2021 Australian & New Zealand Control Conference (ANZCC)
25-26 November 2021. Gold Coast, Australia

Regularized impulse response estimation for systems with colored output noise

Emerson C. Boeira[1] and Diego Eckhard[2]

Abstract— **This paper addresses the use of the regularization feature on impulse response estimation for systems with colored output noise. Firstly, it is shown that the optimal regularization matrix for this scenario is quite different than the optimal for the white noise case and that there is a direct relationship between the Regularized Weighted Least-Squares with a Bayesian perspective of the identification problem for such case. Also, a new Empirical Bayes method, based on the Bayesian perspective, is introduced to estimate the regularization and noise covariance matrices from data. Finally, a numerical example demonstrates that this new methodology outperforms the traditional Regularized Least-Squares, producing better statistical properties and better results for a model fit measure.**

I. INTRODUCTION

The study of System Identification has its origins on the field of Statistics and it deals with the problem of modeling dynamic systems based on observed data [1]. Particularly, within the Control Systems field, there is a continuous demand for model quality improvement in order to reduce the system's uncertainties and result in higher efficiency and better control loop performances.

That being said, over the past decade, a new paradigm has been emerging in the System Identification literature, based on recent and brilliant ideas that were brought from the Machine Learning community, and exposed in [2], [3], which are deeply connected to the use of regularization on impulse response estimation methods [4], [5]. As stated in [5], the great advantage of these new regularized methodologies is that they can outperform the classical system identification paradigm of Prediction Error Methods (PEM), producing a better bias-variance trade-off on estimated models.

Up until this moment in this new regularized system identification context, in order to identify the process' model, the noise transfer function is usually considered as a unitary gain, which means that the output noise is usually considered as white noise, a condition that seems fairly restrictive. However, when the noise filter assumes a rational transfer function structure, which means that the output noise is colored, the current regularized approach is to identify the process' optimal predictor using a FIR (Finite Impulse Response) MISO (Multiple-Input Single-Output) model [6], instead of the process' model itself, which can be an essential feature for some kind of applications such as model-based controller design.

Based on this scenario, this work addresses the analysis of the regularized impulse response identification of the process' model when the system is contaminated with output colored noise, which can be considered as more wide an generic formulation for the regularized identification problem. We propose a theoretical analysis on the optimal regularization matrix, alongside with new ideas to solve the problem, which are based on the Bayesian perspective for system identification, and that results on the use of the Regularized Weighted Least-Squares (RWLS) method and the Empirical Bayes to estimate some prior information concerning the system and the noise characteristics.

This work is organized as follows. Section II presents the current state-of-the-art regularized algorithm for impulse response estimation along with its properties and the differences that emerge by considering the colored noise scenario. In the sequence, Section III demonstrates the RWLS method, and its properties, and Section IV demonstrates the Bayesian perspective of impulse response estimation. In Section V, hyperparameters estimation are discussed for the regularization and the noise covariance matrices, and in Section VI a numerical example is presented. Finally, section VII ends the paper with the concluding remarks.

II. REGULARIZED LEAST-SQUARES FOR IMPULSE RESPONSE ESTIMATION

The regularized least-squares identification technique aims to identify the typical SISO (Single-Input Single-Output) linear time-invariant system, described by [1]

$$y(t) = G_0(q)u(t) + H_0(q)e(t) \tag{1}$$

where q represents the forward-shift operator, $y(t) \in \mathbb{R}$ denotes the system's measured output, $G_0(q)$ represents its true transfer function, $u(t) \in \mathbb{R}$ is the system's input signal, $H_0(q)$ represents the noise model and $e(t) \in \mathbb{R}$ is a zero-mean white noise signal with variance denoted by σ_e^2. Also, the expansion of $G_0(q)$ and $H_0(q)$ on q^{-1} results in

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

$$H_0(q) = 1 + \sum_{k=1}^{\infty} h_0(k)q^{-k}, \tag{3}$$

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq), and by PETROBRAS S/A.

