Matrix Rank Optimization Problems in System Identification via Nuclear Norm Mimization

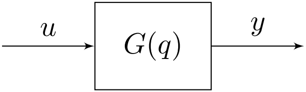
Bo Wahlberg and Cristian Rojas,
Department of Automatic Control
ACCESS Linnaeus Center,
School of Electrical Engineering,
KTH Royal Institute of Technology,
Stockholm, Sweden

Tutorial on sparse and low-rank representation methods in control, estimation and system identification, ECC 2013

ROYAL INSTITUTE
OF TECHNOLOGY

Dynamical System

Stable Input-Output System



Transfer Function

$$G(q) = \sum_{k=1}^{\infty} g_k q^{-k}$$

State-Space Model

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) \end{aligned}$$

$$\dim\{x(t)\} = n \quad g_k = CA^{k-1}B$$

Observability, Controllability and the Hankel Matrix

Extended Observability Matrix $\mathcal{O}_p = [C; CA; \dots CA^{p-1}]$
has rank less or equal to n.

Extended Controllability Matrix $\mathcal{C}_q = [B\ AB\ \ \dots A^{q-1}B]$
has rank less or equal to n.

Hankel Matrix

$$\mathcal{H}_{p,q} = \begin{pmatrix} g_1 & g_2 & \cdots & g_p \\ g_2 & g_3 & \cdots & g_{p+1} \\ & & \ddots & \\ g_q & g_{q+1} & \cdots & g_{q+p} \end{pmatrix}$$

$\boxed{\mathcal{H} = \mathcal{O}\mathcal{C} \text{ has rank less or equal to } n!}$

Flexible Models in System Identification

FIR (high order):

$$y(t) = \sum_{k=1}^{n_g} g_k u(t-k),$$

ARX (high order):

$$y(t) = -\sum_{k=1}^{n_a} a_k y(t-k) + \sum_{k=1}^{n_b} b_k u(t-k)$$

4SID Linear Regression (Jansson & Wahlberg 1996):

$$y_\alpha(t) = L_1 y_\beta(t-\beta) + L_2 u_\beta(t-\beta) + L_3 u_\alpha(t)$$

where $y_\alpha(t) = [y(t), y(t+1), \ldots y(t+\alpha-1)]^T$, ...

Constraints - A Priori Information

- Stability
- Rank $\{\mathcal{H}\} = n$,
  Range Space $\{\mathcal{H}\} =$ Range Space $\{\mathcal{O}_q\}$
- Rank $\{L_1 \; L_2]\} = n$,
  Range Space $\{[L_1 \; L_2]\} =$ Range Space $\{\mathcal{O}_\alpha\}$

$\Rightarrow$ Subspace Methods!

The Matrix Rank Tool in SI 1970 – 2000: SVD

The truncated Singular Value Decomposition (SVD) gives the optimal **un-structured** rank r approximation of a given matrix.

S.Y. Kung. A new identification and model reduction algorithm via singular value decomposition. In 12th Asilomar Conference on Circuits, Systems and Computers, 1978.

- SVD forms the base for Subspace System Identification. Very efficient numerical algorithms!
- A key to success was **weighted** subspace fitting!

(Optimal Hankel Norm Model Reduction goes far beyond SVD.)

The Matrix Rank Tool 2001 –: The Nuclear Norm

Difficult Problem

$$\begin{array}{rl} \text{minimize} & \mathrm{Rank}\{X\} \\ \text{subject to} & X \in \text{Convex Set} \end{array}$$

Heuristic Convex Optimization Problem:

$$\begin{array}{rl} \text{minimize} & \|X\|_* \\ \text{subject to} & X \in \text{Convex Set} \end{array}$$

where $\|X\|_*$ is the nuclear norm.

Motivation of the Nuclear Norm

The singular values of X, $\{\sigma_i \geq 0\}$, are the square root of eigenvalues of $X^T X$.

Induced norm: $\|X\| = \sigma_{max}$

Frobenius norm: $\|X\|_F = \sqrt{\mathrm{tr}\{X^T X\}} = \sqrt{\sum_{i=1}^r \sigma_i^2}$
(l_2-norm of the singular values)

Nuclear norm: $\|X\|_* = \sum_{i=1}^r \sigma_i$
(l_1-norm of the singular values)

Recent overviews:

Recht, Fazel and Parrilo: Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization, 2010

Fazel, Pong, Sun, and Tseng: Hankel matrix rank minimization with applications in system identification and realization, 2013

SDP Formulation

Nuclear norm equivalent problem:

$$\begin{array}{rl} \text{minimize} & (\mathrm{tr}\{Y\} + \mathrm{tr}\{Z\})/2 \\ \textsf{subject to} & \begin{pmatrix} Y & X \\ X^T & Z \end{pmatrix} \succeq 0 \\ & X \in \textsf{Convex Set} \end{array}$$

(Expensive to solve via general-purpose solvers)

M. Fazel, H. Hindi, and S. Boyd: A Rank Minimization Heuristic with Application to Minimum Order System Approximation, ACC, 2001. Over 300 citations!

