# COVARIANCE SHAPING APPROACH TO LINEAR LEAST-SQUARES ESTIMATION

*Yonina C. Eldar*

Department of Electrical Engineering
Technion–Israel Institute of Technology
Haifa, 32000, Israel

**ABSTRACT**

*A new biased linear estimator, referred to as the covariance shaping least-squares (CSLS) estimator, is proposed for estimating a set of unknown deterministic parameters in a linear model. The CSLS estimator is directed at improving the performance of the least-squares (LS) estimator by choosing the estimate to minimize the error variance in the observations subject to a constraint on its covariance. The CSLS estimator is shown to achieve the Cramer-Rao bound for biased estimators. Furthermore, the covariance of the estimate can be chosen such that there is a threshold SNR, below which the CSLS estimator yields a lower MSE than the LS estimator, for all values of the parameters.*

## 1. INTRODUCTION

A generic estimation problem that has been studied extensively in the literature is that of estimating the unknown deterministic parameters $\mathbf{x}$ in the linear model

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

where $\mathbf{H}$ is a known $n \times m$ matrix, and $\mathbf{w}$ is a zero-mean random vector with positive definite covariance $\mathbf{C}_w$. For simplicity of exposition we assume that rank$(\mathbf{H}) = m$.

A common approach to estimating $\mathbf{x}$ is to find the linear estimate of $\mathbf{x}$ that results in an estimated data vector $\hat{\mathbf{y}}$ that is as close as possible to the given data vector $\mathbf{y}$ in a (weighted) least-squares (LS) sense. Thus the LS estimate of $\mathbf{x}$, denoted $\hat{\mathbf{x}}_{\text{LS}}$, is chosen such that $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}_{\text{LS}} = \mathbf{H}\mathbf{G}\mathbf{y}$ minimizes the total squared error

$$\varepsilon_{\text{LS}} = (\mathbf{y} - \mathbf{H}\mathbf{G}\mathbf{y})^* \mathbf{A} (\mathbf{y} - \mathbf{H}\mathbf{G}\mathbf{y}), \tag{2}$$

where $\mathbf{A}$ is an arbitrary positive definite weighting matrix. If we choose $\mathbf{A} = \mathbf{C}_w^{-1}$, then the LS estimate is given by

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}_w^{-1} \mathbf{y}. \tag{3}$$

The Gauss-Markov theorem [1] states that with $\mathbf{A} = \mathbf{C}_w^{-1}$, the LS estimator minimizes the total variance defined by $V(\hat{\mathbf{x}}) = E((\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))^*(\hat{\mathbf{x}} - E(\hat{\mathbf{x}})))$ from all linear *unbiased* estimators. Furthermore, if $\mathbf{w}$ is a zero-mean Gaussian random vector, then the LS estimator (with optimal weighting) acheives the Cramer-Rao lower bound (CRLB) for unbiased estimators [1, 2], so that it minimizes the total variance from all linear and nonlinear *unbiased* estimators.

The LS estimator has a variety of optimality properties in the class of unbiased estimators. However, an unbiased estimator does not necessarily lead to minimum mean-squared error (MSE), where the MSE of an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ is defined by

$$\text{MSE}(\hat{\mathbf{x}}) = E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = V(\hat{\mathbf{x}}) + \|B(\hat{\mathbf{x}})\|^2. \tag{4}$$

Here $B(\hat{\mathbf{x}}) = E(\hat{\mathbf{x}}) - \mathbf{x}$ denotes the bias of the estimator $\hat{\mathbf{x}}$.

The LS method is widely employed in diverse fields, both as an estimation criterion and as a method for parametric modelling of data (see *e.g.,* [1, 3]). Various modifications of the LS estimator for the case in which the model (1) is assumed to hold perfectly have been proposed [4]. Among the more prominent alternatives are the ridge estimator [5] (also known as Tikhonov regularization [6]) and the shrunken estimator [7].

In our method we assume that the data model holds *i.e.,* $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ with $\mathbf{H}$ and $\mathbf{y}$ known exactly, and our objective is to minimize the error between $\mathbf{x}$ and the estimate of $\mathbf{x}$. In many cases the data vector $\mathbf{y}$ is not very sensitive to changes in $\mathbf{x}$, so that a large error in estimating $\mathbf{x}$ may translate into a small error in estimating $\mathbf{y}$, in which case the LS estimate may result in a poor estimate of $\mathbf{x}$. This effect is especially predominant at low to moderate SNR, where the data vector $\mathbf{y}$ is typically affected more by the noise than by changes in $\mathbf{x}$; the exact SNR range will depend on the properties of the model matrix $\mathbf{H}$. A difficulty often encountered in this estimation problem is that the error in the estimation can have a covariance structure with a very high dynamic range.

