Geodesic convexity & covariance estimation

Ami Wiesel

School of Engineering and Computer Science
Hebrew University of Jerusalem, Israel

June 28, 2013
Acknowledgments

- Teng Zhang (Princeton).
- Maria Greco (Università di Pisa).
- Ilya Soloveychik (Hebrew University).
- Alba Sloin (Hebrew University).
1. Geodesic convexity

2. Covariance estimation

3. Non Gaussian

4. Kronecker models

5. Symmetry constraints
Outline

1. Geodesic convexity
2. Covariance estimation
3. Non Gaussian
4. Kronecker models
5. Symmetry constraints
Convexity

Convex function

\[ f(x_t) \leq tf(x_1) + (1 - t)f(x_0) \]

\[ x_t = tx_1 + (1 - t)x_0 \]

- Local solutions are easy to find and globally optimal!
- Easy to generalize:
  - Building bricks: linear, quadratic, norms...
  - Rules: convex + convex = convex,...
Convex optimization with positive variables

**Power control** [Chiang:07]

\[
\begin{align*}
\text{minimize} & \quad \prod_{i=1}^{N} \frac{1}{1 + \text{SIR}_i} \\
\text{subject to} & \quad (2^{T}R_{i,min} - 1) \frac{1}{\text{SIR}_i} \leq 1, \quad \forall i, \\
& \quad (\text{SIR}_{th})^{N-1}(1 - P_{o,i,\text{max}}) \prod_{j \neq i}^{N} \frac{G_{ij}P_{j}}{G_{ii}P_{i}} \leq 1, \quad \forall i, \\
& \quad P_{i}(P_{i,\text{max}})^{-1} \leq 1, \quad \forall i.
\end{align*}
\]

Variables: powers.

**Circuit design** [Hershenson:01]

Variables: transistors widths, lengths, currents, capacitors,...

The Geometric Programming (GP) trick

- The above problems are non-convex.
- Can be convexified by a change of variables \( q_{i} = e^{Z_{i}} \).
Convexity with positive variables

- Exp: $e^{z_i}$ are convex in $z_i$.
- Log-sum-exp: $\log \sum_i e^{z_i}$ is convex in $z_i$.
- If $f(e^z)$ is convex in $z$ then $f(e^{z_1+z_2})$ is convex in $z_1, z_2$.
- $e^z$ transforms sums into products!

The Geometric Programming (GP) trick
- Minimize products of positive numbers $q_i \geq 0$ using $e^{z_i}$.
Today: GP with positive definite matrices $Q_i \succeq 0$

- Can we minimize powers $a^T Q^{\pm 1} a$?
- Can we minimize log determinants $\log|Q|$?
- Can we minimize products $Q_1 \otimes Q_2$?

- The answers are YES!
- But the solution is not a simple change of variables.
- Instead, we turn to geodesic convexity.
Revisiting the GP trick

Convexity

\[
\text{line}
\]
\[
f(tz_1 + (1 - t)z_0) \leq tf(z_1) + (1 - t)f(z_0)
\]

Geodesic convexity \( \tilde{f}(q) = f(\log q) \)

\[
\text{geodesic}
\]
\[
\tilde{f}(q_1^{t} q_0^{1-t}) \leq t\tilde{f}(q_1) + (1 - t)\tilde{f}(q_0)
\]
For any \( q_1, q_0 \in D \) we define a geodesic \( q_t \in D \) parameterized by \( t \in [0, 1] \).

A function \( f(q) \) is g-convex in \( q \in D \) if

\[
f(q_t) \leq tf(q_1) + (1 - t)f(q_0) \quad \forall \quad t \in [0, 1].
\]

**Properties**

- Any local minimizer of \( f(q) \) over \( D \) is a global minimizer.
- \( g\)-convex + \( g\)-convex = \( g\)-convex.
From scalars to matrices

- We do not know the matrix version of $e^x$.
- We do know how to generalize the geodesics $q_t = q_1^t q_0^{1-t}$.

