

# BALANCED LEAST SQUARES: ESTIMATION IN LINEAR SYSTEMS WITH NOISY INPUTS AND MULTIPLE OUTPUTS

Javier Vía and Ignacio Santamaría

University of Cantabria, Spain  
{jvia,nacho}@gtas.dicom.unican.es

## ABSTRACT

This paper revisits the linear model with noisy inputs, in which the performance of the total least squares (TLS) method is far from acceptable. Under the assumption of Gaussian noises, the maximum likelihood (ML) estimation of the system response is reformulated as a general balanced least squares (BLS) problem. Unlike TLS, which minimizes the trace of the product between the empirical and inverse theoretical covariance matrices, BLS promotes solutions with similar values of both the empirical and theoretical error covariance matrices. The general BLS problem is reformulated as a semidefinite program with a rank constraint, which can be relaxed in order to obtain polynomial time algorithms. Moreover, we provide new theoretical results regarding the scenarios in which the relaxation is tight, as well as additional insights on the performance and interpretation of BLS. Finally, some simulation results illustrate the satisfactory performance of the proposed method.

**Index Terms**— Balanced Least Squares (BLS), Errors in Variables (EIV), Total Least Squares (TLS), Semidefinite Programming (SDP), Rank Constrained Optimization.

## 1. INTRODUCTION

The linear model is a key ingredient for signal processing. In this paper we revisit the linear model with noisy inputs, which can be seen as a particular instance of the Errors in Variables (EIV) model [1, 2]. Focusing on the estimation of the system parameters in the EIV model, total least squares (TLS) [3–6] is by far the most popular method among the signal processing community. However, the TLS performance is very far from satisfactory, among others, in the scenario considered in this paper.

Interestingly, the linear model with noisy inputs is mathematically equivalent to the scenario with a Gaussian model matrix, which in the case of scalar outputs has been addressed in [7, 8]. However, the general case with vector outputs remains as an open problem. Recently, and focusing on the case with Gaussian noisy inputs and scalar outputs, a new technique called balanced least squares (BLS) has been proposed as the maximum likelihood (ML) estimator of the system response [9], thus rediscovering the results in [7, 8]. This technique promotes solutions with similar empirical and theoretical variances of the residual error, which clearly contrasts with the LS and TLS solutions. In particular, LS exclusively focus on the minimization of the empirical variance, whereas TLS minimizes the ratio between the empirical and theoretical variances of the residual error, thus promoting solutions with small empirical and large theoretical variances.

This paper extends the BLS method to the general case of vector outputs, where the ML estimation problem is reformulated as a rank constrained semidefinite program [10]. Fortunately, the relaxation of the non-convex rank constraint provides satisfactory solutions in most practical situations. Moreover, we provide some interesting insights regarding the performance, interpretation and relationship to the TLS method, as well as a sufficient condition for the tightness (with probability one) of the relaxation.

Throughout the paper we will use boldfaced uppercase letters for matrices and boldfaced lowercase letters for (column) vectors. Given a matrix  $\mathbf{A}$ ,  $\mathbf{A}^T$  and  $|\mathbf{A}|$  denote the transpose and determinant, respectively.  $\mathbf{I}$  and  $\mathbf{0}$  denote the identity and zero matrices of the required dimensions, and  $\mathbf{A} \succeq \mathbf{0}$  means that  $\mathbf{A}$  is symmetric positive semidefinite.  $\lambda_k(\mathbf{A}, \mathbf{B})$  refers to the  $k$ -th largest generalized eigenvalue of the matrix pencil  $(\mathbf{A}, \mathbf{B})$ . That is  $\mathbf{A}\mathbf{v} = \lambda_k\mathbf{B}\mathbf{v}$ , with  $\mathbf{v}$  the  $k$ -th generalized eigenvector. For a symmetric block matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{1,1} & \mathbf{M}_{1,2} \\ \mathbf{M}_{1,2}^T & \mathbf{M}_{2,2} \end{bmatrix}, \quad (1)$$

$\mathbf{S}(\mathbf{M}, \mathbf{M}_{2,2}) = \mathbf{M}_{1,1} - \mathbf{M}_{1,2}\mathbf{M}_{2,2}^{-1}\mathbf{M}_{2,1}$  denotes the Schur complement of  $\mathbf{M}_{2,2}$  [11, 12]. Finally, given two sequences of vectors  $\mathbf{x}_k, \mathbf{y}_k$ , we define the sample (cross-)covariance matrices as  $\mathbf{R}_{\mathbf{x}} = \sum_k \mathbf{x}_k\mathbf{x}_k^T$ ,  $\mathbf{R}_{\mathbf{x},\mathbf{y}} = \sum_k \mathbf{x}_k\mathbf{y}_k^T$ .

