

CONSTRAINED NONLINEAR MINIMUM MSE ESTIMATION

Tomer Michaeli and Yonina C. Eldar

Department of Electrical Engineering
Technion—Israel Institute of Technology, Haifa, Israel
tomermic@tx.technion.ac.il, yonina@ee.technion.ac.il

ABSTRACT

We address the problem of minimum mean-squared error (MMSE) estimation where the estimator is constrained to belong to a pre-defined set of functions. We derive a simple closed form formula that reveals the structure of the restricted estimator for a wide class of constraints. Using this formula we study various types of constrained estimation problems that arise commonly in the fields of signal processing and communication.

Index Terms— Constrained estimation, Nonlinear estimation, MMSE estimation.

1. INTRODUCTION

Constrained Bayesian estimation refers to the problem of estimating a random vector (r.v.) $\mathbf{x}$ based on a realization of the r.v. $\mathbf{y}$ subject to a restriction on the types of estimators we can use. Specifically, we wish to design an estimator ϕ that minimizes the mean squared error (MSE) $E[\|\mathbf{x} - \phi(\mathbf{y})\|^2]$ under the constraint that ϕ belongs to a certain family of functions.

If no constraint is imposed on ϕ , then it is well known that the minimum MSE (MMSE) estimator is $\phi_0(\mathbf{y}) = E[\mathbf{x}|\mathbf{y}]$. As for the constrained case, a few problems were studied in the past. Perhaps the most famous restriction for which a formula is available is that ϕ be a linear function. This is known as the linear MMSE (LMMSE) estimator [1]. Other classic results include the constraint that ϕ be a lower triangular matrix (the finite dimensional version of the causal Wiener filter [2]), a low rank matrix [1], a matrix that causes the covariance of the estimated vector to possess a predefined structure [3], and a few more.

In all of the examples above, the estimator ϕ is linear. In [4] we addressed the general problem of constrained LMMSE estimation, namely the problem of designing an MMSE linear estimate subject to linear or nonlinear constraints on the estimate coefficients. We presented a generic formula that applies to a wide range of such problems. These include all of the results above and many more. Nevertheless, it seems that very little is known on constrained *non-linear* estimation problems (i.e. where ϕ is nonlinear). Our goal here is to study these types of problems.

One mathematical tool for designing a constrained estimator is the orthogonality principle, which is adequate only for linear restrictions¹. Constraining ϕ to be a linear transformation or a lower triangular matrix may be handled using the orthogonality principle, but a restriction of the type $E[\|\phi(\mathbf{y})\|] \leq \varepsilon$ is nonlinear and must be handled differently. In a recent paper [5] we provided a generalization of the orthogonality principle to the case of convex restric-

tions², called the extended orthogonality principle. This principle is an inequality that the optimal constrained estimator must satisfy. Its disadvantage, thus, is that it is not constructive in the sense that it does not lead to an equation whose solution is the desired estimator. Furthermore, among the various problems studied in [5] using this principle, only in one example did we consider a nonlinear ϕ . Specifically, we constructed an estimator ϕ that minimizes the MSE subject to the constraint $\phi(\mathbf{y}) \in \mathcal{A}$ for all $\mathbf{y}$, where $\mathcal{A}$ is a closed convex set in $\mathbb{R}^n$.

In this paper we prove a simple, yet powerful, theorem which reveals the structure of a very large class of constrained nonlinear estimators. This class contains, for example, the deterministic restriction on $\phi(\mathbf{y})$ studied in [5]. Our approach also allows the treatment of stochastic constraints on $\phi(\mathbf{y})$, i.e. constraints on the statistical properties of the r.v. $\hat{\mathbf{x}} = \phi(\mathbf{y})$. We demonstrate the theorem in the contexts of two types of problems which commonly arise in signal processing and communication applications. The first is restricting the estimated vector. Specifically, we obtain a closed form solution to the minimization of the MSE subject to the following constraints: $E[\|\phi(\mathbf{y})\|^2] \leq \varepsilon$, $E[\|\phi(\mathbf{y})\|] \leq \varepsilon$ and $\|\phi(\mathbf{y})\| \leq \varepsilon$. The second type of problems is the design of an estimator which is resistant to an interference $\mathbf{z}$ which may be present at its input. In this scheme we minimize $E[\|\mathbf{x} - \phi(\mathbf{y})\|^2]$ subject to the constraints $E[\|\phi(\mathbf{z})\|^2] \leq \varepsilon$, $E[\|\phi(\mathbf{z})\|] \leq \varepsilon$ and $f_{\mathbf{z}}(\mathbf{z})\|\phi(\mathbf{z})\| \leq \varepsilon$, where $f_{\mathbf{z}}(\mathbf{z})$ is the probability density function (pdf) of the r.v. $\mathbf{z}$.

