Online Robust Reduced-Rank Regression

Yangzhuoran Fin Yang
Monash University
Melbourne, Australia

Ziping Zhao
ShanghaiTech University
Shanghai, China

The Eleventh IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)
Hangzhou, China
8th–10th June, 2020

Yangzhuoran Fin Yang
Monash University

Prof. Ziping Zhao
ShanghaiTech University

Outline

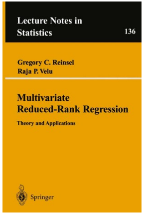
Outline

The Reduced-Rank Regression Models in Data Science

- Reduced-rank regression (RRR) [VelRei'13] is a multivariate linear regression model with a reduced-rank coefficient matrix

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{B}^T\mathbf{x} + \mathbf{D}\mathbf{z} + \boldsymbol{\epsilon}, \qquad \text{(RRR)}$$

  where $\boldsymbol{\mu} \in \mathbb{R}^P$, loading matrix $\mathbf{A} \in \mathbb{R}^{P\times r}$, factor matrix $\mathbf{B} \in \mathbb{R}^{Q\times r}$, $\mathbf{D} \in \mathbb{R}^{Q\times R}$ are coefficients, and $\boldsymbol{\epsilon}$ denotes the model innovation.



- The $\mathbf{A}\mathbf{B}^T$ forms a Stiefel (low-rank) manifold to realize dimension reduction via factors $\mathbf{B}^T\mathbf{x}$.

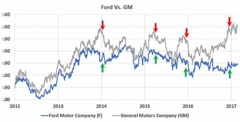
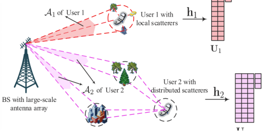


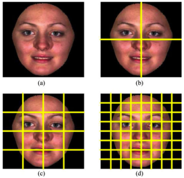
Figure from Internet

Financial econometrics [Johansen'92]



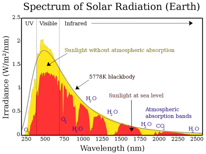
[XieGaoJin'16]

Wireless systems [StankovicHaardt'08]



[Zheng'14]

Computer vision [DongTorre'10]



[WalterSaskaFranchi'18]

Environmental engineering [Glasbey'92]

. . .

Estimation of RRR Models

- Classical ways for RRR parameter estimation (learning) is via Gaussian assumption on innovations $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, i.e.,

$$f(\boldsymbol{\epsilon}) = (2\pi)^{-\frac{k}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}\}.$$

Problem formulation

Given N observed samples $\{\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\}_{i=1}^N$,

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^N \left\|\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}^T\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\right\|_2^2 \qquad \text{(OLSE)}$$

where $\boldsymbol{\theta} \triangleq \{\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}, \mathbf{D}\}$, or

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \frac{N}{2}\log\det(\boldsymbol{\Sigma}) + \frac{1}{2}\sum_{i=1}^N \left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}^T\mathbf{x}_i - \mathbf{D}\mathbf{z}_i)\right\|_2^2 \\ \text{subject to} \quad & \boldsymbol{\Sigma} \succeq \mathbf{0} \end{aligned} \qquad \text{(GMLE)}$$

where $\boldsymbol{\theta} \triangleq \{\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}, \mathbf{D}, \boldsymbol{\Sigma}\}$.

Estimation of RRR Models

- Classical ways for RRR parameter estimation (learning) is via Gaussian assumption on innovations $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, i.e.,

$$f(\boldsymbol{\epsilon}) = (2\pi)^{-\frac{k}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}\}.$$

Problem formulation

Given N observed samples $\{\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\}_{i=1}^N$,

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \tfrac{1}{2}\sum_{i=1}^N \left\|\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}^T\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\right\|_2^2 \qquad \text{(OLSE)}$$

where $\boldsymbol{\theta} \triangleq \{\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}, \mathbf{D}\}$, or

$$\begin{aligned} &\underset{\boldsymbol{\theta}}{\text{minimize}} && \tfrac{N}{2}\log\det(\boldsymbol{\Sigma}) + \tfrac{1}{2}\sum_{i=1}^N \left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{B}^T\mathbf{x}_i - \mathbf{D}\mathbf{z}_i)\right\|_2^2 \\ &\text{subject to} && \boldsymbol{\Sigma} \succeq \mathbf{0} \end{aligned} \qquad \text{(GMLE)}$$

where $\boldsymbol{\theta} \triangleq \{\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}, \mathbf{D}, \boldsymbol{\Sigma}\}$.

Estimation of RRR Models (cont.)

