MAXIMUM LIKELIHOOD ESTIMATION IN RANDOM LINEAR MODELS: GENERALIZATIONS AND PERFORMANCE ANALYSIS

Ami Wiesel and Yonina C. Eldar

Department of Electrical Engineering,
Technion – Israel Institute of Technology
amiw@tx.technion.ac.il, yonina@ee.technion.ac.il

ABSTRACT

We consider the problem of estimating an unknown deterministic parameter vector in a linear model with a Gaussian model matrix. The matrix has a known mean and independent rows of equal covariance matrix. Our problem formulation also allows for some known columns within this model matrix. We derive the maximum likelihood (ML) estimator associated with this problem and show that it can be found using a simple line-search over a unimodal function which can be efficiently evaluated. We then analyze its asymptotic performance using the Cramer Rao bound. Finally, we discuss the similarity between the ML, total least squares (TLS), and regularized TLS estimators.

1. INTRODUCTION

A generic estimation problem that has received much attention in the estimation literature is that of estimating an unknown, deterministic vector parameter $\mathbf{x}$ in the linear model $\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w}$, where $\mathbf{G}$ is a linear transformation and $\mathbf{w}$ is a Gaussian noise vector. The importance of this problem stems from the fact that a wide range of problems in communications, array processing, and many other areas of signal processing and statistics can be cast in this form.

Most of the literature concentrates on the simplest case, in which it is assumed that the model matrix $\mathbf{G}$ is completely specified. In this setting, the celebrated least squares (LS) estimator coincides with the maximum likelihood (ML) estimator and is known to minimize the mean squared error (MSE) among all unbiased estimators of $\mathbf{x}$ [1]. The estimation problem when $\mathbf{G}$ is not completely specified received much less attention. It can be divided into two main categories in which $\mathbf{G}$ is either deterministically unknown or random. In the standard errors in variables (EIV) model, $\mathbf{G}$ is considered as a deterministic unknown matrix, and the estimate is based on noisy observations of this matrix. The ML estimator for $\mathbf{x}$ in this case was derived in [2], and coincides with the well known total LS (TLS) estimator [3, 4]. Interestingly, the resulting estimator is a deregularized LS estimator. Thus, in order to stabilize the solution, regularized TLS (RTLS) estimators were derived [5, 6]. An opposite strategy is the robust LS estimator which is designed for the worst case $\mathbf{G}$ within a known deterministic set [7]. When $\mathbf{G}$ is assumed to be random, an intuitive approach is to minimize the expected LS (ELS) criterion with respect to $\mathbf{G}$ [8].

In this paper, we address the ML estimation of $\mathbf{x}$ when the model matrix $\mathbf{G}$ is a random matrix with known second order statistics. We recently solved this problem for the special case in which $\mathbf{G}$ has statistically independent and equal variance elements [9]. We found that the ML estimator in this setting is the solution of a multidimensional, non-linear and non-convex optimization problem. We reformulated it and solved it using a simple line-search over a unimodal function which can be efficiently evaluated. One of our interesting results was that the ML estimator can be interpreted as a TLS estimator with a logarithmic penalty. This result provides an important motivation to the RTLS estimator, and suggests a particular choice of regularization function.

Here, we develop new results on ML estimation in a linear model with a random model matrix. First, we generalize the problem setting. One of the important contributions in the TLS literature was the derivation of the generalized TLS (GTLS) estimator which is designed for the case in which only some of the columns of $\mathbf{G}$ are subject to error. This estimator also allows correlation between the elements in each of the rows of $\mathbf{G}$ (See [10, 4] and references within). In the sequel, we will derive the ML estimator associated with this channel model, and show that it can be solved using the same techniques as in [9].

Our second contribution is with respect to the performance analysis of the ML estimator. Since we do not have a closed form solution for the estimator, it is difficult to analytically asses its performance. Instead, we provide an asymptotic performance analysis using the Cramer Rao Bound (CRB). We derive a closed form expression of the CRB. Interestingly, the bound shows that the degradation in performance due to the randomness of the model matrix is not as severe as one may suspect. Actually, the randomness may even improve the performance in terms of MSE. Another interesting result is that when $\mathbf{G}$ is random the performance depends on the particular value of $\mathbf{x}$. This is in contrast to the well known CRB of estimation with known $\mathbf{G}$ that does not depend on $\mathbf{x}$.

