IMPROVED LINEAR LEAST SQUARES ESTIMATION USING BOUNDED DATA UNCERTAINTY

Tarig Ballal[1], Tareq Y. Al-Naffouri[1, 2]

[1]Electrical Engineering Department, King Abdullah University of Science & Technology (KAUST)
[2]Electrical Engineering Department, King Fahd University of Petroleum & Minerals

ABSTRACT

This paper addresses the problem of linear *least squares* (LS) estimation of a vector $\boldsymbol{x}$ from linearly related observations. In spite of being unbiased, the original LS estimator suffers from high mean squared error, especially at low signal-to-noise ratios. The *mean squared error* (MSE) of the LS estimator can be improved by introducing some form of regularization based on certain constraints. We propose an improved LS (ILS) estimator that approximately minimizes the MSE, without imposing any constraints. To achieve this, we allow for perturbation in the measurement matrix. Then we utilize a *bounded data uncertainty* (BDU) framework to derive a simple iterative procedure to estimate the regularization parameter. Numerical results demonstrate that the proposed BDU-ILS estimator is superior to the original LS estimator, and it converges to the best linear estimator, the *linear-minimum-mean-squared error estimator* (LMMSE), when the elements of $\boldsymbol{x}$ are statistically white.

Index Terms— linear estimation, least squares, regularization, mean squared error, bounded data uncertainty.

1. INTRODUCTION

This paper addresses the linear estimation problem of a vector quantity $\boldsymbol{x}$ from an observation vector

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times M}$ is a *known* linear transformation matrix and $\boldsymbol{v}$ is white noise with unknown variance. The vector $\boldsymbol{x}$ is assumed to be stochastic with unknown distribution. The elements of $\boldsymbol{x}$ are assumed to be independent but not necessarily identically distributed. Such problem arises in may areas of science and engineering including communications, economics, signal processing, seismology, and control [1].

Owing to the lack of prior information on $\boldsymbol{x}$, usually, a *least squares* (LS) approach is used. Ordinary LS estimation attempts to minimize the squared error $||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2$, where $||.||_2$ is the l_2 norm [2]. Contrary to ordinary LS estimation, the objective in a typical estimation scenario is to minimize *mean squared error* (MSE) between $\boldsymbol{x}$ and its estimate, $\hat{\boldsymbol{x}}$. In general, this goal cannot be achieved by applying the LS approach. However, the attractiveness of the LS approach stems from its agnostic nature to the probabilistic structure of the data and its ease of implementation [2].

Throughout the years, several methods have been proposed to improve the MSE performance of the LS estimator. Among the alternatives are the regularization known in the statistical literature as the ridge estimator [3], the shrunken estimator [4], the covariance shaping LS estimator [5], and the weighted LS [6]. A common feature in all these methods is the need for some prior information in a form of a *constraint*. Depending on the nature of this constraint, the solution of the LS problem takes one of various forms. In this paper, we are particularly interested in the simple regularized LS form given by [3, 7]

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{A} + \gamma\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{y}, \tag{2}$$

where γ is a real positive value, $\boldsymbol{I}$ is the identity matrix of appropriate dimension, and $(.)^T$ is the matrix transpose operation. For $\gamma = 0$, (2) gives the *ordinary* LS estimator, which is unbiased. For $\gamma > 0$, we have a biased estimator [3].

It is proven in [3] that there always exists a positive value γ such that the estimator (2) offers lower MSE than the ordinary LS estimator. However, ways of finding a suitable γ that minimizes the MSE are not precisely known. The goal of this paper is to find an approach through which we can find a value of γ that approximately minimizes the MSE. The difficulty arises from the fact that this value of γ is always dependent on the unknown vector $\boldsymbol{x}$, as will be shown subsequently.

