Conditioning and illposedness of tensor approximations

and why engineers should care

Lek-Heng Lim

MSRI Summer Graduate Workshop

July 7–18, 2008

Slight change of plans

- **Week 1**
  - Mon: *Overview: tensor approximations* (**LH**)
  - Wed: *Conditioning and ill-posedness of tensor approximations, nonnegative and symmetric tensors, some applications* (**LH**)
- **Week 2**
  - Tue: *Computations: Gauss-Seidel method, semidefinite programming, Hilbert's 17th problem, optimization on Stiefel and Grassmann manifolds* (**LH**)

Tensor rank is hard to compute

- Eugene L. Lawler: "The Mystical Power of Twoness."
  - 2-SAT is easy, 3-SAT is hard;
  - 2-dimensional matching is easy, 3-dimensional matching is hard;
  - Order-2 tensor rank is easy, order-3 tensor rank is hard.

Theorem (Håstad)

Computing $\operatorname{rank}_{\otimes}(\mathcal{A})$ *for* $\mathcal{A} \in \mathbb{F}^{l \times m \times n}$ *is NP-hard for* $\mathbb{F} = \mathbb{Q}$ *and NP-complete for* $\mathbb{F} = \mathbb{F}_q$*?*

- **Open question:** Is tensor rank NP-hard/NP-complete over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ in the sense of BCSS?
  - L. Blum, F. Cucker, M. Shub, S. Smale, *Complexity and real computation*, Springer-Verlag, New York, NY, 1998.

Tensor rank depends on base field

For $A \in \mathbb{R}^{m \times n} \subset \mathbb{C}^{m \times n}$, $\operatorname{rank}_{\mathbb{R}}(A) = \operatorname{rank}_{\mathbb{C}}(A)$. Not true for tensors.

Theorem (Bergman)

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

- $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ linearly independent and let $\mathbf{z} = \mathbf{x} + i\mathbf{y}$.

$$\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} - \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{y} + \mathbf{y} \otimes \mathbf{x} \otimes \mathbf{y} + \mathbf{y} \otimes \mathbf{y} \otimes \mathbf{x} = \frac{1}{2}(\mathbf{z} \otimes \bar{\mathbf{z}} \otimes \bar{\mathbf{z}} + \bar{\mathbf{z}} \otimes \mathbf{z} \otimes \mathbf{z}).$$

- May show that $\operatorname{rank}_{\otimes,\mathbb{R}}(\mathcal{A}) = 3$ and $\operatorname{rank}_{\otimes,\mathbb{C}}(\mathcal{A}) = 2$.
- $\mathbb{R}^{2 \times 2 \times 2}$ has 8 distinct orbits under $\mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R})$.
- $\mathbb{C}^{2 \times 2 \times 2}$ has 7 distinct orbits under $\mathrm{GL}_2(\mathbb{C}) \times \mathrm{GL}_2(\mathbb{C}) \times \mathrm{GL}_2(\mathbb{C})$.

Recall: fundamental problem of multiway data analysis

- $\mathcal{A}$ hypermatrix, symmetric hypermatrix, or nonnegative hypermatrix. Want

$$\operatorname{argmin}_{\operatorname{rank}(\mathcal{B})\le r}\|\mathcal{A}-\mathcal{B}\|.$$

- $\operatorname{rank}(\mathcal{B})$ 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 $\sigma_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$, $i=1,\dots,r$, that minimizes

$$\|\mathcal{A}-\sigma_1\mathbf{u}_1\otimes\mathbf{v}_1\otimes\mathbf{w}_1-\sigma_2\mathbf{u}_2\otimes\mathbf{v}_2\otimes\mathbf{w}_2-\cdots-\sigma_r\mathbf{u}_r\otimes\mathbf{v}_r\otimes\mathbf{w}_r\|$$

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

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

- May assume $\mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$ unit vectors and U, V, W orthonormal columns.

