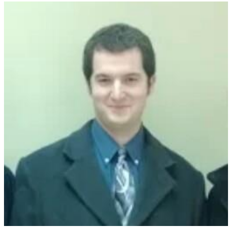
The centroid body: algorithms and statistical estimation for heavy-tailed distributions

Luis Rademacher

Athens GA, March 2016

Joint work with Joseph Anderson, Navin Goyal and Anupama Nandi



Challenge.

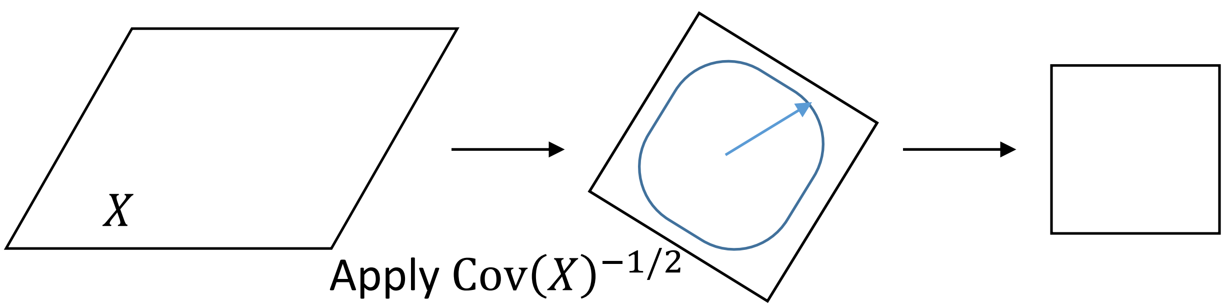
- Covariance matrix is frequently used in algorithmic statistical analysis of data.
- What if data is heavy-tailed?
  What if data seems to follow a distribution with infinite second moment?
- Our work:
  Finite first moment $\Rightarrow$ replace covariance matrix by centroid body in certain algorithmic application: Independent Component Analysis (ICA)

Independent Component Analysis (ICA)

- INPUT: samples $X^{(1)}, X^{(2)}, \ldots$ from random vector $X = AS$, where:
  - S is d-dimensional random vector with independent coordinates. Assume 0-mean for simplicity.
  - A is square invertible matrix.
- GOAL: estimate (directions of columns of) A.
- S, A are not observed. Distribution of S is unknown.

Example: How to learn a parallelepiped? [Frieze Jerrum Kannan]

- Illustrative case:
  Estimate a parallelepiped from uniformly random samples $X^{(1)}, X^{(2)}, \ldots$
  Model:
  S: uniform in axis aligned cube.
  $X = AS$: uniform in a parallelepiped
- By estimating covariance and applying $\mathrm{Cov}(X)^{-1/2}$, can assume it is a rotated cube centered at 0.
- To estimate rotation: Enumerate all local minima of directional 4th moment on unit sphere.
  Theorem: Normals to facets are a complete set of local minima.

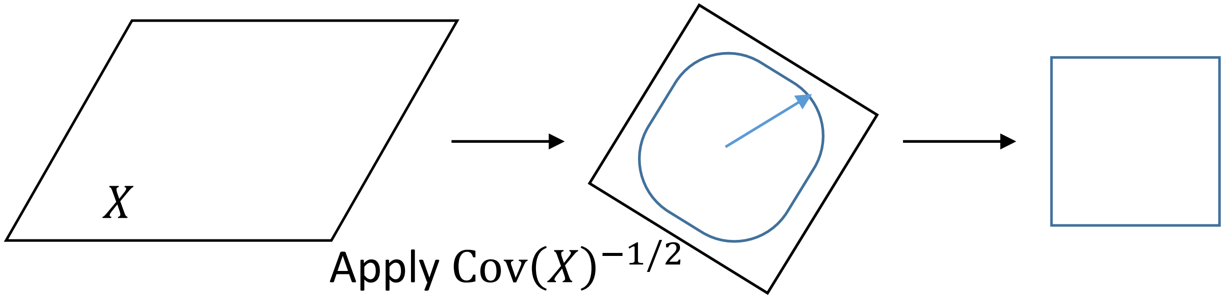


$$F(v) = E\left((X \cdot v)^4\right)$$

An ICA algorithm: unexpected usefulness of local optima

[Delfosse-Loubaton SignalProcessing95] [Frieze-Jerrum-Kannan FOCS96] [Hyvarinen IEEE NeuralNets99]

