Application of Rank-Constrained Optimisation to Nonlinear System Identification ⋆

Ramón A. Delgado * **Juan C. Agüero** **
Graham C. Goodwin * **Eduardo M. A. M. Mendes** ***

* *School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*
Graham.Goodwin@newcastle.edu.au, Ramon.Delgado@uon.edu.au
** *Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile*
juan.aguero@usm.cl
*** *Departamento de Engenharia Electrônica, Universidade Federal de Minas Gerais, Brazil*
emmendes@cpdee.ufmg.br

Abstract: Nonlinear System identification has a rich history spanning at least 5 decades. A very flexible approach to this problem depends upon the use of Volterra series expansions. Related work includes Hammerstein models, where a static nonlinearity is followed by a linear dynamical system, and Wiener models, where a static nonlinearity is inserted after a linear dynamical model. A problem with these methods is that they inherently depend upon series type expansions and hence it is difficult to know which terms should be included. In this paper we present a possible solution to this problem using recent results on rank-constrained optimization. Simulation results are included to illustrate the efficacy of the proposed strategy.

1. INTRODUCTION

System identification tools are frequently used to fit models to experimentally obtained data from a system. There is a vast literature in the topic (Ljung, 1999). Predominantly the literature deals with the case of linear systems. However, some systems are known to exhibit inherently nonlinear behaviour.

Since *nonlinear* includes anything other than *linear* there are endless options for what structure model should be chosen. A broad classification of model types includes Black Box, White Box and Grey Box. The key distinction is that black box methods use a flexible structure under the "one-size-fits-all" philosophy. On the other hand, white box methods build the model using physical/Biological understanding. Grey Box lies between these extremes and utilises elements of both Black and White Box methods.

An advantage of Black Box methods is that they are flexible and can be used in situations where it is difficult to obtain a good model structure from physical reasoning. The current paper focuses on this class of models.

An inherent feature of Black Box methods is that they depend upon series type expansions. Hence, there is a problem of how to choose which terms to include and which to exclude in the series. Indeed, if used naively, then there is the potential for including too many terms leading to overfitting. This is a well known phenomenon which is often termed the Bias-Variance problem. Specifically, including too few terms will lead to deterministic bias errors whereas including too many terms will lead to variance errors due to the impact of noise on the parameter estimates.

The topic of bias-variance trade-off has been addressed in many recent papers, see for example (Chen et al., 2012; Mendes and Billings, 2001). Standard tools for restricting the complexity of a model include the ℓ_1 heuristic (Candes et al., 2006), nuclear norm (Fazel et al., 2001). and rank constrained optimisation, see e.g. (Delgado et al., 2014a; Markovsky, 2013).

Here we adopt the rank-constrained optimisation approach. We apply this idea to the problem of choosing which terms of a series expansion to include in a nonlinear model. We believe that this is the first time that this class of methods has been applied in the context of nonlinear system identification.

Methods for imposing rank constraints are closely related to the problem of rank-minimisation. The latter problem has received considerable attention over the past few decades. The focus has centred on various approximations such as trace, nuclear norm and log-det heuristics (see e.g. (Fazel et al., 2001, 2003)). These developments have been applied to several system identification problems, e.g. to identify a moving bed process from incomplete data sets (Grossmann et al., 2009), identification of Box-Jenkins models (Hjalmarsson et al., 2012), and for system identification with missing inputs and outputs (Liu et al., 2013). Other approaches use the related idea of structured

⋆ J. C. Agüero was partially supported by the Chilean Research Council (CONICYT) through Basal Project FB0008, and FONDECYT (grant no. 1150954).

Peer-reviewed author's copy of:
R.A. Delgado, J.C. Agero, G.C. Goodwin and Eduardo M.A.M Mendes. **Application of Rank-Constrained Optimization to Nonlinear System Identification**. In *1st IFAC Conference on Modelling, Identification and Control of Nonlinear Systems (MICNON 2015)*, Saint-Petersburg, Russia, 2015.
Available at http://dx.doi.org/10.1016/j.ifacol.2015.09.290

low-rank approximation. This idea has been applied to several system identification problems. For example, it has been used for identification of periodically time-varying systems (Markovsky et al., 2014) and system identification in the behavioural setting (Markovsky, 2013).

There also exist approaches that solve rank-minimization problems exactly, see e.g (d'Aspremont, 2003). However, computational complexity of the associated methods is formidable even for small-size problems. Most heuristics developed for rank-minimization can also be applied to rank-constrained problems. However, when using these heuristics, the condition on the rank is not set as a hard constraint.

