On Re-weighting, Regularization Selection, and Transient in Nuclear Norm based Identification ⋆

Mohamed Abdalmoaty * **Håkan Hjalmarsson** *

* *Automatic Control Department and ACCESS Linnaeus Center, KTH Royal institute of Technology, Stockholm, Sweden (e-mail: {mohamed.abdalmoaty, hakan.hjalmarsson}@ee.kth.se).*

Abstract: In this contribution, we consider the classical problem of estimating an Output Error model given a set of input-output measurements. First, we develop a regularization method based on the re-weighted nuclear norm heuristic. We show that the re-weighting improves the estimate in terms of better fit. Second, we suggest an implementation method that helps in eliminating the regularization parameters from the problem by introducing a constant based on a validation criterion. Finally, we develop a method for considering the effect of the transient when the initial conditions are unknown. A simple numerical example is used to demonstrate the proposed method in comparison to classical and another recent method based on the nuclear norm heuristic.

Keywords: Output error identification; low-rank estimation; nuclear norm; least squares; regularization.

1. INTRODUCTION

In this paper we revisit the classical problem of estimating the transfer function of an output error model, in open-loop, from a set of input-output measurements. The classical approach is to use the Prediction-Error Method (PEM); Ljung (1999), which coincides with the Maximum-Likelihood (ML) method under the assumption of Gaussian noise. The estimation problem under this framework is non-convex and one may be trapped into local minima or ill-conditioned problems. A major difficulty associated with this approach is the problem of model order selection. One possible technique that can be used to select a suitable model order is cross validation. Using this technique, we evaluate the objective function for the different model orders for another independent set of data, validation data, and select the model order which gives the best fit to this independent data set. Alternatively, one may treat the problem as a hypothesis test or from an information-theoretic point of view. Among the most common methods are the Akaike's Information Criterions (AIC), Schwarz's Bayesian Infromation Criterion (BIC), and Rissanen's Minimum Description Length (MDL) criterion.

On the other hand, under stability conditions, the simplest approach that can be used to estimate the transfer function is to truncate its infinite expansion to a finite number of impulse response coefficients. The resulting high-order FIR model can be estimated using the least-squares method. The only drawback of this approach is that the estimate may suffer from high variance. This issue may be solved by regularization techniques as shown, for example, in Pillonetto et al. (2014).

In order to regularize the estimate, we add a regularization term to the least-squares objective function. So far, regularized least-squares methods have been developed using the l_1-norm, the l_2-norm and the nuclear-norm. In this paper, we focus on nuclear-norm regularization methods.

The nuclear-norm is a unitarily invariant matrix function defined as the sum of the singular values. It is also known as the Schatten 1-norm, Ky-Fan r-norm, and the trace norm. It has been used to produce (exact or approximate) convex formulations of problems including a rank constraint (Recht et al., 2010). The heuristic was first proposed in Fazel et al. (2001) where it was shown that the nuclear-norm of a matrix is the best convex approximation of its rank and it gave the representation of nuclear-norm optimization problems as a Semidefinite-Program when the feasibility set is given by Linear-Matrix-Inequalities.

In the system identification community, there has been some interest in the nuclear norm heuristic as a surrogate for the rank. Several contributions based on this idea have been published already. In Liu and Vandenberghe (2010), Liu and Vandenberghe (2009), Mohan and Fazel (2010), Gebraad et al. (2011), Hansson et al. (2012), Liu et al. (2013), and Verhaegen and Hansson (2014), the heuristic has been used to formulate and solve subspace identification methods in terms of nuclear norms. For example, Liu et al. (2013) suggests a subspace identification method denoted by N4SID-NN, in which the low rank approximation step is done using the nuclear norm heuristic. In Grossmann et al. (2009), nuclear norm regularization has been used to estimate high order FIR models. In Hjalmarsson et al. (2012), nuclear-norm regularization is extended to

⋆ This work was supported by the European Research Council under the advanced grant LEARN, contract 267381 and by the Swedish Research Council under contract 621-2009-4017.

estimate high order ARX models with an extensive simulation study. In both cases the regularization parameter was determined by cross validation techniques. So far, there is no rigorous analytical analysis of the performance of the nuclear norm heuristic as a tool for system identification. However, as noted in Markovsky (2012), the results obtained by nuclear norm heuristics could be suboptimal - especially when it is used to approximate a fixed rank.

