Balanced Least Squares:
Linear Model Estimation with Noisy Inputs

Javier Vía and Ignacio Santamaría
University of Cantabria. Spain
{jvia,nacho}@gtas.dicom.unican.es

Abstract—This paper focuses on a linear model with noisy inputs in which the performance of the conventional Total Least Squares (TLS) approach is (maybe surprisingly) far from satisfactory. Under the typical Gaussian assumption, we obtain the maximum likelihood (ML) estimator of the system response. This estimator promotes a reasonable balance between the empirical and theoretical variances of the residual errors, which suggests the name of Balanced Least Squares (BLS). The solution of the associated optimization problem is based on its reformulation as a rank constrained semidefinite program (SDP), for which we show that the relaxation is tight with probability one. Both TLS and BLS can be seen as regularized LS estimators, but the (possibly negative) regularization in BLS is softer than its TLS counterpart, which avoids the inconsistency of TLS in our particular model.

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

I. INTRODUCTION: ERRORS IN VARIABLES AND TOTAL LEAST SQUARES

Let us consider a general linear model with m real[1] inputs $\mathbf{x} = [x_1, \ldots, x_m]^T \in \mathbb{R}^{m \times 1}$ and one output y related as

$$y = (\mathbf{x} + \mathbf{u})^T \mathbf{h} + n, \qquad (1)$$

where $\mathbf{x}$ is known, $\mathbf{h} \in \mathbb{R}^{m \times 1}$ is deterministic unknown, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{C_u})$ represents the input noise, and $n \sim \mathcal{N}(0, \sigma^2)$ is the measurement noise, which is assumed to be independent of $\mathbf{u}$. We also assume that both n and $\mathbf{u}$ are temporally white. This model can be seen as a particular case of the structural Errors In Variables (EIV) model [1], and it is of particular importance in signal processing, communications, and econometrics. For instance, in wireless communications, it represents the training sequence transmitted from a multiantenna amplify and forward relay and observed by a single antenna receiver [2], or the observations of an eavesdropper in a discriminatory channel estimation scenario [3], [4].

A classical problem in EIV consists in the estimation of the unknown parameter vector $\mathbf{h}$ from a set of N independent observations $\mathbf{y} = [y_1, \ldots, y_N]^T$. Unfortunately, the estimation of $\mathbf{h}$ results in difficult non-convex optimization problems, which translates into local minima problems for typical approaches such as the Expectation Maximization (EM) algorithm [5]. A well known approach to parameter estimation in EIV models consists in the application of the Total Least Squares (TLS)

method [6]–[9]. In particular, for the model in (1), the posterior pdf of $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]^T$ can be written as

$$p(\mathbf{U}|\mathbf{y}; \mathbf{h}) \propto e^{-\frac{\|\mathbf{y} - (\mathbf{X} + \mathbf{U})\mathbf{h}\|^2}{2\sigma^2} - \frac{1}{2}\left\| \mathbf{U}\mathbf{C_u}^{-\frac{1}{2}} \right\|^2},$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T$. Therefore, an hybrid[2] maximum likelihood (ML) - maximum a posteriori (MAP) approach for the joint estimation of $\mathbf{h}$ (ML) and $\mathbf{U}$ (MAP) reduces to

$$\underset{\mathbf{U}, \mathbf{h}}{\text{minimize}} \quad \frac{\|\mathbf{y} - (\mathbf{X} + \mathbf{U})\mathbf{h}\|^2}{\sigma^2} + \left\| \mathbf{U}\mathbf{C_u}^{-\frac{1}{2}} \right\|^2, \qquad (2)$$

whose solution is given by the TLS algorithm [8], [9]. Specifically, the TLS solution can be obtained from the following generalized eigenvalue problem (GEV)

$$\mathbf{R}\tilde{\mathbf{h}} = \lambda_{\text{TLS}} \mathbf{C}\tilde{\mathbf{h}}, \qquad (3)$$

where λ_{TLS} is the smallest generalized eigenvalue,

$$\mathbf{R} = \begin{bmatrix} r_y & -\mathbf{r}_{\mathbf{x}y}^T \\ -\mathbf{r}_{\mathbf{x}y} & \mathbf{R_x} \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C_u} \end{bmatrix}, \qquad (4)$$

