Optimal Stochastic Parameter Design for Estimation Problems

Hamza Soganci, *Student Member, IEEE*, Sinan Gezici*, *Senior Member, IEEE*, and Orhan Arikan, *Member, IEEE*

Abstract

In this study, the aim is to perform optimal stochastic parameter design in order to minimize the cost of a given estimator. Optimal probability distributions of signals corresponding to different parameters are obtained in the presence and absence of an average power constraint. It is shown that the optimal parameter design results in either a deterministic signal or a randomization between two different signal levels. In addition, sufficient conditions are obtained to specify the cases in which improvements over the deterministic parameter design can or cannot be achieved via the stochastic parameter design. Numerical examples are presented in order to provide illustrations of theoretical results.

Index Terms– SSP-PARE: Parameter estimation, Bayes risk, stochastic parameter design, randomization.

I. INTRODUCTION

In parametric estimation problems, an unknown parameter is estimated based on observations, the probability distribution of which is known as a function of the unknown parameter [1], [2]. In the presence of prior information about the parameter, Bayesian estimators, such as the minimum mean-squared error (MMSE) estimator and the minimum mean-absolute error (MMAE) estimator, are commonly employed [1]. On the other hand, in the absence of prior information about the parameter, the minimum variance unbiased estimator (MVUE), if it exists, or the maximum likelihood estimator (MLE) can be used [2]. In these conventional formulations of the parameter estimation problem, the aim is to obtain an optimal estimator that minimizes a certain cost function, such as the mean-squared error. In this study, we consider a different formulation in which the aim is to minimize the cost of a given estimator by performing stochastic parameter design under certain constraints. Motivations for this seemingly counterintuitive formulation will be provided in the next section.

Recently, various studies have employed signal randomization in order to improve the performance of detection and estimation systems (e.g., [3]-[7]). For example, an additive noise component that is randomized between two

The authors are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey, Tel: +90 (312) 290-3139, Fax: +90 (312) 266-4192, e-mails: {hsoganci,gezici,oarikan}@ee.bilkent.edu.tr. This research was supported in part by the National Young Researchers Career Development Programme (project no. 110E245) of the Scientific and Technological Research Council of Turkey (TUBITAK).

signal values can increase the detection probability of certain detectors under a false alarm constraint [3], [4]. Also, for power constrained communications systems, transmitting stochastic signals that are randomized among at most three different signal values can provide reductions in the average probability of error compared to the conventional case in which deterministic signal values are transmitted for each symbol [5]. In [6], it is shown that performance of some suboptimal estimators can be enhanced via additive "noise" that is injected into the observations before the estimation process. It is observed that this noise component can be a constant signal or a randomization between two signal values.

Motivated by the investigation of signal randomization in recent works [3]-[7], we consider the concept of stochastic parameter design for estimation problems in this study. Specifically, we try to answer the following question: If a fixed estimator is used at the receiver, what should be the optimal distribution of the signal sent from the transmitter for each possible parameter value? Referring to Fig. 1, the aim is to design the optimal *stochastic* signal $\mathbf{s}_{\boldsymbol{\theta}}$ for each $\boldsymbol{\theta}$ in order to minimize the cost (specifically, the Bayes risk) of a given estimator, which performs estimation based on the noise corrupted version of $\mathbf{s}_{\boldsymbol{\theta}}$, that is, $\mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n}$. Since there can exist power limits for transmitted signals in practice, this design problem needs to be solved under certain constraints.

As a specific example, consider a scenario in which the receiver employs the sample mean estimator to estimate a parameter θ based on a number of independent and identically distributed (i.i.d.) observations. The aim is to find the optimal random variable for each parameter value at the transmitter in order to minimize the Bayes risk of the sample mean estimator at the receiver. For instance, we would like to determine if sending i.i.d. Gaussian or Laplacian random variables with mean θ and variance 1 results in a lower Bayes risk. Or, more generally, among all continuous and discrete random variables, we would like to determine the one that minimizes the Bayes risk of the sample mean estimator.

In this study, after providing some motivations (Section II), we formulate this optimal stochastic parameter design problem, and prove that the optimal $\mathbf{s}_{\boldsymbol{\theta}}$ can be represented by either a deterministic signal value or a randomization between two different signal values (Section III). In addition, a convex relaxation of the optimal parameter design problem (resulting in linearly constrained linear programming) is presented (Section III), and sufficient conditions under which stochastic parameter design can or cannot provide improvements over the deterministic parameter design are obtained (Section IV). Also, numerical examples are presented to investigate the theoretical results (Section V).

II. Motivation

In conventional estimation problems, the aim is to design an optimal estimator for a given distribution of the observations. However, motivations can also be provided for the stochastic parameter design problem investigated in this study. For example, consider the design of a generic device (Device A in Fig. 1) which needs to output a certain parameter. This output is to be measured by a measurement device (the dashed box in Fig. 1) which employs a certain estimation algorithm for determining the parameter (e.g., averages various measurements). Then,

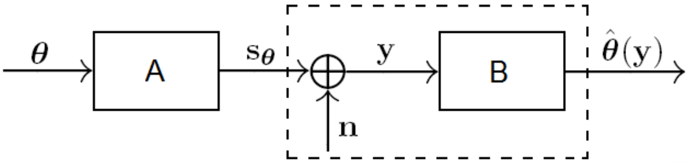

Fig. 1. System model. Device A transmits a stochastic signal $\mathbf{s}_{\boldsymbol{\theta}}$ for each value of parameter $\boldsymbol{\theta}$, and Device B estimates $\boldsymbol{\theta}$ based on the noise corrupted version of $\mathbf{s}_{\boldsymbol{\theta}}$. One interpretation is to consider the dashed box as a measurement device, in which case $\mathbf{n}$ denotes the measurement noise.

the aim is to design a stochastic signal $\mathbf{s}_{\boldsymbol{\theta}}$ for each $\boldsymbol{\theta}$ so that the accuracy (i.e., estimation performance) of the given measurement device is optimized. In other words, considering a certain type of a measurement device, the estimation performance of the overall system is to be optimized by designing stochastic signals for different parameters. Such a system model, in which estimation is performed based on measurements obtained by a number of measurement devices, is considered also in [8]. However, a different problem is considered in that study, and the optimal linear estimator is obtained in the presence of cost-constrained measurements. It should also be mentioned that most measurement devices are designed under a certain measurement noise assumption, such as Gaussian. They are typically non-adaptive devices, hence, in the presence of noise that deviates from the assumed noise distribution, their performance may degrade significantly. To improve the performance, the measurement device can be replaced with a more capable one; however, such a replacement may be very costly in some cases. To avoid the replacement cost and associated complications, the proposed stochastic parameter design approach can be used, which designs optimal signals for each parameter so that the performance of the suboptimal measurement device can be improved.