In this paper, we focus on the case where $N \geq M$ and the problem has a unique solution. More specifically, the emphasis is on the case where N is not much larger than M, which is the case where the ordinary LS estimator suffers most. We adopt a *bounded data uncertainty* (BDU) approach [8]. In the BDU model, the measurement matrix is not precisely known due to an unknown perturbation. In other words, the actual measurement matrix is $\boldsymbol{A} + \boldsymbol{\Delta}$; we know $\boldsymbol{A}$ and a bound on the 2-induced norm of the perturbation $\boldsymbol{\Delta}$. This formulation leads to the same LS solution (2), where γ is estimated by solving a *secular equation*. The solution (2), under BDU assumptions, targets minimizing a maximum error criterion, which is not necessarily minimiz-

ing the MSE. Although the BDU approach is proposed as a robust estimation method, in this paper, we employ the technique as a regularization tool that allows us to obtain a value of γ to minimize the MSE of the regularized LS estimator (2). It will be shown that by using a number of approximations, the required value of γ can be obtained without any prior conditions on $\boldsymbol{x}$, or the perturbation $\boldsymbol{\Delta}$. In the case where $\boldsymbol{x}$ has a *white* statistic, it will be shown that the improved LS solution converges to the superior *linear minimum mean squared error* (LMMSE) estimator.

2. THE BASIC BDU APPROACH

The data model for the BDU approach is [8]

$$\boldsymbol{y} = (\boldsymbol{A} + \boldsymbol{\Delta})\boldsymbol{x} + \boldsymbol{v}, \tag{3}$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times M}$ is full column rank, and $\boldsymbol{\Delta} \in \mathbb{R}^{N\times M}$ is a perturbation or uncertainty in $\boldsymbol{A}$. To find an estimate of $\boldsymbol{x}$, a min-max approach is pursued in [8]

$$\begin{aligned} &\min_{x} \max ||(\boldsymbol{A} + \boldsymbol{\Delta})\boldsymbol{x} - (\boldsymbol{y} - \boldsymbol{v})||_2 \\ &\text{subject to: } ||\boldsymbol{\Delta}||_2 \leq \eta, ||\boldsymbol{v}_d||_2 \leq \eta_v, \end{aligned} \tag{4}$$

where $||.||_2$ denotes the l_2 norm in the case of a vector, and the 2-induced norm in the case of a matrix, η is a known upper bound on the 2-induced norm of $\boldsymbol{\Delta}$, and η_v is the upper bound on the l_2 norm of $\boldsymbol{v}_d$ that turns out to be irrelevant and disappears from the final solution. The solution of (4) is shown to be exactly the regularized LS solution (2). Using the *singular value decomposition* (SVD) $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T$, the parameter γ is obtained by solving the secular equation [8]

$$\boldsymbol{y}^T\boldsymbol{U}(\boldsymbol{\Lambda}^2 - \eta^2\boldsymbol{I})(\boldsymbol{\Lambda}^2 + \gamma\boldsymbol{I})^{-2}\boldsymbol{U}^T\boldsymbol{y} = 0. \tag{5}$$

It is shown in [8] that, under certain conditions on η, (2) is the unique solution of the problem in (4).

In this work, we utilize the BDU model for regularization purpose. We want to solve the problem of estimating $\boldsymbol{x}$ based on the model (1). As a form of regularization, we seek to find a perturbation of the singular values of $\boldsymbol{A}$ that improves the LS solution. Equivalently, we want to perturb $\boldsymbol{A}$ into a matrix $\boldsymbol{A} + \boldsymbol{\Delta}$ and use the latter matrix to reduce the MSE. Since, we have no knowlege of $\boldsymbol{\Delta}$, the problem is equivalent to that of finding $\boldsymbol{x}$ based on the BDU model (3).

The BDU estimator suffers from the shortcomings shared with other LS estimators. First, the method is not guaranteed to minimize the MSE, as it is optimizing the error in a min-max sense. Second, the method is dependent on the choice of the parameter η, whose value determines the quality of the estimator. In the robust estimation problem, η is used to quantify the uncertainty in $\boldsymbol{A}$ which is dictated by the real world. In the regularization case, since we have no specific bound on the perturbation we want to add, η needs to be judiciously controlled such that the MSE is minimized. In the following sections, we show how the BDU approach can be used as a regularizer to approximately minimize the MSE, without having to make any assumptions on the parameter η.

