Maximum A-Posteriori Estimation in Linear Models With a Gaussian Model Matrix

Ido Nevat*, Ami Wiesel†, Jinhong Yuan* and Yonina C. Eldar†
*School of Electrical Engineering and Telecommunications, University of New South Wales
Email: ido@student.unsw.edu.au, J.Yuan@unsw.edu.au
†Department of Electrical Engineering, Technion - Israel Institute of Technology
Email: amiw@tx.technion.ac.il, yonina@ee.technion.ac.il

***Abstract*—We consider the Bayesian inference of a random Gaussian vector in a linear model with a Gaussian model matrix. We derive the maximum a-posteriori (MAP) estimator for this model and show that it can be found using a simple line search over a unimodal function that can be efficiently evaluated. Next, we discuss the application of this estimator in the context of near-optimal detection of near-Gaussian-digitally modulated signals and demonstrate through simulations that the MAP estimator outperforms the standard linear MMSE estimator in terms of mean square error (MSE) and bit error rate (BER).**

I. Introduction

A generic problem in many different fields is the estimation of a random Gaussian vector $\mathbf{x}$ in the linear model

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

where $\mathbf{G}$ is a linear transformation matrix and $\mathbf{w}$ is a Gaussian noise vector. Three standard methods for estimating $\mathbf{x}$ in this Bayesian framework are the minimum mean square error (MMSE), the linear minimum mean squared error (LMMSE) and the maximum a-posteriori (MAP) estimators. The first two approaches are based on a quadratic cost function whereas the third minimizes a hit-or-miss risk. From a detection point of view, the MAP method is also related to the minimum error probability criterion.

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 MMSE, LMMSE and MAP estimators coincide and have a simple closed form solution. However, in our model, the matrix $\mathbf{G}$ is a random matrix with independent and identically distributed elements and known second-order statistics. A typical scenario in which $\mathbf{G}$ is random is estimation under uncertainty conditions. For example, in communication systems this setting is appropriate when only partial channel state information is available. In this model, the MMSE, LMMSE and MAP approaches lead to different estimators. In fact, we will show that the solution of the MMSE leads to an intractable integration, whereas the MAP estimator can be efficiently found.

A possible application is digital communication systems employing near-Gaussian constellation sets. It is well known that in order to achieve capacity in linear Gaussian channels, powerful coding schemes must be combined with shaping methods which result in near-Gaussian symbols [1], [2]. Two practical schemes that obtain shaping gain are "trellis shaping" [3] and "shell mapping" [4]. Another example is the interleave-division-multiplexing space-time (IDM-ST) scheme, in which multiple independent data streams are encoded with forward error correction (FEC), interleaved and multiplexed simultaneously into different antennas. The superposition of multiple independent symbols generates a Gaussian distributed signal that is capacity achieving [5].

In [6]-[7], the authors derived the maximum likelihood (ML) solution for estimating a deterministic unknown vector $\mathbf{x}$ in model (1). In this paper, we extend these results and incorporate prior information about $\mathbf{x}$ by assuming that it is a Gaussian random vector. In fact, the MAP estimator can be interpreted as a regularized ML estimator which utilizes this prior information. When the variance of the elements in $\mathbf{x}$ are infinite, or when $\mathbf{x}$ is uniformly distributed, the MAP criterion reduces to the ML objective. Thus, the main contribution of this paper is to generalize the results in [6]-[7] to the case in which $\mathbf{x}$ is a Gaussian random vector with independent elements of known finite variance. An additional contribution is the application of this technique to near-optimal detection of near-Gaussian digital constellations.

This paper is organized as follows. In Section II, we formulate the problem and introduce the MMSE, LMMSE and MAP estimators. Next, in Section III we provide the numerical solution to the MAP optimization problem. For comparison purposes, we review the MMSE and LMMSE methods in Section IV. The application of our results in near-optimal detection are discussed in Section V, and some simulation results are offered in VI. Finally, concluding remarks are given in Section VII.