[1]Emerson C. Boeira is with the Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Rio Grande do Sul, Av. Osvaldo Aranha 103, Porto Alegre, Brazil, emerson.boeira@ufrgs.br

[2]Diego Eckhard is with the Departamento de Matemática Pura e Aplicada, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, Brazil, diegoeck@ufrgs.br

with $H_0(\infty) = 1$ [1], and where $g_0(k)$ and $h_0(k)$ represent the k-th coefficients of process and noise impulse responses.

The model structure considered in the regularized impulse response identification scenario is a FIR model[1] [4], [6]:

$$G(q, \theta) = \sum_{k=1}^{n} g(k) q^{-k}, \tag{4}$$

where $g(k)$ represents the impulse response coefficients of the model and n denotes its order. Such model belongs to a category of linear regression models, where the predicted system's output can be described as

$$\hat{y}(t) = \varphi(t)^T \theta, \tag{5}$$

with $\varphi(t) \in \mathbb{R}^n$ being the regressor vector and $\theta \in \mathbb{R}^n$ the parameter vector:

$$\varphi(t) = \begin{bmatrix} u(t-1) & u(t-2) & \dots & u(t-n) \end{bmatrix}^T, \tag{6}$$

$$\theta = \begin{bmatrix} g(1) & g(2) & \dots & g(n) \end{bmatrix}^T. \tag{7}$$

Most theoretical analysis of the model assumes that it can describe the true transfer function, such that the next assumption is respected.

Assumption 1: The true process' transfer function $G_0(q)$ belongs to the chosen model structure $G(q, \theta)$. This means that there exists θ_0 such that $G(q, \theta_0) = G_0(q)$.

When Assumption 1 is satisfied then the true system presented in (1) can be described by the following linear regression:

$$y(t) = \varphi(t)\theta_0 + v(t), \tag{8}$$

with $\theta_0 = \begin{bmatrix} g_0(1) & g_0(2) & \dots & g_0(n) \end{bmatrix}^T$ and $v(t) = H_0(q)e(t)$

The Regularized Least-Squares (RLS) estimation for the FIR model exposed above is very similar compared to the traditional least-squares method but with the addition of a quadratic regularization term [4], [6]

$$\hat{\theta}_R = \arg\min_{\theta} \; (Y - \Phi\theta)^2 + \theta^T P^{-1} \theta, \tag{9}$$

$$\hat{\theta}_R = \left(P\Phi^T\Phi + I\right)^{-1} \left(P\Phi^T Y\right), \tag{10}$$

where $P \in \mathbb{R}^{n \times n}$ is the regularization matrix and with $\Phi \in \mathbb{R}^{N_t \times n}$ and $Y \in \mathbb{R}^{N_t}$ being written as

$$\Phi = \begin{bmatrix} \varphi(n+1) & \varphi(n+2) & \dots & \varphi(N) \end{bmatrix}^T, \tag{11}$$

$$Y = \begin{bmatrix} y(n+1) & y(n+2) & \dots & y(N) \end{bmatrix}^T. \tag{12}$$

It can be noticed that the vector Y can be written as

$$Y = \Phi\theta_0 + V, \tag{13}$$

with $V = \begin{bmatrix} v(n+1) & v(n+2) & \dots & v(N) \end{bmatrix}^T \in \mathbb{R}^{N_t}$.

As demonstrated in [4], [6], the greatest advantage of the regularized impulse response estimation provided by (10) is that the matrix P balances the variance and the bias error in order to achieve better identification properties, especially concerning the Mean-Square Error (MSE) matrix.

In this work, we will expand the results [4], [6] to include systems with colored output noise. We will present expressions for bias, variance and MSE of the estimate along with an analysis of the optimal regularization matrix. We will also show that there is a close relation between the Regularized Weighted Least-Squares estimate and the Bayesian estimate, such that the Empirical Bayes method can be used to estimate both regularization and weight matrices.

