Numerical Multilinear Algebra I

Lek-Heng Lim

University of California, Berkeley

January 5–7, 2009
Past 50 years, Numerical Linear Algebra played indispensable role in

- the statistical analysis of two-way data,
- the numerical solution of partial differential equations arising from vector fields,
- the numerical solution of second-order optimization methods.

Next step — development of Numerical Multilinear Algebra for

- the statistical analysis of multi-way data,
- the numerical solution of partial differential equations arising from tensor fields,
- the numerical solution of higher-order optimization methods.
**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 $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $d$ can be expressed as

$$f(x) = a_0 + a_1^T x + x^T A_2 x + A_3(x, x, x) + \cdots + A_d(x, \ldots, x).$$

$$a_0 \in \mathbb{R}, a_1 \in \mathbb{R}^n, A_2 \in \mathbb{R}^{n \times n}, A_3 \in \mathbb{R}^{n \times n \times n}, \ldots, A_d \in \mathbb{R}^{n \times \cdots \times n}.$$

- Numerical linear algebra: $d = 2$.
Motivation

Why multilinear:
- “Classification of mathematical problems as linear and nonlinear is like classification of the Universe as bananas and non-bananas.”

Why numerical:
- Different from Computer Algebra.
- Numerical rather than symbolic: floating point operations — cheap and abundant; symbolic operations — expensive.
- Like other areas in numerical analysis, will entail the approximate solution of approximate multilinear problems with approximate data but under controllable and rigorous confidence bounds on the errors involved.
Tensors: mathematician’s definition

- $U, V, W$ vector spaces. Think of $U \otimes V \otimes W$ as the vector space of all formal linear combinations of terms of the form $u \otimes v \otimes w$,

$$\sum \alpha u \otimes v \otimes w,$$

where $\alpha \in \mathbb{R}, u \in U, v \in V, w \in W$.

- One condition: $\otimes$ decreed to have the multilinear property

$$(\alpha u_1 + \beta u_2) \otimes v \otimes w = \alpha u_1 \otimes v \otimes w + \beta u_2 \otimes v \otimes w,$$

$$u \otimes (\alpha v_1 + \beta v_2) \otimes w = \alpha u \otimes v_1 \otimes w + \beta u \otimes v_2 \otimes w,$$

$$u \otimes v \otimes (\alpha w_1 + \beta w_2) = \alpha u \otimes v \otimes w_1 + \beta u \otimes v \otimes w_2.$$

- Up to a choice of bases on $U, V, W$, $A \in U \otimes V \otimes W$ can be represented by a $3$-hypermatrix $A = [a_{ijk}]^{l,m,n}_{i,j,k=1} \in \mathbb{R}^{l \times m \times n}$. 
Tensors: physicist’s definition

- “What are tensors?” ≡ “What kind of physical quantities can be represented by tensors?”
- Usual answer: if they satisfy some ‘transformation rules’ under a change-of-coordinates.

**Theorem (Change-of-basis)**

Two representations $A, A'$ of $A$ in different bases are related by

$$(L, M, N) \cdot A = A'$$

with $L, M, N$ respective change-of-basis matrices (non-singular).

- Pitfall: tensor fields (roughly, tensor-valued functions on manifolds) often referred to as tensors — stress tensor, piezoelectric tensor, moment-of-inertia tensor, gravitational field tensor, metric tensor, curvature tensor.
Tensors: data analyst’s definition

- **Data structure:** $k$-array $A = \left[ a_{ijk} \right]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l \times m \times n}$

