IEEE Transactions on Signal Processing, accepted October 2005, in press

The Gaussian Transform of distributions: definition, computation and application

Teodor Iulian Alecu*, Sviatoslav Voloshynovskiy and Thierry Pun

1

Abstract—**This paper introduces the general purpose Gaussian Transform of distributions, which aims at representing a generic symmetric distribution as an infinite mixture of Gaussian distributions. We start by the mathematical formulation of the problem and continue with the investigation of the conditions of existence of such a transform. Our analysis leads to the derivation of analytical and numerical tools for the computation of the Gaussian Transform, mainly based on the Laplace and Fourier transforms, as well as of the afferent properties set (e.g. the transform of sums of independent variables).**

The Gaussian Transform of distributions is then analytically derived for the Gaussian and Laplacian distributions, and obtained numerically for the Generalized Gaussian and the Generalized Cauchy distribution families.

In order to illustrate the usage of the proposed transform we further show how an infinite mixture of Gaussians model can be used to estimate/denoise non-Gaussian data with linear estimators based on the Wiener filter. The decomposition of the data into Gaussian components is straightforwardly computed with the Gaussian Transform, previously derived. The estimation is then based on a two-step procedure, the first step consisting in variance estimation, and the second step in data estimation through Wiener filtering. To this purpose we propose new generic variance estimators based on the Infinite Mixture of Gaussians prior. It is shown that the proposed estimators compare favorably in terms of distortion with the shrinkage denoising technique, and that the distortion lower bound under this framework is lower than the classical MMSE bound.

Index Terms—**Gaussian mixture, Gaussian Transform of distributions, Generalized Gaussian, denoising, shrinkage**

I. Introduction

Gaussian distributions are extensively used in the (broad sense) signal processing community, mainly for computational benefits. For instance, in an estimation problem Gaussian priors yield quadratic functionals and linear solutions. In rate-distortion and coding theories, closed form results are mostly available for Gaussian source and channel descriptions [11]. However, real data is most often non–Gaussian distributed, and is best described by other types of distribution (e.g in image processing, the most commonly used model for the wavelet coefficients distribution is the Generalized Gaussian distribution [6]). The goal of the work presented in this paper is to describe non-Gaussian distributions as an infinite mixture of Gaussian distributions, through direct computation of the mixing functions from the non-Gaussian distribution.

In a related work [4], it was proven that any distribution can be approximated through a mixture of Gaussian up to an arbitrary level of precision. However, no hint was given by the author on how to obtain the desired mixture in the general case. In the radar community the problem of modeling non-Gaussian radar clutters led to the theory of Spherically Invariant Random Processes (SIRP) [13], which aimed at generating non-Gaussian multivariate distributions using univariate Gaussian distributions. However, the proposed solutions bypassed the mixing function ([12] and references therein), performing direct computation of the multivariate non-Gaussian distribution from the marginal probability distribution function and the desired covariance matrix. In [5], an analytical formula is given for an infinite mixture of Gaussians equivalent to the Laplacian distribution, and used in a source coding application. Unfortunately, no generalization was attempted by the authors. The work presented here has the roots in their proof and extends the concept through the introduced Gaussian Transform of distributions, which permits straightforward derivation of exact mixing functions for a wide range of symmetric distributions, including the Generalized Cauchy and the Generalized Gaussian distributions.

In order to exemplify the possible usage of the transform we consider then the problem of estimation/denoising, motivated by works such as [9] and [10], where, in order to preserve the computational advantages provided by Gaussian modelling, non-Gaussian distributions are approximated with finite Gaussian mixtures obtained through iterative numerical optimization techniques. Relying on the Gaussian Transform, we propose new generic algorithms for denoising non-Gaussian data, based on the description of non-Gaussian distributions as an Infinite Mixture of Gaussians (IMG).

The rest of the paper is divided in three main sections. In Section II we define the Gaussian Transform, analyze its existence, investigate its properties and derive the mathematical tools for analytical and/or numerical computation. In Section III we exemplify both the transform for some typical distributions such as Generalized Gaussian and Generalized Cauchy, and some of the properties deduced in Section II. The last section tackles the problem of

This work is supported in part by the Swiss NCCR IM2 (Interactive Multimodal Information Management, http://www.im2.ch), the EU Network of Excellence Similar (http://www.similar.cc) and the SNF grant N° PP002-68653/1 .

T. I. Alecu, S. Voloshynovskiy, T. Pun are with the Computer Vision and Multimedia Laboratory, University of Geneva, 24 Rue Général-Dufour, 1204 Geneva, Switzerland (corresponding author: Teodor.Alecu@cui.unige.ch, phone: +41(22)3791084; fax: +41(22)3797780).

estimation/denoising and proposes new Gaussian Transform based denoising schemes, exemplifying them for the Generalized Gaussian Distribution (GGD) family. The obtained results are compared in terms of distortion with the state-of-the-art denoising method of Moulin and Liu [7].