Recall: 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

$$\|\mathcal{A} - \lambda_1 \mathbf{u}_1^{\otimes k} - \lambda_2 \mathbf{u}_2^{\otimes k} - \cdots - \lambda_r \mathbf{u}_r^{\otimes k}\|$$

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

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

- May assume $\mathbf{u}_i$ unit vector and U orthonormal columns.
- Pierre's lectures in Week 2.

Best low rank approximation of a matrix

- Given $A \in \mathbb{R}^{m \times n}$. Want

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

- More precisely, find $\sigma_i, \mathbf{u}_i, \mathbf{v}_i$, $i = 1, \dots, r$, that minimizes

$$\|\mathcal{A} - \sigma_1 \mathbf{u}_1 \otimes \mathbf{v}_1 - \sigma_2 \mathbf{u}_2 \otimes \mathbf{v}_2 - \cdots - \sigma_r \mathbf{u}_r \otimes \mathbf{v}_r\|.$$

Theorem (Eckart–Young)

Let $A = U \Sigma V^\top = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ *be singular value decomposition. For* $r \le \operatorname{rank}(A)$, *let*

$$A_r := \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top.$$

Then

$$\|A - A_r\|_F = \min_{\operatorname{rank}(B) \le r} \|A - B\|_F.$$

- No such thing for hypermatrices of order 3 or higher.

Lemma

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* $\mathscr{S}_r(d_1, \ldots, d_k) := \{\mathcal{A} \mid \operatorname{rank}_\otimes(\mathcal{A}) \leq r\}$ *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

$\operatorname{rank}_\otimes(\mathcal{A}) = 3$ *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 slightly 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

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 $\min\{d_1, \dots, d_k\}$.
- Vin's next three lectures.

Outer product approximations are ill-behaved

- Tensor rank can jump over an arbitrarily large gap:

Theorem

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

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

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

In $\mathbb{R}^{2\times2\times2}$, 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

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.
- Details: Vin's lectures. Not possible in general: JM's lectures.

Conditioning of 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 $A\mathbf{x} = \mathbf{b}$.
- $\mathscr{M} = \{A \in \mathbb{R}^{n\times n} \mid \det(A) = 0\}$ is the manifold of ill-posed problems.
- $A \in \mathscr{M}$ 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 **condition number**

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

- Note that

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

Conditioning of linear systems

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

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

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 $\sigma_n^{-1} = \|A^{-1}\|_2$.
- Rule of thumb: $\log_{10}\kappa_2(A) \approx$ loss in number of digits of precision.

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. $\mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}, \mathbf{z} \neq \mathbf{0}$?

$2 \times 2 \times 2$ hyperdeterminant

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

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

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

$2 \times 2 \times 3$ hyperdeterminant

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

$$\operatorname{Det}_{2,2,3}(\mathcal{A}) = \det\begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{100} & a_{101} & a_{102} \\ a_{010} & a_{011} & a_{012} \end{bmatrix} \det\begin{bmatrix} a_{100} & a_{101} & a_{102} \\ a_{010} & a_{011} & a_{012} \\ a_{110} & a_{111} & a_{112} \end{bmatrix} - \det\begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{100} & a_{101} & a_{102} \\ a_{110} & a_{111} & a_{112} \end{bmatrix} \det\begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{010} & a_{011} & a_{012} \\ a_{110} & a_{111} & a_{112} \end{bmatrix}$$

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 $\operatorname{Det}_{2,2,3}(\mathcal{A}) = 0$.

Cayley hyperdeterminant and tensor rank

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

Theorem

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

Theorem (Kruskal)

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

- Vin's next three lectures.

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 $\mathscr{M}$.

Theorem

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

Nonnegative hypermatrices and nonnegative tensor rank

- Let $0 \leq \mathcal{A} \in \mathbb{R}^{l \times m \times n}$. The nonnegative rank of $\mathcal{A}$ is

$$\operatorname{rank}_+(\mathcal{A}) := \min\bigl\{r \bigm| \sum\nolimits_{p=1}^{r} \mathbf{x}_p \otimes \mathbf{y}_p \otimes \mathbf{z}_p,\ \mathbf{x}_p, \mathbf{y}_p, \mathbf{z}_p \geq 0\bigr\}$$

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

Theorem

Let $\mathcal{A} = [\![a_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$ be nonnegative. Then

$$\inf\bigl\{\bigl\|\mathcal{A} - \sum\nolimits_{p=1}^{r} \mathbf{x}_p \otimes \mathbf{y}_p \otimes \mathbf{z}_p\bigr\| \bigm| \mathbf{x}_p, \mathbf{y}_p, \mathbf{z}_p \geq 0\bigr\}$$

is always attained.

