Principal Cumulant Component Analysis

Lek-Heng Lim

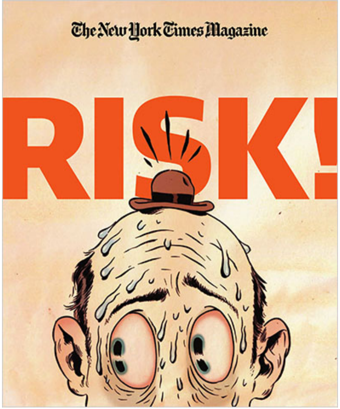
University of California, Berkeley

July 4, 2009

(Joint work with: Jason Morton, Stanford University)

Blaming the math

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



- **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:

$$\infty = 2.$$

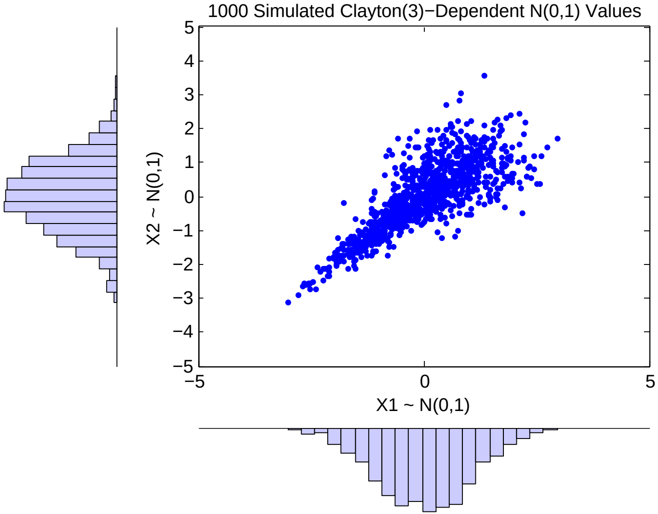
- If $\mathbf{x}$ is multivariate Gaussian, then

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

- $\mathcal{K}_1(\mathbf{x})$ mean, $\mathcal{K}_2(\mathbf{x})$ (co)variance, $\mathcal{K}_3(\mathbf{x})$ (co)skewness, $\mathcal{K}_4(\mathbf{x})$ (co)kurtosis,. . . .
- **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.



Why not VaR

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

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

  $\operatorname{grad} f(\mathbf{x}) \in \mathbb{R}^n$, $\operatorname{Hess} f(\mathbf{x}) \in \mathbb{R}^{n \times n}$, $\dots$, $D^{(k)} f(\mathbf{x}) \in \mathbb{R}^{n \times \cdots \times n}$.
- **Hooke's law in 1D:** x extension, F force, k spring constant,

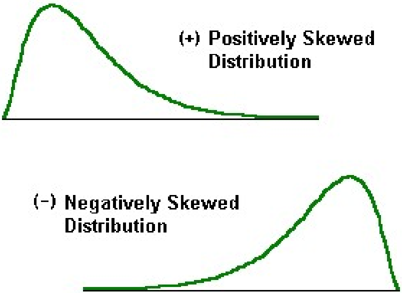
$$F = -kx.$$

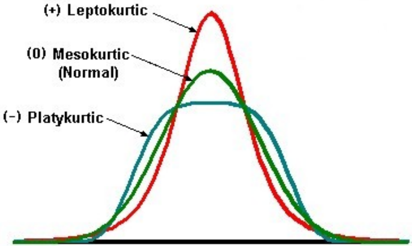
- **Hooke's law in 3D:** $\mathbf{x} = (x_1, x_2, x_3)$, elasticity tensor $\mathcal{C} \in \mathbb{R}^{3\times 3\times 3\times 3}$, stress $\Sigma \in \mathbb{R}^{3\times 3}$, strain $\Gamma \in \mathbb{R}^{3\times 3}$

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

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.