II. The Gaussian Transform of distributions

A. Definition and existence

We consider a generic symmetric continuous distribution *p(x)*. As we are aiming at representing it through an infinite mixture of Gaussians, we can safely disregard the mean, and assume for simplicity reasons that *p(x)* is zero-mean. We are looking for an integral representation in the form:

$$\int_0^{\infty} G(\sigma^2)\mathcal{N}(x \mid \sigma^2)\,d\sigma^2 = p(x), \qquad (1)$$

where $\mathcal{N}(x \mid \sigma^2)$ is the zero-mean Gaussian distribution:

$$\mathcal{N}(x \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2},$$

and $G(\sigma^2)$ is the mixing function that should reproduce the original *p(x)*. We can now introduce the Gaussian Transform.

Definition 1: Gaussian Transform. The direct Gaussian Transform $\mathcal{G}$ is defined as the operator which transforms $p(x)$ into $G(\sigma^2)$, and the Inverse Gaussian Transform $\mathcal{G}^{-1}$ is defined as the operator which maps $G(\sigma^2)$ back to $p(x)$:

$$\mathcal{G}: p(x) \mapsto G(\sigma^2);\ \mathcal{G}^{-1}: G(\sigma^2) \mapsto p(x).$$

Obviously, $\mathcal{G}^{-1}$ is simply given by (1):

$$\mathcal{G}^{-1}\left(G(\sigma^2)\right) = \int_0^{\infty} G(\sigma^2)\mathcal{N}(x \mid \sigma^2)\,d\sigma^2 = p(x). \qquad (2)$$

The Gaussian Transform of a distribution exists if, given $p(x)$, a mixing distribution $G(\sigma^2)$ can be found such as to comply with (1). This can be summarized in three conditions.

Condition 1. For a given *p(x)*, a function $G(\sigma^2)$ defined according to (1) exists.

Condition 2. This function is non-negative.

Condition 3. Its integral $\int_0^{\infty} G(\sigma^2)\,d\sigma^2$ is equal to 1.

The last condition is a consequence of Condition 1. Indeed, if $G(\sigma^2)$ exists, then integrating both sides of (1) with respect to *x* and inverting the integration order one obtains:

$$\int_{-\infty}^{\infty}\left(\int_0^{\infty} G(\sigma^2)\mathcal{N}(x \mid \sigma^2)\,d\sigma^2\right)dx = \int_{-\infty}^{\infty} p(x)\,dx,$$

$$\int_0^{\infty}\left(G(\sigma^2)\int_{-\infty}^{\infty} \mathcal{N}(x \mid \sigma^2)\,dx\right)d\sigma^2 = 1.$$

Finally, since $\mathcal{N}(x \mid \sigma^2)$ is a distribution:

$$\int_0^{\infty} G(\sigma^2)\,d\sigma^2 = 1.$$

In order to investigate *Condition 1*, the existence of $G(\sigma^2)$, we perform the following variable substitutions: $s = x^2$ and $t = \frac{1}{2\sigma^2}$. Since *p(x)* is symmetric, it can be rewritten as:

$$p(x) = p(|x|) = p(\sqrt{s}).$$

The left hand side of (1) transforms to:

$$\int_0^{\infty} G(\sigma^2)\mathcal{N}(x \mid \sigma^2)\,d\sigma^2 = \int_0^{\infty} G\left(\frac{1}{2t}\right)\frac{1}{2t}\frac{1}{\sqrt{\pi t}} e^{-st}\,dt.$$

According to the definition of the Laplace Transform $\mathcal{L}$ [1], equation (1) finally takes the form:

$$\mathcal{L}\left(G\left(\frac{1}{2t}\right)\frac{1}{2t}\frac{1}{\sqrt{\pi t}}\right) = p(\sqrt{s}). \qquad (3)$$

Thus, $G(\sigma^2)$ is linked to the original probability distribution *p(x)* through the Laplace Transform and can be computed using the Inverse Laplace Transform $\mathcal{L}^{-1}$. The direct Gaussian Transform is therefore given by:

$$\mathcal{G}(p(x)) = \frac{1}{\sigma^2}\sqrt{\frac{\pi}{2\sigma^2}}\left(\mathcal{L}^{-1}\left(p(\sqrt{s})\right)(t)\right)_{t=\frac{1}{2\sigma^2}}, \qquad (4)$$

and the Inverse Gaussian Transform can be computed as:

$$\mathcal{G}^{-1}\left(G(\sigma^2)\right) = \left(\mathcal{L}\left(G\left(\frac{1}{2t}\right)\frac{1}{2t}\frac{1}{\sqrt{\pi t}}\right)(s)\right)_{s=x^2}. \qquad (5)$$

Consequently, the existence of the Gaussian Transform is conditioned by the existence of the Inverse Laplace Transform of $p(\sqrt{s})$. Using the general properties of the Laplace Transform, it is sufficient [1] to prove that the limit at infinity of $s \cdot p(\sqrt{s})$ is bounded, or equivalently :

$$\lim_{x\to\infty} x^2 \cdot p(x) < \infty. \qquad (6)$$

The above condition is satisfied by all the distributions from the exponential family, as well as by all the distributions with finite variance.