- To solve OLSE/GMLE for RRR, we first examine the first-order optimality conditions for $\{\boldsymbol{\mu}, \mathbf{D}, \boldsymbol{\Sigma}\}$.
- The optimum for $\{\boldsymbol{\mu}, \mathbf{D}, \boldsymbol{\Sigma}\}$ as functions of $\{\mathbf{A}, \mathbf{B}\}$ is

$$\begin{cases} [\boldsymbol{\mu}, \mathbf{D}]\,(\mathbf{A}, \mathbf{B}) = (\mathbf{Y} - \mathbf{A}\mathbf{B}^T\mathbf{X})[\mathbf{1}, \mathbf{Z}^T]\left([\mathbf{1}, \mathbf{Z}^T]^T[\mathbf{1}, \mathbf{Z}^T]\right)^{-1} \\ \boldsymbol{\Sigma}\,(\mathbf{A}, \mathbf{B}) = \frac{1+P}{N}(\mathbf{Y}\mathbf{P} - \mathbf{A}\mathbf{B}^T\mathbf{X}\mathbf{P})(\mathbf{Y}\mathbf{P} - \mathbf{A}\mathbf{B}^T\mathbf{X}\mathbf{P})^T, \end{cases}$$

where $\mathbf{Y} \triangleq [\mathbf{y}_1, \ldots, \mathbf{y}_N]$, $\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, $\mathbf{Z} \triangleq [\mathbf{z}_1, \ldots, \mathbf{z}_N]$, and $\mathbf{P} \triangleq \mathbf{I}_N - \mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{Z}$.

- Substituting $\{\boldsymbol{\mu}, \mathbf{D}, \boldsymbol{\Sigma}\}$ into the objective, we have the subproblems w.r.t. $\{\mathbf{A}, \mathbf{B}\}$

$$\begin{aligned} &\underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad \mathrm{tr}\left[(\mathbf{N} - \mathbf{A}\mathbf{B}^T\mathbf{M})(\mathbf{N} - \mathbf{A}\mathbf{B}^T\mathbf{M})^T\right] \quad \mathbf{A}\mathbf{B}^T\text{-subprob. in OLSE} \\ &\underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad \det\left[(\mathbf{N} - \mathbf{A}\mathbf{B}^T\mathbf{M})(\mathbf{N} - \mathbf{A}\mathbf{B}^T\mathbf{M})^T\right] \quad \mathbf{A}\mathbf{B}^T\text{-subprob. in GMLE,} \end{aligned}$$

where $\mathbf{N} \triangleq \mathbf{Y}\mathbf{P}$ and $\mathbf{M} \triangleq \mathbf{X}\mathbf{P}$.

Estimation of RRR Models (cont.)

Proposition (Johansen'91, Lütkepohl'05)

Define $\mathbf{R}_{mm} \triangleq \mathbf{M}\mathbf{M}^T$, $\mathbf{R}_{mn} \triangleq \mathbf{M}\mathbf{N}^T$, $\mathbf{R}_{nm} \triangleq \mathbf{N}\mathbf{M}^T = \mathbf{R}_{mn}^T$, and $\mathbf{R}_{nn} \triangleq \mathbf{N}\mathbf{N}^T$. The optimum in OLSE/GMLE subproblems with respect to $\mathbf{A}$ and $\mathbf{B}$ is obtained at

$$\mathbf{A}^\star = \mathbf{R}_{nm}\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{U}_r \quad \text{and} \quad \mathbf{B}^\star = \mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{U}_r, \tag{1}$$

where $\mathbf{U}_r \in \mathbb{R}^{Q \times r}$ contains the left singular vectors corresponding to the r largest singular values of matrix $\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{R}_{mn}\mathbf{R}_{nn}^{-\frac{1}{2}}$ sorted in nonincreasing order.

- As expected, for deterministic estimation scheme under Gaussianity, OLSE is equivalent to GMLE.
  - ☺ closed-form solution
  - ☹ not adaptive to outliers and large-scale data

In data analytics, deterministic Gaussian estimation is inefficient with **outliers** and **data proliferation**.

Estimation of RRR Models (cont.)

Proposition (Johansen'91, Lütkepohl'05)

Define $\mathbf{R}_{mm} \triangleq \mathbf{M}\mathbf{M}^T$, $\mathbf{R}_{mn} \triangleq \mathbf{M}\mathbf{N}^T$, $\mathbf{R}_{nm} \triangleq \mathbf{N}\mathbf{M}^T = \mathbf{R}_{mn}^T$, and $\mathbf{R}_{nn} \triangleq \mathbf{N}\mathbf{N}^T$. The optimum in OLSE/GMLE subproblems with respect to $\mathbf{A}$ and $\mathbf{B}$ is obtained at

$$\mathbf{A}^\star = \mathbf{R}_{nm}\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{U}_r \quad \text{and} \quad \mathbf{B}^\star = \mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{U}_r, \tag{1}$$

where $\mathbf{U}_r \in \mathbb{R}^{Q \times r}$ contains the left singular vectors corresponding to the r largest singular values of matrix $\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{R}_{mn}\mathbf{R}_{nn}^{-\frac{1}{2}}$ sorted in nonincreasing order.

