Analysis of Performance Degradation using Convex Optimization for a Mismatched Receiver

Kung Yao
Electrical Engineering Dept., UCLA, Los Angeles, CA USA
Email: yao@ee.ucla.edu

Flavio Lorenzelli
Electrical Engineering Dept., UCLA, Los Angeles, CA USA
Email: flavio.lorenz@gmail.com

***Abstract* - Consider a binary baseband vector-valued communication channel modeled by a zero-mean CGN vector N with a non-singular covariance matrix Λ. We study the maximum loss of system performance using the metric of a decrease of P_D for a fixed P_{FA}. Under H_0, the observed vector is given by x = n, while under H_1, x = s + n. The optimum receiver compares the statistic $x^T\Lambda^{-1}s$ to a threshold γ determined by P_{FA}. However, the sub-optimum mismatched receiver assumes a WGN with a statistic given by $x^T s$, with its P_D^{sub} satisfying $P_D^{sub} \le \mathrm{Q}\left(\mathrm{Q}^{-1}(P_{FA}) - \sqrt{\mathbf{s}^T\mathbf{\Lambda}^{-1}\mathbf{s}}\right) = P_D^{opt}$, which is equal to $\|\mathbf{s}\|^4 \le \left(\mathbf{s}^T\mathbf{\Lambda}\mathbf{s}\right)\left(\mathbf{s}^T\mathbf{\Lambda}^{-1}\mathbf{s}\right)$, and is satisfied by Schwarz Inequality. For a given Λ, the solution for finding the maximum degraded (i.e., smallest) P_D^{sub} is equivalent to finding the signal vector s that attains the maxium of $\left(\mathbf{s}^T\mathbf{\Lambda}\mathbf{s}\right)\left(\mathbf{s}^T\mathbf{\Lambda}^{-1}\mathbf{s}\right)$, which uses the convex optimization method based on the Karush-Kuhn-Tucker conditions. The mismatched system performance degradation is formulated as a margin loss in SNR(dB) useful for robust communication system engineering analysis and design. Some explicit examples illustrating the margin loss are given.**

***Index Terms* - Digital communication system performance degradation, mismatched receiver, margin loss, convex optimization, KKT condition.**

I. INTRODUCTION

In the study of digital communication theory/system, one basic problem is the understanding of the maximum loss of performance of an optimized system assuming certain channel parameters when these assumed parameters are invalid. Practical wireless communication systems may encounter complicated fading phenomena and various copper wired communication systems operating in crowded environments (e.g., with transmission cables inside wiring bundles in an airplane or on a ship), the transmission channels may experience severe cross-talk interferences. In all these scenarios, realistic modeling of channel parameters may be difficult. In order to formulate an analytically tractable problem, we start with the assumption of the simplest model of a binary antipodal digital communication system with an additive white Gaussian noise (WGN) transmission channel. In reality, the transmission channel maybe quite complicated, but we model it as an arbitrary additive colored Gaussian noise (CGN) channel. Then we want to investigate the worst system performance degradation, when the receiver is designed for WGN disturbance, but in practice it is facing a CGN disturbance. Thus, we are studying the robustness of a complex communication system mismatch problem. A potential application of this study is that we can formulate the concept of requiring additional margin in the signal-to-noise (SNR) operating point of the system in order to maintain the original desired system performance in the presence of this mismatched receiver design. We will show the explicit use of these SNR margins in Ex. 7 considered in Sec. 3. The use of a SNR margin due to propagation path loss, as well as the use of a SNR margin due to received mismatch, can all be considered to be standard tools in practical radio engineering analysis and design [2].

Consider an uncoded binary baseband communication system with an additive CGN channel. Specifically, we study the loss of system performance, as the decrease of probability of detection (P_D) for a fixed probability of false alarm (P_{FA}). We assume

Manuscript received May 27, 2009; revised March 8, 2010; accepted March 15, 2010.

doi:10.4304/jcm.5.6.483-492

under hypothesis H_0 there is only the n×1 zero-mean CGN vector **N** with a n×n non-singular covariance matrix Λ, and under hypothesis H_1, there is a n×1 deterministic signal vector **s** in addition to the noise vector **N**. Thus, under H_0, the observed vector is denoted by $\mathbf{x} = \mathbf{n}$, while under H_1, $\mathbf{x} = \mathbf{s} + \mathbf{n}$. However, the receiver assumes the noise is a zero-mean white Gaussian noise (WGN) with a covariance matrix $\Lambda = \sigma^2 \boldsymbol{I}_n$. There are various reasons for this possible mismatch when the channel is truly CGN, but the receiver operates as if it is WGN. An obvious reason is that the receiver is not capable of estimating the statistics of the CGN or the low-cost receiver is willing to accept the loss of performance. In any case, it is interesting to evaluate the worst case loss of performance under any given CGN scenario.