As another motivation of the setup in Fig. 1, a wireless sensor network [9], in which a parameter value (such as temperature or pressure) is sent from one device to another, can be considered. When the transmitter (Device A) knows the probability distribution of the channel noise, $\mathbf{n}$ (which can be obtained via feedback), it can perform stochastic parameter design in order to optimize the performance of the estimator at the receiver (Device B). If the probability distribution of $\mathbf{n}$ is unknown, then the results can be considered to provide a theoretical upper bound on the estimation performance. It is important to note that the additive noise is used to model all the operations/effects between Device A and Device B in Fig. 1. For example, signal values can be quantized, and encoded symbols can be sent via a specific digital communications method in some cases. Then, the additive noise model in Fig. 1 can be considered to provide an abstraction for all the blocks between Device A and Device B, such as quantizer, encoder/decoder, modulator/demodulator, and additive noise channel, as discussed in [11]. It should also be noted that noise $\mathbf{n}$ in Fig. 1 is modeled to have generic PDFs, not necessarily Gaussian, in the theoretical investigations in this study.

III. Stochastic Parameter Design

Consider a parameter estimation scenario as in Fig. 1, where the aim is to send the information about parameter $\boldsymbol{\theta}$ from Device A to Device B over an additive noise channel. For that purpose, Device A can transmit a (random) function of $\boldsymbol{\theta}$, say $\mathbf{s}_{\boldsymbol{\theta}}$, to Device B. Then, the received signal (observation) at Device B is expressed as

$$\mathbf{y} = \mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ denotes the channel noise, which has a probability density function (PDF) represented by $p_{\mathbf{n}}(\cdot)$. It is assumed that Device B employs a fixed estimator specified by $\hat{\boldsymbol{\theta}}(\mathbf{y})$ in order to estimate $\boldsymbol{\theta}$. In addition, the prior distribution of $\boldsymbol{\theta}$ is denoted by $w(\boldsymbol{\theta})$, and the parameter space in which $\boldsymbol{\theta}$ resides is represented by Λ.

In this study, the problem is to find the optimal probability distribution of $\mathbf{s}_{\boldsymbol{\theta}}$ for each $\boldsymbol{\theta} \in \Lambda$ in order to minimize the Bayes risk of a given estimator. It should be noted that, in conventional estimation problems, the aim is to design the optimal estimator for a given probability distribution of the observation [2]. However, we consider a different problem in which the aim is to optimize the information carrying parameters in order to optimize the performance of a given estimator. Another important point is that unlike conventional estimation problems, $\mathbf{s}_{\boldsymbol{\theta}}$ in (1) is modeled as a random variable for each value of $\boldsymbol{\theta}$; that is, a *stochastic parameter design* approach is considered in this study.

A. Unconstrained Optimization

First, no constraints are considered in the selection of $\mathbf{s}_{\boldsymbol{\theta}}$. Then, the optimal stochastic parameter design problem can be formulated as

$$\{p^{\text{opt}}_{\mathbf{s}_{\boldsymbol{\theta}}},\, \boldsymbol{\theta} \in \Lambda\} = \underset{\{p_{\mathbf{s}_{\boldsymbol{\theta}}},\, \boldsymbol{\theta}\in\Lambda\}}{\arg\min}\; r(\hat{\boldsymbol{\theta}}) \tag{2}$$

where $\{p_{\mathbf{s}_{\boldsymbol{\theta}}},\, \boldsymbol{\theta} \in \Lambda\}$ denotes the set of PDFs for $\mathbf{s}_{\boldsymbol{\theta}}$ for all possible values of parameter $\boldsymbol{\theta}$, and $r(\hat{\boldsymbol{\theta}})$ is the Bayes risk of the estimator. In order to obtain a more explicit formulation of the problem, the Bayes risk can be expressed as

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) \int \mathrm{C}[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]\, p_{\boldsymbol{\theta}}(\mathbf{y})\, d\mathbf{y}\, d\boldsymbol{\theta} \tag{3}$$

where $p_{\boldsymbol{\theta}}(\mathbf{y})$ denotes the PDF of $\mathbf{y}$, which is indexed by $\boldsymbol{\theta}$, and $\mathrm{C}[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]$ represents a cost function [2]. For example, $\mathrm{C}[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}] = (\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^2$ corresponds to the squared-error cost function, for which $r(\hat{\boldsymbol{\theta}})$ becomes the mean-squared error (MSE). In this study, a generic cost function $\mathrm{C}[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]$ is considered in all the derivations.

If $\mathbf{s}_{\boldsymbol{\theta}}$ were modeled as a deterministic quantity for each value of $\boldsymbol{\theta}$, $p_{\boldsymbol{\theta}}(\mathbf{y})$ in (3) could be expressed in terms of the PDF of $\mathbf{n}$ as $p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_{\boldsymbol{\theta}})$ (see (1)). However, we consider a stochastic parameter design framework and model $\mathbf{s}_{\boldsymbol{\theta}}$ as a stochastic variable for each $\boldsymbol{\theta}$. Then, assuming that the noise and $\mathbf{s}_{\boldsymbol{\theta}}$ are independent, $p_{\boldsymbol{\theta}}(\mathbf{y})$ is calculated as $\int p_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) p_{\mathbf{n}}(\mathbf{y} - \mathbf{x})\, d\mathbf{x}$. Therefore, (3) becomes

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta}) \int p_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) \int \mathrm{C}[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]\, p_{\mathbf{n}}(\mathbf{y} - \mathbf{x})\, d\mathbf{y}\, d\mathbf{x}\, d\boldsymbol{\theta}\ . \tag{4}$$

Defining an auxiliary function $g_{\boldsymbol{\theta}}(\mathbf{x})$ as

$$g_{\boldsymbol{\theta}}(\mathbf{x}) \triangleq \int \mathrm{C}[\hat{\boldsymbol{\theta}}(\mathbf{y}), \boldsymbol{\theta}]\, p_{\mathbf{n}}(\mathbf{y} - \mathbf{x})\, d\mathbf{y}\ , \tag{5}$$

the relation in (4) can be stated as

$$r(\hat{\boldsymbol{\theta}}) = \int_{\Lambda} w(\boldsymbol{\theta})\, \mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}\, d\boldsymbol{\theta} \tag{6}$$

where each expectation operation is over the PDF of $\mathbf{s}_{\boldsymbol{\theta}}$ for a given value of $\boldsymbol{\theta}$. From (6), it is observed that $r(\hat{\boldsymbol{\theta}})$ can be minimized if, for each $\boldsymbol{\theta}$, the PDF of $\mathbf{s}_{\boldsymbol{\theta}}$ assigns all the probability to the minimizer of $g_{\boldsymbol{\theta}}$.[1] Namely, the solution of the optimization problem in (2) can be expressed as

$$p^{\mathrm{opt}}_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}^{\mathrm{unc}}_{\boldsymbol{\theta}})\ , \quad \mathbf{s}^{\mathrm{unc}}_{\boldsymbol{\theta}} = \arg\min_{\mathbf{x}} g_{\boldsymbol{\theta}}(\mathbf{x}) \tag{7}$$

for all $\boldsymbol{\theta} \in \Lambda$. Therefore, it is concluded that the optimal stochastic parameter design results in optimal PDFs that have single point masses. Hence, deterministic parameter design is optimal and no stochastic modeling is needed when there are no constraints in the design problem. However, in practice, the values of $\mathbf{s}_{\boldsymbol{\theta}}$ cannot be chosen without any constraints (such as an average power constraint), and it will be shown in the next section that the stochastic parameter design can result in performance improvements in the presence of constraints on the moments of $\mathbf{s}_{\boldsymbol{\theta}}$. Another important observation from (7) is that the solution does not require the knowledge of the prior distribution $w(\boldsymbol{\theta})$, since the optimal solution is obtained for each $\boldsymbol{\theta}$ separately.