The equation (4) allows for straightforward identification of Gaussian Transforms for distributions whose Laplace Transforms are known, by simply using handbook tables. Unfortunately, the condition (6) does not guarantee compliance with the Condition 2: non-negativity. As it is rather difficult to verify a priori this constraint, the test should be performed a posteriori, either analytically or numerically. However, by imposing the non-negativity constraint in (1), one can see that it is necessary to have a strictly decreasing distribution function *p(x)* on positive *x*, as *p(x)* is described by a sum of strictly decreasing zero-mean Gaussians functions.

From the same reason, but not related to non-negativity, the distribution function *p(x)* should also be of class C^{∞}.

B. Properties of the Gaussian Transform

Property 1: Uniqueness. If two functions $G_1(\sigma^2)$ and

$G_2(\sigma^2)$ exist such that $\mathcal{G}^{-1}(G_1(\sigma^2)) = p(x)$ and $\mathcal{G}^{-1}(G_2(\sigma^2)) = p(x)$ then they are identical $G_1(\sigma^2) = G_2(\sigma^2)$. Conversely, if two distribution probabilities $p_{X_1}(x)$ and $p_{X_2}(x)$ have the same Gaussian Transform $G(\sigma^2)$ then they are identical.

Proof. The second proposition of the uniqueness property is proved straightforwardly using (1). In what concerns the first proposition, it is a direct consequence of the uniqueness of the Laplace Transform [1], which states that if two functions have the same Laplace transform, then they are identical. It is then sufficient to observe that $\mathcal{G}^{-1}$ is computed as a Laplace Transform (5).

Property 2: Mean Value (variance conservation). The mean value of the Gaussian Transform is equal to the variance of the original distribution.

$$\int_{-\infty}^{\infty} x^2 p(x)\,dx = \int_0^{\infty} \sigma^2 G(\sigma^2)\,d\sigma^2 . \tag{7}$$

Proof. $\int_{-\infty}^{\infty} x^2 p(x)\,dx = \int_{-\infty}^{\infty} x^2 \left(\int_0^{\infty} G(\sigma^2)\mathcal{N}(x\,|\,\sigma^2)\,d\sigma^2 \right) dx =$

$= \int_0^{\infty} G(\sigma^2)\left(\int_{-\infty}^{\infty} x^2 \mathcal{N}(x\,|\,\sigma^2)\,dx \right) d\sigma^2 = \int_0^{\infty} G(\sigma^2)\sigma^2\,d\sigma^2$ q.e.d.

Property 3: Value at infinity. If $\lim_{x\to\infty} x^2 \cdot p(x) < \infty$ then the Gaussian Transform tends to 0 when σ^2 tends to infinity:

$$\lim_{\sigma^2\to\infty} G(\sigma^2) = 0 . \tag{8}$$

Proof. We use the initial value theorem for the Laplace Transform [1] and the direct formula (3):

$$\lim_{t\to 0} G\left(\frac{1}{2t}\right)\frac{1}{2t}\frac{1}{\sqrt{\pi t}} = \lim_{s\to\infty} s\cdot p\left(\sqrt{s}\right) = \lim_{x\to\infty} x^2\cdot p(x),$$
$$\text{then } \lim_{\sigma^2\to\infty} G(\sigma^2)(\sigma^2)^{3/2} < \infty \Rightarrow \lim_{\sigma^2\to\infty} G(\sigma^2) = 0 .$$

We study now the influence of basic operations on the Gaussian Transform, such as scalar multiplication and addition of independent variables.