II. OPTIMUM AND SUB-OPTIMUM SYSTEM PERFORMANCES

First, the optimum receiver compares the observed statistic $\eta_{\mathrm{CGN}} = \mathbf{x}^{\mathbf{T}}\Lambda^{-1}\mathbf{s}$ to a threshold determined by the P_{FA}. From detection theory [1], for a specifed P_{FA} given by

$$P_{FA} = \mathrm{Q}\left(\frac{\gamma + (1/2)\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}} \right), \tag{1}$$

it yields γ and the optimum P_D is then given by

$$P_D^{opt} = \mathrm{Q}\left(\frac{\gamma - (1/2)\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}} \right), \tag{2}$$

where Q(.) is the complementary Gaussian distribution function. If the P_{FA} is fixed at some acceptable value (e.g., set $P_{FA} = 10^{-3}$, which yields $\mathrm{Q}^{-1}(P_{FA}) = \mathrm{Q}^{-1}(10^{-3}) = 3.09$), then the threshold constant γ can be solved from (1) as

$$\gamma = \sqrt{\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}\,\mathrm{Q}^{-1}(P_{FA}) - (1/2)\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}\,, \tag{3}$$

where $\mathrm{Q}^{-1}(.)$ is the inverse of complementary Gaussian distribution function. This γ is then used to obtain the optimum P_D in (2) to yield

$$\begin{aligned} P_D^{opt} &= \mathrm{Q}\left(\frac{\sqrt{\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}\,\mathrm{Q}^{-1}(P_{FA}) - \mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}}} \right) \\ &= \mathrm{Q}\left(\mathrm{Q}^{-1}(P_{FA}) - \sqrt{\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}} \right). \end{aligned} \tag{4}$$

(4) shows the optimum P_D (for a fixed P_{FA}) depends only on the factor $\mathbf{s}^T \mathbf{\Lambda}^{-1}\mathbf{s}$, which depends on the signal vector **s** and Λ. From **s** and Λ, they define the SNR (in dB) as

$$SNR(dB) = 10 \times \log_{10}\left(\frac{\| \mathbf{s} \|^2}{Trace(\mathbf{\Lambda})} \right). \tag{5}$$

On the other hand, suppose the observation noise is colored with a covariance matrix Λ, but we assume the noise is "white" with a covariance matrix $\Lambda = \sigma^2\, \mathbf{I}_{\mathrm{M}}$. Then the suboptimum "white matched filter" receiver operating in colored noise using a decision statistic of $\eta_{\mathrm{WGN}} = \mathbf{x}^T\mathbf{s}$, resulting in a P_{FA} of

$$P_{FA} = \mathrm{Q}\left(\frac{\gamma_0}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}\mathbf{s}}} \right), \tag{6}$$

where γ_0 is taken to yield the desired P_{FA}, and the suboptimum P_D is given by

$$P_D^{sub} = \mathrm{Q}\left(\frac{\gamma_0 - \mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}\mathbf{s}}} \right). \tag{7}$$

If the P_{FA} is fixed at some value, then the threshold constant γ_0 can be solved from (6) as

$$\gamma_0 = \sqrt{\mathbf{s}^T \mathbf{\Lambda}\mathbf{s}}\;\mathrm{Q}^{-1}(P_{FA}). \tag{8}$$

Substituting γ_0 from (8) into (7), we obtain

$$\begin{aligned} P_D^{sub} &= \mathrm{Q}\left(\frac{\sqrt{\mathbf{s}^T \mathbf{\Lambda}\mathbf{s}}\;\mathrm{Q}^{-1}(P_{FA}) - \mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}\mathbf{s}}} \right) \\ &= \mathrm{Q}\left(\mathrm{Q}^{-1}(P_{FA}) - \frac{\mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T \mathbf{\Lambda}\mathbf{s}}} \right). \end{aligned} \tag{9}$$

Since the expression in (2) represents the optimum (i.e., largest) P_D^{opt}, we want to show that expression is greater or equal to the suboptimum P_D^{sub} of (9). Thus, we want to show

$$P_D^{sub} = \mathrm{Q}\left(\mathrm{Q}^{-1}(P_{FA}) - \frac{\mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}}} \right) \le \mathrm{Q}\left(\mathrm{Q}^{-1}(P_{FA}) - \sqrt{\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}} \right) = P_D^{opt}. \quad (10)$$

Since Q(.) is a monotonically decreasing function, for the inequality in (10) to hold, it is equivalent to show that for any **s** and Λ, the parameters χ^{opt} associated with P_D^{opt} and χ^{sub} associated with P_D^{sub} in (11) must satisfy

$$\chi^{\mathrm{sub}} = \frac{\mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}}} \le \sqrt{\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}} = \chi^{opt} \quad , \quad (11)$$

or equivalently,

$$\|\mathbf{s}\|^4 \le \left(\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}\right)\left(\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}\right) . \quad (12)$$

However, Schwarz Inequality shows the inequalities in (11) - (12) are always valid (Appendix 1). In particular, when Λ is a WGN covariance matrix satisfying $\Lambda = \sigma^2 I_M$, then $\Lambda^{-1} = (1/\sigma^2) I_M$, and equality is attained in (12). We also note that any signal vector **s** in (12) is invariant to a scalar multiplication factor.

Furthermore, from (11) denote the mismatched factor γ by

$$\gamma = \chi^{opt} / \chi^{sub} = \sqrt{\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}}\sqrt{\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}} / \mathbf{s}^T\mathbf{s} \ge 1. \quad (13)$$

When Λ is a WGN covariance matrix, then $\gamma = 1$ for all signal vector **s**. In general, γ is a function of the CGN Λ and **s**. For a fixed but arbitrary non-singular Λ, different **s** vectors yield values of $\gamma \ge 1$. Denote $\chi_{\mathbf{s}}^{opt} = \sqrt{\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}}$ when we use the vector **s** in the optimum system. Suppose we use $\tilde{\mathbf{s}} = \gamma\mathbf{s}$ in the suboptimum system. Then

$$\chi_{\tilde{\mathbf{s}}}^{sub} = \tilde{\mathbf{s}}^T\tilde{\mathbf{s}} / \sqrt{\tilde{\mathbf{s}}^T\boldsymbol{\Lambda}\tilde{\mathbf{s}}} = \gamma^2\mathbf{s}^T\mathbf{s} / \gamma\sqrt{\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}} = \gamma\chi_{\mathbf{s}}^{sub} = \chi_{\mathbf{s}}^{opt}. \quad (14)$$