## 2. PROBLEM FORMULATION AND PREVIOUS WORKS

This paper focuses on a general linear model in which  $m$  deterministic real<sup>1</sup> inputs  $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathbb{R}^{m \times 1}$  and  $n$  real outputs  $\mathbf{y} = [y_1, \dots, y_n]^T \in \mathbb{R}^{n \times 1}$  are related as

$$\mathbf{y} = \mathbf{H}^T(\mathbf{x} + \mathbf{u}) + \mathbf{n}, \quad (2)$$

where  $\mathbf{H} \in \mathbb{R}^{m \times n}$  represents the deterministic channel response,  $\mathbf{u} \in \mathbb{R}^{m \times 1}$  is the zero-mean input noise with covariance matrix  $\mathbf{C}_{\mathbf{u}} \succ \mathbf{0}$ , and  $\mathbf{n} \in \mathbb{R}^{n \times 1}$  is a zero-mean noise vector with covariance matrix  $\mathbf{C}_{\mathbf{n}} \succ \mathbf{0}$ , and independent of  $\mathbf{u}$ .

Eq. (2) represents a particular case<sup>2</sup> of the structural EIV model [2], which finds application in econometrics, computer vision, signal processing, and communications. For instance, in wireless communications, eq. (2) represents the training sequence transmitted from a multiantenna amplify and forward relay and observed by a multiantenna receiver [13]; the observations of an eavesdropper in a discriminatory channel estimation scenario [14, 15]; or the received signals in a channel estimation system based on superimposed training [16].

<sup>1</sup>All the results in this paper can be easily extended to the complex case.

<sup>2</sup>The general EIV model includes a reference channel providing some additional information about  $\mathbf{u}$ .

This work has been supported by the Spanish Government, Ministerio de Ciencia e Innovación, under project RACHEL (TEC2013-47141-C4-3-R).

A classical problem in EIV consists in the estimation, from a set of  $N \geq m$  independent observations  $\mathbf{y}_k$  ( $k = 0, \dots, N-1$ ), of the deterministic parameter  $\mathbf{H}$  in eq. (2). Unfortunately, the estimation generally results in difficult non-convex optimization problems, which translates into local minima for typical approaches such as the Expectation Maximization (EM) algorithm [8, 17]. Moreover, the performance of classical approaches such as TLS [5] is far from satisfactory in several important cases [7, 8], including the model considered in this paper.

## 2.1. Total Least Squares (TLS)

For the signal processing community, TLS emerges as the main tool for estimation in linear systems with imperfect knowledge of the input signals, which justifies several recent research efforts related to the TLS method [18–21]. However, the original formulation of the TLS method is based on the following model

$$\mathbf{y} = \mathbf{H}^T \mathbf{x} + \mathbf{n}, \quad (3)$$

$$\mathbf{z} = \mathbf{x} + \mathbf{u}, \quad (4)$$

where the deterministic input  $\mathbf{x}$  is now unknown and only observed through the reference channel  $\mathbf{z}$ . Obviously, the observation model in (3)-(4) differs from eq. (2) because the errors  $\mathbf{u}$  are not inputs of the channel  $\mathbf{H}$ .

The TLS method is originally obtained as the ML estimate of  $\mathbf{H}$  in model (3)-(4) under the assumption of Gaussian noises  $\mathbf{n}$  and  $\mathbf{u}$ . However, focusing on the model in eq. (2), TLS can be recovered as an hybrid ML-MAP estimator of  $\mathbf{H}$  and  $\mathbf{U} = [\mathbf{u}_0, \dots, \mathbf{u}_{N-1}]$ , which can be easily verified by writing the posterior pdf of  $\mathbf{U}$

$$p(\mathbf{U}|\mathbf{Y}; \mathbf{H}) \propto e^{-\frac{1}{2} \left\| [\mathbf{Y} - (\mathbf{X} + \mathbf{U})\mathbf{H}] \mathbf{C}_n^{-\frac{1}{2}} \right\|^2 - \frac{1}{2} \left\| \mathbf{U} \mathbf{C}_u^{-\frac{1}{2}} \right\|^2},$$

where  $\mathbf{Y}$  and  $\mathbf{X}$  are defined analogously to  $\mathbf{U}$ . Thus, the estimation problem reduces to

