Four problems for the experts gathered here

Lek-Heng Lim

University of California, Berkeley

December 22, 2008

Problem 1

Problem (Minimal Rank-1 Matrix Subspace)

Let $A_1, \dots, A_l \in \mathbb{R}^{m \times n}$. *Find smallest* r *such that there exist rank-*1 *matrices* $\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top$ *with*

$$A_1, \dots, A_l \in \operatorname{span}\{\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top\}.$$

Problem 1

Problem (Minimal Rank-1 Matrix Subspace)

Let $A_1, \dots, A_l \in \mathbb{R}^{m \times n}$*. Find smallest* r *such that there exist rank-*1 *matrices* $\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top$ *with*

$$A_1, \dots, A_l \in \operatorname{span}\{\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top\}.$$

- Expected to be difficult: NP-complete over $\mathbb{F}_q$, NP-hard over $\mathbb{Q}$.

Problem 1

Problem (Minimal Rank-1 Matrix Subspace)

Let $A_1, \dots, A_l \in \mathbb{R}^{m\times n}$. *Find smallest* r *such that there exist rank-*1 *matrices* $\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top$ *with*

$$A_1, \dots, A_l \in \operatorname{span}\{\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top\}.$$

- Expected to be difficult: NP-complete over $\mathbb{F}_q$, NP-hard over $\mathbb{Q}$.
- Convex relaxation along the lines of compressive sensing/sparse decoding?

$$\|\cdot\|_1 \approx \|\cdot\|_0, \qquad \|\cdot\|_* \approx \operatorname{rank}.$$

Problem 1

Problem (Minimal Rank-1 Matrix Subspace)

Let $A_1, \dots, A_l \in \mathbb{R}^{m \times n}$. *Find smallest* r *such that there exist rank*-1 *matrices* $\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top$ *with*

$$A_1, \dots, A_l \in \operatorname{span}\{\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top\}.$$

- Expected to be difficult: NP-complete over $\mathbb{F}_q$, NP-hard over $\mathbb{Q}$.
- Convex relaxation along the lines of compressive sensing/sparse decoding?

$$\|\cdot\|_1 \approx \|\cdot\|_0, \qquad \|\cdot\|_* \approx \operatorname{rank}.$$

- Ky Fan/nuclear/Schatten/trace norm,

$$\|A\|_* = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_i(A).$$

Problem 1

Problem (Minimal Rank-1 Matrix Subspace)

Let $A_1, \dots, A_l \in \mathbb{R}^{m\times n}$. Find smallest r such that there exist rank-1 matrices $\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top$ with

$$A_1, \dots, A_l \in \operatorname{span}\{\mathbf{u}_1\mathbf{v}_1^\top, \dots, \mathbf{u}_r\mathbf{v}_r^\top\}.$$

- Expected to be difficult: NP-complete over $\mathbb{F}_q$, NP-hard over $\mathbb{Q}$.
- Convex relaxation along the lines of compressive sensing/sparse decoding?

$$\|\cdot\|_1 \approx \|\cdot\|_0, \qquad \|\cdot\|_* \approx \operatorname{rank}.$$

- Ky Fan/nuclear/Schatten/trace norm,

$$\|A\|_* = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_i(A).$$

- [Fazel, Hindi, Boyd; 01], [Recht, Fazel, Parrilo; 09], [Candès, Recht; 09], [Ma, Goldfarb, Chen; 09].

Tensors as hypermatrices

Up to choice of bases on U, V, W, a tensor $A \in U \otimes V \otimes W$ may be represented as a hypermatrix

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

where $\dim(U) = l, \dim(V) = m, \dim(W) = n$ if

1. we give it coordinates;
2. we ignore covariance and contravariance.

Henceforth, tensor = hypermatrix.