A. MSE error for RLS

To quantify the error between the estimate and the true parameter value the MSE matrix is a reasonable measure [4]:

$$\mathcal{Q}_R(P) = E\left[\left(\hat{\theta}_R - \theta_0\right)\left(\hat{\theta}_R - \theta_0\right)^T\right], \tag{14}$$

which comprehends both the bias and the covariance errors as [4], [6]

$$\mathcal{Q}_R(P) = \mathcal{B}_R(P)\mathcal{B}_R(P)^T + \mathcal{V}_R(P), \tag{15}$$

where $\mathcal{B}_R(P)$ denotes the bias error and $\mathcal{V}_R(P)$ the variance error:

$$\mathcal{B}_R(P) = E[\hat{\theta}_R] - \theta_0, \tag{16}$$

$$\mathcal{V}_R(P) = E\left[\left(\hat{\theta}_R - E[\hat{\theta}_R]\right)\left(\hat{\theta}_R - E[\hat{\theta}_R]\right)^T\right], \tag{17}$$

which are all functions of the regularization matrix P.

For the estimator considered in this paper, presented in (10) and considering that $H_0(q)$ is a rational transfer function, the expected value of $\hat{\theta}_R$ can be computed using the fact that $Y = \Phi\theta_0 + V$ as follows:

$$E[\hat{\theta}_R] = E\left[(PR + I)^{-1} PR\theta_0\right] + E\left[(PR + I)^{-1} PRV\right],$$
$$E[\hat{\theta}_R] = (PR + I)^{-1} PR\theta_0, \tag{18}$$

with $R = \Phi^T\Phi$ and because $E[V] = 0$. From this result, the following expression for the bias error can be achieved:

$$\mathcal{B}_R(P) = -(PR + I)^{-1}\theta_0. \tag{19}$$

The covariance of the referred estimate can be obtained in a similar procedure. It can be seen, for this scenario, that

$$\hat{\theta}_R - E[\hat{\theta}_R] = \left(P\Phi^T\Phi + I\right)^{-1} P\Phi^T Y - \left(P\Phi^T\Phi + I\right)^{-1} P\Phi^T\Phi\theta_0, \tag{20}$$

$$\hat{\theta}_R - E[\hat{\theta}_R] = \left(P\Phi^T\Phi + I\right)^{-1} P\Phi^T V, \tag{21}$$

then, it can be seen from (17) that $\mathcal{V}_R(P)$ is given by

$$\mathcal{V}_R(P) = (PR + I)^{-1}\left(P\Lambda P^T\right)\left(RP^T + I\right)^{-1}, \tag{22}$$

[1]This model could be generalized with the addition of the input delay without loss of generality under the method's properties and analysis.

with

$$\Lambda = \Phi^T \Sigma \Phi, \tag{23}$$

$$\Sigma = E\left[VV^T\right] \tag{24}$$

$$\Sigma = \begin{bmatrix} r_{vv}(0) & r_{vv}(1) & \dots & r_{vv}(N_t-1) \\ r_{vv}(-1) & r_{vv}(0) & \dots & r_{vv}(N_t-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{vv}(-N_t+1) & r_{vv}(-N_t+2) & \dots & r_{vv}(0) \end{bmatrix}, \tag{25}$$

with $r_{vv}(\tau)$ being the autocorrelation function of $v(t)$, i.e.

$$r_{vv}(\tau) = E\left[v(t+\tau)v(t)\right]. \tag{26}$$

Now, from the bias and the covariance errors exposed above, the MSE error matrix can be computed from (15), producing the following result:

$$\mathcal{Q}_R(P) = (PR+I)^{-1}\left(\theta_0\theta_0^T + P\Lambda P^T\right)\left(RP^T+I\right)^{-1}, \tag{27}$$

which is a more generic expression compared with the one achieved in [4], [6] which is specific for white output noise.