Property 4: Scaling. If X_1 and X_2 are random variables such that $X_2 = \alpha X_1$ and $\mathcal{G}(p_{X_1}(x)) = G_{X_1}$, then the Gaussian Transform corresponding to the scaled variable X_2 is a scaled version of G_{X_1}:

$$G_{X_2}(\sigma^2) = \frac{1}{\alpha^2} G_{X_1}\left(\frac{\sigma^2}{\alpha^2}\right) \tag{9}$$

Proof. We begin by observing that the distribution probability of the random variable X_2 can be expressed as:

$$p_{X_2}(x) = \frac{1}{\alpha} p_{X_1}\left(\frac{x}{\alpha}\right).$$

Then we compute the Inverse Gaussian Transform of the expression proposed in (9) using (2):

$$\mathcal{G}^{-1}(G_2(\sigma^2)) = \frac{1}{\alpha^2}\int_0^{\infty} G_1\left(\frac{\sigma^2}{\alpha^2}\right)\mathcal{N}(x\,|\,\sigma^2)\,d\sigma^2;$$

$$\mathcal{G}^{-1}(G_2(\sigma^2)) = \int_0^{\infty} G_1\left(\frac{\sigma^2}{\alpha^2}\right)\frac{1}{\alpha\sqrt{2\pi\frac{\sigma^2}{\alpha^2}}} e^{-\frac{x^2}{\alpha^2}*\frac{\alpha^2}{2\sigma^2}}\, d\left(\frac{\sigma^2}{\alpha^2}\right)$$

$$\mathcal{G}^{-1}(G_2(\sigma^2)) = \frac{1}{\alpha} p_{X_1}\left(\frac{x}{\alpha}\right) = p_{X_2}(x).$$

Finally using the uniqueness property we conclude that the Gaussian Transform $\mathcal{G}(p_{X_2}(x))$ is unique and given by (9).

Property 5: Convolution. If X_1 and X_2 are two independent random variables with $\mathcal{G}(p_{X_1}(x)) = G_{X_1}(\sigma^2)$ and $\mathcal{G}(p_{X_2}(x)) = G_{X_2}(\sigma^2)$, then the Gaussian Transform of their sum is the convolution of their respective Gaussian Transforms (the result can be generalized for the sum of multiple variables):

$$\mathcal{G}(p_{X_1+X_2}(x)) = G_{X_1}(\sigma^2) * G_{X_2}(\sigma^2). \tag{10}$$

Proof. Consider the random variable $X = X_1 + X_2$. Since G_{X_1} exists, X_1 is a random Gaussian variable with variance $\sigma_{X_1}^2$ distributed according to the distribution probability G_{X_1}. Similarly, X_2 is a random Gaussian variable with variance $\sigma_{X_2}^2$ distributed according to G_{X_2}. Then X is also a random Gaussian variable with variance $\sigma_X^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2$. But $\sigma_{X_1}^2$ and $\sigma_{X_2}^2$ are independent variables drawn from G_{X_1} and G_{X_2}. It follows that σ_X^2 is a random variable described by the probability distribution $G_X = G_{X_1} * G_{X_2}$, q.e.d.

*Corollary***.** If X_1 and X_2 are independent random variables and X_2 is Gaussian distributed $p_{X_2}(x) = \mathcal{N}(x\,|\,\sigma_{X_2}^2)$, then the Gaussian transform of their sum is a shifted version of G_{X_1}:

$$G_{X_1+X_2}(\sigma^2) = \begin{cases} 0, \sigma^2 < \sigma_{X_2}^2 \\ G_{X_1}(\sigma^2 - \sigma_{X_2}^2), \sigma^2 \geq \sigma_{X_2}^2 \end{cases}. \tag{11}$$

C. *Numerical computation*

The computation of the Gaussian Transform for distributions not available in handbooks is still possible through the complex inversion method for Laplace Transforms known as the Bromwich integral [3]:

$$\mathcal{L}^{-1}(f(s)) = \frac{1}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty} f(u)e^{ut}\,du , \tag{12}$$

where ε is a real positive constant satisfying $\varepsilon > \sup(\mathrm{Re}(poles(f)))$ and u is an auxiliary variable. The equation (12) can be rearranged as a Fourier Transform, allowing the use of the numerous numerical and/or symbolical packages available (ω is the variable in the Fourier space):

$$\mathcal{L}^{-1}(f(s)) = e^{\varepsilon t}\mathcal{F}^{-1}(f(\varepsilon + i\omega)). \tag{13}$$

Very often *p(x)* has no poles, being a continuous and bounded function, and in this case it might be very practical to evaluate (13) in the limit case $\varepsilon \to 0$. Using (4) and (13):

$$G\left(\sigma^2\right)=\frac{1}{\sigma^2}\sqrt{\frac{\pi}{2\sigma^2}}\left(\mathcal{F}^{-1}\left(p\left(\sqrt{i\omega}\right)\right)(t)\right)_{t=\frac{1}{2\sigma^2}}. \quad (14)$$

When the original distribution is only numerically known, approximation of $G\left(\sigma^2\right)$ is still possible either through analytical approximations of *p(x)*, followed by (4) or (14), or through solving the inverse problem yielded by (1), or through real inversion of the Laplace transform using (4). The accuracy of the obtained transforms can then be assessed using an appropriate metric or distance such as Kullback-Leibler. However, the abovementioned approximation methods and the accuracy study constitute stand-alone problems out of the scope of the present paper.

