A MINIMAX APPROACH FOR MEAN SQUARE DENOISING

Jean-Christophe Pesquet

Institut Gaspard Monge and UMR-CNRS 8049,
Université de Marne la Vallée,
77454 Marne la Vallée Cedex 2, France
email : pesquet@univ-mlv.fr

Yonina Eldar

Department of Electrical Engineering,
Technion, Israel Institute of Technology,
Haifa, Israel 32000
email : yonina@ee.technion.ac.il

ABSTRACT

Minimax estimation aims at finding optimal estimators in the worst case situation compatible with the available information. In the present work, we consider the minimax mean square denoising of a random vector using a nonlinear estimator. The data set over which the minimax estimator is looked for takes the general form of a convex set where the correlation matrix of the data is constrained to lie. Also, additional convex contraints on the weights defining the estimator can be taken into account in the proposed approach.

1. INTRODUCTION

Minimax estimation is a cautious approach aiming at finding optimal estimators in the worst case situation compatible with the available information. Minimax problems have been the focus of many works in statistics, especially in the former Soviet Union. More recently, a renewal of interest has been observed for minimax signal estimation problems and, among many interesting new results, connections between these problems and Linear Matrix Inequalities have been shown in [1, 2, 3] for some kind of minimax mean square estimation problems.

In the meantime, nonlinear estimation has gained popularity in signal processing problems. For example, wavelet regression methods using thresholding operators have been developed and they have been shown to be optimal from an asymptotic minimax viewpoint for certain classes of regular signals [4]. A different approach for optimizing nonlinear estimors relies on the use of Stein's Unbiased Risk Estimator (SURE) [5] in problems involving additive Gaussian noise. A SURE method was used in [6] to build adaptive thresholding estimators. A similar approach was used in [7] to develop more sophisticated multivariate estimates for multicomponent image denoising. The SURE principle was also applied in [8] for signal restoration problems formulated as constrained convex optimization problems. More precisely, convex constraints on the signal to be recovered have been derived from empirical moments of the observed signal.

In the present work, we consider the minimax mean square estimation of a random vector through a nonlinear estimator. To be more precise the considered estimators are weighted sums of nonlinear functions of the observations. Unlike many previous works, the data set over which the minimax estimator is looked for is not restricted to a ball in some normed space but it takes the more general form of a convex set where the correlation matrix of the data is constrained to lie. Also, additional convex contraints on the weights defining the estimator can be taken into account in the proposed approach. Having the ability to introduce more flexible constraints in the definition of the minimax estimator may be important in order to increase its effectiveness in practical situations.

The structure of the considered nonlinear mean square estimator is described in Section 2. The general minimax formulation of the problem is investigated in Section 3. The algorithmic issues are investigated in Section 4. Section 5 describes how this framework can be applied to signal denoising in the presence of additive Gaussian noise, by exploiting Stein's principle. Finally, some conclusions are drawn in Section 6.

2. PROBLEM STATEMENT

We consider the following problem, which is classical in statistical signal processing.

A random vector $\mathbf{x} \in \mathbb{C}^n$ is to be estimated from an observed random vector $\mathbf{y} \in \mathbb{C}^m$. The two vectors $\mathbf{x}$ and $\mathbf{y}$ are assumed to be defined on the same probability space. To allow for non-linear estimators without considerably complicating the problem, in our development, we consider estimators of the form

$$\widehat{\mathbf{x}} = \boldsymbol{L}\boldsymbol{\varphi}(\mathbf{y}), \tag{1}$$

where $\boldsymbol{\varphi}$ is a given function from $\mathbb{C}^m$ to $\mathbb{C}^p$, and $\boldsymbol{L}$ is a matrix in $\mathbb{C}^{n\times p}$. Note that the estimator (1) is linear with respect to the matrix $\boldsymbol{L}$, whereas it may be nonlinear in the observation vector.

Our problem is to design the matrix $\boldsymbol{L}$ to be optimal in some sense. To be able to address constrained optimization problems, we will allow the choice of this matrix to be limited by convex constraints. This will be accounted for by imposing that $\boldsymbol{L}$ belongs to a non-empty closed convex set $\mathcal{L}$ of $\mathbb{C}^{n\times p}$.