Recently in (Delgado et al., 2014a,b), a novel approach to dealing with sparsity and rank constraints has been proposed. The current paper builds on this earlier work and applies the approach to nonlinear system identification.

The layout of the remainder of the paper is as follows: In Section 2, we formulate the problem of interest. In Section 3 we discuss rank and cardinality constraints. Section 4 shows how the approach can be adapted to solve a nonlinear system identification problem. Simulated examples are given in Section 5. Finally, conclusions are drawn in Section 6.

Notation and basic definitions: $\text{rank}(A)$ denotes the rank of a matrix A. $\lambda_i(A)$ denotes the i-th largest eigenvalue of a symmetric matrix A, $A \circ B$ denotes the Hadamard product of A and B, $A \succeq 0$ denotes that A is positive semidefinite, and $A \succeq B$ denotes that $A - B \succeq 0$. We represent the transpose of a given matrix A as $A^\top$. $\mathbb{S}^n$ denotes the set of symmetric matrices of size $n \times n$, and $\mathbb{S}^n_+$ the set of symmetric positive semidefinite matrices, i.e. $\mathbb{S}^n_+ := \{A \in \mathbb{S}^n | A \succeq 0\}$. $\mathbf{1}$ denotes a vector with ones as entries. $\|\|_F$ denotes the Frobenius norm.

2. NONLINEAR BLACK-BOX MODELS

There are many options for nonlinear models including Volterra expansions and neural networks. For example, the Volterra model has the following generic structure

$$\begin{aligned} y(k) = &\sum_{n_1=1}^{N} h_1(n_1)u(k-n_1) \\ &+ \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} h_2(n_1,n_2)u(k-n_1)u(k-n_2) + \dots \\ &\cdots + n(k) \end{aligned} \tag{1}$$

where $\{h(n_1, \dots)\}$ are generalised impulse response terms, and $y(k)$, $u(k)$ and $n(k)$ denote the output, input and additive noise, respectively.

It is readily seen from (1) that the model potentially has a huge number of parameters, namely $N + N^2 + \cdots + N^P$ where P is the order of the expansion. There is thus clearly a potential for over-fitting especially when little data is available.

A special case of (1) occurs when the choice

$$h(n_1, n_2, \dots, n_P) = \begin{cases} h_j(n_1) & \text{if } n_1 = n_2 = \cdots = n_P \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

This simplification leads to a generalised Hammerstein model structure

$$y(k) = \sum_{j=1}^{P}\sum_{n_1=0}^{N} h_j(n_1)u(k-n_1)^j + n(k) \tag{3}$$

In the sequel, and for sake of simplicity of exposition, we will focus on the above structure. However, the extension to more general instances of the model 1 are straightforward.

Our goal will be to use rank-constrained optimisation ideas to restrict the class of impulse response functions and the set of terms of $\{u(k)^j\}$ used in the model.

3. OVERVIEW OF RANK-CONSTRAINED OPTIMISATION

We have previously applied rank constrained optimisation to a class of problems in linear system identification (see (Delgado et al., 2015)). Our goal here is to extend these tools to the nonlinear case. We first summarise the key elements of rank constrained optimisation that we will use.

3.1 Rank-Constrained optimisation

Consider the following rank-constrained optimisation problem

$$\begin{aligned} \mathcal{P}_{rco}: \quad & \min_{x \in \mathbb{R}^p} \ f(x) \\ & \text{subject to } \ x \in \Omega \\ & \qquad\qquad \text{rank}(G(x)) \leq r \end{aligned}$$

Also, consider the following optimisation problem involving bilinear constraints

$$\begin{aligned} \mathcal{P}_{rcoequiv}: \quad & \min_{x \in \mathbb{R}^p, W \in \mathbb{S}^n} \ f(x) \\ & \text{subject to } \ x \in \Omega \\ & \qquad\qquad G(x)W = 0_{m\times n} \\ & \qquad\qquad 0 \preceq W \preceq I_n \\ & \qquad\qquad \text{trace}(W) = n - r \end{aligned}$$

where $\Omega \subset \mathbb{R}^p$ is a constraining set, $f : \mathbb{R}^p \to \mathbb{R}$ is the objective function and $G : \mathbb{R}^p \to \mathbb{R}^{m\times n}$.

The following result shows that $\mathcal{P}_{rco}$ and $\mathcal{P}_{rcoequiv}$ are equivalent.

Corollary 1. (Delgado et al., 2015) $x^\star \in \mathbb{R}^p$ is a global solution of $\mathcal{P}_{rco}$ if and only if there exist a $W^\star$ such that the pair $(x^\star, W^\star)$ is a global solution of $\mathcal{P}_{rcoequiv}$.