In this paper, we revisit the idea of nuclear norm regularization, as in Hjalmarsson et al. (2012), with three contributions. We first develop a regularization method based on the re-weighted trace heuristic (Fazel et al., 2003), as an alternative heuristic for the rank. We show by a numerical example that the re-weighted heuristic is better than the unweighted approach. Then, we use the idea of SPARSEVA method (Rojas and Hjalmarsson, 2011), to eliminate the regularization parameter from the problem by introducing a constant based on a validation criterion. By doing this we reduce the computational burden significantly. Finally, we extend the method to estimate the transient in cases with unknown initial conditions.

2. THE PROBLEM

Let a linear-time-invariant (LTI), causal, discrete-time system with single input $u(t)$ and single output $y(t)$ be described by

$$y(t) = G_0(q)u(t) + e(t), \quad t = 1, 2, 3, \dots \tag{1}$$

where

$$G_0 = \sum_{k=1}^{\infty} g_0(k) q^{-k} \tag{2}$$

is the transfer operator, q^{-1} is the backward shift operator. We will assume that the system is stable in the sense that

$$\sum_{k=1}^{\infty} |g_0(k)| < \infty. \tag{3}$$

The input $u(t)$ is deterministic, and the additive term $e(t)$ represents additive white (measurement) noise, independent from $u(t)$, with zero mean and variance σ^2. In this sense, the LTI system is represented by the infinite sequence $\{g(k)\}_{k=1}^{\infty}$, which we shall call the impulse-response of the system.

The problem can be stated as follows. Given a finite set of observations $Z^N := \{(y(t), u(t)) \mid t = 1, \dots, N\}$, construct an estimator $\hat{G}(Z^N)$ of the transfer operator $G_0(\cdot)$ in terms of an estimate of the impulse-response

$$\hat{G} : Z^N \to \{\hat{g}(k)\}_{k=1}^{\infty}. \tag{4}$$

2.1 Finite Impulse-Response (FIR) approximation

Given any arbitrary real number ϵ, there exists an integer n such that the impulse-response can be split into two parts, such that

$$\sum_{k=0}^{\infty} |g_0(k)| = S_n + T_n = \sum_{k \le n} |g_0(k)| + \sum_{k > n} |g_0(k)|, \tag{5}$$

with a finite sum S_n and a tail $T_n < \epsilon$. This means that if n is large enough, the elements of the impulse response $g_0(k)$ for all $k > n$ will be approximately zero and

$$G_0(q) \approx \sum_{k=1}^{n} g_0(k) q^{-k}. \tag{6}$$

Therefore, one way to estimate the transfer operator is by truncating the impulse-response at $k = n$, and estimating the vector $\theta_0 := [g_0(1) \dots g_0(n)]^T$. This can be done by solving a linear regression problem of the form

$$Y = \Phi\theta_0 + E \tag{7}$$

where

$$Y = [y(n+1) \dots y(N)]^T$$
$$E = [e(n+1) \dots e(N)]^T,$$
$$\Phi = \begin{bmatrix} u(n) & u(n-1) & \dots & u(1) \\ u(n+1) & u(n-2) & \dots & u(2) \\ \vdots & \vdots & \ddots & \vdots \\ u(N-1) & u(N-2) & \dots & u(N-n) \end{bmatrix}.$$

Observe that since the initial conditions $u(-n+1), \dots, u(0)$ are not known, the first n outputs, $y(1), y(2), \dots, y(n)$ in the data set Z^N are not used.

The least-squares estimator of θ is given by the solution of the normal equation

$$\Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y. \tag{8}$$

However, the resulting estimate will lack accuracy due to the large size of the unknown vector θ. This can be solved by using regularization methods which prevent over-fitting by penalizing model flexibility. A general regularized linear-regression problem can be written in the form

$$\underset{\theta}{\text{minimize}} \;\; \|Y - \Phi\theta\|_2^2 + \mathcal{J}(\theta, \rho) \tag{9}$$

in which $\mathcal{J}$ is a scalar function of θ and the regularization vector $\rho \in R^m$. Two common variants for $\mathcal{J}$ are the weighted l_1 and l_2 norms in R^n corresponding to the LASSO and the kernel-based estimators. In this paper, we will use the nuclear norm of the Hankel matrix of θ.