Geodesic between $Q_0 \succ 0$ and $Q_1 \succ 0$

$$Q_t = Q_0^{\frac{1}{2}} \left( Q_0^{-\frac{1}{2}} Q_1 Q_0^{-\frac{1}{2}} \right)^t Q_0^{\frac{1}{2}}, \quad t \in [0, 1].$$
Theorem

The function

\[ f(Q) = a^T Q^{\pm 1} a \]

is g-convex in \( Q \succ 0 \).

Proof: eigenvalue decomposition reduces to scalar case.
Theorem

The function

\[ f(Q) = \log \left| \sum_{i=1}^{n} H_i Q H_i^T \right| \]

is g-convex in \( Q \succ 0 \).

- Similarly, \( \log |Q| \) is g-linear.
- Proof: eigenvalue decomposition reduces to scalar case.
Products (matrix case)

**Theorem**

If $f(W)$ is $g$-convex in $W \succ 0$, then

$$g(Q_1, \cdots, Q_n) = f(Q_1 \otimes Q_2 \otimes \cdots \otimes Q_n)$$

is $g$-convex in $Q_i \succ 0$.

- The operation $\otimes$ is a Kronecker product.

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1p}B \\ a_{21}B & a_{22}B & \cdots & a_{21}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}B & a_{p2}B & \cdots & a_{pp}B \end{bmatrix}$$
A set $S$ is $g$-convex if

$$Q_0, Q_1 \in S \implies Q_t \in S.$$ 

Local minimas over $g$-convex sets are global.

**Theorem**

For orthonormal $U$, the set $\{Q : Q = UQU^T\}$ is $g$-convex.

- Proof: Matrix commutativity properties $QU = UQ$.
- Trivial in scalar case.
Summary

- $a^T Q^{\pm 1} a$ is $g$-convex.

- $\log \left| \sum_{i=1}^{n} H_i Q H_i^T \right|$ is $g$-convex.

- $Q_i \otimes \cdots \otimes Q_j$ preserves $g$-convexity.

- $\{ Q : Q = UQU^T \}$ is $g$-convex.
Outline

1. Geodesic convexity
2. Covariance estimation
3. Non Gaussian
4. Kronecker models
5. Symmetry constraints
Covariance estimation

- $\mathbf{x}$: $p$-dimensional random vector.
- Mean $E\{\mathbf{x}\} = \mathbf{0}$, covariance $\Sigma = E [\mathbf{x}\mathbf{x}^T]$.
- $\{\mathbf{x}_i\}_{i=1}^n$: $n$ independent & identically distributed realizations.

Goal

- Problem: Derive $\hat{\Sigma} (\{\mathbf{x}_i\}_{i=1}^n)$ to estimate $\Sigma$.
- Solution: Maximum likelihood.
- Emphasis on the hard non-Gaussian and structured cases.
I work on other stuff: comm, radar, sensor networks...
I was told this can also be used with images [Zhang:2012].
Outline

1. Geodesic convexity
2. Covariance estimation
3. Non Gaussian
4. Kronecker models
5. Symmetry constraints
A popular robust covariance estimator

- Elliptical distributions, Spherically Invariant Random processes, Compound Gaussian, Multivariate Student, etc..

\[
\begin{bmatrix}
\vdots \\
x_i \\
\vdots
\end{bmatrix} = \sqrt{q_i} \begin{bmatrix}
\vdots \\
u_i \\
\vdots
\end{bmatrix}
\]
\[
\mathcal{N}(0, Q)
\]

- Non-convex ML via fixed point iteration:

\[
Q_{k+1} = \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^T}{x_i^T Q_k^{-1} x_i}
\]
A bit of background

\[ Q_{k+1} = \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^T}{x_i^T Q_k^{-1} x_i} \]

- [Tyler:87] Introduction, fixed point iteration, existence, uniqueness, convergence analysis.
- [Gini:95], [Conte:02] Analysis, array processing.
- [Chen:10] Regularization analysis via Perron Frobenius.