The paper is organized as follows. In Section 2 we introduce the problem formulation and derive the ML estimator. We then compare our estimator to existing estimators in Section 3. An asymptotic performance analysis using the CRB is provided in Section 4. The advantage of the ML estimator is demonstrated in Section 5 using computer simulations. Finally, in Section 6, we provide concluding remarks.

The following notation is used. Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and standard lower case letters denote scalars. The superscript $(\cdot)^T$ denotes the transpose, and the superscript $(\cdot)^\dagger$ denotes the pseudoinverse. By $\mathbf{I}$ we denote the identity matrix. $\|\cdot\|$ is the standard Euclidean norm, and $\lambda_{\min}(\mathbf{X})$ is the smallest eigenvalue of $\mathbf{X}$. Finally, $\mathbf{X} \succ 0$ means that the matrix $\mathbf{X}$ is a symmetric positive definite matrix.

This work was supported by the EU 6th framework programme, via the NEWCOM network of excellence, and by the ISRAEL SCIENCE FOUNDATION.

2. THE MAXIMUM LIKELIHOOD ESTIMATOR

Consider the problem of estimating an unknown deterministic parameter vector $\mathbf{x} = \left[\mathbf{x}_1^T\ \mathbf{x}_2^T\right]^T$ in the linear model

$$\mathbf{y} \quad = [\mathbf{G}_1\ \mathbf{G}_2] \left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \end{array}\right] + \mathbf{w}, \tag{1}$$

where $\mathbf{G} = [\mathbf{G}_1\ \mathbf{G}_2]$ is an $N \times (K_1 + K_2)$ random matrix partitioned into a random part $\mathbf{G}_1$ with mean $\mathbf{H}_1$ and independent rows of covariance $\mathbf{C}_H \succ 0$, and a known part $\mathbf{G}_2 = \mathbf{H}_2$. For convenience, we define $\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2]$. The vector $\mathbf{w}$ is a zero mean Gaussian vector with independent elements of variance $\sigma_w^2 > 0$. In addition, $\mathbf{G}_1$ and $\mathbf{w}$ are statistically independent.

An estimator $\widehat{\mathbf{x}}(\mathbf{y}, \mathbf{H}, \mathbf{C}_H, \sigma_w^2)$ of $\mathbf{x}$ is defined as a function of the observations vector and the given statistics that is close to $\mathbf{x}$ in some sense. One of the standard approaches for designing $\widehat{\mathbf{x}}$ is ML estimation, where the estimate is chosen as the parameter vector $\mathbf{x}$ that maximizes the likelihood of the observations. Mathematically, the ML estimate of $\mathbf{x}$ is the solution to

$$\max_{\mathbf{x}} \log p\left(\mathbf{y}; \mathbf{x}\right), \tag{2}$$

where $p\left(\mathbf{y}; \mathbf{x}\right)$ is the probability density function of $\mathbf{y}$ parameterized by $\mathbf{x}$. It is easy to see that in our model $\mathbf{y}$ is a Gaussian vector with mean $\mathbf{Hx}$ and covariance $\left(\mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 + \sigma_w^2\right)\mathbf{I}$. Therefore, the ML estimator of $\mathbf{x}$ can be found by solving

$$\min_{\mathbf{x}_1, \mathbf{x}_2} \left\{ \frac{\left\| \mathbf{y} - [\mathbf{H}_1\ \mathbf{H}_2] \left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \end{array}\right] \right\|^2}{\mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 + \sigma_w^2} + N \log\left(\mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 + \sigma_w^2\right) \right\}. \tag{3}$$

Problem (3) is a K-dimensional, non-linear and non-convex optimization program, and is therefore considered difficult. Our main result is that we can transform it into a tractable form and solve it efficiently as we now show.