2. CONSTRAINED MMSE ESTIMATION

We now present our main result regarding the imposition of restrictions on the MMSE estimator. We assume that $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ and thus the estimator is a function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$. To this end we need to introduce a few definitions regarding such mappings. We denote the set of all squared integrable functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ as

$$\mathcal{L}_2 = \{\psi: \mathbb{R}^n \rightarrow \mathbb{R}^m \mid \int_{\mathbb{R}^n} \|\psi(\mathbf{y})\|^2 d\mathbf{y} < \infty\}, \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$. In the following derivations we use the concept of projections of functions in $\mathcal{L}_2$ onto closed sets. Let g be a function in $\mathcal{L}_2$ and let $\mathcal{W} \subseteq \mathcal{L}_2$ be a closed set. Then the projection of g onto $\mathcal{W}$ is defined by

$$P_{\mathcal{W}}(g) = \arg \min_{\psi \in \mathcal{W}} \int_{\mathbb{R}^n} \|g(\mathbf{y}) - \psi(\mathbf{y})\|^2 d\mathbf{y}. \quad (2)$$

In our constrained estimation setup, we are interested in confining ϕ to belong to a certain family of functions. Note that if $\mathbf{y}$ is such that $f_{\mathbf{y}}(\mathbf{y}) = 0$ then the value of $\phi(\mathbf{y})$ can be chosen arbitrarily since it does not affect the MSE. As for the behavior of ϕ on

¹A linear constraint is a set of functions $\mathcal{S}$ that form a subspace, i.e. if $\phi_1, \phi_2 \in \mathcal{S}$ then also $\alpha\phi_1 + \beta\phi_2 \in \mathcal{S}$ for every $\alpha, \beta \in \mathbb{R}$.

²A convex constraint is a set of functions $\mathcal{S}$ that form a convex set, i.e. if $\phi_1, \phi_2 \in \mathcal{S}$ then also $\alpha\phi_1 + (1 - \alpha)\phi_2 \in \mathcal{S}$ for every $\alpha \in [0, 1]$.

the rest of $\mathbb{R}^n$, we are interested in constraints that can be cast as $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, where $f_{\mathbf{y}}(\mathbf{y})$ is the pdf of $\mathbf{y}$ and $\mathcal{W}$ is a closed set in $\mathcal{L}_2$. This somewhat unnatural representation has two reasons. First, the solution to this constrained estimation problem has a very simple structure which involves the projection operator $P_{\mathcal{W}}$, as presented in Theorem 1 below. Second, this representation is very convenient for the purpose of imposing stochastic constraints on $\hat{\mathbf{x}}$, as we show in the sequel.

We emphasize that the class of constraints that can be handled within this framework is very large. From the viewpoint of the estimator ϕ , this representation includes all restrictions of the form $\phi \in \mathcal{V}$ given that the induced set $\mathcal{W} = \{\phi \mid \phi(\mathbf{y}) = \sqrt{f_{\mathbf{y}}}(\mathbf{y})\psi(\mathbf{y}), \psi \in \mathcal{V}\}$ is closed. The solution to this problem is presented in the following theorem.

Theorem 1 *Let $\mathbf{x} \in \mathbb{R}^m$ be a finite variance r.v. (i.e. $E[\|\mathbf{x}\|^2] < \infty$), let $\mathbf{y} \in \mathbb{R}^n$ be a r.v. with marginal pdf $f_{\mathbf{y}}(\mathbf{y})$ and let $\mathcal{W}$ be a closed set in $\mathcal{L}_2$. Then among all estimators of the form $\hat{\mathbf{x}} = \phi(\mathbf{y})$ that satisfy $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, an estimator that minimizes the MSE $E[\|\mathbf{x} - \phi(\mathbf{y})\|^2]$ is given by*

