Modeling and learning with tensors

Lek-Heng Lim

University of California, Berkeley

February 20, 2009

(Thanks: Charlie Van Loan, National Science Foundation; Collaborators: Jason Morton, Berkant Savas, Yuan Yao)

Why tensors?

Question

What lesson about tensor modeling did we learn from the current global financial crisis?

- **One answer:** Better understanding of tensor-valued quantities (in this case, measures of risk) might have at least forewarned one to the looming dangers.
- Expand multivariate $f(x_1, \dots, x_n)$ in power series

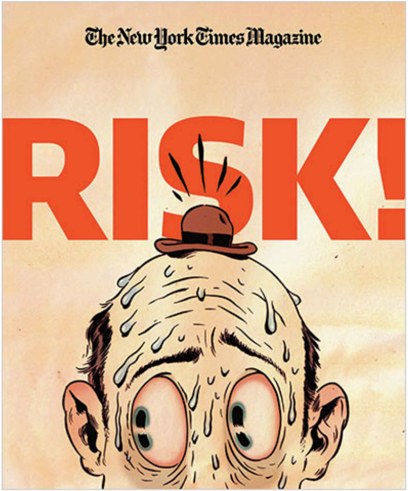
$$f(\mathbf{x}) = a_0 + \mathbf{a}_1^\top \mathbf{x} + \mathbf{x}^\top A_2 \mathbf{x} + \mathcal{A}_3(\mathbf{x}, \mathbf{x}, \mathbf{x}) + \cdots + \mathcal{A}_d(\mathbf{x}, \dots, \mathbf{x}) + \cdots .$$

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

- **Examples:** Taylor expansion, asymptotic expansion, Edgeworth expansion.
- a_0 scalar, $\mathbf{a}_1$ vector, A_2 matrix, $\mathcal{A}_d$ tensor of order d.
- **Lesson:** Important to look beyond the quadratic term.

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.

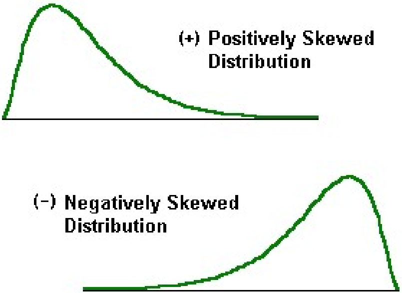
“Why do people measure risks against events that took place in 1987?” he asked, referring to Black Monday, the October day when the U.S. market lost more than 20 percent of its value and has been used ever since as the worst-case scenario in many risk models. “Why is that a benchmark? I call it future-blindness.

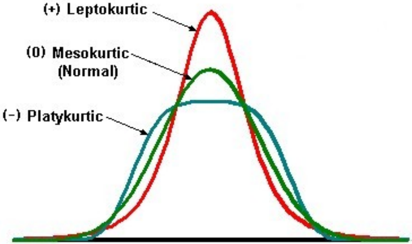
“If you have a pilot flying a plane who doesn’t understand there can be storms, what is going to happen?” he asked. “He is not going to have a magnificent flight. Any small error is going to crash a plane. This is why the crisis that happened was predictable.”

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

Cumulants

- **Univariate distribution:** First four cumulants are
  - mean $\mathcal{K}_1(x) = \mathrm{E}(x) = \mu$,
  - variance $\mathcal{K}_2(x) = \mathrm{Var}(x) = \sigma^2$,
  - skewness $\mathcal{K}_3(x) = \sigma^3 \, \mathrm{Skew}(x)$,
  - kurtosis $\mathcal{K}_4(x) = \sigma^4 \, \mathrm{Kurt}(x)$.





- **Multivariate distribution:** Covariance matrix *partly* describes the dependence structure — enough for Gaussian. Cumulants describe higher order dependence among random variables.

