Symmetric eigenvalue decompositions for symmetric tensors

Lek-Heng Lim

University of California, Berkeley

January 29, 2009

(Contains joint work with Pierre Comon, Jason Morton, Bernard Mourrain, Berkant Savas)

Why symmetric tensors?

Question

What lesson in algebra did we learn from the current global economic/financial crisis?

Why symmetric tensors?

Question

What lesson in algebra did we learn from the current global economic/financial crisis?

- **One answer:** it's important to look beyond the quadratic term.

Why symmetric tensors?

Question

What lesson in algebra did we learn from the current global economic/financial crisis?

- **One answer:** it's important to look beyond the quadratic term.
- **Taylor approximation:** multivariate $f(x_1, \dots, x_n)$ approximated as

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

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

Why symmetric tensors?

Question

What lesson in algebra did we learn from the current global economic/financial crisis?

- **One answer:** it's important to look beyond the quadratic term.
- **Taylor approximation:** multivariate $f(x_1, \dots, x_n)$ approximated as

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

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

- A_2 symmetric matrix: $d = 2$.

Why symmetric tensors?

Question

What lesson in algebra did we learn from the current global economic/financial crisis?

- **One answer:** it's important to look beyond the quadratic term.
- **Taylor approximation:** multivariate $f(x_1, \dots, x_n)$ approximated as

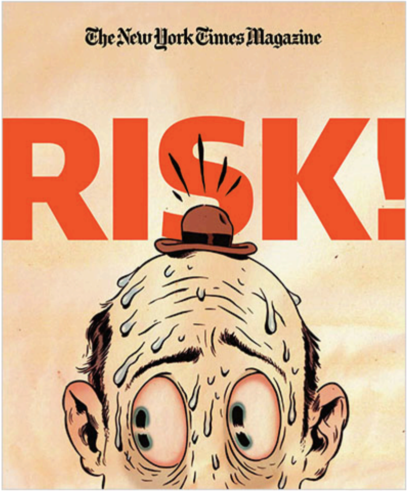
$$f(\mathbf{x}) \approx a_0 + \mathbf{a}_1^\top \mathbf{x} + \mathbf{x}^\top A_2 \mathbf{x} + \mathcal{A}_3(\mathbf{x}, \mathbf{x}, \mathbf{x}) + \cdots + \mathcal{A}_d(\mathbf{x}, \dots, \mathbf{x}) + \cdots .$$

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

- A_2 symmetric matrix: $d = 2$.
- $\mathcal{A}_d$ symmetric hypermatrix: $d > 2$.

The New York Times Magazine

1.4.2009



Zohar Lazar

Risk Mismanagement

By JOE NOCERA

Were the measures used to evaluate Wall Street trades flawed? Or was the mistake ignoring them?

· Times Topics: Credit Crisis

The Way We Live Now

The Senator Track

By LISA BELKIN

Why Caroline Kennedy's "experience" counts.

· Times Topics: Caroline Kennedy
· Post a Comment
· More at The Motherlode Blog

THE MEDIUM

We Interrupt This Program

By VIRGINIA HEFFERNAN

Hulu, the streaming-video service, offers a new (old?) paradigm for watching TV and movies online.

· Post a Comment

QUESTIONS FOR JOAN RIVERS

Cutup

Interview by DEBORAH SOLOMON

The comedian talks about plastic surgery as a business decision, Barack Obama's ears and getting a little work done in a recession.

· Times Topics: Joan Rivers
· Past Questions For ... Columns

ON LANGUAGE

Bleeping Expletives

By WILLIAM SAFIRE

Bonfire of the profanities.

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 billions of dollars in mortgage-backed derivatives and other so-called toxic securities. This was well before Bear Stearns collapsed, before Fannie Mae and Freddie Mac were taken over by the federal government, before Lehman fell and Merrill Lynch was sold and A.I.G. saved, before the $700 billion bailout bill was rushed into law. Before, that is, it became obvious that the risks taken by the largest banks and investment firms in the United States — and, indeed, in much of the Western world — were so excessive and foolhardy that they threatened to bring down the financial system itself. On the contrary: this was back when the major investment firms were still assuring investors that all was well, these little speed bumps notwithstanding — assurances based, in part, on their fantastically complex mathematical models for measuring the risk in their various portfolios.

