Optimal solutions to non-negative PARAFAC/multilinear NMF always exist

Lek-Heng Lim
Stanford University

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

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

Rasmus Bro
Chemometrics Group
Royal Veterinary and Agricultural University

Richard Harshman
Department of Psychology
University of Western Ontario
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

Gene Golub
Department of Computer Science
Stanford University
Hi Rasmus, Lek-Heng, and Lieven,

Here is an interesting exchange between Rasmus and I that happened earlier today, but now seems relevant to our latest discussion. There are questions for Rasmus as well.

Please see below about Rasmus’s lack of experiences with the "nonexistence of a solution" problem (also known as degenerate solutions) -- then an interesting apparent explanation and then an issue regarding positivity [non-negativity] constraints and degeneracies.

> -----Original Message-----
> From: Rasmus Bro (KVL) [mailto:rb@kvl.dk]
> Sent: Thursday, June 09, 2005 4:23 PM
> To: harshman@uwo.ca
> Subject: RE: hello, holiday greetings, asking about exchanges with
> Jos tenBerge
> 
> Hi Richard
> 
> Once I get degeneracies, they mostly reflect some problems
in my specification of the problem. Once that problem is solved, the degeneracy usually disappear. Do you ever enforce *positivity* [nonnegativity] on all modes, and if so when? On some modes? I actually mostly do nonnegativity on all modes because it makes physical sense. Well, I always start without them, and then I actually mostly do nonnegativity on all modes because it makes physical sense. Well, I always start without them, and then if I run into problems, I start imposing nonnegativity. But I seldom think too much about what he problems is. Sometimes it’s degeneracy, sometimes just ‘ugly-looking’ models.

[----- Richard A. Harshman:]

So doesn’t this absolutely prevent degenerate solutions from arising when you do that? It’s no wonder you don’t have a problem with degeneracies.

Perhaps Lek-Heng could work out what the characteristics of the modified problem do about the existence of an exact minimum rather than an infimum.

> Best
> Rasmus

Best regards,
--richard

\[ A = \left[ a_{j_1 \ldots j_k} \right] \in \mathbb{R}^{d_1 \times \cdots \times d_k} \text{ non-negative}, \]  
\[ \text{denoted } A \geq 0, \text{ if all } a_{j_1 \ldots j_k} \geq 0. \]

\( a_{ijk} = \) fluorescence emission intensity at wavelength \( \lambda_{j}^{\text{em}} \) of \( i \)th sample excited with light at wavelength \( \lambda_{k}^{\text{ex}} \). Get 3-way data \( A = \left[ a_{ijk} \right] \in \mathbb{R}^{l \times m \times n} \).

Decomposing \( A \) into a sum of outer products, \[ A = x_1 \otimes y_1 \otimes z_1 + \cdots + x_r \otimes y_r \otimes z_r. \]
yield the true chemical factors responsible for the data (in practice: *approximation* instead of *decomposition* because of the presence of noise).
• $r$: number of pure substances in the mixtures,

• $x_\alpha = (x_{1\alpha}, \ldots, x_{l\alpha})$: relative concentrations of $\alpha$th substance in samples $1, \ldots, l$,

• $y_\alpha = (y_{1\alpha}, \ldots, y_{m\alpha})$: emission spectrum of $\alpha$th substance,

• $z_\alpha = (z_{1\alpha}, \ldots, z_{n\alpha})$: excitation spectrum of $\alpha$th substance.

$x_\alpha, y_\alpha, z_\alpha \geq 0$ — concentration and intensity cannot be negative.
Non-negative Matrix Factorization


Central idea behind NMF (everything else is fluff): the way ‘basis functions’ combine to build ‘target objects’ is an exclusively additive process and should not involve any cancellations between the basis functions.