- **Algebraic structure:**
  1. **Addition/scalar multiplication:** for $[b_{ijk}] \in \mathbb{R}^{l \times m \times n}$, $\lambda \in \mathbb{R}$,
     $$[a_{ijk}] + [b_{ijk}] := [a_{ijk} + b_{ijk}] \quad \text{and} \quad \lambda [a_{ijk}] := [\lambda a_{ijk}] \in \mathbb{R}^{l \times m \times n}$$
  2. **Multilinear matrix multiplication:** for matrices $L = \left[ \lambda_{i'i} \right] \in \mathbb{R}^{p \times l}$, $M = \left[ \mu_{j'j} \right] \in \mathbb{R}^{q \times m}$, $N = \left[ \nu_{k'k} \right] \in \mathbb{R}^{r \times n}$,
     $$(L, M, N) \cdot A := \left[ c_{i'j'k'} \right] \in \mathbb{R}^{p \times q \times r}$$
     where
     $$c_{i'j'k'} := \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \lambda_{i'i} \mu_{j'j} \nu_{k'k} a_{ijk}.$$ 

Think of $A$ as 3-dimensional **hypermatrix**. $(L, M, N) \cdot A$ as multiplication on ‘3 sides’ by matrices $L, M, N$.

Generalizes to arbitrary order $k$. If $k = 2$, ie. matrix, then $(M, N) \cdot A = MAN^T$. 

---

L.-H. Lim (ICM Lecture)  
Numerical Multilinear Algebra I  
January 5–7, 2009  7 / 55
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 $a = [a_1, \ldots, a_n]^\top \in \mathbb{R}^n$.

- 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 $A = [a_{ijk}]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l \times m \times n}$.

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

- \( a \in \mathbb{R}^n \) 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 \( V^* \times W^* \rightarrow \mathbb{R} \) (contravariant), a bilinear form \( V \times W \rightarrow \mathbb{R} \) (covariant), or a linear operator \( V \rightarrow W \) (mixed).

- \( A \in \mathbb{R}^{l \times m \times n} \) can represent trilinear form \( U \times V \times W \rightarrow \mathbb{R} \) (covariant), bilinear operators \( V \times W \rightarrow U \) (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

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

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

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

  where

  $$c_{\alpha\beta\gamma} = \sum_{i,j,k=1}^{l,m,n} x_{\alpha i} y_{\beta j} z_{\gamma k} a_{ijk}.$$
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 \( x \in \mathbb{R}^l, y \in \mathbb{R}^m, z \in \mathbb{R}^n, \)

\[ \mathcal{A}(x, y, z) := \mathcal{A} \cdot (x, y, z) = \sum_{i,j,k=1}^{l,m,n} a_{ijk} x_i y_j z_k, \]

\[ \mathcal{A}(l, y, z) := \mathcal{A} \cdot (l, y, z) = \sum_{j,k=1}^{m,n} a_{ijk} y_j z_k. \]
Segre outer product

If $U = \mathbb{R}^l$, $V = \mathbb{R}^m$, $W = \mathbb{R}^n$, $\mathbb{R}^l \otimes \mathbb{R}^m \otimes \mathbb{R}^n$ may be identified with $\mathbb{R}^{l \times m \times n}$ if we define $\otimes$ by

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

A tensor $A \in \mathbb{R}^{l \times m \times n}$ is said to be decomposable if it can be written in the form

$$A = u \otimes v \otimes w$$

for some $u \in \mathbb{R}^l$, $v \in \mathbb{R}^m$, $w \in \mathbb{R}^n$.

The set of all decomposable tensors 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:

$$\text{Seg}(\mathbb{R}^l, \mathbb{R}^m, \mathbb{R}^n) = \{ A \in \mathbb{R}^{l \times m \times n} \mid a_{i_1 i_2 i_3} a_{j_1 j_2 j_3} = a_{k_1 k_2 k_3} a_{l_1 l_2 l_3}, \{i_\alpha, j_\alpha\} = \{k_\alpha, l_\alpha\} \}.$$
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 \(\sigma \in \mathcal{S}_k\) on indices.
- \(S^k(\mathbb{R}^n)\) denotes set of all order-\(k\) symmetric hypermatrices.

Example

Higher order derivatives of multivariate functions.