and

$$r_y = \frac{\mathbf{y}^T\mathbf{y}}{N}, \qquad \mathbf{r}_{\mathbf{x}y} = \frac{\mathbf{X}^T\mathbf{y}}{N}, \qquad \mathbf{R_x} = \frac{\mathbf{X}^T\mathbf{X}}{N}, \qquad (5)$$

can be seen as sample correlation matrices. Once obtained the minor generalized eigenvector $\tilde{\mathbf{h}} = [\tilde{h}_1, \ldots, \tilde{h}_{m+1}]^T$, the estimate of $\mathbf{h}$ is recovered as $\hat{\mathbf{h}} = [\tilde{h}_2, \ldots, \tilde{h}_{m+1}]^T / \tilde{h}_1$.

At this point, we need to note some important facts related to the model in (1), including two important drawbacks of the TLS approach which motivate the research on alternative estimation approaches.

Remark 1: In order to identify $\mathbf{h}$, we need a tall full column rank matrix $\mathbf{X}$, or equivalently, a positive definite matrix $\mathbf{R_x}$, which implies $N \geq m$. Otherwise, the smallest generalized eigenvector in (3) is $\lambda_{\text{TLS}} = 0$, and the associated eigenvector $\tilde{\mathbf{h}}$ has its first component equal to zero, which translates into a non-generic TLS problem [8], [10]. Therefore, from now on, we will assume $\mathbf{R_x} \succ \mathbf{0}$.

Remark 2: It is well known [9] that the TLS estimate can be written as $\hat{\mathbf{h}} = (\mathbf{R_x} - \lambda_{\text{TLS}} \mathbf{C_u})^{-1} \mathbf{r}_{\mathbf{x}y}$ with $\lambda_{\text{TLS}} \geq 0$. That is, the TLS estimate can be seen as the result of a regularized

[1]All the results in this paper can be easily extended to the complex case.

[2]Note than in other related models, the TLS approach can be simply formulated as the ML estimator. A typical scenario with a satisfactory TLS performance is that with $y = \mathbf{x}^T\mathbf{h} + n$, where $\mathbf{x}$ is unknown and we have access to the reference vector $\mathbf{z} = \mathbf{x} + \mathbf{u}$.

least squares (LS) problem with a negative regularization factor.

Remark 3: In the assymptotic regime ($N \to \infty$) we have $\mathbf{r}_{\mathbf{x}y} = \mathbf{R}_{\mathbf{x}}\mathbf{h}$, and the TLS estimate is given by $\hat{\mathbf{h}} = (\mathbf{R}_{\mathbf{x}} - \lambda_{\text{TLS}}\mathbf{C}_{\mathbf{u}})^{-1}\mathbf{R}_{\mathbf{x}}\mathbf{h}$. Moreover, assuming $\mathbf{C} \succ \mathbf{0}$ it is easy to prove that $\lambda_{\text{TLS}} > 0$, and therefore the TLS estimate $\hat{\mathbf{h}}$ is not consistent.

Remark 4: The GEV in (3) allows us to see the TLS estimate as the solution of the following optimization problem

$$\underset{\mathbf{h}}{\text{minimize}} \quad \frac{\frac{1}{N}\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2}{\sigma^2 + \mathbf{h}^T\mathbf{C}_{\mathbf{u}}\mathbf{h}}. \tag{6}$$

That is, TLS aims to minimize the ratio between the empirical residual variance $\frac{1}{N}\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2$ and its theoretical value $\sigma^2 + \mathbf{h}^T\mathbf{C}_{\mathbf{u}}\mathbf{h}$. However, this fact can be seen as a drawback of the TLS approach for the model considered in this paper. In particular, it seems reasonable to think that a good estimate of $\mathbf{h}$ should result into similar empirical and theoretical residual variances. Instead of that, TLS promotes solutions with small empirical and large theoretical variances, which might result in overfitting. This 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 $\mathbf{h} \in \mathbb{R}^{m \times 1}$.

