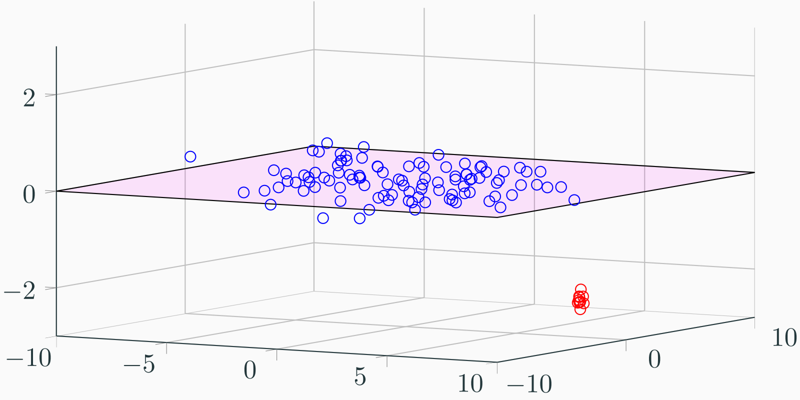
About subspace learning

Arnaud Breloy, Aalto University, 12 May 2021

Université Paris Nanterre

Subspace learning grounds

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



Plan overview

$$\underset{\mathbf{U}\in\mathrm{St}(p,k)}{\text{minimize}} \quad f(\mathbf{U})$$

- **Design** the model/objective function f
- **Solve** the constrained minimization problem
- **Analyze** the estimation problem (performance)
- **Apply** the result to some task

Plan overview

$$\underset{\mathbf{U}\in\mathrm{St}(p,k)}{\text{minimize}} \quad f(\mathbf{U})$$

- **Design** the model/objective function f
- **Solve** the constrained minimization problem
- **Analyze** the estimation problem (performance)
- **Apply** the result to some task

Principal component analysis (PCA)

"Vanilla" PCA of **rank** k

- Singular value decomposition (**SVD**) of the data matrix $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_n] \in \mathbb{C}^{p \times n}$

$$\mathbf{Z} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^{\perp}]\,\mathbf{D}\mathbf{V}^H$$

- **Loading vectors** $\mathbf{U} \in \mathrm{St}(p, k)$
- **Principal components** $\mathbf{z}_i^k = \mathbf{U}^H \mathbf{z}_i \in \mathbf{C}^k$, projected data $\tilde{\mathbf{z}}_i = \mathbf{U}\mathbf{z}_i^k$

Solution of **multiple underlying problems** (frameworks)

$\rightarrow$ each point of view offers interesting **tools** and **extensions**

PCA: geometric point of view

- **Euclidean distance** $\mathrm{dist}(\mathbf{U}, \mathbf{z}) = \sqrt{\mathbf{z}^H\mathbf{z} - \mathbf{z}^H\mathbf{U}\mathbf{U}^H\mathbf{z}}$

- **Geometric PCA** [Pearson, 1901]

$$\underset{\mathbf{U}\in\mathrm{St}(p,k)}{\text{minimize}} \sum_{i=1}^{n} \mathrm{dist}^2(\mathbf{U}, \mathbf{z}_i)$$

  a solution $\mathbf{U}^\star$ is the k leading eigenvectors of $\mathbf{Z}\mathbf{Z}^H = \sum_{i=1}^n \mathbf{z}_i\mathbf{z_i}^H/n \Leftrightarrow$ **PCA**

- **Extensions** using **alternate distances**
  - Robust costs: $f(\mathbf{U}) = \sum_{i=1}^n \rho(\mathrm{dist}^2(\mathbf{U}, \mathbf{z}_i))$ [Ding, 2006]
  - Other objects: $f(\mathbf{U}) = \sum_{i=1}^n \mathrm{dist}^2_{\mathcal{G}(p,k)}(\mathbf{U}, \mathbf{U}_i)$ [Marrinan, 2014]

PCA: statistical point of view (1/2)

- **Covariance matrix** $\mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] = \mathbf{\Sigma}$
- **Statistical PCA** a.k.a. "maximizing expected variance" [Hotelling, 1933]

$$\underset{\mathbf{U}\in\mathrm{St}(p,k)}{\text{maximize}} \quad \mathrm{Tr}\left\{\mathbf{U}^H \underbrace{\left(\sum_{k=1}^{n} \mathbf{z}_i\mathbf{z}_i^H / n\right)}_{\hat{\mathbf{\Sigma}}} \mathbf{U}\right\}$$

  a solution $\mathbf{U}^\star$ is the k leading eigenvectors of $\hat{\mathbf{\Sigma}} \Leftrightarrow$ **PCA**

- **Extensions** using **alternate plug-in estimates**
  - M-estimators, R-estimators, ... [Drašković, 2019]
  - Structure priors (Toeplitz, persymetric, ...) [Mériaux, 2019]
- **Tools**: notion of **uncorrelated** principal components

PCA: statistical point of view (2/2)

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

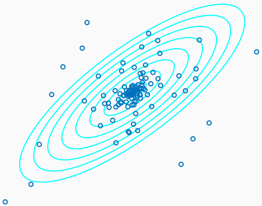
$$\mathbf{z}_i = \mathbf{U}\mathbf{D}^{1/2}\mathbf{s}_i + \mathbf{n}_i \quad \text{with} \quad \mathbf{s} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_k) \quad \text{and} \quad \mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_p)$$

ML estimator of $\mathbf{U}$ is the k leading eigenvectors of $\sum_{i=1}^{n} \mathbf{z}_i \mathbf{z_i}^H \Leftrightarrow$ **PCA**

- **Extensions** using **alternate distributions**

$$\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \underbrace{\mathbf{U}\mathbf{D}\mathbf{U}^H + \sigma^2\mathbf{I}_p}_{\mathbf{\Sigma}}, g)$$ [Bouchard, 2021]

$$\mathcal{L}(\{\mathbf{z}_i\}; \mathbf{\Sigma}) = n \log |\mathbf{\Sigma}| + p \sum_{i=1}^{n} \log g(\mathbf{z}_i^H \mathbf{\Sigma}^{-1} \mathbf{z}_i)$$



or **mixtures** of independent contributions [Sun, 2016] [Hong, 2018]

- **Tools**: statistical analysis, performance bounds, missing data, ...

