Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Linking regularization and low-rank approximation for impulse response modeling

Anna Marconato * **Lennart Ljung** ** **Yves Rolain** *
Johan Schoukens *

* *Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. (e-mail: anna.marconato@vub.ac.be)*
** *Division of Automatic Control, Department of Electrical Engineering, Linköping University, SE-581 83, Linköping, Sweden.*

Abstract:
In the last years, nonparametric linear dynamical systems modeling has regained attention in the system identification world. In particular, the application of regularization techniques that were already widely used in statistics and machine learning, has proven beneficial for the estimation of the impulse response of linear systems. The low-rank approximation of the impulse response obtained by the truncated singular value decomposition (SVD) also leads to reduced complexity estimates. In this paper, the link between regularization and SVD truncation for finite impulse response (FIR) model estimation is made explicit. The SVD truncation is reformulated as a regularization problem with a specific choice of the regularization matrix. Both approaches (regularization and SVD truncation) are applied on a FIR modeling example and compared with the classic prediction error method/maximum likelihood approach. The results show the advantage of these techniques for impulse response estimation.

1. INTRODUCTION

The aim of system identification is to obtain a mathematical description (a *model*) of (non)linear systems starting from a set of input-output observations.

For linear dynamic modeling in particular, a well-established theoretical framework exists. Specific methods are available to tackle this estimation problem (Ljung [1999], Söderström and Stoica [1989], Pintelon and Schoukens [2012]).

New possibilities and challenges are offered by the introduction of regularization in system identification. See Section IV in Ljung et al. [2011] for an overview of recent advances in the field. Regularization is of particular interest in the bias-variance trade-off that characterizes model estimation from noisy observations (Ljung and Chen [2013]).

The key idea in regularization is the introduction of a penalty term to the (least squares) cost function to limit the model complexity (see e.g. Hastie et al. [2009]). Several regularization schemes exist in the literature. They differ in the definition of the norm used to define the penalty term (e.g. AIC, Akaike [1974], Lasso, Tibshirani [1994], ridge regression, Hoerl and Kennard [1970]). The common feature in all these methods is to reduce the model complexity by pulling as many parameters as possible to zero.

In the last years, the drive to remove noise sensitivity has brought new attention to the nonparametric identification of linear dynamical systems. The beneficial impact of regularization schemes in these simple problems has been pointed out in several papers.

In Chen et al. [2012], regularization is employed for the estimation of finite impulse response (FIR) models to address high variance issues of the estimates. For noisy and short data records, the regularized least squares approach turns out to be more robust and even slightly more accurate than the standard prediction error method/maximum likelihood (PEM/ML) approach. The underlying reason for such behavior is that it may be beneficial to allow some bias to reduce the variance. Based on a Bayesian interpretation of the problem, one gets insight in the choice of the regularization matrix. Prior knowledge about the distribution of the parameters here boils down to imposing the presence of an exponentially decaying impulse response. More details about how to tune the hyperparameters that steer the regularized FIR model estimation can be found in Chen and Ljung [2013].

The approach discussed in Chen et al. [2012] is strongly linked with the results presented in Pillonetto and De Nicolao [2010]. In that work, the impulse response of a stable linear system is estimated using a Gaussian process model. A Bayesian framework is invoked to produce a nonparametric estimate in an infinite-dimensional space;

⋆ This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), by the Belgian Government through the Inter university Poles of Attraction (IAP VII) Program, the ERC Advanced Grant SNLSID, and by the ERC advanced grant LEARN, no 287381, funded by the European Research Council.

in a second step a low-order model can be obtained by projecting this solution onto a finite-dimensional space.

In Hjalmarsson et al. [2012], the regularization for linear model estimation is obtained including two additional terms in the least squares cost function. A structured high-order ARX model is obtained by including the nuclear norm of two Hankel matrices derived from the noise and the system models.

In Rojas and Hjalmarsson [2011], sparse solutions for linear regression problems are obtained by minimizing the ℓ_1-norm of the parameter vector. The resulting performance is similar to that of a validated least-squares estimate on a separate set of validation data.

Another possible approach to reduce the intrinsic complexity of a model has recently been discussed in Marconato et al. [2013]. In that work, reduced rank solutions are obtained based on the truncated singular value decomposition (SVD) of the regressor matrix. This approach is applied to both linear and nonlinear in the parameters problems. By explicitly forcing the smallest singular values to zero, a low-rank model is estimated that does not rely on the parameterization of the model.