We begin by solving for $\mathbf{x}_2$ using the well known LS solution

$$\mathbf{x}_2 = \mathbf{H}_2^\dagger \left(\mathbf{y} - \mathbf{H}_1 \mathbf{x}_1\right). \tag{4}$$

We then plug the optimal $\mathbf{x}_2$ back into the objective:

$$\min_{\mathbf{x}_1} \left\{ \frac{\left\| \widetilde{\mathbf{y}} - \widetilde{\mathbf{H}}_1 \mathbf{x}_1 \right\|^2}{\mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 + \sigma_w^2} + N \log\left(\mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 + \sigma_w^2\right) \right\}, \tag{5}$$

where $\widetilde{\mathbf{y}} = \left(\mathbf{I} - \mathbf{H}_2 \mathbf{H}_2^\dagger\right) \mathbf{y}$ and $\widetilde{\mathbf{H}}_1 = \left(\mathbf{I} - \mathbf{H}_2 \mathbf{H}_2^\dagger\right) \mathbf{H}_1$. The problem in (5) is similar to the optimization program which we solved in [9]. Therefore, we rely on the results in [9] and obtain the following theorem:

Theorem 1 ([9]). *For any $t \geq 0$, let*

$$f(t) \quad = \left\{ \begin{array}{ll} \min_{\mathbf{x}_1} & \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{H}}_1 \mathbf{x}_1\|^2 \\ \text{s.t.} & \mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 = t, \end{array} \right. \tag{6}$$

and denote the optimal argument by $\mathbf{x}_1(t)$. Then, the solution to (5) is $\mathbf{x}_1(t^*)$ where t^* is the minimizer of the following unimodal optimization problem

$$\min_{t \geq 0} \left\{ \frac{f(t)}{t + \sigma_w^2} + N \log\left(t + \sigma_w^2\right) \right\}. \tag{7}$$

At first sight, Theorem 1 looks trivial. It is just a different way of writing (5) using a slack variable t. However, it allows for an efficient solution of the ML problem due to two important observations. The first is that there are standard methods for evaluating $f(t)$ in (6) for any $t \geq 0$. The second is that the line-search in (7) is unimodal in $t \geq 0$, and therefore any simple one-dimensional search algorithm, such as bisection, can efficiently find its global minima.

We will now discuss the methods for evaluating $f(t)$ in (6). This is a quadratically constrained LS problem whose solution can be traced back to [11]:

Lemma 1 ([11, 12]). *The solution of*

$$f(t) \quad = \left\{ \begin{array}{ll} \min_{\mathbf{x}_1} & \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{H}}_1 \mathbf{x}_1\|^2 \\ \text{s.t.} & \mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 = t, \end{array} \right. \tag{8}$$

is

$$\mathbf{x}_1(t) = \left(\widetilde{\mathbf{H}}_1^T \widetilde{\mathbf{H}}_1 + \alpha \mathbf{C}_H\right)^\dagger \widetilde{\mathbf{H}}_1^T \widetilde{\mathbf{y}}, \tag{9}$$

where $\alpha \geq -\lambda_{\min}\left(\mathbf{C}_H^{-\frac{1}{2}} \widetilde{\mathbf{H}}_1^T \widetilde{\mathbf{H}}_1 \mathbf{C}_H^{-\frac{1}{2}}\right)$ is the unique root of the equation

$$\mathbf{x}_1(t)^T \mathbf{C}_H \mathbf{x}_1(t) = t. \tag{10}$$

Using the eigenvalue decomposition of $\mathbf{C}_H^{-\frac{1}{2}} \widetilde{\mathbf{H}}_1^T \widetilde{\mathbf{H}}_1 \mathbf{C}_H^{-\frac{1}{2}}$ we can easily calculate $\mathbf{x}_1(t)^T \mathbf{C}_H \mathbf{x}_1(t)$ for different values of α. The monotonicity in α enables us to find the α which satisfies (10) using a simple line-search. Once this α is found, $f(t)$ can be evaluated by plugging the appropriate $\mathbf{x}_1(t)$ into $\|\widetilde{\mathbf{y}} - \widetilde{\mathbf{H}}_1 \mathbf{x}_1(t)\|^2$. Moreover, the function can be efficiently evaluated also in large scale problems, such as those arising in image processing applications, where the eigenvalue decomposition is not practical. More details on this procedure and the related "Trust Region Subproblem" can be found in [13] and references within.