B. Optimal regularization matrices

Expression (27) makes it clear that the MSE depends on the regularization matrix P such that an optimal matrix can be computed. We propose to minimize the trace of the MSE matrix (27), such that:

$$P_{0cn} = \arg\min_P \, tr\left[\mathcal{Q}_R(P)\right]. \tag{28}$$

The solution for the problem above is reached through the following matrix equation

$$\frac{\partial tr\left[\mathcal{Q}_R(P)\right]}{\partial P} = 0, \tag{29}$$

which can be computed with matrix calculus:

$$\begin{aligned} \frac{\partial tr\left[\mathcal{Q}_R(P)\right]}{\partial P} = & -RL_1^{-1}\left(\theta_0\theta_0^T + P\Lambda P^T\right)L_2^{-1}L_1^{-1} + \\ & \Lambda P^T L_2^{-1}L_1^{-1} + \Lambda P^T L_1^{-T} L_2^{-T} - \\ & RL_2^{-T}\left(\theta_0\theta_0^T + P\Lambda P^T\right)L_1^{-T}L_2^{-T}, \end{aligned} \tag{30}$$

with $L_1 = (PR+I)$ and $L_2 = (RP^T+I)$, which can be simplified by using the following matrix symmetries:

$$\begin{cases} \left(RP^T+I\right)^{-T} = \left[\left(RP^T+I\right)^T\right]^{-1} = (PR+I)^{-1}, \\ (PR+I)^{-T} = \left[(PR+I)^T\right]^{-1} = \left(RP^T+I\right)^{-1}. \end{cases} \tag{31}$$

to hold:

$$\begin{aligned} \frac{\partial tr\left[\mathcal{Q}_R(P)\right]}{\partial P} = & -2RL_1^{-1}\left(\theta_0\theta_0 + P\Lambda P^T\right)L_2^{-1}L_1^{-1} + \\ & \Lambda P^T L_2^{-1}L_1^{-1}. \end{aligned} \tag{32}$$

Now, using the expression (32) and some degree of matrix algebra, the solution to (29) can be found as

$$P_{0cn} = \theta_0\theta_0^T R\Lambda^{-1}. \tag{33}$$

Observe that if the noise is white ($H_0(q)=1$), then $\Lambda = \sigma_e^2 R$ and

$$P_{0cn} = \frac{\theta_0\theta_0^T}{\sigma_e^2} \tag{34}$$

which is the solution demonstrated in [4], [6] for white output noise, such that (33) is a generalization of the optimal matrix presented there.

Observe that also that P_{0cn} is a semi-definite matrix with only one positive eigenvalue (the other are all equals to zero), and it depends on unknown quantities in practical applications, such as θ_0 and Σ.

III. Regularized weighted least-squares for impulse response estimation

As an alternative to identify the impulse response and establish a connection with the Bayesian perspective for the colored output noise scenario, the Regularized Weighted Least-Squares algorithm must be studied. The RWLS estimate is computed by solving the following optimization problem:

$$\hat{\theta}_{RW} = \arg\min_\theta \, ||(Y-\Phi\theta)||_M^2 + \theta^T P^{-1}\theta, \tag{35}$$

where $||x||_M^2$ denotes the quadratic norm $x^T M x$, with $M \in \mathbb{R}^{N_t\times N_t}$ being the weighting matrix and P the aforementioned regularization matrix. Finally, the solution to the problem above results in the following parameter estimate:

$$\hat{\theta}_{RW} = \left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T MY. \tag{36}$$

A. MSE error for RWLS

Following the same idea exposed in previous sections, it's important to compute the MSE error matrix for the RWLS estimate, in order to further optimize it in some sense with the use of regularization and also to get new insights on P's optimal structure and dependencies. To do so, the bias and covariance errors are, again, used as follows:

$$\mathcal{Q}_{RW}(P) = \mathcal{B}_{RW}(P)\mathcal{B}_{RW}(P)^T + \mathcal{V}_{RW}(P), \tag{37}$$

where

$$\mathcal{B}_{RW}(P) = E[\hat{\theta}_{RW}] - \theta_0, \tag{38}$$

$$\mathcal{V}_{RW}(P) = E\left[\left(\hat{\theta}_{RW} - E[\hat{\theta}_{RW}]\right)\left(\hat{\theta}_{RW} - E[\hat{\theta}_{RW}]\right)^T\right]. \tag{39}$$

