16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

REVISITING THE DENOISING PROBLEM IN THE CONTEXT OF ELLIPTICAL DISTRIBUTIONS

S. Zozor [1] *and C. Vignat* [2]

[1] GIPSA-Lab, Departement Image et Signal, 961 Rue de la Houille Blanche, 38402 St Martin d'Hères, France
email: steeve.zozor@gipsa-lab.inpg.fr

[2] Institut Gaspard Monge, Université de Marne-la-Vallée 5 Boulevard Descartes, 77454 Marne-la-Vallée, France
email: vignat@univ-mlv.fr

ABSTRACT

In this paper, we revisit the denoising problem in the context of elliptical distributions; this approach extends a work by Alecu *et al.* [1] in two directions: (i) we address the multivariate case and (ii) our study is not restricted to the scale mixtures of Gaussians, but extends to the whole family of elliptical distributions. We consider the framework where both the vector to be estimated and the additive independent corrupting noise are elliptically distributed. The MMSE and Wiener filters are computed and their performances are compared in terms of the value of the SNR in several illustrations.

1. INTRODUCTION

Recently, Alecu *et al.* have introduced the notion of "Gaussian transform" of a random variable with symmetric probability density function (pdf) [1]. They deal in fact with random variables that are known since decades as scale mixtures of Gaussians [2, 3, 4, 5, 6, 7]. Such a variable can be stochastically expressed as the product of a positive random variable A, by an independent Gaussian random variable N, namely $X \stackrel{d}{=} AN$, where $\stackrel{d}{=}$ stands for the equality in distribution sense. In image processing, this kind of random variable is widely used for modeling images: in this context, variable A^2 generally models the texture of images. The so-called "Gaussian transform" of the pdf of X as defined in [1] is in fact nothing but the pdf of the square A^2 of the mixing variable A.

In the scale mixture of Gaussians context, Alecu *et al.* proposed to apply their concept of Gaussian transform to the denoising problem: from an observation $Y = X + Z$, one wishes to estimate random variable X, assumed scale mixture of Gaussians of known mixing disribution, where Z is an additive Gaussian noise with known variance and independent of X . In this paper, we propose to revisit this problem in the more general d-dimensional case, where both X and Z are scale mixtures of Gaussians. More generally, we consider the problem where X and Z are elliptically distributed: this class of random vectors contains the class of scale mixtures of Gaussians, but it is wider since there exists elliptically distributed random vectors that are not scale mixtures of Gaussians (multivariate Pearson type II for example).

In section 2 we briefly recall some basics about elliptically invariant random vectors. Then, section 3 is devoted to the denoising problem in such a context. In this section we will derive the minimum mean-squared error (MMSE) estimator in the elliptically distributed framework. Moreover, we will show that under an additional assumption on the covariance matrices of X and Z, the MMSE estimator simplifies to a one dimensional estimator. Finally, the shape of the MMSE estimator and its performance are exhibited in several illustrations and for various values of the dimension d. They are compared to those of the best linear estimator. Some conclusions are drawn from these examples, and comments on Alecu's results are provided [1].

2. BASICS ON ELLIPTICALLY DISTRIBUTED VECTORS

A d-dimensional random vector X is elliptically distributed if its probability density function p_x is a function of the quadratic form $(x-\mu)^t R^{-1}(x-\mu)$, that is

$$p_x(x) = |R|^{-1/2} d_x((x-\mu)^t R^{-1}(x-\mu))$$

where d_x is a function $\mathbb{R}_+ \to \mathbb{R}_+$ and R a symmetric definite positive matrix [4, 8, 9, 10, 11]. Matrix R is called characteristic matrix of distribution p_x and μ is a location parameter. When defined, these quantities are respectively the covariance matrix and the mean of X. When $R \propto I$, the $(d \times d)$ identity matrix, X is said spherically or orthogonally invariant (or distributed).