$$\underset{\mathbf{H}, \mathbf{U}}{\text{minimize}} \quad \left\| [\mathbf{Y} - (\mathbf{X} + \mathbf{U})\mathbf{H}] \mathbf{C}_n^{-\frac{1}{2}} \right\|^2 + \left\| \mathbf{U} \mathbf{C}_u^{-\frac{1}{2}} \right\|^2,$$

or equivalently

$$\begin{aligned} & \underset{\mathbf{H}, \Delta_X, \Delta_Y}{\text{minimize}} \quad \left\| \begin{bmatrix} \Delta_Y \mathbf{C}_n^{-\frac{1}{2}} & \Delta_X \mathbf{C}_u^{-\frac{1}{2}} \end{bmatrix} \right\| \\ & \text{subject to} \quad \mathbf{Y} + \Delta_Y = (\mathbf{X} + \Delta_X)\mathbf{H}, \end{aligned}$$

which is the most common formulation of the TLS method [5]. Specifically, the TLS solution is obtained from the following generalized eigenvalue (GEV) problem

$$\mathbf{R}\tilde{\mathbf{H}} = \tilde{\mathbf{C}}\tilde{\mathbf{H}}\Lambda^{\text{TLS}}, \quad (5)$$

where  $\Lambda^{\text{TLS}} = \text{diag}([\lambda_{m+1}(\mathbf{R}, \mathbf{C}), \dots, \lambda_{m+n}(\mathbf{R}, \mathbf{C})])$  contains the  $n$  smallest generalized eigenvalues, the columns of  $\tilde{\mathbf{H}}$  are the corresponding generalized eigenvectors, and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_y & -\mathbf{R}_{x,y}^T \\ -\mathbf{R}_{x,y} & \mathbf{R}_x \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_u \end{bmatrix}.$$

Finally, if we partition  $\tilde{\mathbf{H}}$  as  $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1^T \quad \tilde{\mathbf{H}}_2^T]^T$ , with  $\tilde{\mathbf{H}}_1 \in \mathbb{R}^{n \times n}$ , the estimate of  $\mathbf{H}$  is recovered as<sup>3</sup>  $\hat{\mathbf{H}}^{\text{TLS}} = \tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_1^{-1}$ . However, the TLS estimates under the model in eq. (2) are far from satisfactory, mainly due to the reason in the following remark.

<sup>3</sup>Obviously, this requires an invertible  $\tilde{\mathbf{H}}_1^{-1}$ , and the cases in which this condition is not satisfied are referred to as non-generic TLS problems [22].

**Remark 1 (Unbalanced TLS Solutions)** From the GEV in eq. (5), TLS can be seen as the solution of the following optimization problem

$$\underset{\tilde{\mathbf{H}}}{\text{minimize}} \quad \text{trace} \left[ \left( \tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}} \right) \left( \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right)^{-1} \right]. \quad (6)$$

That is, TLS amounts to minimizing the *ratio* between the empirical covariance ( $\tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}}$ ) of the residual error ( $\mathbf{y} - \mathbf{H}^T \mathbf{x}$ ) and its theoretical value ( $\tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}}$ ). However, this fact can be seen as an important drawback of the TLS approach because it seems reasonable to think that a good estimate of  $\mathbf{H}$  should result into similar values of the empirical and theoretical covariance matrices. Instead of that, TLS promotes solutions with small empirical and large theoretical covariance matrices, which might result in overfitting. This fact, which can be also interpreted as a direct consequence of estimating a large number of nuisance variables  $\mathbf{U} \in \mathbb{R}^{N \times m}$ , when we are only interested in  $\mathbf{H} \in \mathbb{R}^{m \times n}$ , provides an alternative explanation for the TLS inconsistency<sup>4</sup> under the model in eq. (2).

## 3. BALANCED LEAST SQUARES (BLS)

Due to the problems pointed out in the previous section, here we will follow a true ML approach to estimate  $\mathbf{H}$  under the model in eq. (2). In particular, the maximization of the log-likelihood function yields the following non-convex optimization problem

$$\underset{\tilde{\mathbf{H}} = [\mathbf{I} \quad \mathbf{H}^T]^T}{\text{minimize}} \quad \text{trace} \left[ \left( \tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}} \right) \left( \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right)^{-1} \right] + \log \left| \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right|. \quad (7)$$

**Remark 2 (Balanced Least Squares)** As can be seen, the proposed approach results in an optimization problem similar to that in eq. (6), but introducing a logarithmic regularization term which promotes similar values of the empirical and theoretical residual covariance matrices. In particular, if we were free to choose the theoretical covariance  $\tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}}$  for a fixed value of the empirical covariance  $\tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}}$ , the solution would be given by  $\tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} = \tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}}$ . Although the coupling of both terms by means of  $\mathbf{H}$  yields different values of the covariance matrices at the solution of the optimization problem in (7), it is clear that the proposed approach will result into a better balance than TLS.

