Digital Signal Processing **11**, 80–90 (2001)
doi:10.1006/dspr.2000.0371, available online at http://www.idealibrary.com on IDEAL®

Linear Regression Constrained to a Ball [1]

Petre Stoica and Girish Ganesan[2]

Systems and Control Group, P.O. Box 27, S-751 03, Uppsala, Sweden

Published online November 7, 2000

Stoica, P., and Ganesan, G., Linear Regression Constrained to a Ball, *Digital Signal Processing* **11** (2001), 80–90.

A worst case lower bound (WCLB) result obtained by Nemirovskii suggests that a potentially significant estimation accuracy enhancement may be achieved provided the true parameter vector is known to belong to a ball. In this paper we discuss the many facets and implications of Nemirovskii's result by using linear regression as a vehicle for illustration. In particular, we address briefly such issues as biased versus unbiased estimation, minimax optimal estimation, tightness of the WCLB, and comparison of WCLB with the performance of the least squares estimator constrained to the ball and that of the linear minimax estimator.

Key Words: linear regression; minimax estimation; biased estimation; Nemirovskii's worst case lower bound; constrained parameter estimation.

1. INTRODUCTION

Let θ denote the unknown parameter vector in a given estimation problem and let J be the corresponding Fisher information matrix defined in the usual way. (See, e.g., [2, 4].) Assume that

$$\|\theta\| \leq \rho \tag{1}$$

for a given $\rho > 0$. Henceforth $\|\cdot\|$ denotes the Euclidean norm for vectors as well as the two-norm (also called spectral norm) for matrices. Let

$$\text{MSE}_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \tag{2}$$

denote the mean-squared error (MSE) of an estimate $\hat{\theta}$ of θ. Whenever $\hat{\theta}$ is an unbiased estimate, (2) above is simply the covariance matrix of $\hat{\theta}$. However, $\hat{\theta}$ is *not* constrained to be unbiased in what follows.

[1] This work was partly supported by the Senior Individual Grant Program of the Swedish Foundation for Strategic Research.

[2] All correspondence may be addressed to Girish Ganesan. E-mail: Girish.Ganesan@syscon.uu.se.

1051-2004/01 $35.00

Let

$$\mathrm{MSE}_{\sup} = \sup_{\|\theta\| \leq \rho} \mathrm{MSE}_{\hat{\theta}}(\theta) \tag{3}$$

denote the MSE matrix corresponding to the worst parameter vector(s) in the ball. The sup operation in (3) is defined with respect to the ordering relationship of positive (semi)definite matrices. Finally, let

$$\mathrm{WCLB} = \frac{\inf_{\|\theta\| \leq \rho} J^{-1}}{\left(1 + \sup_{\|\theta\| \leq \rho} \frac{\|J^{-1/2}\|}{\rho}\right)^2}. \tag{4}$$

Using the previous notations we can state Nemirovskii's worst-case lower bound (WCLB) result (that holds under regularity conditions; see [2] for details) as follows:

$$\mathrm{MSE}_{\sup} \geq \mathrm{WCLB}. \tag{5}$$

Does this result have any potential applications in the signal processing area? Should we really be interested in a worst-case lower bound result, such as (5) above? Is the WCLB in (4) achievable? Is there any guideline about how to obtain an estimate that achieves WCLB or at least gets close to it? These are the types of questions that we address briefly in the following sections by making use of the linear regression problem as a vehicle for illustration.

2. CRITICAL DISCUSSION OF WCLB AND RELATED ISSUES

For the sake of clarity we restrict the following discussion to linear regression, even though most of the comments made in this section are valid for general estimation problems as well.

Consider the liner regression model

$$y = \Phi\theta + e, \tag{6}$$

where $y \in \mathbb{R}^{N \times 1}$ and $\Phi \in \mathbb{R}^{N \times n}$ are given, θ is the parameter vector to be estimated, and e is a noise term assumed to be Gaussian distributed with mean zero and covariance matrix equal to $\sigma^2 I$. Let

$$R = \Phi^T \Phi; \qquad r = \Phi^T y \tag{7}$$

and assume that R is a nonsingular matrix. The least-squares estimate (LSE) of θ in (6) is given by (see, e.g., [4]):

$$\hat{\theta} = R^{-1} r. \tag{8}$$

It is well known that, under the assumptions made above, the LSE is Gaussian distributed with mean equal to θ and covariance matrix equal to J^{-1}, where the