II. BALANCED LEAST SQUARES (BLS)

Due to the problems pointed out in the previous section, we follow true ML approach to estimate $\mathbf{h}$. In particular, the maximization of the log-likelihood function yields the following non-convex optimization problem

$$\underset{\mathbf{h}}{\text{minimize}} \quad \frac{\frac{1}{N}\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2}{\sigma^2 + \mathbf{h}^T\mathbf{C}_{\mathbf{u}}\mathbf{h}} + \log\left(\sigma^2 + \mathbf{h}^T\mathbf{C}_{\mathbf{u}}\mathbf{h}\right). \tag{7}$$

Remark 5: As can be seen, the proposed approach introduces a logarithmic regularization term which promotes similar values of the empirical and theoretical residual variances. In particular, if we were free to choose the theoretical variance for a fixed value of the empirical variance, the solution would be given by $\sigma^2 + \mathbf{h}^T\mathbf{C}_{\mathbf{u}}\mathbf{h} = \frac{1}{N}\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2$. Although the coupling of both terms by means of $\mathbf{h}$ results into different values of the variances at the solution of the optimization problem in (7), it is clear that the proposed approach will result into a better balance than TLS.

A. Reformulation as a Convex Optimization Problem

In order to solve the optimization problem in (7), we start by rewriting it as

$$\underset{\tilde{\mathbf{h}}}{\text{minimize}} \quad \frac{\tilde{\mathbf{h}}^T\mathbf{R}\tilde{\mathbf{h}}}{\tilde{\mathbf{h}}^T\mathbf{C}\tilde{\mathbf{h}}} + \log\left(\frac{\tilde{\mathbf{h}}^T\mathbf{C}\tilde{\mathbf{h}}}{\tilde{h}_1^2}\right), \tag{8}$$

where $\tilde{\mathbf{h}} = [\tilde{h}_1, \dots, \tilde{h}_{m+1}]^T = c[1, \mathbf{h}^T]^T$ and c is an arbitrary scaling factor. Moreover, taking into account the invariance of the cost function in (8) to scalings of $\tilde{\mathbf{h}}$, we can focus on the equivalent problem

$$\begin{aligned} \underset{\tilde{\mathbf{h}}}{\text{minimize}} \quad & \tilde{\mathbf{h}}^T\mathbf{R}\tilde{\mathbf{h}} - \log\left(\tilde{h}_1^2\right) \\ \text{subject to} \quad & \tilde{\mathbf{h}}^T\mathbf{C}\tilde{\mathbf{h}} = 1. \end{aligned} \tag{9}$$

Althoug the above problem is still non-convex, it is already in a form suitable for its reformulation as a semidefinite program (SDP). Thus, defining

$$\mathbf{M} = \tilde{\mathbf{h}}\tilde{\mathbf{h}}^T = \begin{bmatrix} \mathbf{M}_{1,1} & \mathbf{M}_{2,1}^T \\ \mathbf{M}_{2,1} & \mathbf{M}_{2,2} \end{bmatrix}, \tag{10}$$

with $\mathbf{M}_{1,1} \in \mathbb{R}^{1 \times 1}$, $\mathbf{M}_{2,1} \in \mathbb{R}^{m \times 1}$, and $\mathbf{M}_{2,2} \in \mathbb{R}^{m \times m}$, we can rewrite (9) as

$$\begin{aligned} \underset{\mathbf{M} \succeq \mathbf{0}}{\text{minimize}} \quad & \text{trace}(\mathbf{R}\mathbf{M}) - \log\left(\mathbf{M}_{1,1}\right) \\ \text{subject to} \quad & \text{trace}(\mathbf{C}\mathbf{M}) = 1 \\ & \text{rank}(\mathbf{M}) = 1, \end{aligned} \tag{11}$$

and since the only non-convexity is due to the last constraint, we propose to solve the relaxed problem

$$\begin{aligned} \underset{\mathbf{M} \succeq \mathbf{0}}{\text{minimize}} \quad & \text{trace}(\mathbf{R}\mathbf{M}) - \log\left(\mathbf{M}_{1,1}\right) \\ \text{subject to} \quad & \text{trace}(\mathbf{C}\mathbf{M}) = 1. \end{aligned} \tag{12}$$

B. Solution of the BLS Problem

The optimization problem in (12) is convex and satisfies the Slater's constraint qualification [11]. Therefore, strong duality holds and we can obtain its solution by means of standard convex optimization tools such as CVX [12]. Moreover, the following lemma shows that the relaxation from (11) to (12) is tight.