3. COMPARISON TO THE ML ESTIMATOR IN THE ERROR IN VARIABLES MODEL

In this section, we compare our estimation problem with the standard EIV problem formulation which is a similar approach for handling the uncertainty in the model matrix [2]. The EIV model is

$$\left\{ \begin{array}{l} \mathbf{y} = [\mathbf{G}_1\ \mathbf{H}_2] \left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \end{array}\right] + \mathbf{w} \\ \mathbf{H}_1 = \mathbf{G}_1 + \mathbf{W}_1, \end{array} \right. \tag{11}$$

where $\mathbf{y}$ and $\mathbf{H}_1$ are the observed vector and matrix, $\mathbf{w}$ is a zero mean Gaussian vector of covariance $\sigma_w^2 \mathbf{I}$, and $\mathbf{W}_1$ is a zero mean Gaussian matrix with independent rows of covariance $\mathbf{C}_H$. As before, we assume that $\mathbf{w}$ and $\mathbf{W}_1$ are statistically independent.

Models (1) and (11) are very similar. In both, we have access to the observations $\mathbf{y}$ and to $\mathbf{H}$, and the true channel $\mathbf{G}_1$ is equal to $\mathbf{H}_1$ plus some Gaussian noise. The main difference is that in model (1) the matrix $\mathbf{G}_1$ itself is random, whereas in (11) the matrix $\mathbf{G}_1$ is deterministically unknown. Thus, the ML estimator in (11) estimates both $\mathbf{x}$ and $\mathbf{G}_1$ by solving

$$\max_{\mathbf{x}, \mathbf{G}_1} \log p\left(\mathbf{y}, \mathbf{H}_1; \mathbf{x}, \mathbf{G}_1\right), \tag{12}$$

where $p(\mathbf{y}, \mathbf{H}_1; \mathbf{x}, \mathbf{G}_1)$ is the joint probability density function of $\mathbf{y}$ and $\mathbf{H}_1$ parameterized by $\mathbf{x}$ and $\mathbf{G}_1$. Due to the Gaussian assumption, (12) is equivalent to

$$\min_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{G}_1} \left\{ \frac{1}{\sigma_w^2} \left\| \mathbf{y} - [\mathbf{G}_1 \ \mathbf{H}_2] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right\|^2 + \|\mathbf{H}_1 - \mathbf{G}_1\|_{\mathbf{C}_H^{-1}} \right\}, \quad (13)$$

where $\|\mathbf{X}\|_{\mathbf{W}} = \mathrm{Tr}\left\{\mathbf{X}\mathbf{W}\mathbf{X}^T\right\}$. In our context, we are not really interested in the nuisance parameter $\mathbf{G}_1$. Instead, using tedious algebraic manipulations, we analytically solve for $\mathbf{G}_1$, substitute it into (13), and obtain the following optimization problem

$$\min_{\mathbf{x}_1,\mathbf{x}_2} \frac{\left\| \mathbf{y} - [\mathbf{H}_1 \ \mathbf{H}_2] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right\|^2}{\mathbf{x}_1^T \mathbf{C}_H \mathbf{x}_1 + \sigma_w^2}. \quad (14)$$

Comparing (3) and (14) we see that the ML in (3) can be considered as the ML of (14) with an additional logarithmic penalty.

In the signal processing literature (14) is usually known as a generalized (or weighted) variant of the TLS estimator [10, 4]. The TLS is an extension of the LS solution for the problem $\mathbf{y} \approx \mathbf{H}\mathbf{x}$ when both $\mathbf{y}$ and $\mathbf{H}$ are subject to measurement errors. It tries to find $\mathbf{x}$ and $\mathbf{G}_1$ that minimize the squared errors in $\mathbf{y}$ and in $\mathbf{H}_1$ as expressed by (13). Thus, our ML estimator can also be interpreted as a regularized (or penalized) generalized TLS estimator.