- As expected, for deterministic estimation scheme under Gaussianity, OLSE is equivalent to GMLE.
  - ☺ closed-form solution
  - ☹ not adaptive to outliers and large-scale data

In data analytics, deterministic Gaussian estimation is inefficient with **outliers** and **data proliferation**.

Estimation of RRR Models (cont.)

Proposition (Johansen'91, Lütkepohl'05)

Define $\mathbf{R}_{mm} \triangleq \mathbf{M}\mathbf{M}^T$, $\mathbf{R}_{mn} \triangleq \mathbf{M}\mathbf{N}^T$, $\mathbf{R}_{nm} \triangleq \mathbf{N}\mathbf{M}^T = \mathbf{R}_{mn}^T$, and $\mathbf{R}_{nn} \triangleq \mathbf{N}\mathbf{N}^T$. The optimum in OLSE/GMLE subproblems with respect to $\mathbf{A}$ and $\mathbf{B}$ is obtained at

$$\mathbf{A}^\star = \mathbf{R}_{nm}\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{U}_r \quad \text{and} \quad \mathbf{B}^\star = \mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{U}_r, \tag{1}$$

where $\mathbf{U}_r \in \mathbb{R}^{Q\times r}$ contains the left singular vectors corresponding to the r largest singular values of matrix $\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{R}_{mn}\mathbf{R}_{nn}^{-\frac{1}{2}}$ sorted in nonincreasing order.

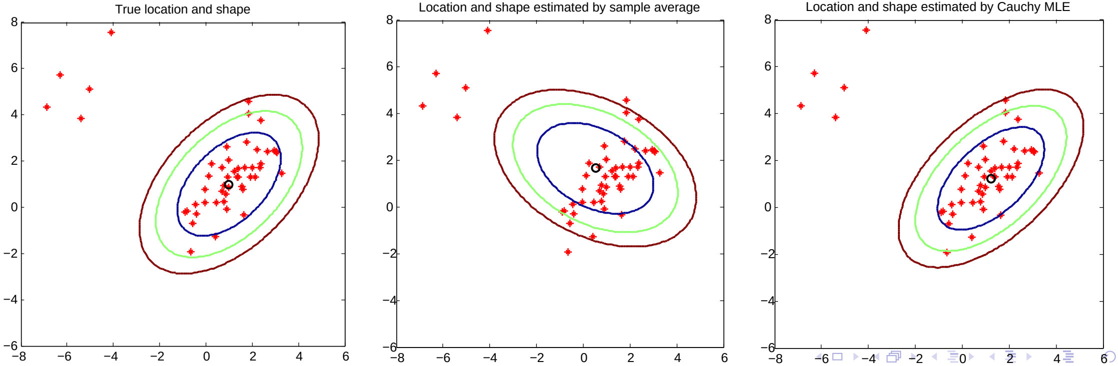
- As expected, for deterministic estimation scheme under Gaussianity, OLSE is equivalent to GMLE.
  - ☺ closed-form solution
  - ☹ not adaptive to outliers and large-scale data

In data analytics, deterministic Gaussian estimation is inefficient with **outliers** and **data proliferation**.

Outline

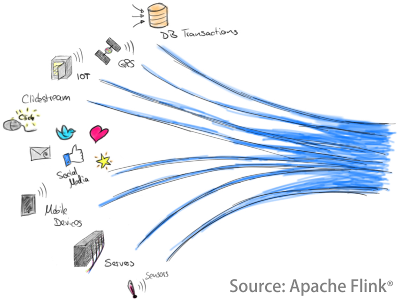
Target 1: Robust Against Heavy-tails and Outliers

- For applications especially involved in big data, the data to analyze often exhibit features of heavy-tails and outliers [George'14].
- Robust pursuit for RRR is essential in scenarios like
  - statistical robust channel estimation [Garcia'06],
  - erratic seismic noise attenuation [ChenSacchi'15],
  - genetic analysis [SheChen'17], etc.



Target 2: Online Estimation under Large-Scale Dataset

- With the proliferation of data, data dimension can be huge or the data is continuously collected as streams.
- Dealing with data batch can be computationally expensive.
- Model needs to be updated with new information coming in.
  - online channel estimation [Venkateswaran'10]
  - online cointegration detection in finance [ZhaoPalomar'17]
  - real-time functional magnetic resonance imaging [Ulfarsson'10]
  - etc.