PCA: one of many Bayesian point of views [Besson, 2011]

- **Bayesian PCA**: prior on $\mathbf{U}$ in $\mathbf{z}_i \overset{d}{=} \mathbf{U}\mathbf{s}_i + \mathbf{n}_i$

- **Bingham-Langevin prior** [Ben Abdallah, 2020]

$$\mathbf{U} \sim \mathrm{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$$

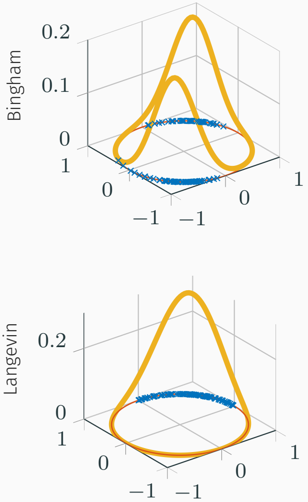
$$\mathcal{L}_{\mathbf{U}}(\mathbf{U}) \propto \exp(\textstyle\sum_{r=1}^{k} \left[\Re\mathfrak{e}\{\mathbf{c}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H \mathbf{A}_r \mathbf{u}_r\right])$$

- **MMSD estimator**

$$f(\hat{\mathbf{U}}) = \mathbb{E}\left[||\hat{\mathbf{U}}\hat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H||_F^2 \right]$$

- **MAP**

$$f(\mathbf{U}) = \underbrace{\mathcal{L}(\{\mathbf{z}_i\}|\mathbf{U})}_{\text{data fitting}} + \underbrace{\mathcal{L}_{\mathbf{U}}(\mathbf{U})}_{\text{shrinkage}}$$



PCA: algebraic point of view

- **Low-rank approximation**

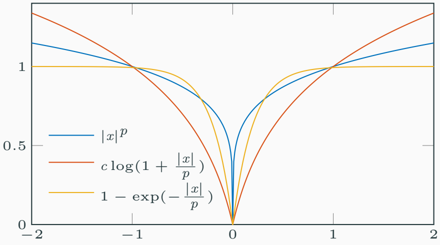
$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & ||\mathbf{Z} - \mathbf{X}||_F^2 \\ \text{subject to} \quad & \text{rank}(\mathbf{X}) = k \end{aligned}$$

$\mathbf{X}^\star$ is the rank-k truncation of the SVD $\Leftrightarrow$ **subspace recovered by PCA**

- **Extensions** using **alternate decompositions/structures**
  - Low-rank plus sparse recovery (*Robust PCA*) [Candès, 2011]
  - Matrix completion (missing entries) [Boumal, 2016]
  - Additional structure in the principal components [Uematsu, 2017]
  - Non-negative matrix factorization, ...

Sparse PCA

- **Sparse PCA**: variable selection through the loading vectors
- **In practice** add sparsity-promoting penalties $\rho_{\mathrm{S}}(\mathbf{U}) = \sum_{i=1}^{p} \sum_{j=1}^{k} \ell_\epsilon([\mathbf{U}]_{i,j})$



Entry-wise sparse penalty ℓ_ϵ

"Design" part: concluding overview

Motivations: accurate fitting, robustness, introducing prior, regularization

Statistics

- **Likelihood & Covariance**

$$\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}(\mathbf{U}, \boldsymbol{\theta}), g)$$

- **Bayesian priors**

$$\mathbf{U} \sim \mathrm{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$$

Geometry

- **Distances**

$$\mathrm{dist}(\mathbf{U}, \mathbf{z}) = \sqrt{\mathbf{z}^H\mathbf{z} - \mathbf{z}^H\mathbf{U}\mathbf{U}^H\mathbf{z}}$$

- **Sparsity**

ℓ_1-, $\ell_{2,1}$-norm

ℓ_0-norm proxies

Matrix algebra: $\mathbf{U}$ hidden in a low-rank matrix decomposition

Plan overview

$$\underset{\mathbf{U}\in\mathrm{St}(p,k)}{\text{minimize}} \quad f(\mathbf{U})$$

- **Design** the model/objective function f
- **Solve** the constrained minimization problem
- **Analyze** the estimation problem (performance)
- **Apply** the result to some task

Introduction

- **Problem**: solving

$$\underset{\mathbf{U}\in \mathrm{St}(p,k)}{\text{minimize}} \ f(\mathbf{U})$$

on the Stiefel manifold

$$\mathrm{St}(p,k) = \{\, \mathbf{U} \in \mathbb{C}^{p\times k} \mid \mathbf{U}^H\mathbf{U} = \mathbf{I} \,\},\ \ k < p$$

- **Examples**: all flavors of PCA, subspace recovery, low-rank matrix recovery, ...
- **Issue**: orthonormality constraint is not friendly ! (non-convex, bi-linear)