Lemma 1: The convex optimization problem in (12) has (at least) one rank-one global solution, which provides the solution of the ML estimation problem in (7).

Proof: The proof reduces to consider a global solution $(\mathbf{M}^*)$ not necessarily with rank one. Obviously, the solutions of

$$\begin{aligned} \underset{\mathbf{M} \succeq \mathbf{0}}{\text{minimize}} \quad & \text{trace}(\mathbf{R}\mathbf{M}) \\ \text{subject to} \quad & \text{trace}(\mathbf{C}\mathbf{M}) = 1 \\ & \mathbf{M}_{1,1} = \mathbf{M}_{1,1}^* \end{aligned} \tag{13}$$

are also solutions of the problem in (12). However, the above problem is a real linear SDP with only two linear constraints, which admits at least one rank-one solution [11]. ∎

Although the previous lemma shows that the relaxation is tight, it does not provide a specific algorithm for finding the solution of (7), and it seems to suggest that in general one needs a *purification* technique following the lines in [11], [13]. In order to shed some light on these facts, we will solve the dual problem of (12). Let us start by writing the Lagrangian as

$$\begin{aligned} \mathcal{L}(\mathbf{M}; \lambda, \mathbf{\Theta}) = {} & \text{trace}(\mathbf{R}\mathbf{M}) - \log\left(\mathbf{M}_{1,1}\right) \\ & + \lambda\left[\text{trace}(\mathbf{C}\mathbf{M}) - 1\right] - \text{trace}(\mathbf{\Theta}\mathbf{M}), \end{aligned} \tag{14}$$

whose gradient is given by

$$\nabla_{\mathbf{M}}\mathcal{L}(\mathbf{M}; \lambda, \mathbf{\Theta}) = \mathbf{R} - \begin{bmatrix} \mathbf{M}_{1,1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \lambda\mathbf{C} - \mathbf{\Theta}. \tag{15}$$

Thus, equating the gradient to zero yields

$$\mathbf{M}_{1,1} = (r_y + \lambda\sigma^2 - \mathbf{\Theta}_{1,1})^{-1}, \tag{16}$$

$$\mathbf{\Theta}_{2,1} = -\mathbf{r}_{\mathbf{x}y}, \tag{17}$$

$$\mathbf{\Theta}_{2,2} = \mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}}, \tag{18}$$

where the partition for $\boldsymbol{\Theta}$ is defined analogously to (10). This allows us to write the dual problem as

$$\begin{aligned} \underset{\lambda, \boldsymbol{\Theta}_{1,1}}{\text{maximize}} \quad & \log\left(r_y + \lambda\sigma^2 - \boldsymbol{\Theta}_{1,1}\right) - \lambda \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{\Theta}_{1,1} & -\mathbf{r}_{\mathbf{x}y}^T \\ -\mathbf{r}_{\mathbf{x}y} & \mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \quad (19)$$

Here, it is important to note that the semidefinite constraint requires a null projection of $\mathbf{r}_{\mathbf{x}y}$ onto the null space of $\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}}$. Nevertheless, if $\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}}$ is rank deficient, the projection of $\mathbf{r}_{\mathbf{x}y}$ (a random vector from a continuous distribution) onto its null space will be non-null with probability one. Therefore, we can claim that the solution of (19) satisfies $\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}} \succ \mathbf{0}$ with probability one. Moreover, taking into account that the objective function decreases with $\boldsymbol{\Theta}_{1,1}$ we have $\boldsymbol{\Theta}_{1,1} = \mathbf{r}_{\mathbf{x}y}^T (\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}})^{-1} \mathbf{r}_{\mathbf{x}y}$, which allows us to rewrite the dual problem as

$$\begin{aligned} \underset{\lambda}{\text{maximize}} \quad & \log\left(r_y + \lambda\sigma^2 - \mathbf{r}_{\mathbf{x}y}^T (\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}})^{-1} \mathbf{r}_{\mathbf{x}y}\right) - \lambda \\ \text{subject to} \quad & \mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}} \succ \mathbf{0}, \end{aligned} \quad (20)$$

or equivalently

$$\underset{\lambda \geq \lambda_{\min}}{\text{maximize}} \quad \log\left(r_y + \lambda\sigma^2 - \sum_{k=1}^m \frac{w_k^2}{1 + \lambda\gamma_k}\right) - \lambda \quad (21)$$