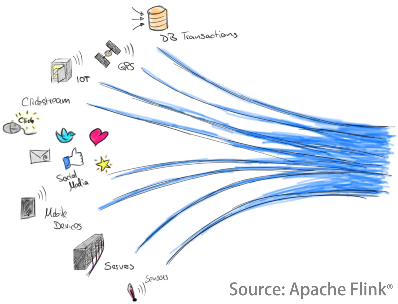
Source: Apache Flink®

Research Question: Design an RRR estimation scheme which is robust and amenable to large datasets?

Yes, this paper!

Target 2: Online Estimation under Large-Scale Dataset

- With the proliferation of data, data dimension can be huge or the data is continuously collected as streams.
- Dealing with data batch can be computationally expensive.
- Model needs to be updated with new information coming in.
  - online channel estimation [Venkateswaran'10]
  - online cointegration detection in finance [ZhaoPalomar'17]
  - real-time functional magnetic resonance imaging [Ulfarsson'10]
  - etc.



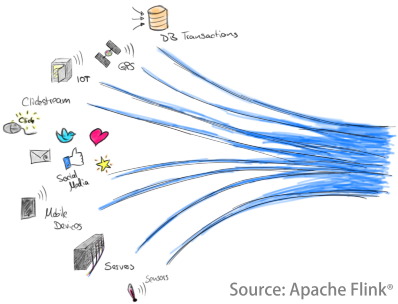
Source: Apache Flink®

Research Question: Design an RRR estimation scheme which is robust and amenable to large datasets?

Yes, this paper!

Target 2: Online Estimation under Large-Scale Dataset

- With the proliferation of data, data dimension can be huge or the data is continuously collected as streams.
- Dealing with data batch can be computationally expensive.
- Model needs to be updated with new information coming in.
  - online channel estimation [Venkateswaran'10]
  - online cointegration detection in finance [ZhaoPalomar'17]
  - real-time functional magnetic resonance imaging [Ulfarsson'10]
  - etc.



Source: Apache Flink®

Research Question: Design an RRR estimation scheme which is robust and amenable to large datasets?

Yes, this paper!

Online RRRR: Robust Reduced-Rank Regression with Data Streams

- To promote robustness, Cauchy distribution assumption is used [Huber'11].
- Assume $\boldsymbol{\epsilon} \sim \text{Cauchy}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \in \mathbb{S}_{++}^{P}$, then its p.d.f. is

$$f(\boldsymbol{\epsilon}) = \frac{\Gamma[(1+P)/2]}{\Gamma(1/2)\pi^{P/2}\det(\boldsymbol{\Sigma})^{1/2}}\left(1+\boldsymbol{\epsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\right)^{-\frac{1+P}{2}},$$

and the negative log-likelihood function of a single observation $\boldsymbol{\xi} \triangleq \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is defined as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\xi}) \triangleq \frac{1}{2}\log\det(\boldsymbol{\Sigma}) + \frac{1+P}{2}\log\left[1 + (\mathbf{y}-\boldsymbol{\mu}-\mathbf{A}\mathbf{B}^T\mathbf{x}-\mathbf{D}\mathbf{z})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}-\mathbf{A}\mathbf{B}^T\mathbf{x}-\mathbf{D}\mathbf{z})\right].$$

Problem formulation for Online RRRR

Given the loss function $\ell(\boldsymbol{\theta}, \boldsymbol{\xi})$ for one sample $\boldsymbol{\xi}$, the online estimation problem is given as

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \left[L(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{\xi}}[\ell(\boldsymbol{\theta}, \boldsymbol{\xi})]\right] \\ \text{subject to} \quad & \boldsymbol{\Sigma} \succeq \mathbf{0}. \end{aligned} \qquad \text{(CMLE)}$$

- This is a highly nonconvex constrained stochastic optimization problem.

Outline

Stochastic Estimation via SMM

- A classical approach for solving the "CMLE" problem is the SAA method [Plambeck'96].

Sample Average Approximation (SAA)

For function $\ell(\cdot)$ of parameter $\boldsymbol{\theta}$ and data $\boldsymbol{\xi}_i$, the optimization step at iteration k with $N^{(k)}$ samples is

$$\boldsymbol{\theta}^{(k)} \leftarrow \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \ell(\boldsymbol{\theta}, \boldsymbol{\xi}_i).$$

- Since the function $\ell(\cdot)$ is non-convex, per-iteration SAA computation is computationally expensive.