To improve the performance over the LS estimator at low to moderate SNR we choose the estimator of $\mathbf{x}$ to minimize the total error variance in the observations $\mathbf{y}$, subject to a constraint on the covariance of the error in the estimate of $\mathbf{x}$, so that we control the dynamic range and spectral shape of the covariance of the estimation error. The resulting estimator of $\mathbf{x}$ is referred to as the covariance shaping LS (CSLS) estimator, and is developed in Section 2. In Section 3 we show that both the ridge estimator and the shrunken estimator can be formulated as CSLS estimators.

In Section 4 we show that the CSLS estimator has a property analogous to the LS estimator. Specifically, it is shown to achieve the CRLB for biased estimators [1, 2, 8] when the noise is Gaussian. This implies that for Gaussian noise, there is no linear or nonlinear estimator with a smaller variance, or MSE, and the same bias as the CSLS estimator. In Section 5 we analyze the MSE in estimating $\mathbf{x}$ of both the CSLS estimator and the LS estimator, and show that in many cases the CSLS estimator can result in lower MSE than the LS estimator by allowing for a bias.

Based on the concept of CSLS estimation, in Section 6 we propose a new linear receiver for synchronous CDMA systems which we refer to as the *covaraicne shaping multiuser (CSMU) recvier*. The CSMU receiver depends only on the users' signatures and does not require knowledge of the channel parameters. Nonetheless, as we demonstrate, over a wide range of these parameters the performance of the CSMU receiver can approach the performance of the linear minimum mean-squared error (MMSE) receiver [9] which is the optimal linear receiver that assumes knowledge of the channel parameters.

## 2. THE COVARIANCE SHAPING LS ESTIMATOR

Since the bias of an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$, and consequently the MSE of (4), depend explicitly on the unknown parameters $\mathbf{x}$, $\hat{\mathbf{x}}$ cannot be chosen to directly minimize the MSE. A common approach is to restrict $\hat{\mathbf{x}}$ to be linear and unbiased, and then seek the estimator of this form that minimizes the variance or the MSE, which leads to the LS estimator. In our development, the estimator is not constrained to be unbiased.

The CSLS estimate of $\mathbf{x}$, denoted $\hat{\mathbf{x}}_{\mathrm{CSLS}}$, is chosen to minimize the total variance of the weighted error between $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}_{\mathrm{CSLS}} = \mathbf{HGy}$ and $\mathbf{y}$, subject to the constraint that the covariance of the error in $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ is proportional to a given covariance matrix $\mathbf{R}$. From (1) it follows that the covariance of $\mathbf{y}$ is equal to $\mathbf{C}_w$, so that the covariance of $\hat{\mathbf{x}}_{\mathrm{CSLS}}$, which is equal to the covariance of the error in $\hat{\mathbf{x}}_{\mathrm{CSLS}}$, is $\mathbf{G}\mathbf{C}_w\mathbf{G}^*$. Thus, $\hat{\mathbf{x}}_{\mathrm{CSLS}} = \mathbf{Gy}$ is chosen to minimize

$$\varepsilon_{\mathrm{CSLS}} = E\left((\mathbf{y}' - \mathbf{HGy}')^*\mathbf{C}_w^{-1}(\mathbf{y}' - \mathbf{HGy}')\right) \qquad (5)$$

where $\mathbf{y}' = \mathbf{y} - E(\mathbf{y})$, subject to

$$\mathbf{G}\mathbf{C}_w\mathbf{G}^* = c^2\mathbf{R}. \qquad (6)$$

Here $c > 0$ is a constant that is either specified, or chosen to minimize $\varepsilon_{\mathrm{CSLS}}$ of (5).

This minimization problem is a special case of the *weighted minimum mean-squared error (WMMSE) shaping problem* considered in [10]. Specifically, the problem of (5) and (6) can be restated as the problem of finding the transformation $\mathbf{W}$ to minimize

$$E\left((\mathbf{a} - \mathbf{b})^*\mathbf{C}_a^{-1}(\mathbf{a} - \mathbf{b})\right), \qquad (7)$$

where $\mathbf{b} = \mathbf{Wa}$, subject to

$$\mathbf{C}_b = \mathbf{W}\mathbf{C}_a\mathbf{W}^* = c^2\mathbf{Q}. \qquad (8)$$

The solution to the WMMSE shaping problem has been obtained in [10], and is incorporated in the following theorem:

**Theorem 1 (WMMSE covariance shaping)** *Let $\mathbf{a} \in \mathbb{C}^m$ be a random vector with positive definite covariance matrix $\mathbf{C}_a$. Let $\widehat{\mathbf{W}}$ be the optimal covariance shaping transformation that minimizes the weighted MSE defined by (7), between the input $\mathbf{a}$ and the output $\mathbf{b} = \mathbf{Wa}$ with covariance $\mathbf{C}_b = c^2\mathbf{Q}$ where $\mathbf{Q}$ is a given covariance matrix and $c > 0$. Then*

$$\widehat{\mathbf{W}} = \beta(\mathbf{Q}\mathbf{C}_a^{-1})^{1/2} = \beta\mathbf{Q}(\mathbf{C}_a^{-1}\mathbf{Q})^{1/2}\mathbf{C}_a^{-1},$$

