Tensor approximations

and why are they of interest to engineers

Lek-Heng Lim

MSRI Summer Graduate Workshop

July 7–18, 2008

Synopsis

- **Week 1**
  - Mon: *Tensor approximations* (**LH**)
  - Tue: *Notions of tensor ranks: rank, border rank, multilinear rank, nonnegative rank* (**Vin**)
  - Wed: *Conditioning, computations, applications* (**LH**)
  - Thu: *Constructibility of the set of tensors of a given rank* (**Vin**)
  - Fri: *Hyperdeterminants and optimal approximability* (**Vin**)
- **Week 2**
  - Mon: *Uniqueness of tensor decompositions, direct sum conjecture* (**Vin**)
  - Tue: *Nonnegative hypermatrices, symmetric tensors* (**LH**)
  - Wed: *Linear mixtures of random variables, cumulants, and tensors* (**Pierre**)
  - Thu: *Independent component analysis of invertible mixtures* (**Pierre**)
  - Fri: *Independent component analysis of underdetermined mixtures* (**Pierre**)

Hypermatrices

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

- Vector or n-tuple

$$f : [n] \to \mathbb{R}.$$

  If $f(i) = a_i$, then f is represented by $\mathbf{a} = [a_1, \dots, a_n]^\top \in \mathbb{R}^n$.

- Matrix

$$f : [m] \times [n] \to \mathbb{R}.$$

  If $f(i, j) = a_{ij}$, then f is represented by $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{R}^{m \times n}$.

- Hypermatrix (order 3)

$$f : [l] \times [m] \times [n] \to \mathbb{R}.$$

  If $f(i, j, k) = a_{ijk}$, then f is represented by $\mathcal{A} = [\![a_{ijk}]\!]_{i,j,k=1}^{l,m,n} \in \mathbb{R}^{l \times m \times n}$.

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

Hypermatrices and tensors

Up to choice of bases

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

A hypermatrix is the same as a tensor if

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

Basic operation on a hypermatrix

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

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

where

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

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

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

where

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

Basic operation on a hypermatrix

- Covariant version:

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

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

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

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

Symmetric hypermatrices

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

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

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

Example

Higher order derivatives of multivariate functions.

Example

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

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

Symmetric hypermatrices

Example

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

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

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

- Important in Independent Component Analysis (ICA).
- Pierre's lectures in Week 2.

Inner products and norms

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

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

- Frobenius norm

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

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

DARPA mathematical challenge eight

One of the twenty three mathematical challenges announced at DARPA Tech 2007.

Problem

Beyond convex optimization: *can linear algebra be replaced by algebraic geometry in a systematic way?*

- **Algebraic geometry in a slogan:** polynomials are to algebraic geometry what matrices are to linear algebra.
- Polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ of degree d can be expressed as

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

$a_0 \in \mathbb{R}, \mathbf{a}_1 \in \mathbb{R}^n, A_2 \in \mathbb{R}^{n \times n}, \mathcal{A}_3 \in \mathbb{R}^{n \times n \times n}, \dots, \mathcal{A}_d \in \mathbb{R}^{n \times \cdots \times n}$.

- Numerical linear algebra: $d = 2$.
- Numerical multilinear algebra: $d > 2$.

Tensor ranks (Hitchcock, 1927)

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

$$\begin{aligned}\operatorname{rank}(A) &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet 1},\dots,A_{\bullet n}\}) && \text{(column rank)}\\ &= \dim(\operatorname{span}_{\mathbb{R}}\{A_{1\bullet},\dots,A_{m\bullet}\}) && \text{(row rank)}\\ &= \min\{r \mid A = \textstyle\sum_{i=1}^r \mathbf{u}_i\mathbf{v}_i^\top\} && \text{(outer product rank).}\end{aligned}$$

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

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

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

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

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

Eigenvalue and singular value decompositions of a matrix

- Swiss Army knife of engineering applications.
- **Symmetric eigenvalue decomposition** of $A \in \mathsf{S}^2(\mathbb{R}^n)$,

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

  where $\operatorname{rank}(A) = r$, $V \in \mathsf{O}(n)$ eigenvectors, Λ eigenvalues.
- **Singular value decomposition** of $A \in \mathbb{R}^{m \times n}$,

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i \tag{1}$$

  where $\operatorname{rank}(\mathcal{A}) = r$, $U \in \mathsf{O}(m)$ left singular vectors, $V \in \mathsf{O}(n)$ right singular vectors, Σ singular values.
- Rank-revealing decompositions.

Eigenvalue and singular value decompositions

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

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i \tag{2}$$

  where $\operatorname{rank}_{\mathsf{S}}(A) = \min\{r \mid \mathcal{A} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i\} = r$.
  - LH's lecture in Week 2, Pierre's lectures in Week 2.
- **Singular value decomposition** of $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$,

$$\mathcal{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i \tag{3}$$

  where $\operatorname{rank}_{\otimes}(\mathcal{A}) = r$.
  - Vin's lecture on Tue.
- (2) used in applications of ICA to signal processing; (3) used in applications of the PARAFAC model to analytical chemistry.