There are many such models, but by far the most widely used is called VaR — Value at Risk. Built around statistical ideas and probability theories that have been around for centuries, VaR was developed and popularized in the early 1990s by a handful of scientists and mathematicians — "quants," they're called in the business — who went to work for JPMorgan. VaR's great appeal, and its great selling point to people who do not happen to be quants, is that it expresses risk as a single number, a dollar figure, no less.

VaR isn't one model but rather a group of related models that share a mathematical framework. In its most common form, it measures the boundaries of risk in a portfolio over short durations, assuming a "normal" market. For instance, if you have $50 million of weekly VaR, that means that over the course of the next week, there is a 99 percent chance that your portfolio won't lose more than $50 million. That portfolio could consist of equities, bonds, derivatives or all of the above; one reason VaR became so popular is that it is the only commonly used risk measure that can be applied to just about any asset class. And it takes into account a head-spinning variety of variables, including diversification, leverage and volatility, that make up the kind of market risk that traders and firms face every day.

Another reason VaR is so appealing is that it can measure both individual risks — the amount of risk contained in a single trader's portfolio, for instance — and firmwide risk, which it does by combining the VaRs of a given firm's trading desks and coming up with a net number. Top executives usually know their firm's daily VaR within minutes of the market's close.

with “Fooled by Randomness,” which was published in 2001 and became an immediate cult classic on Wall Street, and more recently with “The Black Swan: The Impact of the Highly Improbable,” which came out in 2007 and landed on a number of best-seller lists. He also went from being primarily an options trader to what he always really wanted to be: a public intellectual. When I made the mistake of asking him one day whether he was an adjunct professor, he quickly corrected me. “I’m the Distinguished Professor of Risk Engineering at N.Y.U.,” he responded. “It’s the highest title they give in that department.” Humility is not among his virtues. On his Web site he has a link that reads, “Quotes from ‘The Black Swan’ that the imbeciles did not want to hear.”

“How many of you took statistics at Columbia?” he asked as he began his lecture. Most of the hands in the room shot up. “You wasted your money,” he sniffed. Behind him was a slide of Mickey Mouse that he had put up on the screen, he said, because it represented “Mickey Mouse probabilities.” That pretty much sums up his view of business-school statistics and probability courses.

Taleb’s ideas can be difficult to follow, in part because he uses the language of academic statisticians; words like “Gaussian,” “kurtosis” and “variance” roll off his tongue. But it’s also because he speaks in a kind of brusque shorthand, acting as if any fool should be able to follow his train of thought, which he can’t be bothered to fully explain.

“This is a Stan O’Neal trade,” he said, referring to the former chief executive of Merrill Lynch. He clicked to a slide that showed a trade that made slow, steady profits — and then quickly spiraled downward for a giant, brutal loss.

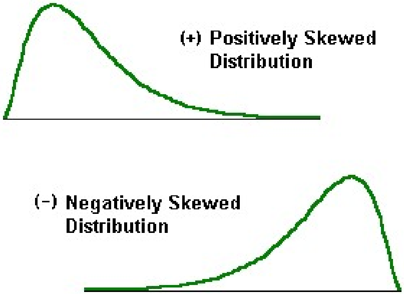
“Why do people measure risks against events that took place in 1987?” he asked, referring to Black Monday, the October day when the U.S. market lost more than 20 percent of its value and has been used ever since as the worst-case scenario in many risk models. “Why is that a benchmark? I call it future-blindness.

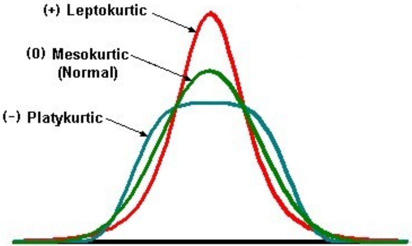
“If you have a pilot flying a plane who doesn’t understand there can be storms, what is going to happen?” he asked. “He is not going to have a magnificent flight. Any small error is going to crash a plane. This is why the crisis that happened was predictable.”

Eventually, though, you do start to get the point. Taleb says that Wall Street risk models, no matter how

Cumulants

- **Univariate distribution:** Mean, variance, skewness and kurtosis describe the *shape*.





- **Multivariate distribution:** (Co)variance matrix *partly* describes the dependence structure.
- But if the variables are not multivariate Gaussian, not the whole story. Covariance matrix analogs: **cumulants**.

