Maximum Likelihood Estimation in Linear Models With a Gaussian Model Matrix

Ami Wiesel, Yonina C. Eldar, and Amir Beck

***Abstract*—We consider the problem of estimating an unknown deterministic parameter vector in a linear model with a Gaussian model matrix. We derive the maximum likelihood (ML) estimator for this problem and show that it can be found using a simple line-search over a unimodal function that can be efficiently evaluated. We then discuss the similarity between the ML, the total least squares (TLS), the regularized TLS, and the expected least squares estimators.**

***Index Terms*—Errors in variables (EIV), linear models, maximum likelihood (ML) estimation, random model matrix, total least squares (TLS).**

I. 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]. In ill-posed problems, the regularized LS estimator due to Tikhonov [2] can often outperform the LS strategy in terms of MSE. An alternative approach is taken in [3], where the minimax MSE estimator is derived.

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 [4] and coincides with the well-known total LS (TLS) estimator [5], [6]. Interestingly, the resulting estimator is a deregularized LS estimator. Thus, in order to stabilize the solution, regularized TLS (RTLS) estimators were derived [7], [8]. An opposite strategy is the robust LS estimator that is designed for the worst-case $\mathbf{G}$ within a known deterministic set [9], [10]. When $\mathbf{G}$ is assumed to be random, an intuitive approach is to minimize the expected LS (ELS) criterion with respect to $\mathbf{G}$ [11], [12]. Finally, the minimax MSE estimator was also generalized to the case of a deterministic $\mathbf{G}$ subject to uncertainties in [3] and to the case of a random $\mathbf{G}$ matrix in [12].

In this letter, we address the ML estimation of $\mathbf{x}$ in a linear model, when the model matrix $\mathbf{G}$ is a random matrix with independent and identically distributed Gaussian elements and known second-order statistics. The ML estimator in this case is the solution of a multidimensional, nonlinear, and nonconvex optimization problem. We reformulate it and solve it using a simple line-search over a unimodal function that can be efficiently evaluated. The resulting estimator may be interpreted as a TLS estimator with a logarithmic penalty or as an approximate ELS estimator. These results provide an important motivation to the these well-known estimators and suggest a particular choice of regularization function.

This letter is organized as follows. In Section II, we introduce the problem formulation and derive the ML estimator. Next, we compare our estimator with existing estimators in Section III. The advantage of the ML estimator is demonstrated in Section IV using computer simulations. Finally, in Section V, we provide concluding remarks.

The following notation is used. Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and standard lowercase letters denote scalars. The superscript $(\cdot)^T$ denotes the transpose, the superscripts $(\cdot)'$ and $(\cdot)''$ denote the first and second derivatives, respectively, and the superscript $(\cdot)^\dagger$ denotes the pseudoinverse. By $\mathbf{I}$, we denote the identity matrix. $\|\cdot\|_F$ is the Frobenius matrix norm, $\|\cdot\|$ is the standard Euclidean norm, $\mathcal{R}(\mathbf{X})$ is the range of $\mathbf{X}$, and $\lambda_{\min}(\mathbf{X})$ is the smallest eigenvalue of $\mathbf{X}$. Finally, $\mathbf{X} \succeq 0$ means that the matrix $\mathbf{X}$ is a symmetric positive semidefinite matrix.

II. ML Estimation

Consider the problem of estimating an unknown deterministic parameter vector $\mathbf{x}$ in the linear model

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{G}$ is an $N \times K$ Gaussian matrix with known mean $\mathbf{H}$ and independent elements of variance $\sigma_h^2 > 0$, and $\mathbf{w}$ is a zero-mean Gaussian vector with independent elements of variance $\sigma_w^2 > 0$. In addition, $\mathbf{G}$ and $\mathbf{w}$ are statistically independent.

Manuscript received September 7, 2005; revised December 5, 2005. This work was supported in part by the Israel Science Foundation and in part by the European Union 6th Framework Program, via the NEWCOM network of excellence. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian Sadler.