Multilinear matrix multiplication

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

$$C = (X, Y) \cdot A = XAY^\top \in \mathbb{R}^{p\times q},$$
$$c_{\alpha\beta} = \sum\nolimits_{i,j=1}^{m,n} x_{\alpha i} y_{\beta j} a_{ij}.$$

- 3-tensors can be multiplied on three sides: $\mathcal{A} \in \mathbb{R}^{l\times m\times n}$, $X \in \mathbb{R}^{p\times l}$, $Y \in \mathbb{R}^{q\times m}$, $Z \in \mathbb{R}^{r\times n}$,

$$\mathcal{C} = (X, Y, Z) \cdot \mathcal{A} \in \mathbb{R}^{p\times q\times r},$$
$$c_{\alpha\beta\gamma} = \sum\nolimits_{i,j,k=1}^{l,m,n} x_{\alpha i} y_{\beta j} z_{\gamma k} a_{ijk}.$$

- Correspond to change-of-bases transformations for tensors.
- Define 'right' (covariant) multiplication by $(X, Y, Z) \cdot \mathcal{A} = \mathcal{A} \cdot (X^\top, Y^\top, Z^\top)$.

Tensor rank = Problem 1

- Segre **outer product** is $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} := [\![u_i v_j w_k]\!]_{i,j,k=1}^{l,m,n}$.
- A **decomposable tensor** is one that can be expressed as $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$.

Definition (Hitchcock, 1927)

Let $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$. Tensor rank is defined as

$$\operatorname{rank}_\otimes(\mathcal{A}) := \min\Bigl\{ r \Bigm| \mathcal{A} = \sum\nolimits_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i \Bigr\}.$$

- NP-complete over $\mathbb{F}_q$, NP-hard over $\mathbb{Q}$ [Håstad; 90].
- $U \otimes V \otimes W \simeq \operatorname{Hom}(U, V \otimes W)$.
- Write $\mathcal{A} = [A_1, \dots, A_l]$ where $A_1, \dots, A_l \in \mathbb{R}^{m \times n}$. Then

$$\operatorname{rank}_\otimes(\mathcal{A}) = \min\Bigl\{ r \Bigm| A_1, \dots, A_l \in \operatorname{span}\{\mathbf{u}_1 \mathbf{v}_1^\top, \dots, \mathbf{u}_r \mathbf{v}_r^\top\} \Bigr\}$$

- [Bürgisser, Clausen, Shokrollahi; 97]

Symmetric tensors

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

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

- For order p, invariant under all permutations $\sigma \in \mathfrak{S}_p$ on indices.
- $\mathsf{S}^p(\mathbb{R}^n)$ denotes set of all order-p symmetric tensors.
- Symmetric multilinear matrix multiplication $\mathcal{C} = (X, X, X) \cdot \mathcal{A}$ where

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

- $[\![a_{j_1 \cdots j_p}]\!] = [\![a_{\mathbf{j}}]\!] \in \mathsf{S}^p(\mathbb{R}^n)$ can be associated with a unique homogeneous polynomial $F \in \mathbb{R}[x_1, \dots, x_n]_p$ via

$$F(x_1, \dots, x_n) = \sum\nolimits_{\mathbf{j}} a_{\mathbf{j}} x_1^{q_1(\mathbf{j})} \cdots x_n^{q_n(\mathbf{j})}.$$

For $\mathbf{j} = (j_1, \dots, j_p)$, $q_j(\mathbf{j})$ counts number of times j appears in $\mathbf{j}$.

Symmetric tensor rank

Definition

Let $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$. Symmetric tensor rank is defined as

$$\operatorname{rank}_{\mathsf{S}}(\mathcal{A}) := \min\Bigl\{r \Bigm| \mathcal{A} = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i\Bigr\}.$$

- Similar for arbitrary order. May be viewed as **eigenvalue decomposition** for symmetric tensors.
- Equivalently, given $F(x_1, \dots, x_n) \in \mathbb{R}[x_1, \dots, x_n]_p$, want smallest r such that F is linear combination of r **powers of linear forms**,

$$F(x_1, \dots, x_n) = \sum\nolimits_{j=1}^{r} \lambda_j (\alpha_{j1} x_1 + \dots + \alpha_{jr} x_r)^p.$$

- Difficult even to determine generic value of r given (n, p) — Waring's Problem for homogeneous polynomials.
- Over $\mathbb{C}$, result is **Alexander-Hirschowitz theorem**. Over $\mathbb{R}$, very little is known. See [Comon, Golub, L, Mourrain; 08] for a survey.

Friends of Problem 1

Problem 1(a): **Effective Alexander-Hirshchowitz.** Given F, determine the $r = \operatorname{rank}_{\mathsf{S}}(F)$ linear forms explicity.

Problem 1(b): **Principal directions of *m*-ellipse.** $F =$ polynomial defining m-ellipse [Nie, Sturmfels; 09].