Note that in this last case the solution still lives in the original space, but it lies in a lower dimensional subspace. This subspace depends on the coloring of the excitation signal. Following this approach, it becomes possible to reduce the model complexity without pulling individual parameters to zero.

This paper shows the link between regularization and low-rank approaches to reduce the model complexity.

More specifically, the focus is put on a specific estimation example: the FIR modeling problem. For the sake of simplicity, only the single-input single-output (SISO) case is considered.

The impulse response of an unknown linear dynamic system is estimated with a FIR model. This representation assumes that the impulse response is a finite length sequence b_k:

$$\hat{y}(t) = \sum_{k=0}^{n_b-1} b_k u(t-k) \tag{1}$$

The n_b parameters b_k for $k = 0, \cdots, n_b - 1$ represent the estimated impulse response.

A set of N input-output observations $\{(u(t), y(t))\}_{t=1}^{N}$ is assumed to be available to estimate the FIR model. Note that the order n_b needs to be selected by the user based on some metric (e.g. based on cross-validation results). Its choice is directly related to the bias-variance trade-off.

The regularization approach and the SVD truncation ideas are applied to this FIR modeling example. Their relative performance is compared with the result obtained using a classic PEM/ML approach. Remember that this does not use prior knowledge about the impulse response. To show the usefulness of the compared techniques, a very simple example was selected on purpose, as this stresses the relevance of the methods even for a problem that is as simple as the FIR model estimation.

The paper shows the link between the regularization and the low-rank approximation via SVD truncation approaches explicitly. The two problem formulations are proven to be equivalent, provided that the regularization matrices are selected in a specific way.

Section 2 presents a summary of the regularization ideas applied specifically to the FIR modeling problem. Section 3 applies the truncated SVD approach to the same problem. In Section 4, the SVD truncation is reformulated as a regularization problem. This requires one to define a specific choice of the regularization matrix.

Despite this close connection between the two techniques, there is an important distinction in their derivation. The regularization approach exploits some assumptions about the prior knowledge of the system to be modeled (exponential decay and smoothness of the impulse response). The truncated SVD method relies entirely on the nature of the excitation signal (colored excitation). A comparison of the different results obtained on a specific FIR modeling example is provided in Section 5. Final remarks and considerations are summarized in Section 6.

2. REGULARIZATION FOR SYSTEM IDENTIFICATION – FIR CASE

2.1 The LS Estimate

Consider the finite impulse response model (1). Introduce the column vector Y for all the output measurements, and the column vector θ for the impulse response coefficients b_k. Collect the input measurements in the Toeplitz matrix $K \in \mathbb{R}^{N \times n_b}$ (regressor matrix)

$$K = \begin{pmatrix} u(1) & u(0) & \cdots & u(-n_b+2) \\ u(2) & u(1) & \cdots & u(-n_b+3) \\ \vdots & \vdots & \ddots & \vdots \\ u(N) & u(N-1) & \cdots & u(N-n_b+1) \end{pmatrix} \tag{2}$$

Eq.(1) can be written compactly as the linear regression model

$$Y = K\theta + E \tag{3}$$

where the vector E denotes the additive errors in the measurements of the output y. The well known least squares estimate of the parameters is defined to be

$$\hat{\theta} = \arg\min_{\theta} \|Y - K\theta\|^2 = (K^T K)^{-1} K Y \tag{4}$$

This estimate could be troublesome (have high variance) if the matrix $K^T K$ is ill-conditioned. This will happen e.g. if the FIR order n_b is large or the input u is poorly exciting.

2.2 The Regularized LS Estimate

A well known remedy for such bad conditioning is to use *regularization*, i.e. to add a penalty to the criterion that depends on the parameter size, see e.g. page 221 in Ljung [1999]:

$$V(\theta) = \|Y - K\theta\|^2 + \lambda(\theta - \theta^*)^T R(\theta - \theta^*) \tag{5}$$

This modifies the estimate to

$$\hat{\theta}_{reg} = (K^T K + \lambda R)^{-1} K^T Y \tag{6}$$

if $\theta^* = 0$. This estimate for FIR models is discussed in some detail in Chen et al. [2012]. We may note that the

regularization only depends on the product λR, but it is useful to think of the scalar λ as a simple tuning variable for the "amount" of regularization.