In the case where function $s \mapsto d_x(s)$ admits an inverse Laplace transform (see [9]), the pdf p_x can be expressed as

$$p_x(x) = \int_0^{+\infty} f_a(a)\, \mathcal{N}(x-\mu|_{a^2R})\, da$$

where $\mathcal{N}(\cdot|_R)$ denotes the d-dimensional zero-mean Gaussian pdf with covariance matrix R. Integrating p_x over x shows that, provided that integration signs can be exchanged, function f_a sums to 1. Moreover, under the condition that d_x is an absolutely monotone function, function f_a is nonnegative [10, 11, 12, 13, 14] and hence is a pdf. In other words, there exists a positive scalar random variable[1] $A \sim f_a \equiv p_a$ such that X has the stochastic representation $X \stackrel{d}{=} AN$ where $N \sim p_N(x) = \mathcal{N}(x-\mu|_R)$ is independent of A, and equality is in the sense of distributions: vector X is then called a scale mixture of Gaussians. In the following, we will denote by p a pdf, while f will represent a mixing function, that can be a pdf ($f \geq 0$) or not. We note that matrix R is defined up to a constant scaling factor that can be included in A; choosing R as the covariance matrix of X removes this indeterminacy and imposes that $\int_0^{+\infty} a^2 f_a(a)da = 1$ (whether $f_a \geq 0$ or not).

Several properties of scale mixtures of Gaussians or, more generally, of elliptical distributions can be found in the above cited papers. Note however that the stability property mentioned in [1, Property 5] holds only in the scalar context. Indeed, in the d-dimensional context, the sum of two

[1] $\sim$ means "distributed according to"

independent elliptically distributed random vectors X and Y remains elliptically distributed if and only if either their covariance matrix R_x and R_y are proportional ($R_x \propto R_y$) or they are both Gaussian. Indeed, one can show that the Fourier transform of an elliptical pdf with characteristic matrix R is elliptical with characteristic matrix R^{-1} [8]. Hence, the characteristic function $\Phi_{x+y}(u) = E[e^{iu^t(X+Y)}]$ writes $\Phi_{x+y}(u) = \Phi_x(u)\Phi_y(u) = \phi_x(u^tR_xu)\phi_y(u^tR_yu)$. The only possibility for $\phi_x(u^tR_xu)\phi_y(u^tR_yu)$ to be of the form $\phi(u^tRu)$ is that either $R_x \propto R_y$, or both ϕ_x and ϕ_y are exponential. In both cases, obviously $R = R_x + R_y$ – which is the covariance matrix of $X+Y$ in any case – is the characteristic matrix of the elliptically distributed vector $X+Y$. In the case of scale mixture of Gaussians with $R_x = \alpha R_y$ and with the mixing random variable of X and Y equal to $A \sim f_a \equiv p_a$ and $B \sim f_b \equiv p_b$ respectively, the mixing variable is given by $C = \sqrt{A^2 + \alpha^2 B^2} \sim f_c \equiv p_c$. The relation between f_a, f_b and f_c remains true even if f_a or f_b is not a pdf.

In the rest of the paper, we focus without loss of generality on zero-mean random vectors, *i.e.* $\mu = 0$, and restrict our study to vectors that admit a covariance matrix (*e.g.* we exclude cases such as Cauchy distributed vectors). In the next section, we turn now to the revisit of the denoising problem in the here-introduced framework.

3. DENOISING IN THE ELLIPTICAL CONTEXT

Let us consider the estimation problem of a d-dimensional random vector X with elliptical distribution and covariance matrix R_x, from a noisy observation

$$Y = X + Z \tag{1}$$

where X and Z are assumed independent. Let us moreover assume that the observation noise Z is also elliptically distributed with covariance matrix R_z. We know that pdfs p_x and p_z of X and Z can be written respectively as

$$\begin{aligned} p_x(u) &= \int_0^{+\infty} f_a(a)\mathcal{N}(u|_{a^2R_x})da \quad \text{and} \\ p_z(u) &= \int_0^{+\infty} f_b(b)\mathcal{N}(u|_{b^2R_z})db, \end{aligned}$$

where f_a and f_b are supposed known. Remember that the mixing (or weighting) functions f_a and f_b can be negative (see [9]). Finally, we also assume that both covariance matrices R_x and R_z are known. We deal with the same context as in [1], but generalized to any dimension d and to the general elliptical framework.