Problem 1(c): **Nonnegative tensor decomposition.** Problem 1 over $\mathbb{R}_+$.

As well as suitably relaxed versions of these.

Tensor ranks

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

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

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

$$\operatorname{rank}_{\otimes}(\mathcal{A}) = \min\Bigl\{ r \Bigm| \mathcal{A} = \sum\nolimits_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i, \quad \sigma_i \in \mathbb{R} \Bigr\}.$$

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

$$\operatorname{rank}_{\mathsf{S}}(\mathcal{A}) = \min\Bigl\{ r \Bigm| \mathcal{A} = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i, \quad \lambda_i \in \mathbb{R} \Bigr\}.$$

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

$$\operatorname{rank}_{+}(\mathcal{A}) = \min\Bigl\{ r \Bigm| \mathcal{A} = \sum\nolimits_{i=1}^{r} \delta_i \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i, \quad \delta_i \in \mathbb{R}_+ \Bigr\}.$$

SVD, EVD, NMF of a matrix

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

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

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

- **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 $\operatorname{rank}(A) = r$, $V \in \mathrm{O}(n)$ eigenvectors, Λ eigenvalues.

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

$$A = X\Delta Y^\top = \sum_{i=1}^{r} \delta_i \mathbf{x}_i \otimes \mathbf{y}_i$$

where $\operatorname{rank}_+(A) = r$, $X, Y \in \mathbb{R}_+^{m \times r}$ unit column vectors (in the 1-norm), Δ positive values.

SVD, EVD, NMF of a tensor

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

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

  where $\operatorname{rank}_{\otimes}(\mathcal{A}) = r$, $\mathbf{u}_i \in \mathbb{R}^l, \mathbf{v}_i \in \mathbb{R}^m, \mathbf{w}_i \in \mathbb{R}^n$ unit vectors, $\sigma_i \in \mathbb{R}$.
- **Symmetric outer product decomposition** of $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$,

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

  where $\operatorname{rank}_{\mathsf{S}}(A) = r$, $\mathbf{v}_i$ unit vector, $\lambda_i \in \mathbb{R}$.
- **Nonnegative outer product decomposition** for tensor $\mathcal{A} \in \mathbb{R}_+^{l \times m \times n}$ is

$$\mathcal{A} = \sum_{i=1}^{r} \delta_i \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i$$

  where $\operatorname{rank}_+(A) = r$, $\mathbf{x}_i \in \mathbb{R}_+^l, \mathbf{y}_i \in \mathbb{R}_+^m, \mathbf{z}_i \in \mathbb{R}_+^n$ unit vectors, $\delta_i \in \mathbb{R}_+$.

Best low rank approximation of a matrix

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

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

- 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 $r \leq \operatorname{rank}(A)$, let

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

Then

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

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

Plausible candidates for Problem 2

- Polynomial optimization problems if $\|\cdot\|$ = sum of squares:
  - Given $\mathcal{A} \in \mathbb{R}^{l\times m\times n}$, find $\sigma_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$, $i = 1, \dots, r$, that minimizes

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

  - Given $\mathcal{A} \in \mathsf{S}(\mathbb{R}^n)$, find $\lambda_i, \mathbf{v}_i$, $i = 1, \dots, r$, that minimizes

$$\|\mathcal{A} - \lambda_1\mathbf{v}_1 \otimes \mathbf{v}_1 \otimes \mathbf{v}_1 - \lambda_2\mathbf{v}_2 \otimes \mathbf{v}_2 \otimes \mathbf{v}_2 - \cdots - \lambda_r\mathbf{v}_r \otimes \mathbf{v}_r \otimes \mathbf{v}_r\|.$$

  - Given $\mathcal{A} \in \mathbb{R}_+^{l\times m\times n}$, find nonnegative $\delta_i, \mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i$, $i = 1, \dots, r$, that minimizes

$$\|\mathcal{A} - \delta_1\mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1 - \delta_2\mathbf{x}_2 \otimes \mathbf{y}_2 \otimes \mathbf{z}_2 - \cdots - \delta_r\mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r\|.$$

- **Surprise:** Only the last problem has a solution in general.
- **Explanation:** Set of tensors (resp. symmetric tensors) of rank (resp. symmetric rank) $\leq r$ is not closed.