*where*

1. *if $c$ is specified then $\beta = c$;*
2. *if $c$ is chosen to minimize the weighted MSE then $\beta = \hat{c}$ where $\hat{c} = Tr((\mathbf{C}_a^{-1}\mathbf{Q})^{1/2})/Tr(\mathbf{C}_a^{-1}\mathbf{Q})$.*

In the problem of (5), $\mathbf{a} = \mathbf{y}'$, $\mathbf{C}_a = \mathbf{C}_w$, $\mathbf{W} = \mathbf{HG}$, and $\mathbf{Q} = \mathbf{HRH}^*$. Denoting $\widetilde{\mathbf{G}} = (1/c)\mathbf{G}$ it follows from Theorem 1 that the optimal value of $\widetilde{\mathbf{G}}$, denoted $\widehat{\widetilde{\mathbf{G}}}$, satisfies

$$\mathbf{H}\widehat{\widetilde{\mathbf{G}}} = (\mathbf{HRH}^*\mathbf{C}_w^{-1})^{1/2}. \qquad (9)$$

Using straightforward matrix manipulations it can be shown that

$$\widehat{\widetilde{\mathbf{G}}} = (\mathbf{RH}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1/2}\mathbf{RH}^*\mathbf{C}_w^{-1}. \qquad (10)$$

If the scaling $c$ in (6) is specified, then the CSLS estimator is given by $\hat{\mathbf{x}}_{\mathrm{CSLS}} = c\widehat{\widetilde{\mathbf{G}}}\mathbf{y}$. If $c$ is chosen to minimize $\varepsilon_{\mathrm{CSLS}}$, then $\hat{\mathbf{x}}_{\mathrm{CSLS}} = \hat{c}\widehat{\widetilde{\mathbf{G}}}\mathbf{y}$, where from Theorem 1,

$$\hat{c} = \frac{\mathrm{Tr}((\mathbf{RH}^*\mathbf{C}_w^{-1}\mathbf{H})^{1/2})}{\mathrm{Tr}(\mathbf{RH}^*\mathbf{C}_w^{-1}\mathbf{H})}. \qquad (11)$$

Note that $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ is a biased estimator of $\mathbf{x}$, so that when $\sigma^2 \to 0$, $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ does not converge to $\mathbf{x}$. The advantage of the CSLS estimator is at low to moderate SNR, where we reduce the MSE of the estimator by allowing for a bias. Indeed, as we show in Section 5, for many choices of $\mathbf{R}$, regardless of the value of $\mathbf{x}$ there is always a threshold SNR, so that for SNR values below this threshold the CSLS estimator yields a lower MSE than the LS estimator. Examples considered in [10] indicate that in a variety of applications the threshold values are pretty large. Furthermore, simulations presented in [10] and in Sections 5 and 6 strongly suggest that the CSLS estimator can significantly decrease the MSE of the estimation error over the LS estimator for a wide range of SNR values.

The CSLS estimator with optimal scaling is invariant to an overall gain in $\mathbf{C}_w$. Thus if $\mathbf{C}_w = \sigma^2\mathbf{C}$ for some covariance matrix $\mathbf{C}$, then the CSLS estimator with optimal scaling does not depend on $\sigma$. This property does not hold in the case in which $c$ is chosen as a constant, independent of $\sigma$. In this case the CSLS estimator depends explicitly on $\sigma$ which therefore must be known. Alternatively, if we let $c = \sigma$, then the CSLS estimator will again not depend on $\sigma$.

The CSLS estimator is summarized in the following theorem:

**Theorem 2 (CSLS estimator)** *Let $\mathbf{x}$ denote the deterministic unknown parameters in the model $\mathbf{y} = \mathbf{Hx}+\mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank $m$, and $\mathbf{w}$ is a zero-mean random vector with positive definite covariance $\mathbf{C}_w$. Let $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ denote the covariance shaping least-squares estimator of $\mathbf{x}$ that minimizes the error (5) subject to (6), for some $c > 0$. Then*

$$\hat{\mathbf{x}}_{\mathrm{CSLS}} = \beta(\mathbf{RH}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1/2}\mathbf{RH}^*\mathbf{C}_w^{-1}\mathbf{y},$$

*where*

1. *if $c$ is specified then $\beta = c$;*
2. *if $c$ is chosen to minimize (5) then $\beta = \hat{c}$ given by (11).*

In [10] it is shown that the CSLS estimator can be expressed as an LS estimator followed by a WMMSE shaping transformation, that optimally shapes the covariance of the LS estimate of $\mathbf{x}$.

## 3. CONNECTION WITH OTHER LS MODIFICATIONS

We now compare the CSLS estimator with the ridge estimator proposed by Hoerl and Kennard [5], and Tikhonov [6], and with the shrunken estimator proposed by Mayer and Willke [7].

