

$2 \times 2 \times 3$ hyperdeterminant

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

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

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 $\operatorname{Det}(\mathcal{A}) = 0$.

Cayley hyperdeterminant and tensor rank

- The Cayley hyperdeterminant $\text{Det}_{2,2,2}$ may be extended to any $\mathcal{A} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with $\text{rank}_{\otimes}(\mathcal{A}) \leq 2$.

Theorem (de Silva, L)

Let $d_1, d_2, d_3 \geq 2$*.* $\mathcal{A} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ *is a weak solution, i.e.*

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

iff $\text{Det}_{2,2,2}(\mathcal{A}) = 0$.

Theorem (Kruskal)

Let $\mathcal{A} \in \mathbb{R}^{2 \times 2 \times 2}$*. Then* $\text{rank}_{\otimes}(\mathcal{A}) = 2$ *if* $\text{Det}_{2,2,2}(\mathcal{A}) > 0$ *and* $\text{rank}_{\otimes}(\mathcal{A}) = 3$ *if* $\text{Det}_{2,2,2}(\mathcal{A}) < 0$.

- See de Silva-L for a proof via the Cayley hyperdeterminant.

Symmetric hypermatrices for blind source separation

Problem

Given $\mathbf{y} = M\mathbf{x} + \mathbf{n}$. *Unknown: source vector* $\mathbf{x} \in \mathbb{C}^n$, *mixing matrix* $M \in \mathbb{C}^{m\times n}$, *noise* $\mathbf{n} \in \mathbb{C}^m$. *Known: observation vector* $\mathbf{y} \in \mathbb{C}^m$. *Goal: recover* $\mathbf{x}$ *from* $\mathbf{y}$.

- Assumptions:
  1. components of $\mathbf{x}$ statistically independent,
  2. M full column-rank,
  3. $\mathbf{n}$ Gaussian.
- Method: use cumulants

$$\kappa_k(\mathbf{y}) = (M, M, \dots, M) \cdot \kappa_k(\mathbf{x}) + \kappa_k(\mathbf{n}).$$

- By assumptions, $\kappa_k(\mathbf{n}) = 0$ and $\kappa_k(\mathbf{x})$ is diagonal. So need to diagonalize the symmetric hypermatrix $\kappa_k(\mathbf{y})$.

Diagonalizing a symmetric hypermatrix

- A best symmetric rank approximation may not exist either:

Example (Comon, Golub, L, Mourrain)

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be linearly independent. Define for $n \in \mathbb{N}$,

$$\mathcal{A}_n := n\left(\mathbf{x} + \frac{1}{n}\mathbf{y}\right)^{\otimes k} - n\mathbf{x}^{\otimes k}$$

and

$$\mathcal{A} := \mathbf{x} \otimes \mathbf{y} \otimes \cdots \otimes \mathbf{y} + \mathbf{y} \otimes \mathbf{x} \otimes \cdots \otimes \mathbf{y} + \cdots + \mathbf{y} \otimes \mathbf{y} \otimes \cdots \otimes \mathbf{x}.$$

Then $\operatorname{rank}_{\mathsf{S}}(\mathcal{A}_n) \leq 2$, $\operatorname{rank}_{\mathsf{S}}(\mathcal{A}) = k$, and

$$\lim_{n\to\infty} \mathcal{A}_n = \mathcal{A}.$$

Nonnegative hypermatrices and nonnegative tensor rank

- Let $0 \leq \mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. The nonnegative rank of $\mathcal{A}$ is

$$\operatorname{rank}_+(\mathcal{A}) := \min\left\{ r \mid \sum\nolimits_{i=1}^{r} \mathbf{u}_i \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{z}_i,\ \mathbf{u}_i, \dots, \mathbf{z}_i \geq 0 \right\}$$

- Clearly nonnegative decomposition exists for any $\mathcal{A} \geq 0$.
- Arises in the Naïve Bayes model.

Theorem (L)

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

$$\inf\left\{ \left\| \mathcal{A} - \sum\nolimits_{i=1}^{r} \mathbf{u}_i \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{z}_i \right\| \mid \mathbf{u}_i, \dots, \mathbf{z}_i \geq 0 \right\}$$

is always attained.

Advertisement

Geometry and representation theory of tensors for computer science, statistics, and other areas

1. MSRI Summer Graduate Workshop
   - July 7 to July 18, 2008
   - Organized by J.M. Landsberg, L.-H. Lim, J. Morton
   - Mathematical Sciences Research Institute, Berkeley, CA
   - `http://msri.org/calendar/sgw/WorkshopInfo/451/show_sgw`
2. AIM Workshop
   - July 21 to July 25, 2008
   - Organized by J.M. Landsberg, L.-H. Lim, J. Morton, J. Weyman
   - American Institute of Mathematics, Palo Alto, CA
   - `http://aimath.org/ARCC/workshops/repnsoftensors.html`