Example

Moments of a random vector \(x = (X_1, \ldots, X_n)\):

\[
m_k(x) = \left[ E(x_{i_1}x_{i_2} \cdots x_{i_k}) \right]_{i_1, \ldots, i_k=1}^n = \left[ \int \cdots \int x_{i_1}x_{i_2} \cdots x_{i_k} \ d\mu(x_{i_1}) \cdots d\mu(x_{i_k}) \right]_{i_1, \ldots, i_k=1}^n.
\]
Symmetric hypermatrices

Example

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

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

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

- Important in Independent Component Analysis (ICA).
Inner products and norms

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

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

- Norm topology often more directly relevant to engineering applications than Zariski topology.
Other norms

- Let $\| \cdot \|_{\alpha_i}$ be a norm on $\mathbb{R}^{d_i}$, $i = 1, \ldots, k$. Then **operator norm** of multilinear functional $A : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} \to \mathbb{R}$ is

\[
\| A \|_{\alpha_1, \ldots, \alpha_k} := \sup \frac{|A(x_1, \ldots, x_k)|}{\| x_1 \|_{\alpha_1} \cdots \| x_k \|_{\alpha_k}}.
\]

- Deep and important results about such norms in functional analysis.
- **$E$-norm** and **$G$-norm**:

\[
\| A \|_E = \sum_{i_1, \ldots, i_k = 1}^{d_1, \ldots, d_k} \left| a_{j_1 \ldots j_k} \right|
\]

and

\[
\| A \|_G = \max \left\{ \left| a_{j_1 \ldots j_k} \right| \mid j_1 = 1, \ldots, d_1; \ldots; j_k = 1, \ldots, d_k \right\}.
\]

- Multiplicative on rank-1 tensors:

\[
\| u \otimes v \otimes \cdots \otimes z \|_E = \| u \|_1 \| v \|_1 \cdots \| z \|_1,
\]

\[
\| u \otimes v \otimes \cdots \otimes z \|_F = \| u \|_2 \| v \|_2 \cdots \| z \|_2,
\]

\[
\| u \otimes v \otimes \cdots \otimes z \|_G = \| u \|_\infty \| v \|_\infty \cdots \| z \|_\infty.
\]
Tensor ranks (Hitchcock, 1927)

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

  \[
  \text{rank}(A) = \dim(\text{span}_\mathbb{R}\{A_{\cdot 1}, \ldots, A_{\cdot n}\}) \quad \text{(column rank)}
  \]
  \[
  = \dim(\text{span}_\mathbb{R}\{A_{1\cdot}, \ldots, A_{m\cdot}\}) \quad \text{(row rank)}
  \]
  \[
  = \min\{r \mid A = \sum_{i=1}^{r} u_i v_i^\top\} \quad \text{(outer product rank)}.
  \]

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

  \[
  r_1(A) = \dim(\text{span}_\mathbb{R}\{A_{1\cdot\cdot}, \ldots, A_{l\cdot\cdot}\})
  \]
  \[
  r_2(A) = \dim(\text{span}_\mathbb{R}\{A_{\cdot 1\cdot}, \ldots, A_{\cdot m\cdot}\})
  \]
  \[
  r_3(A) = \dim(\text{span}_\mathbb{R}\{A_{\cdot\cdot 1}, \ldots, A_{\cdot\cdot n}\})
  \]

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

  \[
  \text{rank}_\otimes(A) = \min\{r \mid A = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i\}
  \]

  where \(u \otimes v \otimes w := \left[ u_i v_j w_k \right]_{i,j,k=1}^{l,m,n}\).
Properties of matrix rank

1. Rank of \( A \in \mathbb{R}^{m \times n} \) easy to determine (Gaussian elimination)

2. Best rank-\( r \) approximation to \( A \in \mathbb{R}^{m \times n} \) always exist (Eckart-Young theorem)

3. Best 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. \( A = L_1 DL_2^T \)) and \( \text{rank}(A) \) from an orthogonal rank-revealing decomposition (e.g. \( A = Q_1 RQ_2^T \)) are equal