III. Examples of Gaussian Transforms

A. *Analytic Gaussian transforms*

The most obvious and natural example is the Gaussian Transform of a *Gaussian* distribution $\mathcal{N}\left(x \mid \sigma_0^2\right)$. One would expect to have (with δ the Dirac function):

$$\mathcal{G}\left(\mathcal{N}\left(x \mid \sigma_0^2\right)\right)=\delta\left(\sigma^2-\sigma_0^2\right). \quad (15)$$

Proof. It is trivial that (15) verifies the equation (1). Then, by using the uniqueness of the Gaussian Transform one can conclude that the Gaussian Transform of a Gaussian distribution $\mathcal{N}\left(x \mid \sigma_0^2\right)$ is a Dirac function centered in σ_0^2.

The convolution property (10) can now be used to prove a well known result in statistics: the sum of two independent Gaussian variables, with respective probability laws $\mathcal{N}\left(x \mid \sigma_{X_1}^2\right)$ and $\mathcal{N}\left(x \mid \sigma_{X_2}^2\right)$, is another Gaussian variable with probability distribution $\mathcal{N}\left(x \mid \sigma_{X_1}^2+\sigma_{X_2}^2\right)$ (the extension to non-zero mean distributions is trivial). *Proof*:

$$\mathcal{G}\left(p_{X_1}(x)\right)=\delta\left(\sigma^2-\sigma_{X_1}^2\right);\ \mathcal{G}\left(p_{X_2}(x)\right)=\delta\left(\sigma^2-\sigma_{X_2}^2\right)$$
$$\mathcal{G}\left(p_{X_1+X_2}(x)\right)=\mathcal{G}\left(p_{X_1}(x)\right)*\mathcal{G}\left(p_{X_2}(x)\right)=\delta\left(\sigma^2-\left(\sigma_{X_1}^2+\sigma_{X_2}^2\right)\right)$$

Then, inverting the Gaussian Transform:

$$p_{X_1+X_2}=\mathcal{G}^{-1}\left(\delta\left(\sigma^2-\left(\sigma_{X_1}^2+\sigma_{X_2}^2\right)\right)\right)=\mathcal{N}\left(x \mid \sigma_{X_1}^2+\sigma_{X_2}^2\right).$$

Similarly to (15), it is possible to compute the Gaussian Transforms of other usual symmetric distributions using the Laplace transform tables [1]. We exemplify with the *Laplacian* and *Cauchy* distributions.
Laplacian distribution:

$$p_x^L(x \mid \lambda)=\frac{\lambda}{2}e^{-\lambda x};\ p_x^L\left(\sqrt{s} \mid \lambda\right)=\frac{\lambda}{2}e^{-\lambda\sqrt{s}}$$
$$\mathcal{G}\left(p_x^L\left(\sqrt{s} \mid \lambda\right)\right)=\frac{\lambda^2}{2}e^{-\frac{\lambda^2}{2}\sigma^2}. \quad (16)$$

As mentioned, the result (16) was already proven in [5].
Cauchy distribution:

$$p_x^C(x \mid b)=\frac{1}{\pi}\frac{b}{b^2+x^2};\ p_x^C\left(\sqrt{s} \mid b\right)=\frac{1}{\pi}\frac{b}{b^2+s}$$
$$\mathcal{G}\left(p_x^C(x \mid b)\right)=\frac{1}{\sigma^2}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{b}{2\sigma^2}}. \quad (17)$$

The results (15), (16) and (17) are plotted in Fig. 1.
To exemplify the C*orollary* to the **Convolution Property**, consider now Cauchy data contaminated with independent additive white Gaussian (AWG) noise. Then the Gaussian Transform of the measured data (Fig. 2) is a shifted version of the original data transform (11).

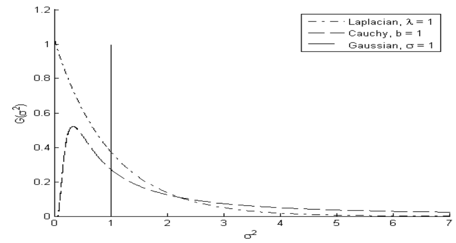

Figure 1. Gaussian Transforms of the Laplacian, Gaussian and Cauchy distributions

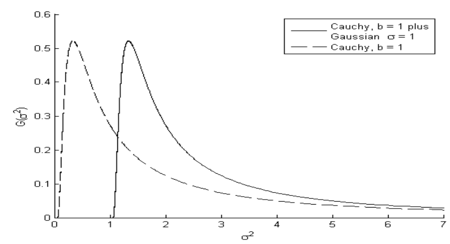

Figure 2. Shift of the Gaussian Transform following data contamination with AWG noise

The previous analytical results will be generalized in subsection *B* through numerical computations.