Let us define the symmetric definite positive matrix

$$R = R_x^{-\frac{1}{2}} R_z R_x^{-\frac{1}{2}}$$

where $R_x^{\frac{1}{2}}$ denote the (unique) symmetric definite positive square root of the symmetric definite positive matrix R_x [15, th. 7.2.6], and write the eigenvalue factorization of R as

$$R = V\Delta V^t$$

where Δ is a diagonal matrix of the eigenvalues of R and where V is the orthogonal matrix of the corresponding eigenvectors. The denoising problem can then be simplified by multiplying model (1) by $V^tR_x^{-1/2}$ on the left:

$$Y_0 = X_0 + Z_0 \tag{2}$$

with $\{X,Y,Z\}_0 = V^tR_x^{-\frac{1}{2}}\{X,Y,Z\}$. It is immediate that X_0 has identity covariance matrix and that Z_0 has covariance matrix Δ. Moreover, vectors X_0 and Z_0 remain independent and since p_x is orthogonally invariant, X_0 has the same mixing function/distribution f_a as X. Similarly Z_0 has the same mixing function f_b as Z. This is obvious when dealing with scale mixtures of Gaussians vectors : $X_0 = V^tR_x^{-\frac{1}{2}}X \overset{d}{=} AV^tR_x^{-\frac{1}{2}}N$ and $V^tR_x^{-\frac{1}{2}}N \overset{d}{=} N_0 \sim \mathcal{N}(\cdot|_I)$ and similarly $Z_0 \overset{d}{=} BN_0'$ with $N_0' \sim \mathcal{N}(\cdot|_\Delta)$.

As a conclusion, we can assume without loss of generality in the following that $R_x = I$ and $R_z = \Delta$.

3.1 Minimum Mean Square Estimation

The well-known Minimum Mean Square Error (MMSE) estimator of X based on the observation Y, *i.e.* the vector $\widehat{X}$ that minimizes the quadratic error $E[\|\widehat{X} - X\|^2]$, is given by the conditional mean [16]

$$\widehat{X}_{\text{mmse}} = E[X|Y].$$

In the rest of the paper, $\widehat{X}_{\text{mmse}}$ will denote $\widehat{X}_{\text{mmse}}(y) = E[X|Y=y]$. Using Bayes' rule, and remarking that, by the independence assumption, $p_{y|x}(y,x) = p_z(y-x)$, this expression can be equivalently written as

$$\widehat{X}_{\text{mmse}} = \int_{\mathbb{R}^d} x p_{x|y}(x,y)dx = \frac{\int x p_z(y-x)p_x(x)\,dx}{p_y(y)} \tag{3}$$

Both pdfs p_x and p_z can be expressed via their mixing functions f_a and f_b to achieve

$$\begin{aligned} \widehat{X}_{\text{mmse}} &= \frac{1}{p_y(y)}\int_{\mathbb{R}^d} x \times \\ &\quad \int_0^{+\infty} \mathcal{N}(y-x|_{b^2\Delta})f_b(b)\,db \int_0^{+\infty} \mathcal{N}(x|_{a^2I})f_a(a)\,da\,dx \\ &= \frac{1}{p_y(y)}\int_{\mathbb{R}_+^2} \mathcal{N}(y|_{a^2I+b^2\Delta})f_a(a)f_b(b) \times \\ &\quad \frac{\int_{\mathbb{R}^d} x\mathcal{N}(y-x|_{b^2\Delta})\mathcal{N}(x|_{a^2I})\,dx}{\mathcal{N}(y|_{a^2I+b^2\Delta})}\,da\,db \end{aligned}$$

Now, the fraction of the last line is recognized as the MMSE estimator of a Gaussian vector N_1 with covariance matrix a^2I embedded in an independent Gaussian noise N_2 with covariance matrix $b^2\Delta$: the best estimator of N_1 is thus linear and by the Wiener theory [16], it expresses as $a^2I(a^2I+b^2\Delta)^{-1}y$. Replacing this expression in $\widehat{X}_{\text{mmse}}$ and expressing p_y as the convolution of p_x and p_z, namely

$$\begin{aligned} p_y(y) &= \int p_z(y-x)p_x(x)\,dx \\ &= \int_{\mathbb{R}_+^2}\left(\int \mathcal{N}(y-x|_{b^2\Delta})\mathcal{N}(x|_{a^2I})\,dx\right) f_a(a)f_b(b)\,da\,db \end{aligned}$$