where $\lambda_{\min}$ is the minimum value of λ satisfying $\lambda \geq -1/\gamma_1$ and $r_y + \lambda\sigma^2 - \sum_{k=1}^m \frac{w_k^2}{1+\lambda\gamma_k} \geq 0$, $\mathbf{w} = [w_1, \ldots, w_m]^T = \mathbf{V}^T \mathbf{R}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{r}_{\mathbf{x}y}$, and

$$\mathbf{R}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{C}_{\mathbf{u}} \mathbf{R}_{\mathbf{x}}^{-\frac{1}{2}} = \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T \quad (22)$$

denotes the eigenvalue decomposition with the eigenvalues $\boldsymbol{\Gamma} = \text{diag}([\gamma_1, \ldots, \gamma_m])$ in decreasing order.

Since (21) is a one-dimensional concave optimization problem, it can be easily solved by means of the Golden Section Algorithm [14], or bisection in the derivative of the objective function. This last approach is summarized in Algorithm 1.

Finally, from the complementary slackness [11] condition $\boldsymbol{\Theta}\mathbf{M} = \mathbf{0}$, we get

$$\hat{\mathbf{h}} = (\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}})^{-1} \mathbf{r}_{\mathbf{x}y}. \quad (23)$$

Moreover, the combination of (15) and $\boldsymbol{\Theta}\mathbf{M} = \mathbf{0}$ yields

$$\text{trace}(\mathbf{R}\mathbf{M}) - 1 + \lambda = 0, \quad (24)$$

which provides the upper bound $\lambda \leq 1 - \lambda_{\text{TLS}} \leq 1$ used in Algorithm 1.

C. Further Comments

Let us conclude this section with two important observations.

Remark 6: Both TLS and BLS can be seen as regularized LS approaches with different regularization factors. In particular, while the TLS regularization factor is always negative, the factor λ of the BLS can be positive or negative. More specifically, from (3) it is easy to see that the TLS solution satisfies

$$r_y - \lambda_{\text{TLS}}\sigma^2 - \mathbf{r}_{\mathbf{x}y}^T (\mathbf{R}_{\mathbf{x}} - \lambda_{\text{TLS}}\mathbf{C}_{\mathbf{u}})^{-1} \mathbf{r}_{\mathbf{x}y} = 0, \quad (25)$$

Algorithm 1 Balanced Least Squares Algorithm.

Input: Observations r_y, $\mathbf{r}_{\mathbf{x}y}$, parameters $\mathbf{R}_{\mathbf{x}}$, σ^2, $\mathbf{C}_{\mathbf{u}}$, and precision μ.
Output: BLS estimate $\hat{\mathbf{h}}$ and dual variable λ.
Initialize: Obtain $\mathbf{w} \neq \mathbf{0}$ and eigenvalues $\gamma_1, \ldots, \gamma_m$. Set $[\lambda_{\min}, \lambda_{\max} = 1]$ and current value $\lambda = (\lambda_{\min} + \lambda_{\max})/2$.
while $\lambda_{\max} - \lambda_{\min} > \mu$ **do**
    if $(1 - \lambda)\sigma^2 - r_y + \sum_{k=1}^m \frac{w_k^2[1+(\lambda+1)\gamma_k]}{(1+\lambda\gamma_k)^2} > 0$ **then**
        Set $\lambda_{\min} = \lambda$.
    else
        Set $\lambda_{\max} = \lambda$.
    end if
    Set current value $\lambda = (\lambda_{\min} + \lambda_{\max})/2$.
end while
Set the final estimate $\hat{\mathbf{h}} = (\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}})^{-1} \mathbf{r}_{\mathbf{x}y}$.

---

whereas the BLS solution maximizes

$$\log\left(r_y + \lambda\sigma^2 - \mathbf{r}_{\mathbf{x}y}^T (\mathbf{R}_{\mathbf{x}} + \lambda\mathbf{C}_{\mathbf{u}})^{-1} \mathbf{r}_{\mathbf{x}y}\right) - \lambda. \quad (26)$$

That is, BLS never cancels the term inside the logarithm, which is precisely what TLS does. Since this term increases with λ, we can easily conclude that $\lambda > -\lambda_{\text{TLS}}$. In other words, the (possibly) negative regularization in BLS is always *less aggressive* than its TLS counterpart.

