Learning graphical factor models with Riemannian optimization

Hippert-Ferrer, Bouchard, Mian, Vayer, Breloy, arXiv preprint arXiv:2210.11950, 2022

Arnaud Breloy, Séminaire équipe Sigma (CRIStAL)

Université Paris Nanterre

Statistics in signal processing and machine learning

Statistical point of view is ubiquitous:

- **Data** appears as the result of a **random processes** (uncertainties)
- Cast **statistical models** that reasonably fit **empirical histograms**
- Derive **processes** that achieve certain **average performance** for a **task**

(fitting, estimation, detection, classification, prediction)

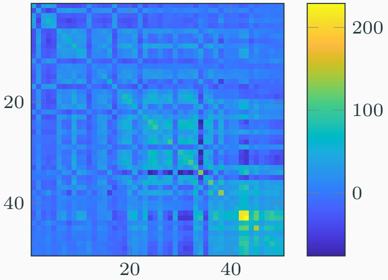
Scharf, Demeure, "Statistical signal processing: detection, estimation, and time series analysis," PrenticeHall, 1991

Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning," Springer-Verlag, 2009

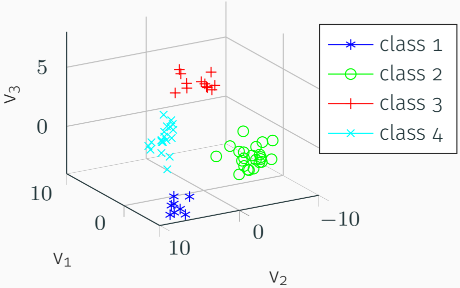
Parametric approach

Represent or analyze the data $\mathbf{x}$ through some statistical parameter $\boldsymbol{\theta}$

Example with $p \simeq 7k$ genes of $n = 63$ patients with $k = 4$ classes [Khan2001] represented by



Covariance of 50 selected genes



3 principal components

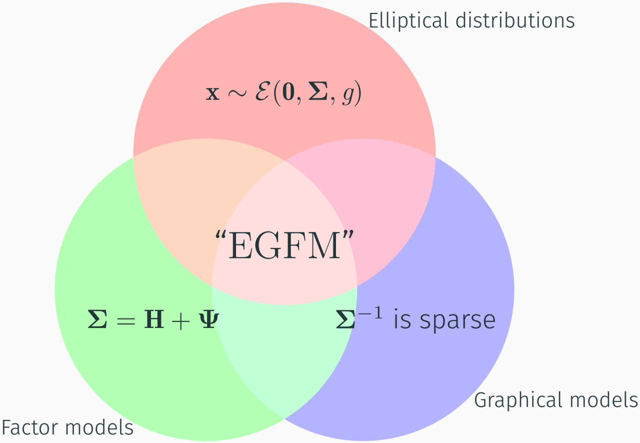
Statistical approach

"Assume $\mathbf{x} \sim f(\mathbf{x}, \boldsymbol{\theta})$, then do stuff"

- **Design** a meaningful pdf f and parameter $\boldsymbol{\theta}$
- **Analyze** model properties, performance bounds...
- **Solve** related optimization problems (MLEs, barycenters...)
- **Apply** the results to a task

Today's talk: **design** and **solve** for "**elliptical graphical factor models**"

WTF is this that?



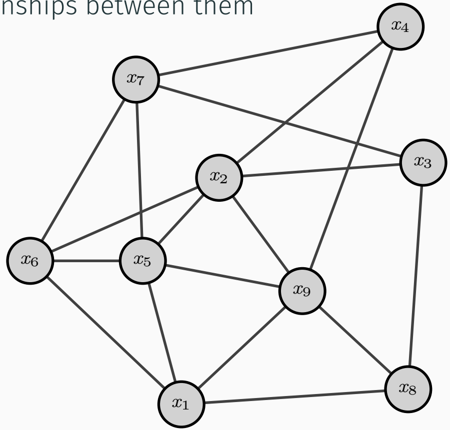
Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

Graphical models

Graphs help visualizing **relationships between entities**:

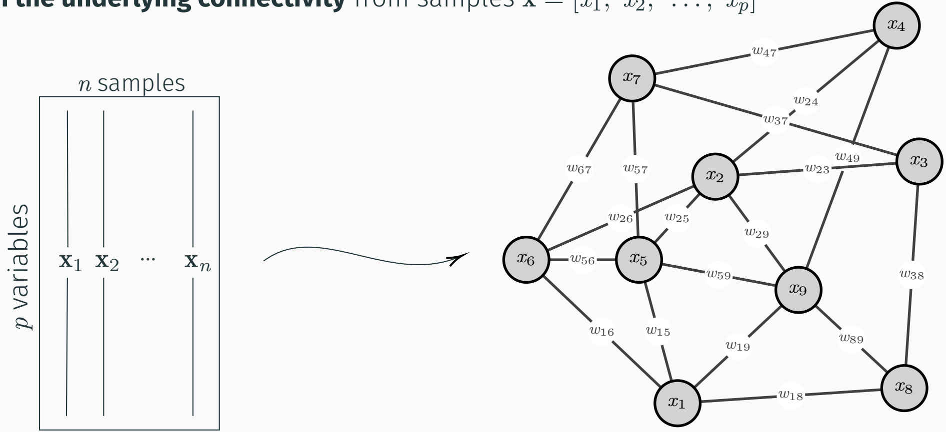
- **Nodes** correspond to entities, or **variables**
- **Edges** encode relationships between them