$$\phi(\mathbf{y}) = \begin{cases} \frac{1}{\sqrt{f_{\mathbf{y}}(\mathbf{y})}} \left\{ P_{\mathcal{W}} \left(\sqrt{f_{\mathbf{y}}} \phi_0 \right) \right\}(\mathbf{y}) & f_{\mathbf{y}}(\mathbf{y}) \neq 0 \\ 0 & f_{\mathbf{y}}(\mathbf{y}) = 0, \end{cases} \quad (3)$$

where $P_{\mathcal{W}}$ is the projection operator onto the set $\mathcal{W}$ defined by (2) and $\phi_0(\mathbf{y}) = E[\mathbf{x}|\mathbf{y}]$ is the unconstrained MMSE estimator.

Note that since $\phi(\mathbf{y})$ can be chosen arbitrarily wherever $f_{\mathbf{y}}(\mathbf{y}) = 0$, of all the possible solutions, $\phi(\mathbf{y})$ in (3) has the minimal variance $E[\|\phi(\mathbf{y})\|^2]$. As for vectors $\mathbf{y}$ at which $f_{\mathbf{y}}(\mathbf{y}) > 0$, the solution is guaranteed to be unique if $\mathcal{W}$ is a convex set. Theorem 1 can be extended to the case where $\mathbf{x}$ and $\mathbf{y}$ do not have densities by using the Radon–Nikodym derivative of measures.

Theorem 1 implies two key stages in the design of a constrained estimator. The first is to present the restriction in the form $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, where $\mathcal{W}$ is a closed set in $\mathcal{L}_2$. The second is the derivation of a formula for the projection operator $P_{\mathcal{W}}$. Once these two ingredients are available, the solution is readily obtained by Theorem 1.

As a special case of Theorem 1, we can obtain the following.

Corollary 2 *Let $\mathcal{A}$ be a closed set in $\mathbb{R}^m$, then among all functions $\phi : \mathbb{R}^n \rightarrow \mathcal{A}$, the MMSE estimator is*

$$\phi(\mathbf{y}) = P_{\mathcal{A}}(E[\mathbf{x}|\mathbf{y}]), \quad (4)$$

where $P_{\mathcal{A}}$ here is the projection of a vector in $\mathbb{R}^m$ onto the set $\mathcal{A}$.

The proof of the corollary relies on Theorem 1 and is omitted due to lack of space. It can be seen that a deterministic restriction of the type $\hat{\mathbf{x}} \in \mathcal{A}$ (for every realization) leads to a simple and intuitive result. The constrained estimate is the projection of the unrestricted estimate $E[\mathbf{x}|\mathbf{y}]$ onto the set $\mathcal{A}$. Corollary 2 provides a generalization of [5, Theorem 2], where $\mathcal{A}$ was assumed to be a convex set.

In the following sections we give a few examples of Theorem 1, where the constraint is not necessarily deterministic.

3. BOUNDING THE ESTIMATED VECTOR

An estimator is many times one block in a larger scheme. One example is when a noisy signal is first to be cleaned and then transmitted under certain limitations or coded efficiently. These tasks usually require that the estimated signal be bounded in some sense. In the

following subsections we give a few examples for such situations and derive appropriate estimators using our constrained estimation framework.

3.1. Squared Norm Limitation

Suppose that after estimating $\mathbf{x}$ from $\mathbf{y}$, the estimated vector is transmitted under a power limitation. We may take into account this power constraint by designing an estimator that minimizes the MSE under the restriction that the variance of $\hat{\mathbf{x}}$ does not exceed a given threshold ε . Our problem is thus

$$\begin{aligned} \arg \min_{\phi} E[\|\mathbf{x} - \phi(\mathbf{y})\|^2] \\ \text{s.t. } E[\|\phi(\mathbf{y})\|^2] \leq \varepsilon. \end{aligned} \quad (5)$$

To tackle this problem within our framework we need to express the constraint as a restriction on $\sqrt{f_{\mathbf{y}}}\phi$. This can be done by writing

$$E[\|\phi(\mathbf{y})\|^2] = \int_{\mathbb{R}^n} \left\| \sqrt{f_{\mathbf{y}}}(\mathbf{y})\phi(\mathbf{y}) \right\|^2 d\mathbf{y}. \quad (6)$$