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

Tensors as hypermatrices

- Choose bases, ignore contra/covariance, write $\mathbf{A} \in U \otimes V \otimes W$ as

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

- **Matrix rank.** $A \in \mathbb{R}^{m \times n}$.
  - $\operatorname{rank}(A) = \min\{r \mid A = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{v}_i^\top\}$
    $= \dim(\operatorname{span}_{\mathbb{R}}\{A_{\bullet 1}, \dots, A_{\bullet n}\}) = \dim(\operatorname{span}_{\mathbb{R}}\{A_{1\bullet}, \dots, A_{m\bullet}\})$.
- **Tensor rank.** $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$.
  - **outer-product rank:** $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} := [\![u_i v_j w_k]\!]_{i,j,k=1}^{l,m,n}$,

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

  - **multilinear rank:** generalizes row and column ranks,

$$\operatorname{rank}_{\boxplus}(\mathcal{A}) = (r_1(\mathcal{A}), r_2(\mathcal{A}), r_3(\mathcal{A})).$$

- Generalizing $A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i$: either keep Σ diagonal or U, V orthonormal but not both.

Humans cannot understand ‘raw’ tensors

Humans cannot make sense out of more than $O(n)$ numbers. For most people, $5 \leq n \leq 9$ [Miller; '56].

- VaR: single number
  - Readily understandable.
  - Not sufficiently informative and discriminative.
- Covariance matrix: $O(n^2)$ numbers
  - Hard to make sense of without further processing.
  - Eigenvalue decomposition: PCA, MDS, ISOMAP, LLE, Laplacian Eigenmap, etc.
- Cumulant of order d: $O(n^d)$ numbers
  - How to make sense of these?
  - Want analogue of ‘eigenvalue decomposition’ for symmetric tensors.

SVD for tensors

- **Linear combination** of decomposable tensors

$$\mathcal{A} = (X, Y, Z) \cdot \Sigma = \sum\nolimits_{i=1}^{r} \sigma_i \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i.$$

  - **Computational complexity:** Strassen matrix multiplication/inversion

$$\inf\{\omega \mid \operatorname{rank}_\otimes(\textstyle\sum_{i,j,k=1}^{n} \varphi_{ik} \otimes \varphi_{kj} \otimes E_{ij}) = O(n^\omega)\} = 2?$$

  - **Quantum computing:** algebraic measure of entanglement

$$|\mathrm{GHZ}\rangle = |0\rangle \otimes |0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle \otimes |1\rangle \in \mathbb{C}^{2\times 2\times 2}.$$

  - **Geometry:** secant varieties of Segre and Veronese varieties.

- **Multilinear combination** of orthonormal U, V, W

$$\mathcal{A} = (U, V, W) \cdot \mathcal{C} = \sum\nolimits_{i,j,k=1}^{r_1,r_2,r_3} c_{ijk} \mathbf{u}_i \otimes \mathbf{v}_j \otimes \mathbf{w}_k.$$

  - **Geometry:** subspace varieties, symmetric subspace varieties

$$\mathrm{Gr}(l, p) \times \mathrm{Gr}(m, q) \times \mathrm{Gr}(n, r) \times \mathbb{R}^{p\times q\times r} \to \mathrm{Sub}_{p,q,r}(\mathbb{R}^l, \mathbb{R}^m, \mathbb{R}^n).$$

Eliminating the impossible

- Computing 3-tensor rank is NP hard [Håstad; 1990].
- Just about every tensor problem is NP hard in both the Cook-Karp-Levin and the Blum-Shub-Smale sense [L & Hillar; 2009]:
  - best rank-1 approximation of a 3-tensor;
  - best rank-1 approximation of a symmetric 3-tensor;
  - singular values/vectors of a 3-tensor [L; 2005];
  - eigenvalues/vectors of a symmetric 3-tensor [L; 2005], [Qi; 2005];
  - spectral norm of a 3-tensor;
  - feasibility of a system of bilinear equations;
  - solving a system of bilinear equations in both the exact and least squares sense.
- Best rank-r tensor approximation problems are unsolvable in general [de Silva & L; 2008], [Comon, Golub, L, Mourrain; 2008].