Cumulants

- For multivariate $\mathbf{x}$, $\mathcal{K}_d(\mathbf{x}) = [\![\kappa_{j_1\cdots j_d}(\mathbf{x})]\!]$ are symmetric tensors of order d.
- In terms of Edgeworth expansion,

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

  $\alpha = (j_1, \dots, j_n)$ is a multi-index, $\mathbf{t}^\alpha = t_1^{j_1}\cdots t_n^{j_n}$, $\alpha! = j_1!\cdots j_n!$.
- Provide a natural measure of non-Gaussianity: If $\mathbf{x}$ Gaussian,

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

- Gaussian assumption equivalent to quadratic approximation.
- **Non-Gaussian data:** Not enough to look at just mean and covariance.

Tensors inevitable in multivariate problems

- Mathematics
  - **Derivatives of univariate functions:** $f : \mathbb{R} \to \mathbb{R}$ smooth, $f'(x), f''(x), \dots, f^{(k)}(x) \in \mathbb{R}$.
  - **Derivatives of multivariate functions:** $f : \mathbb{R}^n \to \mathbb{R}$ smooth, $\operatorname{grad} f(\mathbf{x}) \in \mathbb{R}^n, \operatorname{Hess} f(\mathbf{x}) \in \mathbb{R}^{n \times n}, \dots, D^{(k)} f(\mathbf{x}) \in \mathbb{R}^{n \times \dots \times n}$.
- Statistics
  - **Cumulants of random variables:** $\mathcal{K}_d(x) \in \mathbb{R}$.
  - **Cumulants of random vectors:** $\mathcal{K}_d(\mathbf{x}) = [\![\kappa_{j_1 \cdots j_d}(\mathbf{x})]\!] \in \mathbb{R}^{n \times \dots \times n}$.
- Physics
  - **Hooke's law in 1D:** x extension, F force, k spring constant,

    $$F = -kx.$$

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

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

Humans cannot understand tensors

Humans cannot make sense out of more than $O(n)$ numbers. For most people, $5 \leq n \leq 9$ [Miller, 1956].

- VaR: single number
  - Readily understandable.
  - Not sufficiently informative and discriminative.
- Covariance matrix: $O(n^2)$ numbers
  - Hard to make sense of without further processing.
  - For symmetric matrices, may perform eigenvalue decomposition.
  - Basis for PCA, MDS, ISOMAP, LLE, Laplacian Eigenmap, etc.
  - Used in clustering, classification, dimension reduction, feature identification, learning, prediction, visualization, etc.
- Cumulant of order d: $O(n^d)$ numbers
  - How to make sense of these?
  - Want analogue of 'eigenvalue decomposition' for symmetric tensors.
  - Principal Cumulant Component Analysis: finding components that simultaneously account for variation in cumulants of all orders (cf. Jason Morton's talk).

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: $d > 2$.

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.

Probably the source

Woldemar Voigt, *Die fundamentalen physikalischen Eigenschaften der Krystalle in elementarer Darstellung*, Verlag Von Veit, Leipzig, 1898.

"An abstract entity represented by an array of components that are functions of co-ordinates such that, under a transformation of co-ordinates, the new components are related to the transformation and to the original components in a **definite way**."

Definite way: multilinear matrix multiplication

- Correspond to change-of-bases transformations for tensors.
- 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: $\mathcal{A} \in \mathbb{R}^{l\times m\times n}$, $X \in \mathbb{R}^{p\times l}$, $Y \in \mathbb{R}^{q\times m}$, $Z \in \mathbb{R}^{r\times n}$,

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

- Define 'right' (covariant) multiplication by $(X, Y, Z) \cdot \mathcal{A} = \mathcal{A} \cdot (X^\top, Y^\top, Z^\top)$.

Not every 3-array of numbers is a 3-tensor

- 3-way array: data structure; 3-tensor: algebraic object.
- Saying that a measured or observed 3-array of numbers is a 3-tensor is a modeling process.
- Should have some reason to believe that these numbers transform as expected under change-of-bases, i.e. via multilinear matrix multiplications.
- **Not a 3-tensor:**
  - Take $n \times 3n$ matrix representing a linear operator from an $3n$-dimensional vector space to an n-dimensional vector space and write it as $n \times n \times n$ array of numbers.
  - iPod sales figures stored in a ZIP code-by-model number-by-month array.
  - Phone directory — page-by-row-by-column of phone numbers.