Lots of applications! Lots of difficult theory! But specific and hard to follow and generalize.
Revisiting Tyler’s estimator

The negative log likelihood is

\[ L(Q) = \frac{p}{n} \sum_{i=1}^{n} \log \left( x_i^T Q^{-1} x_i \right) + \log |Q| \]

- Non-convex optimization problem.
- 25 years of methods that converge to the global solution.

Theorem

[Auderset:05] The negative log likelihood is g-convex. Actually, jointly g-convex in \( q \) and \( Q \). Also for other elliptical distributions, e.g., MGGD.
Often, we need regularization / prior.

[Abramovich:07], [Chen:10] difficult design and analysis.

We propose to use g-convex regularization schemes

\[
\min L(\cdot) + \lambda h(\cdot)
\]

needs to be g-convex

Guaranteed to be g-convex, and can be solved efficiently. We can put priors on both the covariance and the scalings.
Prior knowledge on the scaling factors via g-convex functions:

- Bounded peak values $L \leq \log q_i \leq U$.
- Bounded second moments $\sum_i \log^2 q_i \leq U$.
- Sparsity (outliers) $\sum_i |\log q_i| \leq U$.
- Smooth time series $|\log q_i - \log q_{i-1}| \leq U$.

Without g-convexity [Bucciareli:96], [Wang:06], [Chitour:08].

We can also change variables and use convex penalties.
G-convex matrix penalties

- Shrinkage to identity ($\mathbf{T} = \mathbf{I}$) or arbitrary target
  \[
  h(Q) = p \log \left( \text{Tr} \left\{ Q^{-1} \mathbf{T} \right\} \right) + \log |Q|
  \]

- Shrinkage to diagonal
  \[
  h(Q) = \log \prod_{i=1}^{p} \left[ (Q^{-1})_{ii} \right] + \log |Q|
  \]

- Regularization of condition number
  \[
  h(Q) = \frac{\lambda_{\text{max}}(Q)}{\lambda_{\text{min}}(Q)}
  \]

Non-Gaussian versions of [Stoica:08], [Schafer:05], [Won:09].
Experiments

Shrink to diag Toeplitz $p = 10$

$\Sigma_{ij} = 0.4|i−j|$

Factor 2 on 1st cross validation

Condition number Toeplitz $p = 10$

$\Sigma_{ij} = 0.4|i−j|$

$\kappa = 4.98 \in [1, 10]$

cross validation

Correlated scalings Toeplitz $p = 10$

$\Sigma_{ij} = 0.8|i−j|$

MA(2) with

$\|Lz\|_2 \leq 7$. 
Outline

1. Geodesic convexity
2. Covariance estimation
3. Non Gaussian
4. Kronecker models
5. Symmetry constraints
Kronecker (separable, transposable) model $Q_1 \otimes Q_2$

- Estimating covariances of random $p_2 \times p_1$ matrices.
- A standard approach is to impose structure

$$X = Q_2^{\frac{1}{2}} W Q_1^{\frac{1}{2}}$$

- $W_{ij}$ are i.i.d. $\mathcal{N}(0, 1)$.
- $Q_2$ correlates the columns.
- $Q_1$ correlates the rows.

- In vector notations, $E[xx^T] = Q_1 \otimes Q_2$
- Examples: $Tx \otimes Rx$, products $\otimes$ costumers, etc...
A bit of background

\[ Q_1 \otimes Q_2 \]

- [Kermoal:02] Experiments in MIMO radio channels.
- [Lu:05], [Srivastava:08] Testing, uniqueness.
- [Tsiligkaridis:12] COMING UP COLLOQUIUM.

Lots of applications! Lots of difficult theory!
But specific and hard to follow and generalize.
Revisiting the Kronecker model

The Kronecker likelihood function is

\[ L(Q_1, Q_2) = \sum_{i=1}^{n} x_i^T (Q_1 \otimes Q_2)^{-1} x_i + \log |Q_1 \otimes Q_2| \]

- Non-convex optimization problem.
- 20 years of methods that converge to the global solution.