By using $\tilde{\mathbf{s}} = \gamma\mathbf{s}$ in the suboptimum system, which represents an increase of its energy by a factor of γ^2 relative to the energy of $\|\mathbf{s}\|^2$ in the optimum system, (14) is achieved resulting in identical values of $P_D^{sub}|_{\tilde{\mathbf{s}}} = P_D^{opt}|_{\mathbf{s}}$. Thus, we can consider the Margin Loss in SNR (dB) of performance due to the mismatched of the CGN covariance matrix Λ and the signal vector **s**

$$\text{Margin Loss (dB)} = 10 \log_{10}(\gamma^2). \quad (15)$$

From the above discussions, in practice given an arbitrary CGN covariance matrix Λ, how much loss of performance may be realized for various possible values of the signal vector **s**?

Now, considered an explicit example to illustrate the advantage of using the colored matched filter over a white matched filter in the presence of a colored noise.

Ex. 1. Consider a Markov covariance matrix Λ

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & r & \cdots & r^{n-1} \\ r & 1 & r & \vdots \\ \vdots & \cdots & \ddots & r \\ r^{n-1} & \cdots & r & 1 \end{bmatrix}, \quad (16)$$

Suppose we select 5 pseudo-randomly generated n×1 signal vector **s** for n = 2, 3, and 50, with r = 0.1 and 0.9, with the constraint of $\mathrm{Q}^{-1}(P_{FA}) = 3.09$. Now, we tabulate 30 simulation results (i.e., 5 signal vectors × 3 values of n × 2 values of r) in terms of the mismatched factor $\gamma = \chi^{opt} / \chi^{sub}$ of (16). Table 1 shows one set of such 30 simulation results for the five realizations.

n = 2

SNR(dB)	-1.96	-9.93	-1.76	-9.60	-1.51

$\gamma = \chi^{opt} / \chi^{sub}$

r	#1	#2	#3	#4	#5
0.1000	1.0006	1.0021	1.0002	1.0006	1.0000
0.9000	1.2139	1.6591	1.0826	1.2158	1.0108

n = 3

SNR(dB)	-2.66	-10.58	-1.48	-7.66	-3.16

$\gamma = \chi^{opt} / \chi^{sub}$

r	#1	#2	#3	#4	#5
0.1000	1.0021	1.0047	1.0022	1.0047	1.0015
0.9000	1.2834	1.8340	1.1951	1.7868	1.8814

n = 50

SNR(dB)	-5.00	-5.42	-3.50	-5.58	-6.31
$\gamma = \chi^{opt} / \chi^{sub}$					
r	#1	#2	#3	#4	#5
0.1000	1.0047	1.0094	1.0037	1.0050	1.0067
0.9000	4.5241	6.2867	3.9924	4.8558	5.2242

Table 1. Values of *SNR(dB)* and $\gamma = \chi^{opt} / \chi^{sub}$ for a Markov covariance matrix with r = 0.1 and 0.9 of dimensions n = 2, 3, and 50 for five rand uniform pseudo-random generated signal vectors **s** with the $P_{FA} = 10^{-3}$ constraint.

From (11) - (12), we note if r = 0, then the CGN problem reduces to the WGN problem. Thus, for the small value of r = 0.1, all the $\gamma = \chi^{opt} / \chi^{sub}$ values are only slightly greater than unit value for all n = 2, 3, and 50 cases. However, as r increases to r = 0.9, we note the ratios of $\gamma = \chi^{opt} / \chi^{sub}$ increase greatly as the dimension increases to n = 50 for all five realizations, showing the system performance degradations due to the mismatch. □

III. CONVEX OPTIMIZATION FOR EVALUATING THE MAXIMUM LOSS OF PERFORMANCES

The results shown in Table1 of **Ex. 1** were obtained from specific realizations in the simulations. An interesting question is for any given n×n colored non-singular covariance matrix Λ, what n×1 signal vector **s** of unit norm (i.e., $\|\mathbf{s}\|^2 = 1$) will theoretically yield the smallest (worst performing) P_D^{sub} in (10)? This is asking what signal vector **s** provides the largest advantage in using the colored matched detector over a white detector. Equivalently, what **s** will yield the largest ratio $\gamma = \chi^{opt} / \chi^{sub}$ in (11)? Or equivalently, what **s** will attain the maximum of the product of the two factors $(\mathbf{s}^T \mathbf{\Lambda} \mathbf{s})(\mathbf{s}^T \mathbf{\Lambda}^{-1} \mathbf{s})$ in (12)? We already know Schwarz Inequality was able to provide readily a lower bound on $(\mathbf{s}^T \mathbf{\Lambda} \mathbf{s})(\mathbf{s}^T \mathbf{\Lambda}^{-1} \mathbf{s})$ in (12). As we will show below, finding the maximum of $(\mathbf{s}^T \mathbf{\Lambda} \mathbf{s})(\mathbf{s}^T \mathbf{\Lambda}^{-1} \mathbf{s})$ is more complicated.

