Most tensor problems are NP hard

(preliminary report)

Lek-Heng Lim

University of California, Berkeley

August 25, 2009

(Joint work with: Chris Hillar, MSRI)

Innocent looking problem

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

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

- NP-complete over $\mathbb{F}_q$, NP-hard over $\mathbb{Q}$ [Håstad; 90].
- Slight extension: NP-hard over $\mathbb{R}$ and $\mathbb{C}$ [L & Hillar; 09].
- Convex relaxation along the lines of compressive sensing?

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

- Ky Fan/nuclear/Schatten/trace norm,

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

- [Fazel, Hindi, Boyd; 01], [Recht, Fazel, Parrilo; 09], [Candès, Recht; 09], [Zhu, So, Ye; 09], [Toh, Yun; 09].

Tensors as hypermatrices

Up to choice of bases on U, V, W, a 3-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.

- Scalar = 0-tensor, vector = 1-tensor, matrix = 2-tensor.
- Cubical 3-tensor $[\![a_{ijk}]\!] \in \mathbb{R}^{n \times n \times n}$ is **symmetric** if

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

  Set of symmetric 3-tensors denoted $\mathsf{S}^3(\mathbb{R})$.

Examples

- Higher order derivatives of real-valued multivariate functions:

$$D^{(p)}f(\mathbf{x}) = \left[\!\!\left[\frac{\partial^p f}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_p}} \right]\!\!\right]_{j_1,\dots,j_p=1}^{n}.$$

grad $f(\mathbf{x}) \in \mathbb{R}^n$, Hess $f(\mathbf{x}) \in \mathbb{R}^{n\times n}$, $\dots$, $D^{(p)}f(\mathbf{x}) \in \mathbb{R}^{n\times\dots\times n}$.

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

$$\mathcal{S}_p(\mathbf{x}) = \left[\!\!\left[E(x_{j_1} x_{j_2} \cdots x_{j_p}) \right]\!\!\right]_{j_1,\dots,j_p=1}^{n}.$$

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

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

Blaming the math

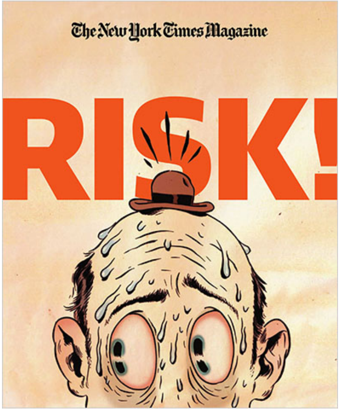
- **Wired:** Gaussian copulas for CDOs.

WIRED

THE
SECRET FORMULA
That Destroyed Wall Street

$$P = \phi(A, B, \gamma)$$

- **NYT:** normal market in VaR.



The New York Times

January 4, 2009

Risk Mismanagement

By JOE NOCERA

THERE AREN'T MANY widely told anecdotes about the current financial crisis, at least not yet, but there's one that made the rounds in 2007, back when the big investment banks were first starting to write down

Why not Gaussian

- Log characteristic function

$$\log \mathrm{E}(\exp(i\langle \mathbf{t}, \mathbf{x}\rangle)) = \sum_{|\alpha|=1}^{\infty} i^{|\alpha|} \kappa_\alpha(\mathbf{x}) \frac{\mathbf{t}^\alpha}{\alpha!}.$$

- Gaussian assumption equivalent to quadratic approximation:

$$\infty = 2.$$

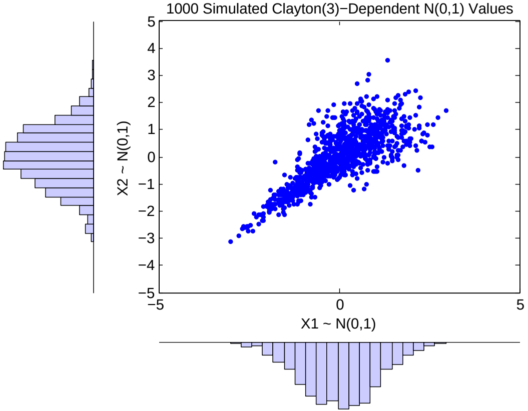
- If $\mathbf{x}$ is multivariate Gaussian, then

$$\log \mathrm{E}(\exp(i\langle \mathbf{t}, \mathbf{x}\rangle)) = i\langle \mathrm{E}(\mathbf{x}), \mathbf{t}\rangle + \frac{1}{2}\mathbf{t}^\top \operatorname{Cov}(\mathbf{x})\mathbf{t}.$$

- $\mathcal{K}_1(\mathbf{x})$ mean (vector), $\mathcal{K}_2(\mathbf{x})$ variance (matrix), $\mathcal{K}_3(\mathbf{x})$ skewness (3-tensor), $\mathcal{K}_4(\mathbf{x})$ kurtosis (4-tensor),. . . .
- **Non-Gaussian data:** Not enough to look at just mean and covariance.

Why not copulas

- Nassim Taleb: "Anything that relies on correlation is charlatanism."
- Even if marginals normal, dependence might not be.