**Theorem**

The negative log likelihood is jointly g-convex in \( Q_1 \) and \( Q_2 \)!
Also holds for multiway models with \( Q_1 \otimes \cdots \otimes Q_n \).

Thus, every local minima is global, and we have lots of extensions.
Why is this helpful? Regularized ML

- Kronecker models do not require many samples.
- [Allen:10] one sample + regularization via SVD.
- We propose

\[
\min_{Q_1, Q_2} L(Q_1, Q_2) + \alpha \text{Tr} \left\{ Q_1^{-1} \right\} \text{Tr} \left\{ Q_2^{-1} \right\}
\]

which is jointly g-convex.

\[p_1 = p_2 = 5\]
\[\Sigma_{ij} = 0.8|i-j|\]
Why is this helpful? Non-Gaussian & Kronecker ML

- Just for fun: hybrid robust Kronecker model:

\[ q_i Q + Q_1 \otimes Q_2 \Rightarrow q_i \cdot Q_1 \otimes Q_2 \]

- We propose

\[
\min_{q, Q_1, Q_2} \sum_{i=1}^{n} x_i^T (q_i \cdot Q_1 \otimes Q_2)^{-1} x_i + \log |q_i \cdot Q_1 \otimes Q_2|
\]

which is jointly g-convex.

\[ p_1 = 10 \text{ and } p_2 = 2 \]
\[ \Sigma_{ij} = 0.8|i-j| \]
Outline

1. Geodesic convexity
2. Covariance estimation
3. Non Gaussian
4. Kronecker models
5. Symmetry constraints
Common symmetry constraints

\[ Q = UQU^T \quad \forall \quad U \in \mathcal{K} \]

Applications:

- Circulant, used for approximating Toeplitz = stationary
- Persymmetric, e.g., radar systems using a symmetrically spaced linear array with constant pulse repetition interval
More symmetry constraints - properness

Symmetry

\[ Q = UQU^T \quad \forall \quad U \in \mathcal{K} \]

Applications:

- **Complex normal** = double real normal \((CN_p = \mathcal{N}_{2p})\)
- Plus a symmetry constraint \(x \sim e^{i\theta} x\).

\[
\text{cov} \begin{bmatrix}
\text{Re}(x) \\
\text{Im}(x)
\end{bmatrix} = \begin{bmatrix}
A & B \\
-B & A
\end{bmatrix}
\]

- Recently, proper Gaussian quaternions \(x = a + ib + jc + kd\).
- For example, in radar with I/Q phase and polarizations
- Here too: \(QN_p = \mathcal{N}_{4p} + \text{special symmetry } x \sim e^{\nu\theta} x\).
A bit of background

\[ Q = UQU^T \]

- **Gaussian**
  - General symmetry groups [Shah & Chandrasekaran 2012]
  - Everybody knows proper complex (circularly symmetric)
  - Proper quaternion [Miron:06], [Bukhari:11], [Via:11]...

- **Non Gaussian**
  - Persymmetric [Pailloux:11]
  - Complex elliptical distributions [Bombrun:11], [Ollila:12]

Lots of applications! But specific and hard to follow and generalize. Easy in the Gaussian case (linear constraint).
Theorem

The set $Q = UQU^T$ is g-convex!

- Can be combined with any g-convex negative-log-likelihood.
- Can be combined with Kronecker models.
- Symmetrically constrained Tyler, MGGD....
- Any descent algorithm should find the global solution.
Proper quaternion multivariate $T$ distribution, dimension 10.
Discussion

- Geodesic convexity in positive definite matrices
- Similar to geometric programming in scalars.
- Powers and log determinants are g-convex.
- G-convexity is preserved in Kronecker products.
- Symmetry sets are g-convex.
- Unifies and generalizes many previous results.
- Lots of applications....

Take home message

If you always find the global solution, maybe its (g-)convex!