It remains to decide how to choose the regularization λR. It is useful to realize that the estimate $\hat{\theta}_{reg}$ equals the Bayesian *maximum a posteriori, MAP* estimate, in case the prior distribution of $\theta \in N(0, \Pi)$, that is, it is a zero mean Gaussian with covariance matrix

$$\Pi = \sigma^2 R^{-1}/\lambda \quad \text{or } \lambda R = \sigma^2 \Pi^{-1} \tag{7}$$

where σ^2 is the variance of the (white) measurement errors.

2.3 Parameterization of the Regularization Matrix and Empirical Bayes

If we believe that $\theta \in N(0, \Pi)$ and the additive noise $E \in N(0, \sigma^2 I)$, then $Y = K\theta + E \in N(0, Z(\Pi))$ where $Z(\Pi) = K\Pi K^T + \sigma^2 I$. That means that the distribution of Y is known up to (parameters present in) Π. So they could be estimated by maximum likelihood:

$$\hat{\Pi} = \arg\max_{\Pi} \log P(Y|\Pi) \tag{8}$$

$$= \arg\min_{\Pi} Y^T Z(\Pi)^{-1} Y + \log\det Z(\Pi) \tag{9}$$

This method is known as *empirical Bayes*. Supposing that the impulse response of the system is exponentially stable and smooth, it is natural to parameterize Π as so that its k, r element is

$$\Pi_{k,r} = C\alpha^{k+r}\rho^{|k-r|}, \quad C > 0,\ 0 < \alpha < 1,\ |\rho| < 1 \tag{10}$$

So, eqs.(6)-(10) give a complete algorithm to estimate an FIR model for a stable linear system using a well tuned regularization.

Note that here the added prior knowledge is the exponential decay and the smoothness of the impulse response.

3. LOW-RANK SOLUTION WITH SVD TRUNCATION – FIR CASE

Consider the same linear least squares solution $\hat{\theta}$ given by eq.(4):

$$\hat{\theta} = (K^T K)^{-1} K Y$$

Substitute the SVD of matrix K, $K = U\Sigma V^T$, herein:

$$\hat{\theta} = V\Sigma^{-1} U^T Y \tag{11}$$

Starting from this expression, a low-rank solution is obtained whenever the singular value spectrum of the matrix Σ is truncated. The truncated matrix Σ_n is obtained by keeping the first n singular values $\sigma_1, \cdots, \sigma_n$ and setting the remaining $n_b - n$ singular values $\sigma_{n+1}, \cdots, \sigma_{n_b}$ to zero.

To propagate the rank reduction to $\hat{\theta}$, the matrix Σ_n^{-1} is defined as:

$$\Sigma_n^{-1} = \text{diag}\left(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_n}, 0, \cdots, 0\right) \tag{12}$$

The resulting low-rank solution $\hat{\theta}_n$ becomes:

$$\hat{\theta}_n = V\Sigma_n^{-1} U^T Y \tag{13}$$

Note that despite $\hat{\theta}_n$ is still in $\mathbb{R}^{n_b}$ (all parameters are non-zero), it lies in a n-dimensional subspace of $\mathbb{R}^{n_b}$, with $n < n_b$.

The tuning parameter here is the truncation level n which is user selectable. This n can for example be determined using *cross validation*, that is generate a separate validation data set with output $y(t)$ and computing a model output $\hat{y}(t)$ using the tested model and the validation data input. Then evaluate the validation cost function:

$$e_{\text{RMS}}^{val} = \sqrt{\frac{1}{N_{val}} \sum_{t=1}^{N_{val}} (y(t) - \hat{y}(t))^2} \tag{14}$$

The value of n corresponding to the lowest validation error is then selected.

The quality of the truncated solution strictly depends on the nature of the singular value spectrum of the input dependent regressor matrix K.

If $u(t)$ is a white noise sequence, the columns of K are linearly independent. The magnitude of all singular values is therefore asymptotically equal. A truncation of Σ therefore becomes impossible without compromising the quality of the estimates significantly.

