Tensors for chemists and psychologists†

Lek-Heng Lim
Institute for Computational and Mathematical Engineering
Stanford University

6th ERCIM Workshop on Matrix Computations and Statistics
Copenhagen, Denmark
April 1–3, 2005

†: with apologies to Rasmus Bro and Richard Harshman

Acknowledgement

Vin de Silva
Department of Mathematics
Stanford University

Rasmus Bro
Chemometrics Group
Royal Veterinary and Agricultural University

Richard Harshman
Department of Psychology
University of Western Ontario

Overview

- definition
- ranks
- decompositions
- norms and inner products
- approximations
- hyperdeterminants
- covariance and contravariance
- contraction products
- multilinear functions
- eigenvalues and singular values
- technical stuff best left to the end

What is a Matrix

Question: What makes a matrix a matrix as opposed to merely a 2-array of numbers?

Answer: The algebraic operations of matrix addition, scalar multiplication, and, most importantly, matrix multiplication:

1. $A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{m \times n}$, $\lambda \in \mathbb{R}$,

   $$A + B := [a_{ij} + b_{ij}] \qquad \text{and} \qquad \lambda A := [\lambda a_{ij}].$$

2. $A = [a_{ij}] \in \mathbb{R}^{l \times m}, B = [b_{jk}] \in \mathbb{R}^{m \times n}$, $AB := [c_{ik}] \in \mathbb{R}^{l \times n}$ where

   $$c_{ik} := \sum_{j=1}^{n} a_{ij} b_{jk}.$$

We are so used to seeing these operations performed on 2-arrays of numbers that we sometimes forget that they are defined by us and not something that comes automatically with a 2-array.

Tensors in a nutshell

A matrix is an order-2 tensor.

An order-k tensor is simply a k-array of numbers with natural generalizations of the aforementioned algebraic operations.

Caution: What physicists and geometers call tensors are really tensor fields (ie. tensor-valued functions on manifolds). E.g. stress tensor, moment-of-intertia tensor, Einstein tensor, metric tensor, curvature tensor, Ricci tensor, etc.

Two-sided matrix multiplication

Before coming to that, observe that matrix multiplication is a special case of a more general algebraic operation: a matrix may be simultaneously multiplied on both sides by two matrices.

Given $A = [a_{jk}] \in \mathbb{R}^{m\times n}$, $L_1 = [\ell^1_{ij}] \in \mathbb{R}^{r\times m}$ and $L_2 = [\ell^2_{lk}] \in \mathbb{R}^{s\times n}$:

$$L_1 A L_2^t = C$$

where $C = [c_{il}] \in \mathbb{R}^{r\times s}$ has entries

$$c_{il} = \sum_{j=1}^{m} \sum_{k=1}^{n} \ell^1_{ij} \ell^2_{lk} a_{jk}.$$

The result is independent of the order we perform the left and right matrix multiplications, ie. $L_1(AL_2^t) = (L_1 A)L_2^t$ — a property known as associativity.

Matrix-matrix multiplications (ie. AB, BA), matrix-vector multiplications (ie. $A\mathbf{x}$, $\mathbf{y}^t A\mathbf{x}$) are all special cases of this.

Order-3 Tensors

A tensor of *order* 3 is a 3-way array $A = [\![a_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$ on which the following algebraic operations are defined:

1. Addition/Scalar Multiplication: for $[\![b_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$, $\lambda \in \mathbb{R}$,

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

2. Multilinear Matrix Multiplication: for matrices $L = [\lambda_{i'i}] \in \mathbb{R}^{p \times l}, M = [\mu_{j'j}] \in \mathbb{R}^{q \times m}, N = [\nu_{k'k}] \in \mathbb{R}^{r \times n}$,

   $$(L, M, N)A := [\![c_{i'j'k'}]\!] \in \mathbb{R}^{p \times q \times r}$$

   where

   $$c_{i'j'k'} := \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \lambda_{i'i}\mu_{j'j}\nu_{k'k}a_{ijk}.$$

Tensors

A tensor of order k and size $(d_1, \ldots, d_k)$ is a k-array of real numbers with two properties:

1. Addition/Scalar Multiplication: We may add two arrays of the same size or multiply an array by a scalar.
2. Multilinear Matrix Multiplication: We may multiply an array in each 'mode' by matrices.