Graph learning

The graph **topology is unknown** but **each node generates data**

Learn the underlying connectivity from samples $\mathbf{x} = [x_1,\ x_2,\ \ldots,\ x_p]$



Graphical model: a statistical point of view

"Connection in the graph = conditional dependence"

For $\mathbf{x} = [\underbrace{x_1,\ x_2}_{\mathbf{x}_\top},\ \mathbf{x}_\perp]$, **conditional independence** $x_1 \perp\!\!\!\perp x_2$ holds if $\mathcal{L}(x_1|x_2, \mathbf{x}_\perp) = \mathcal{L}(x_1|\mathbf{x}_\perp)$

"knowing $\mathbf{x}_\perp$ makes x_2 irrelevant for predicting x_1"

Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\top\top} & \boldsymbol{\Sigma}_{\top\perp} \\ \boldsymbol{\Sigma}_{\perp\top} & \boldsymbol{\Sigma}_{\perp\perp} \end{bmatrix}$ and $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Theta}_{\top\top} & \boldsymbol{\Theta}_{\top\perp} \\ \boldsymbol{\Theta}_{\perp\top} & \boldsymbol{\Theta}_{\perp\perp} \end{bmatrix}$

Then $\mathbf{x}_\top|\mathbf{x}_\perp \sim \mathcal{N}(\xi_{\top|\perp},\ \boldsymbol{\Sigma}_{\top|\perp})$ with $\boldsymbol{\Sigma}_{\top|\perp} = \boldsymbol{\Sigma}_{\top\top} - \boldsymbol{\Sigma}_{\top\perp}\boldsymbol{\Sigma}_{\perp\perp}^{-1}\boldsymbol{\Sigma}_{\perp\top} = \boldsymbol{\Theta}_{\top\top}^{-1}$

So $x_1 \perp\!\!\!\perp x_2$ (no edge w_{12} on the graph) $\Leftrightarrow$ $\boldsymbol{\Sigma}_{\top|\perp}$ is diagonal $\Leftrightarrow$ $\boldsymbol{\Theta}_{12} = 0$

Learning Gaussian graphical models (GGM)

Assume a **Gaussian Markov Random Field** $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$

A **Gaussian graphical model** implies a **sparse precision matrix** $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$

Graphical Lasso (GLasso) $\Leftrightarrow$ regularized MLE of $\mathbf{\Theta}$

$$\underset{\mathbf{\Theta} \in \mathcal{S}_p^{++}}{\text{maximize}} \quad \log\det(\mathbf{\Theta}) - \text{Tr}\{\mathbf{S}\mathbf{\Theta}\} - \lambda h(\mathbf{\Theta})$$

$\longrightarrow$ Graph drawn from $\mathbf{\Theta}$'s support

Some limitations

- Gaussian model assumption $\longrightarrow$ sensitive to heavy tails
- No structure in $\mathbf{\Sigma}$ $\longrightarrow$ poor estimates when $n \simeq p$ or $n < p$

Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

Motivation for elliptical distributions

Objective: find a model $f(\mathbf{x}, \theta)$

- $\mathbf{x}$ is a **sample** in $\mathbb{R}^p$ or $\mathbb{C}^p$ (unstructured)
- f is a **pdf**
- θ **parameterizes** the pdf

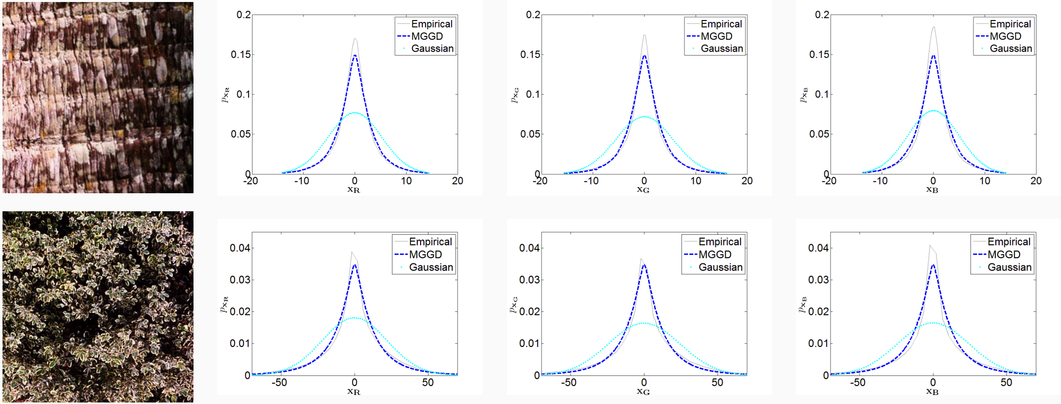
Challenges from real data:

- **Non-Gaussian**, **heavy-tailed** distributions
- **Outliers**

Elliptical models good entry point for this tutorial =)

- **Large family** that that generalizes the multivariate Gaussian distribution
- Still parameterized through **1st and 2nd order moments** (mean, covariance)
- **Better fit** to empirical histograms $\rightarrow$ **better results**