we finally achieve the following expression for the MMSE estimator:

$$\widehat{X}_{\text{mmse}} = \frac{\int_{\mathbb{R}_+^2} \mathcal{N}(y|_{a^2I+b^2\Delta})\,a^2 f_a(a)f_b(b)(a^2I+b^2\Delta)^{-1}\,da\,db}{\int_{\mathbb{R}_+^2} \mathcal{N}(y|_{a^2I+b^2\Delta})f_a(a)f_b(b)\,da\,db}\,y \tag{4}$$

which makes use of integration over $\mathbb{R}_+^2$ instead of the integration over $\mathbb{R}^d$ required in (3). Again, p_y can be viewed as $\int p_{y,a,b}(y,a,b)\,da\,db = \int p_{y|a,b}(y,a,b) f_a(a) f_b(b)\,da\,db$ in the case of scale mixture of Gaussians, but this expression still holds if f_a or f_b take negative values.

When dimension $d = 1$ and when Z is Gaussian ($f_b(b) = \delta(b-1)$), this result coincides with that of Alecu [1, eq. (34)], but, again, the nonnegativity assumption of f_a is not required here and the expression is extended to the d-dimensional elliptically distributed context. As for the estimator proposed by Alecu, numerical integration is generally needed. But the numerical evaluation is easy since the matrix inversion and the determinant evaluation involve low complexity computations.

Note also that Chu [9] addressed a similar problem, but under the assumptions that the concatenated vector $[X\ Z]$ is elliptically distributed and that random vectors X and Z are uncorrelated. In this particular context, the solution $\widehat{X}_w$ is given by the Wiener theory

$$\widehat{X}_w = (I+\Delta^{-1})^{-1}\Delta^{-1} y = (I+\Delta)^{-1} y \qquad (5)$$

Since the two parts X and Z of an elliptical vector $[X\ Z]$ can be independent only in the case where X and Z are Gaussian, we conclude that estimator (4) and Chu's estimator (5) coincide only in the Gaussian case. Moreover, the Wiener estimator is the best linear estimator in the MSE sense whatever X and Z, provided they are uncorrelated.

In the elliptical framework considered here, an interesting case arises when $\Delta = \sigma^2 I$. In this case, both Y and $\widehat{X}_{\text{mmse}}(Y)$ are spherically distributed. Immediately from (4), for any rotation (or orthogonal) matrix C_θ, one has $C_\theta^t \widehat{X}_{\text{mmse}}(C_\theta y) = \widehat{X}_{\text{mmse}}(y)$. Now, if C_θ is so that $C_\theta y = [y_0\ 0\ \ldots\ 0]^t$, then $\widehat{X}_{\text{mmse}}(y) = C_\theta^t[\overline{X}_{\text{mmse}}(y_0)\ 0\ \ldots\ 0]^t$ where $\overline{X}_{\text{mmse}}(y_0)$ is a scalar function under the form (4), in which $\mathcal{N}(y|_{a^2+\sigma^2 b^2})$ is replaced by $(a^2+\sigma^2 b^2)^{-\frac{d-1}{2}}\mathcal{N}(y|_{a^2+\sigma^2 b^2})$. The $2d$ double integrations required in (4) reduce to only 2 double integrations, *i.e.* to the computation load involved in the one-dimensional context, at the extra cost of two rotations (one applied to vector y and one to the estimator).

The following subsection aims to illustrate the performance of the MMSE estimator as a function of the signal-to-noise ratio (SNR). This performance will be compared to those of the Wiener estimator, and some comments will be drawn in regards to suboptimal approaches that can be found in the literature.