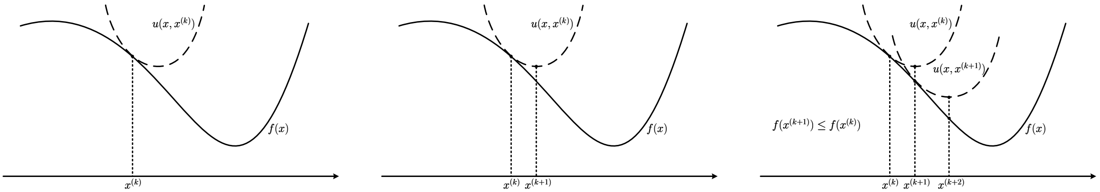
Stochastic Majorization Minimization (SMM) [Razaviyayn'16]

SMM in each iteration optimizes over a surrogate majorizing (upper-bound) function $\bar{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}, \boldsymbol{\xi}_i)$ as

$$\boldsymbol{\theta}^{(k)} \leftarrow \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \bar{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}, \boldsymbol{\xi}_i),$$

The majorizing function is commonly chosen to be strictly convex or one leading to a closed-form solution.

Find a Majorizing Function in SMM



- It is easy to see that the key in using SMM is to find a good majorizing function $\bar{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}, \boldsymbol{\xi}_i)$.

Lemma: Linear majorization [ZhaoPalomar'17]

Given $\boldsymbol{\theta}^{(k-1)}$, the loss function $\ell(\boldsymbol{\theta}, \boldsymbol{\xi}_i)$ can be majorized as

$$\bar{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}, \boldsymbol{\xi}_i) \triangleq \frac{1}{2}\log\det(\boldsymbol{\Sigma}) + \frac{1+P}{2}\left(\bar{\mathbf{y}}_i^{(k)} - \sqrt{w_i^{(k)}}\boldsymbol{\mu} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{x}}_i^{(k)} - \mathbf{D}\bar{\mathbf{z}}_i^{(k)}\right)^T \times \boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{y}}_i^{(k)} - \sqrt{w_i^{(k)}}\boldsymbol{\mu} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{x}}_i^{(k)} - \mathbf{D}\bar{\mathbf{z}}_i^{(k)}\right) + const.,$$

where $\bar{\mathbf{y}}_i^{(k)} \triangleq \sqrt{w_i^{(k)}}\mathbf{y}_i$, $\bar{\mathbf{x}}_i^{(k)} \triangleq \sqrt{w_i^{(k)}}\mathbf{x}_i$, $\bar{\mathbf{z}}_i^{(k)} \triangleq \sqrt{w_i^{(k)}}\mathbf{z}_i$, and the weight $w_i^{(k)}$ is a function of $\boldsymbol{\theta}^{(k-1)}$.

Solving the Subproblem in SMM

- The majorized objective function becomes

$$\bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \triangleq \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \bar{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_i)$$

The subproblem in SMM in each iteration k is

$$\begin{array}{ll} \underset{\boldsymbol{\theta}}{\text{minimize}} & \bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \\ \text{subject to} & \boldsymbol{\Sigma} \succeq \mathbf{0}, \end{array} \qquad \text{(Majorized Subprob. in CMLE)}$$

- Examining the first order optimality condition for $\boldsymbol{\mu}$, $\mathbf{D}$, and $\boldsymbol{\Sigma}$, we have

$$\begin{cases} [\boldsymbol{\mu}, \mathbf{D}]\,(\mathbf{A}, \mathbf{B}) = (\bar{\mathbf{Y}}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)})\mathbf{Q}^{(k)T}(\mathbf{Q}^{(k)}\mathbf{Q}^{(k)T})^{-1}, \\ \boldsymbol{\Sigma}\,(\mathbf{A}, \mathbf{B}) = \frac{1+P}{N^{(k)}}(\bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)})(\bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)})^T, \end{cases}$$

where $\mathbf{Q}^{(k)} \triangleq \left[\sqrt{\mathbf{w}^{(k)}},\ \bar{\mathbf{Z}}^{(k)T}\right]^T$ and $\mathbf{P} \triangleq \mathbf{I}_N - \mathbf{Q}^{(k)T}(\mathbf{Q}^{(k)}\mathbf{Q}^{(k)T})^{-1}\mathbf{Q}^{(k)}$.

Solving the Subproblem in SMM

- The majorized objective function becomes

$$\bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \triangleq \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \bar{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}_i)$$

The subproblem in SMM in each iteration k is

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \\ \text{subject to} \quad & \boldsymbol{\Sigma} \succeq \mathbf{0}, \end{aligned} \qquad \text{(Majorized Subprob. in CMLE)}$$

- Examining the first order optimality condition for $\boldsymbol{\mu}$, $\mathbf{D}$, and $\boldsymbol{\Sigma}$, we have

$$\begin{cases} [\boldsymbol{\mu}, \mathbf{D}]\,(\mathbf{A}, \mathbf{B}) = \left(\bar{\mathbf{Y}}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\right)\mathbf{Q}^{(k)T}\left(\mathbf{Q}^{(k)}\mathbf{Q}^{(k)T}\right)^{-1}, \\ \qquad \boldsymbol{\Sigma}\,(\mathbf{A}, \mathbf{B}) = \frac{1+P}{N^{(k)}}\left(\bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)}\right)\left(\bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)}\right)^T, \end{cases}$$

where $\mathbf{Q}^{(k)} \triangleq \left[\sqrt{\mathbf{w}^{(k)}},\ \bar{\mathbf{Z}}^{(k)T}\right]^T$ and $\mathbf{P} \triangleq \mathbf{I}_N - \mathbf{Q}^{(k)T}(\mathbf{Q}^{(k)}\mathbf{Q}^{(k)T})^{-1}\mathbf{Q}^{(k)}$.