We first note, in a matrix eigenvalue problem, a non-zero constant c times each eigenvector is still an eigenvector. Thus, there is no loss of generality, if we constrain all the vectors **s** under consideration in (11) or (12) to have unit norms (i.e., $\|\mathbf{s}\|^2 = 1$). Given any n×n non-singular covariance matrix Λ, perform an eigenvalue decomposition resulting in **ΛU** = **UD**, where **U** is an orthogonal matrix defined by $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$, where $\mathbf{u}_i$, i = 1,..., n, are the orthonormal eigenvectors of **Λ**, and **D** is a diagonal matrix of the eigenvalues of **Λ** with

$$\mathrm{diag}(D) = [\lambda_1, \ldots, \lambda_n] = \boldsymbol{\lambda}^T , \tag{17}$$

where we order the positive-valued eigenvalues to be in an descending manner satisfying

$$\lambda_1 > \lambda_2 > \ldots \lambda_n > 0 . \tag{18}$$

Thus,

$$\mathbf{\Lambda} = \mathbf{U}\mathbf{D}\mathbf{U}^T , \tag{19}$$

and

$$\mathbf{\Lambda}^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T. \tag{20}$$

We note, **U** is an orthogonal matrix. Define a new n×1 vector

$$\mathbf{z} = \mathbf{U}^T \mathbf{s} . \tag{21}$$

In particular,

$$\begin{aligned} \|\mathbf{z}\|^2 &= (\mathbf{z}, \mathbf{z}) = \mathbf{z}^T \mathbf{z} = \mathbf{s}^T \mathbf{U} \mathbf{U}^T \mathbf{s} \\ &= \mathbf{s}^T \mathbf{s} = (\mathbf{s}, \mathbf{s}) = \|\mathbf{s}\|^2 = 1 . \end{aligned} \tag{22}$$

In other words, the vector **z** has the same norm as that of **s** which was earlier set equal to 1. From (20), we have

$$\mathbf{s} = \mathbf{U} \mathbf{z} \tag{23}$$

or

$$\mathbf{s}^T = \mathbf{u}^T \mathbf{U}^T . \tag{24}$$

Now, substitute (23) and (24) into (11) and use **Λ** of (19) and Λ^{-1} of (20). Thus, (11) becomes

$$\begin{aligned}&\left(\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}\right)\ \left(\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}\right)\\&=\left(\mathbf{z}^T\mathbf{U}^T\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T\mathbf{U}\mathbf{z}\ \right)\left(\mathbf{z}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{U}^T\mathbf{U}\mathbf{z}\right)\\&=\left(\mathbf{z}^T\mathbf{D}^{-1}\mathbf{z}\ \right)\left(\mathbf{z}^T\mathbf{D}\mathbf{z}\right)\\&=\left(\sum_{i=1}^{M} z_i^2/\lambda_i\right)\left(\sum_{i=1}^{M} z_i^2\lambda_i\right)\geq 1.\end{aligned} \tag{25}$$

From (25), we can define n new variables

$$y_i = z_i^2 \ \geq\ 0\ ,\ i=1,...,n. \tag{26}$$

Then (25) becomes

$$\left(\sum_{i=1}^{M} y_i/\lambda_i\right)\left(\sum_{i=1}^{M} y_i\lambda_i\right)\geq 1\ , \tag{27}$$

where

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i^2 = 1\ . \tag{28}$$

The optimization problem is to maximize the left-hand-side (lhs) of the expression of (27) with respect to the $\{y_i,\ i = 1,\dots,M\}$ subject to the two constraints that they are non-negative-valued (i.e., (26)) whose sum is equal to 1 (i.e., (28)). In other words, we want the lhs of (27) to be as large as possible subject to the constraints of that they are negative-valued whose sum is equal to 1. If the equality in (27) is attained, this means the performance of the sub-optimum WGN sufficient statistic detector (when the noise is CGN) is equal to the performance of the optimum CGN sufficient statistic detector. That is, equality in (27) is equivalent to equality in (10) or equivalently equality in (9) is attained. Now, consider three new simple examples.

Ex. 2. Consider the case when Λ is a WGN autocorrelation matrix, where all the eigenvalues $\lambda_i = \lambda$, $i = 1, ..., n$. From (26), the lhs of (25) is equal to one since it is the product of two terms each equal to one. No maximization of the lhs of (25) is possible. Thus, the lhs of (25) attains the equality on the rhs of (25), and and $\mathrm{P}_{\mathrm{D}}^{\mathrm{sub}} = \mathrm{P}_{\mathrm{D}}^{\mathrm{opt}}$. □

Ex. 3. Take $y_1 = 1$ and all the other $y_i = 0$, $i = 2,\dots,n$. Then (25) reduces to

$$(1/\lambda_1)\ (\lambda_1) = 1. \tag{29}$$

Then equality in (25) is attained. However , this is not an optimum choice for $\{y_i,\ i = 1,\dots,n\}$ to maximizes the lhs of (25). □

Ex. 4. Take $y_M = 1$ and all the other $y_i = 0$, $i = 1,\dots, n-1$. Then (25) reduces to

$$(1/\lambda_n)\ (\lambda_n) = 1. \tag{30}$$

Then equality in (25) is attained. However, this is also not an optimum choice for $\{y_i,\ i = 1,\dots,n\}$ to maximizes the lhs of (25). □

Lemma. Consider the case of $n = 2$. Then $\hat{y}_1 = \hat{y}_2 = 1/2$ is the optimum choice of the $\{y_i,\ i = 1, 2\}$ that maximizes the lhs of (25).