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 superscripts $(\cdot)^T$, $(\cdot)'$, $(\cdot)''$, and $(\cdot)^\dagger$ denote the transpose, the first and second derivatives, and the pseudoinverse, respectively. By $\mathbf{I}$ we denote the identity matrix. $\|\cdot\|$ is the standard Euclidean norm, $\mathbb{R}(\mathbf{X})$ is the range of $\mathbf{X}$, $\lambda_{\min}(\mathbf{X})$ is the smallest eigenvalue of $\mathbf{X}$ and $\mathbf{X} \succeq 0$ means that the matrix is a symmetric positive semidefinite matrix. The functions $p(\mathbf{x})$, $p(\mathbf{x}|\mathbf{y})$ and $E\{\cdot\}$ denote the probability distribution function (PDF) of $\mathbf{x}$, the PDF of $\mathbf{x}$ given $\mathbf{y}$, and the expectation, respectively.

1-4244-1037-1/07/$25.00 C2007 IEEE.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on October 31, 2009 at 06:27 from IEEE Xplore. Restrictions apply.

II. PROBLEM FORMULATION

Consider the problem of estimating a random vector $\mathbf{x}$ in the linear model

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{w}, \tag{2}$$

where $\mathbf{G}$ is an $N \times K$ Gaussian matrix with known mean $\mathbf{H}$ and variance $\sigma_g^2 > 0$, $\mathbf{x}$ is a zero-mean Gaussian vector with independent elements of variance $\sigma_x^2 > 0$ and $\mathbf{w}$ is a zero-mean Gaussian vector with independent elements of variance $\sigma_w^2 > 0$. In addition, $\mathbf{x}$, $\mathbf{G}$ and $\mathbf{w}$ are statistically independent. It is desired to find an estimator $\widehat{\mathbf{x}}(\mathbf{y})$ which is a function of the observation vector $\mathbf{y}$ and the given statistics, that is optimal in some sense. Under the Bayesian framework, a typical procedure for choosing $\widehat{\mathbf{x}}(\mathbf{y})$ is to define a nonnegative cost function $C(\mathbf{x}, \widehat{\mathbf{x}}(\mathbf{y}))$ and to minimize its expected value [8]. The most common objective is the quadratic error which is defined as (See Fig. 1)

$$C(\mathbf{x}, \widehat{\mathbf{x}}(\mathbf{y})) = \|\mathbf{x} - \widehat{\mathbf{x}}(\mathbf{y})\|^2. \tag{3}$$

Minimizing this objective leads to the well known MMSE estimator

$$\widehat{\mathbf{x}}_{MMSE}(\mathbf{y}) = E\{\mathbf{x}|\mathbf{y}\}. \tag{4}$$

In many problems, the computational complexity of the MMSE estimator is impractical. In this case, a common approach is to resort to a linear MMSE (LMMSE) estimator which satisfies the following closed form solution:

$$\widehat{\mathbf{x}}_{LMMSE}(\mathbf{y}) = E\left\{\mathbf{x}\mathbf{y}^T\right\} E^{-1}\left\{\mathbf{y}\mathbf{y}^T\right\}\mathbf{y}, \tag{5}$$

where we have used the fact that $\mathbf{x}$ and $\mathbf{y}$ are zero mean random vectors. Alternatively, one may choose to minimize the hit-or-miss cost given by (See Fig. 1)

$$C(\mathbf{x}, \widehat{\mathbf{x}}(\mathbf{y})) = \begin{cases} 0, & \|\mathbf{x} - \widehat{\mathbf{x}}(\mathbf{y})\| \leq \epsilon \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

where $\epsilon \to 0$ is a positive scalar. Optimizing this risk yields the MAP estimator:

$$\widehat{\mathbf{x}}_{MAP}(\mathbf{y}) = \arg\max_{\mathbf{x}} \log p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}). \tag{7}$$

The main goal of this paper is to derive the MAP estimator of $\mathbf{x}$ in model (2), and compare its performance to the LMMSE method.

III. MAP ESTIMATION

In this section, we provide an efficient algorithm for finding the MAP estimator in model (2). The estimator is defined as

$$\begin{aligned} \widehat{\mathbf{x}}_{MAP}(\mathbf{y}) &= \arg\max_{\mathbf{x}} \left\{\log p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})\right\} \\ &= \arg\max_{\mathbf{x}} \left\{\log p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) + \log p_{\mathbf{x}}(\mathbf{x})\right\}. \end{aligned} \tag{8}$$

Due to the Gaussian assumptions, $\widehat{\mathbf{x}}_{MAP}(\mathbf{y})$ is the solution to

$$\min_{\mathbf{x}} \left\{ \frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma_g^2 \|\mathbf{x}\|^2 + \sigma_w^2} + N\log(\sigma_g^2\|\mathbf{x}\|^2 + \sigma_w^2) + \frac{\|\mathbf{x}\|^2}{\sigma_x^2} \right\}. \tag{9}$$