Nonnegative matrix factorization

- D.D. Lee and H.S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, **401** (1999), pp. 788–791.
- **Main idea behind NMF** (everything else is fluff): the way dictionary functions combine to build 'target objects' is an exclusively additive process and should not involve any cancellations between the dictionary functions.
- **NMF in a nutshell**: given nonnegative matrix A, decompose it into a sum of outer-products of nonnegative vectors:

$$A = XY^{\top} = \sum\nolimits_{i=1}^{r} \mathbf{x}_i \otimes \mathbf{y}_i.$$

- **Noisy situation**: approximate A by a sum of outer-products of nonnegative vectors

$$\min_{X \geq 0, Y \geq 0} \|A - XY^{\top}\|_F = \min_{\mathbf{x}_i \geq 0, \mathbf{y}_i \geq 0} \left\| A - \sum\nolimits_{i=1}^{r} \mathbf{x}_i \otimes \mathbf{y}_i \right\|_F .$$

Generalizing to hypermatrices

- **Nonnegative outer-product decomposition** for hypermatrix $\mathcal{A} \geq 0$ is
$$\mathcal{A} = \sum\nolimits_{p=1}^{r} \mathbf{x}_p \otimes \mathbf{y}_p \otimes \mathbf{z}_p$$
where $\mathbf{x}_p \in \mathbb{R}_+^l, \mathbf{y}_p \in \mathbb{R}_+^m, \mathbf{z}_p \in \mathbb{R}_+^n$.
- Clear that such a decomposition exists for any $\mathcal{A} \geq 0$.
- **Nonnegative outer-product rank**: minimal r for which such a decomposition is possible.
- Best nonnegative outer-product rank-r approximation:
$$\operatorname{argmin}\left\{\left\|\mathcal{A} - \sum\nolimits_{p=1}^{r} \mathbf{x}_p \otimes \mathbf{y}_p \otimes \mathbf{z}_p\right\|_F \;\middle|\; \mathbf{x}_p, \mathbf{y}_p, \mathbf{z}_p \geq 0\right\}.$$

Recap: outer product decomposition in spectroscopy

- Application to fluorescence spectral analysis by [Bro; 1997].
- Specimens with a number of pure substances in different concentration
  - a_{ijk} = fluorescence emission intensity at wavelength λ_j^{em} of ith sample excited with light at wavelength λ_k^{ex}.
  - Get 3-way data $\mathcal{A} = [\![a_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$.
  - Get outer product decomposition of $\mathcal{A}$

$$\mathcal{A} = \mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1 + \cdots + \mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r.$$

- Get the true chemical factors responsible for the data.
  - r: number of pure substances in the mixtures,
  - $\mathbf{x}_p = (x_{1p}, \dots, x_{lp})$: relative concentrations of pth substance in specimens $1, \dots, l$,
  - $\mathbf{y}_p = (y_{1p}, \dots, y_{mp})$: excitation spectrum of pth substance,
  - $\mathbf{z}_p = (z_{1p}, \dots, z_{np})$: emission spectrum of pth substance.
- Noisy case: find best rank-r approximation (CANDECOMP/PARAFAC).