Proof:
From (26), we can express $y_2 = 1 - y_1$. Then the lhs of (25) can be expressed as

$$\begin{aligned}G(y_1) &= \left(y_1/\lambda_1 + (1-y_1)/\lambda_2\right)\times\\&\left(y_1\lambda_1 + (1-y_1)\lambda_2\right)\\&= y_1^2 + y_1(\lambda_1/\lambda_2) - y_1^2(\lambda_1/\lambda_2) + 1\\&-2y_1 + y_1^2 + y_1(\lambda_2/\lambda_1) - y_1^2(\lambda_2/\lambda_1)\\&= y_1^2(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1))\\&-y_1(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1))+1\ .\end{aligned} \tag{31}$$

We note $G(y_1)$ is a quadratic function of y_1. Then by setting its first derivative to zero,

$$\begin{aligned}G^{'}(y_1) &= 2y_1(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1))\\&-(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1)) = 0\ ,\end{aligned} \tag{32}$$

we find

$$\hat{y}_1 = \frac{(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1))}{2(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1))} = \frac{1}{2}. \tag{33}$$

Furthermore, since the second derivative of $G(y_1)$ is negative

$$\begin{aligned} G''(y_1) &= 2(2-(\lambda_1/\lambda_2)-(\lambda_2/\lambda_1)) \\ &= \frac{-2(\lambda_1^2+\lambda_2^2-2\lambda_1\lambda_2)}{\lambda_1\lambda_2} \\ &= \frac{-2(\lambda_1-\lambda_2)^2}{\lambda_1\lambda_2} < 0 \,, \end{aligned} \tag{34}$$

then the local maximum solution of (31) for $\hat{y}_1 = \hat{y}_2 = 1/2$ is a global maximum of the quadratic function on the lhs of (25). □

Theorem. The maximum of the lhs of (35) given by

$$\left(\sum_{i=1}^{n} y_i / \lambda_i\right)\left(\sum_{i=1}^{n} y_i \lambda_i\right) \geq 1 \,, \tag{35}$$

subject to

$$\sum_{i=1}^{n} y_i = 1 \,,\ y_i \geq 0 \,,\ i = 1,\ldots,n, \tag{36}$$

$$\lambda_1 > \lambda_2 > \ldots \lambda_n > 0,$$

$$\hat{y}_1 = \hat{y}_n = 1/2,\ \hat{y}_i = 0 \,,\ i = 2,\ldots,n-1 \,. \tag{37}$$

Proof: The proof is given in Appendix 2. □

Ex. 5. Consider the $n = 2$ case where the CGN covariance matrix spectral decomposition yields an orthogonal matrix **U** and diagonal $\tilde{\mathbf{\Lambda}}$ given by

$$\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tilde{\mathbf{\Lambda}} = \frac{\sigma^2}{100.01}\begin{bmatrix} 100 & 0 \\ 0 & 0.01 \end{bmatrix}. \tag{38}$$

For $P_{FA} = 10^{-3}$ at SNR(dB) = 10 dB, using the optimum $\hat{\mathbf{s}}$ of

$$\hat{\mathbf{s}} = (\mathbf{\theta}_1 + \mathbf{\theta}_n)/\sqrt{2} \,, \tag{39}$$

we obtain $P_D^{opt} = 1$ and $P_D^{sub} = 0.9165$. □

Ex. 6. Consider the $n = 3$ case where the CGN covariance matrix spectral decomposition yields an orthogonal matrix **U** and diagonal $\tilde{\mathbf{\Lambda}}$ given by

$$\begin{aligned} \mathbf{U} &= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & -1 \end{bmatrix}, \\ \tilde{\mathbf{\Lambda}} &= \frac{\sigma^2}{101.01}\begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}. \end{aligned} \tag{40}$$

For $P_{FA} = 10^{-3}$ at SNR(dB) = 10 dB, using the optimum $\hat{\mathbf{s}}$ of (39), we obtain $P_D^{opt} = 1$ and $P_D^{sub} = 0.9199$. □

Ex. 7. Consider the problem using the Markov covariance matrix Λ of (13) in Ex. 1. By using the optimum $\hat{\mathbf{s}}$ of (23), (26), and (37) to evaluate the ratio of $\gamma\,|_{\mathbf{s}=\hat{\mathbf{s}}} = \left(\chi^{opt}/\chi^{sub}\right)|_{\mathbf{s}=\hat{\mathbf{s}}}$, we obtained the following results:

	$\gamma\,\|_{\mathbf{s}=\hat{\mathbf{s}}} = \left(\chi^{opt}/\chi^{sub}\right)\|_{\mathbf{s}=\hat{\mathbf{s}}}$	Margin Loss (dB)
n = 2		
r = 0.10	1.0050	0.043
r = 0.90	2.2942	7.213
n = 3		
r = 0.10	1.0101	0.087
r = 0.90	3.2233	10.16
n = 50		
r = 0.10	1.0196	0.169
r = 0.90	8.7236	18.81

Table 2. Values of $\gamma\,|_{\mathbf{s}=\hat{\mathbf{s}}} = \left(\chi^{opt}/\chi^{sub}\right)|_{\mathbf{s}=\hat{\mathbf{s}}}$ for n = 2, 3, and 50 for r = 0.10 and r = 0.90 .

We note while the values of χ^{opt} and χ^{sub} depend on the value of the SNR, $\gamma\,|_{\mathbf{s}=\hat{\mathbf{s}}} = \left(\chi^{opt}/\chi^{sub}\right)|_{\mathbf{s}=\hat{\mathbf{s}}}$ is independent of the SNR value. We also note that for any given n and r, all the $\gamma\,|_{\mathbf{s}=\hat{\mathbf{s}}} = \left(\chi^{opt}/\chi^{sub}\right)|_{\mathbf{s}=\hat{\mathbf{s}}}$ in Table 2 are greater than all five of the pseudo-randomly generated corresponding values of $\gamma = \chi^{opt}/\chi^{sub}$ in Table 1 of Ex. 1. The results in Table 2 also show the corresponding margin loss as a function of n and r. Specifically, for n = 50 and r = 0.9, for this Markov covariance matrix, by using the worst signal vector $\hat{\mathbf{s}}$ (obtained from the "optimum" solution of the

Theorem), there is a margin loss of 18.81 dB. In other words, by using this signal vector $\hat{\mathbf{s}}$, we need to increase its energy by a factor of $(8.7236)^2$ or 18.21 dB in order to achieve the same P_D^{opt} as in the optimum colored matched filter receiver. □

Ex. 8. Consider again the problem using the 50×50 Markov covariance matrix $\boldsymbol{\Lambda}$ of (15) in Ex. 1.