2.2 Nuclear norm regularization for FIR systems

Consider the LTI system defined in (1), and define the square Hankel matrix

$$\mathcal{H}_n(g) := \begin{bmatrix} g(1) & g(2) & \dots & g(k+1) \\ g(2) & g(3) & \dots & g(k+2) \\ \vdots & \vdots & \ddots & \vdots \\ g(k+1) & g(k+2) & \dots & g(2k+1) \end{bmatrix}, \tag{10}$$

with $n = 2k + 1$. It is known from system theory that if the transfer operator can be written as a rational function in q; i.e,

$$G_0(q) = \frac{B(q)}{F(q)} \tag{11}$$

where B, and F are coprime polynomials in q, and $\deg(F) = r$ then

$$\text{rank}(H_{r+i}(g)) = r \;\; \text{for all } i = 0, 1, 2 \dots \tag{12}$$

In this situation, with $n > r$ the estimate

$$\begin{aligned} \hat{\theta} := \underset{\theta}{\text{argmin}} \quad & \|Y - \Phi\theta\|_2^2 \\ \text{such that} \quad & \text{rank}(\mathcal{H}_n(\theta)) = r \end{aligned} \tag{13}$$

will have better accuracy compared to the ordinary least-squares estimate defined in equation (8), because the

model order is known and taken into account. Using Lagrange relaxation, the problem is equivalent to

$$\hat{\theta} := \underset{\theta}{\text{argmin}} \quad \|Y - \Phi\theta\|_2^2 + \lambda \text{ rank}(\mathcal{H}_n(\theta)) \tag{14}$$

for some λ.

Unfortunately, there are two difficulties with this formulation. Firstly, the rank function is a non-convex function on the unconstrained domain, and secondly, the positive integer r is unknown and hence also λ. The problem can be relaxed to a convex problem by introducing the nuclear norm as a convex heuristic for rank approximation.

The nuclear norm of a matrix $X \in \mathbb{R}^{n\times m}$ is equal to the sum of its singular values, i.e,

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X) \tag{15}$$

where $\sigma_i(X)$ denotes the i^{th} largest singular value of X and it is equal to the square-root of the i^{th} largest eigenvalue of XX^T. The relaxed problem is

$$\underset{\theta}{\text{minimize}} \quad \|Y - \Phi\theta\|_2^2 + \lambda\|\mathcal{H}_n(\theta)\|_*. \tag{16}$$

This is a regularized linear-regression in which the function $\mathcal{J}$ is given by the nuclear norm of the Hankel matrix of θ. The solution of this problem is given by the solution of the semidefinite program

$$\begin{aligned} \underset{\theta,X}{\text{minimize}} \quad & \|Y - \Phi\theta\|_2^2 + \lambda \text{ trace}(X) \\ \text{such that} \quad & \begin{bmatrix} X & \mathcal{H}_n(\theta) \\ \mathcal{H}_n(\theta) & X \end{bmatrix} \succeq 0 \end{aligned} \tag{17}$$

in which X is a symmetric slack matrix. The regularization parameter λ is determined by using the cross validation technique. The data set is divided into two parts. The first part is used for estimating θ for a grid of values of λ, and the second part is used for evaluating the quality of the corresponding models. The parameter λ corresponding to the model with the minimum prediction-error on the validation data is selected. Then the whole data set is used to re-estimate a model with the selected λ. Let us denote this method by NN-CV, in which the NN stands for nuclear norm, and the CV stands for cross validation.