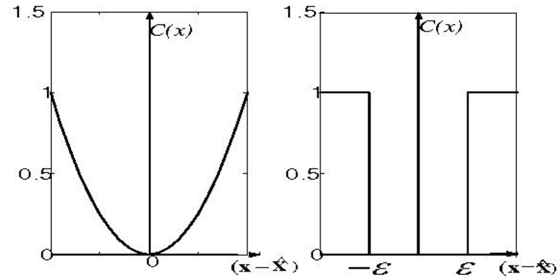

Fig. 1. quadratic cost function (left) and hit-or-miss cost function (right)

Problem (9) is a K-dimensional, nonlinear and nonconvex optimization program. In [6]-[7] the authors have presented a method to transform a similar problem into a tractable form and solve it efficiently. Under their setting, the vector $\mathbf{x}$ was treated as an *unknown deterministic* vector. In our setting, the vector $\mathbf{x}$ is treated as a *random Gaussian* vector. This difference results in an additional quadratic term in the MAP objective function, namely $\|\mathbf{x}\|^2/\sigma_x^2$, which incorporates the a-priori information about the random vector $\mathbf{x}$. The following theorem shows that the technique in [6]-[7] can also be applied in the MAP problem.

Theorem 3.1: For any $t \geq 0$, let

$$f(t) = \min_{\mathbf{x}: \|\mathbf{x}\|^2 = t} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \tag{10}$$

and denote the optimal argument by $\mathbf{x}(t)$. Then, the MAP estimator of $\mathbf{x}$ in the model (2) is $\mathbf{x}(t^*)$, where t^* is the solution to the following unimodal optimization problem:

$$\arg\min_{t\geq 0} \left\{ \frac{f(t)}{\sigma_g^2 t + \sigma_w^2} + N\log(\sigma_g^2 t + \sigma_w^2) + \frac{t}{\sigma_x^2} \right\} \tag{11}$$

Proof: By introducing a slack variable $t = \|\mathbf{x}\|^2$, we can rewrite (9) as (11) using $f(t)$ defined in (10). In [7] it was shown that the line search is unimodal in $t \geq 0$ when there is no prior information on $\mathbf{x}$. In the appendix, we prove that this property holds also in the MAP case. ■

The change of variables in Theorem 3.1 allows for an efficient solution of the MAP problem since there are standard methods for evaluating $f(t)$ in (10) for any $t \geq 0$. Moreover, the unimodality ensures that a simple one dimensional search can find the global optimum. In the rest of this section, we discuss these two properties.

First we provide a simple method for evaluating $f(t)$ in (10). This is a quadratically constrained LS problem whose solution can be traced back to [9]:

Lemma 3.2: ([9], [10]): The solution to

$$f(t) = \min_{\mathbf{x}: \|\mathbf{x}\|^2 = t} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \tag{12}$$

is

$$\mathbf{x}(t) = \left(\mathbf{H}^T\mathbf{H} + \eta\mathbf{I}\right)^{\dagger}\mathbf{H}^T\mathbf{y}, \tag{13}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on October 31, 2009 at 06:27 from IEEE Xplore. Restrictions apply.

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

$$\|\mathbf{x}(t)\|^2 = t. \tag{14}$$

Using the eigenvalue decomposition of $\mathbf{H}^T\mathbf{H}$, we can calculate $\left\|\left(\mathbf{H}^T\mathbf{H}+\eta\mathbf{I}\right)^{\dagger}\mathbf{H}^T\mathbf{y}\right\|^2$ for different values of η. The monotonicity of this squared norm in η enables us to find η that satisfies (14) using a simple line-search. The search range is $-\lambda_{min}\left(\mathbf{H}^T\mathbf{H}\right) \leq \eta \leq \eta_{max}$, where η_{max} is some sufficiently large upper bound. Next, $f(t)$ can be evaluated by plugging the appropriate $\mathbf{x}(t)$ into $\|\mathbf{y}-\mathbf{H}\mathbf{x}(t)\|^2$. This algorithm is presented in **Algorithm 1**.