Interestingly, the concept of regularizing the TLS estimator is not new [5, 6]. It is well known that the TLS solution is not stable when it is applied to ill posed problems. In such cases, a regularization of some sort is required. Two standard regularization methods are:

$$\min_{\mathbf{x}} \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\mathbf{x}^T\mathbf{C}_H\mathbf{x}+\sigma_w^2} \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{T}\mathbf{x} \leq 1 \quad (15)$$

$$\min_{\mathbf{x}} \left\{ \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\mathbf{x}^T\mathbf{C}_H\mathbf{x}+\sigma_w^2} + \mathbf{x}^T\mathbf{T}\mathbf{x} \right\}, \quad (16)$$

where $\mathbf{T} \succeq 0$ is a regularization matrix. It has been shown that in many applications these heuristic approaches may significantly improve the performance of the TLS estimator in terms of MSE. Our new ML estimator provides a statistical reasoning to this phenomena and suggests an inherent logarithmic penalty scheme. Furthermore, using $\log(1+a) \leq a$, which is tight for sufficiently small a, we obtain the following upper bound on our ML criterion in (3)

$$\min_{\mathbf{x}} \left\{ \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\mathbf{x}^T\mathbf{C}_H\mathbf{x}+\sigma_w^2} + \frac{N}{\sigma_w^2}\mathbf{x}^T\mathbf{C}_H\mathbf{x} \right\}, \quad (17)$$

which is exactly the RTLS estimator in (16) with $\mathbf{T} = \frac{N}{\sigma_w^2}\mathbf{C}_H$. Thus, our ML estimator also provides a reasonable choice for the regularization matrix of the RTLS estimator.

4. ASYMPTOTIC PERFORMANCE ANALYSIS AND THE CRAMER RAO BOUND

In the previous section (and in [9]) we presented a numerical algorithm for finding the ML estimator in a linear model when the model matrix is Gaussian. Unfortunately, it is difficult to analytically evaluate the performance of this estimator. Instead, we now provide an asymptotic performance analysis using the CRB. The CRB is a lower bound for the MSE of any unbiased estimator [1]. Moreover, it is well known that under a number of regularity conditions the MSE of the ML estimator asymptotically attains this bound. For simplicity, in this section, we restrict ourselves to the case where $\mathbf{G} = \mathbf{G}_1$ and $\mathbf{x} = \mathbf{x}_1$, although the generalization to the model in (1) is straight forward. The CRB for this special case is provided in the following theorem.

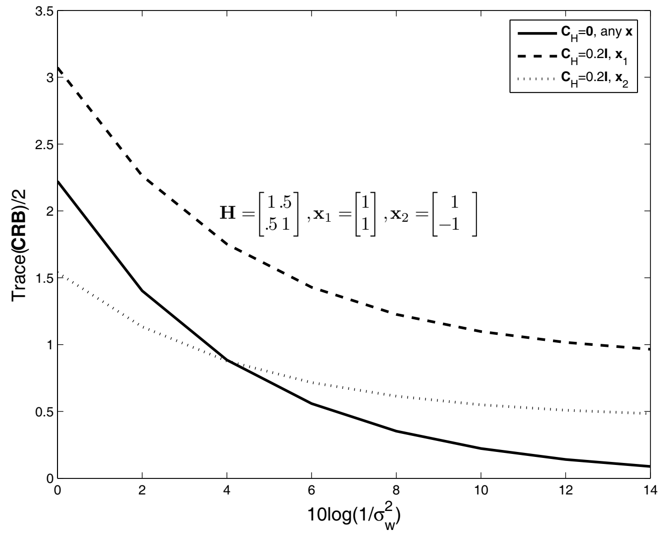

Fig. 1. The CRB in (18) for a specific choices of $\mathbf{H}$ and $\mathbf{x}$.