3. R2NEVA

3.1 Re-weighted nuclear norm regularization for FIR systems

A better heuristic for rank minimization is the logarithm of the determinant (Fazel et al., 2003). This is a smooth concave approximation for the rank function. Using the function $\log\det(\mathcal{H}_n(\theta) + \delta I_n)$, the regularization problem in (17) becomes

$$\begin{aligned} \underset{\theta,X}{\text{minimize}} \quad & \|Y - \Phi\theta\| + \lambda \log\det(X + \delta I_n) \\ \text{such that} \quad & \begin{bmatrix} X & \mathcal{H}_n(\theta) \\ \mathcal{H}_n(\theta) & X \end{bmatrix} \succeq 0 \end{aligned} \tag{18}$$

where $\delta > 0$ is a small regularization parameter. The minimization is done locally via a sequence of convex problems using local minimization techniques. In the k^{th} step of the algorithm we solve

$$\begin{aligned} \underset{\theta,X}{\text{minimize}} \quad & \|Y - \Phi\theta\|_2^2 + \lambda \text{ trace}(X^k + \delta I_n)^{-1}X \\ \text{such that} \quad & \begin{bmatrix} X & \mathcal{H}_n(\theta) \\ \mathcal{H}_n(\theta) & X \end{bmatrix} \succeq 0 \end{aligned} \tag{19}$$

to get θ^{k+1}. In the first step of the algorithm $X^0 = I_n$ which is equivalent to solving the nuclear norm regularization problem in (17). Therefore, the algorithm is equivalent to solving the re-weighted nuclear norm regularization

$$\underset{\theta}{\text{minimize}} \quad \|Y - \Phi\theta\|_2^2 + \lambda\|W^k\mathcal{H}_n(\theta)W^k\|_* \tag{20}$$

in which each iteration builds on the top of the nuclear norm regularization without re-weighting. As shown in Mohan and Fazel (2010), we have

$$\begin{aligned} X^{k+1} &= (W^k)^{(-1)}U\Sigma U^T(W^k)^{(-1)} \\ W^{k+1} &= (X^{k+1} + \delta I_n)^{-1} \end{aligned} \tag{21}$$

in which $U\Sigma V^T$ is the singular value decomposition of the symmetric matrix $W^k\mathcal{H}_n(\theta^{k+1})W^k$.

One way to select the value of the regularization parameter λ is by using cross validation. Let us denote this method by RNN-CV, in which the RNN stands for re-weighted nuclear norm. Unfortunately, since the problem is solved iteratively, the cross validation technique will be time-consuming and will add to the computational complexity of the method. In the next section, we suggest a method for solving this issue.

3.2 R2NEVA

An alternative idea for dealing with the regularization parameter in (20) is to combine the estimation and the validation problems into one problem. This idea was first introduced in Rojas and Hjalmarsson (2011) for l_1-sparse estimation.

The idea of the SPARSEVA method is to relax the least-squares objective to what the least-squares method would achieve on validation data and then use this as a constraint in the minimization of the regularization term. Therefore, instead of solving the problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \|Y - \Phi\theta\|_2^2 \\ \text{such that} \quad & \|\theta\|_1 \leq \eta \end{aligned} \tag{22}$$

we solve

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \|\theta\|_1 \\ \text{such that} \quad & V(\theta) \leq (1 + \epsilon_N)V(\hat{\theta}_{LS}) \end{aligned} \tag{23}$$

where the constant ϵ_N corresponds to the loss of fit that can be expected on validation data. The choice $\frac{2n}{N}$ corresponds to the AIC, $\frac{n\log N}{N}$ to the BIC, and

$$V(\theta) = \|Y - \Phi\theta\|_2^2. \tag{24}$$

This allows us to eliminate all the tuning parameters from the problem.

Using this approach, instead of solving the problem in (20), we solve the problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \|W^k\mathcal{H}_n(\theta)W^k\|_* \\ \text{such that} \quad & V(\theta) \leq (1 + \epsilon_N)V(\hat{\theta}_{LS}). \end{aligned} \tag{25}$$

In the k^{th} step of the algorithm we solve the semidefinite program

$$\begin{aligned} \underset{\theta,X}{\text{minimize}} \quad & \text{trace}(X^k + \delta I_n)^{-1} X \\ \text{such that} \quad & \begin{bmatrix} X & \mathcal{H}_n(\theta) \\ \mathcal{H}_n(\theta) & X \end{bmatrix} \succeq 0 \\ & (1+\epsilon_N) V(\hat{\theta}_{LS}) - V(\theta) \geq 0 \end{aligned} \tag{26}$$

and update the weight matrix W^k and the matrix X^k according to (21). We will call this re-weighted nuclear-norm estimation method *R2NEVA* (Re-weighted Nuclear-Norm Estimation based on VAlidation criteria). When the number of iterations is set to one ($k = 0$ with $X^0 = I$), we will call the method *2NEVA* (Nuclear-Norm Estimation based on VAlidation criteria).