Among whatever remains

- **Principal Component Analysis:** components accounting for variation in covariance.
- **Principal Cumulant Component Analysis:** components accounting for variation in all cumulants simultaneously [L & Morton; 2008], [Morton & L; 2009],

$$\min_{Q \in \mathrm{O}(n,r),\, \mathcal{C}_p \in \mathsf{S}^p(\mathbb{R}^r)} \sum_{p=1}^{\infty} \alpha_p \|\hat{\mathcal{K}}_p(\mathbf{y}) - (Q, \dots, Q) \cdot \mathcal{C}_p\|_F^2.$$

- Surprising relaxation: optimization over a single Grassmannian $\mathrm{Gr}(n, r)$ of dimension $r(n - r)$,

$$\max_{Q \in \mathrm{Gr}(n,r)} \sum_{p=1}^{\infty} \alpha_p \|(Q^\top, \dots, Q^\top) \cdot \hat{\mathcal{K}}_p(\mathbf{y})\|_F^2.$$

- Efficient algorithm exists: limited memory BFGS on Grassmannian [Savas & L; 2009].

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 component analysis

Linear generative model:

$$\mathbf{y} = A\mathbf{s} + \varepsilon.$$

Principal component 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 component 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].

Principal cumulant component analysis

- 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

$$\min_{Q \in \mathrm{O}(n,r),\, \mathcal{C}_p \in \mathsf{S}^p(\mathbb{R}^r)} \sum_{p=1}^{\infty} \alpha_p \|\hat{\mathcal{K}}_p(\mathbf{y}) - (Q, \dots, Q) \cdot \mathcal{C}_p\|_F^2,$$

- We have assumed $A = Q \in \mathrm{O}(n, r)$ since otherwise $A = QR$ and

$$\mathcal{K}_p(A\mathbf{s}) = (Q, \dots, Q) \cdot [(R, \dots, R) \cdot \mathcal{K}_p(\mathbf{s})].$$

- Recover orthonormal basis of subspace spanned by A.
- $\mathcal{C}_p \approx (R, \dots, R) \cdot \hat{\mathcal{K}}_p(\mathbf{s})$ not necessarily diagonal.

Newton/quasi-Newton on a Grassmannian

- Objective $\Phi : \mathrm{Gr}(n, r) \to \mathbb{R}$.
- $\mathbf{T}_X$ tangent space at $X \in \mathrm{Gr}(n, r)$

$$\mathbb{R}^{n \times r} \ni \Delta \in \mathbf{T}_X \qquad \Longleftrightarrow \qquad \Delta^\top X = 0$$