3. MINIMIMIZING THE MSE OF THE REGULARIZED LS ESTIMATOR

Starting from the LS estimator (2) for the problem in (1), we define the overall MSE of the estimator as

$$\text{MSE} = \text{tr}\left\{\mathbb{E}\left[(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^T\right]\right\}, \tag{6}$$

$\mathbb{E}$ and tr(.) are the expectation and matrix trace operators, respectively. By substituting (1) in (2), substituting the result in (6) and manipulating, we obtain

$$\begin{aligned} &\text{MSE} = \\ &\sigma_v^2\text{tr}\left[\boldsymbol{B}_\gamma^{-1}\boldsymbol{B}_o\boldsymbol{B}_\gamma^{-1}\right] + \gamma^2\text{tr}\left[\boldsymbol{B}_\gamma^{-1}\boldsymbol{C}_x\boldsymbol{B}_\gamma^{-1}\right], \end{aligned} \tag{7}$$

where $\boldsymbol{B}_\gamma \triangleq \left(\boldsymbol{A}^T\boldsymbol{A} + \gamma\boldsymbol{I}\right)$, $\boldsymbol{B}_o = \boldsymbol{B}_\gamma|_{\gamma=0}$, and $\boldsymbol{C}_x$ is the covariance matrix of $\boldsymbol{x}$. The symbol γ in (7) actually represents the expected value of γ in (2) (γ varies for each realization of $\boldsymbol{y}$); for simplicity the same notation γ is reused.

To obtain insightful results from the analysis, we assume that $\boldsymbol{x}$ is statistically *white*, with variance equal to σ_x^2. Thus, we can replace $\boldsymbol{C}_x$ with $\sigma_x^2\boldsymbol{I}$. This results in

$$\text{MSE} = \sigma_v^2\text{tr}\left(\boldsymbol{B}_\gamma^{-1}\boldsymbol{B}_o\boldsymbol{B}_\gamma^{-1}\right) + \sigma_x^2\gamma^2\text{tr}\left(\boldsymbol{B}_\gamma^{-2}\right). \tag{8}$$

In the case where the statistic of $\boldsymbol{x}$ is not truly white, (8) can still hold in an approximate sense when $\boldsymbol{C}_x$ is diagonally dominant. In such case, σ_x^2 is given by the average $\sigma_x^2 = \text{tr}(\boldsymbol{C}_x)/M$. The accuracy of such an approximation is due to the *trace* operation in (7), i.e., $\text{tr}[\boldsymbol{B}_\gamma^{-1}\boldsymbol{C}_x\boldsymbol{B}_\gamma^{-1}]$. Replacing $\boldsymbol{C}_x$ with the *average* diagonal matrix $\sigma_x^2\boldsymbol{I}$, when $\boldsymbol{x}$ is non-white, is found to be a good approximation in many cases. It is easy to see that the MSE has two components; a component that varies with σ_v^2, and a component that is dependent on σ_x^2. Using $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T$, (8) can be written as

$$\begin{aligned} \text{MSE} &= \sigma_v^2\text{tr}\left[\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2 + \gamma\boldsymbol{I}\right)^{-2}\right] \\ &+ \sigma_x^2\gamma^2\text{tr}\left[\left(\boldsymbol{\Lambda}^2 + \gamma\boldsymbol{I}\right)^{-2}\right]. \end{aligned} \tag{9}$$

Now, we can readily obtain the gradient of the MSE by differentiating (9) with respect to γ, which, after some algebraic manipulations, yields the simple form

$$\overline{\text{MSE}} = 2\text{tr}\left[\left(2\sigma_x^2\gamma - 2\sigma_v^2\right)\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2 + \gamma\boldsymbol{I}\right)^{-3}\right]. \tag{10}$$

By setting $\overline{\text{MSE}} = 0$, we find the location of the stationary point:

$$\gamma_{opt} = \frac{\sigma_v^2}{\sigma_x^2}. \tag{11}$$

By analyzing the sign of $\overline{\text{MSE}}$ to the left and right of γ_{opt}, it can be shown that this stationary point is a minimum. The result in (11) signifies that, under whiteness conditions, when the LS estimator (2) attains its

best performance, it actually converges to the Bayesian LMMSE estimator [2]. For white noise and non-white $\boldsymbol{x}$, the LMMSE estimator of $\boldsymbol{x}$ is given by [2]

$$\hat{\boldsymbol{x}}_{LMMSE} = (\boldsymbol{A}^T\boldsymbol{A} + \sigma_v^2\boldsymbol{C}_x^{-1})^{-1}\boldsymbol{A}^T\boldsymbol{y}, \qquad (12)$$

in which case the best LS estimation performance will not coverage exactly to that of the LMMSE estimator (which is the optimal linear estimator for the problem under consideration). However, the improved (regularized) LS estimator will always outperform the ordinary LS, as will be demonstrated later in this paper.