3.3 Likelihood approach for validation

In this section, we will suggest an alternative choice for the constant ϵ_N based on a test statistic and the maximum likelihood principle. Observe that according to the assumed linear regression model we have

$$\hat{\theta}_{LS} = \theta_0 + (\Phi^T \Phi)^{-1} \Phi^T E \tag{27}$$

and

$$\begin{aligned} V_N(\theta) &= \|Y - \Phi\theta\|_2^2 \\ &= (\theta - \hat{\theta}_{LS})^T \Phi^T \Phi (\theta - \hat{\theta}_{LS}) + V(\hat{\theta}_{LS}). \end{aligned} \tag{28}$$

Therefore

$$\frac{V(\theta_0) - V(\hat{\theta}_{LS})}{\sigma^2} = \frac{E^T \Phi (\Phi^T \Phi)^{-1} \Phi^T E}{\sigma^2} \sim \chi^2(n) \tag{29}$$

by the assumption on E and the properties of the matrix $\Phi(\Phi^T\Phi)^{-1}\Phi^T$.

An estimate of the variance σ^2 can be found using the LS solution of the linear regression problem to be

$$\hat{\sigma}^2 = \frac{V(\hat{\theta}_{LS})}{N - 2n}, \tag{30}$$

and we notice that

$$V(\hat{\theta}_{LS}) = E^T (I - \Phi(\Phi^T\Phi)^{-1}\Phi^T) E, \tag{31}$$

therefore

$$\frac{(N-2n)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(N-2n). \tag{32}$$

From (29) and (32) we get the statistic

$$\frac{V(\theta_0) - V(\hat{\theta}_{LS})}{n\hat{\sigma}^2} \sim \mathcal{F}(n, N-2n) \tag{33}$$

Given that $\theta_0 = \theta$, the value

$$\frac{V(\theta) - V(\hat{\theta}_{LS})}{n\hat{\sigma}^2} = \frac{V(\theta) - V(\hat{\theta}_{LS})}{n\frac{V(\hat{\theta}_{LS})}{N-2n}} = \frac{(n-2)(N-2n)}{n(N-2n+2)} \tag{34}$$

maximizes the probability distribution function of the distribution $\mathcal{F}(n, N-n)$. Thus with maximum likelihood we have

$$V(\theta) = \left(1 + \frac{(n-2)}{(N-2n+2)}\right) V(\hat{\theta}_{LS}). \tag{35}$$

This suggests taking

$$\epsilon_N = \frac{(n-2)}{(N-2n+2)}.$$

In what follows, we will call this choice of ϵ_N, the maximum-likelihood approach for validation.

4. ESTIMATING THE TRANSIENT

When solving the linear regression problem in equation (7), the first n inputs and outputs are not used because the initial conditions are not known. If n is relatively large with respect to the size of the data set Z_N, a considerable amount of information is lost by removing the first part of the data. In this section, we show how it is possible to make use of the full data set by estimating an additional FIR model of size n.

Assume that we have a rational model as described in (11), with

$$\begin{aligned} F(q) &= 1 + f_1 q^{-1} + \cdots + f_r q^{-r} \\ B(q) &= b_1 q^{-1} + \cdots + b_r q^{-r} \end{aligned} \tag{36}$$

Then we have the following difference equation

$$\begin{aligned} y(t) + f_1 y(t-1) + \cdots + f_r y(t-r) = \\ b_1 u(t-1) + \cdots + b_r u(t-r) \\ + b_0^\tau \delta_t + b_1^\tau \delta_{t-1} + \cdots + b_r^\tau \delta_{t-r} \end{aligned} \tag{37}$$