The authors are with the Department of Electrical Engineering, The Technion, Israel Institute of Technology, Haifa 32000, Israel (e-mail: amiw@tx.technion.ac.il; yonina@ee.technion.ac.il; becka@tx.technion.ac.il).

Digital Object Identifier 10.1109/LSP.2006.870377

An estimator $\widehat{\mathbf{x}}(\mathbf{y}, \mathbf{H}, \sigma_h^2, \sigma_w^2)$ of $\mathbf{x}$ is defined as a function of the observations vector and the given statistics that are close to $\mathbf{x}$ in some sense. One of the standard approaches for desiging $\widehat{\mathbf{x}}(\cdot)$ 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 $\left(\sigma_h^2\|\mathbf{x}\|^2 + \sigma_w^2\right)\mathbf{I}$. Therefore, the ML estimator of $\mathbf{x}$ can be found by solving

$$\min_{\mathbf{x}} \left\{ \frac{\|\mathbf{y} - \mathbf{Hx}\|^2}{\sigma_h^2\|\mathbf{x}\|^2 + \sigma_w^2} + N \log\left(\sigma_h^2\|\mathbf{x}\|^2 + \sigma_w^2\right) \right\}. \tag{3}$$

Problem (3) is a K-dimensional, nonlinear, and nonconvex optimization program and is therefore considered difficult. Our main result is that we can transform it into a tractable form and solve it efficiently,[1] as summarized in the following theorem.

Theorem 1: For any $t \geq 0$, let

$$f(t) = \min_{\mathbf{x}:\|\mathbf{x}\|^2 = t} \|\mathbf{y} - \mathbf{Hx}\|^2 \tag{4}$$

and denote the optimal argument by $\mathbf{x}(t)$. Then, the ML estimator of $\mathbf{x}$ in the model (1) is $\mathbf{x}(t^*)$, where t^* is the solution to the following unimodal optimization problem:

$$\min_{t \geq 0} \left\{ \frac{f(t)}{\sigma_h^2 t + \sigma_w^2} + N \log\left(\sigma_h^2 t + \sigma_w^2\right) \right\}. \tag{5}$$

Proof: See the Appendix . □

At first sight, Theorem 1 looks trivial. It is just a different way of writing (3) 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 (4) for any $t \geq 0$. The second is that the line-search in (5) 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 (4). This is a norm-constrained LS problem whose solution can be traced back to [13].

Lemma 1 ([13], [14]): The solution of

$$f(t) = \min_{\mathbf{x}:\|\mathbf{x}\|^2 = t} \|\mathbf{y} - \mathbf{Hx}\|^2 \tag{6}$$

is

$$\mathbf{x}(t) = \left(\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}\right)^{\dagger} \mathbf{H}^T\mathbf{y} \tag{7}$$

where $\alpha \geq -\lambda_{\min}\left(\mathbf{H}^T\mathbf{H}\right)$ is the unique root of the equation

$$\|\mathbf{x}(t)\|^2 = t. \tag{8}$$

[1]A similar approach is taken in [8] for deriving the Tikhonov regularization of the TLS estimator.

Using the eigenvalue decomposition of $\mathbf{H}^T\mathbf{H}$, we can easily calculate $\|\left(\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}\right)^{\dagger}\mathbf{H}^T\mathbf{y}\|^2$ for different values of α. The monotonicity of this squared norm in α enables us to find the α that satisfies (8) using a simple line-search. Once this α is found, $f(t)$ can be evaluated by plugging the appropriate $\mathbf{x}(t)$ into $\|\mathbf{y} - \mathbf{Hx}(t)\|^2$. 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 [15] and references within.

III. Comparison to Similar Problems

In this section, we compare our problem with similar estimation problems in statistical signal processing.