Existing solutions

- **Solution #1**: Riemannian optimization on $\mathrm{St}(p, k)$

ABS09 Absil, Mahony, Sepulchre, "Optimization algorithms on matrix manifolds," Princeton Univ. Press, 2009

EDE98 Edelman, Arias, Smith, "The geometry of algorithms with orthogonality constraints," SIMAX, 1998

MAN02 Manton, "Optimization algorithms exploiting unitary constraints," IEEE Tans. on SP, 2002

- **Solution #2**: artful ADMM tricks involving $f = f_{\mathrm{u}} + f_{\mathrm{v}}$

$$\begin{array}{ll} \underset{\mathbf{U},\mathbf{V}}{\text{minimize}} & f_{\mathrm{u}}(\mathbf{U}) + f_{\mathrm{v}}(\mathbf{V}) \\ \text{subject to} & \mathbf{U} \in \mathrm{St}(p, k) \\ & \mathbf{U} = \mathbf{V} \end{array}$$

UEM19 Uematsu, Fan, Chen, Lv, Lin, "SOFAR: Large-Scale Association Network Learning," IEEE Trans. on IT, 2019

- **Solution #3**: Majorization-Minimization ticks ?

Some references on Majorization-Minimization (MM)

- **Tutorial articles**:

HUN04 Hunter, Lange, "A Tutorial on MM Algorithms", Amer. Statistician, 2004

SUN17 Sun, Babu, Palomar, "Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning", IEEE Trans. on SP, 2017

- **Courses slides**:

LAN07 Lange, "The MM Algorithm", Departments of Biomathematics, UCLA, 2007

SUN16 Sun, Palomar, "Majorization-Minimization Algorithm Theory and Applications", Department of Electronic and Computer Engineering, HKUST, 2016

- **MM for $\mathrm{St}(p, k)$**:

BRE21 Breloy, Kumar, Sun, Palomar, "Majorization-Minimization on the Stiefel Manifold With Application to Robust Sparse PCA", IEEE Trans on SP, 2021

The MM Algorithm principle (1/3)

- Consider the optimization problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X}, \end{array}$$

where f is too complex to be handled directly

- The idea is to successively minimize an approximation $g(\mathbf{x}|\mathbf{x}_t)$

$$\mathbf{x}_{t+1} = \underset{\mathbf{x} \in \mathcal{X}}{\arg\min} \; g(\mathbf{x}|\mathbf{x}_t)$$

hoping the sequence $\{\mathbf{x}_t\}$ will converge to a critical point of f

- The MM algorithm provides
  - The guidelines for the construction of such function g
  - The conditions to ensure the success of this method

The MM Algorithm principle (2/3)

Construction rules for the surrogate function g

(A1) Equality at the considered point

$$g(\mathbf{y}|\mathbf{y}) = f(\mathbf{y}) \ \ \forall\, \mathbf{y} \in \mathcal{X}$$

(A2) “Majorization”

$$f(\mathbf{x}) \leq g(\mathbf{x}|\mathbf{y}) \ \ \forall\, \mathbf{x}, \mathbf{y} \in \mathcal{X}$$

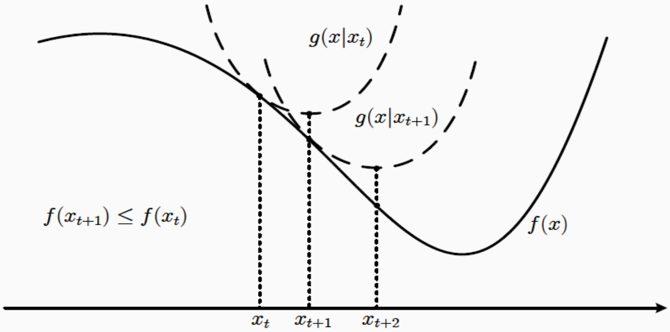
(A3) Equality of directional derivatives

$$g'(\mathbf{x}, \mathbf{y}; \mathbf{d})|_{\mathbf{x}=\mathbf{y}} = f'(\mathbf{y}; \mathbf{d}) \ \ \forall\, \mathbf{d} \text{ with } \mathbf{y} + \mathbf{d} \in \mathcal{X}$$

(A4) $g(\mathbf{x}|\mathbf{y})$ is continuous in $\mathbf{x}$ and in $\mathbf{y}$

The MM Algorithm principle (3/3)

"Iteratively minmizing a smooth local tight upperbound of the objective"



$$\mathbf{x}_{t+1} = \underset{\mathbf{x}\in\mathcal{X}}{\text{argmin}} \; g(\mathbf{x}|\mathbf{x}_t)$$

MM for $\mathrm{St}(p, k)$ (1/3)

- General idea:
  - Apply the MM principle for $\mathrm{St}(p, k)$
  - Formulate iterations as orthogonal Procrustes problems
  - Iterations under orthonormality constraint are hence easily solved !
- Unified view and generalizations of a well known trick

KOS91 Koschat, Swayne, “A weighted Procrustes criterion,” Psychometrika, 1991

KIE02 Kiers, “Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems,” Computational statistics & data analysis, 2002

MM for $\mathrm{St}(p, k)$ (2/3)

- **New rule**

  (A5) Linearity: restricting to $\mathrm{St}(p, k)$, g can be expressed as

$$g\left(\mathbf{U}|\mathbf{U}^t\right) = -\mathrm{Tr}\left\{(\mathbf{R}(\mathbf{U}^t))^H\mathbf{U}\right\} - \mathrm{Tr}\left\{\mathbf{U}^H\mathbf{R}(\mathbf{U}^t)\right\} + \mathrm{const.},$$

  where $\mathbf{R}\left(\mathbf{U}^t\right)$ is a matrix function of $\mathbf{U}^t$.