3.2 Illustrative examples

In the following, we restrict our attention to the case where $\Delta = \sigma^2 I$ and where the corrupting noise Z is Gaussian. The performance in terms of Mean Square Error (MSE) of both estimators $\widehat{X}_{\text{mmse}}$ and $\widehat{X}_w$ will be shown versus the signal-to-noise ratio [2] $SNR = \frac{E[\|X\|^2]}{\|Z\|^2} = \frac{d}{\text{Tr}(\Delta)} = \sigma^{-2}$ and for several values of the dimension d. Writing for any estimator $\widehat{X} = \widehat{X}(y)$ the Mean Square Error as $MSE = E_Y[E[\|\widehat{X}-X\|^2|Y]]$, one easily achieves

$$MSE = d + \int_{\mathbb{R}^d}\|\widehat{X}\|^2 p_y(y)\,dy - 2\int_{\mathbb{R}^d}\widehat{X}^t\widehat{X}_{\text{mmse}}\, p_y(y)\,dy \quad (6)$$

This error has no analytical expression in the general case, except in the Gaussian case and dealing with a linear estimator (e.g. Wiener). However, in the particular case when (i) $\Delta = \sigma^2 I$, (ii) noise Z is Gaussian and (iii) the estimator is of the form $\widehat{X} = \widehat{f}(y^t y)\,y$ (as this is the case for both MMSE and Wiener estimators), both integrands in the MSE reduce to scalar functions of $y^t y$ and, from [17, 4.642], the computation of (6) simplifies to an integration over $\mathbb{R}_+$, what makes the numerical integration tractable. In all figures below, the theoretical results are depicted, using respectively (4), (5) and (6).

The first case deals with a vector X following an exponential power distribution $X \sim p_x(x) \propto e^{-(\gamma x^t x)^{\frac{p}{2}}}$ where $\gamma = \frac{\Gamma\left(\frac{d+2}{p}\right)}{d\,\Gamma\left(\frac{d}{p}\right)}$. We limit here our investigation to the case where $0 < p < 2$. For $d = 1$, this class of distributions is that studied by Alecu *et al.* and generally known as p-generalized Gaussian distributions [1]. In the general d-dimensional case, one can show that the mixing function is $f_a(a) \propto a^{d-3}\,\mathscr{P}_{\frac{p}{2}}\left(a^{-2}(2\gamma)^{-1}\cos^{-\frac{2}{p}}(p\pi/4)\right)$ where $\mathscr{P}_\alpha$ is the pdf of an α-stable variable of stability index α and totally skewed to the right (with skew parameter $\beta = 1$) [18]. Except in the particular case of the Lévy distribution given by $\alpha = \frac{1}{2}$, $\mathscr{P}_\alpha$ has no analytical expression [18]. However its values can be numerically computed and codes are available at the following address [19].

Figure 1 describes the behavior of $\widehat{X}_{\text{mmse}}$ versus y, for $SNR = -5$ dB, and for $p = .7$ (left) and $p = 1$ (right). The solid line corresponds to the dimension $d = 1$, the dashed line to $d = 5$ (with $y = [y_0, 0, \ldots, 0]^t$) and the dashed-dotted line to $d = 10$. The dotted line represents the Wiener estimator. We first note that for any value of y, the MMSE estimator differs dramatically from the Wiener estimator. However, for any dimension d, it appears that the MMSE, although nonlinear, is roughly linear in 3 (or at least 2) domains; this suggests a possible local approximation by such a suboptimal estimator (*e.g.* by the Wiener filter for small values of y). This approach is left as a perspective since it requires that a threshold between "small" ("medium") and "large" values of y should be determined. Finally, we observe that the dimension has a small influence on the shape of the estimator.

Figure 2 depicts the behavior of the MSE normalized by the dimension, $\frac{MSE}{d}$, of the MMSE estimator, as a function of the SNR, compared to that of the Wiener estimator. It appears that, in general, for small SNR values the performance of the Wiener filter degrades, while for large SNR values both estimators behave similarly. For large values of p, the performance of the Wiener filter is close to that of the MMSE estimator, and degrades more and more as p decreases. The performances of the estimators proposed in [1] are not depicted here, but they appear worse than the Wiener filter in general (and obviously worse than the MMSE). Because of its simplicity, the Wiener filter is to be preferred in general for exponential power distributed X. Only for "small" values of p, the MMSE estimator should be preferred, despite it requires numerical integration. Note however that the estimator proposed by Alecu requires also numerical integration which does not seem simpler than that required to implement the MMSE.