Algorithm 1 Constrained Least Squares (*Lemma 3.2*)

Input: t, $\mathbf{H}$, $\mathbf{y}$, $\lambda_{min}\left(\mathbf{H}^T\mathbf{H}\right)$, η_{max}
Output: $f(t)$,$\mathbf{x}$

$\eta_L = -\lambda_{min}\left(\mathbf{H}^T\mathbf{H}\right)$
$\eta_R = \eta_{max}$
repeat
  $\eta_M = \frac{\eta_L+\eta_R}{2}$
  $\mathbf{x} = \left(\mathbf{H}^T\mathbf{H}+\eta_M\mathbf{I}\right)^{-1}\mathbf{H}^T\mathbf{y}$
  $\epsilon = \mathbf{x}^T\mathbf{x} - t$
  if $\epsilon > 0$ **then**
    $\eta_L = \eta_M$
  else
    $\eta_R = \eta_M$
  end if
until $|\epsilon| \leq \epsilon_{min}$
$f(t) = \|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2$

Now that we have an efficient method for evaluating $f(t)$, it remains to solve (11). The unimodality property ensures that this line search can be efficiently implemented using the Golden Section search [11]. Theoretically, the search region is defined to be over $0 \leq t \leq \infty$. However, in practice, the search can be confined to $0 \leq t \leq t_{max}$ where t_{max} is a sufficiently large upper bound. This algorithm is presented in **Algorithm 2**.

Algorithm 2 MAP Estimation - Solution of eq. (11)

Input: $\mathbf{y}$, $\mathbf{H}$, σ_x^2, σ_w^2, σ_g^2, N, t_{max}
Output: $\mathbf{x}$

$t_L = 0$
$t_R = t_{max}$
$\rho = \frac{\left(\sqrt{5}-1\right)^2}{4}$
repeat
  $\Delta = t_R - t_L$
  $t_A = t_L + \rho\Delta$
  $t_B = t_R - \rho\Delta$
  $r(t_A) = \frac{f(t_A)}{\sigma_g^2 t_A+\sigma_w^2} + N\log(\sigma_g^2 t_A + \sigma_w^2) + \frac{t_A}{\sigma_x^2}$
  $r(t_B) = \frac{f(t_B)}{\sigma_g^2 t_B+\sigma_w^2} + N\log(\sigma_g^2 t_B + \sigma_w^2) + \frac{t_B}{\sigma_x^2}$
  if $r(t_A) < r(t_B)$ **then**
    $t_R = t_B$
  else
    $t_L = t_A$
  end if
until $|r(t_A) - r(t_B)| < \epsilon$

IV. MMSE ESTIMATION

For comparison, we now discuss the MMSE and LMMSE estimators of $\mathbf{x}$ in model (2). The MMSE estimator can be written as

$$\begin{aligned}\widehat{\mathbf{x}}_{MMSE}(\mathbf{y}) &= E\{\mathbf{x}|\mathbf{y}\}\\ &= E\{E\{\mathbf{x}|\mathbf{y},\mathbf{G}\}|\,\mathbf{y}\}\\ &= E\left\{\left(\mathbf{G}^T\mathbf{G}+\frac{\sigma_w^2}{\sigma_x^2}I\right)^{-1}\mathbf{G}^T\mathbf{y}\,\middle|\,\mathbf{y}\right\}.\end{aligned} \tag{15}$$

Unfortunately, it is easy to see that the computational complexity involved in solving (15) is too high for practical applications. Instead, a common approach is to consider the LMMSE estimator given by

$$\begin{aligned}\widehat{\mathbf{x}}_{LMMSE}(\mathbf{y}) &= \mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + K\sigma_g^2\mathbf{I} + \frac{\sigma_w^2}{\sigma_x^2}\mathbf{I}\right)^{-1}\mathbf{y}\\ &= \left(\mathbf{H}^T\mathbf{H}+\eta_{LMMSE}\mathbf{I}\right)^{-1}\mathbf{H}^T\mathbf{y},\end{aligned} \tag{16}$$

where $\eta_{LMMSE} = K\sigma_g^2 + \frac{\sigma_w^2}{\sigma_x^2}$. Note that (16) has the same form as (13) with a different regularization term η_{LMMSE}. However, while η_{LMMSE} is deterministic and always positive, the MAP regularization parameter η is a random variable which may be negative.