Theorem 2. *Let $\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w}$ where $\mathbf{G}$ is Gaussian with a full rank mean $\mathbf{H}$ and independent rows of covariance $\mathbf{C}_H$, $\mathbf{w}$ is a Gaussian vector with zero mean and covariance $\sigma_w^2\mathbf{I}$, and $\mathbf{G}$ is statistically independent of $\mathbf{w}$. Then, the CRB for estimating the deterministic vector $\mathbf{x}$ based on the observations $\mathbf{y}$ is*

$$\mathbf{B}(\mathbf{C}_H) = \left(\mathbf{x}^T\mathbf{C}_H\mathbf{x} + \sigma_w^2\right)\left(\mathbf{H}^T\mathbf{H}\right)^{-1} - \boldsymbol{\Delta} \quad (18)$$

where

$$\boldsymbol{\Delta} = \frac{\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{C}_H\mathbf{x}\mathbf{x}^T\mathbf{C}_H\left(\mathbf{H}^T\mathbf{H}\right)^{-1}}{\frac{1}{2N} + \frac{\mathbf{x}^T\mathbf{C}_H\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{C}_H\mathbf{x}}{\mathbf{x}^T\mathbf{C}_H\mathbf{x}+\sigma_w^2}} \quad (19)$$

Proof. The CRB is obtained using the closed form expression of the Fisher information matrix for the multivariate Gaussian distribution in [1] and by applying the matrix inversion lemma. ☐

In the special case where $\mathbf{C}_H = \mathbf{0}$, the CRB in (18) is identical to the well known CRB of estimating $\mathbf{x}$ in a known linear model and is equal to

$$\mathbf{B}(\mathbf{0}) = \sigma_w^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}. \quad (20)$$

Comparing (18) with (20) reviles that the randomness of the matrix model has two complementary effects on the CRB. First, as expected, it can be interpreted as an additional independent noise term with variance $\mathbf{x}^T\mathbf{C}_H\mathbf{x}$. This is easy to explain by writing (1) as $\mathbf{y} = \mathbf{H}\mathbf{x} + \widetilde{\mathbf{w}}$ where $\widetilde{\mathbf{w}}$ is a zero mean Gaussian vector with independent elements with variance $\mathbf{x}^T\mathbf{C}_H\mathbf{x} + \sigma_w^2$. However, the randomness has another positive effect on the CRB due to the term $\boldsymbol{\Delta}$. This effect tries to compensate for the additional noise, and decreases the CRB. Thus, the degradation in performance due to the randomness is not as severe as one may intuitively suspect. Actually, plotting (18) in Fig. 1 for different values of $\mathbf{C}_H$ and $\mathbf{x}$ reveals

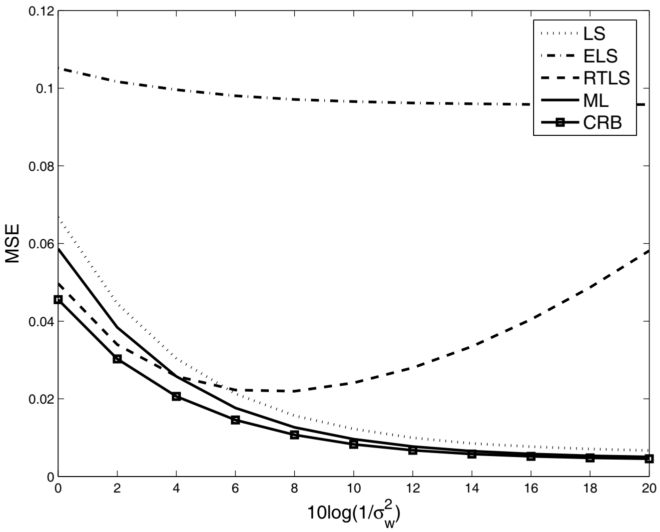

Fig. 2. MSE in estimating $\mathbf{x}$ using different estimators in comparison to the CRB.

that adding randomness may even improve the MSE. A possible explanation is that the randomness in $\mathbf{G}$ increases the signal to noise ratio and is therefore beneficial. Another interesting observation is that unlike (20) the CRB in a random model matrix depends on the specific value of $\mathbf{x}$.