### 3.1. Relaxation to a Convex Optimization Problem

In order to solve the optimization problem in (7), we recover the partition of  $\tilde{\mathbf{H}}$  as  $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1^T \quad \tilde{\mathbf{H}}_2^T]^T$  and rewrite (7) as

$$\begin{aligned} & \underset{\tilde{\mathbf{H}} = [\mathbf{I} \quad \mathbf{H}^T]^T}{\text{minimize}} \quad \text{trace} \left[ \left( \tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}} \right) \left( \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right)^{-1} \right] \\ & \quad + \log \left| \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right| - \log \left| \tilde{\mathbf{H}}_1^T \tilde{\mathbf{H}}_1 \right|. \end{aligned}$$

Moreover, taking into account the invariance of the cost function under linear transformations of the form  $\tilde{\mathbf{H}}\mathbf{A}$  (with an invertible  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ), we can focus on the equivalent problem

$$\begin{aligned} & \underset{\tilde{\mathbf{H}}}{\text{minimize}} \quad \text{trace} \left( \tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}} \right) - \log \left| \tilde{\mathbf{H}}_1^T \tilde{\mathbf{H}}_1 \right| \\ & \text{subject to} \quad \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} = \mathbf{I}. \end{aligned} \quad (8)$$

<sup>4</sup>The straightforward proof of the inconsistency of the TLS solution is omitted here due to the lack of space.

Although the above problem is still non-convex, it is already in a form suitable for its reformulation as a semidefinite program (SDP). Thus, defining  $\mathbf{M} = \tilde{\mathbf{H}}\tilde{\mathbf{H}}^T \in \mathbb{R}^{(m+n) \times (m+n)}$  and introducing a partition similar to that in eq. (1) with  $\mathbf{M}_{1,1} \in \mathbb{R}^{n \times n}$ , we can rewrite (8) as

$$\begin{aligned} & \underset{\mathbf{M} \succeq \mathbf{0}}{\text{minimize}} && \text{trace}(\mathbf{R}\mathbf{M}) - \log |\mathbf{M}_{1,1}| \\ & \text{subject to} && \text{trace}(\mathbf{C}\mathbf{M}) = n \\ & && \mathbf{I} \succeq \mathbf{C}^{\frac{1}{2}}\mathbf{M}\mathbf{C}^{\frac{1}{2}} \\ & && \text{rank}(\mathbf{M}) = n, \end{aligned} \quad (9)$$

and since the non-convexity of (9) is solely due to the last constraint, we propose to solve the relaxed problem

$$\begin{aligned} & \underset{\mathbf{M} \succeq \mathbf{0}}{\text{minimize}} && \text{trace}(\mathbf{R}\mathbf{M}) - \log |\mathbf{M}_{1,1}| \\ & \text{subject to} && \text{trace}(\mathbf{C}\mathbf{M}) = n \\ & && \mathbf{I} \succeq \mathbf{C}^{\frac{1}{2}}\mathbf{M}\mathbf{C}^{\frac{1}{2}}. \end{aligned} \quad (10)$$

### 3.2. Main Results

The relaxed problem in (10) is convex and satisfies the Slater's constraint qualification [10]. Therefore, its solution can be efficiently found by means of standard convex optimization tools [23,24]. However, the optimal<sup>5</sup>  $\mathbf{M}^{\text{BLS}}$  does not necessarily satisfy the constraint  $\text{rank}(\mathbf{M}^{\text{BLS}}) = n$ , and therefore it does not provide the solution of the ML problem in (7). In this subsection we summarize our main findings in the form of five lemmas, whose proofs are omitted due to the lack of space, and will be included in a future journal paper [25].

**Lemma 1 (Consistency)** In the asymptotic regime ( $N \rightarrow \infty$ ), the solution of (10) is given by

$$\begin{aligned} \mathbf{M}^{\text{BLS}} &= \tilde{\mathbf{H}} \left( \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right)^{-1} \tilde{\mathbf{H}}^T, \\ \lambda^{\text{BLS}} &= 0, \\ \Phi^{\text{BLS}} &= \mathbf{0} \end{aligned}$$

where  $\lambda^{\text{BLS}}$  and  $\Phi^{\text{BLS}}$  represent the optimal values of the dual variables for the constraints  $\text{trace}(\mathbf{C}\mathbf{M}) = n$  and  $\mathbf{I} \succeq \mathbf{C}^{\frac{1}{2}}\mathbf{M}\mathbf{C}^{\frac{1}{2}}$ , and  $\tilde{\mathbf{H}} = [\mathbf{I} \quad \mathbf{H}^T]^T$  contains the true value of the channel  $\mathbf{H}$ . In other words, the estimator in (10) is consistent.