Using the same procedure as before, the expected value for $\hat{\theta}_R$ is calculated as

$$\begin{aligned} E[\hat{\theta}_{RW}] = & E\left[\left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T M\Phi\theta_0\right] + \\ & E\left[\left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T M\Phi V\right], \end{aligned} \tag{40}$$

$$E[\hat{\theta}_{RW}] = \left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T M\Phi\theta_0, \tag{41}$$

So, from the above expression for the expected value of the RWLS estimate, the bias can be found as

$$\mathcal{B}_R(P) = -\left(P\Phi^T M\Phi + I\right)^{-1}\theta_0. \tag{42}$$

Also, the covariance of the RWLS estimate is found similarly as before. Firstly, the following difference is taken:

$$\hat{\theta}_{RW} - E[\hat{\theta}_{RW}] = \left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T MY - \left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T M\Phi\theta_0, \quad (43)$$

$$\hat{\theta}_{RW} - E[\hat{\theta}_{RW}] = \left(P\Phi^T M\Phi + I\right)^{-1} P\Phi^T MV, \quad (44)$$

then, it results in the variance error given by:

$$\mathcal{V}_{RW}(P) = \left(P\Phi^T M\Phi + I\right)^{-1} W \left(\Phi^T M^T \Phi P^T + I\right)^{-1}, \quad (45)$$

with $W = P\Phi^T M\Sigma M^T \Phi P^T$, and where, finally, the MSE error is achieved:

$$\mathcal{Q}_{RW}(P) = \left(P\Phi^T M\Phi + I\right)^{-1} \left(\theta_0\theta_0^T + W\right) \left(\Phi^T M^T \Phi P^T + I\right)^{-1}. \quad (46)$$

IV. The Bayesian perspective

The Bayesian perspective of the impulse response estimation is quite different from the traditional system identification perspective, but it's essential to produce ideas for the estimation of the regularization matrix and the use of the Regularized Least-Squares method. In the Bayesian perspective, the true parameter vector θ itself is considered a gaussian distributed random variable, with zero mean and covariance matrix denoted by Π, i.e. [4], [6]:

$$\theta \sim \mathcal{N}(0, \Pi), \quad (47)$$

which differs from the traditional identification perspective, where θ is considered deterministic but unknown. So, if (47) holds and the system's output noise is considered as gaussian and colored, then, the output vector Y and the parameter vector θ can now be described as two jointly gaussian variables as

$$\begin{bmatrix}\theta \\ Y\end{bmatrix} = \left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\Pi & \Pi\Phi^T \\ \Phi\Pi & \Phi\Pi\Phi^T + \Sigma\end{bmatrix}\right). \quad (48)$$

The parameter estimate from this stochastic point of view can be accomplished by maximizing the conditional distribution of θ given that the output Y is measured from the process, which represents the posterior distribution of θ given Y, and that can be obtained with Bayes theorem. For jointly gaussian variables the posterior distribution is well-known [7]:

$$\theta | Y \sim \mathcal{N}\left(\hat{\theta}_B, \Pi_B\right), \quad (49)$$

$$\hat{\theta}_B = \left(\Pi\Phi^T\Sigma^{-1}\Phi + I\right)^{-1} \Pi\Phi^T\Sigma^{-1}Y, \quad (50)$$

$$\Pi_B = \Pi - \Pi\Phi^T\left(\Phi\Pi\Phi^T + \Sigma\right)^{-1}\Phi\Pi, \quad (51)$$

where the Bayesian parameter estimate is denoted by $\hat{\theta}_B$.

Now, comparing (50) and (36) it becomes evident that the regularized weighted leas-squares estimates is equivalent to the Bayesian one with the regularization matrix is chosen as $P = \Pi$ and the weighting matrix chosen as $M = \Sigma^{-1}$ regarding the colored output noise scenario for the impulse response identification.