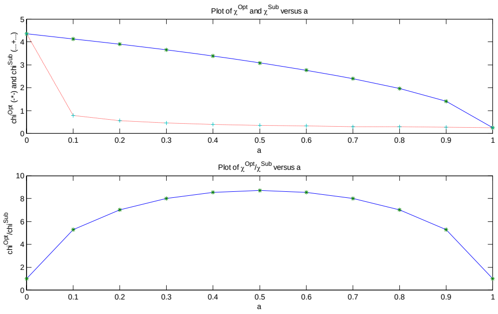

Fig. 1a. Plots of χ^{opt} (solid curve) and χ^{sub} (dotted curve) versus a. Fig. 1b. Plot of χ^{opt} / χ^{sub} versus a. Fig. 1a shows the plots of χ^{opt} (solid curve) and χ^{sub} (dotted curve) versus a, where the signal vector $\mathbf{s} = \sqrt{1\text{-a}}\boldsymbol{\theta}_{50} + \sqrt{\text{a}}\boldsymbol{\theta}_1$, $0 \leq a \leq 1$. As expected χ^{opt} is greater than χ^{sub}, except when they have equal values for a = 0 and a = 1. Fig. 1b shows the plot of χ^{opt} / χ^{sub} versus a. Also for a = ½, $\mathbf{s} = (\boldsymbol{\theta}_{50} + \boldsymbol{\theta}_1)/\sqrt{2}$ equal $\hat{\mathbf{s}}$ of (47). This shows that $\chi^{opt}/\chi^{\text{sub}}$ achieves its maximum $(\chi^{opt}/\chi^{sub})\big|_{\mathbf{s}=\hat{\mathbf{s}}}$ using the optimum $\hat{\mathbf{s}}$ as expected from theory. □

IV. CONCLUSIONS

In this paper, we considered the loss of performance of a binary baseband communication system operating in the presence of a given CGN covariance matrix, when the receiver assumes the noise to have a WGN covariance matrix. Modern convex optimization method based on the Karush-Kuhn-Tucker condition is used to solve this problem. Various examples using simulation and analysis illustrate various aspects of mismatched system performance degradation. The concept of margin loss was also introduced for the analysis and design of robust system in the presence of mismatched parameters in the system.

Appendix 1. Use Schwarz Inequality to obtain the lower bound of $\left(\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}\right)\left(\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}\right)$ in (12). For any column vectors **a** and **b**, Schwarz Inequality states that

`` $$(\mathbf{a}^T\mathbf{b})^2 \leq (\mathbf{a}^T\mathbf{a})(\mathbf{b}^T\mathbf{b}) \,. \tag{A1}$$

Now, pick

$$\begin{aligned} &\mathbf{a} = \boldsymbol{\Lambda}^{1/2}\mathbf{s} \ (\text{or } \mathbf{a}^T = \mathbf{s}^T\boldsymbol{\Lambda}^{1/2}), \\ &\mathbf{b} = \boldsymbol{\Lambda}^{-1/2}\mathbf{s} \ (\text{or } \mathbf{b}^T = \mathbf{s}^T\boldsymbol{\Lambda}^{-1/2}) \,. \end{aligned} \tag{A2}$$

Then

$$\begin{aligned} &\mathbf{a} = \boldsymbol{\Lambda}^{1/2}\mathbf{s} \ (\text{or } \mathbf{a}^T = \mathbf{s}^T\boldsymbol{\Lambda}^{1/2}), \\ &\mathbf{b} = \boldsymbol{\Lambda}^{-1/2}\mathbf{s} \ (\text{or } \mathbf{b}^T = \mathbf{s}^T\boldsymbol{\Lambda}^{-1/2}) \,. \\ &\mathbf{a}^T\mathbf{b} = \mathbf{s}^T\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{s} = \mathbf{s}^T\mathbf{s}, \\ &\mathbf{a}^T\mathbf{a} = \mathbf{s}^T\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{s} = \mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}, \\ &\mathbf{b}^T\mathbf{b} = \mathbf{s}^T\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{s} = \mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s} \,. \end{aligned} \tag{A3}$$

Using (A3) in (A1), yields

$$\|\mathbf{s}\|^4 = (\mathbf{s}^T\mathbf{s})^2 \leq \left(\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}\right)\left(\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}\right), \tag{A4}$$

which is identical to (12) or equivalently

$$\|\mathbf{s}\|^2 = \mathbf{s}^T\mathbf{s} \leq \left(\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}\right)^{1/2}\left(\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}\right)^{1/2}, \tag{A5}$$

or equivalently

$$\frac{\mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T\boldsymbol{\Lambda}\mathbf{s}}} \leq \sqrt{\mathbf{s}^T\boldsymbol{\Lambda}^{-1}\mathbf{s}} \,, \tag{A6}$$

which is identical to (11). □

Appendix 2. Proof of Theorem.