The ridge estimator, denoted by $\hat{\mathbf{x}}_{\mathrm{R}}$, is defined by

$$\hat{\mathbf{x}}_{\mathrm{R}} = (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \delta\mathbf{T})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \qquad (12)$$

where $\mathbf{T}$ is a positive definite matrix and $\delta$ is a regularization parameter. The ridge estimator is equal to a CSLS estimator with covariance $\mathbf{R}_{\mathrm{R}}$, where $\mathbf{R}_{\mathrm{R}}$ is the covariance of $\hat{\mathbf{x}}_{\mathrm{R}}$ and is given by

$$\mathbf{R}_{\mathrm{R}} = (\mathbf{I} + \delta(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{T})^{-1}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \delta\mathbf{T})^{-1}. \qquad (13)$$

Indeed, by direct substitution of (13) into the expression for $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ from Theorem 2, $\hat{\mathbf{x}}_{\mathrm{CSLS}} = \hat{\mathbf{x}}_{\mathrm{R}}$. Based on this connection between $\hat{\mathbf{x}}_{\mathrm{R}}$ and $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ we may interpret $\hat{\mathbf{x}}_{\mathrm{R}}$ as the estimator that minimizes the error $\varepsilon_{\mathrm{CSLS}}$ of (5) from all estimators with covariance $\mathbf{R}_{\mathrm{R}}$.

The shrunken estimator, denoted by $\hat{\mathbf{x}}_{\mathrm{S}}$, is a scaled version of the LS estimator and is defined by

$$\hat{\mathbf{x}}_{\mathrm{S}} = \kappa\hat{\mathbf{x}}_{\mathrm{LS}} = \kappa(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \qquad (14)$$

where $\kappa$ is a regularization parameter. A stochastically (nonlinear) shrunken estimator is a shrunken estimator in which $\kappa$ is a function of the data $\mathbf{y}$, an example of which is the well known James-Stein estimator [11]. The shrunken estimator $\hat{\mathbf{x}}_{\mathrm{S}}$ can be formulated as a CSLS estimator where the covariance of $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ is chosen to be equal to the covariance of $\hat{\mathbf{x}}_{\mathrm{S}}$ given by

$$\mathbf{R}_{\mathrm{S}} = \kappa^2(\mathbf{H}^*\mathbf{C}_w\mathbf{H})^{-1}. \qquad (15)$$

Substituting (15) into the expression for $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ from Theorem 2, we have indeed that $\hat{\mathbf{x}}_{\mathrm{CSLS}} = \hat{\mathbf{x}}_{\mathrm{S}}$. Thus, we may interpret $\hat{\mathbf{x}}_{\mathrm{S}}$ as the estimator that minimizes the error $\varepsilon_{\mathrm{CSLS}}$ of (5) from all estimators with covariance $\mathbf{R}_{\mathrm{S}}$.

In summary, some of the popular alternatives to the LS estimator under the model (1) are in fact CSLS estimators. This provides additional insight and further optimality properties of these estimators. However, the CSLS estimator is more general since we are not constrained to a specific choice of covariance $\mathbf{R}$. By choosing $\mathbf{R}$ to "best" shape the estimator covariance in some sense we can improve the performance over these LS alternatives.

As a final note, suppose we are given an arbitrary linear estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ with covariance $\mathbf{C}_x$. Then we can compute the CSLS estimate $\hat{\mathbf{x}}_{\mathrm{CSLS}}$ with $\mathbf{R} = \mathbf{C}_x$. If $\hat{\mathbf{x}}_{\mathrm{CSLS}} = \hat{\mathbf{x}}$, then the estimate $\hat{\mathbf{x}}$ has the additional property that from all estimators with covariance $\mathbf{C}_x$ it minimizes the (weighted) total error variance in the observations. If, on the other hand, $\hat{\mathbf{x}}_{\mathrm{CSLS}} \neq \hat{\mathbf{x}}$, then we can always improve the total error variance of the estimate without altering its covariance by using $\hat{\mathbf{x}}_{\mathrm{CSLS}}$.

## 4. CRAMER-RAO LOWER BOUND

The variance of an unbiased estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ can be bounded by the Cramer-Rao lower bound (CRLB) [1, 2]. A similar bound is also given for the variance of a biased estimator, which is known as the biased CRLB [8]. Specifically, suppose we want to estimate a set of unknown deterministic parameters $\mathbf{x}$ from given observations $\mathbf{y}$. Let $p(\mathbf{y},\mathbf{x})$ denote the probability density function of $\mathbf{y}$ characterized by $\mathbf{x}$. It is assumed that $p(\mathbf{y},\mathbf{x})$ satisfies the regularity condition $E(\partial p(\mathbf{y},\mathbf{x})/\partial\mathbf{x}) = 0$. Then for any estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ with bias $B(\mathbf{x})$, the covariance of the estimator must satisfy

$$E((\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))^*) \geq \left(\mathbf{I}_m + \frac{\partial B(\mathbf{x})}{\partial \mathbf{x}}\right) J^{-1}(\mathbf{x}) \left(\mathbf{I}_m + \frac{\partial B(\mathbf{x})}{\partial \mathbf{x}}\right)^*, \qquad (16)$$

where $J(\mathbf{x})$ is the Fisher information matrix defined by

$$J(\mathbf{x}) = E\left(\frac{\partial^2 \log p(\mathbf{y},\mathbf{x})}{\partial \mathbf{x}^2}\right). \qquad (17)$$

