MAXIMUM A-POSTERIORI ESTIMATION IN LINEAR MODELS WITH A RANDOM GAUSSIAN MODEL MATRIX: A BAYESIAN-EM APPROACH

Ido Nevat *, *Gareth W. Peters* ‡, *and Jinhong Yuan* *

School of Electrical Engineering and Telecommunications, University of NSW, Australia *
School of Mathematics and Statistics, University of NSW, Australia ‡
email: ido@student.unsw.edu.au, peterga@maths.unsw.edu.au, J.Yuan@unsw.edu.au

ABSTRACT

This paper considers the problem of Bayesian estimation of a Gaussian vector in a linear model with random Gaussian uncertainty in the mixing matrix. The maximum a-posteriori estimator is derived for this model using the Bayesian Expectation-Maximization. It is demonstrated that the solution forms an elegant and simple iteration which can be easily implemented. Finally, the estimator developed is considered in the context of near-Gaussian-digitally modulated signals under channel uncertainty, where it is shown that the MAP estimator outperforms the standard linear MMSE estimator in terms of mean square error (MSE) and bit error rate (BER).

Index Terms— MAP estimation, Bayesian EM

1. 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 function. 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 deterministic and specified a-priori. In this setting, the MMSE, LMMSE and MAP estimators coincide and have a simple closed form solution. The novelty of this paper lies in the specification of the transformation matrix, where we remove the assumption, made in much of the literature that $\mathbf{G}$ is known deterministically. Instead we treat $\mathbf{G}$ as a random matrix and assume weak statistical properties of this matrix, namely that its elements are i.i.d Gaussian distributed with 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 case, 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], this problem was tackled and the MAP solution was derived by transformation of the problem from a multi dimensional into one dimensional optimization program. In this process the objective function becomes convex and this can be exploited in the solution technique. The drawback of this proposed method lies in the fact that to perform this technique, one must determine the eigen values of a potentially high rank matrix, this can lead to computational issues related to scaling the complexity of the system under consideration due to the curse of dimensionality. The proposed technique in this paper bypasses these issues and so scales more effectively with system complexity. Using the BEM procedure, we derive the solution which forms an iterative procedure that can be easily implemented.

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 $(\cdot)^T$ denotes the transpose. By $\mathbf{I}$ we denote the identity matrix and $\|\cdot\|$ is the standard Euclidean norm. The operations $\otimes$ and $\mathbf{vec}(.)$ denote the Kronecker matrix multiplication and the vector obtained by stacking the columns of a matrix one over the other. 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.

2. 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 of $\mathbf{G}$, that is optimal in some sense to be defined next. Under the Bayesian framework, a typical procedure for selecting $\widehat{\mathbf{x}}(\mathbf{y})$ is to define a nonnegative cost function $C(\mathbf{x}, \widehat{\mathbf{x}}(\mathbf{y}))$ and to minimize its expected value [7]. The most common objective function is the quadratic error which is defined as (See Fig. 1)

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

Minimizing this objective function leads to the well known MMSE estimator [7]

$$\begin{aligned}\widehat{\mathbf{x}}_{MMSE}\left(\mathbf{y}\right)&=E\left\{\mathbf{x}|\mathbf{y}\right\}\\&=E\left\{E\left\{\mathbf{x}|\mathbf{y},\mathbf{G}\right\}|\,\mathbf{y}\right\}\\&=E\left\{\left(\mathbf{G}^{T}\mathbf{G}+\frac{\sigma_{w}^{2}}{\sigma_{x}^{2}}I\right)^{-1}\mathbf{G}^{T}\mathbf{y}\middle|\,\mathbf{y}\right\}.\end{aligned} \tag{4}$$

Unfortunately, it is easy to see that the computational complexity involved in solving (4) is too high for practical applications. Instead, a common approach is to consider the LMMSE estimator which satisfies the following closed form solution [7]:

$$\begin{aligned}\widehat{\mathbf{x}}_{LMMSE}\left(\mathbf{y}\right)&=E\left\{\mathbf{x}\mathbf{y}^{T}\right\}E^{-1}\left\{\mathbf{y}\mathbf{y}^{T}\right\}\mathbf{y}\\&=\mathbf{H}^{T}\left(\mathbf{H}\mathbf{H}^{T}+K\sigma_{g}^{2}\mathbf{I}+\frac{\sigma_{w}^{2}}{\sigma_{x}^{2}}\mathbf{I}\right)^{-1}\mathbf{y},\end{aligned} \tag{5}$$

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 function given by (See Fig. 1)