The values $b_0^\tau, \ldots, b_r^\tau$ are used to compensate for the initial conditions, namely the values $u(-r+1), \ldots, u(0)$, and $y(-r+2), \ldots, y(0)$ so that we can consider $y(k) = u(k) = 0$ for all $k \leq 0$. Define the transient polynomial

$$B_\tau(q) = b_0^\tau + b_1^\tau q^{-1} + \cdots + b_r^\tau q^{-r}. \tag{38}$$

Then we have the following output-error model with two-inputs and one-output

$$\begin{aligned} y(t) &= \frac{B(q)}{F(q)} u(t) + \frac{B_\tau(q)}{F(q)} \delta_t + e(t) \\ &= G_0(q) u(t) + G_\tau(q) \delta_t + e(t) \end{aligned} \tag{39}$$

with zero initial conditions.

The second transfer operator $G_\tau(q)$ has an impulse input and it represents the effect of the transient due to the unknown initial conditions. This reformulation maps the unknown initial conditions into a corresponding unknown LTI system with the same order and poles as the original system. To proceed numerically, we approximate both operators G and G_τ, as before, with FIR models of order $n > r$, i.e.

$$G_0(q) \approx \sum_{k=1}^{n} g_0(k) q^{-k}, \quad G_\tau(q) \approx \sum_{k=1}^{n} g_\tau(k) q^{-k} \tag{40}$$

and solve the linear regression problem

$$Y_N = \begin{bmatrix} \Phi_N & I_n \\ & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_\tau \end{bmatrix} + E_N \tag{41}$$

in which I_n is the identity matrix and

$$\begin{aligned} \theta_0 &= [g_0(1) \ldots g_0(n)]^T, \quad \theta_\tau = [g_\tau(1) \ldots g_\tau(n)]^T \\ Y_N &= [y(1) \ldots y(N)]^T, \quad E_N = [e(1) \ldots e(N)]^T, \end{aligned}$$

$$\Phi_N = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ u(1) & 0 & \ldots & 0 \\ u(2) & u(1) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u(n) & u(n-1) & \ldots & u(1) \\ u(n+1) & u(n-2) & \ldots & u(2) \\ \vdots & \vdots & \ddots & \vdots \\ u(N-1) & u(N-2) & \ldots & u(N-n) \end{bmatrix}.$$

Notice that G_0 and G_τ share the same poles. Therefore the block-Hankel rectangle matrix

$$\mathcal{H}_n([\theta_0\ \theta_\tau]) := \begin{bmatrix} g(1) & g(2) & \dots & g(k+1) \\ g(2) & g(3) & \dots & g(k+2) \\ \vdots & \vdots & \ddots & \vdots \\ g(k+1) & g(k+2) & \dots & g(n) \end{bmatrix}, \quad (42)$$

in which $g(m) := [g_0(m)\ g_\tau(m)]$ for all $m = 1, \dots, n$ and $n = 2k+1$, will have rank r whenever $n > r$.

Using this, we formulate a similar problem as before. In the k^{th} step of the regularization problem we solve

$$\begin{aligned} \underset{\theta,\eta,X,Z}{\text{minimize}} \quad & \|Y_N - \Phi_N\theta - \eta\|_2^2 + \lambda(\ \text{trace}(X^k + \delta I_n)^{-1}X \\ & +\ \text{trace}(Z^k + \delta I_n)^{-1}Z) \\ \text{such that} \quad & \begin{bmatrix} X & \mathcal{H}_n(\theta,\eta) \\ \mathcal{H}_n(\theta,\eta)^T & Z \end{bmatrix} \succeq 0 \end{aligned} \quad (43)$$

to get θ^{k+1}, η^{k+1}. Then we have

$$\begin{aligned} X^{k+1} &= (W_1^k)^{(-1)}U\Sigma U^T(W_1^k)^{(-1)} \\ Z^{k+1} &= (W_2^k)^{(-1)}V\Sigma V^T(W_2^k)^{(-1)} \\ W_1^{k+1} &= (X^{k+1} + \delta I_n)^{-1}, \\ W_2^{k+1} &= (Z^{k+1} + \delta I_{2n})^{-1} \end{aligned} \quad (44)$$

in which $U\Sigma V^T$ is the singular value decomposition of the rectangle matrix $W_1^k\mathcal{H}_n(\theta^{k+1}, \eta^{k+1})W_2^k$. The regularization parameter λ is selected by using cross-validation. We will denote this method by RNN-CV+Transient.