Tensor modeling in physics

- **Hooke's law revisited:** At a point $\mathbf{x} = (x_1, x_2, x_3)^\top$ in a linear anisotropic solid,

$$\sigma_{ij} = \sum\nolimits_{k,l=1}^{3} c_{ijkl}\gamma_{kl} - \sum\nolimits_{k=1}^{3} b_{ijk}e_k - ta_{ij}$$

  where elasticity tensor $\mathcal{C} \in \mathbb{R}^{3\times 3\times 3\times 3}$, piezoelectric tensor $\mathcal{B} \in \mathbb{R}^{3\times 3\times 3}$, thermal tensor $A \in \mathbb{R}^{3\times 3}$, stress $\Sigma \in \mathbb{R}^{3\times 3}$, strain $\Gamma \in \mathbb{R}^{3\times 3}$, electric field $\mathbf{e} \in \mathbb{R}^3$, temperature change $t \in \mathbb{R}$.

- **Invariant under change-of-coordinates:** If $\mathbf{y} = Q\mathbf{x}$, then

$$\bar{\sigma}_{ij} = \sum\nolimits_{k,l=1}^{3} \bar{c}_{ijkl}\bar{\gamma}_{kl} - \sum\nolimits_{k=1}^{3} \bar{b}_{ijk}\bar{e}_k - t\bar{a}_{ij}$$

  where

$$\overline{\mathcal{C}} = (Q, Q, Q, Q)\cdot\mathcal{C}, \quad \overline{\mathcal{B}} = (Q, Q, Q)\cdot\mathcal{B}, \quad \overline{A} = (Q, Q)\cdot A,$$
$$\overline{\Sigma} = (Q, Q)\cdot\Sigma, \quad \overline{\Gamma} = (Q, Q)\cdot\Gamma, \quad \bar{\mathbf{e}} = Q\mathbf{e}.$$

Tensor modeling in statistics

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

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

Tensor modeling in computer science

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

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

  where $E_{ij} = \mathbf{e}_i \mathbf{e}_j^{\top} \in \mathbb{R}^{n\times n}$. Let

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

- $\mathcal{T}$ is a tensor of order 3.
- $O(n^{2+\varepsilon})$ algorithm for multiplying two $n \times n$ matrices gives $O(n^{2+\varepsilon})$ algorithm for solving system of n linear equations [Strassen, 1969].
- **Conjecture.** $\log_2(\operatorname{rank}_{\otimes}(\mathcal{T})) \leq 2 + \varepsilon$.

How do tensors arise in modeling?

- Affinity or dissimilarity of triples of objects: symmetric tensors.
  - **Example:** Amit Singer's dissimilarity metric from cryo-EM and NMR applications. $\mathcal{A} = [\![a_{ijk}]\!] \in \mathsf{S}^3(\mathbb{R}^n)$ where

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

  May assume, for simplicity, $a_{ijk} = w_{ij}w_{jk}w_{ki}$ for some nonnegative matrix $W = [w_{ij}] \in \mathsf{S}^2(\mathbb{R}^n)$.
- Measure of higher order dependence: symmetric tensors.
  - **Example:** Cumulants.
- Comparisons of triples of objects: skew-symmetric tensors.
  - **Example:** Triplewise rankings.
- Multilinearity: tensors.
  - **Example:** If all but one factors are kept constant and the quantity you are measuring varies linearly with the changing factor, then that quantity can be modeled by a tensor.