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

where $\epsilon\rightarrow 0$ is a positive scalar. Optimizing this cost function yields the MAP estimator:

$$\begin{aligned}\widehat{\mathbf{x}}_{MAP}\left(\mathbf{y}\right)&=\arg\max_{\mathbf{x}}\left\{\log p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x}|\mathbf{y}\right)\right\}\\&=\arg\max_{\mathbf{x}}\left\{\log p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}|\mathbf{x}\right)+\log p_{\mathbf{x}}\left(\mathbf{x}\right)\right\}.\end{aligned} \tag{7}$$

As a result of the Gaussian assumption,we have that

$$\begin{aligned}p_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}|\mathbf{x}\right)&\sim\mathcal{N}\left(\mathbf{H}\mathbf{x},\left(\sigma_{g}^{2}\left\|x\right\|^{2}+\sigma_{w}^{2}\right)\mathbf{I}\right),\\p_{\mathbf{x}}\left(x\right)&\sim\mathcal{N}\left(0,\sigma_{x}^{2}\mathbf{I}\right),\end{aligned} \tag{8}$$

and $\widehat{\mathbf{x}}_{MAP}\left(\mathbf{y}\right)$ is the solution to

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

Problem (9) is a K-dimensional, nonlinear and nonconvex optimization program. In the rest of the paper we will discuss the solution of (9) using BEM method.

3. REVIEW OF BAYESIAN EM METHODS

In this section we provide a short overview of BEM methods. The Bayesian expectation-maximization algorithm is one of several general techniques for finding the MAP estimates where the model depends on unobserved latent variables.

We have focused on this technique since it utilizes the classical EM algorithm [8], which exploits the property that the EM algorithm is known to converge to a stationary point corresponding to a local optimum of the posterior distribution, though convergence to the global model is not guaranteed.

The BEM algorithm consists of two major steps: an expectation step, followed by a maximization step. The expectation is with respect to the unknown underlying variables, conditioned on the current estimate of the parameters and the observations. During the maximization step one maximizes the complete data likelihood conditioning on the expectation estimates of the previous step. The algorithm is numerically stable and convergence is typically fast, though one should be careful with likelihood functions which are multi-modal. In such cases ad hoc methods such as multiple starting points have been proposed to try to obtain global maxima in this framework. The BEM is implemented in the following steps:

$$Q\left(\mathbf{x},\mathbf{x}_{n}\right)=E\left\{\log p\left(\mathbf{y},\mathbf{G},\mathbf{x}\right)|\mathbf{y};\mathbf{x}_{n}\right\}, \tag{10}$$

$$\mathbf{x}_{n+1}=\arg\max_{\mathbf{x}}\left\{Q\left(\mathbf{x},\mathbf{x}_{n}\right)\right\}. \tag{11}$$

Eq. (10- 11) are repeated until a stopping criterion is attained.

4. MAP ESTIMATION USING BEM

In this section we demonstrate the application of the BEM methodology in solving the MAP estimation problem in context of model (2). At each iteration, the algorithm maximizes the expected log likelihood (The expectation is taken with respect to $\mathbf{G}$ to integrate this parameter out of the target posterior. This allows the maximization step to obtain an updated estimate of the MAP of $\mathbf{x}$ for the target posterior $p(\mathbf{x}\mid\mathbf{y})$)

$$\begin{aligned}\mathbf{x}_{n+1}&=\arg\max_{\mathbf{x}}E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\log p\left(\mathbf{y},\mathbf{G},\mathbf{x}\right)\right\}\\&=\arg\max_{\mathbf{x}}E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\log p\left(\mathbf{y}|\mathbf{G},\mathbf{x}\right)+\log p\left(\mathbf{G}|\mathbf{x}\right)\right.\\&\quad\left.+\log p\left(\mathbf{x}\right)\right\}.\end{aligned} \tag{12}$$

Since $\mathbf{x}$ and $\mathbf{G}$ are statistically independent, $\log p\left(\mathbf{G}|\mathbf{x}\right)=\log p\left(\mathbf{G}\right)$ and is constant w.r.t. $\mathbf{x}$. Therefore (12) simplifies to:

$$\begin{aligned}\mathbf{x}_{n+1}&=\arg\min_{\mathbf{x}}E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\frac{1}{\sigma_{w}^{2}}\left\|\mathbf{y}-\mathbf{G}\mathbf{x}\right\|^{2}+\frac{1}{\sigma_{x}^{2}}\left\|\mathbf{x}\right\|^{2}\right\}\\&=\arg\min_{\mathbf{x}}E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\frac{1}{\sigma_{w}^{2}}\left(-2\mathbf{y}^{T}\mathbf{G}\mathbf{x}+\mathbf{x}^{T}\mathbf{G}^{T}\mathbf{G}\mathbf{x}\right)\right.\\&\quad\left.+\frac{1}{\sigma_{x}^{2}}\left\|\mathbf{x}\right\|^{2}\right\}\\&=\arg\min_{\mathbf{x}}\left\{\frac{1}{\sigma_{w}^{2}}\left(-2\mathbf{y}^{T}E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\mathbf{G}\right\}\mathbf{x}\right)\right.\\&\quad\left.+\frac{1}{\sigma_{w}^{2}}\mathbf{x}^{T}E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\mathbf{G}^{T}\mathbf{G}\right\}\mathbf{x}+\frac{1}{\sigma_{x}^{2}}\mathbf{x}^{T}\mathbf{x}\right\}\\&=\arg\min_{\mathbf{x}}\left\{\frac{1}{\sigma_{w}^{2}}\left(-2\mathbf{y}^{T}\Phi_{1}\left(\mathbf{y},\mathbf{x}_{n}\right)\mathbf{x}+\mathbf{x}^{T}\Phi_{2}\left(\mathbf{y},\mathbf{x}_{n}\right)\mathbf{x}\right)\right.\\&\quad\left.+\frac{1}{\sigma_{x}^{2}}\mathbf{x}^{T}\mathbf{x}\right\},\end{aligned} \tag{13}$$

where

$$\begin{aligned}\Phi_{1}\left(\mathbf{y},\mathbf{x}_{n}\right)&=E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\mathbf{G}\right\},\\\Phi_{2}\left(\mathbf{y},\mathbf{x}_{n}\right)&=E_{\mathbf{G}|\mathbf{y};\mathbf{x}_{n}}\left\{\mathbf{G}^{T}\mathbf{G}\right\}.\end{aligned} \tag{14}$$

Next, we take the derivative with respect to $\mathbf{x}$ and equate to 0 to obtain the following iterative procedure:

$$\mathbf{x}_{n+1}=\left(\Phi_{2}\left(\mathbf{y},\mathbf{x}_{n}\right)+\frac{\sigma_{w}^{2}}{\sigma_{x}^{2}}I\right)^{-1}\Phi_{1}\left(\mathbf{y},\mathbf{x}_{n}\right)^{T}\mathbf{y}. \tag{15}$$

The expectations in (14) can be evaluated based on the jointly Gaussian optimal MMSE estimation theory [7]. First, using the Kronecker product we rewrite (2) as

$$\mathbf{y} = \left(\mathbf{x}^T \otimes I\right) \mathbf{g} + \mathbf{w}, \tag{16}$$

where $\mathbf{g} = \text{vec}(\mathbf{G})$. The Bayesian MMSE estimate of (16) can be expressed as

$$E_{\mathbf{g}|\mathbf{y};\mathbf{x}_n} \{\mathbf{g}\} = \mathbf{vec}(\mathbf{H}) + \frac{\left(\mathbf{x}^T \otimes I\right)(\mathbf{y} - \mathbf{H}\mathbf{x}_n)}{\|\mathbf{x}_n\|^2 + \frac{\sigma_w^2}{\sigma_g^2}}, \tag{17}$$

and

$$\mathbf{COV}_{\mathbf{g}|\mathbf{y};\mathbf{x}_n} \{\mathbf{g}\} = \sigma_g^2 I - \frac{\sigma_g^4 \left(\mathbf{x}_n \mathbf{x}_n^T \otimes I\right)}{\sigma_g^2 \|\mathbf{x}_n\|^2 + \sigma_w^2}. \tag{18}$$

Next, after trivial algebraic manipulations, we can find Φ_1 and Φ_2:

$$\Phi_1(\mathbf{y}, \mathbf{x}_n) = \mathbf{H} + \frac{1}{\|\mathbf{x}_n\|^2 + \frac{\sigma_w^2}{\sigma_g^2}} (\mathbf{y} - \mathbf{H}\mathbf{x}_n) \mathbf{x}_n^T, \tag{19}$$

$$\begin{aligned}\Phi_2(\mathbf{y}, \mathbf{x}_n) &= \Phi_1(\mathbf{y}, \mathbf{x}_n)^T \Phi_1(\mathbf{y}, \mathbf{x}_n) \\ &+ \left(I - \frac{\sigma_g^2 \mathbf{x}_n \mathbf{x}_n^T}{\|\mathbf{x}_n\|^2 + \frac{\sigma_w^2}{\sigma_g^2}}\right) \sigma_g^2 N.\end{aligned} \tag{20}$$