A. Comparison to the ML in the EIV Model and the TLS

One of the standard approaches in the statistical literature for estimating $\mathbf{x}$ in a linear model with model matrix uncertainty is the EIV formulation [4]. The EIV model is

$$\begin{cases} \mathbf{y} = \mathbf{Gx} + \mathbf{w} \\ \mathbf{H} = \mathbf{G} + \mathbf{W} \end{cases} \tag{9}$$

where $\mathbf{y}$ and $\mathbf{H}$ are the observed vector and matrix, $\mathbf{w}$ is a zero-mean Gaussian vector of covariance $\sigma_w^2\mathbf{I}$, and $\mathbf{W}$ is a zero-mean Gaussian matrix with independent elements of variance σ_h^2.

Models (1) and (9) are very similar. In both, we have access to the observations $\mathbf{y}$ and to $\mathbf{H}$, and the true channel $\mathbf{G}$ is equal to $\mathbf{H}$ plus some Gaussian noise. In model (1), the matrix $\mathbf{H}$ is a deterministic parameter, whereas in (9), it is a random observation matrix. Practically, though, its value is known in both cases. In our view, the main difference is that in model (1), the matrix $\mathbf{G}$ is random, whereas in (9), it is a deterministically unknown matrix that must be estimated as well. Thus, the ML estimator in (9) estimates both $\mathbf{x}$ and $\mathbf{G}$ by solving

$$\max_{\mathbf{x},\mathbf{G}} \log p\left(\mathbf{y}, \mathbf{H}; \mathbf{x}, \mathbf{G}\right) \tag{10}$$

where $p\left(\mathbf{y}, \mathbf{H}; \mathbf{x}, \mathbf{G}\right)$ is the joint probability density function of $\mathbf{y}$, and $\mathbf{H}$ parameterized by $\mathbf{x}$ and $\mathbf{G}$. Now, due to the Gaussian assumption, (10) is equivalent to

$$\min_{\mathbf{x},\mathbf{G}} \left\{ \frac{\|\mathbf{y} - \mathbf{Gx}\|^2}{\sigma_w^2} + \frac{\|\mathbf{H} - \mathbf{G}\|_F^2}{\sigma_h^2} \right\}. \tag{11}$$

In our context, we are not really interested in the nuisance parameter $\mathbf{G}$. Instead, we eliminate it by minimizing (11) over $\mathbf{G}$ first and find that the ML estimate of $\mathbf{x}$ in (9) is the solution to

$$\min_{\mathbf{x}} \frac{\|\mathbf{y} - \mathbf{Hx}\|^2}{\sigma_h^2\|\mathbf{x}\|^2 + \sigma_w^2}. \tag{12}$$

Comparing (3) and (12), we see that the ML in (3) can be considered as the ML of (12) with an additional logarithmic penalty.

In the signal processing literature, (12) is usually known as the TLS estimator [6]. The TLS is a generalization of the LS solution for the problem $\mathbf{y} \approx \mathbf{Hx}$ when both $\mathbf{y}$ and $\mathbf{H}$ are subject to measurement errors. It tries to find $\mathbf{x}$ and $\mathbf{G}$ that minimize the squared errors in $\mathbf{y}$ and in $\mathbf{H}$ as expressed by (11). Thus, our ML estimator can also be interpreted as a regularized (or penalized) TLS estimator.

Interestingly, the concept of regularizing the TLS estimator is not new [7], [8]. 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

$$\min_{\mathbf{x}} \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_h^2\|\mathbf{x}\|^2+\sigma_w^2} \quad \text{s.t.} \quad \|\mathbf{x}\|^2 \le \mu \tag{13}$$

$$\min_{\mathbf{x}} \left\{ \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_h^2\|\mathbf{x}\|^2+\sigma_w^2} + \mu\|\mathbf{x}\|^2 \right\}. \tag{14}$$

It has been shown that in many applications, these heuristic regularizations 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

$$\log(1+a) \le a \tag{15}$$

which is tight for sufficiently small a, we obtain the following upper bound on our ML criterion in (3):

$$\min_{\mathbf{x}} \left\{ \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_h^2\|\mathbf{x}\|^2+\sigma_w^2} + \frac{N\sigma_h^2}{\sigma_w^2}\|\mathbf{x}\|^2 \right\} \tag{16}$$

which is exactly the RTLS estimator in (14) with $\mu = N\sigma_h^2/\sigma_w^2$. Thus, (16) is a reasonable approximation of our ML estimator when $(N\sigma_h^2/\sigma_w^2)\|\mathbf{x}\|^2$ is sufficiently small.