Proof. See (Delgado et al., 2015). Background of the proof can be found in (Delgado et al., 2014b).

A key observation is that the problem $\mathcal{P}_{rco}$ is combinatorial in nature whereas problem $\mathcal{P}_{rcoequiv}$ can be solved using standard tools of nonlinear programming. Details are given in (Delgado et al., 2014a).

3.2 Cardinality-constrained optimisation

Problem $\mathcal{P}_{rco}$ can also cover optimisation problems subject to cardinality constraints, i.e. constraints in the number of non-zero elements of a vector. This is achieved, by considering $G(x)$ to have a diagonal structure, i.e. $G(x) = \text{diag}(x)$, then $\text{rank}(G(x)) = \text{card}\, x$. Hence if we consider

the following cardinality-constrained optimisation problem

$$\mathcal{P}_{card}: \quad \min_{x\in\Omega\subseteq\mathbb{R}^n} \ f(x) \quad \text{subject to} \ \ \operatorname{card} x \leq r$$

then using Corollary 1, this problem can be formulated as an optimisation problem subject to bilinear constraints as follows

$$\begin{aligned} \mathcal{P}_{cardequiv}: \quad \min_{x\in\Omega\subseteq\mathbb{R}^n, w\in\mathbb{R}^n} \ & f(x) \\ \text{subject to} \ \ & x_i w_i = 0; \quad i = 1,\dots,n \\ & 0 \leq w_i \leq 1; \quad i = 1,\dots,n \\ & \mathbf{1}^\top w = n - r \end{aligned}$$

The equivalence beween $\mathcal{P}_{card}$ and $\mathcal{P}_{cardequiv}$ has been recently obtained independently in (Burdakov et al., 2015). In addition in (Mitchell et al., 2013; Piga and Tóth, 2013) optimisation problems where the cost contains a term that induces a cardinality constraint has been analysed. The approach considered in these papers can be considered as a special case of the equivalence between $\mathcal{P}_{card}$ and $\mathcal{P}_{cardequiv}$, see (Delgado et al., 2014b).

3.3 Group-constrained optimisation

A problem closely related to that of cardinality-constrained optimisation is that of group-constrained optimisation. The idea of group-handling in sparse representations has received attention in the last decade, see e.g. (Kim et al., 2006; Yuan and Lin, 2006). These methods are based on the ℓ_1-norm heuristic. Here we show how group constraints can be incorporated into cardinality-constrained optimisation problems.

In problem $\mathcal{P}_{cardequiv}$, the variable w, at the optimum is a binary variable taking the value $w_i = 1$ for those elements corresponding to $x_i = 0$. Additional constraints over w can be included in the optimisation problem to manage how the zero and non-zero elements of x interact.

To illustrate these ideas, consider the problem of minimising $f(x)$ subject to $\operatorname{card} x \leq r$, with $x \in \mathbb{R}^m$, with m and r being even numbers. Moreover, say we split the vector x into two groups: Group $G1$ consisting of the first $m/2$ elements of x, and group $G2$ consisting of the last $m/2$ elements of x. An interesting group constraint in this framework is to require that when the i-th element of $G1$ is non-zero, i.e. $x_i \neq 0$, then the i-th element of $G2$ must be zero, i.e $x_{m/2+i} = 0$, and vice versa. Thus, a non-zero element in one group induces a zero in the other group. The associated optimisation problem can be formulated as the following group-constrained optimisation problem:

$$\begin{aligned} \min_{x\in\mathbb{R}^m, w\in\mathbb{R}^m} \ & f(x) \\ \text{subject to} \ \ & x_i w_i = 0 && i = 1,\cdots,m \\ & 0 \leq w_i \leq 1 && i = 1,\cdots,m \\ & \mathbf{1}^\top w = m - r \\ & w_i + w_j \geq 1; && i = 1,\dots,m/2. \\ & && j = m/2 + i. \end{aligned}$$

where the last constraint in the above problem, rules the interaction between the zero and non-zero elements of the vector x.

4. APPLICATION TO HAMMERSTEIN NONLINEAR MODEL

There are multitude of problem formulations that we could adopt to illustrate the above circle of ideas. We adopt one particular strategy to illustrate but anticipate that the same general methodology would apply, mutatis mutandis, to other black-box nonlinear system identification problems.

We begin with generalised Hammerstein model (3). We can then use the ideas of section 3 to apply two constraints, namely

(i) Only a fixed number (say M) of parameters can be used ($M \leq (N+1)P$).