Assuming that the components of $\mathbf{x}$ and $\boldsymbol{\varphi}(\mathbf{y})$ have finite second-order moments, the celebrated MSE estimation approach is to choose the matrix $\boldsymbol{L} \in \mathcal{L}$ to minimize the MSE

$$J(\boldsymbol{L}) = \mathsf{E}[\|\mathbf{x} - \boldsymbol{L}\boldsymbol{\varphi}(\mathbf{y})\|^2]. \tag{2}$$

where $\|\cdot\|$ denotes the usual euclidean norm and $\mathsf{E}[.]$ designates the expectation.

After some straightforward manipulations, the MSE may be rewritten as

$$\begin{aligned} J(\boldsymbol{L}) &= \mathsf{E}[\|\mathbf{x}\|^2 - 2\mathrm{Re}\{\mathbf{x}^{\mathsf{H}}\boldsymbol{L}\boldsymbol{\varphi}(\mathbf{y})\} + \|\mathbf{y}\|^2] \\ &= \mathrm{tr}(\boldsymbol{\Gamma}_{\mathbf{x}} - 2\mathrm{Re}\{\boldsymbol{L}\boldsymbol{\Gamma}^{\mathsf{H}}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}\} + \boldsymbol{L}\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}\boldsymbol{L}^{\mathsf{H}}) \end{aligned} \tag{3}$$

where $\boldsymbol{A}^{\mathsf{H}}$ is the conjugate of the transpose $\boldsymbol{A}^{\mathsf{T}}$ of a matrix $\boldsymbol{A}$, $\mathrm{tr}(\boldsymbol{B})$ is the trace of a square matrix $\boldsymbol{B}$, $\mathrm{Re}\{.\}$ is the real part of its argument and, $\boldsymbol{\Gamma}_{\mathbf{x}}$, $\boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}$ and $\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}$ are the following correlation matrices:

$$\boldsymbol{\Gamma}_{\mathbf{x}} = \mathsf{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}],\ \boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})} = \mathsf{E}[\mathbf{x}\boldsymbol{\varphi}(\mathbf{y})^{\mathsf{H}}], \quad \boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})} = \mathsf{E}[\boldsymbol{\varphi}(\mathbf{y})\boldsymbol{\varphi}(\mathbf{y})^{\mathsf{H}}]. \tag{4}$$

Denoting by

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{x}} & \boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})} \\ \boldsymbol{\Gamma}^{\mathsf{H}}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})} & \boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})} \end{bmatrix} \tag{5}$$

the correlation matrix of the vector $(\mathbf{x}^{\mathsf{T}}\ \ \boldsymbol{\varphi}(\mathbf{y})^{\mathsf{T}})^{\mathsf{T}}$ which belongs to the space $\mathcal{S}^{n+p}$ of hermitian matrices of size $(n+p)\times(n+p)$, we can rewrite (3) as

$$J(\boldsymbol{L}) = \mathrm{tr}([\boldsymbol{I}_n\ \ -\boldsymbol{L}]\boldsymbol{\Gamma}[\boldsymbol{I}_n\ \ -\boldsymbol{L}]^{\mathsf{H}}) \tag{6}$$

where $\boldsymbol{I}_n$ is the identity matrix of size n.

If $\boldsymbol{\Gamma}$ is known, then we can choose $\boldsymbol{L}$ to minimize the MSE given by (6). In practice however, only partial information is often available on $\boldsymbol{\Gamma}$. This typically arises when the individual correlation matrices are empirically estimated, as we discuss further in Section 5, so that they are known up to some estimation errors.

Now, we will show how these situations can be addressed from a minimax viewpoint.

3. MINIMAX APPROACH

3.1. Constraints on the correlation matrices

The information available on $\boldsymbol{\Gamma}$ is assumed to define convex constraints, so that the matrix belongs to a non-empty closed convex subset $\mathcal{R}$ of $\mathcal{S}^{n+p}$. For convenience, $\mathcal{R}$ will be decomposed as

$$\mathcal{R} = \bigcap_{i=0}^{r} \mathcal{R}_i \tag{7}$$

where $r \in \mathbb{N}^* \setminus \{1\}$ and $\mathcal{R}_i$, $i \in \{0,\ldots,r\}$, are geometrically simpler closed convex sets of $\mathcal{S}^{n+p}$. For analytical tractability, we will assume that $\mathcal{R}$ is a compact set. As $\mathcal{R}$ is closed this amounts to saying that

$$\mathcal{R}_0 = \{\boldsymbol{A} \in \mathcal{S}^{n+p}\ \mid\ \|\boldsymbol{A}\|_{\mathrm{F}} \le \rho\} \tag{8}$$

for some $\rho > 0$. This assumption is not restrictive since, in practice, we always have (possibly large) upper bounds on the second-order moments of the components of $\mathbf{x}$ and $\boldsymbol{\varphi}(\mathbf{y})$, respectively.