Proof

- Naive choice of objective: $g : (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^n)^r \to \mathbb{R}$,

$$g(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1, \dots, \mathbf{x}_r, \mathbf{y}_r, \mathbf{z}_r) := \left\| \mathcal{A} - \sum\nolimits_{p=1}^{r} \mathbf{x}_p \otimes \mathbf{y}_p \otimes \mathbf{z}_p \right\|_F^2.$$

- Need to show g attains infimum on $(\mathbb{R}^l_+ \times \mathbb{R}^m_+ \times \mathbb{R}^n_+)^r$.
- Doesn't work because of an additional degree of freedom — $\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i$ may be scaled by non-zero positive scalars that product to 1,

$$\alpha\mathbf{x} \otimes \beta\mathbf{y} \otimes \gamma\mathbf{z} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}, \qquad \alpha\beta\gamma = 1,$$

e.g. $(n\mathbf{x}) \otimes \mathbf{y} \otimes (\mathbf{z}/n)$ can have a diverging loading vector even while the outer-product remains fixed.

Picking the right objective function

- Define $f : \mathbb{R}^r \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^n)^r \to \mathbb{R}$ by

$$f(X) := \left\| \mathcal{A} - \sum_{p=1}^{r} \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p \right\|_F^2$$

  where $X = (\lambda_1, \dots, \lambda_r; \mathbf{u}_1, \mathbf{v}_1, \mathbf{w}_1, \dots, \mathbf{u}_r, \mathbf{v}_r, \mathbf{w}_r)$.
- Let $\mathbb{S}_+^{n-1} := \{\mathbf{x} \in \mathbb{R}_+^n \mid \|\mathbf{x}\|_2 = 1\}$ and

$$\mathscr{P} := \mathbb{R}_+^r \times (\mathbb{S}_+^{l-1} \times \mathbb{S}_+^{m-1} \times \mathbb{S}_+^{n-1})^r.$$

- Global minimizer of f on $\mathscr{P}$, $(\lambda_1, \dots, \lambda_r; \mathbf{u}_1, \mathbf{v}_1, \mathbf{w}_1, \dots, \mathbf{u}_r, \mathbf{v}_r, \mathbf{w}_r) \in \mathscr{P}$, gives required global minimizer (non-unique).

Level sets are compact

- Note $\mathscr{P}$ is closed but unbounded.
- Will show that the level set of f restricted to $\mathscr{P}$,

$$\mathscr{E}_\alpha = \{X \in \mathscr{P} \mid f(X) \leq \alpha\}$$

  is compact for all α.
- $\mathscr{E}_\alpha = \mathscr{P} \cap f^{-1}(-\infty, \alpha]$ closed since f continuous.
- Now to show $\mathscr{E}_\alpha$ bounded.
  - Suppose not, $\{X_n\}_{n=1}^\infty \subset \mathscr{P}$ with $\|X_n\| \to \infty$ but $f(X_n) \leq \alpha$ for all n.
  - Clearly, $\|X_n\| \to \infty$ implies $\lambda_q^{(n)} \to \infty$ for at least one $q \in \{1, \dots, r\}$.
  - Note
$$f(X) \geq \left(\|\mathcal{A}\|_F - \left\|\sum\nolimits_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p\right\|_F\right)^2.$$

Using nonnegativity

- Taking $X \geq 0$ into account,

$$\left\|\sum_{p=1}^{r} \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p\right\|_F^2 = \sum_{i,j,k=1}^{l,m,n} \Bigl(\sum_{p=1}^{r} \lambda_p u_{pi} v_{pj} w_{pk}\Bigr)^2$$
$$\geq \sum_{i,j,k=1}^{l,m,n} (\lambda_q u_{qi} v_{qj} w_{qk})^2$$
$$= \lambda_q^2 \sum_{i,j,k=1}^{l,m,n} (u_{qi} v_{qj} w_{qk})^2$$
$$= \lambda_q^2 \|\mathbf{u}_q \otimes \mathbf{v}_q \otimes \mathbf{w}_q\|_F^2$$
$$= \lambda_q^2$$

  since $\|\mathbf{u}_q\| = \|\mathbf{v}_q\| = \|\mathbf{w}_q\| = 1$.
- Hence, as $\lambda_q^{(n)} \to \infty$, $f(X_n) \to \infty$.
- Contradicts $f(X_n) \leq \alpha$ for all n.

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 $\kappa_k(\mathbf{y})$.
- Pierre's lectures in Week 2.

Diagonalizing a symmetric hypermatrix

- A best symmetric rank approximation may not exist either:

Example

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

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