- Orthogonalization:
  - Apply a linear transformation to reduce to the case where A has orthogonal columns.
  - Implemented by multiplying samples by estimated $\mathrm{Cov}(X)^{-1/2}$.
- Recover rotation (simplified):
  - Enumerate all local minima of directional 4th moment on unit sphere. $\pm$columns of rotation are a complete set of local minima.
- **All previously known provably efficient ICA methods require at least 4 moments.**



$$F(v) = E\left((X \cdot v)^4\right)$$

Heavy-tailed ICA

- All previously known provably efficient ICA methods require at least 4 moments.
- Heavy-tailed distribution ≈ no moments or only a few moments exists.
- Heavy-tailed ICA instances appear naturally in speech and financial data.
- [Anderson Goyal Nandi R.]
  - **Preprocessing:**
    Gaussian damping.
    A provably efficient algorithm that works with no moment assumption when the unknown matrix A is unitary.
  - **Preprocessing:**
    Gaussian damping +
    centroid body orthogonalization.
    A provably efficient algorithm that works assuming finite 1^{st} moment, for any matrix.

Orthogonalization

- For distributions with infinite second moment, $\mathrm{Cov}(X)$ does not make sense. Instead:
- **Orthogonalization:** Given ICA model $X = AS$, find matrix B such that BA has orthogonal columns.
- Idea: think of Legendre's ellipsoid of inertia, having support function

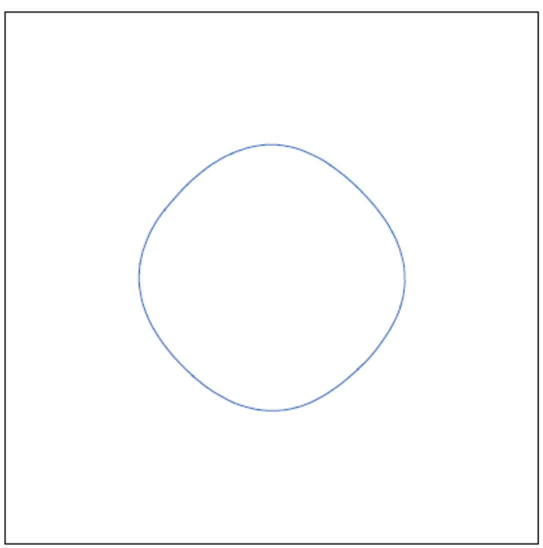
$$h(y) = \sqrt{E(X \cdot y)^2} = \sqrt{y^T Cov(X) y}$$

(Unique ellipsoid having the same covariance matrix as X, up to a constant factor)

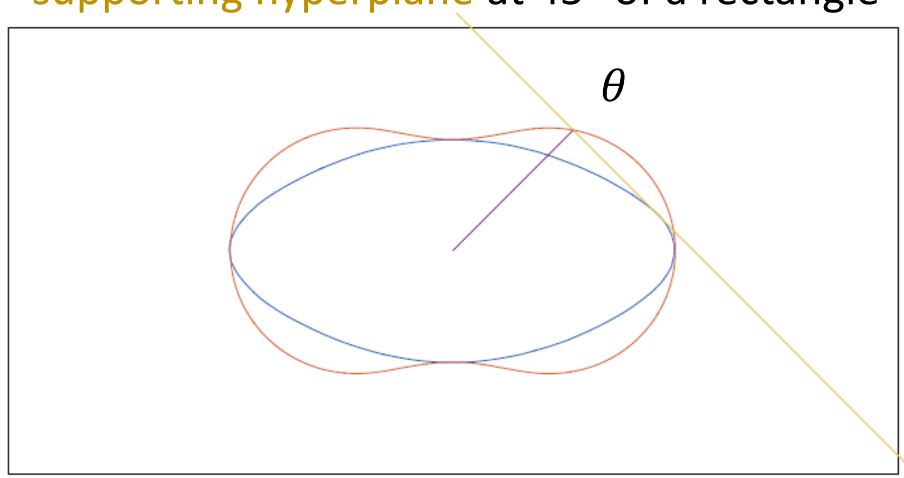
Orthogonalization via the **centroid body**