[2] Tr denote the trace of a matrix

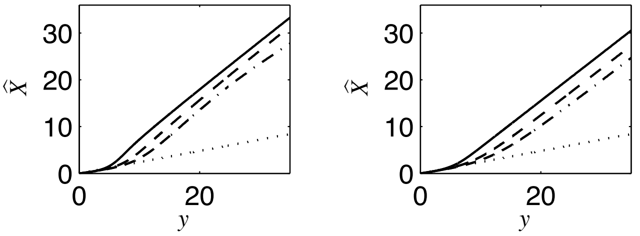

Figure 1: Estimator $\widehat{X}_{\text{mmse}}$ as a function of y for the MMSE estimator of X, exponential power distributed corrupted by Gaussian noise. The SNR is -5dB, and $p = .7$ (left) or $p = 1$ (right) while the dimension is $d = 1$ (solid line), $d = 5$ (dashed line) and $d = 10$ (dashed-dotted line). The dotted line represents the Wiener estimator.

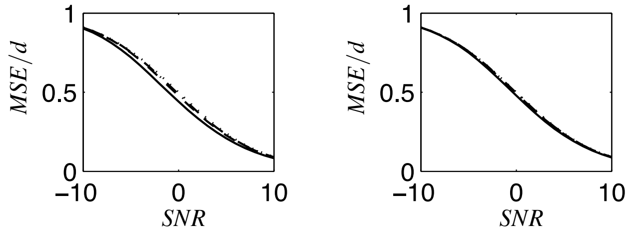

Figure 2: Mean Square Error normalized by dimension d versus the SNR is the same context as in figure 1 and with the same legend ($d = 1$ for the Wiener filter, roughly independent on d).

The second case concerns a vector X with $d-$variate Student -t distribution $p_x(x) \propto \left(1 + \frac{x^t x}{m-2}\right)^{-\frac{d+m}{2}}$ where m is the degree of freedom, assumed larger than 2. [3] Contrarily to the previous example, f_a has an analytical expression given by $f_a(a) \propto a^{-1-m} \mathrm{e}^{-\frac{m-2}{2a^2}}$ [6, 9, 20]. In this case, estimator (4) is evaluated numerically via the change of variable $a = \sigma \tan\theta$. Figure 3 describes the behavior of $\widehat{X}_{\text{mmse}}$ versus y and figure 4 depicts the behavior of the normalized MSE versus *SNR*. The parameters and legends are the same as in the previous example, except that parameter p is replaced by $m = 2.5$ and $m = 5$ respectively. In the illustration, all the previous conclusions and observations still hold.

Although not represented here, we have checked that as m goes to infinity, the MMSE estimator $\widehat{X}_{\text{mmse}}$ converges to the Wiener estimator (X tends in distribution to a Gaussian).

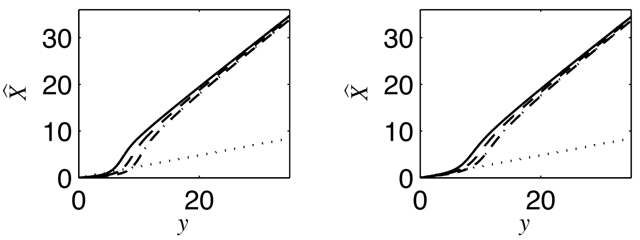

Figure 3: $\widehat{X}_{\text{mmse}}(y)$ for X Student-t distributed and for $m = 2.5$ (left) or $m = 5$ (right) with the same legend as in figure 1.

[3]for values $m \le 2$, $m-2$ is replaced by 1 in the expression of p_x and one can check that X has infinite covariance; for $m = 1$ one finds a Cauchy random vector

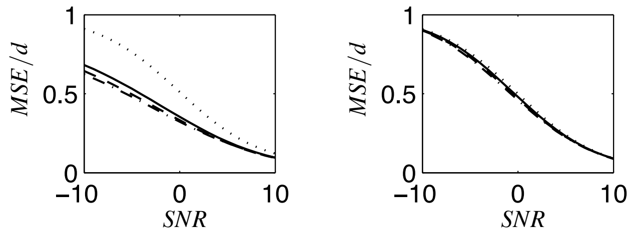

Figure 4: Normalized MSE versus the SNR in the same context as in figure 3 and with the same legend.

