A Rank-Constrained Optimization approach: Application to Factor Analysis

Ramón A. Delgado, Juan C. Agüero, Graham C. Goodwin

School of Electrical Engineering and Computer Science,
The University of Newcastle,
Australia.

THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

Motivation

why include rank constraints?

- Complexity of a model (e.g rank of a Hankel matrix)
- Low-Rank Decomposition (Principal Components Analysis)
- Recent interest on sparse representations.

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Some Difficulties with rank constraints

$$\mathrm{rank}\,\{X\} \leq r$$

- Non-differentiable
- Nonlinear
- Combinatorial nature in optimization

Find a differentiable representation for the rank constraints, and hopefully reduce the number of nonlinearities.

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Some existing representations
By construction

- $\text{rank}\{X\} \leq r \iff X = A \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} B$.

- $\text{rank}\{X\} \leq r \iff X = UV$.

where $A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{n\times n}, U \in \mathbb{R}^{m\times r}, V \in \mathbb{R}^{r\times n}$.

Disadvantage

Already assign a structure to X.

Some existing representations
Using the Characteristic Polynomial

Consider that $c_i(X)$, for $i = 1, \ldots, n$ are the coefficients of the characteristic polynomial of X.

- $X \in \mathbb{S}_+^n$, [Helmersson 2009]
  $\operatorname{rank}\{X\} \leq r \iff c_{n-r-1}(-X) = 0$.

- $X \in \mathbb{S}_+^n$, [d'Aspremont 2003]
  $\operatorname{rank}\{X\} = \min_{v \in \mathbb{R}^n} \sum_{i=1}^n v_i$
  s.t. $c_i(G)(1 - v_i) = 0, \quad v_i \geq 0$ for $i = 1, \ldots, n$.

Disadvantages

- Only valid for $X \in \mathbb{S}_+^n$.
- $c_i(X)$ is, in general, nonlinear.

Some existing representations used in optimization

Closely related to our representation

I. Markovsky $\operatorname{rank}\{X\} \leq r \iff \exists R \in \mathbb{R}^{(m-r)\times m}$ such that
$RX = 0_{m-r\times m}$ and R is full row rank

$X \in \mathbb{R}^{m\times n}$, but a rank constraint is now imposed on an auxiliary matrix.

J. Dattorro $X \in \mathbb{S}^n_+$,
$\operatorname{rank}\{X\} \leq r \iff \exists W \in \Phi_{n,r}$ such that $\operatorname{trace}(WX) = 0$.

where $\Phi_{n,r} = \{W \in \mathbb{S}^n,\ 0 \preceq W \preceq I, \operatorname{trace}(W) = n - r\}$

Only valid for $X \in \mathbb{S}^n_+$.

Main result[1]

Theorem

Let $X \in \mathbb{R}^{m\times n}$ *then*

$$\operatorname{rank}\{X\} \leq r \iff \exists\, W \in \Phi_{n,r},\ \textit{such that}\ GX = 0_{m\times n}$$

where

$$\Phi_{n,r} = \{W \in \mathbb{S}^n,\ 0 \preceq W \preceq I, \operatorname{trace}(W) = n - r\}$$

[1] submitted for publication

Advantages

- Differentiable
- Freedom to impose a desired structure on X (e.g. Hankel).
- Generalisation of Dattorro's result
- Avoid some difficulties in Markovsky and d'Aspremont's results.

Disadvantage

We still have a bilinear condition $WX = 0_{m\times n}$.