B. Constrained Optimization

In practical scenarios, the parameter design cannot be performed without any limitations. For example, in the absence of a power constraint, it would be possible to reduce the Bayes risk arbitrarily by transmitting signals with very high powers compared to the noise power.

In this section, a common design constraint in the form of an average power constraint is considered in the stochastic parameter design problem. Although a specific constraint type is used in the following, it will be discussed that other types of constraints can also be incorporated into the theoretical analysis.

Consider an average power constraint in the form of

$$\mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}} \tag{8}$$

for $\boldsymbol{\theta} \in \Lambda$, where $\|\mathbf{s}_{\boldsymbol{\theta}}\|$ is the Euclidean norm of vector $\mathbf{s}_{\boldsymbol{\theta}}$, and $A_{\boldsymbol{\theta}}$ denotes the average power constraint for $\boldsymbol{\theta}$. It is noted from (8) that a generic model is considered for the constraint $A_{\boldsymbol{\theta}}$, which can depend on the value of $\boldsymbol{\theta}$ in general. For the special case in which the average power constraint is the same for all parameters, $A_{\boldsymbol{\theta}} = A$ for $\boldsymbol{\theta} \in \Lambda$ can be employed.

From (6) and (8), the optimal stochastic parameter design problem can be stated as

$$\min_{\{p_{\mathbf{s}_{\boldsymbol{\theta}}},\, \boldsymbol{\theta} \in \Lambda\}} \int_{\Lambda} w(\boldsymbol{\theta})\, \mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}\, d\boldsymbol{\theta}$$
$$\text{subject to } \mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}}\ ,\ \forall \boldsymbol{\theta} \in \Lambda \tag{9}$$

[1]If there are multiple minimizers, any (combination) of them can be chosen for the optimal solution.

where $g_{\boldsymbol{\theta}}(\cdot)$ is as defined in (5). The investigation of the constrained optimization problem in (9) reveals that the problem can be solved separately for each $\boldsymbol{\theta}$ as follows:

$$\min_{p_{\mathbf{s}_{\boldsymbol{\theta}}}} \mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\} \quad \text{subject to} \quad \mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}} \tag{10}$$

for $\boldsymbol{\theta} \in \Lambda$. In other words, the optimal PDF of $\mathbf{s}_{\boldsymbol{\theta}}$ can be obtained separately for each $\boldsymbol{\theta}$. Therefore, the result does not depend on the prior distribution $w(\boldsymbol{\theta})$, and the solution can be obtained in the absence of prior information.

Optimization problems in the form of (10) have been investigated in different studies in the literature [3], [5], [10]. Specifically, [3] and [10] aim to obtain the optimal additive "noise" PDF that maximizes the detection probability under a constraint on the false-alarm probability, and [5] investigates optimal signal PDFs in a power constrained binary communications systems. Based on similar arguments to those in [3], [5], [10], the following result can be obtained.

Proposition 1: *Suppose $g_{\boldsymbol{\theta}}$ is a continuous function and each component of $\mathbf{s}_{\boldsymbol{\theta}}$ resides in a finite closed interval. Then, an optimal solution to* (10) *can be expressed in the following form:*

$$p^{\mathrm{opt}}_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) = \lambda_{\boldsymbol{\theta}}\, \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},1}) + (1 - \lambda_{\boldsymbol{\theta}})\, \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},2}) \tag{11}$$

for $\lambda_{\boldsymbol{\theta}} \in [0, 1]$.

Proof: Consider the set of all $(g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}), \|\mathbf{s}_{\boldsymbol{\theta}}\|^2)$ pairs and the set of all $(\mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}, \mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\})$ pairs, and denote them as U and W, respectively. Namely, $U = \{(u_1, u_2) : u_1 = g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\,,\, u_2 = \|\mathbf{s}_{\boldsymbol{\theta}}\|^2\,, \forall\, \mathbf{s}_{\boldsymbol{\theta}}\}$ and $W = \{(w_1, w_2) : w_1 = \mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}\,,\, w_2 = \mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\}\,, \forall\, p_{\mathbf{s}_{\boldsymbol{\theta}}}\}$. As discussed in [3] and [5], the convex hull of U can be shown to be equal to W. Then, based on Carathéodory's theorem [12], it is concluded that any point in W can be obtained as a convex combination of at most three points in U. Also, since an optimal PDF should achieve the minimum value, it must correspond to the boundary of W, which results in a convex combination of at most two points in U. (The assumptions in the proposition imply that W is a closed set; therefore, it contains its boundary [5].) Hence, an optimal solution can be expressed as in (11) [13]. □

Proposition 1 states that the optimal solution can be achieved by randomization between at most two different values for each $\boldsymbol{\theta}$. Based on this result, the optimal stochastic parameter design problem in (10) is expressed as

$$\begin{aligned} &\min_{\lambda_{\boldsymbol{\theta}}, \mathbf{s}_{\boldsymbol{\theta},1}, \mathbf{s}_{\boldsymbol{\theta},2}} \; \lambda_{\boldsymbol{\theta}}\, g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},1}) + (1 - \lambda_{\boldsymbol{\theta}})\, g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},2}) \\ &\text{subject to } \lambda_{\boldsymbol{\theta}} \|\mathbf{s}_{\boldsymbol{\theta},1}\|^2 + (1 - \lambda_{\boldsymbol{\theta}}) \|\mathbf{s}_{\boldsymbol{\theta},2}\|^2 \leq A_{\boldsymbol{\theta}}\ ,\ \lambda_{\boldsymbol{\theta}} \in [0, 1] \end{aligned} \tag{12}$$

for $\boldsymbol{\theta} \in \Lambda$. Compared to (10), the formulation in (12) provides a significant simplification as it requires optimization over a finite number of variables instead of over all possible PDFs. Since generic cost functions and noise distributions are considered in the theoretical analysis, $g_{\boldsymbol{\theta}}$ in (5) is quite generic and the optimization problem in (12) can be nonconvex in general. Therefore, global optimization techniques such as particle swarm optimization (PSO) and differential evolution (DE) can be employed to obtain the solution [14], [15].