- **MM steps**: Minimizing (A5) on $\mathrm{St}(p, k) \Leftrightarrow$ orthogonal Procrustes

$$\begin{array}{ll} \underset{\mathbf{U}}{\text{minimize}} & ||\mathbf{R}\left(\mathbf{U}^t\right) - \mathbf{U}||_F^2 \\ \text{subject to} & \mathbf{U}^H\mathbf{U} = \mathbf{I} \end{array} \Rightarrow \begin{array}{l} \mathbf{U}^{(t+1)} = \mathbf{V}_L\mathbf{V}_R^H \\ \mathbf{U}^{(t+1)} \triangleq \mathcal{P}_{\mathrm{St}}\left\{\mathbf{R}(\mathbf{U}^t)\right\} \end{array}$$

with $\mathbf{R}(\mathbf{U}^t) \overset{\mathrm{TSVD}}{=} \mathbf{V}_L\mathbf{D}\mathbf{V}_R^H$

MM for $\mathrm{St}(p, k)$ (3/3)

- **Convergence to the KKT set**:

RAZ13 Razaviyayn, Hong, Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization", SIOPT, 2013

Fu17 Fu, Huang, Hong, Sidiropoulos, Man-Cho So, "Scalable and flexible multiview max-var canonical correlation analysis," IEEE Trans. on SP, 2017

- **Convergence in variable**: case by case study

KIE95 Kiers, "Maximization of sums of quotients of quadratic forms and some generalizations," Psychometrika, 1995

LER17 Lerman, Maunu, "Fast, robust and non-convex subspace recovery," Info. and Inference (IMA), 2017

Finding $\mathrm{R}(\cdot)$: the surrogate catalog

- **Problem**: finding surrogates of the form

$$g\left(\mathbf{U}|\mathbf{U}^t\right) = -\mathrm{Tr}\left\{(\mathbf{R}(\mathbf{U}^t))^H\mathbf{U}\right\} - \mathrm{Tr}\left\{\mathbf{U}^H\mathbf{R}(\mathbf{U}^t)\right\} + \text{const.}$$

- **Atoms covered**:
  - Convex/concave quadratic functions (QFs)
  - Convex/concave composition of a QF and a function ρ
  - Functions that have element-wise quadratic surrogates
  - Ratios of QFs
- **Overall**:
  - Most of the standard costs are covered
  - Easy to build/recognize meaningful costs by combination

Surrogates for convex/concave QFs (1/2)

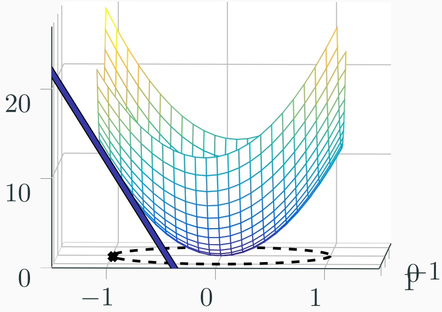
Let $\mathbf{M} \succcurlyeq \mathbf{0}$, $\mathbf{D} \succcurlyeq \mathbf{0}$, and

$$f_{\mathrm{B}}(\mathbf{U}) = \mathrm{Tr}\left\{\mathbf{U}^H\mathbf{M}\mathbf{U}\mathbf{D}\right\}$$

Prop.1 : The function $-f_{\mathrm{B}}$ admits a linear majorizing surrogate with

$$\mathbf{R}(\mathbf{U}^t) = \mathbf{M}\mathbf{U}^t\mathbf{D}.$$

with equality at point $\mathbf{U}^t$

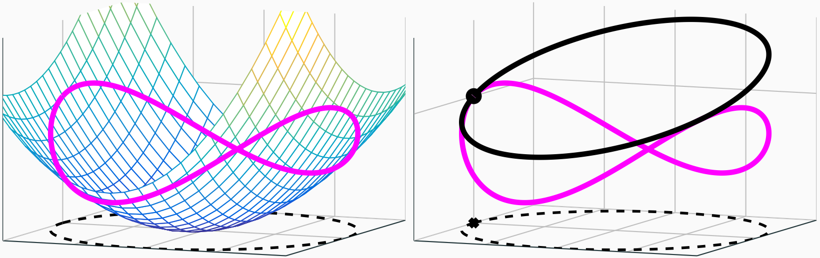


Surrogates for convex/concave QFs (2/2)

Prop.2: The function f_{B} admits on $\mathrm{St}(p,k)$ a linear majorizing surrogate with

$$\mathbf{R}(\mathbf{U}^t) = -\mathbf{K}\mathbf{U}^t\mathbf{D},$$

where $\mathbf{K} = \mathbf{S} - \lambda_{\mathbf{S}}^{\max}\mathbf{I}$ and $\lambda_{\mathbf{S}}^{\max}$ is the largest eigenvalue of $\mathbf{S}$. (equality at $\mathbf{U}^t$)



Surrogates for ratios of QFs

Let $\mathbf{C} \succ \mathbf{0}$, $\mathbf{A} \succcurlyeq \mathbf{0}$ and

$$f_{\mathrm{q}}(\mathbf{U}) = -\mathrm{Tr}\left\{\left(\mathbf{U}^H\mathbf{C}\mathbf{U}\right)^{-1}\mathbf{U}^H\mathbf{A}\mathbf{U}\right\},$$