SI using Hankel Matrix Rank Minimization

Example: Output Error Model with least squares fit using data $\{y(t), u(t), t = 1 \dots N\}$ and model order constraint:

$$
\begin{array}{rl}
\text{minimize} & \displaystyle\sum_{t=1}^{N} [y(t) - \sum_{k=1}^{\infty} g_k u(t-k)]^2 \\
\text{subject to} & \text{Rank}\,\{\mathcal{H}(g)\} = n
\end{array}
$$

Equivalent to $G(q) = \displaystyle\sum_{k=1}^{\infty} g_k q^{-k} = \frac{B(q)}{F(q)}$ of order n.

The rank constraint is as non-convex as the optimization problem minimizing with respect to the coefficients of $B(q)$ and $F(q)$.

Use the nuclear norm relaxation trick!

A high order FIR approximation and the SDP formulation $\Rightarrow$

Heuristic Convex Optimization Problem:

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{N}[y(t) - \sum_{k=1}^{n} g_k u(t-k)]^2 + \lambda(\mathrm{tr}\{Y\} + \mathrm{tr}\{Z\})/2 \\ \text{subject to} & \begin{pmatrix} Y & \mathcal{H}(g) \\ \mathcal{H}(g)^T & Z \end{pmatrix} \succeq 0 \end{array}$$

Ref: Grossmann, Jones and Morari, ECC 2009
Hjalmarsson, Welsh, Rojas, SYSID 2012

Compare with standard regularization techniques in SI where a cost term is added to penalize the model complexity.

ROYAL INSTITUTE OF TECHNOLOGY

Does it work better than PEM/Subspace SI?

- No local minima (seems less ad hoc than the three step Kung algorithm).
- Easy to add constraints on the outputs, or use more complicated penalty terms (for example with missing data, or non-quadratic penalties).
- A lot of simulation benchmarks and real data test cases have been done with good results. In particular, with high order ARX and at low signal to noise ratios.
- Weighting improves the result (work in progress by Hjalmarsson). Compare Indirect PEM.
- Generalized to Box-Jenkins models.

Does it work better than PEM/Subspace SI?

- How effective is the nuclear norm heuristic in solving data approximation problems?

  Many promising examples by Sznaier at SYSID 2012 using real data.

  Some negative answers by Markovsky at SYSID 2012 using simulation examples.

- **No performance analysis**. *E.g.* the effect of λ?

Would compare to the-state-of-the-art in subspace system identification 1990!

A lot of promising algorithms and initial results.
Alternative heuristics, iterative re-weighting, ...

Connection to Balanced Model Reduction

Use $\mathcal{H} = \mathcal{O}\mathcal{C}$ and let $Y = \mathcal{O}\mathcal{O}^T$, and $Z = \mathcal{C}^T\mathcal{C}$.

Feasible Solution: $\begin{pmatrix} Y & \mathcal{H}(g) \\ \mathcal{H}(g)^T & Z \end{pmatrix} = \begin{pmatrix} \mathcal{O} \\ \mathcal{C}^T \end{pmatrix} \begin{pmatrix} \mathcal{O}^T & \mathcal{C} \end{pmatrix} \succeq 0$

Optimality condition:

$$\begin{aligned} \min(\mathrm{tr}\{Y\} + \mathrm{tr}\{Z\})/2 &= \min(\mathrm{tr}\{\mathcal{O}\mathcal{O}^T\} + \mathrm{tr}\{\mathcal{C}^T\mathcal{C}\})/2 \\ &= \min(\mathrm{tr}\{\mathcal{O}^T\mathcal{O}\} + \mathrm{tr}\{\mathcal{C}\mathcal{C}^T\})/2 = \sum_{k=1}^{n} \sigma_i \end{aligned}$$

(the "truncated" observability and controllability grammians, and the sum of the Hankel singular values.)
A balanced realization will satisfy this condition!

This approach to *structured* balanced model reduction was explored in the mid 90s by Carolyn Beck and co-workers.

Subspace System Identification using the Nuclear Norm

ROYAL INSTITUTE OF TECHNOLOGY

Block Hankel matrix formulation of the noise free state space equations:

$$Y = GX + HU \quad \Rightarrow \quad YU^{\perp} = GXU^{\perp} \quad \text{has rank } n$$

Data pre-processing step:

$$\underset{y_m}{\text{minimize}} \quad \sum_{t=1}^{N} [y(t) - y_m(t)]^2 + \lambda ||Y_m U^{\perp}||_*$$

L. Liu and L. Vandenberghe, Interior-point method for nuclear norm approximation with application to system identification, *SIAM Journal on Matrix Analysis and Applications*, 2010.