In the white output noise scenario, addressed in [4], [6], the Bayesian identification perspective produces a direct relationship with the Regularized Least-Squares estimates, where it can be seen that both methodologies are equivalent by choosing $P = \Pi$. However, observing the Bayesian estimate (50) for the colored output noise case, it's not possible to achieve a direct relation between the Regularized Least-Squares estimate (10) and the Bayesian perspective one (50). To do so, we must refer to the Regularized Weighted Least-Squares.

V. Empirical Bayes method for regularization and noise covariance matrices estimation

As demonstrated in [4], the estimation of the regularization matrix P and the covariance matrix Σ from data is performed with the *Empirical Bayes* method [8], observing the marginal distribution of Y from (48), and using a set of hyperparameters $\beta = \begin{bmatrix}\eta^T & \zeta^T\end{bmatrix}^T$:

$$Y \sim \mathcal{N}\left(0, \Phi\Pi(\eta)\Phi^T + \Sigma(\zeta)\right). \quad (52)$$

From this gaussian distribution, a Marginal Likelihood Maximization problem can be formulated to estimate β:

$$\hat{\beta} = \arg\min_{\beta} \left(Y^T\Psi(\beta)^{-1}Y + \ln|\Psi(\beta)|\right), \quad (53)$$

with

$$\Psi(\beta) = \Phi\Pi(\eta)\Phi^T + \Sigma(\zeta). \quad (54)$$

The Marginal Likelihood Maximization expressed above is also fairly similar to the ones proposed on the classical regularization works [4], [6], with the main difference that in our work, the estimation of Σ is held alongside with Π and it plays an important role for the RWLS method.

Furthermore, each matrix Π and Σ has a different interpretation and so, possesses its own parametrization structure. For Π, since it represents the covariance of the system's impulse response coefficients, it should reflect exponential decay and positive correlation among its parameters, for example. So, typical parametrization structures for this matrix are the following [4], [6]:

- Diagonal/Correlated (DC):

$$\Pi_{kj}(\eta) = \lambda\alpha^{(k+j)/2}\rho^{|k-j|}, \quad (55)$$

with $\lambda > 0$, $0 < \alpha < 1$ and $|\rho| < 1$;

- Tuned/Correlated (TC):

$$\Pi_{kj}(\eta) = \lambda\min(\alpha^k, \alpha^j), \quad (56)$$

with $\lambda > 0$, $0 < \alpha < 1$.

On the other hand the matrix Σ is a Toeplitz matrix that represents the covariance of the vector V, as exposed in (25). So, in our work. we choose to identify Σ^{-1} on the Marginal Maximum Likelihood problem, since it can be easily parametrized as:

$$\Sigma^{-1}(\zeta) = S(\zeta)S(\zeta)^T, \quad (57)$$

where

$$S(\zeta)=\begin{bmatrix}\zeta_1 & \zeta_2 & \dots & \zeta_m & 0 & \dots & \dots & 0\\ 0 & \zeta_1 & \zeta_2 & \dots & \zeta_m & 0 & \dots & 0\\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots\\ 0 & \dots & 0 & \zeta_1 & \zeta_2 & \dots & \zeta_m & 0\\ 0 & \dots & \dots & 0 & \zeta_1 & \zeta_2 & \dots & \zeta_m\\ 0 & \dots & \dots & \dots & 0 & \zeta_1 & \zeta_2 & \dots\\ 0 & \dots & \dots & \dots & \dots & 0 & \zeta_1 & \zeta_2\\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & \zeta_1\end{bmatrix}, \quad (58)$$

which comprehend a typical inverse of Σ when $H_0(q)$ is a generic m order IIR filter with no zeros.

Additionally the cost function of the optimization problem (53) is ill-conditioned and has a complex structure which can produce inaccurate results for the hyperparameter estimation [9]. In [9], however, there are some alternatives to compute such cost function efficiently and in this paper some adaptations were made to implement those algorithms for the colored noise scenario addressed here.