B. Comparison to Expected LS

The ML estimator is also related to the ELS estimator derived in [11] and [12]. The ELS criterion is the most intuitive approach for generalizing the LS estimator to the case where $\mathbf{G}$ is random. It optimizes the expected value of the data error

$$\min_{\mathbf{x}} E_{\mathbf{G}}\{\|\mathbf{y}-\mathbf{G}\mathbf{x}\|^2\} \tag{17}$$

where $E_{\mathbf{G}}\{\cdot\}$ denotes the expectation with respect to the distribution of $\mathbf{G}$. Straightforward evaluation of the expectation in our model yields

$$\min_{\mathbf{x}} \left\{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 + N\sigma_h^2\|\mathbf{x}\|^2\right\}. \tag{18}$$

Examining (18) reveals that the standard LS objective should be penalized by $N\sigma_h^2\|\mathbf{x}\|^2$ when $\mathbf{G}$ is random. We will now show that (18) is equivalent to minimizing a lower bound on the objective of our ML estimator. Again, we apply

$$\log(a+b) \le \log(a) + \frac{b}{a} \tag{19}$$

and obtain the following bound:

$$\begin{aligned} &N\log\left(\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 + N\sigma_h^2\|\mathbf{x}\|^2 + N\sigma_w^2\right) \\ &\quad \le \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_h^2\|\mathbf{x}\|^2+\sigma_w^2} + N\log\left(\sigma_h^2\|\mathbf{x}\|^2+\sigma_w^2\right) + \text{const.} \end{aligned} \tag{20}$$

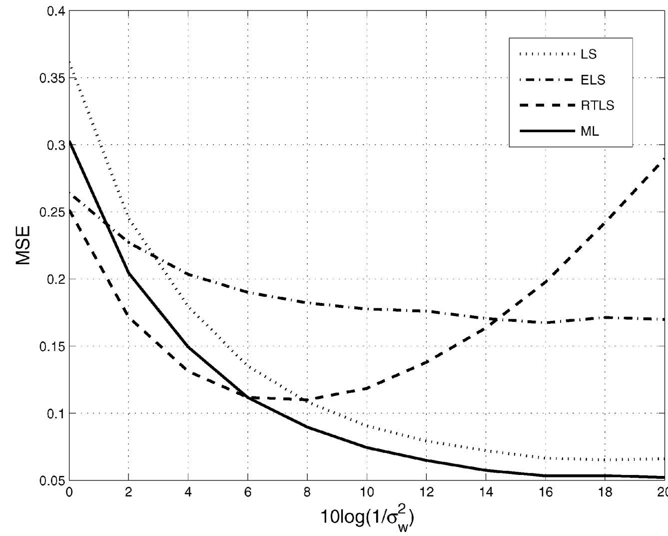

Fig. 1. Comparison of the ML, LS, ELS, and TRLS estimators for $\sigma_h^2 = 0.05$.

This, along with the monotonicity of the logarithm, proves our claim. Thus, the ELS estimator can also be considered as an approximation of the ML estimator.