Remark 1: Although the average power constraint in (8) is considered in obtaining the preceding results, the other types of constraints in the form of $\mathrm{E}\{h_i(\mathbf{s}_{\boldsymbol{\theta}})\} \leq A_{\boldsymbol{\theta},i}$ for $i = 1, \ldots, N_c$ can also be incorporated. Specifically, assuming continuous h_i, the form of the optimal PDF in Proposition 1 becomes $p^{\rm opt}_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) = \sum_{i=1}^{N_c} \lambda_{\boldsymbol{\theta},i}\, \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},i})$, with $\lambda_{\boldsymbol{\theta},i} \geq 0$ for $i = 1, \ldots, N_c$ and $\sum_{i=1}^{N_c} \lambda_{\boldsymbol{\theta},i} = 1$, which can be proven by updating the definitions of sets U and W accordingly in the proof of Proposition 1.

As an alternative approach, a convex relaxation technique can be employed to obtain an approximate solution of (10) in polynomial time [5], [16]. To that aim, it is assumed that $p_{\mathbf{s}_{\boldsymbol{\theta}}}$ can be expressed as $p_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) = \sum_{l=1}^{N_m} \beta_l\, \delta(\mathbf{x}-\tilde{\mathbf{s}}_{\boldsymbol{\theta},l})$, where $\beta_l \geq 0$ for $l = 1, \ldots, N_m$, $\sum_{l=1}^{N_m} \beta_l = 1$, and $\tilde{\mathbf{s}}_{\boldsymbol{\theta},1}, \ldots, \tilde{\mathbf{s}}_{\boldsymbol{\theta},N_m}$ are known possible values for $\mathbf{s}_{\boldsymbol{\theta}}$. Then, by defining $\boldsymbol{\beta} = [\beta_1 \cdots \beta_{N_m}]^T$, $\tilde{\boldsymbol{g}}_{\boldsymbol{\theta}} = [g_{\boldsymbol{\theta}}(\tilde{\mathbf{s}}_{\boldsymbol{\theta},1}) \cdots g_{\boldsymbol{\theta}}(\tilde{\mathbf{s}}_{\boldsymbol{\theta},N_m})]^T$ and $\mathbf{c} = [\|\tilde{\mathbf{s}}_{\boldsymbol{\theta},1}\|^2 \cdots \|\tilde{\mathbf{s}}_{\boldsymbol{\theta},N_m}\|^2]^T$, the convex version of (10) can be obtained as

$$\min_{\boldsymbol{\beta}} \boldsymbol{\beta}^T \tilde{\boldsymbol{g}}_{\boldsymbol{\theta}} \quad \text{subject to } \boldsymbol{\beta}^T \mathbf{c} \leq A_{\boldsymbol{\theta}} \ , \ \boldsymbol{\beta}^T \mathbf{1} = 1 \ , \ \boldsymbol{\beta} \succeq \mathbf{0} \tag{13}$$

where $\mathbf{1}$ and $\mathbf{0}$ denote the vectors of ones and zeros, respectively, and $\boldsymbol{\beta} \succeq \mathbf{0}$ means that each element of $\boldsymbol{\beta}$ is greater than or equal to zero. It is noted that (13) presents a linearly constrained linear optimization problem; hence, it can be solved efficiently in polynomial time [16]. In general, the solution of (13) provides an approximate solution, and the approximation accuracy can be improved by using a large value of N_m.

IV. Optimality Conditions

The deterministic parameter design can be considered as a special case of the stochastic parameter design when $\mathbf{s}_{\boldsymbol{\theta}}$ in (10) is modeled as a deterministic quantity for each $\boldsymbol{\theta}$. Namely, the deterministic parameter design problem can be formulated as

$$\min_{\mathbf{s}_{\boldsymbol{\theta}}} g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}) \quad \text{subject to } \|\mathbf{s}_{\boldsymbol{\theta}}\|^2 \leq A_{\boldsymbol{\theta}} \tag{14}$$

for $\boldsymbol{\theta} \in \Lambda$ (c.f. (10)). Let $\mathbf{s}^{\rm opt}_{\boldsymbol{\theta}}$ denote the minimizer of the optimization problem in (14). Then, the minimum Bayes risk achieved by the optimal deterministic parameter design is given by $r_{\rm det}(\hat{\boldsymbol{\theta}}) = \int_\Lambda w(\boldsymbol{\theta}) g_{\boldsymbol{\theta}}(\mathbf{s}^{\rm opt}_{\boldsymbol{\theta}}) d\boldsymbol{\theta}$ (see (6)). Similarly, let $r_{\rm sto}(\hat{\boldsymbol{\theta}}) = \int_\Lambda w(\boldsymbol{\theta}) \int g_{\boldsymbol{\theta}}(\mathbf{x}) p^{\rm opt}_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) d\mathbf{x}\, d\boldsymbol{\theta}$ represent the minimum Bayes risk achieved by the optimal stochastic parameter design, where $p^{\rm opt}_{\mathbf{s}_{\boldsymbol{\theta}}}$ denotes the optimal solution for $\boldsymbol{\theta}$. In order for the stochastic parameter design to improve over the deterministic parameter design, $r_{\rm sto}(\hat{\boldsymbol{\theta}})$ should be strictly smaller than $r_{\rm det}(\hat{\boldsymbol{\theta}})$. Otherwise, it is concluded that the deterministic parameter design cannot be improved via the stochastic approach; that is, $r_{\rm sto}(\hat{\boldsymbol{\theta}}) = r_{\rm det}(\hat{\boldsymbol{\theta}})$. In the following proposition, sufficient conditions presented for the latter.

Proposition 2: *The deterministic parameter design cannot be improved via the stochastic approach if at least one of the following is satisfied for each* $\boldsymbol{\theta}$ *:*

- $g_{\boldsymbol{\theta}}$ *is a convex function.*
- *The solution of the unconstrained problem (see* (7)*) satisfies the constraint; i.e.,* $\|\mathbf{s}^{\rm unc}_{\boldsymbol{\theta}}\|^2 \leq A_{\boldsymbol{\theta}}$ *.*

Proof: If the second condition is satisfied, that is, if $\|\mathbf{s}^{\rm unc}_{\boldsymbol{\theta}}\|^2 \leq A_{\boldsymbol{\theta}}$, then the solution of (14) coincides with that of the unconstrained problem in Section III-A; namely, $\mathbf{s}^{\rm opt}_{\boldsymbol{\theta}} = \mathbf{s}^{\rm unc}_{\boldsymbol{\theta}}$. Therefore, the solution of the optimal stochastic parameter design problem in (10) becomes $p^{\rm opt}_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{s}^{\rm opt}_{\boldsymbol{\theta}})$. Hence, the deterministic design is optimal in such a scenario, and the stochastic approach is not needed.