VI. Numerical Example

In order to compare our new RWLS algorithm for the colored noise scenario, the RLS proposed for the white noise scenario in [4], [6], and the classical LS method for impulse response estimation when $H(q)\neq 1$, this section demonstrates a numerical example. The system considered for identification here has the following structure:

$$y(t)=\underbrace{\frac{0.1}{q-0.9}}_{G_0(q)}u(t)+\underbrace{\frac{q}{q-0.95}}_{H_0(q)}e(t), \quad (59)$$

where the main objective is to identify the $n=60$ impulse response coefficients of $G_0(q)$. The experimental conditions for the comparison are similar to one case proposed in [4], where $N=500$ data samples are collected in each Monte Carlo run and the Signal-to-Noise Ratio (SNR) is equal to 10, which holds $\sigma_u^2=190$ and $\sigma_w^2=0.0975$. The system was simulated through 1000 Monte Carlo runs and some metrics were evaluated to compare the methods.

The first comparison in this paper regards the bias norm and the trace of the covariance and the MSE matrices achieved by each method on the Monte Carlo simulations. Such results are exhibited in Table I, which compares those metrics for the classical LS method, the RLS method proposed in [4], and our new RWLS algorithm with the TC and DC parametrization. The main aspect that can be seen in Table I is that our new algorithm was able to reduce the trace of the MSE and the covariance matrices in comparison with the RLS method developed in [4].

Another analysis performed to compare each method is the comparison of the impulse responses achieved in each Monte Carlo run of the RLS and the RWLS methods with the DC parametrization, which gives a sample of what happens with the estimated impulse responses for each method. Fig. 1 demonstrate the 1000 estimated impulse responses with the RLS method and Fig. 2 demonstrate the 1000 estimated impulse responses with the RWLS method, both with the

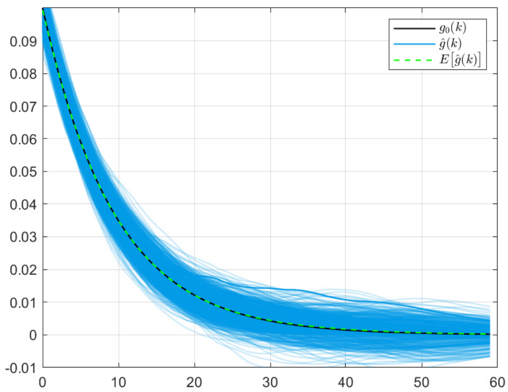

Fig. 1. Comparison of the true impulse response with the ones obtained by the RLS method with DC parametrization and classical hyperparameter estimation.

DC paramatrization structure and its own hyperparameter estimation method.

Comparing both Figures, it can be noticed that the RWLS method proposed here presents better results, since its average impulse response is very similar to the true system's impulse and its covariance is considerable lower than the one achieved with the RLS method.

The last criterion used for comparison in this paper is the distribution of a fit measure between the true impulse response and the estimated ones on the Monte Carlo analysis. This same fit measure was also employed in [4], [6] and it consists on

$$\mathcal{F}=100\left[1-\left(\frac{\sum_{k=0}^{n}|g_0(k)-\hat{g}(k)|^2}{\sum_{k=0}^{n}|g_0(k)-\bar{g}_0(k)|^2}\right)^{1/2}\right], \quad (60)$$

with

$$\bar{g}_0=\frac{1}{n}\sum_{k=0}^{n}g_0(k). \quad (61)$$

Then, a good model, close to the true system's impulse response would present a fit $\mathcal{F}\approx 100$. So, Figure 3 demonstrate the boxplot graphic of the distribution of each fit achieved with all the identification methods compared in this paper. Firstly, it can be noticed, that the lowest average fit is achieved with the traditional LS method. Also, it can be seen that the RWLS method outperform the RLS method in both parametrization structures (TC and DC) employed here, since they average fit are higher.

TABLE I

Bias norm and trace of the covariance and MSE matrices of each impulse response estimation method achieved.