More about cumulants

- For univariate x, cumulants $\mathcal{K}_d(x)$ for $d = 1, 2, 3, 4$ are
  - expectation $\kappa_i = \mathsf{E}(x)$,
  - variance $\kappa_{ii} = \sigma^2$,
  - skewness $\kappa_{iii}/\kappa_{ii}^{3/2}$, and
  - kurtosis $\kappa_{iiii}/\kappa_{ii}^2$.
- For multivariate $\mathbf{x}$, $\mathcal{K}_d(\mathbf{x})$ are symmetric tensors of order d.
- Provide a natural measure of non-Gaussianity: If $\mathbf{x}$ Gaussian,

$$\mathcal{K}_d(\mathbf{x}) = 0 \quad \text{for all } d \geq 3.$$

- Describe higher order dependence among random variables.
- Variance is symmetric matrix, may perform eigenvalue decomposition.
- How to analyze higher-order cumulants?
- Want analogue of 'eigenvalue decomposition' for symmetric tensors.

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

Symmetric tensors as hypermatrices

- 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

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

Symmetric tensors as polynomials

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

$$F(\mathbf{x}) = \sum\nolimits_{j_1,\dots,j_p=1}^{n} \binom{p}{d_1, \dots, d_n} a_{j_1\cdots j_p} x_1^{d_1} \cdots x_n^{d_n},$$

  $d_j =$ number of times index j appears in $j_1, \dots, j_p$,

$$d_1 + \cdots + d_n = p.$$

- $\mathsf{S}^p(\mathbb{C}^n) \cong \mathbb{C}[x_1, \dots, x_n]_p$.

Examples of symmetric tensors

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

$$\mathcal{S}_p(\mathbf{x}) = \llbracket E(x_{j_1} x_{j_2} \cdots x_{j_p}) \rrbracket_{j_1,\dots,j_p=1}^{n}.$$

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

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

Tensor ranks

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

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

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

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

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

- In general, $r_1(\mathcal{A}) \neq r_2(\mathcal{A}) \neq r_3(\mathcal{A}) \neq \operatorname{rank}_{\otimes}(\mathcal{A})$.

Symmetric tensor ranks

- **Multilinear rank.** $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$. Then

$$r_1(\mathcal{A}) = r_2(\mathcal{A}) = r_3(\mathcal{A}).$$

- **Outer product rank.** $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$.

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

Matrix EVD and SVD

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

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

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

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

  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.

One plausible EVD and SVD for hypermatrices

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

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

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

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

  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, $\sigma_i \in \mathbb{R}$.

Another plausible EVD and SVD for hypermatrices

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

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

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

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

  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 $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$.

Eigenvalue decompositions for symmetric tensors

Let $\mathcal{A} \in \mathsf{S}^3(\mathbb{R}^n)$.

- **Symmetric outer product decomposition**

$$\mathcal{A} = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i = (V, V, V) \cdot \Lambda.$$

  where $\operatorname{rank}_{\mathsf{S}}(A) = r$, $V \in \mathbb{R}^{n \times r}$, $\Lambda = \operatorname{diag}[\lambda_1, \dots, \lambda_r] \in \mathsf{S}^3(\mathbb{R}^n)$.

- In general, r can exceed n.
- **Symmetric multilinear decomposition**

$$\mathcal{A} = (U, U, U) \cdot \mathcal{C} = \sum\nolimits_{i,j,k=1}^{s} c_{ijk} \mathbf{u}_i \otimes \mathbf{u}_j \otimes \mathbf{u}_k$$

  where $\operatorname{rank}_{\boxplus}(A) = (s, s, s)$, $U \in \mathrm{O}(n, s)$, $\mathcal{C} = [\![c_{ijk}]\!] \in \mathsf{S}^3(\mathbb{R}^s)$.

- $s \leq n$.