Hence our constraint can be cast as $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, where $\mathcal{W}$ is the $\mathcal{L}_2$ ball

$$\mathcal{W} = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{R}^m \mid \int_{\mathbb{R}^n} \|\psi(\mathbf{y})\|^2 d\mathbf{y} \leq \varepsilon \right\}. \quad (7)$$

The projection operator $P_{\mathcal{W}}(\cdot)$ onto an $\mathcal{L}_2$ ball simply scales its argument to comply with the norm limitation. Thus, using Theorem 1, the solution to this problem is

$$\phi(\mathbf{y}) = cE[\mathbf{x}|\mathbf{y}], \quad (8)$$

where $c \leq 1$ is the largest value for which the power limitation is satisfied. Evidently, constraining the variance of $\hat{\mathbf{x}}$ is achieved by using a scaled version of the unconstrained estimator $E[\mathbf{x}|\mathbf{y}]$.

3.2. Norm Limitation

The squared norm limitation puts a heavy penalty on vectors with large norms. In many cases this may be an over-pessimistic modeling of the problem at hand. Commonly, one is willing to sacrifice a few rare events of $\|\hat{\mathbf{x}}\|$ being large in order to obtain a small norm a majority of the times. This behavior can be achieved by setting the limitation $E[\|\hat{\mathbf{x}}\|] \leq \varepsilon$. Therefore, we are interested in the problem

$$\begin{aligned} \arg \min_{\phi} E[\|\mathbf{x} - \phi(\mathbf{y})\|^2] \\ \text{s.t. } E[\|\phi(\mathbf{y})\|] \leq \varepsilon. \end{aligned} \quad (9)$$

The expectation of the norm of $\hat{\mathbf{x}}$ can be expressed as

$$E[\|\phi(\mathbf{y})\|] = \int_{\mathbb{R}^n} \left\| \sqrt{f_{\mathbf{y}}}(\mathbf{y})\phi(\mathbf{y}) \right\| \sqrt{f_{\mathbf{y}}}(\mathbf{y}) d\mathbf{y}. \quad (10)$$

Therefore, we can identify this as a constrained estimation problem with the restriction $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, where $\mathcal{W}$ is the weighted $\mathcal{L}_1$ ball

$$\mathcal{W} = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{R}^m \mid \int_{\mathbb{R}^n} \|\psi(\mathbf{y})\| \sqrt{f_{\mathbf{y}}}(\mathbf{y}) d\mathbf{y} \leq \varepsilon \right\}. \quad (11)$$

The projection $f = P_{\mathcal{W}}(g)$ of a function $g \in \mathcal{L}_2$ onto this set is given by

$$f(\mathbf{y}) = \begin{cases} 0 & \|g(\mathbf{y})\| \leq \delta \sqrt{f_{\mathbf{y}}}(\mathbf{y}) \\ g(\mathbf{y}) - \delta \sqrt{f_{\mathbf{y}}}(\mathbf{y}) \frac{g(\mathbf{y})}{\|g(\mathbf{y})\|} & \|g(\mathbf{y})\| > \delta \sqrt{f_{\mathbf{y}}}(\mathbf{y}), \end{cases} \quad (12)$$

where $\delta \geq 0$ is the minimum value for which $f \in \mathcal{W}$. Using Theorem 1 along with formula (12), the bounded norm estimator is

$$\phi(\mathbf{y}) = \begin{cases} 0 & \|E[\mathbf{x}|\mathbf{y}]\| \leq \delta \\ E[\mathbf{x}|\mathbf{y}] - \delta \frac{E[\mathbf{x}|\mathbf{y}]}{\|E[\mathbf{x}|\mathbf{y}]\|} & \|E[\mathbf{x}|\mathbf{y}]\| > \delta. \end{cases} \quad (13)$$

Like (8), the above estimator is a modification of the unconstrained estimator $E[\mathbf{x}|\mathbf{y}]$. Only now, the larger $\|E[\mathbf{x}|\mathbf{y}]\|$ is the less the relative modification is. This is balanced by mapping a set of vectors $\mathbf{y}$ to the zero vector so that $\hat{\mathbf{x}} = 0$ with nonzero probability.