5. ILLUSTRATIVE NUMERICAL EXAMPLE

In this section, the algorithm is tested on a 6th order system with poles at $0.5 \pm 0.75i, -0.8 \pm 0.45i, -0.4 \pm 0.7i$ and zeros at $-0.2 \pm 0.85i, 0.4 \pm 0.2i, -0.5, 0.1$ and one delay. The experiment is conducted in MATLAB using YALMIP and the solver SDPT3. We have used a Gaussian white input with variance 1, and the noise variance is 30. The sample size for each run $N = 500$, and the number of estimated parameters is $n = 69$. It was chosen to guarantee that $g_0(k)$ for $k > n$ are close to zero. In the following we compare the different methods discussed in the previous sections. The comparison is done in terms of model fit as measure of accuracy of the estimate. The fit is defined as

$$\text{fit} = 100\left(1 - \sqrt{\frac{\sum_{k=1}^n |g_0(k) - \hat{g}(k)|^2}{\sum_{k=1}^n |g_0(k) - \bar{g}_0|^2}}\right) \quad (45)$$

in which

$$\bar{g}_0 = \frac{1}{n}\sum_{k=1}^n g_0(k) \quad (46)$$

When the fit $= 100$, the estimate coincides with the true parameters. The results are shown in terms of box plots in which the median (of all runs) is represented by the red line, and the mean is given by the diamond.

5.1 NN-CV vs. RNN-CV

We first compared the two methods:

- NN-CV; nuclear-norm regularization with cross validation, and
- RNN-CV; re-weighted nuclear norm regularization with cross validation. The parameter $\delta = 0.1$.

As shown in Fig. 1, RNN-CV performed better that NN-CV for the considered system.

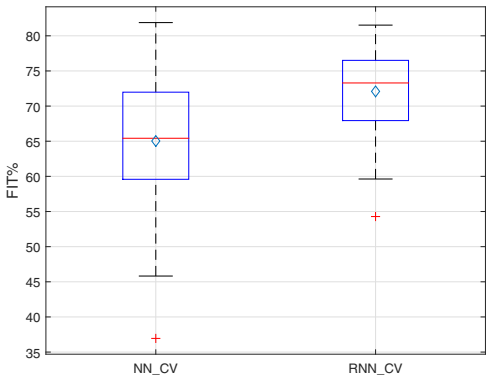

Fig. 1. Box plots for 50 fits comparing NN-CV and RNN-CV

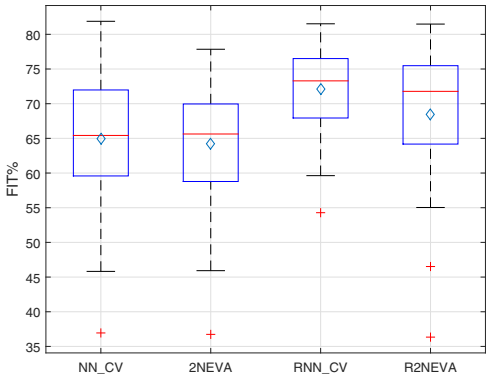

Fig. 2. Box plots for 50 fits comparing NN-CV, 2NEVA, R2NEVA, and RNN-CV methods

5.2 CV vs. EVA

Next we compare the results obtained by the cross-validation techniques to the methods:

- 2NEVA; nuclear norm regularization based on ML approach for validation (see subsection 3.3), and
- R2NEVA; re-weighted nuclear norm regularization based on ML approach for validation. The regularization parameter $\delta = 0.1$. The initial weights W^1 are taken to be the identity.

Fig. 2 shows that 2NEVA gives very close result to that obtained with NN-CV. For this example, R2NEVA gives a slightly less favorable performance compared to RNN-CV, but the computational burden is reduced significantly (40 times in this example).