Solving the Subproblem in SMM

- Given $[\boldsymbol{\mu}, \mathbf{D}]\,(\mathbf{A}, \mathbf{B})$ and $\boldsymbol{\Sigma}\,(\mathbf{A}, \mathbf{B})$, the objective $\bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ becomes

$$\bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{N^{(k)}}{2} \log\det\left[\frac{1+P}{N^{(k)}}\left(\bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)}\right)\left(\bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)} - \mathbf{A}\mathbf{B}^T\bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)}\right)^T\right] + const.$$

The subproblem now becomes

$$\underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad \det\left[\left(\mathbf{N}^{(k)} - \mathbf{A}\mathbf{B}^T\mathbf{M}^{(k)}\right)\left(\mathbf{N}^{(k)} - \mathbf{A}\mathbf{B}^T\mathbf{M}^{(k)}\right)^T\right], \quad \text{(Majorized } \mathbf{A}\mathbf{B}^T\text{-Subprob. in CMLE)}$$

where $\mathbf{N}^{(k)} = \bar{\mathbf{Y}}^{(k)}\mathbf{P}^{(k)}$ and $\mathbf{M}^{(k)} = \bar{\mathbf{X}}^{(k)}\mathbf{P}^{(k)}$.

The solution to "Majorized $\mathbf{A}\mathbf{B}^T$-Subproblem in CMLE" is the solution in "$\mathbf{A}\mathbf{B}^T$-Subproblem in GMLE"!

Solving CMLE is just iteratively solving the GMLE!

SMM-Based Algorithm for Online RRRR

Online RRRR via SMM

Input: Training data $\{\boldsymbol{\xi}_i\}_{i=1}^{\infty}$, the initial parameter $\boldsymbol{\theta}^{(0)} \in \boldsymbol{\Theta}$, and $k = 1$;
For $(i = 1, \ldots)$

1. Calculate $\{\mathbf{w}^{(k)}, \bar{\mathbf{Y}}^{(k)}, \bar{\mathbf{X}}^{(k)}, \bar{\mathbf{Z}}^{(k)}, \mathbf{Q}^{(k)}, \mathbf{P}^{(k)}\}$ based on the parameter $\boldsymbol{\theta}^{(k)}$ and data $\{\boldsymbol{\xi}_i\}_{i=1}^{N^{(k)}}$;
2. Calculate $\{\mathbf{M}^{(k)}, \mathbf{N}^{(k)}, \mathbf{R}_{mm}^{(k)}, \mathbf{R}_{mn}^{(k)}, \mathbf{R}_{nn}^{(k)}, \}$;
3. Compute r left singular vectors of $\mathbf{R}_{mm}^{-\frac{1}{2}}\mathbf{R}_{mn}\mathbf{R}_{nn}^{-\frac{1}{2}}$;
4. Update $\boldsymbol{\theta}^{(k+1)}$; $k \leftarrow k+1$;

Output: $\boldsymbol{\theta}^{(k)} = \{\boldsymbol{\mu}^{(k)}, \mathbf{A}^{(k)}, \mathbf{B}^{(k)}, \mathbf{D}^{(k)}, \boldsymbol{\Sigma}^{(k)}\}$.

Can we design a **unified algorithm framework** for Online RRRR via ~~CMLE~~ GENERAL loss function?

A class of cases can be solved by the SMM algorithm with OLSE/GMLE solutions [ZhouZhaoPal'18]!

SMM-Based Algorithm for Online RRRR

Online RRRR via SMM

Input: Training data $\{\xi_i\}_{i=1}^{\infty}$, the initial parameter $\boldsymbol{\theta}^{(0)} \in \boldsymbol{\Theta}$, and $k = 1$;
For $(i = 1, \ldots)$