Subspace System Identification using the Nuclear Norm

- Easier to select the model order than standard subspace methods.
- Slightly better than standard subspace SI methods on the DaISy benchmarks.
- Weighting improve results and efficient ADMM implementation
  in Hansson, Liu and Vandenberghe, CDC 2012.
- Frequency Domain Subspace SI using the Nuclear Norm:
  R. Smith, CDC 2012.

SI of Hammerstein Models

Let $x, y \in \mathbb{R}^n$ and study

$$x^T y = \mathrm{tr}\{yx^T\} = \mathrm{tr}\{C\}, \quad \text{Rank}\{C\} = 1$$

Can be used to re-formulate certain bilinear optimization problems as rank one optimization problems.

Example: FIR Hammerstein model

$$y(t) = \mathrm{tr}\{C\Psi(t)\} \quad \text{Rank}\{C\} = 1$$

where x corresponds to the FIR parameters and y to the function expansion parameters.

The nuclear norm heuristic for this problem is evaluated in Falck, Suykens, Schoukens and De Moore, CDC 2010.

Blind Identification

Example: Blind identification of FIR models with piecewise constant inputs

$$y(t) = \mathrm{tr}\{b_n^T u_\beta(t-\beta)\} = \mathrm{tr}\{C\}, \quad \text{Rank}\{C\} = 1$$

Estimate both the FIR parameters and the input signal using the nuclear norm heuristic and the total variation l_1 heuristic.
Ref: Ohlsson, Ratliff, Dong and Sastry, ArXiv, 2013.

Can be dangerous to use multi-objective relaxations. One will typically dominate. See Oymak, Jalali, Fazel, Eldar, and Hassibi, ArXiv, 2013.

Take Home Message

- We have lot of rank constraints in systems & control theory, which can be utilized to regularize estimation problems.
- Estimation using the nuclear norm heuristic is an extremely active research area that is based on convex relaxation heuristics.
- Limited analytic results, which typically not are applicable to dynamical system, e.g. the restricted isometry property and random matrix theory.
- A need for analysis beyond simulations to further improve the methods!

A Lot of Exciting Ideas and Promising Methods!

ROYAL INSTITUTE OF TECHNOLOGY

Selected References

P. Bendotti, and C. Beck, C., LFT model reduction and controller synthesis for a pressurized water reactor In *Decision and Control (CDC), 1996, 35th IEEE Conference on*, pages 4480-4485, 1996

T. Falck, J. A K Suykens, Johan Schoukens, and B. De Moor. Nuclear norm regularization for overparametrized hammerstein systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 7202–7207, 2010.

M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings American Control Conference*, 6, pages 4734–4739, Arlingon, Virginia, June 2001.

M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. In *In American Control Conference*, pages 3273–3278. AACC, 2004.

Selected References

M. Fazel, T. K. Pong, D. Sun, and P. Tseng. Hankel matrix rank minimization with applications in system identification and realization. *To appear in SIAM J. on Matrix Analysis and Applications.*, 2013.

C. Grossmann, C.N. Jones, and M. Morari. System identification with missing data via nuclear norm regularization. In *European Control Conference*, Budapest, Hungary, August 2009.

A. Hansson, Zhang Liu, and L. Vandenberghe. Subspace system identification via weighted nuclear norm optimization. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 3439–3444, 2012.

H. Hjalmarsson, J. S. Welsh, and C. R. Rojas. Identification of box-jenkins models using structured arx models and nuclear norm relaxation. In *16th IFAC Symposium on System Identification*, pages 322–327. IFAC, 2012.

Selected References

Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2010.

I. Markovsky. How effective is the nuclear norm heuristic in solving data approximation problems? In *Proc. of the 16th IFAC Symposium on System Identification*, Brussels, 2012.

H. Ohlsson, L. Ratliff, R. Dong, and Sastry S. S. Blind identification of arx models with piecewise constant inputs. Submitted to the 52nd IEEE Conference on Decision and Control, 2013. arXiv:1303.6719

Selected References

S. Oymak, A. Jalali, M. Fazel, Y.C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. Submitted, arXiv:1212.3753

B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, August 2010.

R.S. Smith. Nuclear norm minimization methods for frequency domain subspace identification. In *American Control Conference (ACC), 2012*, pages 2689–2694, 2012. M. Sznaier. Compressive Information Extraction: A Dynamical Systems Approach. In *16th IFAC Symposium on System Identification*, pages 1559-1568. IFAC, 2012.