Tensor approximation is ill-behaved

- 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 example, in a slightly different form, may be found in [Bini, Lotti, Romani; 80]. But see [de Silva, L; 08] for a proof.

Problem is worse than you may think

- Such phenomenon can and will happen for all orders > 2, all norms, and many ranks:

Theorem (de Silva-L)

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 tensor in $\mathbb{R}^{d_1 \times \cdots \times d_k}$ can have rank exceeding $\min\{d_1, \dots, d_k\}$.

Problem is even worse than you may think

- Tensor rank can jump over an arbitrarily large gap:

Theorem (de Silva-L)

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 $\operatorname{rank}_{\otimes}(\mathcal{A}) = r + s$.

- Tensors that fail to have best low-rank approximations are not rare. May occur with non-zero probability; sometimes with certainty.

Theorem (de Silva-L)

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

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

In $\mathbb{R}^{2\times 2\times 2}$, all rank-3 tensors fail to have best rank-2 approximation.

Happens to symmetric tensors ...

- Approximation of a homogeneous polynomial by a sum of powers of linear forms (e.g. Independent Components Analysis).
- Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ be linearly independent. Define for $n \in \mathbb{N}$,

$$A_n := n \left[\mathbf{x} + \frac{1}{n}\mathbf{y} \right]^{\otimes p} - n\mathbf{x}^{\otimes p}$$

- Define

$$\mathcal{A} := \mathbf{x} \otimes \mathbf{y} \otimes \cdots \otimes \mathbf{y} + \mathbf{y} \otimes \mathbf{x} \otimes \cdots \otimes \mathbf{y} + \cdots + \mathbf{y} \otimes \mathbf{y} \otimes \cdots \otimes \mathbf{x}.$$

- Then $\operatorname{rank}_{\mathsf{S}}(\mathcal{A}_n) \leq 2$, $\operatorname{rank}_{\mathsf{S}}(\mathcal{A}) \geq p$, and

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

- See [Comon, Golub, L, Mourrain; 08] for details.

. . . and to operators . . .

- Approximation of an operator by a sum of Kronecker product of lower-dimensional operators (e.g. Numerical Operator Calculus).
- For linearly independent operators $P_i, Q_i : V_i \to W_i$, $i = 1, 2, 3$, let $\mathcal{D} : V_1 \otimes V_2 \otimes V_3 \to W_1 \otimes W_2 \otimes W_3$ be

$$\mathcal{D} := P_1 \otimes Q_2 \otimes Q_3 + Q_1 \otimes Q_2 \otimes P_3 + Q_1 \otimes Q_2 \otimes P_3.$$

- If finite-dimensional, then '$\otimes$' may be taken to be Kronecker product of matrices.
- For $n \in \mathbb{N}$,

$$\mathcal{D}_n := n\left[P_1 + \frac{1}{n}Q_1\right] \otimes \left[P_2 + \frac{1}{n}Q_2\right] \otimes \left[P_3 + \frac{1}{n}Q_3\right] - nP_1 \otimes P_2 \otimes P_3.$$

- Then

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

. . . and functions too

- Approximation of a multivariate function by a sum of separable functions (e.g. Approximation Theory).
- For linearly independent $\varphi_1, \psi_1 : X \to \mathbb{R}$, $\varphi_2, \psi_2 : Y \to \mathbb{R}$, $\varphi_3, \psi_3 : Z \to \mathbb{R}$, let $f : X \times Y \times Z \to \mathbb{R}$ be

$$f(x, y, z) := \varphi_1(x)\psi_2(y)\psi_3(z) + \psi_1(x)\psi_2(y)\varphi_3(z) + \psi_1(x)\psi_2(y)\varphi_3(z).$$

- For $n \in \mathbb{N}$,

$$f_n(x, y, z) := n\left[\varphi_1(x) + \frac{1}{n}\psi_1(x)\right]\left[\varphi_2(y) + \frac{1}{n}\psi_2(y)\right]\left[\varphi_3(z) + \frac{1}{n}\psi_3(z)\right] - n\varphi_1(x)\varphi_2(y)\varphi_3(z).$$

- Then

$$\lim_{n\to\infty} f_n = f.$$

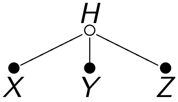
Message

- That the best rank-r approximation problem for tensor 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?