- For heavy tail phenomena, important to examine tensor-valued quantities like kurtosis.

Why not VaR

- Paul Wilmott: "The relationship between two assets can never be captured by a single scalar quantity."
- Multivariate $f : \mathbb{R}^n \to \mathbb{R}$

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

- **Hooke's law in 1D:** x extension, F force, k spring constant,

$$F = -kx.$$

- **Hooke's law in 3D:** $\mathbf{x} = (x_1, x_2, x_3)$, elasticity tensor $\mathcal{C} \in \mathbb{R}^{3\times3\times3\times3}$, stress $\Sigma \in \mathbb{R}^{3\times3}$, strain $\Gamma \in \mathbb{R}^{3\times3}$

$$\sigma_{ij} = \sum\nolimits_{k,l=1}^{3} c_{ijkl}\gamma_{kl}.$$

Numerical linear algebra

$\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. $A \in \mathbb{F}^{m\times n}$, $\mathbf{b} \in \mathbb{F}^m$, and $r \le \min\{m, n\}$.

Rank and numerical rank: Determine $\operatorname{rank}(A)$.

Linear system of equations: Determine if $A\mathbf{x} = \mathbf{b}$ has a solution and if so determine a solution $\mathbf{x} \in \mathbb{F}^n$.

Linear least squares problem: Determine an $\mathbf{x} \in \mathbb{F}^n$ that minimizes $\|A\mathbf{x} - \mathbf{b}\|_2$.

Spectral norm: Determine the value of $\|A\|_{2,2} := \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sigma_{\max}(A)$.

Eigenvalue problem: If $m = n$, determine $\lambda \in \mathbb{F}$ and non-zero $\mathbf{x} \in \mathbb{F}^n$ with $A\mathbf{x} = \lambda\mathbf{x}$.

Singular value problem: Determine $\sigma \in \mathbb{F}$ and non-zero $\mathbf{x} \in \mathbb{F}^n$, $\mathbf{y} \in \mathbb{F}^m$ with $A\mathbf{x} = \sigma\mathbf{y}$, $A^\top\mathbf{y} = \sigma\mathbf{x}$.

Low rank approximation: Determine $A_r \in \mathbb{F}^{m\times n}$ with $\operatorname{rank}(A_r) \le r$ and $\|A - A_r\|_F = \min_{\operatorname{rank}(B)\le r} \|A - B\|_F$.

Numerical multilinear algebra?

- Computing the spectral norm of a 3-tensor:

$$\sup_{\mathbf{x},\mathbf{y},\mathbf{z}\neq\mathbf{0}} \frac{|\mathcal{A}(\mathbf{x},\mathbf{y},\mathbf{z})|}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2\|\mathbf{z}\|_2}.$$

- Computing the spectral norm of a symmeric 3-tensor:

$$\sup_{\mathbf{x}\neq\mathbf{0}} \frac{|\mathcal{S}(\mathbf{x},\mathbf{x},\mathbf{x})|}{\|\mathbf{x}\|_2^3}.$$

- Computing a best rank-1 approximation to a tensor:

$$\min_{\mathbf{x},\mathbf{y},\mathbf{z}}\|\mathcal{A} - \mathbf{x}\otimes\mathbf{y}\otimes\mathbf{z}\|_F.$$

- Computing a best rank-1 approximation to a symmetric tensor:

$$\min_{\mathbf{x}}\|\mathcal{S} - \mathbf{x}\otimes\mathbf{x}\otimes\mathbf{x}\|_F.$$

Numerical multilinear algebra?

- Determine if a given value is a singular value of a 3-tensor

$$\operatorname*{crit}_{\mathbf{x},\mathbf{y},\mathbf{z}\neq\mathbf{0}} \frac{\mathcal{A}(\mathbf{x},\mathbf{y},\mathbf{z})}{\|\mathbf{x}\|_3\|\mathbf{y}\|_3\|\mathbf{z}\|_3}.$$

- Determine if a given value is an eigenvalue of a symmetric 3-tensor

$$\operatorname*{crit}_{\mathbf{x}\neq\mathbf{0}} \frac{\mathcal{S}(\mathbf{x},\mathbf{x},\mathbf{x})}{\|\mathbf{x}\|_3^3}.$$

- Solving a system of bilinear equations in the exact sense

$$\mathcal{A}(\mathbf{x},\mathbf{y},I) = \mathbf{b}$$

  or approximating it in the least-squares sense

$$\min_{\mathbf{x},\mathbf{y}} \|\mathcal{A}(\mathbf{x},\mathbf{y},I) - \mathbf{b}\|_2.$$

- Shorthand:

$$\mathcal{A}(\mathbf{x},\mathbf{y},\mathbf{z}) = \sum\nolimits_{i,j,k=1}^{l,m,n} a_{ijk} x_i y_j z_{j_k}, \quad \mathcal{A}(\mathbf{x},\mathbf{y},I) = \sum\nolimits_{i,j=1}^{l,m} a_{ijk} x_i y_j.$$

Tensor rank = Innocent looking roblem

- 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

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

- $U \otimes V \otimes W \simeq \operatorname{Hom}(U, V \otimes W)$.
- Write $\mathcal{A} = [A_1, \dots, A_l]$ where $A_1, \dots, A_l \in \mathbb{R}^{m \times n}$. Then

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

Best low rank approximation of a matrix

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

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

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

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

Then

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

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

Best low-rank approximation of a tensor

- Given $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$, find $\sigma_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$, minimizing

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

- **Surprise:** In general, existence of solution guaranteed only if $r = 1$.