1. Calculate $\{\mathbf{w}^{(k)}, \bar{\mathbf{Y}}^{(k)}, \bar{\mathbf{X}}^{(k)}, \bar{\mathbf{Z}}^{(k)}, \mathbf{Q}^{(k)}, \mathbf{P}^{(k)}\}$ based on the parameter $\boldsymbol{\theta}^{(k)}$ and data $\{\xi_i\}_{i=1}^{N^{(k)}}$;
2. Calculate $\{\mathbf{M}^{(k)}, \mathbf{N}^{(k)}, \mathbf{R}_{mm}^{(k)}, \mathbf{R}_{mn}^{(k)}, \mathbf{R}_{nn}^{(k)}, \}$;
3. Compute r left singular vectors of $\mathbf{R}_{mm}^{-\frac{1}{2}} \mathbf{R}_{mn} \mathbf{R}_{nn}^{-\frac{1}{2}}$;
4. Update $\boldsymbol{\theta}^{(k+1)}$; $k \leftarrow k + 1$;

Output: $\boldsymbol{\theta}^{(k)} = \{\boldsymbol{\mu}^{(k)}, \mathbf{A}^{(k)}, \mathbf{B}^{(k)}, \mathbf{D}^{(k)}, \boldsymbol{\Sigma}^{(k)}\}$.

Can we design a **unified algorithm framework** for Online RRRR via ~~CMLE~~ GENERAL loss function?

A class of cases can be solved by the SMM algorithm with OLSE/GMLE solutions [ZhouZhaoPal'18]!

SMM-Based Algorithm for Online RRRR

Online RRRR via SMM

Input: Training data $\{\boldsymbol{\xi}_i\}_{i=1}^{\infty}$, the initial parameter $\boldsymbol{\theta}^{(0)} \in \boldsymbol{\Theta}$, and $k=1$;
For $(i = 1, \ldots)$

1. Calculate $\{\mathbf{w}^{(k)}, \bar{\mathbf{Y}}^{(k)}, \bar{\mathbf{X}}^{(k)}, \bar{\mathbf{Z}}^{(k)}, \mathbf{Q}^{(k)}, \mathbf{P}^{(k)}\}$ based on the parameter $\boldsymbol{\theta}^{(k)}$ and data $\{\boldsymbol{\xi}_i\}_{i=1}^{N^{(k)}}$;
2. Calculate $\{\mathbf{M}^{(k)}, \mathbf{N}^{(k)}, \mathbf{R}_{mm}^{(k)}, \mathbf{R}_{mn}^{(k)}, \mathbf{R}_{nn}^{(k)}, \}$;
3. Compute r left singular vectors of $\mathbf{R}_{mm}^{-\frac{1}{2}} \mathbf{R}_{mn} \mathbf{R}_{nn}^{-\frac{1}{2}}$;
4. Update $\boldsymbol{\theta}^{(k+1)}$; $k \leftarrow k+1$;

Output: $\boldsymbol{\theta}^{(k)} = \{\boldsymbol{\mu}^{(k)}, \mathbf{A}^{(k)}, \mathbf{B}^{(k)}, \mathbf{D}^{(k)}, \boldsymbol{\Sigma}^{(k)}\}$.

Can we design a **unified algorithm framework** for Online RRRR via ~~CMLE~~ GENERAL loss function?

A class of cases can be solved by the SMM algorithm with OLSE/GMLE solutions [ZhouZhaoPal'18]!

Outline

Simulation Setup

- A RRR model is specified with $P = Q = 10$ and $r = R = 1$.
- A path of 1000 samples is generated where innovations follow a Student's t-distribution with degree of freedom of 3 mimicking the real data scenarios.
- In the online estimation, we start with 25 samples and 1 sample is added in each iteration.

Comparisons on Convergence and Estimation Accuracy

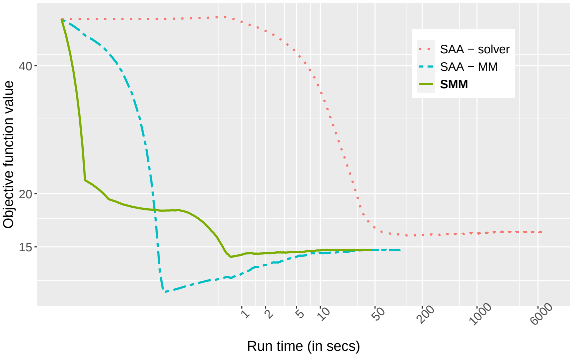


Figure: Alg. convergence (average over 30 MC runs)

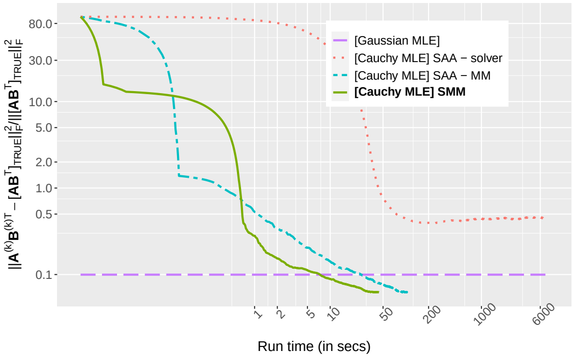


Figure: Estimation accuracy (average over 30 MC runs)

Comparisons on Runtime

- To show the computational efficiency, the estimation time is compared by varying the parameter dimensions where $P = Q$ and $r = R = 1$ based on 100 Monte Carlo simulations.