6. \( \text{rank}(A) \) is base field independent, ie. same value whether we regard \( A \) as an element of \( \mathbb{R}^{m \times n} \) or as an element of \( \mathbb{C}^{m \times n} \).
Properties of outer product rank

1. Computing $\text{rank}_\otimes(A)$ for $A \in \mathbb{R}^{l \times m \times n}$ is NP-hard [Håstad 1990]

2. For some $A \in \mathbb{R}^{l \times m \times n}$, $\arg\min_{\text{rank}_\otimes(B) \leq r} \|A - B\|_F$ does not have a solution

3. When $\arg\min_{\text{rank}_\otimes(B) \leq r} \|A - B\|_F$ 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. $\text{rank}_\otimes(A)$ is base field dependent, ie. value depends 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 $\text{rank}_{\boxplus}(A)$ for $A \in \mathbb{R}^{l \times m \times n}$ is easy

2. Solution to $\arg\min_{\text{rank}_{\boxplus}(B) \leq (r_1, r_2, r_3)} \|A - B\|_F$ always exist

3. Solution to $\arg\min_{\text{rank}_{\boxplus}(B) \leq (r_1, r_2, r_3)} \|A - B\|_F$ easy to find

4. Pick $A \in \mathbb{R}^{l \times m \times n}$ at random, then $A$ has

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

with probability 1

5. If $A \in \mathbb{R}^{l \times m \times n}$ has $\text{rank}_{\boxplus}(A) = (r_1, r_2, r_3)$. 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 $A = (X, Y, Z) \cdot C$. $X, Y, Z$ may be chosen to have orthonormal columns

6. $\text{rank}_{\boxplus}(A)$ is base field independent, ie. same value whether we regard $A \in \mathbb{R}^{l \times m \times n}$ or $A \in \mathbb{C}^{l \times m \times n}$
Algebraic computational complexity

- For $A = (a_{ij}), B = (b_{jk}) \in \mathbb{R}^{n \times n}$,

$$AB = \sum_{i,j,k=1}^{n} a_{ik} b_{kj} E_{ij} = \sum_{i,j,k=1}^{n} \varphi_{ik}(A) \varphi_{kj}(B) E_{ij}$$

where $E_{ij} = e_i e_j^\top \in \mathbb{R}^{n \times n}$. Let

$$T = \sum_{i,j,k=1}^{n} \varphi_{ik} \otimes \varphi_{kj} \otimes E_{ij}.$$ 

- $O(n^{2+\varepsilon})$ algorithm for multiplying two $n \times n$ matrices gives $O(n^{2+\varepsilon})$ algorithm for solving system of $n$ linear equations [Strassen 1969].
- **Conjecture.** $\log_2(\text{rank}_\otimes(T)) \leq 2 + \varepsilon$.
- **Best known result.** $O(n^{2.376})$ [Coppersmith-Winograd 1987; Cohn-Kleinberg-Szegedy-Umans 2005].
More tensor ranks

- For $u \in \mathbb{R}^l$, $v \in \mathbb{R}^m$, $w \in \mathbb{R}^n$,

  $$u \otimes v \otimes w := [u_i v_j w_k]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l \times m \times n}.$$

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

  $$\text{rank}_\otimes(A) = \min\{r \mid A = \sum_{i=1}^{r} \sigma_i u_i \otimes v_i \otimes w_i, \quad \sigma_i \in \mathbb{R}\}.$$

- **Symmetric outer product rank.** $A \in S^k(\mathbb{R}^n),$

  $$\text{rank}_S(A) = \min\{r \mid A = \sum_{i=1}^{r} \lambda_i v_i \otimes v_i \otimes v_i, \quad \lambda_i \in \mathbb{R}\}.$$

- **Nonnegative outer product rank.** $A \in \mathbb{R}_+^{l \times m \times n},$

  $$\text{rank}_+(A) = \min\{r \mid A = \sum_{i=1}^{r} \delta_i x_i \otimes y_i \otimes z_i, \quad \delta_i \in \mathbb{R}_+\}.$$
SVD, EVD, NMF of a matrix

- **Singular value decomposition** of $A \in \mathbb{R}^{m \times n}$,

$$A = U \Sigma V^\top = \sum_{i=1}^{r} \sigma_i u_i \otimes v_i$$

where $\text{rank}(A) = r$, $U \in O(m)$ left singular vectors, $V \in O(n)$ right singular vectors, $\Sigma$ singular values.