1. Compute Grassmann gradient $\nabla\Phi \in \mathbf{T}_X$.
2. Compute Hessian or update Hessian approximation

$$H : \Delta \in \mathbf{T}_X \to H\Delta \in \mathbf{T}_X.$$

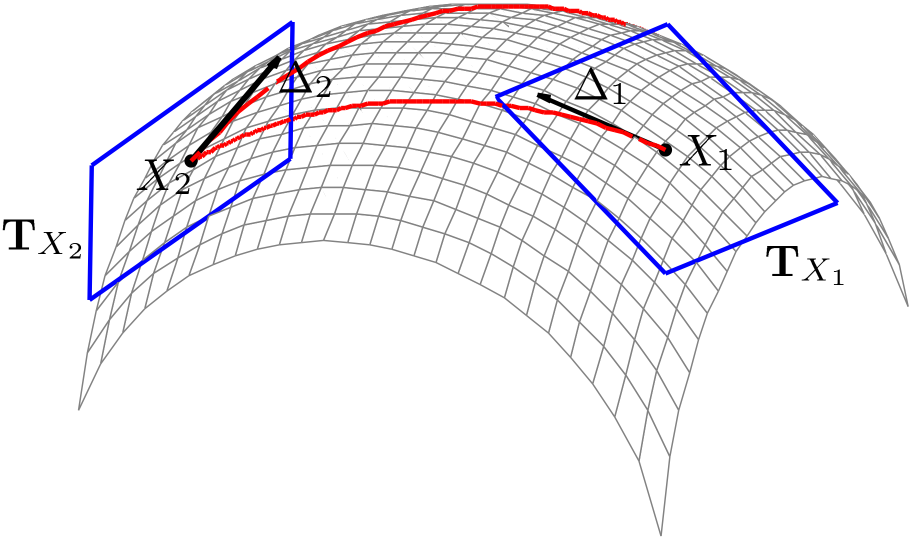
3. At $X \in \mathrm{Gr}(n, r)$, solve

$$H\Delta = -\nabla\Phi$$

   for search direction Δ.
4. Update iterate X: Move along geodesic from X in the direction given by Δ.

- [Arias, Edelman, Smith; 1999], [Savas & L.; 2009].

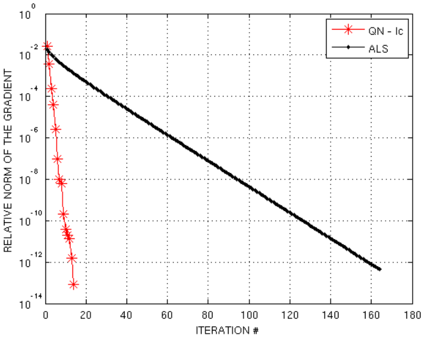
Picture

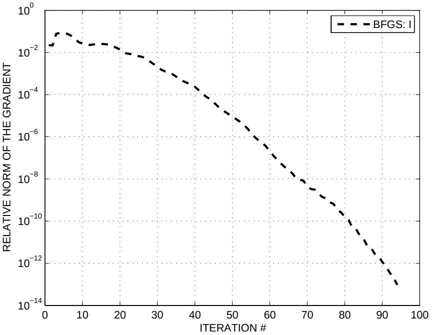


Convergence

Left: $\|(X,X,X)\cdot\mathcal{S}_3\|^2$. Compares favorably with Alternating Least Squares.

Right: $\frac{1}{2!}\|(X,X)\cdot S_2\|^2+\frac{1}{3!}\|(X,X,X)\cdot\mathcal{S}_3\|^2+\frac{1}{4!}\|(X,X,X,X)\cdot\mathcal{S}_4\|^2$.



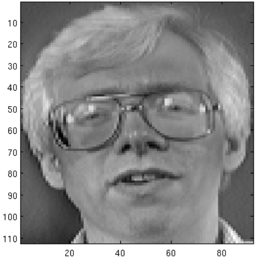


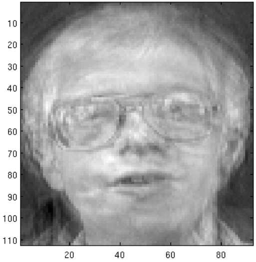
Skew eigenfaces

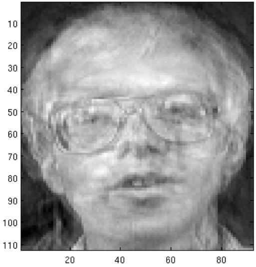
Left: Original.

Center: 30 variance eigenvectors.

Right: 20 variance eigenvectors and 10 skewness eigenvectors.







Higher order portfolio optimization

$$\min \sum_{d=2}^{n} \alpha_d(\mathbf{x}^\top, \dots, \mathbf{x}^\top) \cdot \mathcal{K}_d(\mathbf{y}) \qquad \text{s.t.} \qquad \mathbf{x}^\top \mathrm{E}(\mathbf{y}) > \underline{r}.$$

- $n = 2$: Markowitz mean-variance optimal portfolio theory.
- $n = 4$: mean-variance-skewness-kurtosis optimal portfolio theory.