IV. 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 letter. The parameters in our simulation were $N = 10 \cdot 4$ and $K = 4$. The matrix $\mathbf{H}$ was chosen as a concatenation of ten 4×4 matrices with unit diagonal elements and 0.5 off-diagonal elements. At each realization, $\mathbf{x}$ was randomly generated with independent, equiprobable ± 1 Bernoulli random variables. We estimated the MSEs of each estimator using 10 000 computer simulation. For comparison, we provide the results for the ML estimator of (3), the standard LS estimator, the expected LS estimator of (18), and the RTLS estimator of (16). The results of the TLS estimator were significantly worse than the other estimators and are therefore omitted. The results are presented in Figs. 1 and 2 for variances $\sigma_h^2 = 0.05$ and $\sigma_h^2 = 0.2$, respectively. 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. In addition, when the uncertainty is low (see Fig. 1), then LS works pretty well but gets worse as the uncertainty grows.

V. Conclusion

In this letter, we considered the problem of estimating $\mathbf{x}$ in the model $\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w}$ when $\mathbf{G}$ is Gaussian. We derived the ML estimator 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.

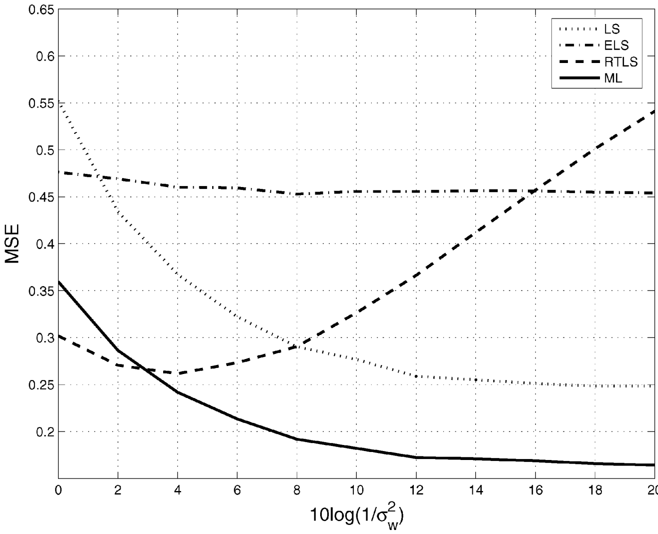

Fig. 2. Comparison of the ML, LS, ELS, and TRLS estimators for $\sigma_h^2 = 0.2$.

Our results motivate the continuing research on this seemingly simple estimation problem. There are still many open questions. The method presented here can be readily extended to the case in which $\mathbf{G}$ has independent rows with a common covariance matrix or the case in which some of the columns in $\mathbf{G}$ are known. The more general case cannot be handled directly using our techniques and is therefore an interesting topic for further research. Another important extension 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 $t = 1, \ldots, T$ and $\mathbf{G}$ is random.

APPENDIX

In this Appendix, we provide the proof of Theorem 1. The main argument of the theorem is obtained by introducing a slack variable $t = \|\mathbf{x}\|^2$ and rewriting (3) as in (5) with $f(t)$ defined in (4). It remains to prove that (5) is unimodal in $t \geq 0$.

First, we will show that $f(t)$ is convex in $t \geq 0$. In [14] and [16], it was shown that strong duality holds in this special case and that $f(t)$ is equal to the value of its dual program

$$f(t) = \begin{cases} \max_\alpha & \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}\left(\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}\right)^\dagger \mathbf{H}^T\mathbf{y} - \alpha t \\ \text{s.t.} & \mathbf{H}^T\mathbf{H} + \alpha\mathbf{I} \succeq \mathbf{0} \\ & \mathbf{H}^T\mathbf{y} \in \mathcal{R}\left(\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}\right). \end{cases} \tag{21}$$

Thus, $f(t)$ is the pointwise maximum of a family of affine functions of t and therefore is convex in $t \geq 0$.