- **Symmetric eigenvalue decomposition** of $A \in S^2(\mathbb{R}^n)$,

$$A = V \Lambda V^\top = \sum_{i=1}^{r} \lambda_i v_i \otimes v_i,$$

where $\text{rank}(A) = r$, $V \in O(n)$ eigenvectors, $\Lambda$ eigenvalues.

- **Nonnegative matrix factorization** of $A \in \mathbb{R}_{+}^{n \times n}$,

$$A = X \Delta Y^\top = \sum_{i=1}^{r} \delta_i x_i \otimes y_i$$

where $\text{rank}_+(A) = r$, $X, Y \in \mathbb{R}_{+}^{m \times r}$ unit column vectors (in the 1-norm), $\Delta$ positive values.
SVD, EVD, NMF of a hypermatrix

- **Outer product decomposition** of $A \in \mathbb{R}^{l \times m \times n}$,

$$A = \sum_{i=1}^{r} \sigma_i u_i \otimes v_i \otimes w_i$$

where $\text{rank}_\otimes(A) = r$, $u_i \in \mathbb{R}^l$, $v_i \in \mathbb{R}^m$, $w_i \in \mathbb{R}^n$ unit vectors, $\sigma_i \in \mathbb{R}$.

- **Symmetric outer product decomposition** of $A \in S^3(\mathbb{R}^n)$,

$$A = \sum_{i=1}^{r} \lambda_i v_i \otimes v_i \otimes v_i$$

where $\text{rank}_S(A) = r$, $v_i$ unit vector, $\lambda_i \in \mathbb{R}$.

- **Nonnegative outer product decomposition** for hypermatrix $A \in \mathbb{R}_+^{l \times m \times n}$ is

$$A = \sum_{i=1}^{r} \delta_i x_i \otimes y_i \otimes z_i$$

where $\text{rank}_+(A) = r$, $x_i \in \mathbb{R}_+^l$, $y_i \in \mathbb{R}_+^m$, $z_i \in \mathbb{R}_+^n$ unit vectors, $\delta_i \in \mathbb{R}_+$. 
Best low rank approximation of a matrix

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

\[
\arg\min_{\text{rank}(B) \leq r} \| A - B \|.
\]

- More precisely, find \( \sigma_i, u_i, v_i, i = 1, \ldots, r \), that minimizes

\[
\| A - \sigma_1 u_1 \otimes v_1 - \sigma_2 u_2 \otimes v_2 - \cdots - \sigma_r u_r \otimes v_r \|.
\]

**Theorem (Eckart–Young)**

Let \( A = U \Sigma V^\top = \sum_{i=1}^{\text{rank}(A)} \sigma_i u_i v_i^\top \) be singular value decomposition. For \( r \leq \text{rank}(A) \), let

\[
A_r := \sum_{i=1}^{r} \sigma_i u_i v_i^\top.
\]

Then

\[
\| A - A_r \|_F = \min_{\text{rank}(B) \leq r} \| A - B \|_F.
\]

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

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

- Hypermatrices that have rank \( > 1 \) are elements on the higher secant varieties of \( \mathcal{I} = \text{Seg}(\mathbb{R}^l, \mathbb{R}^m, \mathbb{R}^n) \).
- E.g. a hypermatrix has rank 2 if it sits on a secant line through two points in \( \mathcal{I} \) but not on \( \mathcal{I} \), rank 3 if it sits on a secant plane through three points in \( \mathcal{I} \) but not on any secant lines, etc.
- Minor technicality: should really be secant quasiprojective variety.
Scientific data mining

- **Spectroscopy**: measure light absorption/emission of specimen as function of energy.
- Typical **specimen** contains $10^{13}$ to $10^{16}$ light absorbing entities or **chromophores** (molecules, amino acids, etc).