Analyzing tensors

- $A \in \mathbb{R}^{m \times n}$.
  - **Singular value decomposition:**
  $$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i$$
  where $\operatorname{rank}(A) = r$, U, V orthonormal columns, $\Sigma = \operatorname{diag}[\sigma_1, \dots, \sigma_r]$.
- $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$. Can either keep diagonality of Σ or orthogonality of U and V but not both.
  - **Linear combination:**
  $$\mathcal{A} = (X, Y, Z) \cdot \Sigma = \sum_{i=1}^{r} \sigma_i \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i$$
  where $\operatorname{rank}_\otimes(A) = r$, X, Y, Z matrices, $\Sigma = \operatorname{diag}_{r \times r \times r}[\sigma_1, \dots, \sigma_r]$; r may exceed n.
  - **Multilinear combination:**
  $$\mathcal{A} = (U, V, W) \cdot \mathcal{C} = \sum_{i,j,k=1}^{r_1, r_2, r_3} c_{ijk} \mathbf{u}_i \otimes \mathbf{v}_j \otimes \mathbf{w}_k$$
  where $\operatorname{rank}_\boxplus(A) = (r_1, r_2, r_3)$, U, V, W orthonormal columns, $\mathcal{C} = [\![c_{ijk}]\!] \in \mathbb{R}^{r_1 \times r_2 \times r_3}$; $r_1, r_2, r_3 \le n$.
  - Ensuing models in Psychometrics: CANDECOMP/PARAFAC and Tucker.

Other forms

- **Approximation theory:** Decomposing function into linear combination of separable functions,

$$f(x, y, z) = \sum_{i=1}^{r} \lambda_i \varphi_i(x) \psi_i(y) \theta_i(z).$$

  Application: separation of variables for PDEs.

- **Operator theory:** Decomposing operator into linear combination of Kronecker products,

$$\Delta_3 = \Delta_1 \otimes I \otimes I + I \otimes \Delta_1 \otimes I + I \otimes I \otimes \Delta_1.$$

  Application: numerical operator calculus (cf. talks by Greg Beylkin, Martin Mohlenkamp).

Other forms

- **Commutative algebra:** Decomposing homogeneous polynomial into linear combination of powers of linear forms,

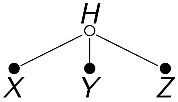
$$p_d(x, y, z) = \sum_{i=1}^{r} \lambda_i (a_i x + b_i y + c_i z)^d.$$

Application: independent components analysis (cf. talks by Philip Regalia, Lieven De Lathauwer).

- **Probability theory:** Decomposing probability density into conditional densities of random variables satisfying naïve Bayes:

$$\Pr(x, y, z) = \sum_{h} \Pr(h) \Pr(x \mid h) \Pr(y \mid h) \Pr(z \mid h).$$

Application: probabilistic latent semantic indexing (cf. talks by Inderjit Dhillon, Haesun Park, Bob Plemmons).



Multilinear spectral theory

- Eigenvalues and eigenvectors of symmetric $A \in \mathbb{R}^{n \times n}$ are critical values and critical points of
$$\mathbf{x}^\top A \mathbf{x} / \|\mathbf{x}\|_2^2.$$
- Define eigenvalues/vectors of symmetric tensor $\mathcal{A}$ as critical values/points of
$$\mathcal{A}(\mathbf{x}, \dots, \mathbf{x}) / \|\mathbf{x}\|_p^p.$$
    - Liqun Qi independently defined essentially the same notion in a different manner.
    - Falls outside Classical Invariant Theory — not invariant under $Q \in \mathrm{O}(n)$, ie. $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.
- Define singular values/vectors of tensor $\mathcal{A}$ as critical values/points of
$$\frac{\mathcal{A}(\mathbf{u}, \mathbf{v} \dots, \mathbf{z})}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \cdots \|\mathbf{z}\|_2}.$$
    - $\sigma_{\max}(\mathcal{A})$ equals spectral norm of $\mathcal{A}$.

Inherent difficulty

- The best r-term approximation problem for tensors has no solution in general (except for the nonnegative case).
- Eugene Lawler: "The Mystical Power of Twoness."
  - 2-SAT is easy, 3-SAT is hard;
  - 2-dimensional matching is easy, 3-dimensional matching is hard;
  - 2-body problem is easy, 3-body problem is hard;
  - 2-dimensional Ising model is easy, 3-dimensional Ising model is hard.
- Applies to tensors too:
  - 2-tensor rank is easy, 3-tensor rank is hard;
  - 2-tensor spectral norm is easy, 3-tensor spectral norm is hard;
  - 2-tensor approximation is easy, 3-tensor approximation is hard;
  - 2-tensor eigenvalue problem is easy, 3-tensor eigenvalue problem is hard.