Consider the nonlinear minimization problem of

$$\text{Min}_{\,y}\ \text{f}(\boldsymbol{y}) \,, \tag{A7}$$

subject to

$$g_i(\boldsymbol{y}) \leq 0 \,,\ i = 1, \ldots, M, \tag{A8}$$

$$h(\boldsymbol{y}) = 0 \,. \tag{A9}$$

The celebrated Karush-Kuhn-Tucker (KKT) [3] necessary conditions for $\mathbf{y} = \hat{\mathbf{y}} = [\hat{y}_1, ..., \hat{y}_M]^T$ to be a local minimum solution of (A7) subject to (A8) and (A9), are such that there exist constants μ_I, $i = 1,\ldots, M$, and υ satisfying

$$1. \ \nabla f(\hat{\mathbf{y}}) + \sum_{i=1}^{M} \mu_i \nabla g_i(\hat{\mathbf{y}}) + \upsilon \nabla h(\hat{\mathbf{y}}) = 0, \quad \text{(A10)}$$

$$2. \ g_i(\hat{\mathbf{y}}) \le 0, \ i = 1, ..., M, \quad \text{(A11)}$$

$$3. \ h(\hat{\mathbf{y}}) = 0, \ i = 1, ..., M, \quad \text{(A12)}$$

$$4. \ \mu_i \ge 0, \ i = 1, ..., M, \quad \text{(A13)}$$

$$5. \ \mu_i g_i(\hat{\mathbf{y}}) = 0, \ i = 1, ..., M. \quad \text{(A14)}$$

In order to use the KKT method for our maximization of (A7) with the constraints of (A8) and (A9), denote

$$f(\hat{\mathbf{y}}) = -A(i)B(i), \quad \text{(A15)}$$

$$g_i(\hat{\mathbf{y}}) = -y_i \ , \ i = 1, ..., M, \quad \text{(A16)}$$

$$h(\hat{\mathbf{y}}) = \sum_{i=1}^{M} y_i - 1, \quad \text{(A17)}$$

$$\lambda_1 > \lambda_2 > \ldots \lambda_M > 0 \ , \quad \text{(A18)}$$

where we define

$$A(i) = \sum_{i=1}^{M} y_i / \lambda_i \,, \quad B(i) = \sum_{i=1}^{M} y_i \lambda_i \,. \quad \text{(A19)}$$

Now, we show the conditions of (1)- (5) of (A10) – (A14) are satisfied for the expressions of (A15) - (A18). From Condition 1, by taking its partial derivative wrt to y_i, $i = 1, \ldots, M$, we have

$$\frac{-1}{\lambda_i} B(j) - \lambda_i A(j) - \mu_i + \upsilon = 0. \quad \text{(A20)}$$

Multiply (A20) by y_i yields

$$\frac{-y_i}{\lambda_i} B(j) - \lambda_i y_i A(j) - \mu_i y_i + \upsilon y_i = 0. \quad \text{(A21)}$$

From Condition 4 and $y_i \ge 0$, then $-\mu_i y_i = 0$. Thus, (A21) becomes

$$\frac{-y_i}{\lambda_i} B(j) - \lambda_i y_i A(j) + \upsilon y_i = 0. \quad \text{(A22)}$$

Summing (A22) over all $i = 1, \ldots, M$, yields

$$\upsilon \sum_{i=1}^{M} y_i = A(i)B(j) + B(i)A(j),$$

and from (A8), we have

$$\upsilon = A(i)B(j) + B(i)A(j) = 2A(j)B(j). \quad \text{(A23)}$$

Substitute (A23) into (A22) yields

$$-\frac{y_i}{\lambda_i} B(j) - \lambda_i y_i A(j) + 2 y_i B(j) A(j) = 0,$$

or

$$y_i \left[\frac{1}{\lambda_i} B(j) + \lambda_i A(j) - 2B(j)A(j) \right] = 0, \quad \text{(A24)}$$

Thus, (A24) shows either $y_i = 0$ or

$$\frac{1}{\lambda_i} B(j) + \lambda_i A(j) - 2B(j)A(j) = 0. \quad \text{(A25)}$$

Multiply (A25) by λ_i, to obtain

$$\lambda_i^2 A(j) - 2\lambda_i B(j)A(j) + B(j) = 0. \quad \text{(A26)}$$

The quadratic equation of (A26) in λ_i, has two non-zero solutions given by

$$\lambda_i = \frac{2B(j)A(j) \pm \sqrt{4B(j)^2 A(j)^2 - 4B(j)A(j)}}{2A(j)}. \quad \text{(A27)}$$

The rest of the (M -2) number of $y_i = 0$. Denote the indices of the two non-zero y_i by a and b. Then the maximization of $A(i)B(i)$ in (A7)) reduces to the maximization of

$$\begin{aligned}
&\left(\frac{y_a}{\lambda_a}+\frac{y_b}{\lambda_b}\right)\left(y_a\lambda_a+y_b\lambda_b\right)\\
&=y_a^2+\frac{\lambda_a}{\lambda_b}y_ay_b+y_b^2+\frac{\lambda_b}{\lambda_a}y_ay_b\\
&=y_a^2+y_b^2+\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}\right)y_ay_b\\
&=\left(y_a+y_b\right)^2+\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)y_ay_b .
\end{aligned} \tag{A28}$$

By denoting $y_b = 1 - y_a$ in (A28), we obtain

$$\begin{aligned}
&\left(\frac{y_a}{\lambda_a}+\frac{y_b}{\lambda_b}\right)\left(y_a\lambda_a+y_b\lambda_b\right)\\
&=1+\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)(1-y_a)y_a\\
&\equiv H(y_a,\lambda_a,\lambda_b).
\end{aligned} \tag{A29}$$