**Fact (Beer’s Law)**

$$A(\lambda) = -\log(I_1/I_0) = \varepsilon(\lambda)c. \quad A = \text{absorbance}, \quad I_1/I_0 = \text{fraction of intensity of light of wavelength } \lambda \text{ that passes through specimen}, \quad c = \text{concentration of chromophores}.$$  

- Multiple chromophores ($f = 1, \ldots, r$) and wavelengths ($i = 1, \ldots, m$) and specimens/experimental conditions ($j = 1, \ldots, n$),

  $$A(\lambda_i, s_j) = \sum_{f=1}^{r} \varepsilon_f(\lambda_i)c_f(s_j).$$

- **Bilinear model aka factor analysis**: $A_{m \times n} = E_{m \times r}C_{r \times n}$

  rank-revealing factorization or, in the presence of noise, low-rank approximation $\min \|A_{m \times n} - E_{m \times r}C_{r \times n}\|$.
Modern data mining

- **Text mining** is the spectroscopy of documents.
- Specimens = documents.
- Chromophores = terms.
- Absorbance = inverse document frequency:
  \[
  A(t_i) = -\log \left( \sum_j \chi(f_{ij}) / n \right).
  \]
- Concentration = term frequency: \(f_{ij}\).
- \(\sum_j \chi(f_{ij}) / n\) = fraction of documents containing \(t_i\).
- \(A \in \mathbb{R}^{m \times n}\) term-document matrix. \(A = QR = U\Sigma V^T\) rank-revealing factorizations.
- Bilinear model aka vector space model.
- Due to Gerald Salton and colleagues: SMART (system for the mechanical analysis and retrieval of text).
Bilinear models

- Bilinear models work on ‘two-way’ data:
  - measurements on object \(i\) (genomes, chemical samples, images, webpages, consumers, etc) yield a vector \(a_i \in \mathbb{R}^n\) where \(n =\) number of features of \(i\);
  - collection of \(m\) such objects, \(A = [a_1, \ldots, a_m]\) may be regarded as an \(m\)-by-\(n\) matrix, e.g. gene \(\times\) microarray matrices in bioinformatics, terms \(\times\) documents matrices in text mining, facial images \(\times\) individuals matrices in computer vision.

- Various matrix techniques may be applied to extract useful information: QR, EVD, SVD, NMF, CUR, compressed sensing techniques, etc.

- Examples: vector space model, factor analysis, principal component analysis, latent semantic indexing, PageRank, EigenFaces.

- Some problems: **factor indeterminacy** — \(A = XY\) rank-revealing factorization not unique; unnatural for \(k\)-way data when \(k > 2\).
Ubiquity of multiway data

- **Batch data**: batch × time × variable
- **Time-series analysis**: time × variable × lag
- **Computer vision**: people × view × illumination × expression × pixel
- **Bioinformatics**: gene × microarray × oxidative stress
- **Phylogenetics**: codon × codon × codon
- **Analytical chemistry**: sample × elution time × wavelength
- **Atmospheric science**: location × variable × time × observation
- **Psychometrics**: individual × variable × time
- **Sensory analysis**: sample × attribute × judge
- **Marketing**: product × product × consumer

**Fact (Inevitable consequence of technological advancement)**

*Increasingly sophisticated instruments, sensor devices, data collecting and experimental methodologies lead to increasingly complex data.*

---

L.-H. Lim  (ICM Lecture)  
Numerical Multilinear Algebra I  
January 5–7, 2009  
30 / 55
Fundamental problem of multiway data analysis

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

Example

Given \( A \in \mathbb{R}^{d_1 \times d_2 \times d_3} \), find \( u_i, v_i, w_i, \ i = 1, \ldots, r \), that minimizes

\[
\| A - u_1 \otimes v_1 \otimes w_1 - u_2 \otimes v_2 \otimes w_2 - \cdots - u_r \otimes v_r \otimes z_r \|
\]

or \( C \in \mathbb{R}^{r_1 \times r_2 \times r_3} \) 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

\[
\| A - (U, V, W) \cdot C \|. 
\]
Fundamental problem of multiway data analysis