3.3. Per-Realization Norm Limitation

There may be cases where the limitation on $\|\hat{\mathbf{x}}\|$ should be enforced for every realization. This precludes the option of using one of the above estimators because neither $E[\|\hat{\mathbf{x}}\|^2] \leq \varepsilon$ nor $E[\|\hat{\mathbf{x}}\|] \leq \varepsilon$ can guaranty that $\|\hat{\mathbf{x}}\|$ is smaller than some constant with probability 1. Thus we are interested in obtaining a closed form solution to the following problem

$$\begin{aligned} \arg \min_{\phi} E[\|\mathbf{x} - \phi(\mathbf{y})\|^2] \\ \text{s.t. } \sup_{\mathbf{y} \in \mathbb{R}^n} \|\phi(\mathbf{y})\| \leq \varepsilon. \end{aligned} \quad (14)$$

Note that this problem is a special case of Corollary 2. To demonstrate how it can be solved directly using Theorem 1, let us express the restriction as

$$\sup_{\mathbf{y} \in \mathbb{R}^n} \|\phi(\mathbf{y})\| = \sup_{\mathbf{y} \in \mathbb{R}^n} \left\{ \left\| \phi(\mathbf{y}) \sqrt{f_{\mathbf{y}}(\mathbf{y})} \right\| \frac{1}{\sqrt{f_{\mathbf{y}}(\mathbf{y})}} \right\}. \quad (15)$$

We see that this is simply a weighted $\mathcal{L}_{\infty}$ constraint on $\sqrt{f_{\mathbf{y}}}\phi$, i.e. $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, where $\mathcal{W}$ is defined by

$$\mathcal{W} = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{R}^m \mid \sup_{\mathbf{y} \in \mathbb{R}^n} \left\{ \|\psi(\mathbf{y})\| \frac{1}{\sqrt{f_{\mathbf{y}}(\mathbf{y})}} \right\} \leq \varepsilon \right\}. \quad (16)$$

The projection $f = P_{\mathcal{W}}(g)$ of a function $g \in \mathcal{L}_2$ onto this set is given by

$$f(\mathbf{y}) = \begin{cases} g(\mathbf{y}) & \|g(\mathbf{y})\| \leq \sqrt{f_{\mathbf{y}}(\mathbf{y})}\varepsilon \\ \frac{g(\mathbf{y})}{\|g(\mathbf{y})\|} \sqrt{f_{\mathbf{y}}(\mathbf{y})}\varepsilon & \|g(\mathbf{y})\| > \sqrt{f_{\mathbf{y}}(\mathbf{y})}\varepsilon. \end{cases} \quad (17)$$

Therefore, employing Theorem 1 and using the projection formula (17), the bounded-realizations estimator is

$$\phi(\mathbf{y}) = \begin{cases} E[\mathbf{x}|\mathbf{y}] & \|E[\mathbf{x}|\mathbf{y}]\| \leq \varepsilon \\ \frac{E[\mathbf{x}|\mathbf{y}]}{\|E[\mathbf{x}|\mathbf{y}]\|} \varepsilon & \|E[\mathbf{x}|\mathbf{y}]\| > \varepsilon. \end{cases} \quad (18)$$

We note that (18) can be thought of as a two-stage estimator. We first compute the unconstrained estimate $E[\mathbf{x}|\mathbf{y}]$ and then project it onto a ball in $\mathbb{R}^m$.

4. RESISTANCE TO INTERFERENCE

Consider the setup where one constructs an estimator to recover $\mathbf{x}$ from the measurements $\mathbf{y}$ but there is uncertainty whether the r.v. $\mathbf{y}$ will indeed be fed to the estimator. Instead, an interference signal $\mathbf{z}$ with pdf $f_{\mathbf{z}}(\mathbf{z})$ might exist at the input. This situation is commonly encountered in speech and image denoising applications where the sparsity of the signal in some transform domain is exploited. In this

context $\mathbf{y}$ corresponds to a coefficient containing signal plus noise and $\mathbf{z}$ corresponds to a noise-only coefficient. A good estimator would be one that outputs the zero vector when applied to a realization of the r.v. $\mathbf{z}$ and outputs the MMSE estimate of $\mathbf{x}$ given $\mathbf{y}$ when applied to a realization of the r.v. $\mathbf{y}$. However this is an unrealizable approach as we do not know in advance whether the vector at the input was drawn from distribution $f_{\mathbf{z}}(\mathbf{z})$ or $f_{\mathbf{y}}(\mathbf{y})$. In the following subsections we suggest a few strategies for solving this problem using our constrained estimation framework.