An order k-array of size $(d_1, \ldots, d_k)$ is denoted by $[\![a_{j_1 \cdots j_k}]\!]_{j_1,\ldots,j_k=1}^{d_1,\ldots,d_k}$, where the entries $a_{j_1 \cdots j_k}$ are understood to be real numbers.

Usually, we just write $[\![a_{j_1 \cdots j_k}]\!]$.

The set of all k-arrays of size $(d_1, \ldots, d_k)$ is denoted by by $\mathbb{R}^{d_1 \times \cdots \times d_k}$.

Quick example on the notation

For $k = 2$, we have $[\![a_{ij}]\!]_{i,j=1}^{m,n} \in \mathbb{R}^{m\times n}$. For example,

$$[\![a_{ij}]\!]_{i,j=1}^{3,2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} 3.511 & -100.231 \\ 34.435 & 0.000 \\ -46.566 & 23.278 \end{bmatrix} \in \mathbb{R}^{3\times 2}.$$

For $k = 3$, we have $[\![a_{ijk}]\!]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l\times m\times n}$. For example,

$$[\![a_{ijk}]\!]_{i,j,k=1}^{3,4,2} = \left[\begin{array}{cccc|cccc} a_{111} & a_{121} & a_{131} & a_{141} & a_{112} & a_{122} & a_{132} & a_{142} \\ a_{211} & a_{221} & a_{231} & a_{241} & a_{212} & a_{222} & a_{232} & a_{242} \\ a_{311} & a_{321} & a_{331} & a_{341} & a_{312} & a_{322} & a_{332} & a_{342} \end{array}\right]$$

$$= \left[\begin{array}{rrrr|rrrr} 3.5 & -1.2 & 3.1 & -1.1 & 3.1 & -1.1 & -1.5 & -0.2 \\ 3.4 & 0.0 & 4.4 & 0.1 & 4.5 & 0.3 & -4.5 & 7.2 \\ 6.5 & -0.2 & -4.6 & 0.8 & 4.6 & 0.7 & -6.6 & 1.2 \end{array}\right] \in \mathbb{R}^{3\times 4\times 2}.$$

The above array should be viewed as a 3-array where the left slab $[a_{ij1}|$ is laying on top of the right slab $|a_{ij2}]$.

Property 1: Vector space structure

Addition/Scalar Multiplication: We may add two arrays of the same size or multiply an array by a scalar — by performing the operations coordinatewise, ie.

$$A + B := [\![a_{j_1 \ldots j_k} + b_{j_1 \ldots j_k}]\!],$$

$$\lambda A := [\![\lambda a_{j_1 \ldots j_k}]\!],$$

for $A = [\![a_{j_1 \ldots j_k}]\!], B = [\![b_{j_1 \ldots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, $\lambda \in \mathbb{R}$.

Property 1 says that $\mathbb{R}^{d_1 \times \cdots \times d_k}$ is a vector space of dimension $d_1 \cdots d_k$.

Property 2: Multilinear structure

Multilinear Matrix Multiplication: We may multiply an array in each 'mode' by matrices.

For an order-k tensor $A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ and matrices

$$L_1 = [\ell^1_{i_1 j_1}] \in \mathbb{R}^{r_1 \times d_1}, \quad \ldots \quad , L_k = [\ell^k_{i_k j_k}] \in \mathbb{R}^{r_k \times d_k},$$

the multiplication is written as $(L_1, \ldots, L_k)A$ and is defined by

$$(L_1, \ldots, L_k)A = C$$

where $C = [\![c_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{r_1 \times \cdots \times r_k}$ has entries

$$c_{i_1 \cdots i_k} = \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} \ell^1_{i_1 j_1} \cdots \ell^k_{i_k j_k} a_{j_1 \cdots j_k}.$$

Property 2 distinguishes $\mathbb{R}^{d_1 \times \cdots \times d_k}$ from being simply a vector space of dimension $d_1 \cdots d_k$. It is the reason why, for instance, $\mathbb{R}^{l \times m \times n}$ (order-3 tensors) is different from $\mathbb{R}^{lm \times n}$ (matrices) or $\mathbb{R}^{lmn}$ (vectors).