V. BAYESIAN DETECTION OF DIGITAL CONSTELLATIONS

We now discuss the application of the MAP estimator in the context of near-optimal detection of near-Gaussian digital constellations. The jointly optimal detector is given by [12]

$$\widehat{\mathbf{x}}_{MAP}(\mathbf{y}) = \arg\max_{\mathbf{x}\in D^K} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}), \tag{17}$$

where D is the modulation alphabet. The complexity of the MAP detector is exponential in K, due to the discrete nature of the support which has $|D|^K$ elements and is usually unrealizable. Instead, we suggest a low complexity suboptimal detector based on the MAP estimator presented in Section III. The jointly optimal detector given by (17) can be written as (9), but this time, the support of $\mathbf{x}$ is $|D|^K$. Thus, it is the solution to

$$\min_{\mathbf{x}\in D^K}\left\{\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_g^2\|\mathbf{x}\|^2+\sigma_w^2} + N\log(\sigma_g^2\|\mathbf{x}\|^2+\sigma_w^2) + \frac{\|\mathbf{x}\|^2}{\sigma_x^2}\right\}. \tag{18}$$

However, as the support of D increases, that is to say, the number of data bins gets larger, the solution of (18) converges to the solution of (9).

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on October 31, 2009 at 06:27 from IEEE Xplore. Restrictions apply.

Therefore, an appealing near-optimal approach for approximating the MAP detector is

$$\widehat{\mathbf{x}}_{D-MAP}(\mathbf{y}) = \text{quantize}\left(\widehat{\mathbf{x}}_{MAP}(\mathbf{y})\right), \tag{19}$$

where $\widehat{\mathbf{x}}_{MAP}(\mathbf{y})$ is the solution to (11). In the limit of infinite number of bins, $\widehat{\mathbf{x}}_{D-MAP}$ is effectively equal to $\widehat{\mathbf{x}}_{MAP}$, and is optimal. In that case, the detection problem, generally considered to be exponential complex, can be solved with linear complexity, given in Section III.

For comparison purposes, the LMMSE detector is given by

$$\widehat{\mathbf{x}}_{D-LMMSE}(\mathbf{y}) = \text{quantize}\left(\widehat{\mathbf{x}}_{LMMSE}(\mathbf{y})\right), \tag{20}$$

where $\widehat{\mathbf{x}}_{LMMSE}(\mathbf{y})$ is given by (16).

VI. Simulation results

We now provide numerical results illustrating the behavior of our new estimator. For this simulation the parameters are $N = 40$, $K = 4$. The matrix $\mathbf{H}$ was chosen as a concatenation of ten 4×4 matrices with unit diagonal elements and 0.5 off-diagonal elements. The vector $\mathbf{x}$ is a zero-mean Gaussian, digitally modulated, stemming 8 different symbols as shown in Fig. 2. The variance of each element is 1.5. Each symbol maps 3 bits using gray labeling, thus, neighboring symbols differ by only one bit. The simulation results for the bit error rate (BER) and mean square error (MSE) of the MAP and LMMSE estimators are presented in Figs. 3 and 4, respectively, for different values of $\sigma_g^2 = \{0, 0.02, 0.04\}$. As expected, in the special case where $\sigma_g^2 = 0$, the MAP and LMMSE estimators are identical. On the other hand, when $\sigma_g^2 = \{0.02, 0.04\}$, the MAP estimator yields better performances in terms of both MSE and BER.

Next, we compare the behavior of the MAP and LMMSE regularization factors for various values of σ_w^2. While the regularization factor of the LMMSE estimator η_{LMMSE}, is deterministic and fully determined by the problem parameters, the regularization factor of the MAP estimator η, is a random variable. Fig 5 depicts the empirical distribution of the regularization factors the MAP and LMMSE estimators, for the case $\sigma_g^2 = 0.04$. The results show that the average value of η is smaller than η_{LMMSE} for every σ_w^2. Since we show that the MSE of the MAP estimator is lower than the LMMSE, we can conclude that the LMMSE is an underestimate of the true value of $\mathbf{x}$. We also observe that both η and η_{LMMSE} decrease as σ_w^2 decreases. This is not surprising since as the observations are less noisy, less regularization is needed. Moreover, we notice that the η can have negative values, that is to say, deregularization form.

VII. Conclusions

In this work, we introduced the MAP estimator of a random Gaussian vector $\mathbf{x}$ in a linear model with random transformation matrix $\mathbf{G}$. We derived the MAP estimator and provided an efficient method for finding it by transforming the multi-dimensional, nonlinear and nonconvex problem into a simple tractable form. Next, we proposed a detection scheme for near-Gaussian-digitally modulated symbols with linear complexity.