B. *Numerical computation of Gaussian Transforms*

This part illustrates the numerical computation of the Gaussian Transform through (14) for the Generalized Gaussian and Generalized Cauchy distributions families.

1) *Generalized Gaussian Distribution (GGD)*

The GGD family is described by an exponential probability density function with parameters γ and σ_γ:

$$p_x^G\left(x \mid \gamma,\sigma_\gamma\right)=\frac{\gamma\eta(\gamma)}{2\Gamma\left(\frac{1}{\gamma}\right)}\frac{1}{\sigma_\gamma}e\left(-\left(\eta(\gamma)\left|\frac{x}{\sigma_\gamma}\right|\right)^\gamma\right), \quad (18)$$

where $\eta(\gamma)=\sqrt{\Gamma\left(\frac{3}{\gamma}\right)\Gamma\left(\frac{1}{\gamma}\right)^{-1}}$. For $\gamma=1$ the GGD particularizes to the Laplacian distribution, while for $\gamma=2$ one obtains the Gaussian distribution. The Gaussian Transform of the Generalized Gaussian distribution can not be obtained in analytical form using (4). However, it does exist (6) and can be calculated numerically through (14):

$$G_{X|\gamma,\sigma_\gamma}\left(\sigma^2\right)=\frac{1}{\sigma^2}\sqrt{\frac{\pi}{2\sigma^2}}\left(\mathcal{F}^{-1}\left(p_{X|\gamma,\sigma_\gamma}^G\left(\sqrt{i\omega}\right)\right)(t)\right)_{t=\frac{1}{2\sigma^2}}$$
$$p_{X|\gamma,\sigma_\gamma}^G\left(\sqrt{i\omega}\right)=\frac{\gamma\eta(\gamma)}{2\Gamma\left(\frac{1}{\gamma}\right)}\frac{1}{\sigma_\gamma}e\left(-\left(\frac{\eta(\gamma)}{\sigma_\gamma}\right)^\gamma(i\omega)^{\gamma/2}\right).$$

The Gaussian Transforms for γ ranging from 0.5 to 2 with fixed $\sigma_\gamma = 1$ are plotted in Fig. 3. The transforms evolve from a Dirac-like distribution centered on 0 for small γ to exponential for $\gamma = 1$, then Rayleigh-like for $\gamma = 1.2$, bell-shaped for $\gamma = 1.5$ and again Dirac-like centered at σ_γ^2 for $\gamma = 2$. As expected, the Gaussian Transform of the Laplacian distribution ($\gamma = 1$) is exponential (16).

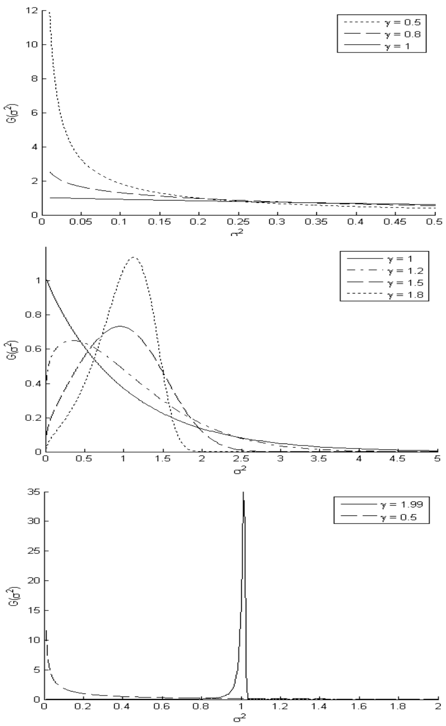

Figure 3. Gaussian Transforms of GGD

Unfortunately, the real part of the complex probability function diverges for periodical values of γ, which impedes the computation of the transform through this method for $\gamma>2$. However, real data (in transform domains such as discrete cosine and wavelet) is mostly confined to $0<\gamma<2$ [6].

2) Generalized Cauchy Distribution (GCD)

The GCD probability density function is given by

$$p_x^C\left(x \mid \nu, b\right) = \frac{\Gamma\left(\nu + \frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(\nu)} \frac{b^{2\nu}}{\left(b^2 + x^2\right)^{\nu+0.5}},$$

and it particularizes to the Cauchy distribution for $\nu = 0.5$. Its Gaussian Transform can be computed through:

$$G_{X|\nu,b}\left(\sigma^2\right) = \frac{b^{2\nu}\Gamma(\nu+0.5)}{\sigma^2\sqrt{2\sigma^2}\,\Gamma(\nu)} \left(\mathcal{F}^{-1}\left(\left(b^2 + i\omega\right)^{-\nu-0.5}\right)\right)_{t=\frac{1}{2\sigma^2}}.$$

Corresponding plots are given in Fig. 4.