Geometry of symmetric outer product decomposition

- Embedding
$$\nu_{n,p} : \mathbb{R}^n \to \mathsf{S}^p(\mathbb{R}^n) \cong \mathbb{R}[x_1, \dots, x_n]_p.$$
- Image $\nu_{n,p}(\mathbb{R}^n)$ is (real affine) **Veronese variety**, set of rank-1 symmetric tensors
$$\mathrm{Ver}_p(\mathbb{R}^n) = \{\mathbf{v}^{\otimes p} \in \mathsf{S}^p(\mathbb{R}^n) \mid \mathbf{v} \in \mathbb{R}^n\}.$$
- As polynomials,
$$\mathrm{Ver}_p(\mathbb{R}^n) = \{L(\mathbf{x})^p \in \mathbb{R}[x_1, \dots, x_n]_p \mid L(\mathbf{x}) = \alpha_1 x_1 + \dots + \alpha_n x_n\},$$
cf. Bernd Sturmfels's talk on Feb. 20.
- $\mathcal{A} \in \mathsf{S}^p(\mathbb{R}^n)$ has rank 2 iff it sits on a bisecant line through two points of $\mathrm{Ver}_p(\mathbb{R}^n)$, rank 3 iff it sits on a trisecant plane through three points of $\mathrm{Ver}_p(\mathbb{R}^n)$, etc.

Secant varieties

- For a nondegenerate variety $X \subseteq \mathbb{R}^n$, write

$$s_r(X) = \text{union of } s\text{-secants to } X \text{ for } s \leq r,$$

  r-**secant quasiprojective variety** of X.

- r-**secant variety**,

$$\sigma_r(X) = \text{Zariski closure of } s_r(X).$$

- Unsymmetric version,

$$X = \operatorname{Seg}(\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_p}) = \{\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \in \mathbb{R}^{d_1 \times \cdots \times d_p} \mid \mathbf{v}_i \in \mathbb{R}^{d_i}\}.$$

- Series difficulty in applications: $s_r(X) \neq \sigma_r(X)$ for $r > 1$.

Outer product approximation is ill-behaved

- 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. For exact decomposition when $r < n$, algorithm of [Comon, Mourrain, Tsigaridas; 09]

Aside: happens to operators too

- 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 $\lim_{n\to\infty} \mathcal{D}_n = \mathcal{D}$.
- More widespread than one may think. See [de Silva, L; 08] for details.

Geometry of symmetric multilinear decomposition

- **Symmetric subspace variety**,

$$\mathrm{Sub}_r^p(\mathbb{R}^n) := \{\mathcal{A} \in \mathsf{S}^p(\mathbb{R}^n) \mid \exists V \leq \mathbb{R}^n \text{ such that } \mathcal{A} \in \mathsf{S}^p(V)\}.$$

- Equivalently,

$$\mathrm{Sub}_r^p(\mathbb{R}^n) := \{\mathcal{A} \in \mathsf{S}^p(\mathbb{R}^n) \mid \mathrm{rank}_{\boxplus}(\mathcal{A}) \leq (r, r, r)\}.$$

- Unsymmetric version, cf. [Landsberg, Weyman; 07],

$$\begin{aligned} &\mathrm{Sub}_{p,q,r}(\mathbb{R}^l, \mathbb{R}^m, \mathbb{R}^n) \\ &\quad = \{\mathcal{A} \in \mathbb{R}^{l\times m\times n} \mid \exists U, V, W \text{ such that } \mathcal{A} \in U \otimes V \otimes W\} \\ &\quad = \{\mathcal{A} \in \mathbb{R}^{l\times m\times n} \mid \mathrm{rank}_{\boxplus}(\mathcal{A}) \leq (p, q, r)\}. \end{aligned}$$

- Symmetric subspace variety in $\mathsf{S}^p(\mathbb{R}^n)$ — closed, irreducible, easier to study.
- Quasiprojective secant variety of Veronese in $\mathsf{S}^p(\mathbb{R}^n)$ — not closed, not irreducible, difficult to study.

Grassmannian parameterization

- Stiefel manifold $\mathrm{O}(n, r)$: set of $n \times r$ real matrices with orthonormal columns. $\mathrm{O}(n, n) = \mathrm{O}(n)$, usual orthogonal group.
- Grassman manifold $\mathrm{Gr}(n, r)$: set of equivalence classes of $\mathrm{O}(n, r)$ under left multiplication by $\mathrm{O}(n)$.
- Parameterization via

$$\mathrm{Gr}(n, r) \times \mathsf{S}^p(\mathbb{R}^r) \to \mathsf{S}^p(\mathbb{R}^n).$$

- Image is $\mathrm{Sub}_r^p(\mathbb{R}^n)$.
- More generally

$$\mathrm{Gr}(n, r) \times \prod_{p=1}^{\infty} \mathsf{S}^p(\mathbb{R}^r) \to \prod_{p=1}^{\infty} \mathsf{S}^p(\mathbb{R}^n).$$

- Image is $\prod_{p=1}^{\infty} \mathrm{Sub}_r^p(\mathbb{R}^n)$.