Rank-Constrained Optimisation

An equivalent representation for rank-constrained optimization problems

$$
\begin{aligned}
\mathcal{P}_{rco} : \quad & \min_{\theta \in \mathbb{R}^p} \; f(\theta) \\
\text{s.t.} \quad & \theta \in \Omega \\
& \operatorname{rank}\{X(\theta)\} \le r
\end{aligned}
\qquad \equiv \qquad
\begin{aligned}
\mathcal{P}_{bi} : \quad & \min_{\theta \in \mathbb{R}^p} \min_{W \in \mathbb{S}^n} \; f(\theta) \\
\text{s.t.} \quad & \theta \in \Omega \\
& X(\theta) W = 0_{m \times n} \\
& W \in \Phi_{n,r}
\end{aligned}
$$

Rank-Constrained Optimisation

We have developed

- a *local optimisation* method
- For the case $X(\theta) \in \mathbb{S}^n_+$, a *global optimisation* method.

The bilinear constraint can be imposed in several ways:

$$X(\theta) \in \mathbb{R}^{m\times n} \qquad \|X(\theta)W\| = 0 \iff X(\theta)W = 0_{m\times n}$$

$$X(\theta) \in \mathbb{S}^n_+ \qquad \operatorname{trace}(X(\theta)W) = 0 \iff X(\theta)W = 0_{n\times n}$$

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Factor Analysis

Consider a measured output $y_k \in \mathbb{R}^N$, factors $f_k \in \mathbb{R}^r$, idiosyncratic noise $v_k \in \mathbb{R}^N$, and a model:

$$y_k = Af_k + v_k \tag{1}$$

where $A \in \mathbb{R}^{N \times n}$ is a tall matrix.

$$f_k \sim \mathcal{N}(0, \Phi) \tag{2}$$
$$v_k \sim \mathcal{N}(0, \Psi) \tag{3}$$

Then, $y_k \sim \mathcal{N}(0, \Sigma)$, where Σ is given by

$$\Sigma = A\Phi A^\top + \Psi$$

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Sparse Noise Covariance

- Existing approaches require Ψ diagonal.
  - $\Psi = \sigma^2 I$, then there is a closed-form solution, e.g. PCA.
  - Ψ diagonal, e.g. Maximum Likelihood (EM algorithm).
- Considering the advances on sparse representations. We propose to assume that Ψ is sparse.

$$
\begin{aligned}
\mathcal{P}_{rcofa}: \quad & \min_{\Psi \in \mathbb{S}^N} && \|\Psi\|_1 \\
& \text{subject to} && \operatorname{rank}\{\Sigma - \Psi\} \leq r \\
& && \Psi \succeq 0 \\
& && \Sigma - \Psi \succeq 0
\end{aligned}
$$

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Numerical Example
Local Optimization method

Consider $r = 3$ factors, $N = 20$ measured outputs and $T = 100$ Samples, $\Psi_{ij} = (0.7)^{|i-j|}$. and the performance index

$$d(P_m) = 1 - \frac{\mathrm{trace}(AP_mA^\top)}{\mathrm{trace}(AA^\top)}$$

Method	$d(\cdot)$	Total Time [s]
PCA	0.1992	0.0136
RCO	0.1002	274.6732
EM	0.1330	0.0198

Table: Mean value over $N_{mc} = 100$ Monte Carlo simulations of $d(\cdot)$.

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Conclusions

- We have developed a new representation of rank constraints.
  - Second-order differentiable.
  - Avoid several nonlinearities, excepting a bilinear constraint.
- We have developed two optimization algorithms.
  - Local Optimisation.
  - Global optimisation.
- We have applied the method:
  - Factor Analysis with Correlated Errors.
  - Sparse Control (tomorrow FrA04.2).

Thanks for your attention!

Any questions?

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Global Optimization Example

Consider $r = 1$ factor, $N = 3$ measured outputs and $T = 100$ Samples.