Next, we will show that

$$r(t) = \frac{f(t)}{\sigma_h^2 t + \sigma_w^2} + N \log\left(\sigma_h^2 t + \sigma_w^2\right) \tag{22}$$

is unimodal in $t \geq 0$. We use the following result from [16]. If $r'(t) = 0$ implies $r''(t) > 0$ for any $t \geq 0$, then $r(t)$ is unimodal in $t \geq 0$. The condition $r'(t) = 0$ states that

$$r'(t) = \frac{f'(t)}{\sigma_h^2 t + \sigma_w^2} - \frac{\sigma_h^2 f(t)}{\left(\sigma_h^2 t + \sigma_w^2\right)^2} + \frac{N\sigma_h^2}{\sigma_h^2 t + \sigma_w^2} = 0. \tag{23}$$

Multiplying by $\sigma_h^2/(\sigma_h^2 t + \sigma_w^2)$ yields

$$\frac{\sigma_h^4 f(t)}{\left(\sigma_h^2 t + \sigma_w^2\right)^3} = \frac{N\sigma_h^4}{\left(\sigma_h^2 t + \sigma_w^2\right)^2} + \frac{\sigma_h^2 f'(t)}{\left(\sigma_h^2 t + \sigma_w^2\right)^2}. \tag{24}$$

The second derivative is

$$r''(t) = \frac{f''(t)}{\sigma_h^2 t + \sigma_w^2} - \frac{\sigma_h^2 f'(t)}{\left(\sigma_h^2 t + \sigma_w^2\right)^2} - \frac{\sigma_h^2 f'(t)}{\left(\sigma_h^2 t + \sigma_w^2\right)^2} + \frac{2\sigma_h^4 f(t)}{\left(\sigma_h^2 t + \sigma_w^2\right)^3} - \frac{N\sigma_h^4}{\left(\sigma_h^2 t + \sigma_w^2\right)^2}. \tag{25}$$

Plugging in the left-hand side of (24) yields

$$r''(t) = \frac{f''(t)}{\sigma_h^2 t + \sigma_w^2} + \frac{N\sigma_h^4}{\left(\sigma_h^2 t + \sigma_w^2\right)^2} \tag{26}$$

when $r'(t) = 0$. Now, $f(t)$ is convex, which means that $f''(t) \geq 0$. Therefore, the first term of $r''(t)$ is non-negative. The second term is positive since $\sigma_w^2 > 0$ and $\sigma_h^2 > 0$. This concludes the proof.

REFERENCES

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

[2] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill Posed Problems*. Washington, DC: Winston and Sons, 1977.

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

[4] W. A. Fuller, *Measurement Error Models*. New York: Wiley, 1987, Series in Probability and Mathematical Statistics.

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

[6] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991, vol. 9, Frontiers in Applied Mathematics.

[7] G. Golub and P. Hansen, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 1, pp. 185–194, Oct. 1999.

[8] A. Beck and A. Ben-Tal, "On the solution of the Tikhonov regularization of the total least squares problem," *SIAM J. Optim.*, to be published.

[9] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM J. Matrix Anal. Appl.*, pp. 1034–1064, Oct. 1997.

[10] S. Chandrasekaran, M. Gu, G. H. Golub, and A. H. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, pp. 235–252, Jan. 1998.

[11] 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. American Control Conf.*, Jun. 1998.

[12] Y. C. Eldar, "Minimax estimation of deterministic parameters in linear models with a random model matrix," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 601–612, Feb. 2006.

[13] G. E. Forsythe and G. H. Golub, "On the stationary values of a second-degree polynomial on the unit sphere," *J. Soc. Ind. Appl. Math.*, vol. 13, no. 4, pp. 1050–1068, 1965.

[14] Y. C. Eldar and A. Beck, "Hidden convexity based near maximum-likelihood CDMA detection," in *Proc. Conf. Information Sciences Systems*, Princeton, NJ, Mar. 2005.

[15] C. Fortin and H. Wolkowicz, "The Trust Region Subproblem and Semidefinite Programming,", Can. Res. Rep. CORR 2002-22.

[16] S. Boyd and L. Vandenberghe, *Introduction to Convex Optimization with Engineering Applications*. Stanford, CA: Stanford Univ. Press, 2003.