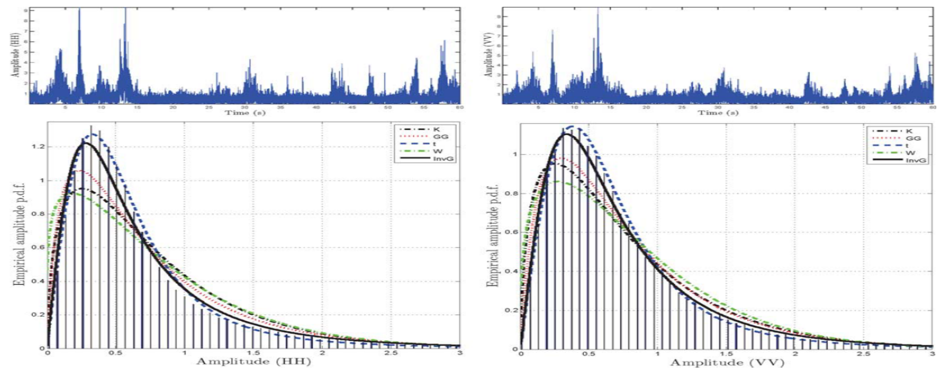
Motivating real-data examples (1/2)



Bark.0000 and Leaves.0008 from VisTex and marginal distributions of wavelet coefficients from RGB channels.

F. Pascal, L. Bombrun, J-Y. Tourneret, Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," IEEE TSP, 2013

Motivating real-data examples (2/2)



Modulus of HH and VV band of Shore of Lake Ontario sensed by McMaster IPIX radar

E. Ollila, D. E. Tyler, V. Koivunen, H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," IEEE TSP, 2012

Elliptical models

Complex elliptically symmetric distributions (CES)

$\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ if its pdf can be written

$$f(\mathbf{x}) \propto |\boldsymbol{\Sigma}|^{-1} g((\mathbf{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})),$$

where $g : [0, \infty) \to [0, \infty)$ is the **density generator** and

- $\boldsymbol{\mu} \in \mathbb{C}^p$ is the symmetry **center**
- $\boldsymbol{\Sigma} \in \mathcal{H}_p^{++}$ is the **scatter matrix**

If $\mathbf{x}$ has finite 2^{nd}-order moment, the **covariance matrix** is $\mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^H\right] = \alpha \boldsymbol{\Sigma}$

- $\alpha = -2\varphi'(0)$,
- φ is defined by the characteristic function $c_{\mathbf{x}}(\mathbf{t}) = \exp(i\mathbf{t}^H \boldsymbol{\mu})\, \varphi(\mathbf{t}^H \boldsymbol{\Sigma} \mathbf{t})$

Practical CES representation

Stochastic representation theorem

$\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ iif it admits the stochastic representation

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\boldsymbol{\Sigma}^{1/2}\mathbf{u}$$

where

- $\mathbf{u} \sim \mathcal{U}\left(\mathbb{C}\mathcal{S}^p\right)$ follow an uniform distribution on unit complex p-sphere
- $\mathcal{Q}$ is the 2^{nd}**-order modular variate**, independent of $\mathbf{u}$, with pdf

$$p(\mathcal{Q}) = \delta_{p,g}^{-1}\mathcal{Q}^{p-1}g(\mathcal{Q})$$

Interpretation:

- $\boldsymbol{\Sigma}$ pilots the **shape of the ellipsoid** (privileged direction)
- $\mathcal{Q}$ (equivalently g) models **amplitude fluctuations** (possibly heavy tails)

Some remarks on CES properties

1. **One-to-one relation** between pdf of $\mathcal{Q}$ and g

2. **Ambiguity**: $(\mathcal{Q}, \mathbf{\Sigma})$ and $(c^{-1}\mathcal{Q}, c\mathbf{\Sigma})$, $c > 0$ are valid stochastic representations of $\mathbf{x}$
$\Rightarrow$ requires normalization constraint

3. **Covariance matrix**: $\mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H\right] = \mathbb{E}[\mathcal{Q}]\mathbf{\Sigma}/p$, if $\mathbb{E}[\mathcal{Q}]$ exists

4. **Sampling**:
   - Draw a 2^{nd}-order modular variate $\mathcal{Q}$ from its pdf $p()$
   - Draw $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_p)$, then $\mathbf{u} \stackrel{d}{=} \mathbf{n}/|\mathbf{n}|\ \mathcal{U} \sim (\mathbb{C}S^p)$
   - Set $\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\mathbf{\Sigma}^{1/2}\mathbf{u}$

Important related distribution families

Compound Gaussian (CG) aka spherically invariant random vectors (SIRV)

$\mathbf{x} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$ iif it admits the stochastic CG-representation

$$\mathbf{x} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{n}$$

where

- $\tau \geq 0$ is called the **texture**, with pdf f_τ that is independent of $\mathbf{n}$
- $\mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is called the **speckle**.