Prop.3: The function f_{q} admits on $\mathrm{St}(p,k)$ a linear majorizing surrogate with

$$\mathbf{R}(\mathbf{U}^t) = \mathbf{A}^{1/2}\mathbf{T}(\mathbf{U}^t) - \left(\mathbf{K}\mathbf{U}^t\widetilde{\mathbf{T}}(\mathbf{U}^t)\right),$$

and

$$\mathbf{T}(\mathbf{U}^t) = \mathbf{A}^{1/2}\mathbf{U}^t\left((\mathbf{U}^t)^H\mathbf{C}\mathbf{U}^t\right)^{-1},$$

$$\widetilde{\mathbf{T}}(\mathbf{U}^t) = \left(\mathbf{T}(\mathbf{U}^t)\right)^H\mathbf{T}(\mathbf{U}^t),$$

$$\mathbf{K} = \mathbf{C} - \lambda_{\mathbf{C}}^{\max}\mathbf{I},$$

where $\lambda_{\mathbf{C}}^{\max}$ is the largest eigenvalue of $\mathbf{C}$. (equality at $\mathbf{U}^t$)

Examples (1/2)

- **A simple example**: let $\mathbf{M} \in \mathcal{H}_M^{++}$ and $\mathbf{u}_1 \in \mathrm{St}(p, 1)$, consider the problem

$$\begin{array}{ll} \underset{\mathbf{u}_1}{\text{minimize}} & -\mathbf{u}_1^H \, \mathbf{M} \, \mathbf{u}_1 \\ \text{subject to} & \mathbf{u}_1^H \mathbf{u}_1 = 1 \end{array}$$

The solution is obviously the strongest eigenvector $\mathbf{M}$. However... applying Prop.1 yields

$$\mathbf{u}_1^H \mathbf{M} \mathbf{u}_1 \mid \mathbf{u}_1^t \;\; \geq \;\; \left(\mathbf{u}_1^{tH} \mathbf{M}\right) \mathbf{u}_1 + \mathbf{u}_1^H \left(\mathbf{M} \mathbf{u}_1^t\right) + \text{const.},$$

so the Procrustes-MM algorithm is

$$\mathbf{u}_1^{t+1} = \mathcal{P}_{\mathrm{St}} \left\{ \mathbf{M} \mathbf{u}_1^t \right\}$$

and we just rediscovered the power method...

Examples (2/2)

Something more complex but still doable

Denote $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_k]$,

$$\underset{\mathbf{U} \in \mathrm{St}(p,k)}{\text{maximize}} \quad \sum_{i=1}^{k} \left[\frac{\mathbf{u}_i^H \mathbf{A}_i \mathbf{u}_i}{\mathbf{u}_i^H \mathbf{C}_i \mathbf{u}_i} + \mathbf{u}_i^H \mathbf{M}_i \mathbf{u}_i + 2\mathfrak{Re}\left\{ \mathbf{u}_i^H \mathbf{c}_i \right\} \right]$$

Hint: $\mathbf{R}(\mathbf{U}^t) = [\mathbf{R}_1^t \mathbf{u}_1^t, \dots, \mathbf{R}_k^t \mathbf{u}_k^t]$

Application to non-convex RSR

- **Definition**: $\rho : \mathbb{R} \longrightarrow \mathbb{R}^+$ is a concave non-decreasing function

$$\underset{\mathbf{U} \in \mathrm{St}(p,k)}{\text{minimize}} \quad \sum_{i=1}^{n} \rho\left(\mathrm{dist}^2\left(\mathbf{U}, \mathbf{z}_i\right)\right)$$

- **Examples**:
  - Least square: $\rho_{\mathrm{LS}}(t) = t$
  - Huber: $\rho_{\mathrm{Hub}}(t) = \begin{cases} t/\sqrt{T} & \text{if } \leq T \\ 2\sqrt{t} - \sqrt{T} & \text{if } t > T \end{cases}$
  - Cauchy-type: $\rho_{\mathrm{C}}(t) = T \ln(T + t)$
  - Geman-McClure: $\rho_{\mathrm{GMC}}(t) = t/(T + t)$

Procrustes-MM algorithm

Prop.4: At a a given point $\mathbf{U}^t$, the objective function majorized by:

$$g\left(\mathbf{U}|\mathbf{U}^t\right) = -\mathrm{Tr}\{\mathbf{U}^{tH}\mathbf{M}(\mathbf{U}^t)\mathbf{U}\} - \mathrm{Tr}\{\mathbf{U}^H\underbrace{\mathbf{M}(\mathbf{U}^t)\mathbf{U}^t}_{\mathbf{R}(\mathbf{U}^t)}\} + \mathrm{const.}$$

with

$$\mathbf{M}\left(\mathbf{U}\right) = \sum_{i=1}^{n} \rho'\left(\mathrm{dist}^2\left(\mathbf{U}, \mathbf{z}_i\right)\right)\mathbf{z}_i\mathbf{z}_i^H$$

MM algorithm: Since g is linear (A5) we have the updates

$$\mathbf{U}^{t+1} = \mathcal{P}_{\mathrm{St}}\left\{\ \mathbf{M}\left(\mathbf{U}^t\right)\mathbf{U}^{tH}\ \right\}$$

Originally proposed as a fixed-point heuristic in

DIN06 Ding, Zhou, He, Zha, “R1-PCA: rotational invariant ℓ_1-norm principal component analysis for robust subspace factorization,” ACM, 2006

Different algorithms (and computational bottlenecks)