5.3 Comparison to other methods

Finally we compare the following three methods

- RNN-CV+Transient; re-weighted nuclear norm regularization with cross validation and estimated transient. The regularization parameter $\delta = 0.1$,
- N4SID-NN; nuclear norm for subspace identification *using the true system order* (Liu et al., 2013). The weighting matrices of the algorithm are chosen according to CVA (canonical variate analysis), and

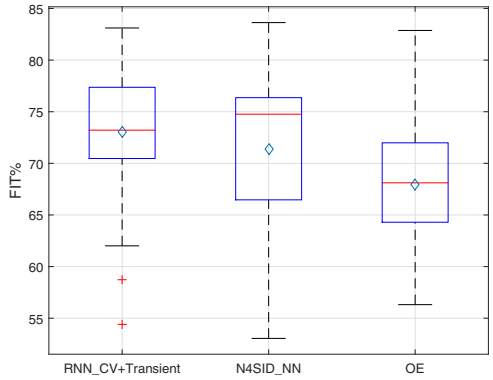

Fig. 3. Box plots for 50 fits comparing RNN-CV+Transient, N4SID-NN, and OE methods

- OE; PEM using an output-error model *using the true system order*. The MATLAB function `oe` is used with tolerance = 0.0001, and the initial conditions are set to `'estimate'`.

Fig. 3 shows that the two nuclear norm based methods perform best on this example. N4SID-NN has a slightly better median behavior while RNN-CV+Transient has slightly better average behavior and less spread.

It is important to emphasize that the two methods N4SID-NN and OE are not performing any order selection. They are equipped with an orcale that gives the true system order. This is in contrast to RNN-CV+Transient which does not use this prior knowledge.

6. CONCLUSIONS

In this contribution we developed a new regularization technique for estimation of output error models. The method is based on a re-weighted nuclear approximation of the rank of the Hankel matrix of the impulse-response. For the implementation, the method from SPARSEVA has been used to remove the regularization parameter from the problem. We suggested a new value for ϵ_N based on a test statistic and the ML-principle. We also developed a method for estimating the transient in cases where the initial conditions are unknown.

The numerical example shows that the re-weighting in the nuclear norm heuristics can improve the performance. By using the methods 2NEVA and R2NEVA we are able to reduce the computational burden of the method on the cost of a slightly inferior performance. The simulation shows that the developed method is capable of achieving competitive performance when compared to other methods.

REFERENCES

Fazel, M., Hindi, H., and Boyd, S. (2001). A rank minimization heuristic with application to minimum order system approximation. In *American Control Conference, 2001. Proceedings of the 2001*, volume 6, 4734–4739 vol.6.

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

Gebraad, P., van Wingerden, J.W., van der Veen, G., and Verhaegen, M. (2011). LPV subspace identification using a novel nuclear norm regularization method. In *American Control Conference (ACC), 2011*, 165–170.

Grossmann, C., Jones, C., and Morari, M. (2009). System identification via nuclear norm regularization for simulated moving bed processes from incomplete data sets. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, 4692–4697.

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

Hjalmarsson, H., Welsh, J.S., and Rojas, C.R. (2012). Identification of box-jenkins models using structured arx models and nuclear norm relaxation. In *16th IFAC Symposium on System Identification*, number 16 in IFAC Proceedings Volumes (IFAC-PapersOnline), 322–327. IFAC. QC 20121122.

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

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

Liu, Z. and Vandenberghe, L. (2009). Semidefinite programming methods for system realization and identification. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, 4676–4681.

Ljung, L. (ed.) (1999). *System Identification (2Nd Ed.): Theory for the User*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

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

Mohan, K. and Fazel, M. (2010). Reweighted nuclear norm minimization with application to system identification. In *American Control Conference (ACC), 2010*, 2953–2959.

Pillonetto, G., Dinuzzo, F., Chen, T., Nicolao, G.D., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3), 657 – 682.

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

Rojas, C. and Hjalmarsson, H. (2011). Sparse estimation based on a validation criterion. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 2825–2830.

Verhaegen, M. and Hansson, A. (2014). Nuclear norm subspace identification (n2sid) for short data batches. *ArXiv e-prints*.