The question now is how to find the value of γ_{opt}. If we have knowledge of the signal and noise second order statistics, we may pursue an optimal LMMSE estimator. The parameter γ_{opt} can also be derived from the *signal-to-noise ratio* (SNR). However, estimating the SNR without prior knowledge on $\boldsymbol{x}$ is a tedious process that requires a large number of observations [9, 10]. In the absence of pertinent information, we need to find a way to estimate γ_{opt} without any assumptions on the signals. In the following section, we show how the BDU framework can be used to facilitate this task.

4. BDU-BASED IMPROVED LS (BDU-ILS) ESTIMATION

As discussed in the preceding sections, the regularization process based on the BDU can be viewed as searching for a perturbation matrix, $\boldsymbol{\Delta}$, with a bounded norm that improves the LS problem solution. The solution of the BDU regularization problem is given by the combination of (2) and (5). The best performance offered by the BDU approach (as well as any other LS approach whose final estimator takes the form (2)) occurs when $\gamma = \gamma_{opt}$ (see Eq. (11). The value of γ is obtained by solving (5), which is dependent on the choice of the parameter η. Eq. (5) can be manipulated to the form

$$\eta^2 = \frac{\text{tr}\left[\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-2}\boldsymbol{U}^T\left(\boldsymbol{y}\boldsymbol{y}^T\right)\boldsymbol{U}\right]}{\text{tr}\left[\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-2}\boldsymbol{U}^T\left(\boldsymbol{y}\boldsymbol{y}^T\right)\boldsymbol{U}\right]}. \qquad (13)$$

It can easily be seen that the value of η, required to produce a certain value of γ, is dependent on the observation vector $\boldsymbol{y}$. To obtain a useful expression for η, let us think of η as an *average* value over many realizations of the observation vector $\boldsymbol{y}$. Based on this perception, $\boldsymbol{y}\boldsymbol{y}^T$ can be replaced with its expected value $\mathbb{E}(\boldsymbol{y}\boldsymbol{y}^T)$, which, based on (1), can be expressed as

$$\mathbb{E}\left(\boldsymbol{y}\boldsymbol{y}^T\right) = \boldsymbol{A}\boldsymbol{C}_x\boldsymbol{A}^T + \sigma_v^2\boldsymbol{I}. \qquad (14)$$

Now, replacing $\boldsymbol{y}\boldsymbol{y}^T$ in (13) with $\mathbb{E}(\boldsymbol{y}\boldsymbol{y}^T)$ from (14), replacing $\boldsymbol{C}_x$ with $\sigma_x^2\boldsymbol{I}$, using the SVD, and manipulating, we obtain

$$\eta^2 = \frac{\text{tr}\left[\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-2}\left(\Lambda^2+\frac{\sigma_v^2}{\sigma_x^2}\boldsymbol{I}\right)\right]}{\text{tr}\left[\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-2}\left(\Lambda^2+\frac{\sigma_v^2}{\sigma_x^2}\boldsymbol{I}\right)\right]}. \qquad (15)$$

Finally, we can insert γ_{opt} to replace σ_v^2/σ_x^2, as in Eq. (11), and manipulate to obtain

$$\eta^2 = \frac{\text{tr}\left[\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-2}\left(\Lambda^2+\gamma_{opt}\boldsymbol{I}\right)\right]}{\text{tr}\left[\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-2}\left(\Lambda^2+\gamma_{opt}\boldsymbol{I}\right)\right]}. \qquad (16)$$

We will use (16) as a surrogate of (5). For the optimal choice $\eta = \eta_{opt}$, we have $\gamma = \gamma_{opt}$. Thus, based on (16), pair of η and γ are optimal if they satisfy

$$\eta^2 = \frac{\text{tr}\left[\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-1}\right]}{\text{tr}\left[\left(\boldsymbol{\Lambda}^2+\gamma\boldsymbol{I}\right)^{-1}\right]}. \qquad (17)$$

Eq. (17) dictates the the relationship between the best choice of the bound on the the perturbation matrix norm (η_{opt}), on the one hands, and the corresponding γ_{opt} that minimizes the MSE, on the other hand. If Eq. (5) is solved for γ when $\eta = \eta_{opt}$, we will obtain $\gamma = \gamma_{opt}$.