Nonnegativity helps

Approximation of joint probability distributions by conditional probability distributions under the Naïve Bayes Hypothesis:

$$\Pr(x, y, z) = \sum_h \Pr(h) \Pr(x \mid h) \Pr(y \mid h) \Pr(z \mid h)$$



Theorem (L-Comon)

The set $\{\mathcal{A} \in \mathbb{R}_+^{l \times m \times n} \mid \operatorname{rank}_+(\mathcal{A}) \leq r\}$ *is closed.*

- *Extends to arbitrary order.*
- *Independent of norms and even Brègman divergences.*
- *Holds more generally over* $C_1 \otimes \cdots \otimes C_p$ *where* $C_1, \dots, C_p$ *are line-free cones.*

Problem 2

Problem (Nonnegative Tensor Approximations)

Let $\mathcal{A} = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}_+^{d_1 \times \cdots \times d_k}$. *Determine*

$$\operatorname{argmin}\Bigl\{\bigl\|\mathcal{A} - \sum\nolimits_{i=1}^{r} \delta_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{z}_i\bigr\| \Bigm| \delta_i, \mathbf{u}_i, \dots, \mathbf{z}_i \geq 0\Bigr\}.$$

Here $\|\cdot\|$ *is the sum of squares norm and* $\mathbf{u}_i, \dots, \mathbf{z}_i$ *are assumed to have unit 2-norms.*

More generally, want:

Problem (Polynomially Constrained Tensor Approximations)

Let $\mathcal{A} = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. *Determine if*

$$\operatorname{argmin}\Bigl\{\bigl\|\mathcal{A} - \sum\nolimits_{i=1}^{r} \delta_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{z}_i\bigr\| \Bigm| P(\delta_1, \dots, \delta_r, \mathbf{u}_1, \dots, \mathbf{z}_r) \geq 0\Bigr\}$$

has a solution and if so find it.

Examples of symmetric tensors

- Higher order derivatives of real-valued multivariate functions.
- Moments of a vector-valued random variable $\mathbf{x} = (x_1, \dots, x_n)$:

$$\begin{aligned}\mathcal{S}_p(\mathbf{x}) &= \llbracket E(x_{j_1} x_{j_2} \cdots x_{j_p}) \rrbracket_{j_1,\dots,j_p=1}^{n} \\ &= \left\llbracket \int \cdots \int x_{j_1} x_{j_2} \cdots x_{j_p} \, d\mu(x_{j_1}) \cdots d\mu(x_{j_p}) \right\rrbracket_{j_1,\dots,j_p=1}^{n}.\end{aligned}$$

- Cumulants of a vector-valued random variable $\mathbf{x} = (x_1, \dots, x_n)$:

$$\mathcal{K}_p(\mathbf{x}) = \left\llbracket \sum_{A_1 \sqcup \cdots \sqcup A_q = \{j_1,\dots,j_p\}} (-1)^{q-1}(q-1)! E\Bigl(\prod_{j \in A_1} x_j\Bigr) \cdots E\Bigl(\prod_{j \in A_q} x_j\Bigr) \right\rrbracket_{j_1,\dots,j_p=1}^{n}.$$

Cumulants

- In terms of log characteristic and cumulant generating functions,

$$\kappa_{j_1\cdots j_p}(\mathbf{x}) = \frac{\partial^p}{\partial t_{j_1}\cdots\partial t_{j_p}} \log \mathbf{E}(\exp(\langle \mathbf{t}, \mathbf{x}\rangle)\Big|_{\mathbf{t}=\mathbf{0}}$$
$$= (-1)^p \frac{\partial^p}{\partial t_{j_1}\cdots\partial t_{j_p}} \log \mathbf{E}(\exp(i\langle \mathbf{t}, \mathbf{x}\rangle)\Big|_{\mathbf{t}=\mathbf{0}}.$$

- In terms of Edgeworth expansion,

$$\log \mathbf{E}(\exp(i\langle \mathbf{t}, \mathbf{x}\rangle) = \sum_{\alpha=0}^{\infty} i^{|\alpha|}\kappa_\alpha(\mathbf{x})\frac{\mathbf{t}^\alpha}{\alpha!}, \quad \log \mathbf{E}(\exp(\langle \mathbf{t}, \mathbf{x}\rangle) = \sum_{\alpha=0}^{\infty} \kappa_\alpha(\mathbf{x})\frac{\mathbf{t}^\alpha}{\alpha!},$$