Examples: Orders 2 and 3

Example. For $A \in \mathbb{R}^{m \times n}$, $(L_1, L_2)A$ is just left and right multiplication by matrices:

$$(L_1, L_2)A = L_1 A L_2^t = L_1(A L_2^t) = (L_1 A) L_2^t.$$

This is equivalent to multiplying every column vector of A by L_1 and then every row vector of the result by L_2. These operations can be done in any order. We may multiply every row of A by L_2 first and then multiply every column of the result by L_1.

Example. For $A \in \mathbb{R}^{l \times m \times n}$, $(L_1, L_2, L_3)A$ is equivalent to multiplying every horizontal slabs of A by L_1, every lateral slabs of the result by L_2, and then every frontal slabs of the result by L_3:

$$B \leftarrow [L_1 A_{1 \bullet \bullet} \mid \cdots \mid L_1 A_{p \bullet \bullet}]; \tag{S-1}$$

$$C \leftarrow [L_2 B_{\bullet 1 \bullet} \mid \cdots \mid L_2 B_{\bullet q \bullet}]; \tag{S-2}$$

$$(L_1, L_2, L_3)A \leftarrow [L_3 C_{\bullet \bullet 1} \mid \cdots \mid L_3 C_{\bullet \bullet r}]; \tag{S-3}$$

As before, (S-1), (S-2), (S-3) may be performed in any order.

Outer product

The outer product of k vectors, $\mathbf{x}^1 = (x_1^1, \ldots, x_{d_1}^1)^t \in \mathbb{R}^{d_1}, \ldots, \mathbf{x}^k = (x_1^k, \ldots, x_{d_k}^k)^t \in \mathbb{R}^{d_k}$ is an order-k tensor of size $(d_1, \ldots, d_k)$:

$$\mathbf{x}^1 \otimes \cdots \otimes \mathbf{x}^k := [\![x_{i_1}^1 \ldots x_{i_k}^k]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}.$$

The outer product of k vector spaces, $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_k}$, is simply

$$\mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_k} := \operatorname{span}_{\mathbb{R}}\{\mathbf{x}^1 \otimes \cdots \otimes \mathbf{x}^k \mid \mathbf{x}^1 \in \mathbb{R}^{d_1}, \ldots, \mathbf{x}^k \in \mathbb{R}^{d_k}\}.$$

By definition, $\mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_k}$ is a subspace of the vector space $\mathbb{R}^{d_1 \times \cdots \times d_k}$. Counting dimensions, we see immediately that

$$\mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_k} = \mathbb{R}^{d_1 \times \cdots \times d_k}.$$

This leads to an alternative definition of tensors.

Property 2': Outer product structure

The fact that $\mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_k} = \mathbb{R}^{d_1 \times \cdots \times d_k}$ tells us that every $A \in \mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_k}$ may be written as

$$A = \sum_{\alpha=1}^{r} \mathbf{x}_\alpha^1 \otimes \cdots \otimes \mathbf{x}_\alpha^k$$

for some $\mathbf{x}_\alpha^j \in \mathbb{R}^{d_j}$ ($\alpha = 1, \ldots, r$; $j = 1, \ldots, k$).

This is exactly what gives a tensor its multilinear structure. Given $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$,

$$(L_1, \ldots, L_k)A = \sum_{\alpha=1}^{r} L_1\mathbf{x}_\alpha^1 \otimes \cdots \otimes L_k\mathbf{x}_\alpha^k.$$

So the multilinear structure (Property 2) and outer product structure (Property 2') are one and the same thing. We could have instead defined a tensor as one that satisfies Properties 1 and 2' — a k-array that can be decomposed into a sum of outer products of k vectors.

Matrix rank

$A \in \mathbb{R}^{m \times n}$. rank($A$) may be defined in either one of the three (among other) ways:

- outer product rank: rank(A) $= r$ iff there exists $\mathbf{x}_1, \dots, \mathbf{x}_r \in \mathbb{R}^m$, $\mathbf{y}_1, \dots, \mathbf{y}_r \in \mathbb{R}^n$ such that

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

  and r is minimal over all such decompositions.

- row rank: rank(A) $= r$ iff

  $$\dim(\operatorname{span}_{\mathbb{R}}\{A_{1\bullet}, \dots, A_{m\bullet}\}) = r$$

  where $A_{i\bullet} \in \mathbb{R}^n$ denotes the ith row vector of A.