Functions and operators on graph

$G = (V, E)$ undirected graph.

- **Functions**
  - **vertices:** $s : V \to \mathbb{R}$, $s(i) = s_i$;
  - **edges:** $X : V \times V \to \mathbb{R}$, $X(i,j) = X_{ij} = 0$ if $\{i,j\} \notin E$,
  $$X_{ij} = -X_{ji};$$
  - **triangles:** $\Phi : V \times V \times V \to \mathbb{R}$, $\Phi(i,j,k) = \Phi_{ijk} = 0$ if $\{i,j,k\} \notin T$,
  $$\Phi_{ijk} = \Phi_{jki} = \Phi_{kij} = -\Phi_{jik} = -\Phi_{ikj} = -\Phi_{kji}.$$
- **Operators**
  - $\operatorname{grad} : L^2(V) \to L^2(E)$, $\operatorname{grad} s(i,j) = s_j - s_i$;
  - $\operatorname{curl} : L^2(E) \to L^2(T)$, $\operatorname{curl} X(i,j,k) = X_{ij} + X_{jk} + X_{ki}$;
  - $\operatorname{div} : L^2(E) \to L^2(V)$, $\operatorname{div} X(i) = \sum_j w_{ij} X_{ij}$;
  - **graph Laplacian:** $\Delta_0 : L^2(V) \to L^2(V)$,
  $$\Delta_0 = \operatorname{div} \circ \operatorname{grad};$$
  - **graph Helmholtzian:** $\Delta_1 : L^2(E) \to L^2(E)$,
  $$\Delta_1 = \operatorname{curl}^* \circ \operatorname{curl} - \operatorname{grad} \circ \operatorname{div}.$$

Ranking with tensors

Theorem (Helmholtz decomposition)

Let $G = (V, E)$ *be an undirected, unweighted graph and* Δ_1 *its Helmholtzian. The space of edge flows on* G *admits an orthogonal decomposition*

$$L^2(E) = \operatorname{im}(\operatorname{grad}) \oplus \ker(\Delta_1) \oplus \operatorname{im}(\operatorname{curl}^*).$$

Furthermore, $\ker(\Delta_1) = \ker(\operatorname{curl}) \cap \ker(\operatorname{div})$.

- For each triangle $\{i, j, k\}$, $\operatorname{curl}(X)(i, j, k)$ measures inconsistency along the loop $i \to j \to k \to i$.
- Bottomline: resolve aggregated pairwise rankings $X \in L^2(E)$ into

$$X = \operatorname{grad} s + H + \operatorname{curl}^* \Phi.$$

  - s gives us a global ranking of the alternatives;
  - the residual $X - \operatorname{grad} s$ is a certificate of reliability for s;
  - sizes of H and $\operatorname{curl}^* \Phi$ tell us whether the inconsistencies are of a global or local nature.

Kernel learning with tensors?

- Given training data (x_i, y_i), $i = 1, \dots, n$, want to 'learn' target functions
  - $f : \{\text{e-mails}\} \to \{-1, 1\}$, $f(x) = -1$ if x is spam, $f(x) = 1$ otherwise;
  - $g : \{\text{SNPs}\} \to [-1, 1]$, $g(x) =$ likelihood that x plays a role in diabetes;
  - $h : \{\text{hand-written digits}\} \to \{0, 1, 2, \dots, 9\}$.
- Take Galerkin approach:
  - assume

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i),$$

    K Mercer kernel, e.g. $K(x, y) = \exp(-\|x - y\|^2/\sigma^2)$;
  - solve regularized least-squares for $\alpha_1, \dots, \alpha_n$,

$$\min \sum_{i=1}^{n} [y - f(x_i)]^2 + \lambda \|f\|^2.$$

- Work in progress (with Jason Morton): extend this to symmetric nuclear forms

$$K(x, y, z) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x) \varphi_k(y) \varphi_k(z).$$