  $\alpha = (\alpha_1, \dots, \alpha_n)$ is a multi-index, $\mathbf{t}^{\boldsymbol{\alpha}} = t_1^{\alpha_1}\cdots t_n^{\alpha_n}$, $\alpha! = \alpha_1!\cdots\alpha_n!$.
- For each $\mathbf{x}$, $\mathcal{K}_p(\mathbf{x}) = [\![\kappa_{j_1\cdots j_p}(\mathbf{x})]\!] \in \mathsf{S}^p(\mathbb{R}^n)$ is a symmetric tensor.
- [Fisher, Wishart; 1932]

Properties of cumulants

Multilinearity: If $\mathbf{x}$ is a $\mathbb{R}^n$-valued random variable and $A \in \mathbb{R}^{m \times n}$

$$\mathcal{K}_p(A\mathbf{x}) = (A, \dots, A) \cdot \mathcal{K}_p(\mathbf{x}).$$

Additivity: If $\mathbf{x}_1, \dots, \mathbf{x}_k$ are mutually independent of $\mathbf{y}_1, \dots, \mathbf{y}_k$, then

$$\mathcal{K}_p(\mathbf{x}_1 + \mathbf{y}_1, \dots, \mathbf{x}_k + \mathbf{y}_k) = \mathcal{K}_p(\mathbf{x}_1, \dots, \mathbf{x}_k) + \mathcal{K}_p(\mathbf{y}_1, \dots, \mathbf{y}_k).$$

Independence: If I and J partition $\{j_1, \dots, j_p\}$ so that $\mathbf{x}_I$ and $\mathbf{x}_J$ are independent, then

$$\kappa_{j_1 \cdots j_p}(\mathbf{x}) = 0.$$

Support: There are no distributions where

$$\mathcal{K}_p(\mathbf{x}) \begin{cases} \neq 0 & 3 \leq p \leq n, \\ = 0 & p > n. \end{cases}$$

Examples of cumulants

Univariate: $\mathcal{K}_p(x)$ for $p = 1, 2, 3, 4$ are mean, variance, skewness, kurtosis (unnormalized)

Discrete: $x \sim \textsc{Poisson}(\lambda)$, $\mathcal{K}_p(x) = \lambda$ for all p.

Continuous: $x \sim \textsc{Uniform}([0, 1])$, $\mathcal{K}_p(x) = B_p/p$ where $B_p = p$th Bernoulli number.

Nonexistent: $x \sim \textsc{Student}(3)$, $\mathcal{K}_p(x)$ does not exist for all $p \geq 3$.

Multivariate: $\mathcal{K}_1(\mathbf{x}) = \mathrm{E}(\mathbf{x})$ and $\mathcal{K}_2(\mathbf{x}) = \mathrm{Cov}(\mathbf{x})$.

Discrete: $\mathbf{x} \sim \textsc{Multinomial}(n, \mathbf{q})$,

$$\kappa_{j_1 \cdots j_p}(\mathbf{x}) = n \frac{\partial^p}{\partial t_{j_1} \cdots \partial t_{j_p}} \log(q_1 e^{t_1 x_1} + \cdots + q_k e^{t_k x_k}) \Big|_{t_1, \dots, t_k = 0}.$$

Continuous: $\mathbf{x} \sim \textsc{Normal}(\mu, \Sigma)$, $\mathcal{K}_p(\mathbf{x}) = 0$ for all $p \geq 3$.

Problem 3

Problem (K-moment problem)

Given an infinite sequence $(\mathcal{A}_p)_{p=1}^{\infty}$ in $\prod_{p=1}^{\infty} \mathsf{S}^p(\mathbb{R}^n)$, can we find a positive Borel meausre μ supported on a compact set $K \subseteq \mathbb{R}^n$ such that

$$\mathcal{S}_p(\mathbf{x}) = \mathcal{A}_p \quad \text{for all } p \in \mathbb{N}?$$

- Resolved in [Lasserre, Laurent, Rostalski; 08].
- **Immediate observation:** False for cumulants without further restrictions on the sequence. E.g. there are no distributions where

$$\mathcal{K}_p(\mathbf{x}) \begin{cases} \neq 0 & 3 \leq p \leq n, \\ = 0 & p > n. \end{cases}$$

Problem (K-cumulant problem)

How could we pose an equivalent of the K-moment problem for cumulants?