**Lemma 2 (BLS and Regularized LS)** If  $\text{rank}(\mathbf{M}^{\text{BLS}}) = n$ , the ML estimate of  $\mathbf{H}$  is recovered as

$$\hat{\mathbf{H}}^{\text{BLS}} = \left( \mathbf{R}_{\mathbf{x}} - \lambda^{\text{BLS}} \mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{u}}^{\frac{1}{2}} \Phi_{2,2}^{\text{BLS}} \mathbf{C}_{\mathbf{u}}^{\frac{1}{2}} \right)^{-1} \left( \mathbf{R}_{\mathbf{x},\mathbf{y}} + \mathbf{C}_{\mathbf{u}}^{\frac{1}{2}} \Phi_{2,1}^{\text{BLS}} \mathbf{C}_{\mathbf{n}}^{\frac{1}{2}} \right),$$

where  $\lambda^{\text{BLS}}$  and  $\Phi^{\text{BLS}}$  are obtained from the solution of the dual problem

$$\begin{aligned} & \underset{\lambda, \Phi \succeq \mathbf{0}, \Psi \succeq \mathbf{0}}{\text{maximize}} && \log |\mathbf{S}(\Psi, \Psi_{2,2})| + n\lambda - \text{trace}(\Phi) \\ & \text{subject to} && \Psi = \mathbf{R} - \lambda \mathbf{C} + \mathbf{C}^{\frac{1}{2}} \Phi \mathbf{C}^{\frac{1}{2}}. \end{aligned} \quad (11)$$

<sup>5</sup>From now on we will use the superindex  $(\cdot)^{\text{BLS}}$  to denote the optimal values of the primal or dual variables of (10).

**Remark 3 (BLS Soft Regularization)** The previous lemma allows us to see the BLS estimate as the solution of a regularized LS problem. This interpretation becomes obvious in the case of scalar outputs ( $n = 1$ ), in which we have  $\Phi^{\text{BLS}} = \mathbf{0}$  and  $\hat{\mathbf{h}}^{\text{BLS}} = (\mathbf{R}_{\mathbf{x}} - \lambda^{\text{BLS}} \mathbf{C}_{\mathbf{u}})^{-1} \mathbf{R}_{\mathbf{x},\mathbf{y}}$ . Interestingly, the TLS solution admits a similar interpretation with a regularized matrix  $\Psi^{\text{TLS}} = \mathbf{R} - \lambda^{\text{TLS}} \mathbf{C} + \mathbf{C}^{\frac{1}{2}} \Phi^{\text{TLS}} \mathbf{C}^{\frac{1}{2}}$  satisfying  $\mathbf{S}(\Psi^{\text{TLS}}, \Psi_{2,2}^{\text{TLS}}) = \mathbf{0}$ . Therefore, since  $\mathbf{S}(\Psi, \Psi_{2,2})$  directly appears in the objective function of the dual problem in (11), we can conclude that the BLS regularization is less aggressive than its TLS counterpart, which in the case of scalar outputs reduces to  $\lambda^{\text{BLS}} < \lambda^{\text{TLS}}$  [8,9].

**Lemma 3 (Structure of  $\Phi^{\text{BLS}}$ )** At the optimal solution of (10) we have  $\text{rank}(\Phi^{\text{BLS}}) < n$ . Moreover, if  $\text{rank}(\mathbf{M}^{\text{BLS}}) = n$  then

$$\mathbf{M}^{\text{BLS}} = \tilde{\mathbf{H}} \mathbf{M}_{1,1}^{\text{BLS}} \tilde{\mathbf{H}}^T, \quad \Phi^{\text{BLS}} = \mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{H}} \mathbf{B} \tilde{\mathbf{H}}^T \mathbf{C}^{\frac{1}{2}},$$

with  $\mathbf{B} \in \mathbb{R}^{n \times n}$  a positive semidefinite matrix and

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{I}_n \\ \hat{\mathbf{H}}^{\text{BLS}} \end{bmatrix}.$$

**Remark 4 (Redundancy of  $\mathbf{I} \succeq \mathbf{C}^{\frac{1}{2}}\mathbf{M}\mathbf{C}^{\frac{1}{2}}$ )** The previous lemma generalizes, for  $n > 1$ , the fact that the constraint  $\mathbf{I} \succeq \mathbf{C}^{\frac{1}{2}}\mathbf{M}\mathbf{C}^{\frac{1}{2}}$  is redundant for  $n = 1$ . In this particular case, BLS admits simpler interpretations, and the dual problem in (11) reduces to a one dimensional convex optimization problem, which can be easily solved by means of the Golden Section Algorithm [26], or directly by bisection on the derivative of the objective function [9].