- column rank: rank(A) $= r$ iff

  $$\dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet 1}, \dots, A_{\bullet n}\}) = r$$

  where $A_{\bullet j} \in \mathbb{R}^m$ denotes the jth column vector of A.

Order-3 tensor rank

For an order-3 tensor $A \in \mathbb{R}^{l \times m \times n}$, we have

- outer product rank: $\operatorname{rank}_{\otimes}(A) = r$ iff there exists $\mathbf{x}_1, \dots, \mathbf{x}_r \in \mathbb{R}^l$, $\mathbf{y}_1, \dots, \mathbf{y}_r \in \mathbb{R}^m$, $\mathbf{y}_1, \dots, \mathbf{y}_r \in \mathbb{R}^n$ such that

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

  and r is minimal over all such decompositions.

- 1-slab rank: $\operatorname{rank}_1(A) = r_1$ iff

$$\dim(\operatorname{span}_{\mathbb{R}}\{A_{1\bullet\bullet}, \dots, A_{l\bullet\bullet}\}) = r_1$$

  where $A_{i\bullet\bullet} \in \mathbb{R}^{m \times n}$ denotes the ith 1-slab of A.

- 2-slab rank: $\operatorname{rank}_2(A) = r_2$ iff

$$\dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet 1\bullet}, \dots, A_{\bullet m\bullet}\}) = r_2$$

  where $A_{\bullet j\bullet} \in \mathbb{R}^{l \times n}$ denotes the jth 2-slab of A.

- 3-slab rank: $\text{rank}_3(A) = r_3$ iff

$$\dim(\text{span}_{\mathbb{R}}\{A_{\bullet\bullet 1}, \ldots, A_{\bullet\bullet n}\}) = r_3$$

  where $A_{\bullet\bullet k} \in \mathbb{R}^{l \times m}$ denotes the kth 3-slab of A.

- trilinear rank: $\text{rank}_{\boxplus}(A) = (r_1, r_2, r_3)$.

Note: In general, $\text{rank}_1(A) \neq \text{rank}_2(A) \neq \text{rank}_3(A) \neq \text{rank}_{\otimes}(A)$.

Tensor Rank

$A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. Different notions of tensor ranks:

- outer product rank: $\operatorname{rank}_{\otimes}(A) = r$ iff there exists $\mathbf{x}_i^j \in \mathbb{R}^{d_j}$, $j = 1, \ldots, k$, such that

$$A = \sum_{i=1}^{r} \mathbf{x}_i^1 \otimes \cdots \otimes \mathbf{x}_i^k$$

  and r is minimal over all such decompositions.

- multilinear rank of A is defined as

$$\operatorname{rank}_{\boxplus}(A) = (\operatorname{rank}_1(A), \ldots, \operatorname{rank}_k(A))$$

- p-slab rank ($p = 1, \ldots, k$): $\operatorname{rank}_p(A) = r_p$ iff

$$\dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet\cdots\bullet 1 \bullet\cdots\bullet}, \ldots, A_{\bullet\cdots\bullet d_p \bullet\cdots\bullet}\}) = r_p$$

  where $A_{\bullet\cdots\bullet i \bullet\cdots\bullet} \in \mathbb{R}^{d_1 \times \cdots \times \widehat{d_p} \times \cdots \times d_k}$ denotes the ith p-slab of A, an order-$(k-1)$ tensor.

Why no bilinear rank

When $k = 2$, then 1-slab = row, 2-slab = column, bilinear rank of a matrix $A \in \mathbb{R}^{m \times n}$ is simply

$$\text{rank}_{\boxplus}(A) = (\text{rowrank}(A), \text{colrank}(A)) = (\text{rank}(A), \text{rank}(A)).$$

When $k \geq 3$, $\text{rank}_p(A) \neq \text{rank}_q(A) \neq \text{rank}_{\otimes}(A)$ in general (for $p \neq q$).