In order to investigate the first condition, it is observed that, for any $\mathbf{s}_{\boldsymbol{\theta}}$, $\mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \geq \|\mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}\|^2$ is satisfied due to Jensen's inequality since norm is a convex function. Therefore, due to the constraint $\mathrm{E}\{\|\mathbf{s}_{\boldsymbol{\theta}}\|^2\} \leq A_{\boldsymbol{\theta}}$ in (10), $\|\mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}\|^2 \leq A_{\boldsymbol{\theta}}$ must hold for any feasible PDF of $\mathbf{s}_{\boldsymbol{\theta}}$. Let $\mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}$ be defined as $\check{\mathbf{s}}_{\boldsymbol{\theta}} \triangleq \mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}$. As the minimizer of (14), $\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}$, achieves the minimum $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})$ among all $\mathbf{s}_{\boldsymbol{\theta}}$ that satisfy $\|\mathbf{s}_{\boldsymbol{\theta}}\|^2 \leq A_{\boldsymbol{\theta}}$, $\|\mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}\|^2 = \|\check{\mathbf{s}}_{\boldsymbol{\theta}}\|^2 \leq A_{\boldsymbol{\theta}}$ implies that $g_{\boldsymbol{\theta}}(\mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}) = g_{\boldsymbol{\theta}}(\check{\mathbf{s}}_{\boldsymbol{\theta}}) \geq g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}})$ is satisfied. When $g_{\boldsymbol{\theta}}$ is a convex function as specified in the proposition, $\mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\} \geq g_{\boldsymbol{\theta}}(\mathrm{E}\{\mathbf{s}_{\boldsymbol{\theta}}\}) \geq g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}})$ is obtained from Jensen's inequality and from the previous relation. Therefore, for convex $g_{\boldsymbol{\theta}}$, $\mathrm{E}\{g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}})\}$ can never be smaller than the minimum value of (14), $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}})$, for any PDF of $\mathbf{s}_{\boldsymbol{\theta}}$ that satisfies the average power constraint. Hence, the minimum value of (10) cannot be smaller than $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}})$, meaning that it is always equal to $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}})$ (since (10) covers (14) as a special case).

All in all, when at least one of the conditions in the proposition are satisfied for all $\boldsymbol{\theta}$, the deterministic and the stochastic approaches achieve the same minimum values for all parameters; that is, $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}) = \int g_{\boldsymbol{\theta}}(\mathbf{x}) p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\mathrm{opt}}(\mathbf{x}) d\mathbf{x}$, $\forall \boldsymbol{\theta}$. Therefore, $r_{\mathrm{det}}(\hat{\boldsymbol{\theta}}) = \int_\Lambda w(\boldsymbol{\theta}) g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}})\, d\boldsymbol{\theta}$ and $r_{\mathrm{sto}}(\hat{\boldsymbol{\theta}}) = \int_\Lambda w(\boldsymbol{\theta}) \int g_{\boldsymbol{\theta}}(\mathbf{x}) p_{\mathbf{s}_{\boldsymbol{\theta}}}^{\mathrm{opt}}(\mathbf{x}) d\mathbf{x}\, d\boldsymbol{\theta}$ become equal. □

In order to present an example application of Proposition 2, consider a scenario in which a scalar parameter θ is to be estimated in the presence of zero-mean additive noise n. The average power constraint is in the generic form of $\mathrm{E}\{|\mathrm{s}_\theta|^2\} \leq A_\theta$ for all θ, and the estimator is specified by $\hat{\theta}(\mathrm{y}) = \mathrm{y}$. In addition, the cost function is modeled as $\mathrm{C}[\hat{\theta}(\mathrm{y}), \theta] = (\hat{\theta}(\mathrm{y}) - \theta)^2$. In this scenario, $g_{\boldsymbol{\theta}}$ in (5) can be calculated as

$$g_\theta(\mathrm{x}) = \int_{-\infty}^{\infty} (\mathrm{y} - \theta)^2 p_{\mathrm{n}}(\mathrm{y} - \mathrm{x})\, dy = \int_{-\infty}^{\infty} (\mathrm{y} + \mathrm{x} - \theta)^2 p_{\mathrm{n}}(\mathrm{y})\, dy = (\mathrm{x} - \theta)^2 + \mathrm{Var}\{\mathrm{n}\} \quad (15)$$

where $\mathrm{Var}\{\mathrm{n}\}$ denotes the variance of the noise. From (15), it is noted that $g_{\boldsymbol{\theta}}$ is a convex function for any value of θ. Therefore, the first condition in Proposition 2 is satisfied for all θ, meaning that the performance of the deterministic parameter design cannot be improved via the stochastic approach.[2] Hence, the optimal solution can be obtained from (14), which yields

$$\mathrm{s}_\theta^{\mathrm{opt}} = \arg \min_{|\mathrm{s}_\theta|^2 \leq A_\theta} (\mathrm{s}_\theta - \theta)^2 .$$

For example, if $A_\theta = \theta^2$, then $\mathrm{s}_\theta^{\mathrm{opt}} = \theta$ for all θ.

In the following proposition, sufficient conditions are presented to specify cases in which the stochastic parameter design provides improvements over the deterministic one.

Proposition 3: *The stochastic parameter design achieves a smaller Bayes risk than the deterministic one if there exists $\boldsymbol{\theta} \in \Lambda$ for which $g_{\boldsymbol{\theta}}(\mathbf{x})$ is second-order continuously differentiable around $\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}$ and a real vector $\mathbf{z}$ can be found such that*

$$\left(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}\right)\left(\mathbf{z}^T \nabla g_{\boldsymbol{\theta}}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}}\right) < 0 \quad \textit{and} \quad (16)$$

$$\|\mathbf{z}\|^2 < \left(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}\right)\left(\mathbf{z}^T \mathbf{H}_{\boldsymbol{\theta}} \mathbf{z}\right) / \left(\mathbf{z}^T \nabla g_{\boldsymbol{\theta}}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}}\right) \quad (17)$$

where $\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}$ is the solution of (14), *$\nabla g_{\boldsymbol{\theta}}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}}$ denotes the gradient of $g_{\boldsymbol{\theta}}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}$, and $\mathbf{H}_{\boldsymbol{\theta}}$ is the Hessian of $g_{\boldsymbol{\theta}}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}}^{\mathrm{opt}}$.*

[2]It can be shown that g_θ is convex for all θ also for the absolute error cost function; i.e., $\mathrm{C}[\hat{\theta}(\mathrm{y}), \theta] = |\hat{\theta}(\mathrm{y}) - \theta|$.