Eigenvalue and singular value decompositions

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

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

  where $\operatorname{rank}_{\boxplus}(A) = (r, r, r)$, $U \in \mathbb{R}^{n \times r}$ has orthonormal columns and $\mathcal{C} \in \mathsf{S}^3(\mathbb{R}^r)$.
    - Pierre's lectures in Week 2.
- **Singular value decomposition** of $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$,

$$\mathcal{A} = (U, V, W) \cdot \mathcal{C} \tag{5}$$

  where $\operatorname{rank}_{\boxplus}(A) = (r_1, r_2, r_3)$, $U \in \mathbb{R}^{l \times r_1}$, $V \in \mathbb{R}^{m \times r_2}$, $W \in \mathbb{R}^{n \times r_3}$ have orthonormal columns and $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$.
    - Vin's lecture on Tue.

Optimal approximation

Best r-term approximation

$$f \approx \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_r f_r.$$

- $f \in \mathcal{H}$ vector space, cone, etc.
- $f_1, \dots, f_r \in \mathscr{D} \subset \mathcal{H}$ dictionary.
- $\alpha_1, \dots, \alpha_r \in \mathbb{R}$ or $\mathbb{C}$ (linear), $\mathbb{R}_+$ (convex), $\mathbb{R} \cup \{-\infty\}$ (tropical).
- $\approx$ some measure of nearness.

Dictionaries

- Number base: $\mathscr{D} = \{10^n \mid n \in \mathbb{Z}\} \subseteq \mathbb{R}$,

$$\frac{22}{7} = 3 \cdot 10^0 + 1 \cdot 10^{-1} + 4 \cdot 10^{-2} + 2 \cdot 10^{-3} + \cdots$$

- Spanning set: $\mathscr{D} = \left\{ \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] \right\} \subseteq \mathbb{R}^2$,

$$\left[\begin{smallmatrix} 2 \\ -3 \end{smallmatrix}\right] = 3\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] - 1\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right].$$

- Taylor: $\mathscr{D} = \{x^n \mid n \in \mathbb{N} \cup \{0\}\}$,

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots$$

- Fourier: $\mathscr{D} = \{\cos(nx), \sin(nx) \mid n \in \mathbb{Z}\} \subseteq L^2(-\pi, \pi)$,

$$\frac{1}{2}x = \sin(x) - \frac{1}{2}\sin(2x) + \frac{1}{3}\sin(3x) - \cdots$$

- $\mathscr{D}$ orthonormal basis, Riesz basis, frames, or just a dense spanning set.

More dictionaries

- Paley-Wiener: $\mathscr{D} = \{\operatorname{sinc}(x - n) \mid n \in \mathbb{Z}\} \subseteq H^2(\mathbb{R})$.
- Gabor: $\mathscr{D} = \{e^{i\alpha nx} e^{-(x-m\beta)^2/2} \mid (m, n) \in \mathbb{Z} \times \mathbb{Z}\} \subseteq L^2(\mathbb{R})$.
- Wavelet: $\mathscr{D} = \{2^{n/2}\psi(2^n x - m) \mid (m, n) \in \mathbb{Z} \times \mathbb{Z}\} \subseteq L^2(\mathbb{R})$.
- Friends of wavelets: $\mathscr{D} \subseteq L^2(\mathbb{R}^2)$ beamlets, brushlets, curvelets, ridgelets, wedgelets.

Question: What about continuously varying families of functions?

- Neural networks: $\mathscr{D} = \{\sigma(\mathbf{w}^\top \mathbf{x} + w_0) \mid (w_0, \mathbf{w}) \in \mathbb{R} \times \mathbb{R}^n\}$, $\sigma : \mathbb{R} \to \mathbb{R}$ sigmoid function, eg. $\sigma(x) = [1 + \exp(-x)]^{-1}$.
- Rank-revealing decompositions:
  - Matrices: $\mathscr{D} = \{\mathbf{u}\mathbf{v}^\top \mid (\mathbf{u}, \mathbf{v}) \in \mathbb{R}^m \times \mathbb{R}^n\}$ (non-unique: LU, QR, SVD).
  - Hypermatrices: $\mathscr{D} = \{\mathcal{A} \mid \operatorname{rank}_\otimes(\mathcal{A}) \leq 1\} = \{\mathcal{A} \mid \operatorname{rank}_\boxplus(\mathcal{A}) \leq 1\}$ (unique under mild conditions).
- Structure other than rank, eg. entropy, sparsity, volume, may be used to define $\mathscr{D}$.