If $u(t)$ is a colored noise excitation, the magnitude of the singular values will decrease. A high quality, low-rank solution can then be obtained using the truncation described above.

In the remainder of the paper the input signal will be assumed to be a colored noise sequence.

Next, the SVD truncation is reformulated as a regularization problem.

Note that here the prior is given by the coloring of the input (which results in the truncation of the Σ matrix).

4. FORMULATING THE SVD TRUNCATION AS A FORMAL REGULARIZATION TASK

Section 2 shows how the regularization approach includes a penalty term in the cost function to constrain the model complexity, eq.(5-6).

For a specific choice of the regularization matrix, we show now that the regularized solution coincides with the low-rank solution (13) obtained by SVD truncation.

Substituting the SVD of the full regressor matrix $K = U\Sigma V^T$ in eq.(6), one obtains:

$$\hat{\theta}_{\text{reg}} = (V\Sigma^2 V^T + \lambda R)^{-1} V\Sigma U^T Y \tag{15}$$

Now define

$$D = \text{diag}(\underbrace{\varepsilon, \cdots, \varepsilon}_{n}, \underbrace{1, \cdots, 1}_{n_b - n}) \tag{16}$$

$$\lambda R = \lambda V D V^T \tag{17}$$

with λ the regularization factor and $\varepsilon \sim O(\frac{1}{\lambda^2})$.

The estimate boils down to

$$\hat{\theta}_{\text{reg}} = (V(\Sigma^2 + \lambda D)V^T)^{-1} V\Sigma U^T Y \tag{18}$$

$$= V(\Sigma^2 + \lambda D)^{-1} \Sigma U^T Y$$

Note that

$$\Sigma^2 + \lambda D = \text{diag}(\sigma_1^2 + \lambda\varepsilon, \cdots, \sigma_n^2 + \lambda\varepsilon, \sigma_{n+1}^2 + \lambda, \cdots, \sigma_{n_b}^2 + \lambda) \tag{19}$$

and consequently Σ_n^{-1} becomes

$$\Sigma_n^{-1} = \lim_{\lambda\to\infty} (\Sigma^2 + \lambda D)^{-1}\Sigma \tag{20}$$

$$= \lim_{\lambda\to\infty} \text{diag}(\sigma_1^2 + \lambda\varepsilon, \cdots, \sigma_n^2 + \lambda\varepsilon, \sigma_{n+1}^2 + \lambda, \cdots, \sigma_{n_b}^2 + \lambda)^{-1}\Sigma \tag{21}$$

$$= \text{diag}\left(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_n}, 0, \cdots, 0\right) \tag{22}$$

Finally, one can write:

$$\hat{\theta}_{\text{reg}} = V\Sigma_n^{-1}U^T Y = \hat{\theta}_n \tag{23}$$

To interpret the SVD truncation as a regularization, one needs to fix the truncation level to n, and subsequently use the regularization matrix of eq.(17).

From a Bayesian point of view, this formulation corresponds to a different prior assumption. A certain fixed projection P of the parameter vector θ is made. This zeroes the contribution of θ in the orthogonal complement $P^\perp$ of the projection space.

As $\lambda R = \Pi^{-1}$, with Π the covariance matrix of θ, eq.(17) shows that D^{-1} is not of full column rank. In the P space, this corresponds to parameters being set to zero. Infinite values in D^{-1} correspond to zeros in D. This means that no prior knowledge is available about the corresponding elements of θ, and one has to rely upon the data information to determine their values.

One can also understand this by considering eq.(17) and the cost function in eq.(5). Elements of θ corresponding to the one's in the matrix D need to be equal to zero in order to keep the cost function from tending to infinity, while the other parameters are free.

5. RESULTS

The behavior of the two frameworks presented in the previous sections is now illustrated on a simulation example.

The goal is to estimate a FIR model that mimics the behavior of a linear dynamical infinite impulse response system, excited by a colored excitation. The system considered in this example is a second order Butterworth filter, and all considered signals are defined as follows:

$$y(t) = G_0 u(t) + e(t) \tag{24}$$

$$G_0(z) = \frac{0.0201 + 0.0402z^{-1} + 0.0201z^{-2}}{1 - 1.5610z^{-1} + 0.6414z^{-2}} \tag{25}$$

$$u(t) = Fr(t) \qquad r(t) \sim \mathcal{N}(0,1) \tag{26}$$

$$F(z) = \frac{0.0675 + 0.1349z^{-1} + 0.0675z^{-2}}{1 - 1.1430z^{-1} + 0.4128z^{-2}} \tag{27}$$

$$e(t) \sim \mathcal{N}(0, \sigma^2) \tag{28}$$

The true impulse response of the system is shown in blue (full line) in Figure 1.