**NMF in a nutshell**: given non-negative matrix $A$, decompose it into a sum of outer-products of non-negative vectors:

$$A = XY^\top = \sum_{i=1}^{r} x_i \otimes y_i.$$ 

**Noisy situation**: approximate $A$ by a sum of outer-products of non-negative vectors

$$\min_{X \geq 0, Y \geq 0} \| A - XY^\top \|_F = \min_{x_i \geq 0, y_i \geq 0} \| A - \sum_{i=1}^{r} x_i \otimes y_i \|_F.$$
Generalizing NMF to Higher Order

Non-negative outer-product decomposition for $A \geq 0$ is

$$A = \sum_{p=1}^{r} x_p^1 \otimes \cdots \otimes x_p^k$$

where $x_p^i \in \mathbb{R}_{+}^{d_i} := \{ x \in \mathbb{R}^{d_i} | x \geq 0 \}$. Clear that such a decomposition exists for any $A \geq 0$.

Non-negative outer-product rank: minimal $r$ for which such a decomposition is possible.

Optimal non-negative outer-product rank-$r$ approximation:

$$\arg\min \left\{ \| A - \sum_{p=1}^{r} x_p^1 \otimes \cdots \otimes x_p^k \|_F \mid x_p^i \in \mathbb{R}_{+}^{d_i} \right\}.$$ (†)

In other words,

**Multilinear NMF = Non-negative PARAFAC**
Since a general tensor can fail to have an optimal low-rank approximation (ie. $A$ is **degenerate** or $(\dagger)$ is **ill-posed**), the first question that one should ask in a multilinear generalization of a bilinear model is whether the generalized problem still has a solution.

We will show that it does.
Degeneracy/Ill-posedness


Let $x, y, z, w$ be linearly independent. Define

$$A := x \otimes x \otimes x + x \otimes y \otimes z + y \otimes z \otimes x + y \otimes w \otimes z + z \otimes x \otimes y + z \otimes y \otimes w$$

and, for $\varepsilon > 0$,

$$B_\varepsilon := (y + \varepsilon x) \otimes (y + \varepsilon w) \otimes \varepsilon^{-1}z + (z + \varepsilon x) \otimes \varepsilon^{-1}x \otimes (x + \varepsilon y)$$

$$- \varepsilon^{-1}y \otimes y \otimes (x + z + \varepsilon w) - \varepsilon^{-1}z \otimes (x + y + \varepsilon z) \otimes x$$

$$+ \varepsilon^{-1}(y + z) \otimes (y + \varepsilon z) \otimes (x + \varepsilon w).$$

Then $\text{rank}_{\otimes}(B_\varepsilon) \leq 5$, $\text{rank}_{\otimes}(A) = 6$ and $\|B_\varepsilon - A\| \to 0$ as $\varepsilon \to 0$.

$A$ has no optimal approximation by tensors of rank $\leq 5$. 


Existence of Global Minimizer

**Theorem.** Let $A = \begin{bmatrix} a_{j_1 \ldots j_k} \end{bmatrix} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be non-negative. Then

$$\inf \left\{ \left\| A - \sum_{p=1}^{r} x_p^1 \otimes \cdots \otimes x_p^k \right\|_F \mid x_p^i \in \mathbb{R}^{d_i}_+ \right\}$$

is attained.

**Idea of proof:** If a continuous real-valued function has a non-empty compact level set, then it has to attain its infimum. We will show that for a suitably redefined non-negative PARAFAC objective function, *all* its level sets are compact.
Naive choice of objective: \( g : (\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k})^r \to \mathbb{R}, \)
\[
g(x^1_1, \ldots, x^k_1, \ldots, x^1_r, \ldots, x^k_r) := \| A - \sum_{p=1}^r x^1_p \otimes \cdots \otimes x^k_p \|_F^2.
\]

Need to show \( g \) attains infimum on \( (\mathbb{R}^{d_1}_+ \times \cdots \times \mathbb{R}^{d_k}_+)^r \).