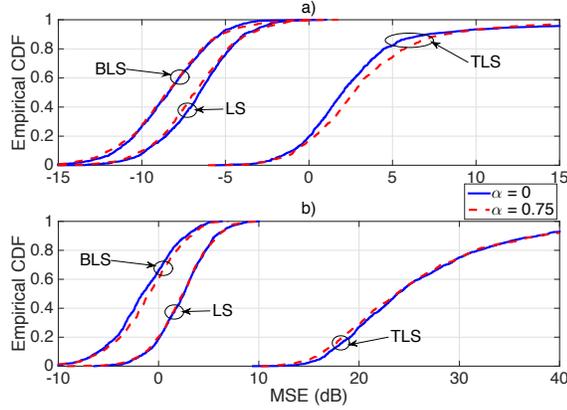
**Lemma 4 (Maximum Rank)** Let us write  $\text{rank}(\mathbf{M}^{\text{BLS}}) = n + r$  and denote the number of unit eigenvalues of  $\mathbf{C}^{\frac{1}{2}}\mathbf{M}^{\text{BLS}}\mathbf{C}^{\frac{1}{2}}$  as  $u$ . Then, there exist an optimal solution of (10) with  $u \geq r$ , which obviously implies  $\text{rank}(\mathbf{M}^{\text{BLS}}) < 2n$ .

**Remark 5 (Purification Method)** Lemma 4 guarantees the existence of an optimal solution  $\mathbf{M}^{\text{BLS}}$  with an upper bounded rank. This solution can be found by means of a purification technique similar to those in [27,28], and which will be reported in a future journal paper [25]. However, although this general upper bound ensures the solution of the ML problem in the case with scalar outputs ( $n = 1$ ), it seems to suggest that the solution of (10) needs to be *purified*. Fortunately, the next lemma provides additional conditions guaranteeing solutions with  $\text{rank}(\mathbf{M}) = n$ , and allows us to get rid (with probability one) of the purification stage.

**Lemma 5 (Solutions with Probability One)** If data is drawn from a continuous distribution, then  $\Pr(\text{rank}(\mathbf{M}^{\text{BLS}}) < 2n) = 1$ . Moreover, if the multiplicity  $\mu$  of  $\lambda_m(\mathbf{R}_{\mathbf{x}}, \mathbf{C}_{\mathbf{u}})$  satisfies  $\mu \geq n$ , then

$$\Pr(\text{rank}(\mathbf{M}^{\text{BLS}}) = n) = 1.$$

**Remark 6 (Practical Design Choices)** Firstly, Lemma 5 proves that the purification approach is not necessary in practice. Secondly, the condition  $\mu \geq n$  can be easily satisfied in many applications (with  $m \geq n$ ) in which matrix  $\mathbf{X}$  is a design parameter. For instance, in channel estimation applications, one sensible choice of the pilot matrix is based on  $\mathbf{R}_{\mathbf{x}} \propto \mathbf{C}_{\mathbf{u}}$ . Moreover, in the case with  $m < n$ , and if the direct solution of (10) provides a high rank  $\mathbf{M}^{\text{BLS}}$ , one sensible approach consists in reducing the dimensionality of the observations by means of principal component analysis (PCA) [29], which would reduce the estimation problem to the case with  $n = m$ .



**Fig. 1.** Empirical CDFs of the normalized MSEs in the estimate of  $\mathbf{H}$ .  $m = n = 4$ ,  $N = 20$ . a)  $\eta^2 = 1$ . b)  $\eta^2 = 10$ .