For the CSLS estimator, the bias is given by

$$B(\hat{\mathbf{x}}_{\mathrm{CSLS}}) = (\beta(\mathbf{R}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{1/2} - \mathbf{I}_m)\mathbf{x}, \qquad (18)$$

and

$$\frac{\partial B(\hat{\mathbf{x}}_{\mathrm{CSLS}})}{\partial \mathbf{x}} = \beta(\mathbf{R}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{1/2} - \mathbf{I}_m. \qquad (19)$$

We now show that if the noise $\mathbf{w}$ in (1) is Gaussian with zero-mean and covariance $\mathbf{C}_w$, then the CSLS estimator achieves the CRLB for biased estimators $\hat{\mathbf{x}}$ with bias $B(\hat{\mathbf{x}})$ given by (18).

For the linear model (1) with Gaussian noise the Fisher information matrix is [1] $J(\mathbf{x}) = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. The CRLB on the variance of any estimator with bias $B(\hat{\mathbf{x}}_{\mathrm{CSLS}})$ is therefore

$$\begin{aligned} & E((\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))^*) \\ \geq \;& \beta^2(\mathbf{R}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{1/2}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{R})^{1/2} \\ = \;& \beta^2\mathbf{R}. \end{aligned} \qquad (20)$$

Now, for the CSLS estimator, $E(\hat{\mathbf{x}}_{\mathrm{CSLS}} - E(\hat{\mathbf{x}}_{\mathrm{CSLS}}))(\hat{\mathbf{x}}_{\mathrm{CSLS}} - E(\hat{\mathbf{x}}_{\mathrm{CSLS}}))^*) = \beta^2\mathbf{R}$, so that the CRLB is achieved. Thus, from all linear and nonlinear estimators with bias given by (18) for some $\beta$ and $\mathbf{R}$, the CSLS estimator minimizes the variance.

## 5. MEAN-SQUARED ERROR PERFORMANCE

In the previous section we showed that the CSLS estimator minimizes the MSE among all estimators with a particular bias. While it would be desirable to analyze the MSE of the CSLS estimator for more general forms of bias, we cannot directly evaluate the MSE of the CSLS estimator since the bias, and consequently the MSE, depend explicitly on the unknown parameters $\mathbf{x}$. To gain some additional insight into the performance of the CSLS estimator, in this section we instead compare its MSE with the MSE of the LS estimator. Our analysis indicates that there are many cases in which the CSLS estimator performs better than the LS estimator in a MSE sense, for all values of the unknown parameters $\mathbf{x}$.

In our analysis we assume that $\mathbf{C}_w = \sigma^2\mathbf{C}$, where the diagonal elements of $\mathbf{C}$ are all equal to 1, so that the variance of each of the noise components of $\mathbf{C}_w$ is $\sigma^2$. To ensure that the estimator does not depend on $\sigma$, which may not be known, we let the scaling of the CSLS estimator be $\beta = \sigma$ or $\beta = \hat{c}$ given by (11).

### 5.1. Fixed Scaling

We first consider the case in which $\beta = \sigma$. The MSE of the CSLS estimator is then given by

$$\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}}) = \|((\mathbf{R}\mathbf{B})^{1/2} - \mathbf{I}_m)\mathbf{x}\|^2 + \sigma^2\mathrm{Tr}(\mathbf{R}), \qquad (21)$$

where $\mathbf{B} = \mathbf{H}^*\mathbf{C}^{-1}\mathbf{H}$. The first term is the squared norm of the bias of $\hat{\mathbf{x}}_{\mathrm{CSLS}}$, and the second term is the total variance of $\hat{\mathbf{x}}_{\mathrm{CSLS}}$.

For large values of $\sigma^2$ in comparison with $\|\mathbf{x}\|^2$, the first term in (21) is negligible and $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}}) \approx \sigma^2\mathrm{Tr}(\mathbf{R})$. Thus, at sufficiently low SNR, where the SNR is defined as $\|x\|^2/\sigma^2$, both $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{LS}})$ and $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}})$ are proportional to $\sigma^2$ where we can always choose $\mathbf{R}$ so that the proportionality constant $\mathrm{Tr}(\mathbf{R})$ of the CSLS estimator is smaller than the proportionality constant $\mathrm{Tr}(\mathbf{B}^{-1})$ of the LS estimator. At sufficiently high SNR, the second term in (21) can be considered negligible and as $\sigma \to 0$, $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}})$ converges to the constant $\|((\mathbf{R}\mathbf{B})^{1/2} - \mathbf{I}_m)\mathbf{x}\|^2$. From this qualitative analysis it is clear that there is a threshold SNR that will depend in general on $\mathbf{x}$ below which, for appropriate choices of $\mathbf{R}$, the CSLS estimator outperforms the LS estimator.