Remark 1: From Eq. (16), it can be proven that η^2 is a *strictly increasing* function of $\gamma \in [0, +\infty)$. The same applies for η^2 and γ in (17). This observation constitutes the core of the proposed procedure for finding γ_{opt} using (17) and (5).

4.1. Finding η_{opt}/γ_{opt}

Let us start with some hypothesized value for γ_{opt} that we denote γ_h. Substituting in (17), we obtain the corresponding value of η_{opt}^2, call it η_h^2. Next, substituting for $\eta^2 = \eta_h^2$ in (5) or (16) and solving for γ, we obtain $\hat{\gamma}$. Based on Remark 1 above and the relationship between (16) and (17), the following results can be proven:

1) $\gamma_h < \gamma_{opt} \Rightarrow \eta_h < \eta_{opt} \Rightarrow \hat{\gamma} > \gamma_h$.

2) $\gamma_h > \gamma_{opt} \Rightarrow \eta_h > \eta_{opt} \Rightarrow \hat{\gamma} < \gamma_h$.

3) $\gamma_h = \gamma_{opt} \Rightarrow \eta_h = \eta_{opt} \Rightarrow \hat{\gamma} = \gamma_h$.

This suggests that, starting from a certain γ_h, we can alternate between (17) and (5) to refine an initial estimate of the regularization parameter until convergence is reached. Convergence is indicated by the fact that $\hat{\gamma} \approx \gamma_h$ or that $\hat{\eta} \approx \eta_h$, where $\hat{\eta}$ is calculated from $\hat{\gamma}$ using (17). The stopping criterion can be derived based on the value of η^2 obtained from (17) by setting $\gamma = 0$ and manipulating to obtain $\eta_0^2 = \frac{M}{\text{tr}(\boldsymbol{\Lambda}^{-2})}$. This represents the smallest η value that corresponds to $\gamma \in [0, +\infty)$. We propose stopping the iterations when $|\hat{\eta}^2 - \eta_h^2|$ is less than a small fraction of η_0^2. A good choice of the initialization point is also found to be $\gamma_h = 0$. The steps of the proposed BDU-ILS estimation procedure can be summarized as follows:

#1 Initialize the threshold ϵ and set $\gamma_h = 0$.

#2 Calculate η_h^2 from (17).

#3 Substitute $\eta^2 = \eta_h^2$ in (5) and solve to obtain $\hat{\gamma}$.

#4 Calculate $\hat{\eta}^2$ using (17).

#5 Test for convergence:

- if $|\hat{\eta}^2 - \eta_h^2| < \epsilon$, go to step #6.
- else, $\gamma_h = \hat{\gamma}$, $\eta_h^2 = \hat{\eta}^2$, go to step #3.

#6 Calculate the estimate of $\boldsymbol{x}$ using (2) for $\gamma = \hat{\gamma}$.

5. NUMERICAL RESULTS

In this section, the performance of the proposed BDU-ILS approach is evaluated via numerical simulations. Four scenarios are considered where $\boldsymbol{A} \in \mathbb{R}^{100\times100}$ and $\boldsymbol{A} \in \mathbb{R}^{120\times100}$ combined with the elements of $\boldsymbol{x}$ being independent and identically distributed (white) and independent but not identically distributed (white noise is modulated by an exponential function to produce a non-white signal). The elements of the matrix $\boldsymbol{A}$ are generated from a Gaussian distribution with zero mean and unity variance. Fig. 1 plots the normalized MSE (NMSE) for each scenario. We use the LMMSE and the *white* LMMSE (WLMMSE) estimators, which assume known signal and noise statistics, as benchmark methods (the LMMSE is the best linear estimator for this problem and the WLMMSE is the best estimator based on (2)). The WLMMSE is obtained by replacing $\boldsymbol{C}_x$ in (12) with $\sigma_x^2\boldsymbol{I}$. Note that when $\boldsymbol{x}$ is white, the LMMSE and the WLMMSE are identical. In such a case, we include only to the LMMSE.