Remind that a matrix is a correlation matrix if and only if it belongs to the closed convex cone $\mathcal{S}^{n+p}_{+}$ of semi-definite positive matrices. This constraint is taken into account by defining $\mathcal{R}_1$ to be equal to

$$\mathcal{R}_1 = \{\boldsymbol{A} \in \mathcal{S}^{n+p}\ \mid\ \boldsymbol{A} \succeq \epsilon \boldsymbol{I}_{n+p}\} \tag{9}$$

where ϵ is a small positive constant which is introduced to avoid singular correlation matrices. Note however that the set $\mathcal{S}^{n+p}_{+}$ can be approximated as closely as desired by choosing ϵ small enough. Additional constraints on $\boldsymbol{\Gamma}$ are related to the available information in the given estimation problem. The case of a random signal corrupted by Gaussian noise is discussed in Section 5.

3.2. Minimax formulation

Since $\boldsymbol{\Gamma}$ is not known exactly, a robust strategy for signal estimation consists in solving the following minimax optimization problem:

$$\inf_{\boldsymbol{L}\in\mathcal{L}}\ \sup_{\boldsymbol{\Gamma}\in\mathcal{R}} f(\boldsymbol{L},\boldsymbol{\Gamma}), \tag{10}$$

where

$$f(\boldsymbol{L},\boldsymbol{\Gamma}) = \mathrm{tr}([\boldsymbol{I}_n\ \ -\boldsymbol{L}]\boldsymbol{\Gamma}[\boldsymbol{I}_n\ \ -\boldsymbol{L}]^{\mathsf{H}}) \tag{11}$$

is the MSE between the complex-valued random vectors $\mathbf{x}$ and $\boldsymbol{L}\boldsymbol{\varphi}(\mathbf{y})$. In other words, we want to find the matrix $\boldsymbol{L}$ that minimizes the MSE in the worst case compatible with the information available on the correlation matrix.

To develop a solution to (10) we invoke the minimax theorem (also called Sion's theorem) that will allow us to replace the minimax problem by a maximin problem which in some cases can lead to more efficient algorithms.

Theorem 1 *The MSE function f defined in (11) possesses a saddle point on $\mathcal{R}\times\mathcal{L}$ and, consequently,*

$$\inf_{\boldsymbol{L}\in\mathcal{L}}\ \sup_{\boldsymbol{\Gamma}\in\mathcal{R}} f(\boldsymbol{L},\boldsymbol{\Gamma}) = \sup_{\boldsymbol{\Gamma}\in\mathcal{R}}\ \inf_{\boldsymbol{L}\in\mathcal{L}} f(\boldsymbol{L},\boldsymbol{\Gamma}). \tag{12}$$

In addition, let g be the convex function defined by

$$\forall \boldsymbol{\Gamma} \succeq \mathbf{0}, \qquad g(\boldsymbol{\Gamma}) = -\inf_{\boldsymbol{L} \in \mathcal{L}} f(\boldsymbol{L}, \boldsymbol{\Gamma}), \tag{13}$$

and let $\boldsymbol{\Gamma}_{\rm o}$ be a minimizer of g on $\mathcal{R}$. The unique optimal minimax solution is $\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma}_{\rm o})$ where, for any given $\boldsymbol{\Gamma} \succ \mathbf{0}$, $\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma})$ is the unique minimizer of $\boldsymbol{L} \mapsto f(\boldsymbol{L}, \boldsymbol{\Gamma})$ on $\mathcal{L}$.