5. NUMERICAL EXAMPLE

We now provide a numerical example illustrating the behavior of our new estimator. The purpose of this example is to demonstrate its performance advantage, rather than a detailed practical application, which is beyond the scope of this paper. The parameters in our simulation were as follows. The matrix $\mathbf{H}$ was chosen as a concatenation of T 4×4 matrices with unit diagonal elements and 0.5 off-diagonal elements. We expect that the ML estimator will attain its asymptotic performance as T increases, therefore we choose $T = 50$. The vector $\mathbf{x}$ was chosen as the normalized eigenvector of $\mathbf{H}^T\mathbf{H}$ associated with its minimal eigenvalue. We estimated the MSEs of each estimator using 50000 computer simulation. For comparison, we provide the results for the ML estimator of (3), the standard LS estimator, the ELS estimator of [8], and the RTLS estimator of (17). The results are presented in Fig. 2 for $\mathbf{C}_H = 0.1\mathbf{I}$. It is easy to see the advantage of the ML estimator over the existing estimators. As expected, when σ_w^2 is relatively high, the RTLS estimator is a good approximation for the ML estimator, and may even result in lower MSEs. On the other hand, the ML estimator approaches the CRB when σ_w^2 is sufficiently low.

6. CONCLUSION

In this paper, we considered the problem of estimating $\mathbf{x}$ in the model $\mathbf{y} = \mathbf{G}\mathbf{x}+\mathbf{w}$ when $\mathbf{G}$ is Gaussian. In continuation to [9], we derived the ML estimator in a generalized problem formulation and provided an efficient method for finding it. We discussed the similarity of the ML estimator with other estimation algorithms, and showed that it can be expressed as a logarithmic regularization of the well known TLS estimator. This result provides a statistical justification for the RTLS that is usually derived based on heuristic considerations. We then analyzed its asymptotic performance using the CRB, and discussed the effect of the channel's randomness on the performance.

Our results motivate the continuing research on this seemingly simple estimation problem. There are still many open questions. For example, an important extension of our work is to consider the problem of estimating $\mathbf{x}$ in a model with multiple observations, i.e., when we observe $\mathbf{y}_t = \mathbf{G}\mathbf{x}_t + \mathbf{w}_t$ for $t = 1, \cdots, T$ and $\mathbf{G}$ is random.

7. REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*, Prentice Hall, 1993.

[2] W. A. Fuller, *Measurement error models*, Wiley Series in Probability and Mathematical Statistics, 1987.

[3] G. H. Golub and C. F. Van Loan, "An analysis of the total least-squares problem," *SIAM Journal of Numerical Analysis*, vol. 17, no. 6, pp. 883–893, 1980.

[4] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers in Applied Mathematics 9, Siam, 1991.

[5] P. C. Hansen G. H. Golub and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM Journal Matrix Analysis and Applications*, Oct. 1999.

[6] A. Beck and A. Ben-Tal, "Regularization of the total least squares," *submitted to SIAM Journal on Optimization*, 2005.

[7] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal Matrix Analysis and Applications*, pp. 1034–1064, Oct. 1997.

[8] H. A. Hindi and S. P. Boyd, "Robust solutions to l_1, l_2, and l_∞ uncertain linear approximation problems using convex optimization," in *Proc. of the 1998 American Control Conference*, June 1998.

[9] A. Wiesel, Y. C. Eldar, and A. Beck, "Maximum likelihood estimation in linear models with a gaussian model matrix," *submitted to IEEE Signal Processing Letters*, Aug. 2005.

[10] S. Van Huffel and J. Vandewalle, "Analysis and properties of the generalized total least squares problem $\mathbf{a}\mathbf{x} \approx \mathbf{b}$ when some or all columns in $\mathbf{a}$ are subject to error," *SIAM Journal Matrix Analysis and Applications*, pp. 294–315, July 1989.

[11] G. E. Forsythe and G. H. Golub, "On the stationary values of a second-degree polynomial on the unit sphere," *Journal of the Society for Industrial and Applied Mathematics*, vol. 13, no. 4, pp. 1050–1068, 1965.

[12] Y. C. Eldar and A. Beck, "Hidden convexity based near maximum-likelihood CDMA detection," in *Proc. of conference on Information Sciences and Systems, Princeton, (CISS-2005)*, March 2005.

[13] C. Fortin and H. Wolkowicz, "The trust region subproblem and semidefinite programming," in *Canada Research Report CORR 2002-22, available at http://www.optimization-online.org*.