Theorem (de Silva-L)

Let $k \geq 3$ and $d_1, \ldots, d_k \geq 2$. For any s such that

$$2 \leq s \leq \min\{d_1, \ldots, d_k\},$$

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.

- **Symmetric variant:** $\mathcal{A} \in \mathsf{S}(\mathbb{R}^n)$, find $\lambda_i, \mathbf{v}_i$ minimizing

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

- Also ill-behaved [Comon, Golub, L, Mourrain; 08].

Best low-rank approximation of a tensor

- **Explanation:** Set of tensors (resp. symmetric tensors) of rank (resp. symmetric rank) $\leq r$ is not closed unless $r = 1$.
- **Another well behaved case:** If $\sigma_i, \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$ constrained to nonnegative orthant, then a solution always exist [L & Comon; 09].
- **Nonnegative tensor approximations:** General non-linear programming algorithms, cf. [Friedlander & Hatz; 08] and many others.
- On-going work with Kojima and Toh: SDP algorithm for convex relaxation of well-behaved cases.
- Best rank-1 approximation of a symmetric tensor also NP-hard [L & Hillar; 09]

$$\min_{\mathbf{x}} \|\mathcal{S} - \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}\|_F.$$

Variational approach to eigenvalues/vectors

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

Eigenvalues/vectors of a tensor

- Extends to tensors [L; 05].
- 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 $\mathcal{S} \in \mathsf{S}^p(\mathbb{R}^n)$ as critical values/points of the multilinear Rayleigh quotient

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

- Lagrangian

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

- At a critical point

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

Some observations

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

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

- Defined in [Qi; 05] and [L; 05] independently.
- Related to rank-1 approximation:

$$\min_{\mathbf{x}} \|\mathcal{S} - \lambda \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}\|_F$$

  attained when

$$\lambda = \max_{\|\mathbf{x}\|=1} \langle \mathcal{S}, \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} \rangle.$$

- For $p = 3$, λ is an eigenvalue of $\mathcal{S}$ iff

$$\mathbf{x}^\top (S_i - \lambda E_i)\mathbf{x} = 0, \quad i = 1, \dots, n,$$

  for some non-zero $\mathbf{x}$. Here $S_i = [s_{ijk}]_{j,k=1}^n$ and $E_i = \mathbf{e}_i \mathbf{e}_i^\top$.

Graph coloring as polynomial system

Found in various forms [Bayer; 82], [Lovász; 94], [de Loera; 95]. Form below from [L & Hillar; 09].

Lemma

Let G be a graph. The $3n + |E|$ polynomials in $2n + 1$ indeterminates

$$\begin{cases} x_i y_i - z^2, \quad y_i z - x_i^2, \quad x_i z - y_i^2, & i = 1, \dots, n, \\ x_i^2 + x_i x_j + x_j^2, & \{i, j\} \in E, \end{cases}$$

has a common nontrivial complex solution if and only if the graph G is 3-colorable.

Tensor eigenvalue is NP-hard

Problem (SymmQuadFeas)

Let $\mathcal{S} = [\![s_{ijk}]\!] \in \mathsf{S}^3(\mathbb{R}^n)$, *i.e.* $s_{ijk} = s_{ikj} = s_{jik} = s_{jki} = s_{kij} = s_{kji}$. *Let* $G_i(\mathbf{x}) = \mathbf{x}^\top S_i \mathbf{x}$ *for* $i = 1, \dots, n$ *be the associated* n *homogeneous, real quadratic forms. Determine if the system of equations* $\{G_i(\mathbf{x}) = c_i\}_{i=1}^n$ *has a nontrivial real solution* $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.

Theorem (L & Hillar; 09)

Graph coloring is polynomial reducible to SymmQuadFeas.

Corollary

Given symmetric tensor $\mathcal{S} \in \mathsf{S}^3(\mathbb{R}^n)$.

1. *Given* $\lambda \in \mathbb{R}$. *NP hard to check if* λ *is eigenvalue of* $\mathcal{S}$.
2. *NP hard to find a best rank-*1 *approximation* $\lambda \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}$ *to* $\mathcal{S}$.

Caveat on complexity theory

- Different notions of NP-hardness:
  - Cook-Karp-Levin (traditional);
  - Blum-Shub-Smale;
  - Valiant.
- Real versus bit complexity
  - **Real complexity**: number of field operations required to determine feasibility as a function of n.
  - **Bit complexity**: if matrices $\mathcal{A}$ has rational entries, number of bit operations necessary as a function of the number of bits required to specify all the a_{ijk}.
  - E.g. multiply two integers of size N, i.e. $\log N$ bits: real complexity is $O(1)$, bit complexity is $O(\log N \log\log N)$.