Note: subclass of CES because if $\mathbf{n}_0 \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I})$, then $\mathbf{n}_0 \overset{d}{=} \sqrt{s}\mathbf{u}$ with $s \sim \Gamma(1, p)$

Mixture of scaled Gaussian distributions (MSG)

$\mathbf{x}_i \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tau_i\boldsymbol{\Sigma})$, where τ_i is unknown deterministic

Main examples (1/2)

Multivariate Gaussian distribution

CG: $f_\tau = \delta_1$ (or CES with $\mathcal{Q} \sim \Gamma(1, p)$)

Multivariate t-distribution with degree of freedom ν

CG: $\tau^{-1} \sim \Gamma(\nu/2, 2/\nu)$, where $\nu > 0$

- Encompasses Complex Cauchy dist. ($\nu = 1$) and CN dist. ($\nu \to \infty$)
- Finite 2nd-order moment for $\nu > 2$

K-distribution with shape parameter ν

CG: $\tau \sim \Gamma(\nu, 1/\nu)$, where $\nu > 0$

- Encompasses heavy-tailed dist. ($\nu \downarrow$) and CN dist. ($\nu \to \infty$)
- $\mathbb{E}[\tau] = 1 \implies \boldsymbol{\Sigma} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$

Main examples (2/2)

GG distribution with parameters s and η

- CES: $\mathcal{Q} =_d G^{1/s}$ where $G \sim \Gamma(m/s, \eta)$, $s, \eta > 0$
- PDF: $f_{\mathbf{x}}(\mathbf{x}) = cte\,|\boldsymbol{\Sigma}|^{-1} \exp\left(-(\eta\, \mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x})^s\right)$

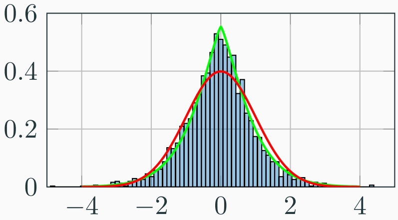
- Complex analog of the exponential power family, also called Box-Tiao distributions
- Subclass of multivariate symmetric Kotz-type distributions
- Case $s = 1 \implies$ CN dist.
- Heavier tailed than normal for $s < 1$ and lighter tailed for $s > 1$
- $s = 1/2 \implies$ generalization of Laplace dist.

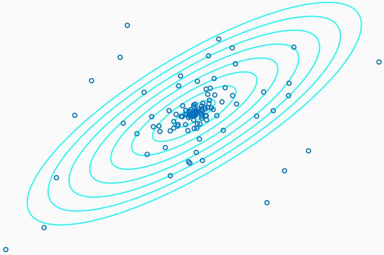
Wrapping-up

Complex elliptically symmetric distributions (CES)

$\mathbf{x} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ if it has for pdf

$$f(\mathbf{x}) \propto |\boldsymbol{\Sigma}|^{-1} g\left((\mathbf{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right)$$



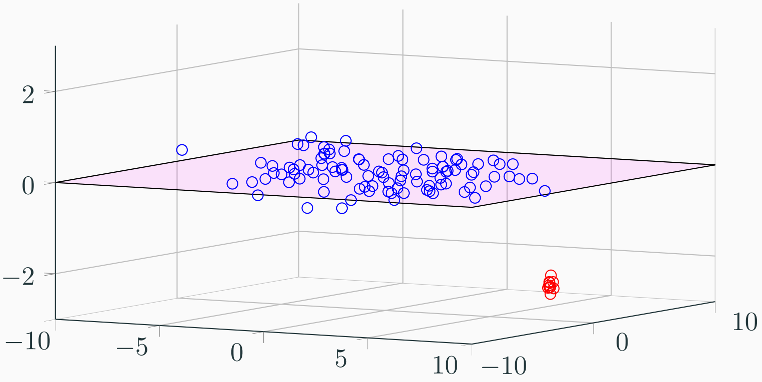


Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

Subspace learning grounds

$$\mathbf{x}_i \simeq \mathbf{U}\mathbf{U}^H\mathbf{x}_i, \text{ with } \mathbf{U} \in \mathrm{St}(p,k) \triangleq \{\mathbf{U} \in \mathbb{C}^{p\times k} \mid \mathbf{U}^H\mathbf{U} = \mathbf{I}\}$$



Probabilistic PCA and low-rank factor models

- **Probabilistic PCA** in Gaussian model [Tipping, 1999]

$$\mathbf{x}_i = \mathbf{U}\mathbf{D}^{1/2}\mathbf{s}_i + \mathbf{n}_i \quad \text{with} \quad \mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k) \quad \text{and} \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

ML estimator of $\mathbf{U}$ is the k leading eigenvectors of $\sum_{i=}^{n} \mathbf{x}_i \mathbf{x}_i^\top \Leftrightarrow$ **PCA**

- **Factor models** generalizes to $[\mathbf{n}]_j \sim \mathcal{N}(\mathbf{0}, \sigma_j^2)$, resulting in the **covariance structure**

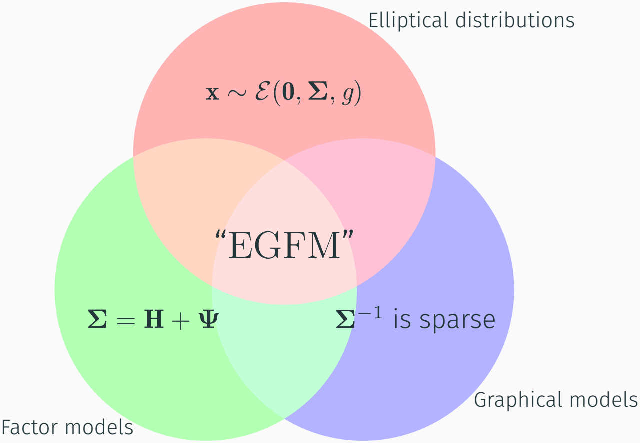
$$\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \mathbf{\Sigma} \in \mathcal{M}_{p,k} = \left\{ \mathbf{\Sigma} = \mathbf{H} + \mathbf{\Psi},\ \mathbf{H} \in \mathcal{S}_{p,k}^{+}, \ \in \mathcal{D}_p^{++} \right\}$$
$$= \text{“rank } k \text{ plus diagonal”}$$

Dimension reduction: from $p(p+1)/2$ to $p(k+1) - k(k-3)/2$ parameters!

Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

Putting everything together



Learning EGFMs

$$\underset{\mathbf{\Sigma} \in \mathcal{S}_p^{++}}{\text{minimize}} \quad \mathcal{L}(\mathbf{\Sigma}) + \lambda h(\mathbf{\Sigma})$$
$$\text{subject to} \quad \mathbf{\Sigma} \in \mathcal{M}_{p,k}$$

- **Likelihood**

$$\mathcal{L}(\mathbf{\Sigma}) \propto \frac{1}{n} \sum_{i=1}^{n} \rho(\mathbf{x}_i^\top \mathbf{\Sigma}^{-1} \mathbf{x}_i) + \frac{1}{2} \log |\mathbf{\Sigma}| + \text{const.}$$

e.g., $g(t) = (1 + t/\nu)^{-\frac{\nu+p}{2}}$ for t-distribution

- **Smooth penalty**

$$h(\mathbf{\Sigma}) = \sum_{q \neq \ell} \phi([\mathbf{\Sigma}^{-1}]_{q\ell})$$

e.g., $\phi(t) = \varepsilon \log(\cosh(t/\varepsilon))$

Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

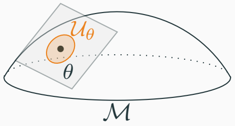
Structured parameter space as a manifold

The **parameter spaces**

- **Covariance matrices**: $\boldsymbol{\Sigma} \in \mathcal{H}_p^{++}$
- **Rank k PSD matrices**: $\mathbf{H} \in \mathcal{S}_{p,k}^{+}$

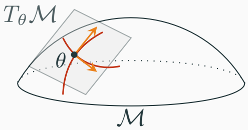
turn out to be **manifolds** $\mathcal{M}$ (locally diffeomorphic to $\mathbb{R}^d$, with $\dim(\mathcal{M}) = d$)

$$\forall \theta \in \mathcal{M},\ \exists \mathcal{U}_\theta \subset \mathcal{M} \text{ and } \varphi_\theta : \mathcal{U}_\theta \to \mathbb{R}^d,\ \text{diffeomorphism}$$



Riemannian manifolds (1/2)

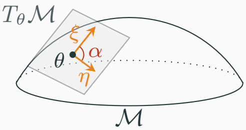
Tangent space $T_\theta\mathcal{M}$ at point $\boldsymbol{\theta}$



- Curve $\gamma : \mathbb{R} \to \mathcal{M}$, $\gamma(0) = \theta$
- Derivative: $\dot{\gamma}(0) = \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t}$
- Tangent space $T_\theta\mathcal{M} = \{\dot{\gamma}(0) : \ \gamma : \mathbb{R} \to \mathcal{M}, \ \gamma(0) = \theta\}$

Equip $T_\theta\mathcal{M}$ with a **Riemannian metric** $\langle\cdot,\cdot\rangle_\theta$ yields a **Riemannian manifold**

- $\langle\cdot,\cdot\rangle_\theta : (T_\theta\mathcal{M} \times T_\theta\mathcal{M}) \to \mathbb{R}$ **inner product** on $T_\theta\mathcal{M}$
  (bilinear, symmetric, positive definite)
- defines length and relative positions of tangent vectors



$$\|\xi\|_\theta^2 = \langle\xi,\xi\rangle_\theta \qquad \alpha(\xi,\eta) = \frac{\langle\xi,\eta\rangle_\theta}{\|\xi\|_\theta \, \|\eta\|_\theta}$$

Riemannian manifolds (2/2)

The Riemannian metric $\langle \cdot, \cdot \rangle_\theta$ induces **a geometry** for $\mathcal{M}$

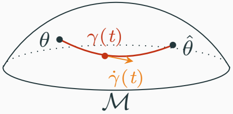
Geodesics $\gamma : [0, 1] \to \mathcal{M}$