The last example aims to show that the previous results still hold even if X, although elliptically distributed, is not a scale mixture of Gaussians. As an example, the multivariate Pearson type II distributions belong to this class. However, as far as we know, there is no analytical form for the mixing function f_a. It can be numerically evaluated (inverse Laplace transform), but for (more simple) illustration purposes only, let us consider an *ad-hoc* example where f_a, taking negative values, is not a pdf. We choose $f_a(a) \propto a^{d-1} J_{m+\frac{d}{2}}(\gamma a^{-2})$ where $m > 1$ and where J_ν is the Bessel function of the first kind and of order ν. From [17, 6.623-3], $p_x(x) \propto \left(\sqrt{(x^t x)^2 + 4\gamma^2} - x^t x\right)^{m+\frac{d}{2}}$ with $\gamma = \frac{2\Gamma\left(\frac{m+d+3}{2}\right)\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m+d+2}{2}\right)\Gamma\left(\frac{m-1}{2}\right)}$. Figure 5 represents then $\widehat{X}_{\text{mmse}}$ versus y and figure 6 depicts the behavior of the normalized MSE versus *SNR*. The parameters and legend are the same as in the previous examples, except that $m = 1.2$ and $m = 2.5$ respectively.

This illustration picks an example that does not belong to the class of scale mixtures of Gaussians. However, in this case, implementing the MMSE still works while for this kind of variable the approach of Alecu seems no more usable since f_a is not a pdf. Furthermore, the whole previous conclusions and observations still hold in this illustration. This suggests that the shape of the MMSE and the generally good performance of the Wiener filter should be generic to the estimation of elliptical distributed vectors embedded in Gaussian noise.

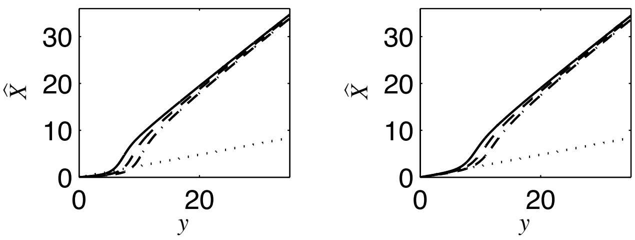

Figure 5: $\widehat{X}_{\text{mmse}}(y)$ for the non scale mixture of Gaussians above-presented and for $m = 1.2$ (left) or $m = 1.5$ (right).

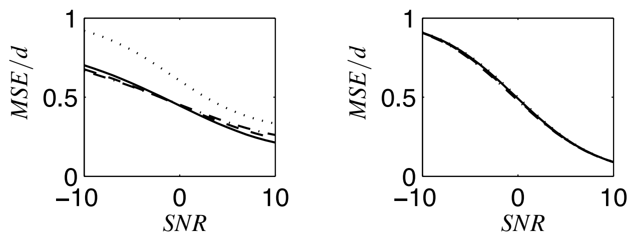

Figure 6: Normalized Mean Square Error versus the SNR is the same context than in figure 5 and with the same legend.

4. DISCUSSION

In this paper we have revisited the denoising problem, extending the approach of [1], that deals only with scalar scale mixture of Gaussians, to the more general class of d-dimensional elliptically distributed vectors. We have shown that when the MMSE estimator (3) cannot be implemented easily and when the mixing functions are known, the minimum mean square estimator can be implemented using numerical integration, at a reasonable computational cost. First, we have seen that the numerical integration is simplified by the introduction of the linear preprocessing (2). The integrations over $\mathbb{R}^d$ required for (3) are then replaced by integrations over $\mathbb{R}_+^2$ in (4). Moreover, we remarked that in the case where the covariance matrices of the vector to be estimated and of the noise are proportional, the problem reduces to a one dimensional estimation problem (up to two additional rotations). The shape of the resulting estimator and its performances are illustrated on several examples where the additive noise is Gaussian; however, the generalization to elliptical additive noise induces only light computational complexity, since the simple integration is replaced by a double one. We note, however, that the suboptimal estimation technique proposed in [1] also required numerical integration. Although not discussed here, refined or fast numerical integration methods can be considered such as quadrature methods.