4.1. Initial guess of $\mathbf{x}_0$

In practice, selection of the initial guess will influence the ability of the algorithm to obtain an optimal solution. Typically the influence is seen trough the following aspects: number of iterations to achieve a solution with a desired tolerance, the ability to find a global solution as opposed to a local solution and computational effort. As discussed previously, the problem addressed in this paper is non-convex, hence convergence to the optimal solution is not guaranteed. There are several ad-hoc procedures for attempting to deal with these issues, which include multiple random starting points, coupled with a criterion that chooses the solution which is obtained most frequently from these starting points. Another alternative includes tempering of the objective function being maximized, in this case the posterior distribution. However, these techniques are not guaranteed to be optimal. As a simple computationally efficient alternative, we propose to choose the initial guess as:

$$\mathbf{x}_0 = \widehat{\mathbf{x}}_{LMMSE}. \tag{21}$$

In summary, a BEM algorithm is composed of (21) and iterating (15) with (19) and (20) until convergence. In practice, the solution of the set of simultaneous equations in (15) does not require matrix inversion and may be solved efficiently as a solution of a set of linear equations, which makes this approach practically viable.

5. BAYESIAN DETECTION OF NEAR-GAUSSIAN DIGITAL CONSTELLATIONS

We now discuss the application of the MAP estimator of near-Gaussian digital constellations. First, we discuss Gaussian-like constellations and then provide a detector for those constellations.

5.1. Nonuniform Constellations

The goal of nonuniform constellations is to create a discrete Gaussian-like distribution of the signals constellation, thus achieving the well known shaping gain [1, 2]. The idea behind constellation shaping is that symbols with small norm (energy) will be used more frequently than symbols with high norm. This achieves an overall reduction in transmitted energy. Theoretically, the signal points should be chosen from a continuous Gaussian distribution. In practice, since the constellations are finite, an optimal gain can not be achieved. The *Maxwell-Boltzman* (M-B) distribution is a good approximation of the optimal solution for the maximum of the mutual information [9] and the symbols' probabilities $P(\mathbf{s}_j), j = 1, ..., |D|$ are attained by

$$P(s_j) = K(\lambda) \exp\left\{-\lambda |s_j|^2\right\}, \ \lambda \geq 0, \tag{22}$$

where

$$K(\lambda) = \left(\sum_{s_j} \exp\left\{-\lambda |s_j|^2\right\}\right)^{-1}, \tag{23}$$

is the distribution normalization factor, and λ controls the tradeoff between the average power and the entropy rate $H(S)$.

5.2. Construction of 8-PAM constellation

We apply the M-B distribution, according to (22) to an 8-ary PAM constellation, using $\lambda = 1/20$.

$$\begin{cases} P(s = \pm 1) & = 0.2425 \\ P(s = \pm 3) & = 0.1625 \\ P(s = \pm 5) & = 0.0730 \\ P(s = \pm 7) & = 0.0220 \end{cases} \tag{24}$$

5.3. Detection Scheme

The jointly optimal detector is given by [7]

$$\widehat{\mathbf{x}}_{MAP}(\mathbf{y}) = \arg \max_{\mathbf{x} \in D^K} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}), \tag{25}$$

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 4. The jointly optimal detector given by (25) can be written as (9), but this time, the support of $\mathbf{x}$ is $|D|^K$. Thus, it is the solution to

$$\min_{\mathbf{x} \in D^K} \left\{ \frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma_g^2 \|\mathbf{x}\|^2 + \sigma_w^2} + N \log(\sigma_g^2 \|\mathbf{x}\|^2 + \sigma_w^2) + \frac{\|\mathbf{x}\|^2}{\sigma_x^2} \right\}. \tag{26}$$

However, as the support of D increases, hence, the quantization error decreases, the solution of (26) converges to the solution of (9).

Therefore, an appealing near-optimal approach for approximating the MAP detector is quantizing the MAP solution for the continuous Gaussian distribution to the nearest lattice point, that is:

$$\widehat{\mathbf{x}}_{D-MAP}(\mathbf{y}) = \text{quantize}\left(\widehat{\mathbf{x}}_{MAP}(\mathbf{y})\right), \tag{27}$$

where $\widehat{\mathbf{x}}_{MAP}(\mathbf{y})$ is attained by (15). 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 exponentially complex, can be solved in linear complexity, given in Section 4. For comparison purposes, the LMMSE detector is:

$$\widehat{\mathbf{x}}_{D-LMMSE}(\mathbf{y}) = \text{quantize}\left(\widehat{\mathbf{x}}_{LMMSE}(\mathbf{y})\right), \tag{28}$$

where $\widehat{\mathbf{x}}_{LMMSE}(\mathbf{y})$ is given by (5).