LER17 Quadratic MM, data matrix version — rank-k SVD($p \times n$)

$$\mathbf{U}^{t+1} = \mathcal{P}_k\{\tilde{\mathbf{Z}}_t\}, \text{ with } [\tilde{\mathbf{Z}}_t]_{:,i} = \sqrt{\rho' \left(\mathrm{dist}^2\left(\mathbf{U}^t, \mathbf{z}_i\right)\right)}\mathbf{z}_i$$

MAR05 Fixed point heuristic, covariance matrix version — rank-k SVD($p \times p$)

$$\mathbf{U}^{t+1} = \mathcal{P}_k\left\{\mathbf{M}\left(\mathbf{U}^t\right)\right\}, \text{ with } \mathbf{M}\left(\mathbf{U}^t\right) = \tilde{\mathbf{Z}}_t\tilde{\mathbf{Z}}_t^H$$

MAN02 Steepest descent on Stiefel — $\times$thin-SVDs($p \times k$)

$$\mathbf{U}^{t+1} = \mathcal{P}_{\mathrm{St}}\left\{\mathbf{U}^t + \gamma\,\nabla_f(\mathbf{U}^t)\right\}, \text{ with the right } \gamma$$

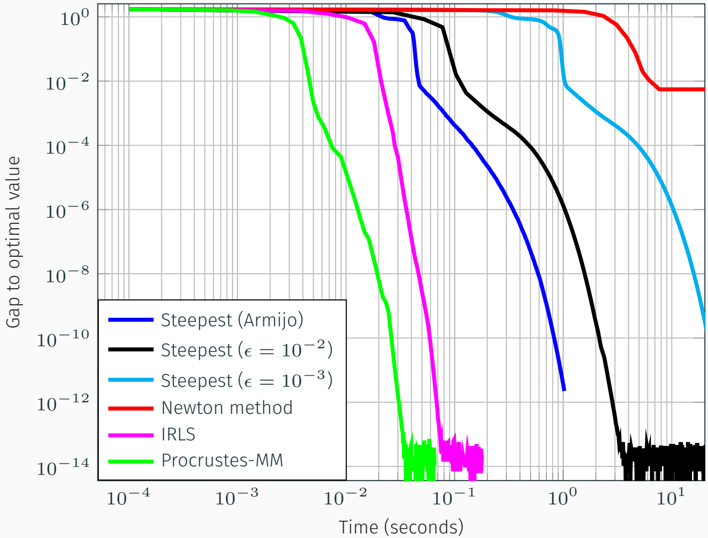
MAN02 Newton method on Stiefel — $(p \times k)^2$ system

$$\mathbf{U}^{t+1} = \mathcal{P}_{\mathrm{St}}\{\mathbf{U}^t + \mathbf{Y}\}, \text{ with } \mathbf{Y} = \mathrm{cpoint}(\mathbf{U}^t, \nabla_f(\mathbf{U}^t), \mathbf{H}_f(\mathbf{U}^t))$$

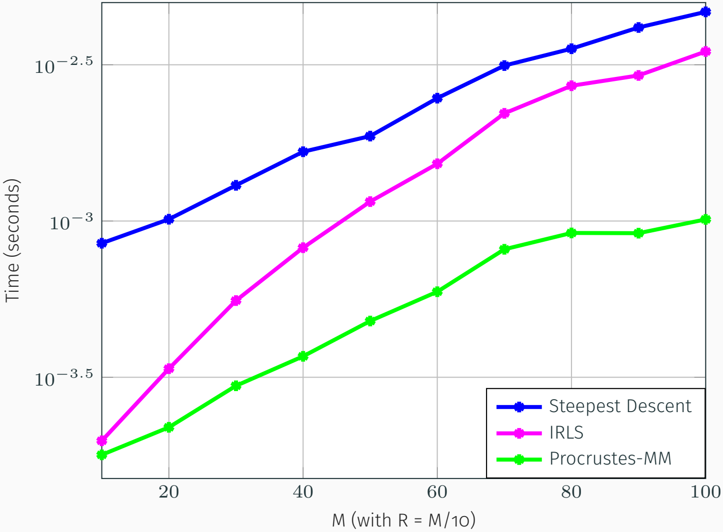
DIN06 Procrustes-MM — thin-SVD($p \times k$)

$$\mathbf{U}^{t+1} = \mathcal{P}_{\mathrm{St}}\left\{\mathbf{M}\left(\mathbf{U}^t\right)\mathbf{U}^t\right\}$$

Objective value (-optimal value) versus CPU time ($p = 30$, $k = 5$, $n = 100$)



Average CPU time of an iteration versus size and rank



Plan overview

$$\underset{\mathbf{U} \in \mathrm{St}(p,k)}{\text{minimize}} \quad f(\mathbf{U})$$

- **Design** the model/objective function f
- **Solve** the constrained minimization problem
- ~~**Analyze** the estimation problem (performance)~~ For another talk ;)
- **Apply** the result to some task

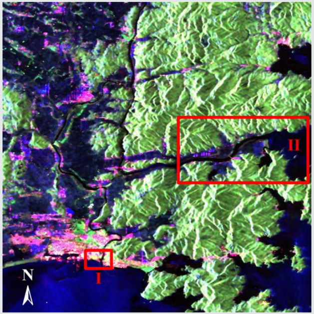
Plan overview

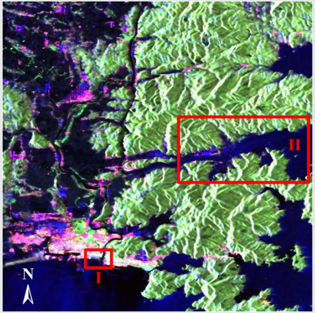
$$\underset{\mathbf{U}\in\mathrm{St}(p,k)}{\text{minimize}} \quad f(\mathbf{U})$$

- **Design** the model/objective function f
- **Solve** the constrained minimization problem
- **Analyze** the estimation problem (performance)
- **Apply** the result to some task

Change detection in satellite image time-series

Monitoring natural disasters:





PolSAR images of Ishinomaki and Onagawa areas [Sato, 2012], Nov.2010 (left), Apr.2011 (right).