	$d(\cdot)$	$\|\widehat{\Psi}\|_1$	time [s]
RCO	0.066	4.46	23.83
RCO-G	0.066	$\underline{\mathbf{4.46}}$	310.19

	$d(\cdot)$	$\|vec(\widehat{\Psi})\|_1$	time [s]
RCO	0.165	11.12	43.20
RCO-G	0.218	$\underline{\mathbf{9.17}}$	1768.31

Solving the Optimisation Problem

Proposed method

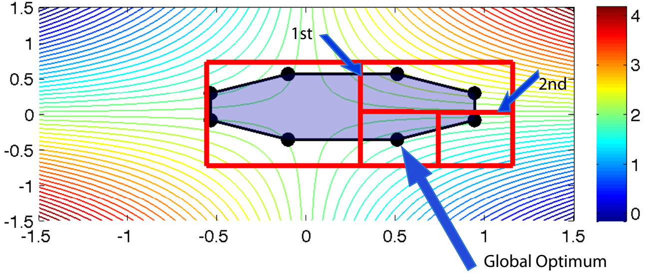
Solve $\mathcal{P}_{bi}$ iteratively.
In each iteration, solves a feasibility problem that deals with the bilinear constraint.

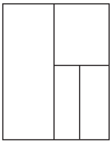
Given the estimate $\theta^m \in \Omega$, at iteration m

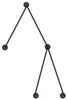
$$\begin{aligned} \min_{\theta \in \mathbb{R}^p} \min_{W \in \mathbb{S}^n} \quad & \|X(\theta)W\| \\ \text{subject to} \quad & f(\theta) \leq f(\theta^m) - \eta_m \\ & \theta \in \Omega \\ & W \in \Phi_{n,r} \end{aligned}$$

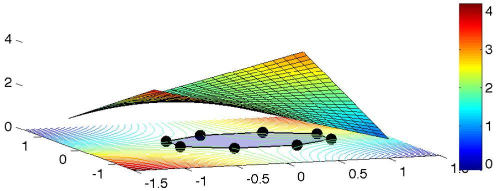
THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Branch and Bound









ℓ_1-norm

- Computationally efficient.
- Unintuitive to chose the sparse parameters
- Poor handling of group-constraints

ℓ_0-norm

- Computationally more expensive.
- ℓ_0-norm constraints have a clear interpretation.
- Can handle Group constraints.

ℓ_0-norm equivalence

$$\mathcal{P}_{\ell_0}: \quad \begin{aligned} \min_{\theta\in\mathbb{R}^p} \ & f(\theta) \\ \text{s.t.} \ & \theta \in \Omega \\ & \|\theta\|_0 \le r \end{aligned} \quad \equiv \quad \mathcal{P}_{eq}: \quad \begin{aligned} \min_{\theta\in\mathbb{R}^p} \min_{w\in\mathbb{R}^p} \ & f(\theta) \\ \text{s.t.} \ & \theta \in \Omega \\ & \theta \circ w = \mathbf{0} \\ & \mathbf{0} \le w \le \mathbf{1} \\ & \mathbf{1}^\top w = p - r \end{aligned}$$

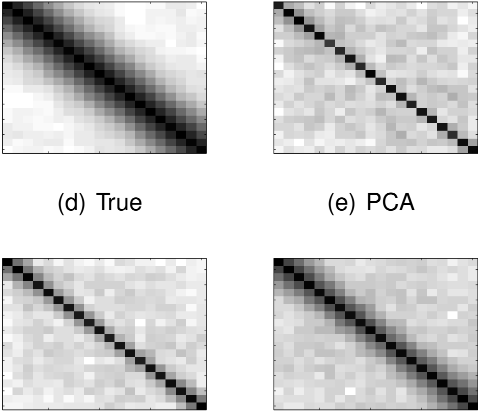
Delgado, Agüero & Goodwin 2014 (IFAC 2014).

Delgado, Agüero & Goodwin submitted to Automatica.

OF LE

Local Optimization

Ψ



(d) True

(e) PCA

(f) EM

(g) RCO

Concave Minimisation

Concave minimization may lead to a combinatorial problem. (e.g. matrix-rank function is quasi-concave on the Positive Semidefinite Cone, thus leads to Concave-minimisation and “reverse-convex” constraints)