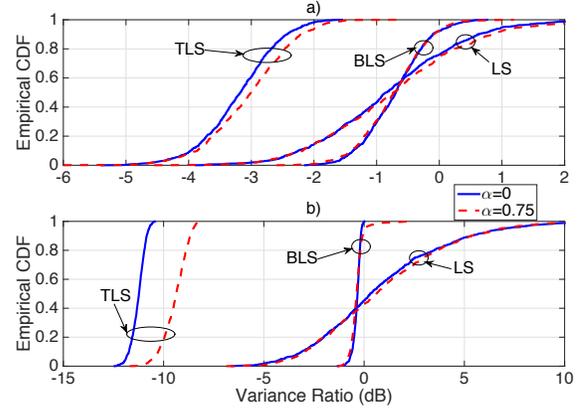
#### 4. NUMERICAL RESULTS

This section evaluates the performance of the BLS, TLS and LS methods by means of some numerical examples. All the results are based on 1000 Monte Carlo simulations of a system with  $m = 4$  inputs,  $n = 4$  outputs,  $N = 20$  samples, and random parameters  $\mathbf{H} \in \mathbb{R}^{4 \times 4}$  with i.i.d. zero-mean and unit variance Gaussian elements. The input noise is Gaussian with zero mean and covariance  $\mathbf{C}_u = \eta^2 \text{diag}(\mathbf{c})$ , where the entries of  $\mathbf{c} \in \mathbb{R}^{4 \times 1}$  are linearly spaced between  $1 - \alpha$  and  $1 + \alpha$ . That is,  $\eta^2$  controls the total variance of the input noise, whereas  $\alpha$  controls the *asymmetry* of the input distribution. The output noise is zero-mean Gaussian with covariance  $\mathbf{C}_n = \mathbf{I}$ , and the deterministic input sequence is selected to satisfy  $\mathbf{R}_x = \mathbf{I}$ . Finally, the implementation of the BLS method is based on the solution of (10) by means of the CVX toolbox [23]. In the very few cases in which a high-rank solution is obtained, the (suboptimal) solution reduces to a rank reduction of  $\hat{\mathbf{H}}^{\text{BLS}}$  by means of PCA [29].

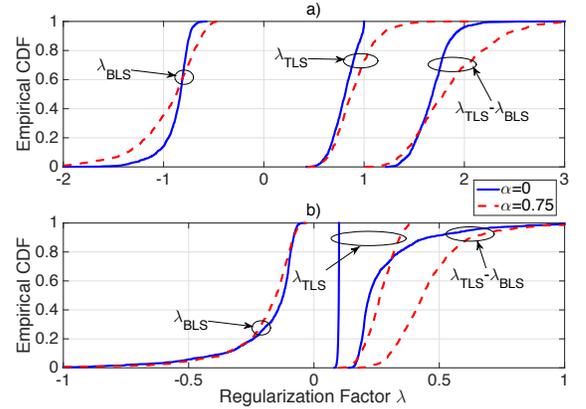
Fig. 1 shows, for two different values of  $\eta^2$  and  $\alpha$ , the empirical cumulative distribution function (CDF) of the normalized (by the number of entries  $mn$ ) mean square error (MSE) on the estimate of  $\mathbf{H}$ . The first obvious conclusion is that the performance of the TLS method is unacceptable, which is due to the reasons outlined in Remark 1, and is also reflected on the bad conditioning of the matrix  $\tilde{\mathbf{H}}_1$  to be inverted (see the comment at the end of Subsection 2.1). Moreover, we can see that BLS outperforms the conventional LS technique, and the gap between them increases with the input noise variance. Finally, from the results in Fig. 1 we can also conclude that the LS and BLS performances are very robust to the value of  $\alpha$ , which in the BLS case is a direct consequence of the fact that  $\hat{\mathbf{H}}^{\text{BLS}}$  satisfies the rank constraint with a very high probability, even when Lemma 5 does not apply ( $\alpha \neq 0$ ).

Fig. 2 shows the empirical CDFs of the Variance Ratio, which we define as  $\text{trace} \left[ \left( \tilde{\mathbf{H}}^T \mathbf{R} \tilde{\mathbf{H}} \right) \left( \tilde{\mathbf{H}}^T \mathbf{C} \tilde{\mathbf{H}} \right)^{-1} \right]$ . As expected (see Remark 2), BLS provides values much closer to one than those of the alternative techniques. Additionally, we can see that the TLS method, which yields very small and far from reasonable values, is much more sensitive to the values of  $\alpha$  than the BLS and LS methods.

Finally, and in order to clarify the implications of Lemma 2 and Remark 3, Fig. 3 shows the empirical CDFs of the *regularization*



**Fig. 2.** Empirical CDFs of the Variance Ratio for the LS, TLS and BLS methods.  $m = n = 4$ ,  $N = 20$ . a)  $\eta^2 = 1$ . b)  $\eta^2 = 10$ .



**Fig. 3.** Empirical CDFs of the TLS and BLS regularization factors  $\lambda$ .  $m = n = 4$ ,  $N = 20$ . a)  $\eta^2 = 1$ . b)  $\eta^2 = 10$ .

factors  $\lambda$  obtained from the TLS and BLS solutions, as well as the empirical CDF of their difference. Clearly, TLS yields much more aggressive regularization factors (larger values of  $\lambda$ ), which can be seen as another explanation for its poor performance.