Outer product decomposition

Let $A \in \mathbb{R}^{l\times m\times n}$ and $\operatorname{rank}_{\otimes}(A) = r$. The outer product or Candecomp/Parafac decomposition of A is

$$A = \sum_{\alpha=1}^{r} \mathbf{x}_\alpha \otimes \mathbf{y}_\alpha \otimes \mathbf{z}_\alpha.$$

In other words,

$$a_{ijk} = \sum_{\alpha=1}^{r} x_{i\alpha} y_{j\alpha} z_{k\alpha}$$

for some $\mathbf{x}_\alpha = (x_{1\alpha}, \dots, x_{l\alpha})^t \in \mathbb{R}^l$, $\mathbf{y}_\alpha = (y_{1\alpha}, \dots, y_{m\alpha})^t \in \mathbb{R}^m$, $\mathbf{z}_\alpha = (z_{1\alpha}, \dots, z_{n\alpha})^t \in \mathbb{R}^n$, $\alpha = 1, \dots, r$.

The vectors $\mathbf{x}_\alpha, \mathbf{y}_\alpha, \mathbf{z}_\alpha$ are sometimes regarded as column vectors of matrices $X = [\mathbf{x}_1, \dots, \mathbf{x}_r] \in \mathbb{R}^{l\times r}$, $Y = [\mathbf{y}_1, \dots, \mathbf{y}_r] \in \mathbb{R}^{m\times r}$, $Z = [\mathbf{z}_1, \dots, \mathbf{z}_r] \in \mathbb{R}^{n\times r}$.

Multilinear decomposition

Let $A \in \mathbb{R}^{l\times m\times n}$ and $\operatorname{rank}_{\boxplus}(A) = (r_1, r_2, r_3)$. Multilinear or Tucker decomposition of A is

$$A = (X, Y, Z)C.$$

In other words,

$$a_{ijk} = \sum_{\alpha=1}^{r_1} \sum_{\beta=1}^{r_2} \sum_{\gamma=1}^{r_3} x_{i\alpha} y_{j\beta} z_{k\gamma} c_{\alpha\beta\gamma}$$

for some full-rank matrices $X = [x_{i\alpha}] \in \mathbb{R}^{l\times r_1}$, $Y = [y_{j\beta}] \in \mathbb{R}^{m\times r_2}$, $Z = [z_{k\gamma}] \in \mathbb{R}^{n\times r_3}$, and core tensor $C = [\![c_{\alpha\beta\gamma}]\!] \in \mathbb{R}^{r_1\times r_2\times r_3}$.

X, Y, Z may be chosen to have orthonormal columns.

For matrices, this is just the $L_1 D L_2^t$ or $Q_1 R Q_2^t$ decompositions.

Norms and inner products

In order to discuss approximations, we need to define a norm on $\mathbb{R}^{d_1 \times \cdots \times d_k}$.

The most convenient one to use is the Frobenius norm, $\| \cdot \|_F$, defined by

$$\| [\![a_{j_1 \ldots j_k}]\!] \|_F^2 = \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} a_{j_1 \cdots j_k}^2.$$

for $[\![a_{j_1 \ldots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$.

It is the norm associated with the trace inner product, $\langle \cdot, \cdot \rangle_{\mathrm{tr}}$, defined by

$$\langle [\![a_{j_1 \ldots j_k}]\!] \mid [\![b_{j_1 \ldots j_k}]\!] \rangle_{\mathrm{tr}} := \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} a_{j_1 \ldots j_k} b_{j_1 \ldots j_k}$$

for $[\![a_{j_1 \ldots j_k}]\!], [\![b_{j_1 \ldots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. Thus $\|A\|_F^2 = \langle A \mid A \rangle_{\mathrm{tr}}$.

Outer product approximation

A Candecomp/Parafac or outer product model has the following form

$$a_{ijk} = \sum_{\alpha=1}^{r} x_{i\alpha} y_{j\alpha} z_{k\alpha} + e_{ijk}$$

where $E = [\![e_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$ denotes the (unknown) error.

To minimize the error, we want an outer product approximation

$$\operatorname{argmin} \|A - \sum_{\alpha=1}^{r} \mathbf{x}_\alpha \otimes \mathbf{y}_\alpha \otimes \mathbf{z}_\alpha\|_F$$

where the minimum is taken over all matrices $X = [\mathbf{x}_1, \dots, \mathbf{x}_r] \in \mathbb{R}^{l \times r}$, $Y = [\mathbf{y}_1, \dots, \mathbf{y}_r] \in \mathbb{R}^{m \times r}$, $Z = [\mathbf{z}_1, \dots, \mathbf{z}_r] \in \mathbb{R}^{n \times r}$.