6. SIMULATION RESULTS

In this section numerical results illustrating the behavior of our new estimator in a MIMO system under a range of variance values for the random matrix $\mathbf{G}$ are provided. 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. We ran 5000 computer simulation for every σ_w^2. The symbols vector $\mathbf{x}$ follows eq. (24) and is shown in Fig. 2. Each symbol maps 3 bits using gray labeling, thus, neighboring symbols differ by only one bit. The simulation results for the BER and the 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. The appearance of the error floor in both Figs 3 and 4 is due to the uncertainty in the mixing matrix $\mathbf{G}$ and can not be mitigated, even in high SNR values.

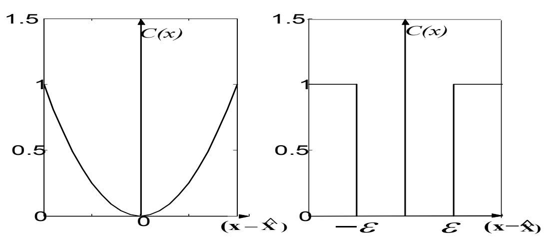

Fig. 1. Quadratic cost function (left), hit-or-miss cost function (right)

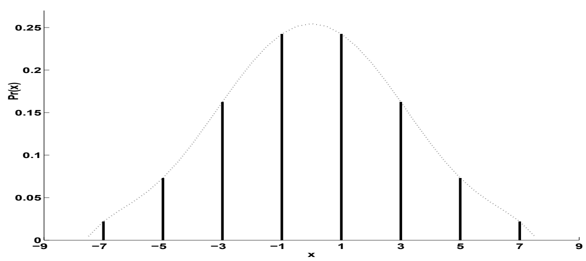

Fig. 2. Near-Gaussian constellation of 8 symbols.

7. 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 using BEM and provided an efficient method for finding it. Next, we proposed a detection scheme for near-Gaussian-digitally modulated symbols with linear complexity. 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.

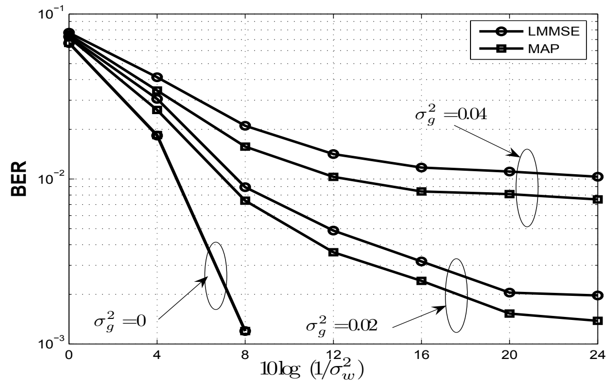

Fig. 3. Bit error rate of a MIMO system with $N = 40$, $K = 4$ for various values of σ_g^2.

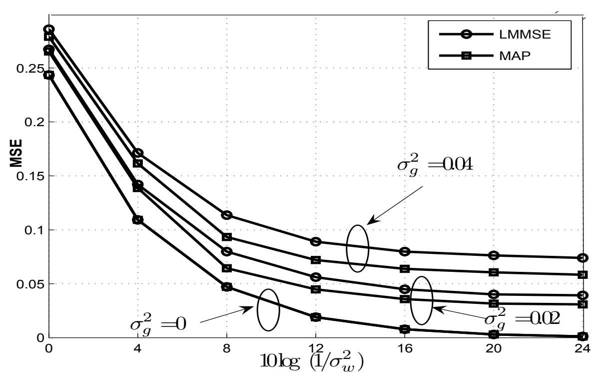

Fig. 4. Mean square error of a MIMO system with $N = 40$, $K = 4$ for various values of σ_g^2.

8. 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*, 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*, 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] I. Nevat, A. Wiesel, J. Yuan, and Y. C. Eldar, "Maximum A-Posteriori estimation in linear models with a gaussian model matrix," *Proc. of conference on Information Sciences and Systems, Baltimore, (CISS-2007)*, March 2007.

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

[8] A. P. Dempster, N . M. Laird, and D . B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, in *Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pp. 138*, December 1977.

[9] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913-929, May 1993.