Proof: In order to prove that a reduced Bayes risk can be achieved via the stochastic parameter design, consider a specific value of $\boldsymbol{\theta}$ for which the conditions in the proposition are satisfied. Also consider two values $\mathbf{s}_{\boldsymbol{\theta},1}$ and $\mathbf{s}_{\boldsymbol{\theta},2}$ around $\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}$, which can be expressed as $\mathbf{s}_{\boldsymbol{\theta},i} = \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}} + \boldsymbol{\epsilon}_i$ for $i = 1, 2$. Then, $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},i})$ can be approximated as $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},i}) \approx g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}) + \boldsymbol{\epsilon}_i^T \tilde{\mathbf{g}}_{\boldsymbol{\theta}} + 0.5\, \boldsymbol{\epsilon}_i^T \mathbf{H}_{\boldsymbol{\theta}} \boldsymbol{\epsilon}_i$ for $i = 1, 2$, where $\tilde{\mathbf{g}}_{\boldsymbol{\theta}} = \nabla g_{\boldsymbol{\theta}}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}}$ is the gradient and $\mathbf{H}_{\boldsymbol{\theta}}$ is the Hessian of $g_{\boldsymbol{\theta}}(\mathbf{x})$ at $\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}$ [17]. Similarly, $\|\mathbf{s}_{\boldsymbol{\theta},i}\|^2$ can be expressed as $\|\mathbf{s}_{\boldsymbol{\theta},i}\|^2 \approx \|\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\|^2 + 2\, \boldsymbol{\epsilon}_i^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}} + \|\boldsymbol{\epsilon}_i\|^2$ for $i = 1, 2$. In order to prove that employing $p_{\mathbf{s}_{\boldsymbol{\theta}}}(\mathbf{x}) = \lambda\, \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},1}) + (1-\lambda)\, \delta(\mathbf{x} - \mathbf{s}_{\boldsymbol{\theta},1})$ results in a lower risk than $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}})$, which is the one achieved by the deterministic parameter design (see (14)), it is sufficient to show that

$$
\begin{aligned}
&\lambda\, g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},1}) + (1-\lambda)\, g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},2}) < g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}) \\
&\lambda\, \|\mathbf{s}_{\boldsymbol{\theta},1}\|^2 + (1-\lambda)\, \|\mathbf{s}_{\boldsymbol{\theta},2}\|^2 < \|\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\|^2 \leq A_{\boldsymbol{\theta}}
\end{aligned} \tag{18}
$$

are satisfied for certain choice of parameters (see (10)). After inserting the expressions for $g_{\boldsymbol{\theta}}(\mathbf{s}_{\boldsymbol{\theta},i})$ and $\|\mathbf{s}_{\boldsymbol{\theta},i}\|^2$ around $\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}$ into (18), it can be obtained that

$$
\begin{aligned}
&\lambda\, \boldsymbol{\epsilon}_1^T \mathbf{H}_{\boldsymbol{\theta}} \boldsymbol{\epsilon}_1 + (1-\lambda)\, \boldsymbol{\epsilon}_2^T \mathbf{H}_{\boldsymbol{\theta}} \boldsymbol{\epsilon}_2 + 2\big(\lambda\, \boldsymbol{\epsilon}_1 + (1-\lambda)\, \boldsymbol{\epsilon}_2\big)^T \tilde{\mathbf{g}}_{\boldsymbol{\theta}} < 0 \\
&\lambda\, \|\boldsymbol{\epsilon}_1\|^2 + (1-\lambda)\, \|\boldsymbol{\epsilon}_2\|^2 + 2\big(\lambda\, \boldsymbol{\epsilon}_1 + (1-\lambda)\, \boldsymbol{\epsilon}_2\big)^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}} < 0
\end{aligned} \tag{19}
$$

Let $\boldsymbol{\epsilon}_1 = \eta\, \mathbf{z}$ and $\boldsymbol{\epsilon}_2 = \nu\, \mathbf{z}$. Then, (19) can be manipulated to obtain

$$
\mathbf{z}^T \mathbf{H}_{\boldsymbol{\theta}} \mathbf{z} + k\big(\mathbf{z}^T \tilde{\mathbf{g}}_{\boldsymbol{\theta}}\big) < 0 \quad \text{and} \quad \|\mathbf{z}\|^2 + k\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big) < 0 \tag{20}
$$

with $k \triangleq 2(\lambda\, \eta + (1-\lambda)\nu)/(\lambda\, \nu^2 + (1-\lambda)\eta^2)$. If the first inequality in (20) is multiplied by $\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big)/\big(\mathbf{z}^T \tilde{\mathbf{g}}_{\boldsymbol{\theta}}\big)$, which is always negative due to the condition (16) in the proposition, (20) becomes

$$
(\mathbf{z}^T \mathbf{H}_{\boldsymbol{\theta}} \mathbf{z})\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big)/\big(\mathbf{z}^T \tilde{\mathbf{g}}_{\boldsymbol{\theta}}\big) + k\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big) > 0 \quad \text{and} \quad \|\mathbf{z}\|^2 + k\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big) < 0\ . \tag{21}
$$

Since k can take any real value by adjusting $\lambda \in [0, 1]$ and infinitesimally small η and ν values, it is guaranteed that both inequalities in (21) can be satisfied if $(\mathbf{z}^T \mathbf{H}_{\boldsymbol{\theta}} \mathbf{z})\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big)/\big(\mathbf{z}^T \tilde{\mathbf{g}}_{\boldsymbol{\theta}}\big)$ is larger than $\|\mathbf{z}\|^2$, which corresponds to (17). □

Remark 2: For the conditions in (16) and (17) to be satisfied, $g_{\boldsymbol{\theta}}(\mathbf{x})$ must be concave at $\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}$ (i.e., $\mathbf{H}_{\boldsymbol{\theta}}$ must be negative-definite) since $\|\mathbf{z}\|^2$ is always nonnegative and $\big(\mathbf{z}^T \mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}\big)/\big(\mathbf{z}^T \nabla g_{\boldsymbol{\theta}}(\mathbf{x})|_{\mathbf{x}=\mathbf{s}_{\boldsymbol{\theta}}^{\text{opt}}}\big)$ is negative due to (16).

Proposition 3 provides a simple approach, based on the first and second order derivatives of $g_{\boldsymbol{\theta}}$, to determine if the stochastic parameter design can provide improvements over the deterministic one. If the conditions are satisfied, the improvements are guaranteed and the optimization problem in (12) or (13) can be solved to obtain the optimal solution. However, since the conditions are sufficient but not necessary, there can also exist certain scenarios in which improvements are observed although the conditions are not satisfied. Examples for various scenarios are provided in the next section.