Table: Average runtime (in secs) with standard error in parentheses

$(P,\ Q)$	$(5,\ 5)$	$(10,\ 10)$	$(20,\ 20)$	$(30,\ 30)$
SAA - MM	48.0 (16.8)	69.9 (17.7)	153.2 (27.6)	264.9 (27.2)
SMM	**17.2 (6.85)**	**25.4 (7.49)**	**54.0 (11.04)**	**88.5 (11.06)**

- SMM consistently runs faster and more stable than the SAA and scales well with the dimension.

An R Package

The proposed algorithm has been provided in open source ☺.

- You are welcome to download it from the GitHub

```
devtools::install_github("finyang/RRRR")
```

or from the Comprehensive R Archive Network (CRAN).

```
install.packages("RRRR")
```

- On CRAN, the package has achieved a total download of downloads 1134.

Outline

Conclusions

- We have discussed the online robust reduced-rank regression problem.
- An efficient algorithm based on the stochastic majorization minimization has been proposed.
- The effectiveness of the model and algorithm has been demonstrated via simulations.

Selected References I

Bernardini, E. and Cubadda, G. (2015).
Macroeconomic forecasting and structural analysis through regularized reduced-rank regression.
International Journal of Forecasting, 31(3):682–691.

Chen, K. and Sacchi, M. D. (2015).
Robust reduced-rank filtering for erratic seismic noise attenuation.
Geophysics, 80(1):V1–V11.

George, G., Haas, M. R., and Pentland, A. (2014).
Big data and management.

Glasbey, C. A. (1992).
A reduced rank regression model for local variation in solar radiation.
Journal of the Royal Statistical Society. Series C (Applied Statistics), 41(2):381.

Gustafsson, T. and Rao, B. D. (2002).
Statistical analysis of subspace-based estimation of reduced-rank linear regressions.
IEEE Transactions on Signal Processing, 50(1):151–159.

Hua, Y., Nikpour, M., and Stoica, P. (2001).
Optimal reduced-rank estimation and filtering.
IEEE Transactions on Signal Processing, 49(3):457–469.

Huang, D. and De la Torre, F. (2010).
Bilinear kernel reduced rank regression for facial expression synthesis.
In *European Conference on Computer Vision*, pages 364–377. Springer.

Selected References II

Huber, P. J. (2011).
Robust statistics.
Springer.

Johansen, S. (1991).
Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models.
Econometrica, 59(6):1551.

Johansen, S. and Juselius, K. (1992).
Some structural hypotheses in a multivariate cointegration analysis of the purchasing power parity and the uncovered interest parity for UK.
Journal of Econometrics, 53(1-3):211–244.

Nicoli, M. and Spagnolini, U. (2005).
Reduced-rank channel estimation for time-slotted mobile communication systems.
IEEE Transactions on Signal Processing, 53(3):926–944.

Plambeck, E. L., Fu, B. R., Robinson, S. M., and Suri, R. (1996).
Sample-path optimization of convex stochastic performance functions.
Mathematical Programming, Series B, 75(2):137–176.

Razaviyayn, M., Sanjabi, M., and Luo, Z. Q. (2016).
A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks.
Mathematical Programming, 157(2):515–545.

She, Y. and Chen, K. (2017).
Robust reduced-rank regression.
Biometrika, 104(3):633–647.

Selected References III

Ulfarsson, M. O. and Solo, V. (2011).
Sparse variable reduced rank regression via stiefel optimization.
In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3892–3895. ieeexplore.ieee.org.

Velu, R. and Reinsel, G. C. (2013).
Multivariate Reduced-Rank Regression: Theory and Applications.
Springer Science & Business Media.

Venkateswaran, V. and others (2010).
Analog beamforming in MIMO communications with phase shift networks and online channel estimation.
IEEE Transactions on.

Walter, V., Saska, M., and Franchi, A. (2018).
Fast mutual relative localization of uavs using ultraviolet led markers.
In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 1217–1226. IEEE.

Xie, H., Gao, F., and Jin, S. (2016).
An overview of low-rank channel estimation for massive mimo systems.
IEEE Access, 4:7313–7321.

Zhao, Z. and Palomar, D. P. (2017).
Robust maximum likelihood estimation of sparse vector error correction model.
In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 913—-917. IEEE.

Zheng, W. (2014).
Multi-view facial expression recognition based on group sparse reduced-rank regression.
IEEE Transactions on Affective Computing, 5(1):71–85.

Thanks!

For more information:

zipingzhao.github.io