Note that the optimization problem involved in (13) for the determination of g is the minimization of a simple convex quadratic function subject to convex constraints. More precisely, consider the weighted Frobenius norm

$$\forall \boldsymbol{A} \in \mathbb{C}^{n \times p}, \qquad \|\boldsymbol{A}\|_{\boldsymbol{\Gamma}} = \Big(\mathrm{tr}(\boldsymbol{A} \boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})} \boldsymbol{A}^{\mathsf{H}})\Big)^{1/2}. \tag{14}$$

We have:

Proposition 1 *The function $\boldsymbol{L} \mapsto f(\boldsymbol{L}, \boldsymbol{\Gamma})$ is minimum on $\mathcal{L}$ at $\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma})$ which is the projection in the space $\big(\mathbb{C}^{n\times p}, \|\cdot\|_{\boldsymbol{\Gamma}}\big)$ of the matrix $\boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}^{-1}$ onto the closed convex set $\mathcal{L}$. Furthermore,*

$$g(\boldsymbol{\Gamma}) = -\|\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma}) - \boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}^{-1}\|_{\boldsymbol{\Gamma}}^2 + \mathrm{tr}(\boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}^{-1}\boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}^{\mathsf{H}} - \boldsymbol{\Gamma}_{\mathbf{x}}). \tag{15}$$

In the special case when $\mathcal{L} = \mathbb{C}^{n\times p}$, the classical Wiener solution is obtained:

$$\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma}) = \boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}^{-1} \tag{16}$$

and $-g(\boldsymbol{\Gamma}) = \mathrm{tr}(\boldsymbol{\Gamma}_{\mathbf{x}} - \boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}\boldsymbol{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}^{-1}\boldsymbol{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}^{\mathsf{H}})$ reduces to the trace of the Schur complement of the matrix $\boldsymbol{\Gamma}$.

Even if an explicit expression of $\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma})$ cannot be found, efficient algorithms can be used to compute the desired projection [9]. These algorithms are applicable whenever $\mathcal{L} = \cap_{i=1}^{\ell}\mathcal{L}_i$ where $\ell \in \mathbb{N}^*$ and, for all $i \in \{1,\dots,\ell\}$, $\mathcal{L}_i$ is a non-empty closed convex set whose projection (or a subgradient projection) onto can be easily calculated.

3.3. Some properties of function g

Even in the unconstrained case when $\mathcal{L} = \mathbb{C}^{n\times p}$, we see that g does not appear as a simple convex function such as a linear or a quadratic function. As $\boldsymbol{\Gamma} \succ \mathbf{0}$, the minimization of g over $\mathcal{R}$ is a nonlinear semi-definite optimization problem. In order to proceed further, we first consider a basic property of the function g defined by (13).

From (13), g can be re-expressed as

$$\forall \boldsymbol{\Gamma} \succ \mathbf{0}, \qquad g(\boldsymbol{\Gamma}) = \sup_{\boldsymbol{K} \in \boldsymbol{S}(\mathcal{L})} \mathrm{Re}\{\mathrm{tr}(\boldsymbol{\Gamma}\boldsymbol{K})\}. \tag{17}$$

where $\boldsymbol{S}(\mathcal{L})$ is the image of the set $\mathcal{L}$ by the application

$$\begin{aligned} \boldsymbol{S} \ : \ \mathbb{C}^{n\times p} &\longrightarrow \mathcal{S}^{n+p} \\ \boldsymbol{L} &\longmapsto \begin{bmatrix} -\boldsymbol{I}_n & \boldsymbol{L} \\ \boldsymbol{L}^{\mathsf{H}} & -\boldsymbol{L}^{\mathsf{H}}\boldsymbol{L}\end{bmatrix}. \end{aligned} \tag{18}$$

So, if $\mathcal{S}^{n+p}$ is endowed with the real part of the usual Frobenius scalar product, g is the so-called support function of the set $\boldsymbol{S}(\mathcal{L})$. We deduce the following property:

Proposition 2 *The function g is differentiable at any $\boldsymbol{\Gamma} \succ \mathbf{0}$ and its gradient is:*

$$\boldsymbol{\nabla} g(\boldsymbol{\Gamma}) = \boldsymbol{S}\big(\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma})\big), \tag{19}$$

where $\boldsymbol{S}$ is the function defined by (18) and $\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma})$ is the minimizer of $f(\boldsymbol{L}, \boldsymbol{\Gamma})$ on $\mathcal{L} \times \mathcal{R}$ as given by (11).