Decomposition approach to data analysis

- $\mathscr{D} \subset \mathcal{H}$, not contained in any hyperplane.
- Let $\mathscr{D}_2 =$ union of bisecants to $\mathscr{D}$, $\mathscr{D}_3 =$ union of trisecants to $\mathscr{D}$, $\dots$, $\mathscr{D}_r =$ union of r-secants to $\mathscr{D}$.
- Define $\mathscr{D}$-rank of $f \in \mathcal{H}$ to be $\min\{r \mid f \in \mathscr{D}_r\}$.
- If $\varphi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is some measure of 'nearness' between pairs of points (e.g. norms, Bregman divergences, etc), we want to find a best low-rank approximation to $\mathcal{A}$:

$$\operatorname{argmin}\{\varphi(f, g) \mid \mathscr{D}\text{-rank}(g) \leq r\}.$$

- In the presence of noise, approximation instead of decomposition

$$f \approx \alpha_1 \cdot f_1 + \cdots + \alpha_r \cdot f_r \in \mathscr{D}_r.$$

$f_i \in \mathscr{D}$ reveal features of the dataset f.

Examples ($\varphi(\mathcal{A}, \mathcal{B}) = \|\mathcal{A} - \mathcal{B}\|_F$)

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

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)

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

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

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

- 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 $\min \|A_{m\times n} - E_{m\times r} C_{r\times n}\|$.

Social data mining

- **Text mining** is the spectroscopy of documents.
- Specimens = **documents**.
- Chromophores = **terms**.
- Absorbance = inverse document frequency:

$$A(t_i) = -\log\left(\sum\nolimits_j \chi(f_{ij})/n\right).$$

- 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. $A = QR = U\Sigma V^{\mathsf{T}}$ 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 $\mathbf{a}_i \in \mathbb{R}^n$ where $n =$ number of features of i;
  - collection of m such objects, $A = [\mathbf{a}_1, \dots, \mathbf{a}_m]$ 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 $k > 2$.

Fundamental problem of multiway data analysis

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

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

- rank may be outer product rank, multilinear rank, symmetric rank (for symmetric hypermatrix), or nonnegative rank (nonnegative hypermatrix).

Example

Given $\mathcal{A}\in\mathbb{R}^{d_1\times d_2\times d_3}$, find $\mathbf{u}_i,\mathbf{v}_i,\mathbf{w}_i$, $i=1,\dots,r$, that minimizes

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

or $\mathcal{C}\in\mathbb{R}^{r_1\times r_2\times r_3}$ and $U\in\mathbb{R}^{d_1\times r_1}, V\in\mathbb{R}^{d_2\times r_2}, W\in\mathbb{R}^{d_3\times r_3}$, that minimizes

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

Fundamental problem of multiway data analysis

Example

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

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

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

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

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,
  - $\mathbf{x}_\alpha = (x_{1\alpha}, \dots, x_{l\alpha})$: relative concentrations of αth substance in specimens $1, \dots, l$,
  - $\mathbf{y}_\alpha = (y_{1\alpha}, \dots, y_{m\alpha})$: excitation spectrum of αth substance,
  - $\mathbf{z}_\alpha = (z_{1\alpha}, \dots, z_{n\alpha})$: emission spectrum of αth substance.
- Noisy case: find best rank-r approximation (CANDECOMP/PARAFAC).

Uniqueness of tensor decompositions

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

Theorem (Kruskal)

$X = [\mathbf{x}_1, \dots, \mathbf{x}_r]$, $Y = [\mathbf{y}_1, \dots, \mathbf{y}_r]$, $Z = [\mathbf{z}_1, \dots, \mathbf{z}_r]$. *Decomposition is unique up to scaling if*

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

- May be generalized to arbitrary order [Sidiroupoulos, Bro; 2000].
- Avoids factor indeterminacy under mild conditions.
- Vin's lecture in Week 2.

Multilinear decomposition in bioinformatics

- Application to cell cycle studies [Omberg, Golub, Alter; 2008].
- Collection of gene-by-microarray matrices $A_1, \dots, A_l \in \mathbb{R}^{m \times n}$ obtained under varying oxidative stress.
  - a_{ijk} = expression level of jth gene in kth microarray under ith stress.
  - Get 3-way data array $\mathcal{A} = [\![a_{ijk}]\!] \in \mathbb{R}^{l \times m \times n}$.
  - Get multilinear decomposition of $\mathcal{A}$

$$\mathcal{A} = (X, Y, Z) \cdot \mathcal{C},$$

  to get orthogonal matrices X, Y, Z and core tensor $\mathcal{C}$ 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; $\mathcal{C}$ governs interactions between these factors.
- Noisy case: approximate by discarding small c_{ijk} (Tucker Model).

Outer product decomposition: separation of variables

Approximation by sum or integral of separable functions

- Continuous

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

- Semi-discrete

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

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

- Discrete

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

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

Separation of variables

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

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

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

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

- Gaussian mixture models

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

f is a sum of separable functions.

Multilinear decomposition: integral kernels

Approximation by sum or integral kernels

- Continuous

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

- Semi-discrete

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

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

- Discrete

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

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