The shapes of the MMSE suggest that suboptimal approaches using estimators linear in different ranges for y, "small" ("medium") or "large", but with different slopes, are to be considered. This should require to determine a threshold between these ranges of observations y. We also observed that, in terms of MSE, the Wiener estimators seems generally almost as good as the MMSE estimator. Except for highly non-Gaussian cases, this estimator should be privileged since it is very elementary to implement. At the opposite, it is possible that an estimator linear in 2 or 3 ranges of y as just suggested should be a good compromise between the MMSE and the Wiener estimator. But these conclusions rely on the illustration and need to be more deeply investigated. The case where the covariance matrices are not proportional is not illustred here; several Monte-Carlo simulations lead to the same conclusions concerning the MSE. However, contrary to the proportional covariance matrix case, estimator $\widehat{X}_{\text{mmse}}$ is no more elliptically distributed and $\widehat{X}_{\text{mmse}}(y)$ must be analyzed in the entire d-dimensional space. Such a study is left to future investigations.

Finally, other estimators can be considered. In [1], the maximum *a posteriori* (MAP) estimation is discussed, although as presented it is equivalent to the MMSE (*i.e.* in the Gaussian context). Preliminary work shows that, as in the scalar case, the MAP estimator is solution of a scalar nonlinear equation (via the scalar functions d_x and d_z). Numerical techniques like Newton-Raphson algorithms can hence be considered to find the solution of the MAP. Although suboptimal, such an approach can be considered when both the mixing functions are unknown or (3) cannot be implemented.

REFERENCES

[1] T. I. Alecu, S. Voloshynovskiy, and T. Pun, "The Gaussian transform of distributions: Definition, computation and application," *IEEE Trans. on Signal Processing*, vol. 54, pp. 2976–2985, August 2006.

[2] H. Teicher, "On the mixture of distributions," *Annals of Math. Stat.*, vol. 31, pp. 55–73, March 1960.

[3] I. F. Blake and J. B. Thomas, "On a class of processes arising in linear estimation theory," *IEEE Trans. on Information Theory*, vol. 14, pp. 12–16, January 1968.

[4] B. Picinbono, "Spherically invariant and compound Gaussian stochastic processes," *IEEE Trans. on Information Theory*, vol. 17, pp. 77–79, January 1970.

[5] D. Kelker, "Infinite divisibility and variance mixture of normal distributions," *The Annals of Math. Stat.*, vol. 42, pp. 802–808, April 1971.

[6] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *Journal of the Royal Stat. Soc. B*, vol. 36, no. 1, pp. 99–102, 1974.

[7] M. West, "On scale mixtures of normal distributions," *Biometrika*, vol. 74, pp. 646–648, September 1987.

[8] R. Lord, "The use of the Hankel transform in statistics I. General theory and examples," *Biometrika*, vol. 41, pp. 44–55, June 1954.

[9] K.-C. Chu, "Estimation and decision for linear systems with elliptical random processes," *IEEE Trans. on Automatic Control*, vol. 18, pp. 499–505, October 1973.

[10] K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Trans. on Information Theory*, vol. 19, pp. 600–608, September 1973.

[11] E. Gómez-Sánchez-Manzano, M. A. Gómez-Villegas, and J. M. Marín, "Sequence of elliptical distributions and mixtures of normal distributions," *Journal of Multivariate Analysis*, vol. 97, pp. 295–310, February 2006.

[12] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2. New-York: John Wiley & Sons, Inc, 1971.

[13] D. V. Widder, *The Laplace Transform*. Princeton Mathematical Series, Princeton University Press, 1946.

[14] R. E. Williamson, "Multiply monotone functions and their Laplace transforms," *Duke Mathematical Journal*, vol. 23, pp. 189–207, June 1956.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

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

[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. San Diego: Academic Press, 1980.

[18] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes. Stochastic Models with infinite Variance*. New-York: Chapman & Hall, 1994.

[19] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Stochastic Models*, vol. 13, no. 4, pp. 759–774, 1997. http://academic2.american.edu/~jpnolan.

[20] C. Vignat, A. O. Hero III, and J. A. Costa, "About closedness by convolution of the Tsallis maximizers," *Physica A*, vol. 340, pp. 147–152, September 2004.