In short, we want an optimal solution $B_\otimes^* = \operatorname{argmin}_{\operatorname{rank}_\otimes(B) \leq r} \|A - B\|_F$.

Multilinear approximation

A Tucker or multilinear model has the following form

$$a_{ijk} = \sum_{\alpha=1}^{r_1} \sum_{\beta=1}^{r_2} \sum_{\gamma=1}^{r_3} x_{i\alpha} y_{j\beta} z_{k\gamma} c_{\alpha\beta\gamma} + e_{ijk}$$

where $E = \llbracket e_{ijk} \rrbracket \in \mathbb{R}^{l \times m \times n}$ denotes the (unknown) error.

To minimize the error, we want a multilinear approximation

$$\text{argmin} \| A - (X, Y, Z) C \|_F$$

where minimum is taken over all full-rank matrices $X \in \mathbb{R}^{l \times r_1}$, $Y \in \mathbb{R}^{m \times r_2}$, $Z \in \mathbb{R}^{n \times r_3}$ and tensor $C \in \mathbb{R}^{r_1 \times r_2 \times r_3}$.

In short, we want an optimal solution

$$B_{\boxplus}^* = \text{argmin}_{\text{rank}_{\boxplus}(B) \le (r_1, r_2, r_3)} \| A - B \|_F.$$

Outer product decomposition: analytical chemistry

Application to fluorescence spectral analysis by Bro.

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

Decomposing A into a sum of outer products,

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

yield the true chemical factors responsible for the data.

- r: number of pure substances in the mixtures,
- $\mathbf{x}_\alpha = (x_{1\alpha}, \dots, x_{l\alpha})$: relative concentrations of αth substance in samples $1, \dots, l$,
- $\mathbf{y}_\alpha = (y_{1\alpha}, \dots, y_{m\alpha})$: excitation spectrum of αth substance,
- $\mathbf{z}_\alpha = (z_{1\alpha}, \dots, z_{n\alpha})$: emission spectrum of αth substance.

Multilinear decomposition: computer vision

Application to facial recognition (TensorFaces) by Vasilescu and Terzopoulos. Facial image database of p male subjects photographed in q poses, r illuminations, s expressions, and stored as a grayscale image with t pixels.

a_{ijklm} = grayscale level of mth pixel of the image of ith person photographed in jth pose, with lth expression, under kth illumination level. Get 5-way data array $A = [\![a_{ijklm}]\!] \in \mathbb{R}^{p\times q\times r\times s\times t}$.

Let multilinear decomposition of A be

$$A = (V, W, X, Y, Z)C,$$

matrices V, W, X, Y, Z chosen to have orthonormal columns.

The column vectors of V, W, X, Y, Z are the 'principal components' or 'parameterizing factors' of the spaces of male subjects, poses, illuminations, expressions, and images respectively. The tensor C governs the interactions between these factors.

Properties of matrix rank

1. Rank of $A \in \mathbb{R}^{m \times n}$ easy to determine (Gaussian Elimination)
2. Optimal rank-r approximation to $A \in \mathbb{R}^{m \times n}$ always exist (Eckart-Young Theorem)
3. Optimal rank-r approximation to $A \in \mathbb{R}^{m \times n}$ easy to find (Singular Value Decomposition)
4. Pick $A \in \mathbb{R}^{m \times n}$ at random, then A has full rank with probability 1, ie. $\text{rank}(A) = \min\{m, n\}$
5. $\text{rank}(A)$ from a non-orthogonal rank-revealing decomposition (e.g. $A = L_1 D L_2^t$) and $\text{rank}(A)$ from an orthogonal rank-revealing decomposition (e.g. $A = Q_1 R Q_2^t$) are equal
6. Let A be a matrix with real entries. Then $\text{rank}(A)$ is the same whether we regard A as an element of $\mathbb{R}^{m \times n}$ or as an element of $\mathbb{C}^{m \times n}$

Outer product rank vs multilinear rank

Every statement on the preceding slide is **false** for the outer product rank of order-k tensors, $k \geq 3$.

Every statement on the preceding slide is **true** for the multilinear rank of order-k tensors, $k \geq 3$.