The system is excited by a 1000 samples colored noise sequence. A small white noise disturbance (with variance $\sigma^2 = 0.0025$) is added to the output. The coloring of the excitation plays a fundamental role in the truncated SVD approach, since it determines the decrease of the magnitude of the singular values of the regressor matrix K.

Following the standard PEM/ML approach, one has to determine the order n_b of the FIR model first. An order estimation criterion combined with the evaluation of a range of model orders is used to select the "best" order. Note that since the FIR model order can grow with the amount of data, this is a nonparametric approach. The true impulse response visually reaches zero within the measurement resolution after less than 50 samples on Figure 1. All model orders $n_b = 1, \cdots, 50$ are hence considered.

A validation cost function is chosen to evaluate the model quality on a portion of the data that was not used to identify the model. Here, a validation data set is generated, using the same colored noise source. $N_{val} = 1000$ samples are drawn and the output corrupted with white noise (variance 0.0025) as in eqs.(24)-(28).

The root mean squared error on the validation set e_{RMS}^{val} as in eq.(14) is computed to evaluate the model performance.

The PEM/ML approach based on e_{RMS}^{val} returns $n_b = 23$ as the best model order. The estimated impulse response is shown in Figure 1. e_{RMS}^{val} is equal to 0.0506 here. As the standard deviation of the output noise is equal to 0.05, the noise component clearly dominates the output variance.

To evaluate the theoretical quality of the considered model, we consider the following performance measure:

$$e_{\text{RMS}} = \sqrt{\int_{-\pi}^{\pi} \left|G_0(\omega) - \hat{G}(\omega,\theta)\right|^2 \phi_u(\omega)\, d\omega} \tag{29}$$

where $G_0(\omega)$ is the true frequency response of the system, $\hat{G}(\omega,\theta)$ is the estimated frequency response, and $\phi_u(\omega)$ is the power spectrum of the colored input signal. In a simulation, $G_0(\omega)$ is known.

To compute e_{RMS} in practice, we generate a very large noisefree validation output signal (100000 samples). In a real measurement based situation such a validation data set will not be available. Here e_{RMS} in eq.(29) is used to better understand the properties of the different models. Note that the model order selection is still based on e_{RMS}^{val}, using a short and noisy validation set. Only the final comparison of the different estimated models uses e_{RMS}.

For the PEM/ML solution $e_{\text{RMS}} = 0.0073$. However, Figure 1 clearly shows that the dashed line representing the PEM/ML estimate is not satisfactory.

This poor result becomes clear looking at the power spectral density (PSD): the low-pass input PSD does not sufficiently excite the system in the measurement band. This motivates the use of the SVD truncation instead.

For the SVD truncation, one needs to select the order n_b of the FIR model and the level of truncation n. The "optimal" combination for n_b and n is found here by scanning e_{RMS}^{val} for each pair (n_b, n) with $n_b = 1, \cdots, 50$ and $n = 1, \cdots, n_b$. The best value for e_{RMS}^{val} is obtained for $n_b = 24$ and $n = 10$. This means that the estimate contains 24 non-zero parameters, but it has a lower rank of only 10. The e_{RMS} value is equal to 0.0049, which shows a

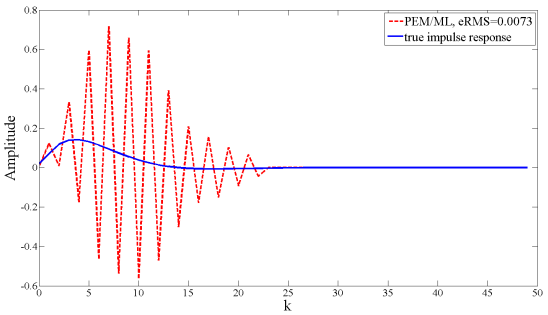

Fig. 1. True impulse response of the linear system under study (blue), and estimated impulse response for $n_b = 23$ (red dashed).