Problems to consider

Huge increase in the number of available acquisitions:

- Sentinel-1: 12 days repeat cycle, since 2014
- TerraSAR-X: 11 days repeat cycle, since 2007
- UAVSAR, ...

Detect changes

- Massive amount of data $\longrightarrow$ Automatic process
- Unlabeled data $\longrightarrow$ Unsupervised detection

Chosen approach: detection with a statistical framework

Change detection with GLRT

Parametric probability model

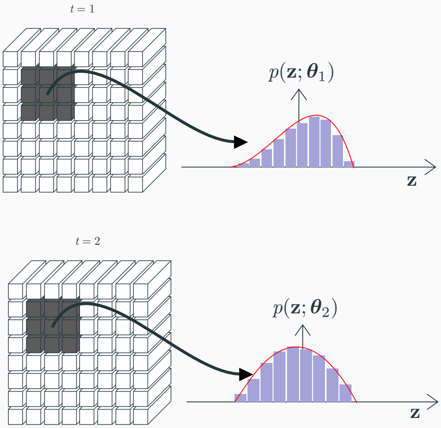
$$\mathbf{Z}_t \sim \mathcal{L}(\mathbf{Z}_t; \boldsymbol{\theta}_t)$$

Hypothesis test

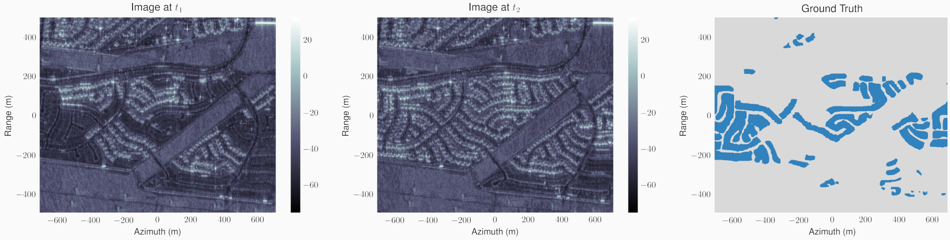
$$\begin{cases} \mathrm{H}_0: & \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \quad (\textit{no change}) \\ \mathrm{H}_1: & \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \quad (\textit{change}) \end{cases}$$

GLRT

$$\frac{\max\limits_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} \mathcal{L}\left(\{\mathbf{Z}_1, \mathbf{Z}_2\}\,;\,\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}\right)}{\max\limits_{\boldsymbol{\theta}_0} \mathcal{L}\left(\{\mathbf{Z}_1, \mathbf{Z}_2\}\,;\,\boldsymbol{\theta}_0\right)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda_{\mathrm{GLRT}}$$



Dataset

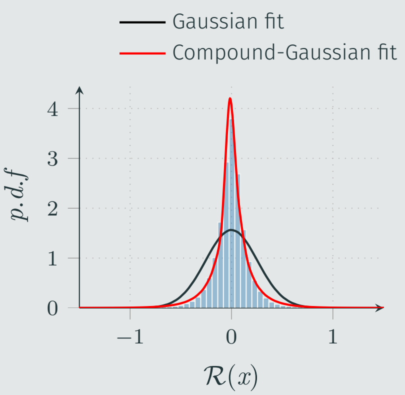


UAVSAR SanAnd_26524_03

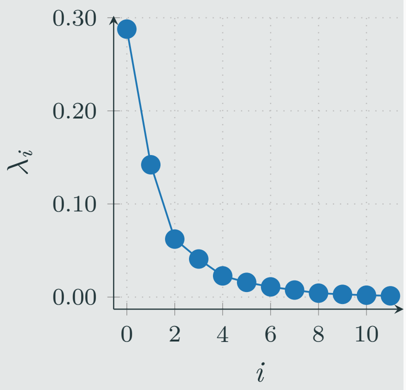
- CD between April 2009 - May 2011 [Nascimento19]
- Polarimetric data $\longrightarrow$ wavelet decomposition $\longrightarrow p = 12$ dim. pixels

Empirical hints for the chosen model

Histogram of UAVSAR data (HH)



Spectrum of UAVSAR data (wavelets)



Covariance based change detection

Models for the GLRT in SAR-ITS: appropriate choice of $\mathcal{L}$ and $\boldsymbol{\theta}$

Gaussian

$$\mathbf{z} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$
$$\boldsymbol{\theta} = \boldsymbol{\Sigma}$$

Low-rank Gaussian

$$\mathbf{z} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_k + \sigma^2\mathbf{I})$$
$$\boldsymbol{\theta} = \boldsymbol{\Sigma}, \text{ with } \mathrm{rank}(\boldsymbol{\Sigma}_k) = k$$