(ii) We could extend the above idea to incorporate the standard Hammerstein structure by adding a group constraint to the parameters. Namely, if any parameter in a group $\{h_j(n_1), n_1 = 1,\dots,N\}$ is non-zero, then all parameters in that group can be non-zero.

To satisfy condition **(i)**, we can use the ideas presented in section 3.2. Then we impose the cardinality constraint by requiring

$$h_j(n_1) w_{j,n_1} = 0 \tag{4}$$

$$0 \leq w_{j,n_1} \leq 1 \tag{5}$$

$$\sum_{j=1}^{P} \sum_{n_1=1}^{N} w_{j,n_1} = (N+1)P - M \tag{6}$$

Additionally, condition **(ii)** can be seen as a group constraint, as the discussed in section 3.3. Specifically, condition **(ii)** can be easily satisfied by including the following additional constraints

$$w_{j,1} = w_{j,2} = \cdots = w_{j,N-1} = w_{j,N} \qquad j = 1,\dots,P. \tag{7}$$

5. SIMULATION STUDY

We simulate the following system

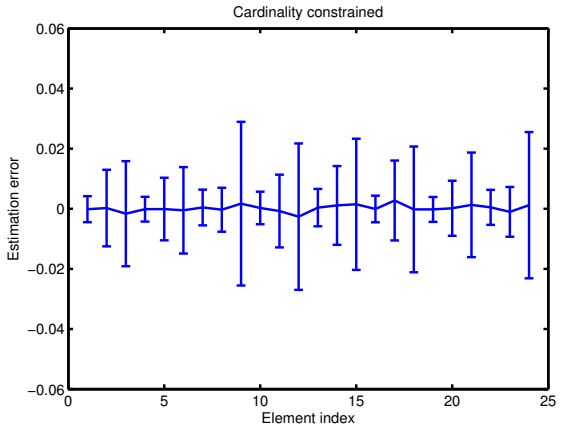
$$y(k) = \varphi(k)^\top \theta + n(k) \tag{8}$$

where

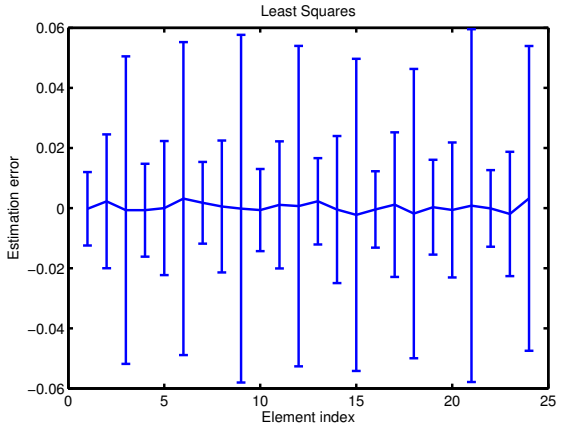
$$\varphi(k) = \begin{bmatrix} u(k) \\ u^2(k) \\ \vdots \\ u^p(k) \\ u(k-1) \\ \vdots \\ u^p(k-1) \\ \vdots \\ u^p(k-N) \end{bmatrix} \tag{9}$$

We note that this system potentially has $(N+1)P$ parameters, but we restrict the model so that only M parameters are non-zero. However, we assume that the specific location of the non-zero elements is unknown. We then choose to fit a generalised Hammerstein model of the form (3) where we set $P = 3$, $N = 7$, but we consider that only $M = 4$ parameters take a non-zero value. Notice that the model potentially contains 24 parameters.

Consider that we are given $N_s = 1000$ input-output samples. The identification problem can be formulated as the following least squares optimisation problem



(a) Proposed approach



(b) Least Squares

Fig. 1. Error bar plot over 100 MonteCarlo simulations.

$$\mathcal{P}_{nsi} \quad \min_{\theta\in\Theta\subseteq\mathbb{R}^{P(N)}} \|Y-\Phi\theta\|_2^2$$
$$\text{subject to} \quad \operatorname{card}\theta \leq M$$

where

$$\Phi = \begin{bmatrix} \varphi(N)^\top \\ \varphi(N+1)^\top \\ \vdots \\ \varphi(N_s)^\top \end{bmatrix} \tag{10}$$

$$Y = \begin{bmatrix} y(N) \\ y(N+1) \\ \vdots \\ y(N_s) \end{bmatrix} \tag{11}$$

We run $N_{mc} = 100$ Monte Carlo simulations, with different realization of the input $\{u(k)\}$, of the noise $\{n(k)\}$ and different θ, all having at most $M = 4$ non-zero elements. The input and the noise are taken to be zero-mean Gaussian with unitary variance. The non-zero elements of θ are draw for a normal distribution.