Variational approach to eigenvalues/vectors

- $A \in \mathbb{R}^{m\times n}$ symmetric.
- Eigenvalues and eigenvectors are critical values and critical points of
$$\mathbf{x}^\top A\mathbf{x}/\|\mathbf{x}\|_2^2.$$
- Equivalently, critical values/points of $\mathbf{x}^\top A\mathbf{x}$ constrained to unit sphere.
- Lagrangian:
$$L(\mathbf{x},\lambda) = \mathbf{x}^\top A\mathbf{x} - \lambda(\|\mathbf{x}\|_2^2 - 1).$$
- Vanishing of ∇L at critical $(\mathbf{x}_c, \lambda_c) \in \mathbb{R}^n \times \mathbb{R}$ yields familiar
$$A\mathbf{x}_c = \lambda_c \mathbf{x}_c.$$

Eigenvalues/vectors of a tensor

- Extends to hypermatrices.
- For $\mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$, write $\mathbf{x}^p := [x_1^p, \dots, x_n^p]^\top$.
- Define the 'ℓ^p-norm' $\|\mathbf{x}\|_p = (x_1^p + \cdots + x_n^p)^{1/p}$.
- Define eigenvalues/vectors of $\mathcal{A} \in \mathsf{S}^p(\mathbb{R}^n)$ as critical values/points of the multilinear Rayleigh quotient

$$\mathcal{A}(\mathbf{x}, \dots, \mathbf{x})/\|\mathbf{x}\|_p^p.$$

- Lagrangian

$$L(\mathbf{x}, \lambda) := \mathcal{A}(\mathbf{x}, \dots, \mathbf{x}) - \lambda(\|\mathbf{x}\|_p^p - 1).$$

- At a critical point

$$\mathcal{A}(I_n, \mathbf{x}, \dots, \mathbf{x}) = \lambda \mathbf{x}^{p-1}.$$

Some observations

- If $\mathcal{A}$ is symmetric,

$$\mathcal{A}(I_n, \mathbf{x}, \mathbf{x}, \dots, \mathbf{x}) = \mathcal{A}(\mathbf{x}, I_n, \mathbf{x}, \dots, \mathbf{x}) = \cdots = \mathcal{A}(\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}, I_n).$$

- Defined in [Qi; '05] and [L; '05] independently.
- For unsymmetric hypermatrices — get different eigenpairs for different modes (unsymmetric matrices have different left/right eigenvectors).
- Falls outside Classical Invariant Theory — not invariant under $Q \in \mathrm{O}(n)$, ie. $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.
- Small stabilizer — $Q \in \mathrm{GL}(n)$ with $\|Q\mathbf{x}\|_p = \|\mathbf{x}\|_p$ for all $\mathbf{x}$ is monomial, i.e. product of a permutation matrix and a diagonal scaling.

Problem 4

- Amit Singer's triplewise affinity tensor $\mathcal{A} = [\![a_{ijk}]\!] \in \mathsf{S}^3(\mathbb{R}^n)$ where

$$a_{ijk} = \exp\left[-\frac{d_{ij}^2 + d_{jk}^2 + d_{ki}^2}{\delta}\right] \times \exp\left[-\frac{1}{\epsilon}\sin^2\left(\frac{\theta_{ij} + \theta_{jk} + \theta_{ki}}{2}\right)\right].$$

- From cryo-EM and NMR applications.
- May assume, for simplicity, $a_{ijk} = w_{ij}w_{jk}w_{ki}$ for some nonnegative matrix $W = [w_{ij}] \in \mathsf{S}^2(\mathbb{R}^n)$.

Problem (Singer's problem)

Find the 10 *largest real eigenvalues and real eigenvectors of* $\mathcal{A}$.

- Equivalent to $\mathbf{x}^\top(A_i - \lambda E_i)\mathbf{x} = 0$ for $i = 1, \dots, n$, and $\|\mathbf{x}\|_3^3 = 1$.
- **Fact:** Solving a system of n quadratic equations in n unknowns over any field is NP-hard.
- Can the special structure of the Singer affinity tensor be exploited?