#### 5. CONCLUSIONS

This paper has extended the BLS method to the case of vector outputs. Although the general formulation results into rank-constrained semidefinite programming problems, we have seen that the relaxation of the non-convex rank constraint is tight in most of the practical cases. Moreover, we have presented several theoretical results providing sufficient conditions for the tightness of the relaxation, as well as guaranteeing the consistency of BLS, and clarifying its relationship with the regularized LS and TLS methods. The satisfactory performance of BLS has been illustrated by means of several simulations, which also corroborate that BLS provides a reasonable balance between the empirical and theoretical covariance matrices of the residual error.

## 6. REFERENCES

- [1] R. J. Adcock, "Note on the method of least squares," *The Analyst*, vol. 4, no. 6, pp. 183–184, Nov 1877.
- [2] J. Gillard, "An overview of linear structural models in errors in variables regression," *REVSTAT–Statistical Journal*, vol. 8, no. 1, pp. 57–80, 2010.
- [3] G. H. Golub, "Some modified matrix eigenvalue problems," *SIAM Review*, vol. 15, no. 2, pp. 318–334, 1973.
- [4] G. H. Golub and C. F. van Loan, "An analysis of the total least squares problem," *SIAM Journal on Numerical Analysis*, vol. 17, no. 6, pp. 883–893, 1980.
- [5] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal processing*, vol. 87, no. 10, pp. 2283–2302, 2007.
- [6] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, Prentice Hall Upper Saddle River, NJ, 2000.
- [7] A. Wiesel, Y.C. Eldar, and A. Beck, "Maximum likelihood estimation in linear models with a Gaussian model matrix," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 292–295, May 2006.
- [8] A. Wiesel, Y.C. Eldar, and A. Yeredor, "Linear regression with Gaussian model uncertainty: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2194–2205, June 2008.
- [9] J. Vía and I. Santamaría, "Balanced least squares: Linear model estimation with noisy inputs," in *IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2015)*, Cancún, Mexico, Dec. 2015.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, March 2004.
- [11] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 2nd edition, 1989.
- [12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
- [13] C. S. Patel and G. L. Stuber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2348–2356, June 2007.
- [14] T. H. Chang, W. C. Chiang, Y. P. Hong, and C.Y. Chi, "Training sequence design for discriminatory channel estimation in wireless MIMO systems," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6223–6237, 2010.
- [15] C.W. Huang, T. H. Chang, X. Zhou, and Y. P. Hong, "Two-way training for discriminatory channel estimation in wireless MIMO systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2724–2738, 2013.
- [16] S. He, J.K. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 3007–3021, June 2007.
- [17] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov 1996.
- [18] R. Arablouei, S. Werner, and K. Dogancay, "Analysis of the gradient-descent total least-squares adaptive filtering algorithm," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1256–1264, March 2014.
- [19] S. Rhode, K. Usevich, I. Markovsky, and F. Gauterin, "A recursive restricted total least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5652–5662, Nov 2014.
- [20] R. Arablouei, K. Dogancay, and S. Werner, "Recursive total least-squares algorithm based on inverse power method and dichotomous coordinate-descent iterations," *IEEE Transactions on Signal Processing*, vol. 63, no. 8, pp. 1941–1949, April 2015.
- [21] S. Huang and C. Li, "Distributed sparse total least-squares over networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2986–2998, June 2015.
- [22] S. Van Huffel and J. Vandewalle, "Analysis and solution of the nongeneric total least squares problem," *SIAM journal on matrix analysis and applications*, vol. 9, no. 3, pp. 360–372, 1988.
- [23] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," <http://cvxr.com/cvx>, 2014.
- [24] Y. Labit, Dimitri Peaucelle, and D. Henrion, "SEDUMI INTERFACE 1.02: a tool for solving LMI problems with SEDUMI," in *IEEE International Symposium on Computer Aided Control System Design*, Sept 2002, pp. 272–277.
- [25] J. Vía and I. Santamaría, "Balanced least squares: ML estimation in the Gaussian linear model with noisy inputs," in preparation, 2015.
- [26] J. Kiefer, "Sequential minimax search for a maximum," *Proceedings of the American Mathematical Society*, vol. 4, no. 3, pp. 502–506, 1953.
- [27] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Mathematics of operations research*, vol. 23, no. 2, pp. 339–358, 1998.
- [28] Y. Huang and D.P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 664–678, Feb 2010.
- [29] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley & Sons, Inc., New York, NY, USA, 1996.