4.1. Squared Norm Resistance

One way to obtain resistance to $\mathbf{z}$ is as follows. Let us design an estimator that minimizes the MSE between $\mathbf{x}$ and $\phi(\mathbf{y})$ under the constraint that the variance at the output of the estimator should be smaller than ε when applied to the r.v. $\mathbf{z}$. Specifically, we are interested in solving

$$\begin{aligned} \arg \min_{\phi} E[\|\mathbf{x} - \phi(\mathbf{y})\|^2] \\ \text{s.t. } E[\|\phi(\mathbf{z})\|^2] \leq \varepsilon. \end{aligned} \quad (19)$$

Let us express the variance of $\phi(\mathbf{z})$ as

$$E[\|\phi(\mathbf{z})\|^2] = \int_{\mathbb{R}^n} \left\| \sqrt{f_{\mathbf{y}}(\mathbf{y})} \phi(\mathbf{y}) \right\|^2 \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} d\mathbf{y}. \quad (20)$$

This representation allows us to cast (19) as a constrained estimation problem with the restriction $\sqrt{f_{\mathbf{y}}}\phi \in \mathcal{W}$, where $\mathcal{W}$ is the weighted $\mathcal{L}_2$ ball

$$\mathcal{W} = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{R}^m \mid \int_{\mathbb{R}^n} \|\psi(\mathbf{y})\|^2 \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} d\mathbf{y} \leq \varepsilon \right\}. \quad (21)$$

The projection $f = P_{\mathcal{W}}(g)$ of a function $g \in \mathcal{L}_2$ onto this set is given by

$$f(\mathbf{y}) = \frac{1}{1 + \lambda \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})}} g(\mathbf{y}), \quad (22)$$

where $\lambda \geq 0$ is the minimum value for which the constraint is satisfied. Therefore, employing Theorem 1 and using the projection formula (22), the optimal squared norm $\mathbf{z}$ -resistant estimator is

$$\phi(\mathbf{y}) = \frac{1}{1 + \lambda \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})}} E[\mathbf{x}|\mathbf{y}]. \quad (23)$$

Interestingly, the function ϕ in (23) is a concatenation of a Bayesian soft-decision rule followed by the MMSE estimator. The term $f_{\mathbf{z}}(\mathbf{y})/f_{\mathbf{y}}(\mathbf{y})$ is the likelihood ratio (LR) which measures how likely is the hypothesis that the interference $\mathbf{z}$ was received over the hypothesis that the measurement $\mathbf{y}$ was received. The scalar function $1/(1 + \lambda LR)$ in (23) is close to 0 if the LR is large, thus achieving resistance to $\mathbf{z}$. On the other hand, this function is close to 1 for vectors with small LR, which makes ϕ act approximately as the unconstrained MMSE estimator $E[\mathbf{x}|\mathbf{y}]$.

4.2. Norm Resistance

The squared norm $\mathbf{z}$ -resistant estimator developed above may be inadequate for certain applications. The reason is that it usually does not output the zero vector even for inputs which are very likely to be interference. To obtain a hard-decision rule, we may consider the following alternative approach. As opposed to bounding the variance of $\phi(\mathbf{z})$ we wish to bound its probability of being

nonzero. Explicitly, we wish to minimize the MSE under the constraint $\Pr \{\phi(\mathbf{z}) \neq 0\} \leq \varepsilon$. Unfortunately, this constraint is non convex and there is no closed form available for the resulting projection.

The restriction $\Pr \{\phi(\mathbf{z}) \neq 0\} \leq \varepsilon$ can be thought of as imposing that $\|\phi(\mathbf{z})\|$ be sparse (as a function over $\mathbb{R}^n$). It is known that an $\mathcal{L}_1$ restriction is a good approximation for sparsity. Thus, we can replace our constraint by the requirement $E[\|\phi(\mathbf{z})\|] \leq \varepsilon$. Our problem is thus

$$\begin{aligned} \arg \min_{\phi} E[\|\mathbf{x} - \phi(\mathbf{y})\|^2] \\ \text{s.t. } E[\|\phi(\mathbf{z})\|] \leq \varepsilon. \end{aligned} \quad (24)$$