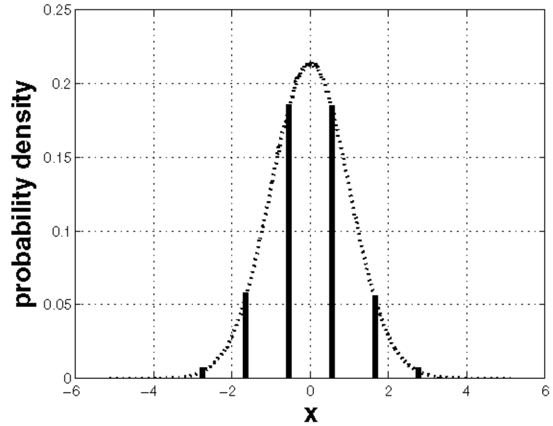

Fig. 2. Near-Gaussian constellation of 8 symbols.

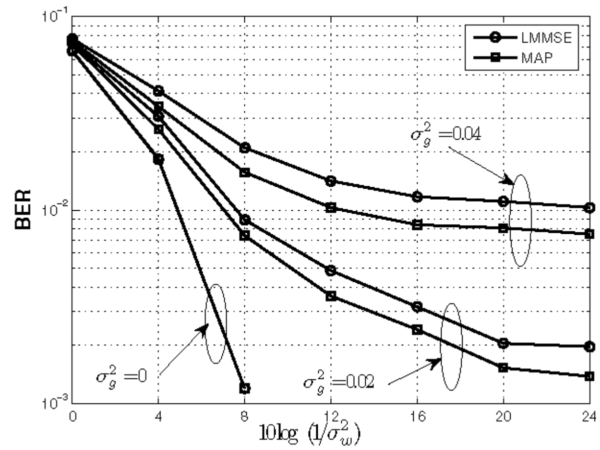

Fig. 3. Bit error rate of a MIMO system with $N = 40$, $K = 4$ for various values of σ_g^2.

Simulation results show the improved performance offered by our new approach in comparison to the standard LMMSE method in terms of both MSE and BER.

Appendix

In this Appendix, we show that (11) is unimodal in $t \geq 0$. First, we will show that $f(t)$ is convex in $t \geq 0$. In [10] and [13], it was shown that strong duality holds in this special case and that is equal to the value of its dual program

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

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

Next, we will show that

$$r(t) = \frac{f(t)}{\sigma_g^2 t + \sigma_w^2} + N\log(\sigma_g^2 t + \sigma_w^2) + \frac{t}{\sigma_x^2} \tag{22}$$

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on October 31, 2009 at 06:27 from IEEE Xplore. Restrictions apply.

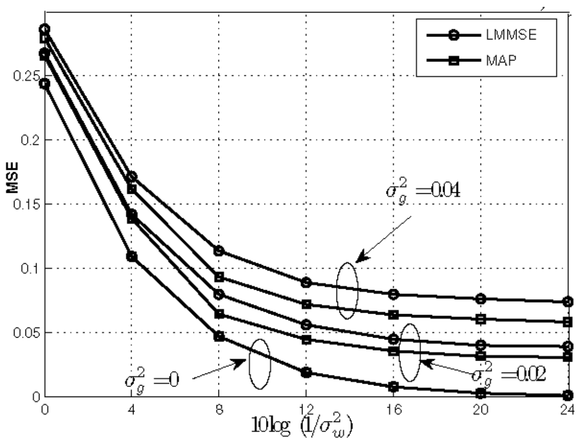

Fig. 4. Mean square error of a MIMO system with $N = 40$, $K = 4$ for various values of σ_g^2.

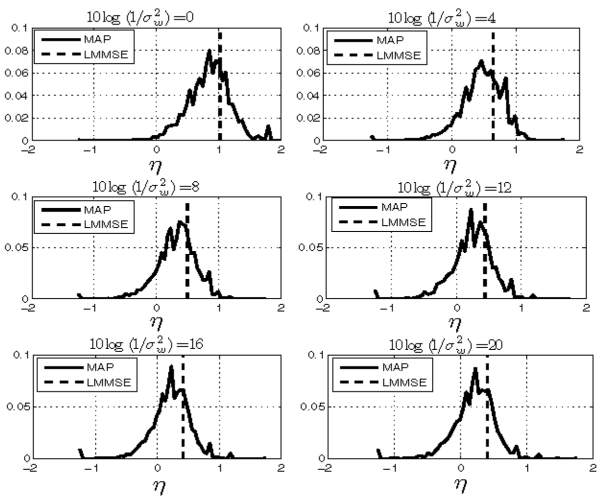

Fig. 5. Distribution of η for $\sigma_g^2 = 0.04$, $N = 40$, $K = 4$ for different values of σ_w^2.