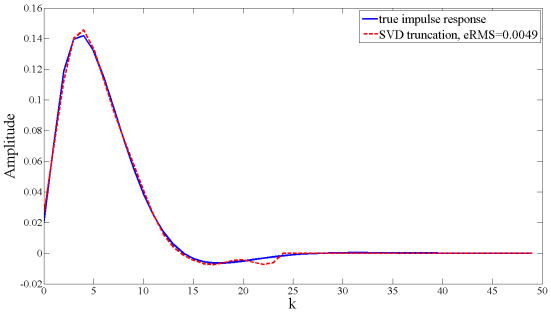

Fig. 2. Estimated impulse response with the SVD truncation approach for $n_b = 24$ and $n = 10$ (red dashed), compared to the true impulse response (blue).

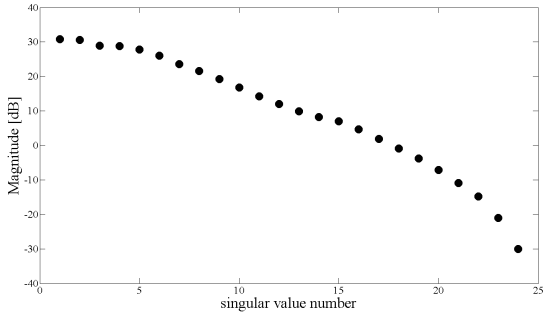

Fig. 3. Magnitude (in dB) of the singular values of matrix K for $n_b = 24$.

significant improvement when compared to the PEM/ML estimate.

Figure 2 shows the impulse response obtained with the truncated SVD. It approximates the true response of the system much more accurately than the PEM/ML estimate. Note that the e_{RMS} of the PEM/ML estimate was less than a factor 2 larger than for the SVD truncation. The big difference in the impulse response estimates (Figures 1 and 2) is due to the fact that the validation cost does not penalize errors in the poorly excited frequency band.

The magnitude of the singular values of the matrix K is shown in Figure 3 for $n_b = 24$. Only the first 10 singular values contribute to the low-rank solution. All the remaining ones are set to zero (see eq.(12)).

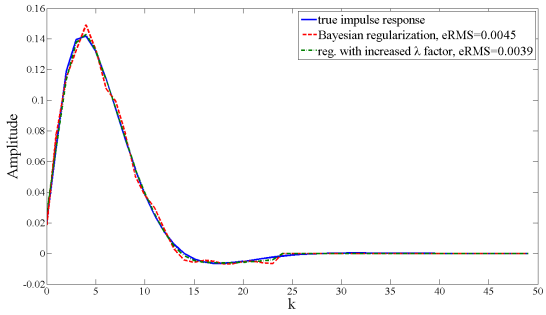

Fig. 4. Estimated impulse response with the tuned regularization approach for $n_b = 24$ (red dashed), and with increased regularization level with $l = 14$ (green dash-dot), compared to the true impulse response (blue).

One obtains the same solution using the regularization approach as above, by calculating D, λ and R as in eqs.(16) and (17) for $n_b = 24$ and $n = 10$.

Next, the problem is solved by regularization using empirical Bayes like in eqs.(6)-(10). A well tuned regularized least squares approach can significantly improve the results of PEM/ML, since it improves the bias-variance trade-off (see e.g. Chen et al. [2012] and Ljung and Chen [2013]).

Using the Bayesian interpretation of Section 2, the smoothness assumption for the exponentially decaying impulse response, one obtains the values of λ and R to be used in eq.(6). This can be repeated for all possible FIR orders $n_b = 1, \cdots, 50$. The best regularized solution is then selected using e^{val}_{RMS}. All results presented below are obtained running the R2013b version of the MATLAB System Identification Toolbox, Ljung [2013].

The best regularized solution in the range $n_b = 1, \cdots, 50$ is obtained for $n_b = 24$ (Figure 4), with a e_{RMS} equal to 0.0045.

This solution can even be improved further by increasing the "amount" of regularization, i.e. by multiplying λ with a factor $l > 0$ that can be tuned by the user using the validation criterion (14) (Figure 4). For $l = 14$ one obtains a e_{RMS} value of 0.0039.