Example

Given $\mathcal{A} \in S^k(\mathbb{C}^n)$, find $u_i$, $i = 1, \ldots, r$, that minimizes

$$\|\mathcal{A} - u_1 \otimes^k - u_2 \otimes^k - \cdots - u_r \otimes^k\|$$

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

$$\|\mathcal{A} - (U, U, U) \cdot C\|.$$
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 $\lambda_j^{em}$ of $i$th sample excited with light at wavelength $\lambda_k^{ex}$.
  - Get 3-way data $A = [a_{ijk}] \in \mathbb{R}^{l \times m \times n}$.
  - Get outer product decomposition of $A$
    $$A = x_1 \otimes y_1 \otimes z_1 + \cdots + x_r \otimes y_r \otimes z_r.$$  
- Get the true chemical factors responsible for the data.
  - $r$: number of pure substances in the mixtures,
  - $x_\alpha = (x_{1\alpha}, \ldots, x_{l\alpha})$: relative concentrations of $\alpha$th substance in specimens 1, $\ldots$, $l$,
  - $y_\alpha = (y_{1\alpha}, \ldots, y_{m\alpha})$: excitation spectrum of $\alpha$th substance,
  - $z_\alpha = (z_{1\alpha}, \ldots, z_{n\alpha})$: emission spectrum of $\alpha$th substance.
- Noisy case: find best rank-$r$ approximation (CANDECOMP/PARAFAC).
Uniqueness of tensor decompositions

- $M \in \mathbb{R}^{m \times n}$, $\text{spark}(M) =$ size of minimal linearly dependent subset of column vectors [Donoho, Elad; 2003].

**Theorem (Kruskal)**

Let $X = [x_1, \ldots, x_r]$, $Y = [y_1, \ldots, y_r]$, $Z = [z_1, \ldots, z_r]$. Decomposition is unique up to scaling if

$$\text{spark}(X) + \text{spark}(Y) + \text{spark}(Z) \geq 2r + 5.$$

- May be generalized to arbitrary order [Sidiroupoulos, Bro; 2000].
- Avoids factor indeterminacy under mild conditions.
Multilinear decomposition in bioinformatics

- Application to cell cycle studies [Omberg, Golub, Alter; 2008].
- Collection of gene-by-microarray matrices $A_1, \ldots, A_l \in \mathbb{R}^{m \times n}$ obtained under varying oxidative stress.
  - $a_{ijk} =$ expression level of $j$th gene in $k$th microarray under $i$th stress.
  - Get 3-way data array $A = [a_{ijk}] \in \mathbb{R}^{l \times m \times n}$.
  - Get multilinear decomposition of $A$
    $$A = (X, Y, Z) \cdot C,$$
    to get orthogonal matrices $X, Y, Z$ and core tensor $C$ by applying SVD to various 'flattenings' of $A$.
- Column vectors of $X, Y, Z$ are ‘principal components’ or ‘parameterizing factors’ of the spaces of stress, genes, and microarrays; $C$ governs interactions between these factors.
- Noisy case: approximate by discarding small $c_{ijk}$ (Tucker Model).
**Code of life is a 3-tensor**

- **Codons**: triplets of nucleotides, \((i, j, k)\) where \(i, j, k \in \{A, C, G, U\}\).
- **Genetic code**: these \(4^3 = 64\) codons encode the 20 amino acids.