- generalizes straight lines on $\mathcal{M}$
- curves on $\mathcal{M}$ with zero acceleration: $\frac{D^2 \gamma}{dt^2} = 0$

defined by $(\gamma(0), \dot{\gamma}(0))$ or $(\gamma(0), \gamma(1))$

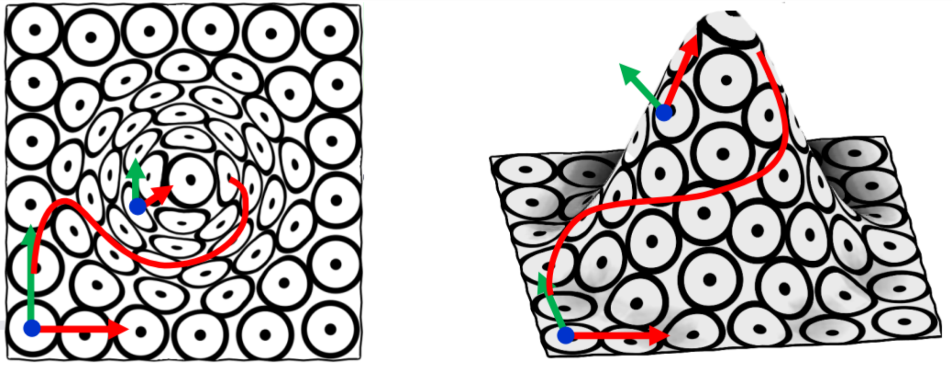
operator $\frac{D^2}{dt^2}$ depends on $\mathcal{M}$ and $\langle \cdot, \cdot \rangle$.

Riemannian distance $\mathrm{dist}(\theta, \hat{\theta}) = \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)} \, dt$



distance = length of γ connecting θ and $\hat{\theta}$

Riemannian geometry – Riemannian metric: intuition



Which metric/geometry to chose ?

The **Fisher information metric** looks like an **ideal driven by the model**

Still, we can chose **alternate metrics** **suited to some needs**

- Availability (**closed-form**) of theoretical objects
- Interesting **invariance** properties
- **Practical results** of the chosen task

Metric	Geodesics	Distance	Retraction	Completeness	Invariance 1	Invariance 2	Perf.
(a)	✗	✗	✓	✓	✗	✓	82%
(b)	✓	✗	✓	✓	✓	✗	**86**%
(c)	✓	✓	✓	✗	✗	✓	79%

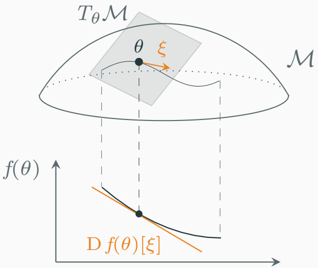
Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

Riemannian optimization

$$\underset{\theta \in \mathcal{M}}{\text{minimize}} \quad f(\theta)$$

Riemannian optimization: a framework for optimization on $\mathcal{M}$ equipped with $\langle \cdot, \cdot \rangle$.



Descent direction of f at θ:

$$\xi \in T_\theta \mathcal{M}, \quad \mathrm{D}\, f(\theta)[\xi] < 0$$

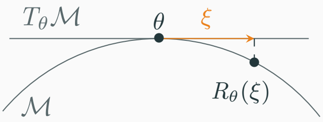
Riemannian gradient of f at θ:

$$\langle \mathrm{grad}\, f(\theta), \xi \rangle_\theta = \mathrm{D}\, f(\theta)[\xi]$$

Riemannian optimization

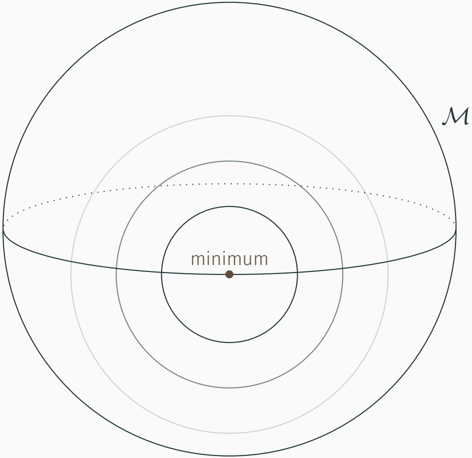
Main ingredients

- Descent direction: $\xi \in T_\theta\mathcal{M}$ so that $\langle \mathrm{grad}\, f(\theta), \xi \rangle_\theta < 0$
- **Retraction** of ξ on $\mathcal{M}$ (smooth mapping)

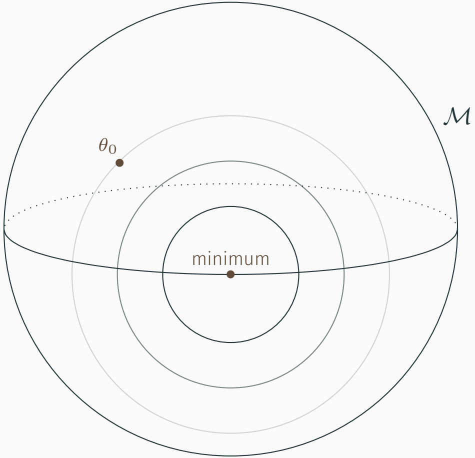


Flexibility: metric, retraction, descent method (gradient, conjugate gradient, BFGS...)