4. OPTIMIZATION ALGORITHM

We now treat the problem of minimizing g. The differentiability of the function g allows the use of a projected gradient algorithm. More precisely, we propose to employ a version of this algorithm involving optimized step-sizes [10]. This can be achieved by building a sequence $(\boldsymbol{\Gamma}^{(i)})_{i\in\mathbb{N}}$ of matrices such that

$$\widetilde{\boldsymbol{\Gamma}}^{(i)} = \mathsf{P}_{\mathcal{R}}[\boldsymbol{\Gamma}^{(i)} - \beta_i\, \boldsymbol{S}\big(\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma}^{(i)})\big)] \tag{20}$$

$$\boldsymbol{\Gamma}^{(i+1)} = \boldsymbol{\Gamma}^{(i)} + \gamma_i(\widetilde{\boldsymbol{\Gamma}}^{(i)} - \boldsymbol{\Gamma}^{(i)}) \tag{21}$$

where $\mathsf{P}_{\mathcal{R}}$ is the projection onto $\mathcal{R}$ and $(\beta_i)_{i\in\mathbb{N}}$ and $(\gamma_i)_{i\in\mathbb{N}}$ are positive step-size sequences. Several variants of the projected gradient algorithm are possible depending on the choice of these step-sizes. In the following, we propose to choose $\gamma_i = 2^{-\hat{\jmath}(i)}$ where $\hat{\jmath}(i)$ is an integer computed through an Armijo search. The algorithm takes the following form:

Algorithm 1

① *Fix $(\underline{\beta}, \overline{\beta}) \in \mathbb{R}^2$ with $0 < \underline{\beta} \le \overline{\beta}$ and $\zeta \in (0,1)$. Set $i = 0$ and choose an initial value $\boldsymbol{\Gamma}^{(0)}$ arbitrarily in $\mathcal{R}$.*

② *Compute $\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma}^{(i)})$.*

③ *Choose $\beta_i \in [\underline{\beta}, \overline{\beta}]$ and calculate $\widetilde{\boldsymbol{\Gamma}}^{(i)}$ using (20).*

④ *Determine $\hat{\jmath}(i) = \min\{j \in \mathbb{N} \mid g(\boldsymbol{\Gamma}^{(i,j)}) \le g(\boldsymbol{\Gamma}^{(i)}) - 2^{-j}\zeta\, \langle \boldsymbol{S}\big(\boldsymbol{L}_{\rm o}(\boldsymbol{\Gamma}^{(i)})\big), \boldsymbol{\Gamma}^{(i)} - \widetilde{\boldsymbol{\Gamma}}^{(i)}\rangle\}$ where, for all $j \in \mathbb{N}$, $\boldsymbol{\Gamma}^{(i,j)} = \boldsymbol{\Gamma}^{(i)} + 2^{-j}(\widetilde{\boldsymbol{\Gamma}}^{(i)} - \boldsymbol{\Gamma}^{(i)})$.*

⑤ *Set $\boldsymbol{\Gamma}^{(i+1)} = \boldsymbol{\Gamma}^{(i,\hat{\jmath}(i))}$.*

⑥ *If $\boldsymbol{\Gamma}^{(i+1)} = \boldsymbol{\Gamma}^{(i)}$, stop.*

⑦ *Set $i \leftarrow i+1$ and go to ②.*

The convergence of the algorithm is secured according to [10, Thm. 1].

An important step in the above algorithm is the computation of the projection onto $\mathcal{R}$, when this one cannot be explicitly calculated. To do this, the efficient block-iterative surrogate constraint splitting method which has been recently developed in [11] can be used, when $\mathcal{R}$ can be decomposed as in (7).

5. NONLINEAR DENOISING IN THE PRESENCE OF GAUSSIAN NOISE

An additive Gaussian model is obtained when the observation equation reads

$$\mathbf{y} = \mathbf{x} + \mathbf{u} \tag{22}$$

where $\mathbf{u}$ is a zero-mean Gaussian vector which is independent of $\mathbf{x}$. For simplification, in this section we will consider that $\mathbf{x}$ and $\mathbf{u}$ are real-valued. The covariance matrix $\mathbf{\Gamma_u} \in \mathcal{S}_+^m$ of the noise is assumed to be known. In this context, the interest of using a nonlinear estimate stems from the possible non-Gaussianity of $\mathbf{x}$. Note that, as already pointed out in Section 1, for denoising problems the use nonlinear techniques is very popular.