It can be shown that $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}}) \leq \mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{LS}})$ if

$$\zeta \leq \frac{\mathrm{Tr}(\mathbf{B}^{-1}) - \mathrm{Tr}(\mathbf{R})}{\sigma_\gamma} \triangleq \zeta_{\mathrm{WC}}, \qquad (22)$$

where $\zeta = \|\mathbf{x}\|^2/(\sigma^2 m)$ denotes the SNR per component, $\gamma = \arg\max \sigma_i$, and $\sigma_i$ are the eigenvalues of $\mathbf{Q} = ((\mathbf{R}\mathbf{B})^{1/2} - \mathbf{I}_m)^*((\mathbf{R}\mathbf{B})^{1/2} - \mathbf{I}_m)$. The bound $\tilde{\zeta}_{\mathrm{WC}}$ given by (22) is a worst case bound, since it corresponds to the worst possible choice of parameters, namely when the unknown vector $\mathbf{x}$ is in the direction of the eigenvector of $\mathbf{Q}$ corresponding to the eigenvalue $\sigma_\gamma$. In practice the CSLS estimator will outperform the LS estimator for higher values of SNR than $\tilde{\zeta}_{\mathrm{WC}}$.

Since we have freedom in designing $\mathbf{R}$, we may always choose $\mathbf{R}$ so that $\tilde{\zeta}_{\mathrm{WC}} > 0$. In this case we are guaranteed that there is a range of SNR values for which the CSLS estimator leads to a lower MSE than the LS estimator for all choices of $\mathbf{x}$.

For example, suppose we wish to design an estimator with covariance proportional to some covariance matrix $\mathbf{Z}$, so that $\mathbf{R} = a\mathbf{Z}$ for some $a > 0$. If we choose $a < \mathrm{Tr}(\mathbf{B}^{-1})/\mathrm{Tr}(\mathbf{Z})$, then there is an SNR range for which the CSLS estimator will have a lower MSE than the LS estimator for all values of $\mathbf{x}$. In specific applications it may not be obvious how to choose a particular proportionality factor $a$. In such cases, we may prefer using the CSLS estimator with optimal scaling, which we now discuss.

### 5.2. Optimal Scaling

In cases in which there is no natural scaling, it may be preferable to use the CSLS estimator with optimal scaling. In this case, the scaling is a function of $\mathbf{R}$ and therefore cannot be chosen arbitrarily, so that in general we can no longer guarantee that there is always an SNR range over which the CSLS performs better than the LS estimator. However, as we show, in the special case in which $\mathbf{R} = \mathbf{I}_m$, there is always such an SNR range.

If $\beta = \hat{c}$ and $\mathbf{R} = \mathbf{I}_m$, then

$$\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}}) = \|(\alpha\mathbf{B}^{1/2} - \mathbf{I}_m)\mathbf{x}\|^2 + m\alpha^2\sigma^2, \qquad (23)$$

where

$$\alpha = \frac{\hat{c}}{\sigma} = \frac{\mathrm{Tr}(\mathbf{B}^{1/2})}{\mathrm{Tr}(\mathbf{B})} = \frac{\sum_{i=1}^m \lambda_i^{1/2}}{\sum_{i=1}^m \lambda_i}, \qquad (24)$$

and $\lambda_i, 1 \le i \le m$ denote the eigenvalues of $\mathbf{B} = \mathbf{H}^*\mathbf{C}^{-1}\mathbf{H}$. It then follows that $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}}) \le \mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{LS}})$ if

$$\zeta \le \frac{(1/m)\sum_{i=1}^m \lambda_i^{-1} - \alpha^2}{|\alpha\lambda_\gamma^{1/2} - 1|^2} \triangleq \zeta_{\mathrm{WC}}, \qquad (25)$$

where $\gamma = \arg\max |\alpha\lambda_i^{1/2} - 1|^2$. Furthermore, when $\hat{\mathbf{x}}_{\mathrm{CSLS}} \neq \hat{\mathbf{x}}_{\mathrm{LS}}$, it can be shown that $\zeta_{\mathrm{WC}} > 0$ so that there is always a range of SNR values for which $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{CSLS}}) \le \mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{LS}})$.

The bound $\zeta_{\mathrm{WC}}$ given by (25) is again a worst case bound, since it corresponds to the worst possible choice of parameters. In practice the CSLS estimator will outperform the LS estimator for higher values of SNR than $\zeta_{\mathrm{WC}}$. In [10] we consider some examples illustrating the threshold values for different matrices $\mathbf{B}$. These examples indicate that in a variety of applications the threshold values are pretty large, as can also be seen from Figs. 1–3.