We compare the estimates obtained by solving problem $\mathcal{P}_{nsi}$ with the estimates obtained via Ordinary Least Squares.

Figures 1a-1b show the estimation mean and variance of the estimation error over the MonteCarlo simulations. Figures 1a shows the estimation error of the proposed approach, and Figure 1b shows the estimation error for the least squares estimates. Notice that the estimation error of the proposed approach has much smaller variance than the estimation error of the least squares estimate. This confirms the principal claim of the paper, namely, that the use of rank constrained optimisation gives a superior bias-variance trade-off in the context of nonlinear system identification. Also, note that in 86% of cases the methodology recovered the correct model structure. Interestingly, in the 13% of the cases where one of the non-zero elements was chosen erroneously, and in 1% of the cases two non-zero elements were chosen erroneously. Note, that the missed parameters were very small making its selection unimportant to the overall model behaviour.

6. CONCLUSIONS

We have applied recently developed ideas for rank constrained optimisation to nonlinear system identification. The advantages of the proposed approach have been illustrated via a numerical example.

REFERENCES

O. Burdakov, C. Kanzow, and A. Schwartz. On a reformulation of mathematical programs with cardinality constraints. In David Gao, Ning Ruan, and Wenxun Xing, editors, *Advances in Global Optimization*, volume 95 of *Springer Proceedings in Mathematics & Statistics*, pages 3–14. Springer International Publishing, 2015.

E. J. Candes, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Th.*, 52(2):489–509, February 2006.

T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and gaussian processes—revisited. *Automatica*, 48(8):1525 – 1535, 2012.

A. d'Aspremont. A semidefinite representation for some minimum cardinality problems. In *42nd IEEE Conference on Decision and Control, 2003*, volume 5, pages 4985–4990, 2003.

R. A. Delgado, J. C. Agüero, and G. C. Goodwin. A rank-constrained optimization approach: Application to factor analysis. In *19th IFAC World Congress*, Cape Town, South Africa, 2014a.

R. A Delgado, J. C. Agüero, and G. C. Goodwin. A novel representation of rank constraints for non-square real matrices. *arXiv preprint arXiv:1410.2317*, 2014b.

R.A. Delgado, J.C. Agüero, and G. C. Goodwin. Imposing structural constraints in system identification via rank-constrained optimisation. *submitted for publication*, 2015.

M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, pages 4734–4739. IEEE, 2001.

M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *American Control Conference*, volume 3, pages 2156–2162, 2003.

C. Grossmann, C. Jones, and M. Morari. System identification with missing data via nuclear norm regular-

ization. In *Presented at: European Control Conference*, volume 23, page 26, 2009.

H. Hjalmarsson, J Welsh, and C Rojas. Identification of Box-Jenkins models using structured ARX models and nuclear norm relaxation. In *Proceedings of the 16th IFAC Symposium on System Identification (SYSID 2012)*, 2012.

Y. Kim, J. Kim, and Y. Kim. Blockwise sparse regression. *Statistica Sinica*, 16(2):375, 2006.

Z. Liu, A. Hansson, and L. Vandenberghe. Nuclear norm system identification with missing inputs and outputs. *Systems & Control Letters*, 62(8):605–612, 2013.

L. Ljung. *System Identification: Theory for the user*. Prentice Hall, 2nd edition, 1999.

I. Markovsky. A software package for system identification in the behavioral setting. *Control Engineering Practice*, 21:1422–1436, 2013. doi: 10.1016/j.conengprac.2013.06.010.

I. Markovsky, J. Goos, K. Usevich, and R. Pintelon. Realization and identification of autonomous linear periodically time-varying systems. *Automatica*, 50 (6):1632 – 1640, 2014. ISSN 0005-1098. doi: http://dx.doi.org/10.1016/j.automatica.2014.04.003.

E. M.A.M. Mendes and S. A. Billings. An alternative solution to the model structure selection problem. *IEEE Transactions on Systems Man and Cybernetics Part A: Systems and Humans*, 31(6):597–608, 2001.

J. Mitchell, J.-S. Pang, A. Waechter, L. Bai, M. Feng, and X. Shen. Complementarity formulations for ℓ_0 norm optimization problems. In *Presentation at the 11-th Workshop on Advances in Continuous Optimization*, Florence, June 2013.

D. Piga and R. Tóth. An SDP approach for ℓ_0-minimization: Application to ARX model segmentation. *Automatica*, 49(12):3646–3653, 2013.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.