Remark 7: Finally, it is easy to prove that, unlike TLS, the BLS estimate is consistent. In order to prove this, we can relax the constraint $\text{trace}(\mathbf{C}\mathbf{M})$ in (12), which is equivalent to fixing $\lambda = 0$. Thus, following the lines in the previous section we can see that, for $N \to \infty$,

$$\boldsymbol{\Theta} = \begin{bmatrix} \mathbf{h}^T \mathbf{R}_{\mathbf{x}} \mathbf{h} & -\mathbf{h}^T \mathbf{R}_{\mathbf{x}} \\ \mathbf{R}_{\mathbf{x}} \mathbf{h} & \mathbf{R}_{\mathbf{x}} \end{bmatrix}, \quad (27)$$

$$\mathbf{M} = \begin{bmatrix} 1 \\ \mathbf{h} \end{bmatrix} (\mathbf{h}^T \mathbf{C}_{\mathbf{u}} \mathbf{h} + \sigma^2)^{-1} \begin{bmatrix} 1 \\ \mathbf{h} \end{bmatrix}^T, \quad (28)$$

which obviously satisfies the constraint $\text{trace}(\mathbf{C}\mathbf{M}) = 1$. Therefore, for $N \to \infty$ we have $\lambda \to 0$, and BLS is consistent and asymptotically equivalent to the conventional LS.

III. NUMERICAL RESULTS

In this section, the performance of BLS is compared with that of TLS and LS by means of numerical examples. Our results are based on the average of 10^5 independent realizations of an experiment with $N = 25$ observations of the output of a system with $m = 10$ inputs, $\mathbf{h} = \mathbf{1}$ (the all-ones vector), an orthogonal matrix $\mathbf{X}$ with $\mathbf{R}_{\mathbf{x}} = \frac{1}{m}\mathbf{I}$, and independent zero-mean Gaussian noises with $\sigma^2 = 1$, and $\mathbf{C}_{\mathbf{u}} = \frac{\eta^2}{m}\mathbf{I}$, where η^2 controls the variance of the input noise.

Figure 1 shows the average value of the regularization factors employed by the BLS and TLS algorithms, which are shown as a function of the input variance noise in dB scale ($20\log_{10}(\eta)$). As can be seen, there exists a big difference in the values of the regularization factor, and as previously pointed out, $\lambda > -\lambda_{\text{TLS}}$. Note also that, although the difference in the regularization factor seems to decrease with η^2, the final performance is more sensitive to λ for stronger input noises (remember that the overall regularization term is $\lambda\mathbf{R}_{\mathbf{u}}$).
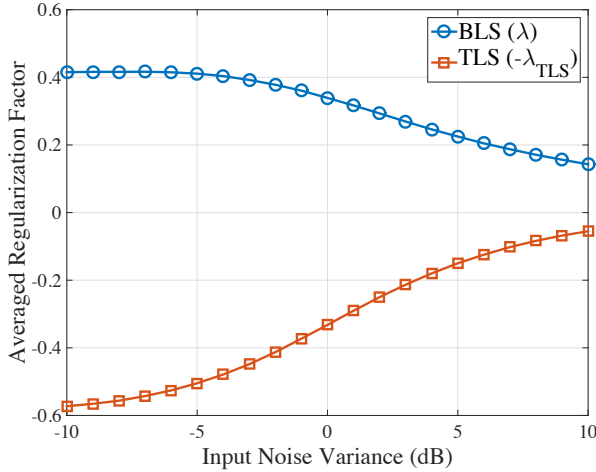
Fig. 1. Average regularization factor for the BLS (λ) and TLS ($-\lambda_{\text{TLS}}$) algorithms. $m = 10$, $N = 25$, $\sigma^2 = 1$.
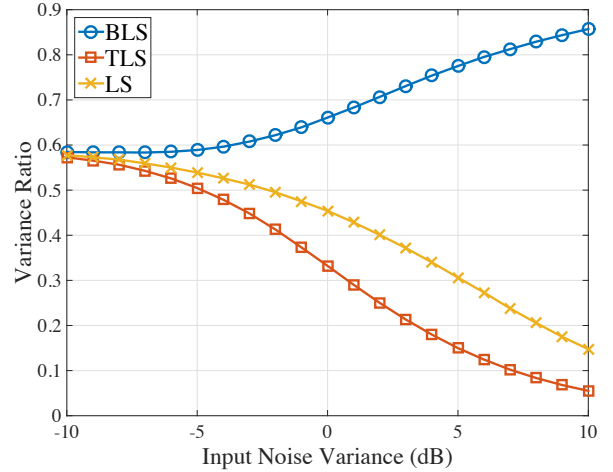