<table>
<thead>
<tr>
<th>First letter</th>
<th>Second letter</th>
<th>Third letter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td><strong>C</strong></td>
<td><strong>A</strong></td>
</tr>
<tr>
<td>UUU</td>
<td>UCU</td>
<td>UAU</td>
</tr>
<tr>
<td>UUC</td>
<td>UCC</td>
<td>UAC</td>
</tr>
<tr>
<td>UUA</td>
<td>UCA</td>
<td>UAA</td>
</tr>
<tr>
<td>UUG</td>
<td>UCG</td>
<td>UAG</td>
</tr>
<tr>
<td></td>
<td><strong>Phe</strong></td>
<td><strong>Tyr</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Leu</strong></td>
<td><strong>Stop</strong></td>
</tr>
<tr>
<td><strong>C</strong></td>
<td><strong>Leu</strong></td>
<td><strong>Gln</strong></td>
</tr>
<tr>
<td>CUU</td>
<td>CCU</td>
<td>CAU</td>
</tr>
<tr>
<td>CUC</td>
<td>CCC</td>
<td>CAC</td>
</tr>
<tr>
<td>CUA</td>
<td>CCA</td>
<td>CAA</td>
</tr>
<tr>
<td>CUG</td>
<td>CCG</td>
<td>CAG</td>
</tr>
<tr>
<td></td>
<td><strong>Leu</strong></td>
<td><strong>His</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Pro</strong></td>
<td><strong>Gln</strong></td>
</tr>
<tr>
<td><strong>A</strong></td>
<td><strong>Ile</strong></td>
<td><strong>Thr</strong></td>
</tr>
<tr>
<td>AUU</td>
<td>ACU</td>
<td>AAU</td>
</tr>
<tr>
<td>AUC</td>
<td>ACC</td>
<td>AAC</td>
</tr>
<tr>
<td>AUA</td>
<td>ACA</td>
<td>AAA</td>
</tr>
<tr>
<td>AUG</td>
<td>ACG</td>
<td>AAG</td>
</tr>
<tr>
<td></td>
<td><strong>Ile</strong></td>
<td><strong>Asn</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Met</strong></td>
<td><strong>Lys</strong></td>
</tr>
<tr>
<td><strong>G</strong></td>
<td><strong>Val</strong></td>
<td><strong>Asp</strong></td>
</tr>
<tr>
<td>GUU</td>
<td>GCU</td>
<td>GAU</td>
</tr>
<tr>
<td>GUC</td>
<td>GCC</td>
<td>GAC</td>
</tr>
<tr>
<td>GUA</td>
<td>GCA</td>
<td>GAA</td>
</tr>
<tr>
<td>GUG</td>
<td>GCG</td>
<td>GAG</td>
</tr>
<tr>
<td></td>
<td><strong>Val</strong></td>
<td><strong>Gly</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Ala</strong></td>
<td><strong>Glu</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Glu</strong></td>
<td><strong>Gly</strong></td>
</tr>
</tbody>
</table>

L.-H. Lim (ICM Lecture)
Numerical Multilinear Algebra I
January 5–7, 2009 36 / 55
Tensors in algebraic statistical biology

Problem (Salmon conjecture)

Find the polynomial equations that defines the set

\[ \{ P \in \mathbb{C}^{4 \times 4 \times 4} \mid \text{rank}_\otimes(P) \leq 4 \}. \]

- Why interested? Here \( P = [p_{ijk}] \) is understood to mean ‘complexified’ probability density values with \( i, j, k \in \{A, C, G, T\} \) and we want to study tensors that are of the form

\[
P = \rho_A \otimes \sigma_A \otimes \theta_A + \rho_C \otimes \sigma_C \otimes \theta_C + \rho_G \otimes \sigma_G \otimes \theta_G + \rho_T \otimes \sigma_T \otimes \theta_T,
\]

in other words,

\[
p_{ijk} = \rho_{Ai} \sigma_{Aj} \theta_{Ak} + \rho_{Ci} \sigma_{Cj} \theta_{Ck} + \rho_{Gi} \sigma_{Gj} \theta_{Gk} + \rho_{Ti} \sigma_{Tj} \theta_{Tk}.
\]

- Why over \( \mathbb{C} \)? Easier to deal with mathematically.
- Ultimately, want to study this over \( \mathbb{R}_+ \).