Moreover, if one performs a double scan minimizing e^{val}_{RMS} over all possible FIR orders $n_b = 1, \cdots, 50$, and with increased regularization factors $l = 1, \cdots, 30$, the best solution ($e_{\text{RMS}} = 0.0027$) is found for $n_b = 28$ and $l = 19$.

Figure 5 shows the amplitude of the error between the true impulse response and the estimated impulse response for the three models obtained with the Bayesian regularization. One can see that when the regularization level is increased (green and black lines) the estimate gets smoother.

These results show that, in a situation where one has a colored excitation, both the SVD truncation and the tuned regularization approach are useful tools to improve the PEM/ML estimate significantly. A smoother approximation of the impulse response results. In particular, including prior knowledge about the system response in the regularization approach, and increasing the amount of

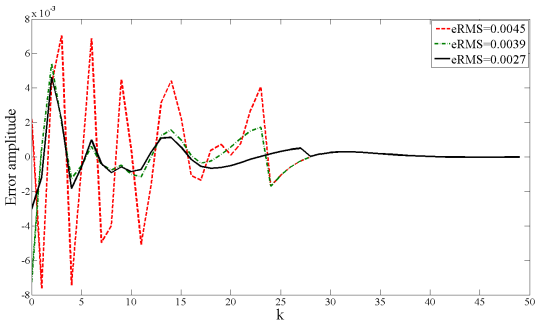

Fig. 5. Error between the true impulse response and the estimated models for the regularization approach. Tuned regularization approach for $n_b = 24$ (red dashed), increased regularization level with $n_b = 24$ and $l = 14$ (green dash-dot), and best solution obtained for $n_b = 28$ and $l = 19$ (black).

regularization by further tuning the value of λ allows one to obtain very accurate models.

6. CONCLUSION

In this paper we have considered the estimation of FIR models to approximate the response of linear dynamical systems, in the situation where the system is excited with a colored noise input. For this problem, tools like regularization and low-rank approximation methods based on the truncated SVD prove successful.

The tuned Bayesian regularization approach uses the prior knowledge that the system response is smooth and exponentially decaying. For the SVD truncation method the prior is instead given by the coloring of the noise input. The link between these two approaches has been shown by reformulating the SVD truncation as a regularization problem given a specific choice of the regularization matrix.

Both methods have been tested on a FIR modeling example with colored excitation. The obtained results outperform significantly the PEM/ML estimate, with a decrease of the validation error of more than 60%. In particular, the regularization approach allows one to obtain a very accurate solution, by incorporating prior knowledge about the system impulse response to tune the regularization matrix R and by further increasing the factor λ. As a future research direction, the possibility of combining the two approaches will be investigated, to include in the prior both the assumptions about the system and the information about the coloring of the input data.

REFERENCES

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716–723, 1974.

T. Chen and L. Ljung. Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. *Automatica*, 49(7):2213–2220, 2013.

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

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2009.

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*, Brussels, Belgium, 2012.

A.E. Hoerl and R.W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.

L. Ljung. *System Identification: Theory for the User (2nd ed.)*. Prentice Hall, New Jersey, 1999.

L. Ljung. *The System Identification Toolbox: The Manual*. The MathWorks Inc., Edition 8.3 2013, Natick, MA, USA, 2013.

L. Ljung and T. Chen. What can regularization offer for estimation of dynamical systems? In *16th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, Caen, France, 2013.

L. Ljung, H. Hjalmarsson, and H. Ohlsson. Four encounters with system identification. *Eur. J. Control*, 17(5-6): 449–471, 2011.

A. Marconato, M. Schoukens, Y. Rolain, and J. Schoukens. Study of the effective number of parameters in nonlinear identification benchmarks. In *52nd IEEE Conference on Decision and Control*, Florence, Italy, 2013.

G. Pillonetto and G. De Nicolao. A new kernel-based approach for linear system identification. *Automatica*, 46(1):81–93, 2010.

R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. Wiley-IEEE Press, 2nd edition, 2012.

C.R. Rojas and H. Hjalmarsson. Sparse estimation based on a validation criterion. In *50th IEEE Conference on Decision and Control and European Control Conference*, pages 2825–2830, 2011.

T. Söderström and P. Stoica. *System identification*. Prentice-Hall, 1989.

R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.