In the above model, an obvious relation between the second-order moments of the observed vector $\mathbf{y}$ and the unknown vector $\mathbf{x}$ is:

$$\mathbf{\Gamma_y} = \mathbf{\Gamma_x} + \mathbf{\Gamma_u}. \tag{23}$$

We assume that some conditions hold on $\boldsymbol{\varphi}$, namely:

- $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^\mathsf{T} \mapsto \boldsymbol{\varphi}(\boldsymbol{\theta})$ is continuous. Besides, for all $\boldsymbol{\theta}$, it is piecewise differentiable w.r.t. each θ_i, $i \in \{1, \ldots, m\}$, and

  $$\forall \boldsymbol{\mu} \in \mathbb{R}^m, \qquad \lim_{\|\boldsymbol{\theta}\| \to \infty} \boldsymbol{\varphi}(\boldsymbol{\theta}) \, \gamma_{\mathbf{\Gamma_u}}(\boldsymbol{\theta} - \boldsymbol{\mu}) = \mathbf{0}.$$

  where $\gamma_{\mathbf{\Gamma_u}}$ denotes the probability density function of a zero-mean Gaussian vector with covariance matrix $\mathbf{\Gamma_u}$;

- $\mathsf{E}[\|\boldsymbol{\varphi}(\mathbf{y})\|^2] < \infty$ and $\mathsf{E}[\|\boldsymbol{\nabla}\boldsymbol{\varphi}(\mathbf{y})\|_\mathrm{F}] < \infty$ where $\boldsymbol{\nabla}\boldsymbol{\varphi}(\mathbf{y})$ is the Jacobian matrix of $\boldsymbol{\varphi}$.

Then, Stein's principle [5] allows us to relate $\mathbf{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})}$ to the expectation of some functions of the observations only. More precisely, we have

$$\mathbf{\Gamma}_{\mathbf{y},\boldsymbol{\varphi}(\mathbf{y})} = \mathbf{\Gamma}_{\mathbf{x},\boldsymbol{\varphi}(\mathbf{y})} + \mathbf{\Gamma_u} \boldsymbol{G}_{\boldsymbol{\varphi}(\mathbf{y})} \tag{24}$$

where $\boldsymbol{G}_{\boldsymbol{\varphi}(\mathbf{y})} = \mathsf{E}[\boldsymbol{\nabla}\boldsymbol{\varphi}(\mathbf{y})]$.

If $k \in \mathbb{N}^*$ realizations $\mathbf{y}_1, \ldots, \mathbf{y}_k$ of $\mathbf{y}$ are available (under weak conditions), empirical estimates of $\mathbf{\Gamma_y}$, $\mathbf{\Gamma}_{\mathbf{y},\boldsymbol{\varphi}(\mathbf{y})}$, $\boldsymbol{G}_{\boldsymbol{\varphi}(\mathbf{y})}$ and $\mathbf{\Gamma}_{\boldsymbol{\varphi}(\mathbf{y})}$ can be computed. For example, we can use the following consistent estimates:

$$\widehat{\mathbf{\Gamma}}_{\mathbf{y}}^{(k)} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{y}_i \mathbf{y}_i^\mathsf{T} \tag{25}$$

$$\widehat{\mathbf{\Gamma}}_{\mathbf{y},\boldsymbol{\varphi}(\mathbf{y})}^{(k)} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{y}_i \boldsymbol{\varphi}(\mathbf{y}_i)^\mathsf{T} \tag{26}$$

$$\widehat{\boldsymbol{G}}_{\boldsymbol{\varphi}(\mathbf{y})}^{(k)} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{\nabla}\boldsymbol{\varphi}(\mathbf{y}_i) \tag{27}$$

$$\widehat{\mathbf{\Gamma}}_{\boldsymbol{\varphi}(\mathbf{y})}^{(k)} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{\varphi}(\mathbf{y}_i) \boldsymbol{\varphi}(\mathbf{y}_i)^\mathsf{T}. \tag{28}$$