is unimodal in $t \geq 0$. We use the following result from [13]: 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

$$r'(t) = \frac{f'(t)}{\sigma_g^2 t + \sigma_w^2} - \frac{f(t)\sigma_g^2}{(\sigma_g^2 t + \sigma_w^2)^2} + \frac{N\sigma_g^2}{\sigma_g^2 t + \sigma_w^2} + \frac{1}{\sigma_x^2} = 0. \tag{23}$$

Multiplying by $\frac{\sigma_g^2}{\sigma_g^2 t + \sigma_w^2}$ and rearranging yields

$$\begin{aligned}\frac{f(t)\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^3} &= \frac{N\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^2} + \frac{\sigma_g^2 f'(t)}{(\sigma_g^2 t + \sigma_w^2)^2} \\ &\quad + \frac{\sigma_g^2}{\sigma_x^2(\sigma_g^2 t + \sigma_w^2)}.\end{aligned} \tag{24}$$

The second derivative is

$$\begin{aligned}r''(t) = {} & \frac{f''(t)}{\sigma_g^2 t + \sigma_w^2} - \frac{f'(t)\sigma_g^2}{(\sigma_g^2 t + \sigma_w^2)^2} - \frac{f'(t)\sigma_g^2}{(\sigma_g^2 t + \sigma_w^2)^2} \\ & + \frac{2f(t)\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^3} - \frac{N\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^2}.\end{aligned} \tag{25}$$

Plugging in (23) results in

$$\begin{aligned}r''(t) = {} & \frac{f''(t)}{\sigma_g^2 t + \sigma_w^2} - \frac{f'(t)\sigma_g^2}{(\sigma_g^2 t + \sigma_w^2)^2} - \frac{f'(t)\sigma_g^2}{(\sigma_g^2 t + \sigma_w^2)^2} \\ & + \frac{2N\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^2} + \frac{2\sigma_g^2 f'(t)}{(\sigma_g^2 t + \sigma_w^2)^2} + \frac{2\sigma_g^2}{\sigma_x^2(\sigma_g^2 t + \sigma_w^2)} \\ & - \frac{N\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^2} \\ = {} & \frac{f''(t)}{\sigma_g^2 t + \sigma_w^2} + \frac{2\sigma_g^2}{\sigma_x^2(\sigma_g^2 t + \sigma_w^2)} + \frac{N\sigma_g^4}{(\sigma_g^2 t + \sigma_w^2)^2}.\end{aligned} \tag{26}$$

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

References

[1] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear gaussian channel," *IEEE Trans. on Information Theory*, vol. 44, pp. 2384–2415, Oct. 1998.

[2] U. Wachsman, R. F. H. Fischer, and J. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.

[3] G. D. Forney, "Trellis shaping," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 281–300, March 1992.

[4] P. Fortier, A. Ruiz, and J. M. Cioffi, "Multidimensional signal sets through the shell construction for parallel channels," *IEEE Trans. on Communication*, vol. 40, no. 500-512, pp. 2384–2415, March 1992.

[5] K. Wu, L. Ping, and J. Yuan, "A quasi-random approach to space-time codes," in *Proc. of International Symposium on Turbo Codes, Munich, Germany*, April 2006.

[6] A. Wiesel, Y. C. Eldar, and A. Beck, "Maximum likelihood estimation in linear models with a gaussian model matrix," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 292–295, May 2006.

[7] A. Wiesel and Y. C. Eldar, "Maximum likelihood estimation in random linear models generalizations and performance analysis," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-2006)*, May 2006.

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

[9] 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.

[10] 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.

[11] W. H. Press, Saul A. Teukolsky, W. T. Vettering, and Brian P. Flannery, *Numerical recipes in C*, Cambridge University Press, Cambridge, second edition, 1992, The art of scientific computing.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing - Detection Theory*, Prentice Hall, 1998.

[13] S. Boyd and L. Vandenberghe, *Introduction to Convex Optimization with Engineering Applications*, Stanford, 2003.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on October 31, 2009 at 06:27 from IEEE Xplore. Restrictions apply.