Thus, the maximization in (A7)) reduces to the maximization of $H(y_a,\lambda_a,\lambda_b)$ in (A29). We note $H(y_a,\lambda_a,\lambda_b)$ is a quadratic function of y_a. Taking the partial derivative of $H(y_a,\lambda_a,\lambda_b)$ wrt to y_a yields

$$\frac{\partial H(y_a,\lambda_a,\lambda_b)}{\partial y_a}=-\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)y_a+\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)(1-y_a)=0,$$

or

$$2\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)y_a=\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right),$$

or

$$y_a=y_b=1/2, \tag{A30}$$

since

$$\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)\neq 0,$$

with $\lambda_a \neq \lambda_b$. The second partial derivative of $H(y_a,\lambda_a,\lambda_b)$ shows

$$\begin{aligned}
\frac{\partial^2 H(y_a,\lambda_a,\lambda_b)}{\partial y_a^2}&=-2\left(\frac{\lambda_a}{\lambda_b}+\frac{\lambda_b}{\lambda_a}-2\right)\\
&=-2\left(\frac{(\lambda_a+\lambda_b)^2}{\lambda_a\lambda_b}\right)<0,
\end{aligned}$$

since $\lambda_a > 0$ and $\lambda_b > 0$. Thus, the local maximum solution of $y_a = y_b = 1/2$ in (A30) is a global maximum solution of $H(y_a,\lambda_a,\lambda_b)$. By using $y_a = y_b = 1/2$ and denoting $\lambda = \lambda_a/\lambda_b > 1$, with the assumption of $\lambda_a > \lambda_b > 0$, then $H(y_a,\lambda_a,\lambda_b)$ of (A29) can be expressed as

$$\begin{aligned}
H(\lambda)&=1+(1/4)(\lambda-2+1/\lambda)\\
&=1+(1/4)\left(\frac{(\lambda+1)(\lambda-1)}{\lambda^2}\right)>0.
\end{aligned} \tag{A31}$$

This shows $H(\lambda)$ is a positive monotonically increasing function of λ. Since $\lambda = \lambda_a/\lambda_b$ and $\lambda_a > \lambda_b > 0$, the maximum of $H(\lambda)$ is attained by using $\lambda_{\max} = \lambda_1/\lambda_M$. This means

$$\hat{y}_1=\hat{y}_M=1/2,\hat{y}_i=0, i=2,\ldots,M-1. \tag{A32}$$

Thus, the solution given by (A32) is the only solution that satisfies the KKT Condition 1 of (A15) that provides the local minimum of

$f(\hat{\mathbf{y}})=-\left(\sum_{i=1}^{M}y_i/\lambda_i\right)\left(\sum_{i=1}^{M}y_i\lambda_i\right)$ or the local maximum of $\left(\sum_{i=1}^{M}y_i/\lambda_i\right)\left(\sum_{i=1}^{M}y_i\lambda_i\right)$.

But $H(y_a,\lambda_a,\lambda_b)$ of (A29) is a quadratic function, thus the solution given by (A32) yields the global maximum of (35). □

ACKNOWLEDGMENT

This work is partially supported by NSF grants EF-0410438 and DBI-0754247. An earlier shortened version of this work was presented at the CMC conference in April 2010. We want to thank conversations with Prof. L. Vandenberghe on various aspects of convex optimization. Comments from students in UCLA EE230A were also appreciated.

REFERENCES

[1] D. Middleton, "An Introduction to Statistical Communication Theory," McGraw-Hill, 1960.
[2] J.B. Anderson, "Digital Transmission Engineering," IEEE Press, 1999.
[3] D.P. Bertsekas, "Nonlinear Programming," 2nd edition, Athena Scientific, 1999.

Kung Yao received his B.S. (Highest Honors) and Ph.D. degrees in Electrical Engineering from Princeton University. Presently, he is a Distinguished Professor in the Electrical Engineering Dept. of UCLA. He was an Assistant Dean at the UCLA School of Engineering and Applied Science from 1985-1988. From May-June 2007, he was a Royal Society Kan Tong Po Visiting Professor at HK Polytechnic Univ. His research interests include acoustic/seismic array processing, sensor system, wireless communication theory and system, and systolic array algorithm and systems. He has published over 250 journal and conference papers. Dr. Yao received the IEEE Signal Processing Society's 1993 Senior Award in VLSI Signal Processing and the 2008 IEEE Communications Society/Information Theory Society Joint Paper Award. He has served as Associate Editors of IEEE Trans. on Information Theory, IEEE Trans. on Signal Processing, IEEE Trans. on Circuits and Systems, IEEE Communication Letters, etc. He is a Life Fellow of IEEE.

Flavio Lorenzelli was awarded a Laurea and a Doctorate degree from Politecnico di Torino, Italy, in 1986 and 1991, and an MS and PhD degree in electrical engineering from UCLA in 1990 and 1993, respectively. He spent two years at CSELT Telecommunications Lab in Torino, Italy between 1987 and 1989. Until 2007 he was with STMicroelectronics, Inc. His last position in that company was manager of the US Research and Innovation team, in the Advanced System Technology division of STMicroelectronics, as well as researcher-in-residence at UCLA. He was also editor of the STMicroelectronics Journal of Research. Flavio is currently with the Aerospace Corp. in the Communication and Networking Subdivision. Flavio has taught graduate and undergraduate courses at the University of Milan and at UCLA in the areas of computer architecture, detection and estimation, stochastic processes, digital communications, system and signals, etc. Flavio's areas of expertise are digital signal processing and telecommunications. He is a senior member of IEEE and was recipient of a Fulbright fellowship between 1989 and 1996.