In the next two slides we will spell these out explicitly for order-3 tensors. The restriction to order-3 tensors is strictly for notational simplicity. All statements generalize to order-k tensors for any $k \geq 3$.

Properties of outer product rank

1. Computing $\text{rank}_{\otimes}(A)$ for $A \in \mathbb{R}^{l\times m\times n}$ is NP-hard
2. For some $A \in \mathbb{R}^{l\times m\times n}$, $\text{argmin}_{\text{rank}_{\otimes}(B)\leq r}\|A - B\|_F$ does not have a solution
3. When $\text{argmin}_{\text{rank}_{\otimes}(B)\leq r}\|A - B\|_F$ does have a solution, computing the solution is an NP-complete problem in general
4. For some l, m, n, if we sample $A \in \mathbb{R}^{l\times m\times n}$ at random, there is no r such that $\text{rank}_{\otimes}(A) = r$ with probability 1
5. An outer product decomposition of $A \in \mathbb{R}^{l\times m\times n}$ with orthogonality constraints on X, Y, Z will in general require a sum with more than $\text{rank}_{\otimes}(A)$ number of terms
6. Let A be a 3-array with real entries. Then $\text{rank}_{\otimes}(A)$ can take different values depending on whether we regard $A \in \mathbb{R}^{l\times m\times n}$ or $A \in \mathbb{C}^{l\times m\times n}$

Properties of multilinear rank

1. Computing $\mathrm{rank}_{\boxplus}(A)$ for $A \in \mathbb{R}^{l \times m \times n}$ is easy
2. Solution to $\mathrm{argmin}_{\mathrm{rank}_{\boxplus}(B) \le (r_1, r_2, r_3)} \|A - B\|_F$ always exist
3. Solution to $\mathrm{argmin}_{\mathrm{rank}_{\boxplus}(B) \le (r_1, r_2, r_3)} \|A - B\|_F$ easy to find
4. Pick $A \in \mathbb{R}^{l \times m \times n}$ at random, then A has

$$\mathrm{rank}_{\boxplus}(A) = (\min(l, mn), \min(m, ln), \min(n, lm))$$

   with probability 1

5. If $A \in \mathbb{R}^{l \times m \times n}$ has $\mathrm{rank}_{\boxplus}(A) = (r_1, r_2, r_3)$. Then there exist full-rank matrices $X \in \mathbb{R}^{l \times r_1}$, $Y \in \mathbb{R}^{m \times r_2}$, $Z \in \mathbb{R}^{n \times r_3}$ and core tensor $C \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ such that $A = (X, Y, Z)C$. X, Y, Z may be chosen to have orthonormal columns
6. Let A be a matrix with real entries. Then $\mathrm{rank}_{\boxplus}(A)$ is the same whether we regard A as an element of $\mathbb{R}^{l \times m \times n}$ or as an element of $\mathbb{C}^{l \times m \times n}$

Generalization to higher order

- It is straight forward to generalize all statements on the last two slides to order-k tensors for any $k \geq 3$; we give two examples:

- Statement 2 for outer product rank:

  - For some $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, $\operatorname{argmin}_{\operatorname{rank}_{\otimes}(B) \leq r} \|A - B\|_F$ does not have a solution

- Statement 4 for multilinear rank:

  - Pick $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ at random, then A has

$$\operatorname{rank}_{\boxplus}(A) = (\min(d_1, d_2 \cdots d_k), \ldots, \min(d_k, d_1 \cdots d_{k-1}))$$

    with probability 1. The p-th slab rank above is just

$$\min(d_p, d_1 \cdots \widehat{d_p} \cdots d_k)$$

What about 'row rank = column rank'

At first glance, this is one property of matrix rank that doesn't seem to generalize to multilinear rank. Actually, it does in a more subtle way. We use the order-3 case as illustration.