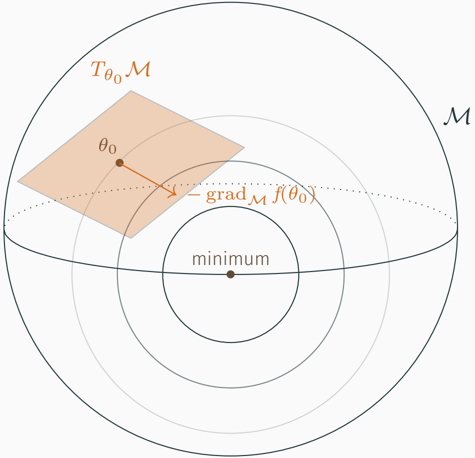
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \, \mathrm{grad} \, f(\theta_i))$



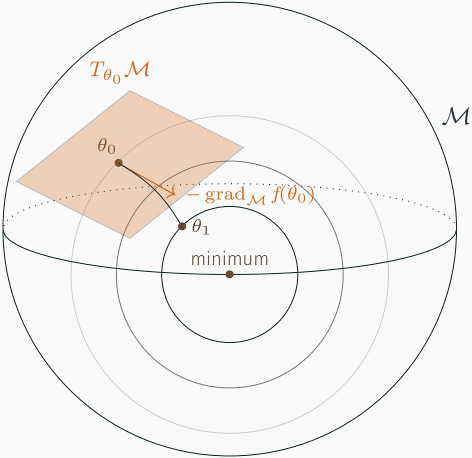
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \,\mathrm{grad}\, f(\theta_i))$



Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \operatorname{grad} f(\theta_i))$



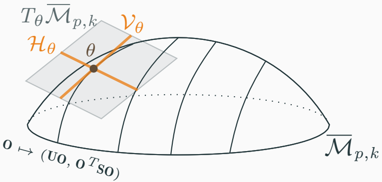
Example: Riemannian gradient descent $\theta_{i+1} = R_{\theta_i}(-t_i \,\mathrm{grad}\, f(\theta_i))$



Optimization on $\mathcal{S}_p^{++}$ and $\mathcal{M}_{p,k}$

Tools for $\Sigma \in \mathcal{S}_p^{++}$

Tangent space	$\forall \boldsymbol{\Sigma} \in \mathcal{S}_p^{++}, \quad T_{\boldsymbol{\Sigma}}\mathcal{S}_p^{++} \simeq \mathcal{S}_p$
Metric	$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle_{\boldsymbol{\Sigma}} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta})$
Gradient	$\mathrm{grad}\, f(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}\, \mathrm{sym}(\mathrm{grad}_{\mathcal{E}}\, f(\boldsymbol{\Sigma}))\boldsymbol{\Sigma}$
Retraction	$R_{\boldsymbol{\Sigma}}(\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}$



Tools for $\mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \in \mathcal{H}_{p,k}^+$ as $(\mathrm{St}(p,k) \times \mathcal{H}_k^{++})/\mathcal{U}_k$

Metric:

$$\langle \bar{\xi}, \bar{\eta} \rangle_{\bar{\theta}} = \underbrace{\mathfrak{Re}(\mathrm{Tr}(\boldsymbol{\xi}_{\mathbf{U}}^H(\mathbf{I}_p - \tfrac{1}{2}\mathbf{U}\mathbf{U}^H)\boldsymbol{\eta}_{\mathbf{U}}))}_{\text{canonical on } \mathrm{St}(p,k)} + \underbrace{\alpha\mathrm{Tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Lambda}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Lambda}}) + \beta\mathrm{Tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\xi}_{\boldsymbol{\Lambda}})\mathrm{Tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Lambda}})}_{\text{affine invariant on } \mathcal{H}_k^{++}}$$

details in [Bouchard21] =)

Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

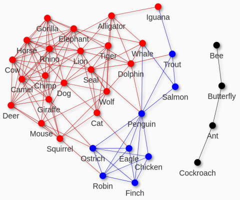
Applications on some data sets

Methods

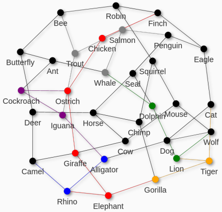
- **GLasso**
- 4 options of **EGFM**: {Gaussian, t-dist.} $\times$ {Full rank, Factor model}
- Laplacian learning: **NGL** (Gauss.), **SGL** (Gauss., K-comp.), **StGL** (t-dist., K-comp.)

Datasets

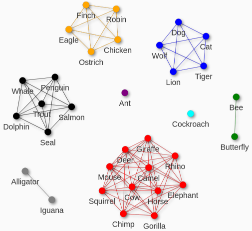
- **Animals**: $p = 33$ animals, $n = 102$ categorical questions
- **GNSS Piton de la Fournaise**: $p = 22$ stations, $n = 1106$ dates
- **Concepts**: $p = 1000$ concepts, $n = 218$ semantic features (5pt scale)