- **Definition** (Petty 1961):
  Given random vector X, the centroid body of X, denoted ΓX, is the convex body with support function
  $$h_{\Gamma X}(\theta) = E(|X \cdot \theta|).$$

centroid body of $[-1,1]^2$

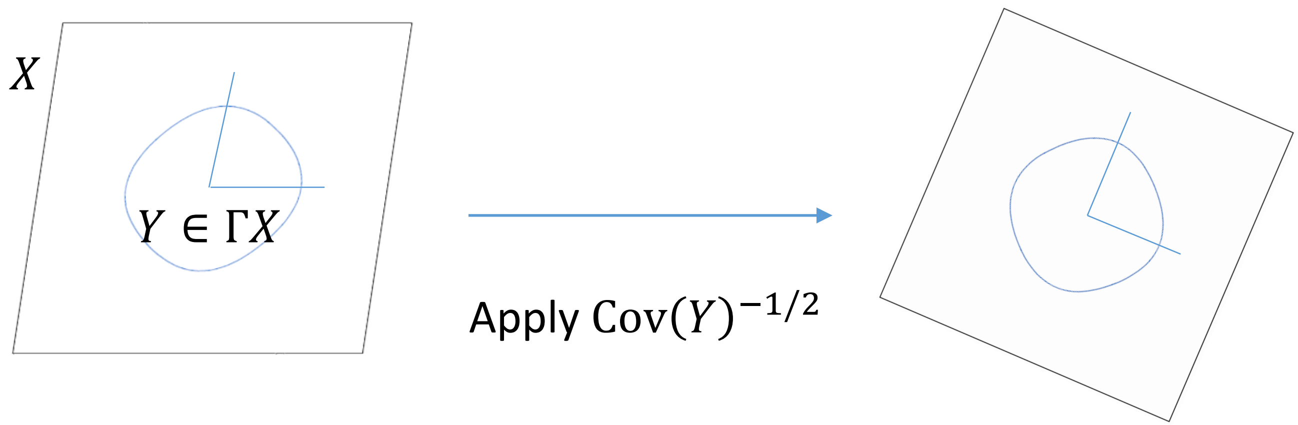


centroid body, support function and supporting hyperplane at 45^0 of a rectangle



Orthogonalization via the **centroid body**

- Idea: Replace covariance of X in orthogonalization step by covariance of uniform distribution in centroid body of X.



Orthogonalization via the **centroid body**

- What property of the ellipsoid of inertia makes the square root of its covariance an orthogonalizer?
- Centroid body ΓX, defined by support function $h_{\Gamma X}(y) = E(|X \cdot y|\ )$
  - If S has a product distribution and is symmetrically distributed, then it is unconditional, and therefore ΓS is unconditional (symmetric around axis-aligned hyperplane reflections).
  - **linear equivariant**: $\Gamma AX = A\Gamma X$, for any invertible matrix A.
- For an algorithm:
  - need to be able to estimate ΓX efficiently.
  - need efficient membership test.
- Trick: If S is not symmetrically distributed then ΓS may not be unconditional. But $\Gamma(S - S')$ is unconditional, as $S - S'$ is symmetrically distributed (where S' is an independent copy of S).