Let $A \in \mathbb{R}^{l\times m\times n}$. Recall that we have defined the p-slab ranks:

$$\begin{aligned}
\operatorname{rank}_1(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{i\bullet\bullet} \mid i = 1,\dots,l\}),\\
\operatorname{rank}_2(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet j\bullet} \mid j = 1,\dots,m\})\\
\operatorname{rank}_3(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet\bullet k} \mid k = 1,\dots,n\}).
\end{aligned}$$

We may also define the (p,q)-slab ranks:

$$\begin{aligned}
\operatorname{rank}_{2,3}(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet jk} \mid j = 1,\dots,m; k = 1,\dots,n\}),\\
\operatorname{rank}_{1,3}(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{i\bullet k} \mid i = 1,\dots,l; k = 1,\dots,n\}),\\
\operatorname{rank}_{1,2}(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{ij\bullet} \mid i = 1,\dots,l; j = 1,\dots,m\}).
\end{aligned}$$

It is easy to see that

$$\begin{aligned}
\operatorname{rank}_1(A) &= \operatorname{rank}_{2,3}(A),\\
\operatorname{rank}_2(A) &= \operatorname{rank}_{1,3}(A),\\
\operatorname{rank}_3(A) &= \operatorname{rank}_{1,2}(A).
\end{aligned}$$

Higher level trilinear rank

The 1st level trilinear rank for an order-3 tensor is what we simply called trilinear rank earlier:

$$\mathrm{rank}^1_{\boxplus}(A) = (\mathrm{rank}_1(A), \mathrm{rank}_2(A), \mathrm{rank}_3(A))$$

The 2nd level trilinear rank for an order-3 tensor is:

$$\mathrm{rank}^2_{\boxplus}(A) = (\mathrm{rank}_{2,3}(A), \mathrm{rank}_{1,3}(A), \mathrm{rank}_{1,2}(A)).$$

Hence the result at the end of the previous slide may be restated for $A \in \mathbb{R}^{l \times m \times n}$ as simply

$$\mathrm{rank}^1_{\boxplus}(A) = \mathrm{rank}^2_{\boxplus}(A).$$

Note that for $A \in \mathbb{R}^{m \times n} = \mathbb{R}^{1 \times m \times n}$, this reduces to

$$(1, \mathrm{rowrank}(A), \mathrm{colrank}(A)) = (1, \mathrm{colrank}(A), \mathrm{rowrank}(A)),$$

and thus $\mathrm{rowrank}(A) = \mathrm{colrank}(A)$.

Higher level multilinear rank

Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. For any $\{p_1, \ldots, p_l\} \subset \{1, \ldots, k\}$, $p_1 < \cdots < p_k$, we may define $(p_1, \ldots, p_l)$-slab rank accordingly.

The $\binom{k}{l}$-tuple of $(p_1, \ldots, p_l)$-slab ranks gives the lth level multilinear rank, for $l = 1, \ldots, k-1$.

May show: The lth level multilinear rank is equal to the $(k-l)$th level multilinear rank, $l = 1, \ldots, k-1$.

Appendix 1: Some technical properties

- Let $A, B \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ and $\lambda, \mu \in \mathbb{R}$. Let $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$. Then

$$(L_1, \ldots, L_k)(\lambda A + \mu B) = \lambda (L_1, \ldots, L_k) A + \mu (L_1, \ldots, L_k) B.$$

- Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. Let $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$, and $M_1 \in \mathbb{R}^{s_1 \times r_1}, \ldots, M_k \in \mathbb{R}^{s_k \times r_k}$. Then

$$(M_1, \ldots, M_k)(L_1, \ldots, L_k) A = (M_1 L_1, \ldots, M_k L_k) A$$

where $M_i L_i \in \mathbb{R}^{s_i \times d_i}$ is simply the matrix-matrix product of M_i and L_i.

- Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ and $\lambda, \mu \in \mathbb{R}$. Let $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_j, M_j \in \mathbb{R}^{r_j \times d_j}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$. Then

$$(L_1, \ldots, \lambda L_j + \mu M_j, \ldots, L_k) A = \lambda (L_1, \ldots, L_j, \ldots, L_k) A + \mu (L_1, \ldots, M_j, \ldots, L_k) A.$$

Appendix 2: NP problems

- NP is the set of problems for which a proposed solution can be verified or rejected in polynomial time
- A problem is NP-hard if an algorithm to solve it in polynomial time would make it possible to solve all NP problems in polynomial time
- NP-complete is the class of problems which are both NP-hard and themselves members of NP
- NP-hard problems are at least as hard as (possibly harder than) any other NP (and thus NP-complete) problems
- The bottom line is that NP-hard and NP (including NP-complete) problems are difficult to solve — no known polynomial-time algorithm exists for finding the solution