 Doesn’t work because of an additional degree of freedom — \( x^1, \ldots, x^k \) may be scaled by non-zero positive scalars that product to 1,
\[
\alpha_1 x^1 \otimes \cdots \otimes \alpha_k x^k = x^1 \otimes \cdots \otimes x^k, \quad \alpha_1 \cdots \alpha_k = 1,
\]
e.g. \((nx) \otimes y \otimes (z/n)\) can have a diverging loading vector even while the outer-product remains fixed.
Modified Objective Function

Define $f : \mathbb{R}^r \times (\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k})^r \to \mathbb{R}$ by

$$f(X) := \|A - \sum_{p=1}^r \lambda_p u_p^1 \otimes \cdots \otimes u_p^k\|_F^2$$

where $X = (\lambda_1, \ldots, \lambda_r; u_1^1, \ldots, u_1^k, \ldots, u_r^1, \ldots, u_r^k)$. Let

$$D := \mathbb{R}_+^r \times (S_{d_1}^{d_1-1} \times \cdots \times S_{d_k}^{d_k-1})^r,$$

$$S_{n-1}^n := \{x \in \mathbb{R}_+^n \mid \|x\| = 1\}.$$

Global minimizer of $f$ on $D$, $(\lambda_1, \ldots, \lambda_r; u_1^1, \ldots, u_1^k, \ldots, u_r^1, \ldots, u_r^k) \in D$, gives required global minimizer (albeit in a non-unique way), e.g. may take $x_1^i = \lambda_i u_1^i$ and $x_p^i = u_p^i$ for $p \geq 2$. 
Compactness of Level Sets

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

$$E_\alpha = \{X \in \mathcal{D} \mid f(X) \leq \alpha\}$$

is compact for all $\alpha$. $E_\alpha = \mathcal{D} \cap f^{-1}(-\infty, \alpha]$ closed since $f$ is polynomial, thus continuous.

Now to show $E_\alpha$ bounded. Suppose there is a sequence $\{X_n\}_{n=1}^\infty \subset \mathcal{D}$ with $\|X_n\| \to \infty$ but $f(X_n) \leq \alpha$ for all $n$. Clearly, $\|X_n\| \to \infty$ implies that $\lambda_q^{(n)} \to \infty$ for at least one $q \in \{1, \ldots, r\}$.

By Cauchy-Schwartz,

$$f(X) \geq (\|A\|_F - \|\sum_{p=1}^r \lambda_p u_p^1 \otimes \cdots \otimes u_p^k\|_F)^2.$$
Taking $X \geq 0$ into account,

$$\left\| \sum_{p=1}^{r} \lambda_p u_p^1 \otimes \cdots \otimes u_p^k \right\|_F^2 = \sum_{i_1, \ldots, i_k = 1}^{d_1, \ldots, d_k} \left( \sum_{p=1}^{r} \lambda_p u_{pi_1}^1 \cdots u_{pi_k}^k \right)^2$$

$$\geq \sum_{i_1, \ldots, i_k = 1}^{d_1, \ldots, d_k} \left( \lambda_q u_{qi_1}^1 \cdots u_{qi_k}^k \right)^2$$

$$= \lambda_q^2 \sum_{i_1, \ldots, i_k = 1}^{d_1, \ldots, d_k} (u_{qi_1}^1 \cdots u_{qi_k}^k)^2$$

$$= \lambda_q^2 \left\| u_q^1 \otimes \cdots \otimes u_q^k \right\|_F^2$$

$$= \lambda_q^2$$

since $\|u_q^1\| = \cdots = \|u_q^k\| = 1$.

Hence, as $\lambda_q^{(n)} \to \infty$, $f(X_n) \to \infty$. This contradicts $f(X_n) \leq \alpha$ for all $n$. 
Uniqueness?

Uniqueness with conditions on Kruskal dimension (hard to check):


Uniqueness with conditions on simplicial cones (bilinear only):