Simulations presented in [10] strongly suggest that the CSLS estimator can significantly decrease the MSE of the estimation error over the LS estimator for a wide range of SNR values. In Fig. 1 we illustrate the performance advantage with one simulation from [10]. In this figure we plot the MSE in estimating a set of AR parameters in an ARMA model contaminated by white noise, using the CSLS estimator with $\mathbf{R} = \mathbf{I}_m$ and optimal scaling and the LS estimator, from 20 noisy observations of the channel averaged over 2000 noise realizations, as a function of $-10\log\sigma^2$ where $\sigma^2$ is the noise variance. As can be seen from the figure, in this example the CSLS estimator significantly outperforms the LS estimator.

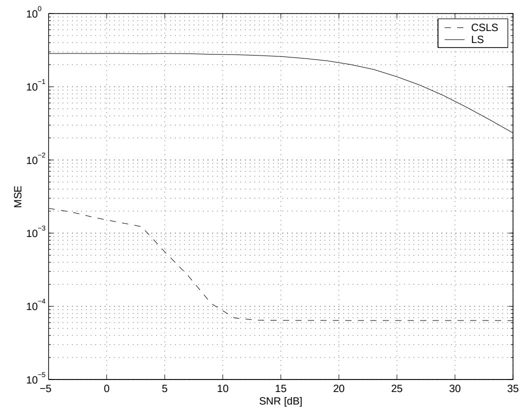


**Fig. 1**. **Mean-squared error in estimating a set of AR parameters using the LS estimator and the CSLS estimator.**

## 6. APPLICATION TO MULTIUSER DETECTION

Based on the concept of CSLS estimation, we now propose a new linear receiver for CDMA systems, which we refer to as the *covariance shaping multiuser (CSMU) receiver*. The CSMU receiver can be viewed as a decorrelator receiver [12] followed by an WMMSE covariance shaping transformation, that optimally shapes the covariance of the decorrelator output prior to detection. The CSMU receiver depends only on the users' signatures and does not require knowledge of the channel parameters. Nonetheless, over a wide range of these parameters its performance can approach the performance of the linear MMSE receiver, that assumes knowledge of these parameters. Simulations suggest that the CSMU receiver often leads to improved performance over the decorrelator and matched filter (MF) receivers. This receiver generalizes the recently proposed orthogonal multiuser receiver [13, 14].

The discrete-time model for the received signal $\mathbf{y}$ in a CDMA system is $\mathbf{y} = \mathbf{S}\mathbf{A}\mathbf{b} + \mathbf{w}$, where $\mathbf{S}$ is the matrix of columns $\mathbf{s}_i$ with $\mathbf{s}_i$ being the signature vector of the $i$th user, $\mathbf{A}$ is the diagonal matrix with diagonal elements $A_i > 0$ with $A_i$ being the received amplitude of the $i$th user's signal, $\mathbf{b}$ is the data vector with components $b_i \in \{1, -1\}$ with $b_i$ being the $i$th user's transmitted symbol, and $\mathbf{w}$ is a noise vector whose elements are independent $\mathcal{CN}(0, \sigma^2)$. We assume for simplicity that the vectors $\mathbf{s}_i$ are linearly independent. Based on the observed signal $\mathbf{y}$, we design a receiver to detect the information transmitted by each user. We restrict our attention to linear receivers that do not require knowledge of the received amplitudes $A_i$ or the noise level $\sigma^2$. The receiver estimates the vector $\mathbf{x} = \mathbf{A}\mathbf{b}$ as $\hat{\mathbf{x}} = \mathbf{Q}^*\mathbf{y}$ for some matrix $\mathbf{Q}$. The $i$th user's symbol is then detected as $\hat{b}_i = \mathrm{sgn}(\hat{x}_i)$ where $\hat{x}_i = \mathbf{q}_i^*\mathbf{y}$ is the $i$th component of $\hat{\mathbf{x}}$, and $\mathbf{q}_i$ are the columns of $\mathbf{Q}$.

If we estimate $\mathbf{x}$ using the LS estimator, then the resulting receiver is equal to the well known decorrelator receiver, introduced by Lupas and Verdu [12]. The decorrelator optimally rejects the multiple-access interference (MAI) but does not compensate for the white noise. Alternatively, we may estimate $\mathbf{x}$ using the MF estimator, $\hat{\mathbf{x}}_{\mathrm{MF}} = \mathbf{S}^*\mathbf{y}$. The resulting receiver is equivalent to the single-user MF receiver. The MF receiver optimally compensates for the white noise on the channel, but it does not take the structure of the MAI into account.