Method		$\|\|\mathcal{B}\|\|_2$	$tr(\mathcal{V})$	$tr(\mathcal{Q})$
LS		0.748×10^{-3}	0.839×10^{-3}	8.40×10^{-4}
RLS	DC	1.25×10^{-3}	0.564×10^{-3}	5.65×10^{-4}
	TC	1.45×10^{-3}	0.638×10^{-3}	6.40×10^{-4}
RWLS	DC	0.610×10^{-3}	0.208×10^{-3}	2.09×10^{-4}
	TC	9.68×10^{-3}	0.398×10^{-3}	4.92×10^{-4}

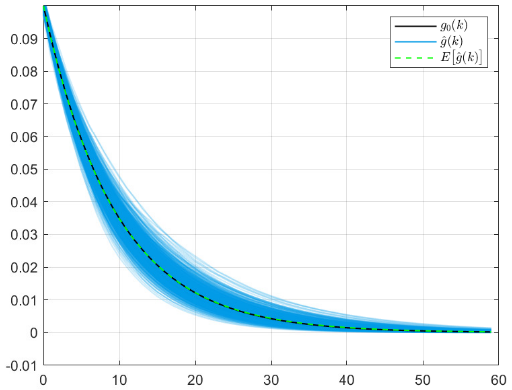

Fig. 2. Comparison of the true impulse response with the ones obtained by the RWLS method with DC parametrization and our hyperparameter estimation.

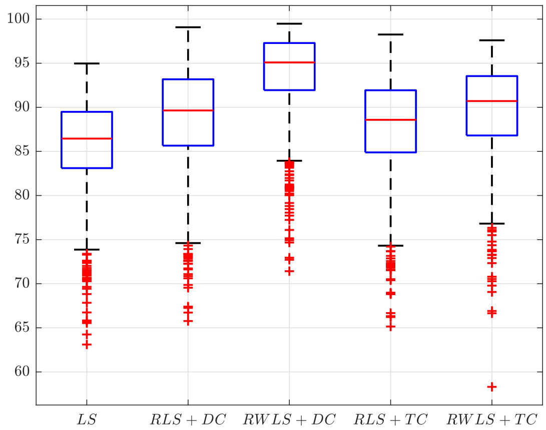

Fig. 3. Comparison of the boxplots obtained for the fit metric for each estimation method. Each column represent a different impulse estimation method.

VII. Concluding Remarks

This paper had demonstrate new insights to the use of regularization to identify the system's impulse response considering the colored output noise scenario. Firstly, it was demonstrated that the optimal regularization matrix is slightly different for the RLS compared to the white noise scenario. Also, we showed that under this output noise condition, the Bayesian perspective has a direct association with the RWLS method, where $P = \Pi$ and $M = \Sigma^{-1}$, and a new algorithm to estimate both P and Σ using this perspective, where the latter has a new specific structure of parametrization.

Finally, the numerical example demonstrates the use of our new methodology, where it can be seen that it has improved the impulse response estimation properties, resulting in a lower trace for the covariance and MSE matrices, compared to the state-of-the-art regularization techniques and the traditional least squares method as well as better mean and variance of the estimated impulse responses and a higher average value for the fit criterion.

References

[1] L. Ljung, *System identification: theory for the user*, 2nd ed. Englewoods Cliff: Prentice Hall, 1999.

[2] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81–93, 2010.

[3] G. Pillonetto, A. Chiuso, and G. De Nicolao, "Prediction error identification of linear systems: a nonparametric gaussian regression approach," *Automatica*, vol. 47, no. 2, pp. 291–305, 2011.

[4] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes - revisited," *Automatica*, vol. 48, no. 8, pp. 1525–1535, 2012.

[5] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *International Journal of Control*, vol. 93, no. 2, pp. 173–180, 2019.

[6] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: a survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs: Prentice-Hall, 1979.

[8] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. New York: Chapman & Hall, 2000.

[9] T. Chen and L. Ljung, "Implementation of algorithms for tuning parameters in regularized least squares problems in system identification," *Automatica*, vol. 49, no. 7, pp. 2213–2220, 2013.