Due to the finite sample size, these estimations are affected by errors. Consequently, one can impose the following constraints on the correlation matrix $\mathbf{\Gamma}$:

$$\left\| \mathbf{\Gamma} - \begin{bmatrix} \widehat{\mathbf{\Gamma}}_{\mathbf{y}}^{(k)} - \mathbf{\Gamma_u} & \widehat{\mathbf{\Gamma}}_{\mathbf{y},\boldsymbol{\varphi}(\mathbf{y})}^{(k)} - \mathbf{\Gamma_u} \widehat{\boldsymbol{G}}_{\boldsymbol{\varphi}(\mathbf{y})}^{(k)} \\ (\widehat{\mathbf{\Gamma}}_{\mathbf{y},\boldsymbol{\varphi}(\mathbf{y})}^{(k)} - \mathbf{\Gamma_u} \widehat{\boldsymbol{G}}_{\boldsymbol{\varphi}(\mathbf{y})}^{(k)})^\mathsf{T} & \widehat{\mathbf{\Gamma}}_{\boldsymbol{\varphi}(\mathbf{y})}^{(k)} \end{bmatrix} \right\|_a \leq \eta_a^{(k)} \tag{29}$$

where $\| \cdot \|_a$ is some semi-norm and $\eta_a^{(k)}$ is a confidence bound on the estimation of of the matrix of interest. From standard results in statistics, for a given probabilistic confidence level, this bound typically decreases as $k^{-1/2}$, when $k \to \infty$.

The equation in (29) defines a closed a closed convex set in $\mathcal{S}^{n+p}$ which can be denoted by $\mathcal{R}_2$. In addition to the sets $\mathcal{R}_0$ and $\mathcal{R}_1$ defined in Section 3.1 and some possible additional constraints on the structure of the matrix $\mathbf{\Gamma}$, they define the set $\mathcal{R}$ for the considered denoising problem. Then, the proposed minimax approach allows us to compute the estimator minimizing the MSE for the worst case correlation matrix complying with these constraints.

6. CONCLUSION

In this paper, we have proposed a framework for minimax nonlinear estimation of random vectors using a MSE criterion. The equivalence between the minimax and maximim solutions has been established. An optimization algorithm has been derived allowing us to cope with a wide class of convex constraints on the correlation matrices of interest. Although the proposed approach is applicable to many statistical signal processing problems such as forecasting of nonlinear time series or identification of nonlinear models, we have briefly mentioned how it can be used for minimax signal denoiling problems.

7. REFERENCES

[1] Y. C. Eldar and N. Merhav, ”A Competitive Minimax Approach to Robust Estimation of Random Parameters,” *IEEE Trans. Signal Processing*, vol. 52, pp. 1931-1946, July 2004.

[2] Y. C. Eldar, A. Ben-Tal and A. Nemirovski, ”Linear Minimax Regret Estimation of Deterministic Parameters with Bounded Data Uncertainties,” *IEEE Trans. Signal Processing*, vol. 52, pp. 2177-2188, Aug. 2004.

[3] Y. C. Eldar, A. Ben-Tal and A. Nemirovski, ”Robust Mean Squared Error Estimation in the Presence of Model Uncertainties,” *IEEE Trans. Signal Processing*, vol. 53, pp.168-181, Jan. 2005.

[4] D.L. Donoho, I.M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425-455, 1994.

[5] C. Stein, “Estimation of the mean of a multivariate normal distribution,” *Annals of Statistics*, vol. 9, no. 6, pp. 1135-1151, 1981.

[6] D.L. Donoho, I.M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200-1224, December 1995.

[7] A. Benazza-Benyahia and J.-C. Pesquet, “Building robust wavelet estimators for multicomponent images Using Stein’s principle,” accepted for publication in *IEEE Trans. on Image Processing*, 2005.

[8] P.L. Combettes and J.-C. Pesquet, “Wavelet-constrained image restoration,” *International Journal on Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 371-389, Dec. 2004.

[9] P. L. Combettes, “The convex feasibility problem in image recovery,” in *Advances in Imaging and Electron Physics*, vol. 95, pp. 155-270. New York: Academic, 1996.

[10] A.N. Iusem, “On the convergence properties of the projected gradient method for convex optimization,” *Mat. Apl. Comput.*, vol. 22, no. 1, p. 37-52, 2003.

[11] P. L. Combettes, “A block-iterative surrogate constraint splitting method for quadratic signal recovery,” *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1771-1782, July 2003.