From Stieffel to Grassmann

- Given $\mathcal{A} \in \mathsf{S}^p(\mathbb{R}^n)$, some $r \ll n$, want

$$\min_{X \in \mathrm{O}(n,r),\, \mathcal{C} \in \mathsf{S}^p(\mathbb{R}^r)} \|\mathcal{A} - (X, \dots, X) \cdot \mathcal{C}\|_F,$$

- Unlike approximation by secants of Veronese, subspace approximation problem always has an globally optimal solution.
- Equivalent to

$$\max_{X \in \mathrm{O}(n,r)} \|(X^\top, \dots, X^\top) \cdot \mathcal{A}\|_F = \max_{X \in \mathrm{O}(n,r)} \|\mathcal{A} \cdot (X, \dots, X)\|_F.$$

- Problem defined on a Grassmannian since

$$\|\mathcal{A} \cdot (X, \dots, X)\|_F = \|\mathcal{A} \cdot (XQ, \dots, XQ)\|_F,$$

  for any $Q \in \mathrm{O}(r)$. Only the subspaces spanned by X matters.
- Equivalent to

$$\max_{X \in \mathrm{Gr}(n,r)} \|\mathcal{A} \cdot (X, \dots, X)\|_F.$$

- Efficient algorithm exists: Limited memory BFGS on Grassmannian [Savas, L; '09]

Cumulants

- In terms of log characteristic and cumulant generating functions,

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

- 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

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

Principal and independent components analysis

Principal components analysis: $\mathbf{s}$ Gaussian,

$$\hat{\mathcal{K}}_2(\mathbf{y}) = Q\Lambda_2 Q^\top = (Q, Q) \cdot \Lambda_2,$$

$\Lambda_2 \approx \hat{\mathcal{K}}_2(\mathbf{s})$ diagonal matrix, $Q \in \mathrm{O}(n, r)$, [Pearson; 1901].

Independent components analysis: $\mathbf{s}$ statistically independent entries, ε Gaussian

$$\hat{\mathcal{K}}_p(\mathbf{y}) = (Q, \dots, Q) \cdot \Lambda_p, \quad p = 2, 3, \dots,$$

$\Lambda_p \approx \hat{\mathcal{K}}_p(\mathbf{s})$ diagonal tensor, $Q \in \mathrm{O}(n, r)$, [Comon; 1994].

What if

- $\mathbf{s}$ not Gaussian, e.g. power-law distributed data in social networks.
- $\mathbf{s}$ not independent, e.g. functional components in neuroimaging.
- ε not white noise, e.g. idiosyncratic factors in financial modelling.

Principal cumulant components analysis [L. & Morton]

- Note that if $\varepsilon = \mathbf{0}$, then
$$\mathcal{K}_p(\mathbf{y}) = \mathcal{K}_p(Q\mathbf{s}) = (Q, \dots, Q) \cdot \mathcal{K}_p(\mathbf{s}).$$
- In general, want principal components that account for variation in all cumulants simultaneously
$$\min_{Q \in \mathrm{O}(n,r),\, \mathcal{C}_p \in \mathsf{S}^p(\mathbb{R}^r)} \sum_{p=1}^{\infty} \alpha_p \|\hat{\mathcal{K}}_p(\mathbf{y}) - (Q, \dots, Q) \cdot \mathcal{C}_p\|_F^2,$$
- $\mathcal{C}_p \approx \hat{\mathcal{K}}_p(\mathbf{s})$ not necessarily diagonal.
- Appears intractable: optimization over infinite-dimensional manifold
$$\mathrm{O}(n, r) \times \prod_{p=1}^{\infty} \mathsf{S}^p(\mathbb{R}^r).$$
- Surprising relaxation: optimization over a single Grassmannian $\mathrm{Gr}(n, r)$ of dimension $r(n - r)$,
$$\max_{Q \in \mathrm{Gr}(n,r)} \sum_{p=1}^{\infty} \alpha_p \|\hat{\mathcal{K}}_p(\mathbf{y}) \cdot (Q, \dots, Q)\|_F^2.$$
- In practice $\infty = 3$ or 4.