The CSMU receiver consists of a CSLS estimator of $\mathbf{Ab}$ followed by detection. Thus the CSMU receiver cross-correlates the received vector $\mathbf{y}$ with each of the columns $\mathbf{q}_i$ of $\mathbf{Q} = \mathbf{SR}(\mathbf{S}^*\mathbf{SR})^{-1/2}$. The $i$th users' bit is then detected as $\hat{b}_i = \mathrm{sgn}\,(\mathbf{q}_i^*\mathbf{y})$. The choice of shaping $\mathbf{R}$ can be tailored to the specific set of signatures. Alternatively, we may view the CSMU receiver as a decorrelator receiver followed by an WMMSE covariance shaping transformation that optimally shapes the covariance of the outputs of the decorrelator prior to detection, so that it compensates for the noise enhancement of the decorrelator. Therefore, in contrast to the MF and the decorrelator, the CSMU receiver takes both the background noise and the MAI into account.

To demonstrate the performance advantage in using the CSMU receiver, we consider the case in which the signature vectors are chosen as PN sequences with norm 1 and equal inner products $-1/N$, and the shaping $\mathbf{R}$ is chosen as a circulant matrix with $[\mathbf{R}]_{ii} = 1$ and $[\mathbf{R}]_{ij} = \rho$ for $i \neq j$.

In Fig. 2 we plot the theoretical probability of bit error of the CSMU receiver in the case of 5 users with $\rho = 0.2$, where the first user, the desired user, has 4 interferers such that $A_i/A_1 = 0.5$ for $i = 2, 3, 4, 5$. The corresponding curves for the decorrelator, MF and linear MMSE receivers are plotted for comparison. We see that the CSMU receiver performs better than the decorrelator and the MF and performs similarly to the linear MMSE receiver, even though it does not rely on knowledge of the channel parameters.

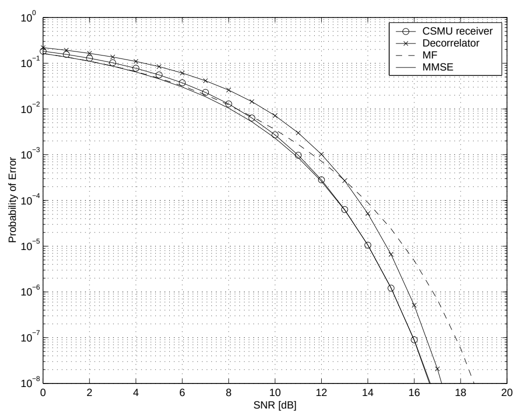


**Fig. 2**. **Probability of bit error with $5$ users and $\rho = 0.2$, as a function of SNR. The amplitude $A_1$ of the desired user is $2$ times greater than the amplitude $A_i$ of any of the other interferers.**

In Fig. 3 we evaluate the probability of bit error of the CSMU receiver in the case of 10 users with $\rho = 0.35$, and with accurate power control so that $A_i = 1$ for all $i$. Here again, the CSMU receiver performs better than the decorrelator and the MF and performs similarly to the linear MMSE receiver.

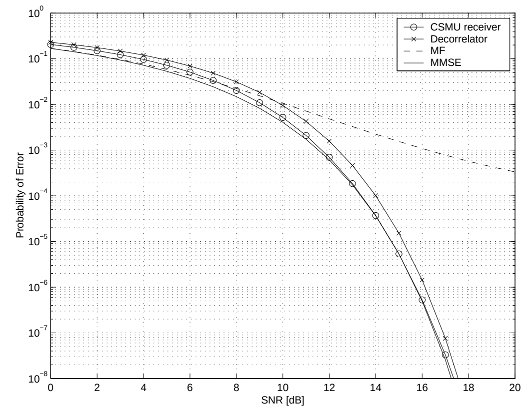


**Fig. 3**. **Probability of bit error with $10$ users, $\rho = 0.35$, and accurate power control, as a function of SNR.**

## 7. REFERENCES


[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Upper Saddle River, NJ: Prentice Hall, Inc., 1993.

[2] C. R. Rao, "Minimum variance and the estimation of several parameters," *Proc. Cambridge Phil. Soc.*, pp. 280–283, 1946.

[3] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1992.

[4] M. H. J. Gruber, *Regression Estimators: A Comparative Study*, San Diego, CA: Academic Press, Inc., 1990.

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

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

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

[8] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Willey and Sons, Inc., 1968.

[9] S. Verdu, *Multiuser Detection*, Cambridge, UK: Cambridge Univ. Press, 1998.

[10] Y. C. Eldar, *Quantum Signal Processing*, Ph.D. thesis, Massachusetts Institute of Technology, Dec. 2001, Available: `http://allegro.mit.edu/dspg/publications/TechRep/index.html`.

[11] W. James and C. M. Stein, "Estimation with quadratic loss," *Proc. 4th Berkeley Symposium Math. Stat. Prob.*, vol. 1, pp. 361–379.

[12] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

[13] Y. C. Eldar and A. V. Oppenheim, "Orthogonal multiuser detection," *Signal Processing*, vol. 82, pp. 321–325, 2002.

[14] Y. C. Eldar and A. M. Chan, "An optimal whitening approach to linear multiuser detection," submitted to *IEEE Trans. Inform. Theory*, Jan. 2001.