Fig. 2. Ratio between the empirical and theoretical values of the residual error variance. $m = 10$, $N = 25$, $\sigma^2 = 1$.

The previous observation is illustrated in Figures 2 and 3. In particular, Figure 2 shows the averaged variance ratio $\frac{\frac{1}{N}\|\mathbf{y}-\mathbf{Xh}\|^2}{\sigma^2+\mathbf{h}^T\mathbf{C_u}\mathbf{h}}$ for the BLS, TLS, and LS algorithms, where we can see that, for large input noise variances, the BLS algorithm exploits the information related to the expected balance of the residual error variances, whereas both TLS and LS suffer from overfitting. This fact is corroborated by the mean square error (MSE) in the estimation of $\mathbf{h}$, which is illustrated in Figure 3.

IV. CONCLUSIONS

Balanced Least Squares (BLS) is a new estimation algorithm for a simple linear model with input noise, in which the performance of TLS is far from satisfactory. BLS provides a reasonable balance between the empirical and theoretical variances of the residual error and, although it can be used for any kind of data, it has been derived as the ML estimate in the case of Gaussian noises. Ongoing work includes the extension of BLS to the case of multivariate outputs, as well as the estimation of other parameters such as the covariance of the input noise.

Fig. 3. Mean Square Error (MSE) in the estimate of $\mathbf{h}$. BLS is much more robust to input noise than LS and TLS. $m = 10$, $N = 25$, $\sigma^2 = 1$.

REFERENCES

[1] J. Gillard, "An overview of linear structural models in errors in variables regression," *REVSTAT–Statistical Journal*, vol. 8, no. 1, pp. 57–80, 2010.

[2] C. Patel and G. 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.

[3] T.-H. Chang, W.-C. Chiang, Y. 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.

[4] C.-W. Huang, T.-H. Chang, X. Zhou, and Y.-W. 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.

[5] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov 1996.

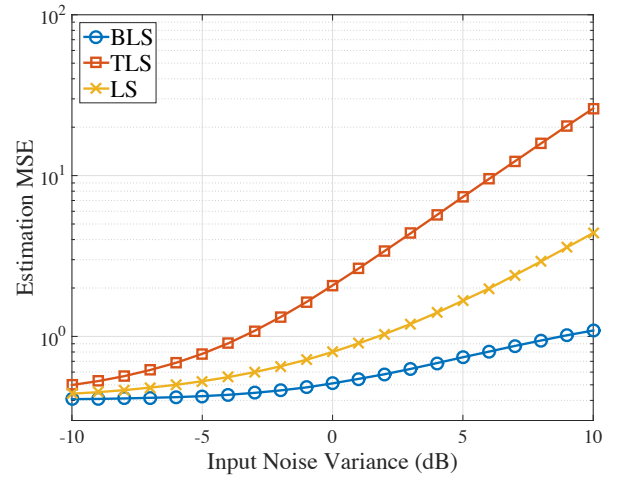[6] G. H. Golub, "Some modified matrix eigenvalue problems," *SIAM Review*, vol. 15, no. 2, pp. 318–334, 1973.

[7] 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.

[8] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal processing*, vol. 87, no. 10, pp. 2283–2302, 2007.

[9] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. Prentice Hall Upper Saddle River, NJ, 2000.

[10] S. Van Huffel and J. Vandewalle, "Analysis and solution of the non-generic total least squares problem," *SIAM journal on matrix analysis and applications*, vol. 9, no. 3, pp. 360–372, 1988.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004.

[12] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Dec. 2010.

[13] 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.

[14] J. Kiefer, "Sequential minimax search for a maximum," *Proceedings of the American Mathematical Society*, vol. 4, no. 3, pp. pp. 502–506, 1953.