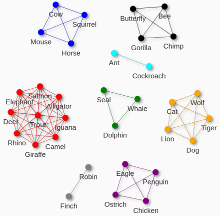
GLasso



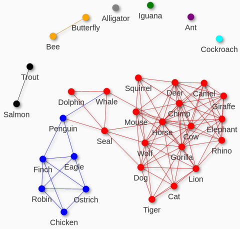
NGL



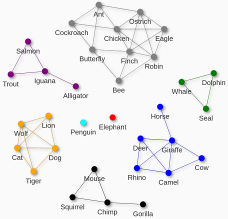
SGL



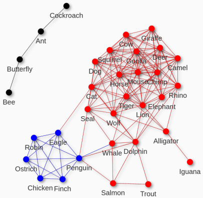
StGL



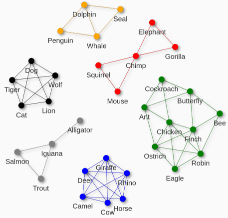
GGM



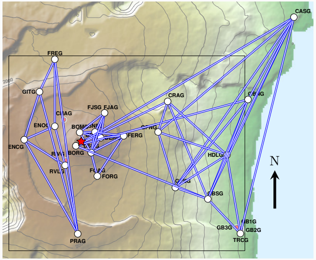
GGFM



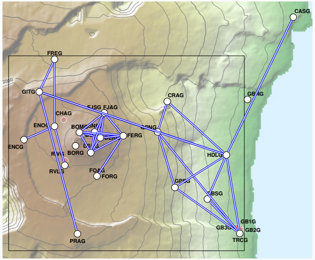
EGM



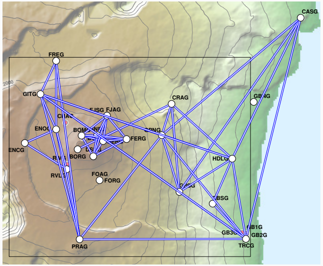
EGFM



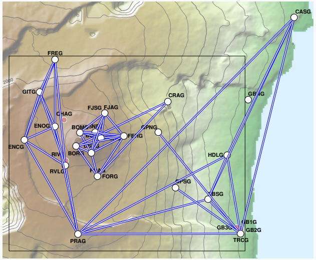
StGL



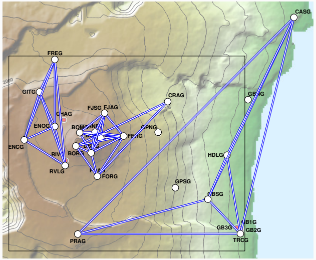
GGM



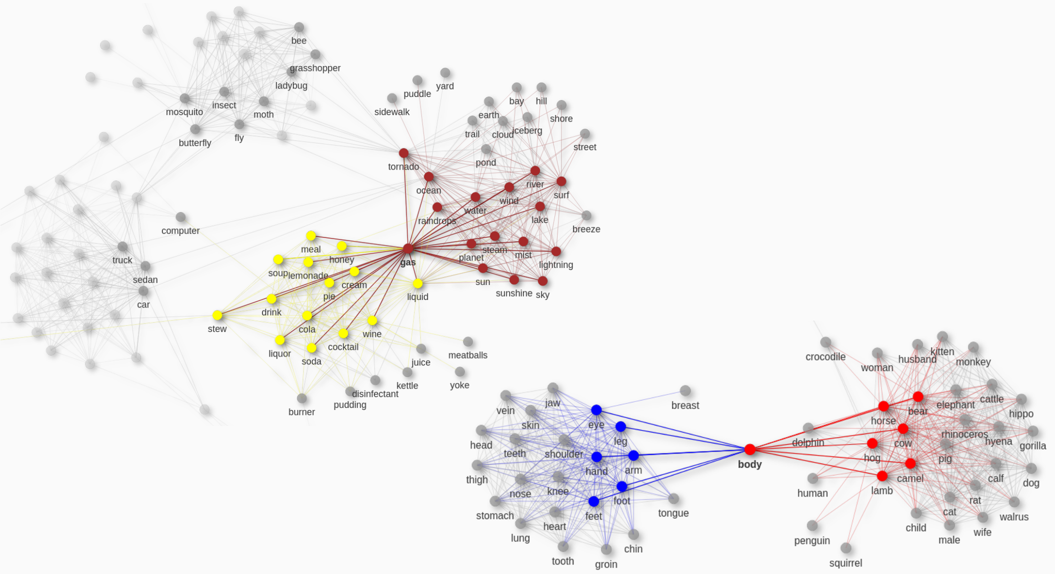
EGM



GGFM



EGFM


bee
grasshopper
ladybug
mosquito
insect
moth
butterfly
fly
puddle
yard
sidewalk
bay
hill
earth
shore
trail
cloud
iceberg
street
pond
tornado
ocean
river
surf
wind
water
lake
raindrops
breeze
computer
meal
honey
gas
steam
mist
planet
lightning
soup
lemonade
cream
sun
truck
sedan
car
pie
liquid
sunshine
sky
drink
stew
cola
wine
liquor
cocktail
soda
juice
meatballs
kettle
yoke
disinfectant
pudding
burner
crocodile
woman
husband
kitten
monkey
jaw
breast
vein
eye
skin
leg
head
teeth
shoulder
hand
arm
body
thigh
nose
knee
foot
tongue
stomach
feet
heart
lung
tooth
groin
chin
bear
elephant
cattle
hippo
horse
rhinoceros
hyena
gorilla
dolphin
cow
hog
pig
camel
calf
dog
human
lamb
rat
cat
walrus
child
male
wife
penguin
squirrel

Outline

- **Design**
  - Gaussian Graphical models
  - Elliptical models
  - Probabilistic PCA and factor models
  - Everything together
- **Solve**
  - Manifolds and Riemannian geometry
  - Riemannian optimization
- **Apply** on real data examples

Perspectives

Graph Learning

- Other statistical models $\mathcal{L}$
- More structures $\mathcal{M}$ (Laplacian, MTP_2, ...)

Geometry of graphs:

- Riemannian geometry of adjacency/Laplacian matrices
- Classify using graphs as features

Links with optimal transport of graphs

Thanks for the invitation! ;D