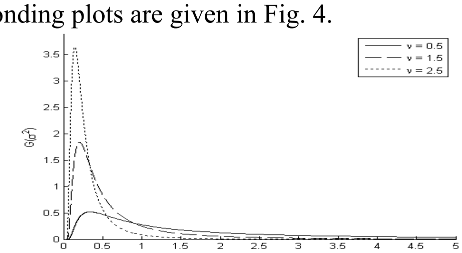

Figure 4. Gaussian Transforms of GCD

As a remark, the variance of the Cauchy distribution being infinite, its Gaussian Transform has infinite mean value (7). The GCD family possesses finite variance only for $\nu > 1$.

IV. Denoising

We are interested in the generic denoising problem of estimating the original scalar data *x* from the measured data *y*, degraded by additive noise *z*:

$$y = x + z. \tag{19}$$

We assume that the noise random variable *Z* is zero-mean Gaussian distributed with variance σ_Z^2:

$$p_Z(z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{z^2}{2\sigma_Z^2}}. \tag{20}$$

We further assume that the probability density function $p_X(x)$, which describes the source random variable *X*, is a zero-mean symmetric distribution, and that its Gaussian Transform $G_X\left(\sigma^2\right) = \mathcal{G}\left(p_X(x)\right)$ exists. For exemplification of the proposed techniques we consider $p_X(x)$ to be the Generalized Gaussian distribution (18).

We base our estimation on the *Maximum a Posteriori* (MAP) principle [14]:

$$\hat{x} = \arg\max_x p_{X|Y}(x \mid y),$$

which yields for i.i.d. zero-mean Gaussian data and noise (with variances σ_X^2 and σ_Z^2) the classical Wiener filter:

$$\hat{x} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Z^2} y. \tag{21}$$

We also note that the estimate (21) is also the *Minimum Mean Square Error* (MMSE) estimate for i.i.d. zero-mean Gaussian data and noise:

$$\begin{aligned} \hat{x}_{\text{MMSE}} &= \arg\min_x \int_x (x - \hat{x})^2 p_{X|Y}(x \mid y)\,dx \\ \hat{x}_{\text{MMSE}} &= \int_x x p_{X|Y}(x \mid y)\,dx \end{aligned}. \tag{22}$$

According to the IMG model, each data sample is interpreted to be drawn from a Gaussian distribution with variance σ_X^2 distributed accordingly to the Gaussian Transform $G_X\left(\sigma^2\right)$. Consequently, using the prior knowledge of $G_X\left(\sigma^2\right)$, we can estimate the variance from *y* and use Wiener filtering (21). Thus, our estimator is given by:

$$\hat{x} = \frac{\sigma_X^2(y)}{\sigma_X^2(y) + \sigma_Z^2} y. \tag{23}$$

The initial problem is now reduced to a variance estimation problem. For this purpose, we begin by introducing the cumulative estimation technique.

A. Cumulative Estimation

We are looking for an estimator that provides variance estimates consistent with the original data distribution, in the sense that the description of the data as a Gaussian mixture should be identical to the non-Gaussian original description. We set the above assertion as a consistency principle.

Principle 1: Consistency. The probability distribution function of the variance estimate should be identical to the Gaussian Transform of the prior distribution $p_X(x)$.

Moreover, since the distributions $p_X(x)$, $p_Z(z)$ are symmetric and zero-mean, we assume an estimator of the form:

$$\hat{\sigma}_X^2(y) = \xi_{\sigma_X^2}^{\mathcal{M}}\left(\mathcal{M}(y)\right),$$

where $\xi_{\sigma_X^2}^{\mathcal{M}}$ denotes the consistent variance estimator depending on the symmetric real non-negative function:

$$\mathcal{M}:\mathbb{R}\mapsto\mathbb{R}^{0+};\ \mathcal{M}(y)=\mathcal{M}(-y),$$

where $\mathbb{R}$ and $\mathbb{R}^{0+}$ denote real, respectively real nonnegative numbers. As an example, a possible $\mathcal{M}$ function is the absolute value of y. We further assume that this estimator is monotonically increasing with the function $\mathcal{M}$:

$$\mathcal{M}(y_1)<\mathcal{M}(y_2)\Rightarrow \xi_{\sigma_X^2}^{\mathcal{M}}\left(\mathcal{M}(y_1)\right)<\xi_{\sigma_X^2}^{\mathcal{M}}\left(\mathcal{M}(y_2)\right). \quad (24)$$

If one chooses the $\mathcal{M}$ function as the absolute value $|y|$, this simply means that if the absolute value of y is larger, then the estimated variance $\hat{\sigma}_X^2(y)$ is also larger. If the variables are vectors (e.g. local estimation paradigm) one can similarly use as the $\mathcal{M}$ function the variance of the measurement vector, and (24) would mean that if the variance of the measurement is larger, then the estimated variance is also larger.