The expectation of the norm of $\phi(\mathbf{z})$ can be written as

$$E[\|\phi(\mathbf{z})\|] = \int_{\mathbb{R}^n} \left\| \sqrt{f_{\mathbf{y}}}(\mathbf{y}) \phi(\mathbf{y}) \right\| \frac{f_{\mathbf{z}}(\mathbf{y})}{\sqrt{f_{\mathbf{y}}}(\mathbf{y})} d\mathbf{y}. \quad (25)$$

This allows us to formulate the constraint as $\sqrt{f_{\mathbf{y}}} \phi \in \mathcal{W}$, where $\mathcal{W}$ is the weighted $\mathcal{L}_1$ ball

$$\mathcal{W} = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{R}^m \left| \int_{\mathbb{R}^n} \|\psi(\mathbf{y})\| \frac{f_{\mathbf{z}}(\mathbf{y})}{\sqrt{f_{\mathbf{y}}}(\mathbf{y})} d\mathbf{y} \leq \varepsilon \right. \right\}. \quad (26)$$

The projection of a function $g \in \mathcal{L}_2$ onto this set is given by (12) with the weighting function $1/\sqrt{f_{\mathbf{y}}}(\mathbf{y})$ replaced by $f_{\mathbf{z}}(\mathbf{y})/\sqrt{f_{\mathbf{y}}}(\mathbf{y})$. Employing Theorem 1 and using the projection operator (12), the norm $\mathbf{z}$ -resistant estimator is

$$\phi(\mathbf{y}) = \begin{cases} 0 & \|E[\mathbf{x}|\mathbf{y}]\| \leq \delta \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} \\ E[\mathbf{x}|\mathbf{y}] - \delta \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} \frac{E[\mathbf{x}|\mathbf{y}]}{\|E[\mathbf{x}|\mathbf{y}]\|} & \|E[\mathbf{x}|\mathbf{y}]\| > \delta \frac{f_{\mathbf{z}}(\mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})}, \end{cases} \quad (27)$$

where $\delta \geq 0$ is the minimum value for which the constraint is satisfied.

We see that the above estimator, as opposed to the variance constrained estimator, comprises a hard-decision rule. This means that it outputs the zero vector with nonzero probability.

As in the variance constrained estimator, here too the LR plays an important role. To understand its influence on $\phi(\mathbf{y})$ let us recall that there are two forces that shape the structure of the estimator. On one hand $\phi(\mathbf{y})$ should be as close as possible to $E[\mathbf{x}|\mathbf{y}]$. Thus if we are to set $\phi(\mathbf{y}) = 0$, it should be only for vectors $\mathbf{y}$ where $\|E[\mathbf{x}|\mathbf{y}]\|$ is small. On the other hand, vectors $\mathbf{y}$ for which the LR is large correspond to the hypothesis that interference was received. Hence, for such vectors we would like the estimator to output the zero vector, i.e. $\phi(\mathbf{y}) = 0$. The affect of these two driving forces is nicely seen in (27). The decision whether to output the zero vector or not is made by comparing $\|E[\mathbf{x}|\mathbf{y}]\|$ against the LR (multiplied by the constant δ). Furthermore, for vectors $\mathbf{y}$ at which $\|E[\mathbf{x}|\mathbf{y}]\| > \delta \text{LR}$, the estimator $\phi(\mathbf{y})$ is a shrunk version of $E[\mathbf{x}|\mathbf{y}]$. The amount of shrinkage is, again, determined by the LR.

4.3. Maximal Rejection

The motivation for the last estimator was the achievement of a threshold rule. We now consider the other extreme. Suppose that the principle that guides us is the following. The more likely the hypothesis that interference was received, the smaller the norm at the output of the estimator is allowed to be. This is a per-realization demand which is very different than the average restrictions $E[\|\phi(\mathbf{z})\|^2] \leq \varepsilon$ and $E[\|\phi(\mathbf{z})\|] \leq \varepsilon$ treated in the previous

subsections. This behavior can be obtained by imposing the constraint $f_{\mathbf{z}}(\mathbf{y}) \|\phi(\mathbf{y})\| \leq \varepsilon, \forall \mathbf{y} \in \mathbb{R}^n$, i.e. we wish to solve

$$\begin{aligned} \arg \min_{\phi} E[\|\mathbf{x} - \phi(\mathbf{y})\|^2] \\ \text{s.t. } \sup_{\mathbf{y} \in \mathbb{R}^n} \{f_{\mathbf{z}}(\mathbf{y}) \|\phi(\mathbf{y})\|\} \leq \varepsilon. \end{aligned} \quad (28)$$