V. Numerical Results and Conclusions

In order to present examples of the theoretical results in the previous sections, consider an estimation problem in which a scalar parameter θ is estimated based on observation y that is modeled as $\mathrm{y} = \mathrm{s}_\theta + \mathrm{n}$, with n denoting the additive noise component. (Although a scalar problem is considered for convenience, vector parameter estimation problems can be treated in a similar fashion (per component) when the noise components are independent and the cost function is additive [2].) The noise n is modeled by a Gaussian mixture distribution, specified as $p_{\mathrm{n}}(\mathrm{n}) = \sum_{l=1}^{L} \gamma_l \exp\{-(\mathrm{n}-\mu_l)^2/(2\sigma_l^2)\}/(\sqrt{2\pi}\,\sigma_l)$, where the parameters are chosen in such a way to generate a zero-mean noise component. In addition, the estimator is given by $\hat{\theta}(\mathrm{y}) = \mathrm{y}$, and the cost function is selected as the uniform cost function, which is expressed as $\mathrm{C}[\hat{\theta}(y), \theta] = 1$ if $|\hat{\theta}(y) - \theta| > \Delta$ and $\mathrm{C}[\hat{\theta}(y), \theta] = 0$ otherwise. Based on this model, g_θ in (5) can be obtained as

$$g_\theta(\mathrm{x}) = \sum_{l=1}^{L} \gamma_l \left(Q\left(\frac{\mathrm{x} - \theta + \mu_l + \Delta}{\sigma_l}\right) + Q\left(\frac{-\mathrm{x} + \theta - \mu_l + \Delta}{\sigma_l}\right)\right) \tag{22}$$

where $Q(\mathrm{x}) = (1/\sqrt{2\pi}) \int_x^\infty \exp\{-t^2/2\} dt$ denotes the Q-function. Regarding the constraint in (8), $\mathrm{E}\{|\mathrm{s}_\theta|^2\} \leq \theta^2$ is considered for each θ.

For the numerical examples, parameter θ is modeled to lie between -10 and 10; that is, the parameter space is specified as $\Lambda = [-10, 10]$. Also, s_θ can take values in the interval $[-10, 10]$ under the average power constraint, $\mathrm{E}\{|\mathrm{s}_\theta|^2\} \leq \theta^2$. In addition, the parameters of the Gaussian mixture noise n are selected as $\gamma_1 = 0.33$, $\gamma_2 = 0.13$, $\gamma_3 = 0.08$, $\gamma_4 = 0.07$, $\gamma_5 = 0.11$, $\gamma_6 = 0.28$, $\mu_1 = -3.8$, $\mu_2 = -1.6$, $\mu_3 = -0.51$, $\mu_4 = 0.4657$, $\mu_5 = 2.42$, $\mu_6 = 4.3$, and $\sigma_l = 0.5$, $\forall l$. With this selection of the parameters, the noise becomes a zero-mean random variable so that $\hat{\theta}(\mathrm{y}) = \mathrm{y}$ can be regarded as a practical estimator.[3] Finally, $\Delta = 1$ is considered for the uniform cost function described in the previous paragraph.

In Fig. 2, the conditional risks (i.e., $\mathrm{E}\{g_\theta(\mathrm{s}_\theta)\}$ in (6)) are plotted versus θ for various parameter design approaches. For the optimal stochastic parameter design, both the exact solution obtained from (12) and the convex relaxation solutions obtained from (13) are plotted. In the convex relaxation approach, the set of possible values for s_θ are selected between -10 and 10 with an increment of D (in short, $-10 : \mathrm{D} : 10$), and the results for $\mathrm{D} = 0.25$ and $\mathrm{D} = 0.5$ are illustrated in the figure. The results for the optimal deterministic parameter design are calculated from (14). In addition, the results obtained from the unconstrained problem (see (7)) and those obtained by using $p_{\mathrm{s}_\theta}(\mathrm{x}) = \delta(\mathrm{x} - \theta)$ (labeled as "Conventional") are shown in the figure to provide performance benchmarks. It is observed that the optimal stochastic parameter design achieves the minimum conditional risks for all θ values in the presence of the average power constraint. It provides performance improvements over the deterministic parameter design for certain range of parameter values, e.g., for $\theta > 2.1$. In addition, both the stochastic and the deterministic design approaches achieve the same conditional risks as the unconstrained solution for some θ values, which is due to the fact that the unconstrained solutions satisfy the average power constraint for those values of θ.

[3]Although this is not an optimal estimator, it can be used in practice due to its simplicity compared to the optimal estimator, which would have high complexity due to the multimodal noise structure.

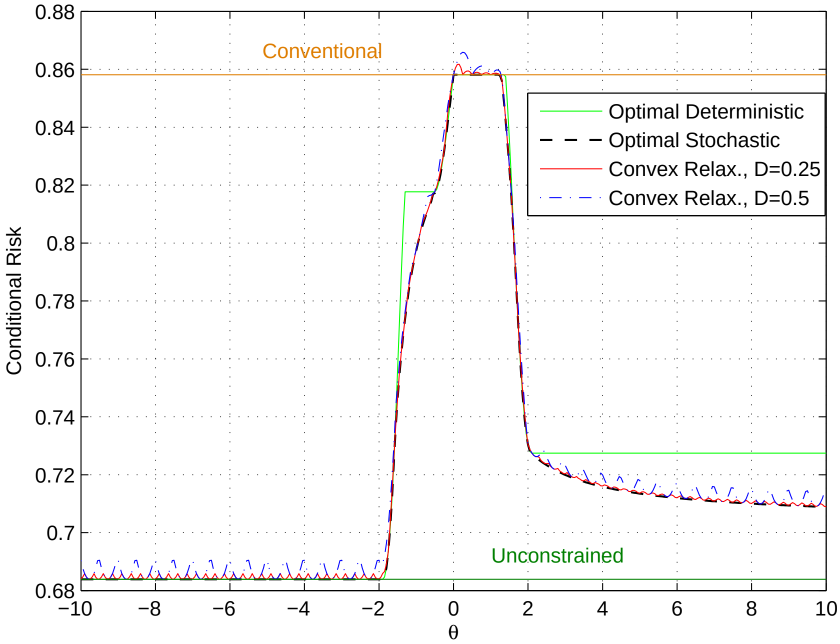

Fig. 2. Conditional risk versus θ for various parameter design approaches.

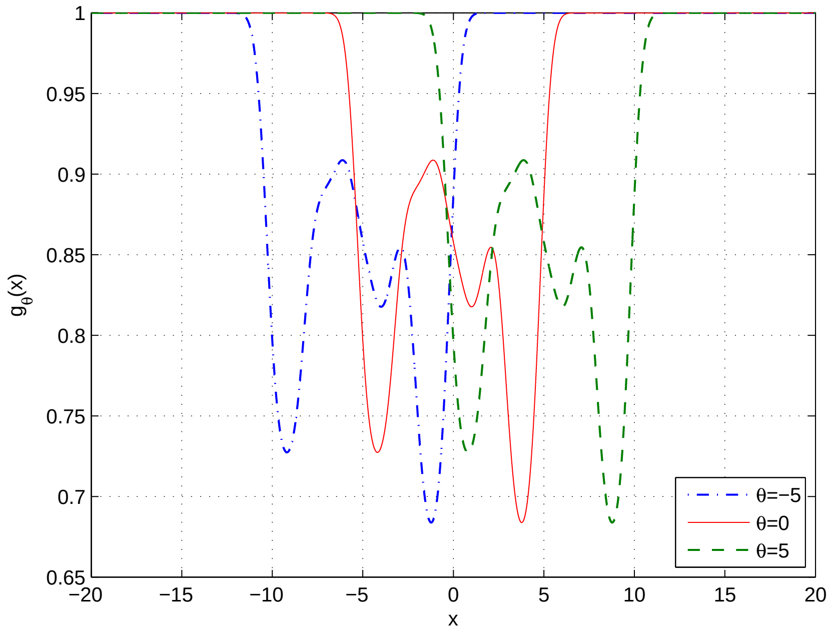

Fig. 3. $g_\theta(\mathrm{x})$ in (22) for various values of θ.