Compound-Gaussian

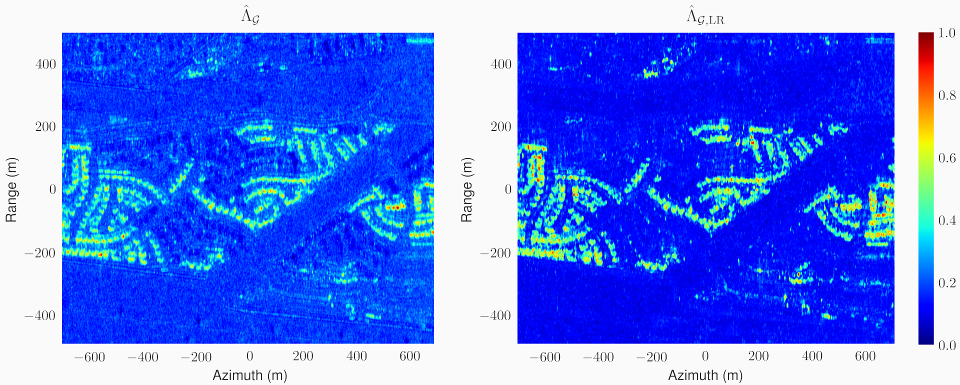
$$\mathbf{z}_i \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tau_i\boldsymbol{\Sigma})$$
$$\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \{\tau_i\}\}$$

Low-rank Compound-Gaussian

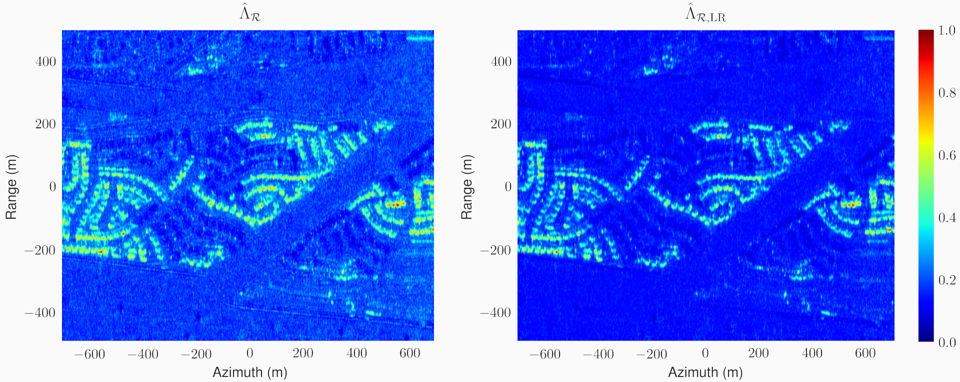
$$\mathbf{z}_i \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tau_i(\boldsymbol{\Sigma}_k + \sigma^2\mathbf{I}))$$
$$\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \{\tau_i\}\}, \text{ with } \mathrm{rank}(\boldsymbol{\Sigma}_k) = k$$

Optimization handled with $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^H$ and previous techniques (MM, Riemannian opt.)

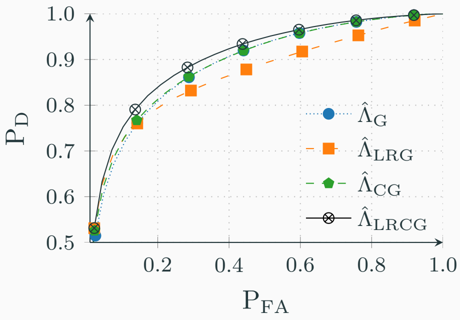
Results with a 5×5 sliding windows: Gaussian detectors



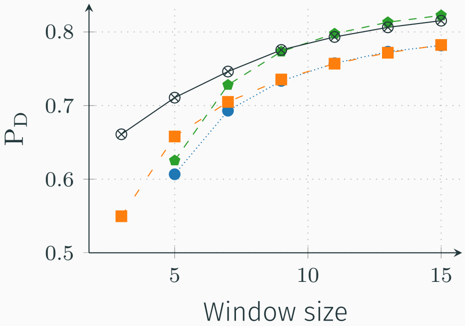
Results with a 5×5 sliding windows: Robust detectors



Performance curves ($p = 12$, $k = 3$)



ROC curves



$\mathrm{P_D}$ *vs window size at* $\mathrm{P_{FA}} = 5\%$

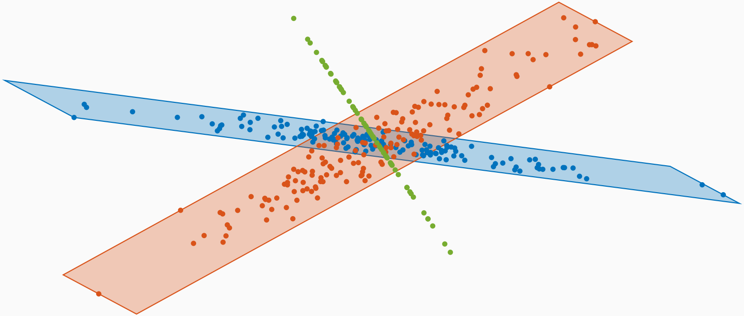
Concluding overview on my works

"Some flavors of PCA"

- **Design**
  - Covariance structures in elliptical models
  - Bayesian priors on orthonormal bases
  - Robust geometric and/or sparse costs
- **Analyze**
  - Intrinsic Cramér-Rao analysis
  - Asymptotic analysis of M-estimators
- **Solve**
  - Majorization-minimization
  - Riemannian optimization
  - $\mathcal{M}$-ADMM
- **Apply**
  - Array processing (detection, DoA)
  - SAR image time-series
  - Clustering w. subspaces as features

Extensions

What if the data looks like this?



Some keywords: mixture of probabilistic PCA, subspace clustering, ...