The above constraint can be cast as a restriction on $\sqrt{f_{\mathbf{y}}} \phi$ by writing

$$\sup_{\mathbf{y} \in \mathbb{R}^n} \{f_{\mathbf{z}}(\mathbf{y}) \|\phi(\mathbf{y})\|\} = \sup_{\mathbf{y} \in \mathbb{R}^n} \left\{ \left\| \phi(\mathbf{y}) \sqrt{f_{\mathbf{y}}}(\mathbf{y}) \right\| \frac{f_{\mathbf{z}}(\mathbf{y})}{\sqrt{f_{\mathbf{y}}}(\mathbf{y})} \right\}. \quad (29)$$

Thus it is associated with the weighted $\mathcal{L}_{\infty}$ ball defined by

$$\mathcal{W} = \left\{ \psi : \mathbb{R}^n \rightarrow \mathbb{R}^m \left| \sup_{\mathbf{y} \in \mathbb{R}^n} \left\{ \left\| \psi(\mathbf{y}) \right\| \frac{f_{\mathbf{z}}(\mathbf{y})}{\sqrt{f_{\mathbf{y}}}(\mathbf{y})} \right\} \leq \varepsilon \right. \right\}. \quad (30)$$

The projection of a function $g \in \mathcal{L}_2$ onto this set is given by (17) with the weighting function $1/\sqrt{f_{\mathbf{y}}}(\mathbf{y})$ replaced by $f_{\mathbf{z}}(\mathbf{y})/\sqrt{f_{\mathbf{y}}}(\mathbf{y})$. Using Theorem 1 with the projection formula (17), the maximal rejection estimator is

$$\phi_{\mathcal{W}}(\mathbf{y}) = \begin{cases} E[\mathbf{x}|\mathbf{y}] & \|E[\mathbf{x}|\mathbf{y}]\| \leq \frac{\varepsilon}{f_{\mathbf{z}}(\mathbf{y})} \\ \frac{\varepsilon}{f_{\mathbf{z}}(\mathbf{y})} \frac{E[\mathbf{x}|\mathbf{y}]}{\|E[\mathbf{x}|\mathbf{y}]\|} & \|E[\mathbf{x}|\mathbf{y}]\| > \frac{\varepsilon}{f_{\mathbf{z}}(\mathbf{y})}. \end{cases} \quad (31)$$

The above estimator acts in a complete different manner than (27). It coincides with the unconstrained estimator wherever $\|E[\mathbf{x}|\mathbf{y}]\|$ is small and acts as a shrunk version of it where $\|E[\mathbf{x}|\mathbf{y}]\|$ is large. Moreover, it interestingly does not depend explicitly on the LR but rather solely on $f_{\mathbf{z}}(\mathbf{y})$. This stems from the fact that we want to ensure that every single realization of $\mathbf{z}$ satisfies $f_{\mathbf{z}}(\mathbf{z}) \|\phi(\mathbf{z})\| \leq \varepsilon$. Therefore, the value of $f_{\mathbf{y}}(\mathbf{y})$ is irrelevant in this respect. The output of the estimator is clipped wherever $\|E[\mathbf{x}|\mathbf{y}]\| f_{\mathbf{z}}(\mathbf{y})$ exceeds ε .

5. CONCLUSIONS

We presented a general framework for solving constrained MMSE estimation problems. We demonstrated our approach on a series of problems encountered in signal processing and communications. We believe that the method developed in this paper can find many more applications.

6. REFERENCES

- [1] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley Pub. Co, 1991.
- [2] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. MIT Press, Cambridge, Mass. and Wiley, New York, 1949.
- [3] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 686–697, 2003.
- [4] T. Michaeli and Y. C. Eldar, "Constrained Linear Minimum MSE Estimation," *CCIT Report 641, EE Dept., Technion-Israel Institute of Technology*, 2007.
- [5] —, "Minimum MSE Estimation with Convex Constraints," *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP'07)*, vol. 3, 2007.