Furthermore, the convex relaxation approaches (which provide low complexity solutions) perform very closely to the exact solutions of the optimal stochastic parameter design problem for small values of D.

In order to provide further explanations of the results in Fig. 2, Fig. 3 illustrates $g_\theta(\mathrm{x})$ in (22) for $\theta = -5$, $\theta = 0$, and $\theta = 5$. As expected from the expression in (22), each function in the figure is a shifted version of the others. Also, this figure can be used to determine when the unconstrained solution coincides with the solutions of the optimal stochastic and the optimal deterministic parameter designs. For example, for $\theta = -5$, the global minimum of $g_\theta(\mathrm{x})$ is achieved at -1.223, which already satisfies the constraint. Therefore, all the three approaches yield the same conditional risk for that parameter (see Fig. 2). On the other hand, for $\theta = 5$, the global minimum is at 8.777; hence, the conditional risk obtained from the unconstrained problem in (7) cannot be achieved by the constrained approaches. Specifically, the optimal deterministic approach in (14) chooses the minimum value in the interval $[-5, 5]$, which results in the optimal signal value of $s_\theta^{\text{opt}} = 0.81$. On the other hand, the solution of the optimal stochastic parameter design problem in (12) results in a randomization between 8.741 and 0.809 with probabilities of 0.321 and 0.679, respectively, and achieves a lower conditional risk than the deterministic approach (see Fig. 2). In Table I, the optimal solutions for the optimal stochastic, the optimal deterministic and

TABLE I

OPTIMAL STOCHASTIC SOLUTION $p_{\mathrm{s}_\theta}^{\text{opt}}(\mathrm{x}) = \lambda_\theta\,\delta(\mathrm{x} - \mathrm{s}_{\theta,1}) + (1 - \lambda_\theta)\,\delta(\mathrm{x} - \mathrm{s}_{\theta,2})$, OPTIMAL DETERMINISTIC SOLUTION $\mathrm{s}_\theta^{\text{opt}}$, AND UNCONSTRAINED SOLUTION $\mathrm{s}_\theta^{\text{unc}}$.

θ	λ_θ	$\mathrm{s}_{\theta,1}$	$\mathrm{s}_{\theta,2}$	$\mathrm{s}_\theta^{\text{opt}}$	$\mathrm{s}_\theta^{\text{unc}}$
-5	1	-1.223	-	-1.223	-1.223
-3	1	0.777	-	0.777	0.777
-1.5	0.295	-0.331	1.774	1.5	2.277
0	1	0	-	0	3.777
1.5	0.42	-0.294	-1.954	-1.5	5.277
3	0.826	-1.177	6.719	-1.19	6.777
5	0.679	0.809	8.741	0.81	8.777

the unconstrained parameter design approaches are presented for various values of θ. Fig. 3 can also be used to explain the oscillatory behavior of the convex relaxation solutions in Fig. 2. Since the convex relaxation approach considers possible s_θ values as $-10 : \mathrm{D} : 10$ and since $g_\theta(\mathrm{x})$ shifts with θ, the signal values obtained from the convex optimization problem in (13) move around the optimal values of the exact solution periodically. Finally, the conditions in Proposition 3 are evaluated for different θ values, and it is observed that they provide sufficient but not necessary conditions for specifying improvements via the stochastic parameter design over the deterministic one. For example, the calculations show that the conditions in Proposition 3 are satisfied for $\theta \in [-1.381, -1.31]$ and $\theta \in [1.397, 1.536]$, and improvements are observed in Fig. 2 for those values of θ.

Future work involves the investigation of the stochastic parameter design problem in the presence of partial knowledge of the noise distribution. The robustness of the stochastic parameter design will be analyzed, and various design approaches will be considered.

Acknowledgments: *The authors would like to thank the editor, Dr. Ta-Hsin Li, for suggesting the example in the fourth paragraph of the introduction.*

REFERENCES

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

[2] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[3] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I–Fixed detectors," *IEEE Trans. Sig. Processing*, vol. 55, no. 7, pp. 3172–3184, July 2007.

[4] S. M. Kay, "Noise enhanced detection as a special case of randomization," *IEEE Sig. Processing Lett.*, vol. 15, pp. 709–712, 2008.

[5] C. Goken, S. Gezici, and O. Arikan, "Optimal stochastic signaling for power-constrained binary communications systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3650–3661, Dec. 2010.

[6] H. Chen, P. K. Varshney, and J. H. Michels, "Noise enhanced parameter estimation," *IEEE Trans. Sig. Processing*, vol. 56, no. 10, pp. 5074–5081, Oct. 2008.

[7] B. Dulek and S. Gezici, "Detector randomization and stochastic signaling for minimum probability of error receivers," *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 923–928, Apr. 2012.

[8] A. Ozcelikkale, H. M. Ozaktas, and E. Arikan, "Signal recovery with cost-constrained measurements," *IEEE Trans. Sig. Processing*, vol. 58, no. 7, pp. 3607–3617, July 2010.

[9] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks: Technology, Protocols, and Applications*. Hoboken, New Jersey: John Wiley & Sons, 2007.

[10] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained signal detection," *IEEE Trans. Sig. Processing*, vol. 57, no. 5, pp. 1655–1669, May 2009.

[11] M. Azizoglu, "Convexity properties in binary detection problems," *IEEE Trans. Inform. Theory*, vol. 42, no. 4, pp. 1316–1321, July 1996.

[12] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Boston, MA: Athena Specific, 2003.

[13] S. Bayram, N. D. Vanli, B. Dulek, I. Sezer, and S. Gezici, "Optimum power allocation for average power constrained jammers in the presence of non-Gaussian noise," *IEEE Commun. Letters*, 2012, in press.

[14] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization method for constrained optimization problems*. IOS Press, 2002, pp. 214–220, in Intelligent Technologies–Theory and Applications: New Trends in Intelligent Technologies.

[15] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. New York: Springer, 2005.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[17] S. Bayram and S. Gezici, "On the improvability and nonimprovability of detection via additional independent noise," *IEEE Sig. Processing Lett.*, vol. 16, no. 11, pp. 1001–1004, Nov. 2009.