From Fig. 1, it can be seen that in all the four scenarios, the proposed BDU-ILS estimator outperforms the ordinary LS in most of the displayed SNR range. In Fig. 1 (a) where $\boldsymbol{x}$ is white and the system is square (100×100), the proposed estimator almost replicates the performance of the LMMSE estimator. The bias of the proposed BDU-ILS is not visible in the depicted SNR range in Fig. 1 (a). When the size of $\boldsymbol{A}$ is increased to 100×120, the performance of the BDU-ILS estimator exhibits a bias in the form of clear deviation from of the LMMSE and LS in the higher SNR regime. Fig. 1 (c) and (d) are the counterparts of Fig. 1 (a) and (b) for non-white $\boldsymbol{x}$. Performance degradation is clearly seen compared to the white signal case. However, for the most part, the performance of the proposed BDU-ILS estimator stays close to that of the WLMMSE estimator, emphasizing the capability of the BDU-ILS estimator to achieve near-optimal performance based on the regularization form (2).

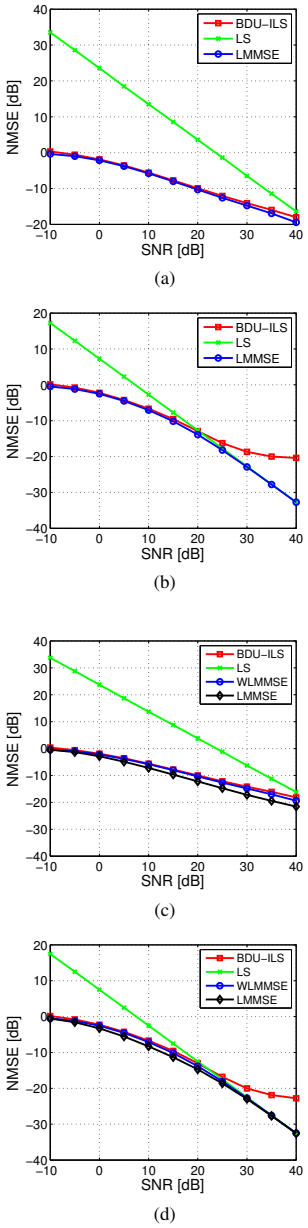

Fig. 1: Normalized MSE (NMSE) versus SNR: a) white $\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{100\times100}$; b) white $\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{120\times100}$; c) non-white $\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{100\times100}$; d) non-white $\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{120\times100}$.

6. CONCLUSIONS

We proposed the BDU-ILS estimator to solve the linear least squares estimation problem. The estimator targets the minimization of the MSE criterion. To do so, perturbation in the measurement matrix was permitted. To find the best regularization parameter, an iterative procedure based on the BDU model was proposed. Numerical results demonstrated the substantial improvement in the mean squared error offered by the proposed estimator. It was also shown that when the elements of the vector being estimated are white, the BDU-ILS estimator converges to the LMMSE estimator.

7. REFERENCES

[1] Y.C. Eldar, A Ben-Tal, and A Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 168–181, Jan 2005.

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Printice Hall, 1993.

[3] A. E. Hoerl and R.W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problem," *Technometrics*, vol. 12, pp. 55–67, Feb 1970.

[4] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometrics*, vol. 15, pp. 497–508, Aug 1973.

[5] Y.C. Eldar and A.V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 686–697, Mar 2003.

[6] Y.C. Eldar, "Improvement of least-squares under arbitrary weighted mse," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2007)*, April 2007, vol. 3, pp. III–837–III–840.

[7] W. Gander, "Least squares with a quadratic constraint," *Numerische Mathematik*, vol. 36, pp. 291–307, 1981.

[8] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM Journal on Matrix Analysis and Applications*, pp. 235–252, 1998.

[9] Johanna Vartiainen, Harri Saarnisaari, Janne J. Lehtomaki, and Markku Juntti, "A blind signal localization and snr estimation method," in *Proceedings of the 2006 IEEE Conference on Military Communications*, 2006, MILCOM'06, pp. 3317–3323.

[10] Hong Shu Liao, Lu Gan, and Ping Wei, "A blind snr estimation method for radar signals," in *2009 IET International Radar Conference*, April 2009, pp. 1–4.