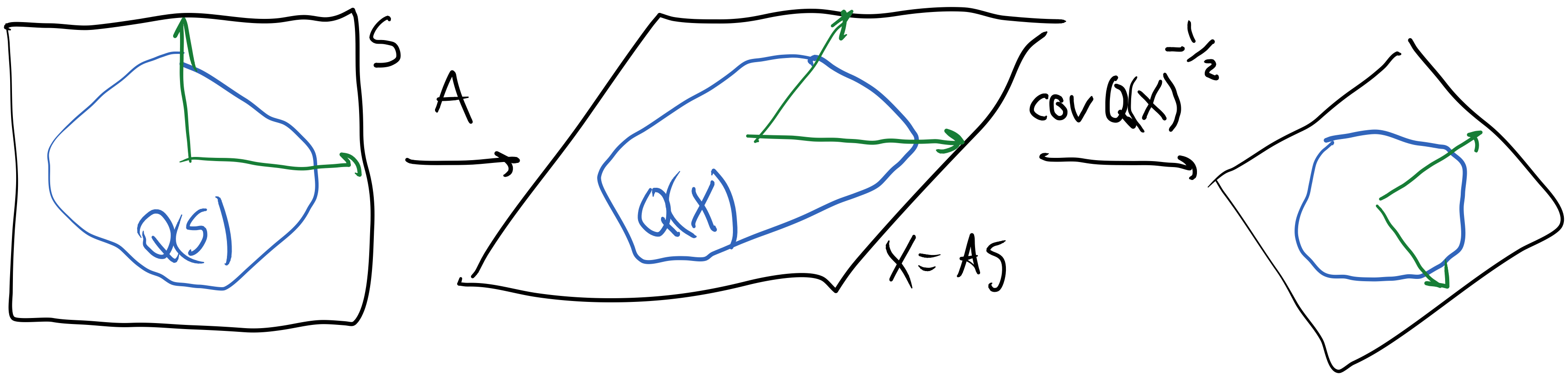
More generally: Lemma

- U: family of d-dim. product distributions.
- $\overline{U}$: closure of U under invertible linear transformations.
- For any $P \in \overline{U}$, pick a distribution $Q(P)$ (e.g. uniform in ΓP)
- If
  1. For all $P \in U$, $Q(P)$ is **unconditional**.
  2. Map Q is **linear equivariant**.
  3. $\mathrm{Cov}(Q(P))$ is positive definite for any $P \in \overline{U}$.
- Then for any ICA model $X = AS$ with $S \in U$ we have $\mathrm{Cov}\big(Q(P)\big)^{-1/2}$ is an orthogonalizer for X.

Proof Idea

1. For all $P \in U$, $Q(P)$ is unconditional.
2. Map Q is linear equivariant.
3. $\mathrm{Cov}(Q(P))$ is positive definite for any $P \in \overline{U}$.

- unconditional ⇒ covariance is diagonal
- unconditional ⇒ axes of $Q(S)$ aligned with axes of independence of S
- equivariance ⇒ axes of independence of transformed $X = AS$ aligned with axes of $Q(X)$.
  Orthogonalizing $Q(X)$ orthogonalizes X. $\mathrm{Cov}\big(Q(X)\big)^{-1/2}$ is an orthogonalizer for $Q(X)$ and therefore for X.



How to estimate $\mathrm{Cov}(\Gamma X)$?

- Use random points from ΓX.
- Given membership oracle for ΓX, use random walk-based methods to generate random points. We use [Kannan Lovasz Simonovits].
- Membership oracle for ΓX: Given finite $1 + \epsilon$ moments of X, can estimate support function of ΓX pointwise efficiently from samples. In theory, use ellipsoid algorithm to decide membership in ΓX from support function.
- More practical: Use "dual" (zonoid) expression of ΓX to get explicit linear program:

$$\Gamma X = \{E(\lambda(X)X)\colon -1 \leq \lambda(x) \leq 1, \lambda\colon R^n \to R\}$$

$$"\Gamma X = E[-X, X]"$$

After orthogonalization: Recover rotation?

- Model: $X = RS$, where R has orthogonal columns. We need no moment assumptions on S.
- **Gaussian Damping**:
  - Construct model $\tilde{X}$ by multiplying density of X by Gaussian e^{-x^2/R^2}, for suitable $R > 0$.
  - $\tilde{X}$ has same axes of independence as X and all moments of $\tilde{X}$ are finite.
  - Implemented by rejection sampling.
- Apply known higher moment-based ICA algorithm to $\tilde{X}$ (e.g. [Goyal Vempala Xiao]).

Orthogonalization with no moment assumption?

- Tempting: Use convex floating body of [Schütt and Werner] in place of centroid body.
  - Also linearly equivariant and unconditional when S is symmetric.
  - Appears to be computationally intractable. No efficient access to support function or membership.