The assumption in (24) implies that the cumulative distribution function of the variance estimates is equal to the cumulative distribution function of $\mathcal{M}$:

$$p\left(\hat{\sigma}_X^2 \le \hat{\sigma}_X^2(y)\right) = p\left(\mathcal{M}\le\mathcal{M}(y)\right). \quad (25)$$

Considering continuous probability density functions and incorporating the consistency condition (the distribution of the estimates should be identical to G_X), (25) can be rewritten as:

$$\begin{gathered} P_{\mathcal{M}}\left(\mathcal{M}(y)\right) = PG_X\left(\xi_{\sigma_X^2}^{\mathcal{M}}\left(\mathcal{M}(y)\right)\right), \\ P_{\mathcal{M}}\left(\mathcal{M}(y)\right) = \int_0^{\mathcal{M}(y)} p_{\mathcal{M}}(u)du;\ PG_X\left(\sigma_X^2\right) = \int_0^{\sigma_X^2} G_X(u)du, \end{gathered} \quad (26)$$

where $p_{\mathcal{M}}$ is the probability density function describing $\mathcal{M}(y)$, and $P_{\mathcal{M}}$ and PG_X are cumulative probability functions, as defined in (26). Then the estimator we are looking for is simply given by:

$$\xi_{\sigma_X^2}^{\mathcal{M}}\left(\mathcal{M}(y)\right) = PG_X^{-1}\left(P_{\mathcal{M}}\left(\mathcal{M}(y)\right)\right), \quad (27)$$

where PG_X^{-1} is the mathematical inverse of the function PG_X:

$$\text{if } PG_X\left(\sigma^2\right)=u \text{ then } PG_X^{-1}(u)=\sigma^2.$$

We denote (27) as the *cumulative estimator*, simply because it uses cumulative probability functions for the estimation.

We further infer that the cumulative estimator is robust with respect to monotonically increasing transformations of $\mathcal{M}$.

Theorem 1. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are two symmetric real non-negative functions $\mathcal{M}_1,\mathcal{M}_2:\mathbb{R}\mapsto\mathbb{R}^{0+}$, and if

$$\mathcal{M}_1(y_1)<\mathcal{M}_1(y_2)\Leftrightarrow\mathcal{M}_2(y_1)<\mathcal{M}_2(y_2),$$

then the associated cumulative estimators are identical:

$$\xi_{\sigma_X^2}^{\mathcal{M}_1}\left(\mathcal{M}_1(y)\right) = \xi_{\sigma_X^2}^{\mathcal{M}_2}\left(\mathcal{M}_2(y)\right).$$

Proof. Since the inequalities are preserved (theorem hypothesis), then the cumulative distribution functions $P_{\mathcal{M}_1}$ and $P_{\mathcal{M}_2}$ are equal:

$$p\left(\mathcal{M}_1\le\mathcal{M}_1(y)\right) = p\left(\mathcal{M}_2\le\mathcal{M}_2(y)\right) \text{ then}$$

$$P_{\mathcal{M}_1}\left(\mathcal{M}_1(y)\right) = \int_0^{\mathcal{M}_1(y)} p_{\mathcal{M}_1}(u)du = \int_0^{\mathcal{M}_2(y)} p_{\mathcal{M}_2}(u)du = P_{\mathcal{M}_2}\left(\mathcal{M}_2(y)\right).$$

The proof is ended by corroborating the above equality with the cumulative estimation formula (27).

B. IMG cumulative estimator

The cumulative estimator (27) can be directly used in (23) by setting the $\mathcal{M}$ function as $\mathcal{M}(y)=|y|$, and the IMG cumulative estimator (IMG CE) is simply:

$$\begin{gathered} \hat{x}_{IMG\_CE} = \frac{\hat{\sigma}_{X\_CE}^2(y)}{\hat{\sigma}_{X\_CE}^2(y)+\sigma_Z^2}\,y \text{ with} \\ \hat{\sigma}_{X\_CE}^2(y) = PG_X^{-1}\left(P_{\mathcal{M}}(y)\right),\ P_{\mathcal{M}}(y) = 2\int_0^{|y|} p_Y(u)du. \end{gathered} \quad (28)$$

As X and Z are assumed to be independent, $p_Y(y)$ can be computed as the convolution of $p_X(x)$ and $p_Z(z)$:

$$p_Y = p_X * p_Z.$$

In the general case the operations in (28) have to be performed numerically. However, if the distribution $p_X(x)$ is Laplacian (GGD with $\gamma=1$), its Gaussian Transform is (16), (18):

$$G_{X|\gamma,\sigma_\gamma}^G\left(\sigma^2\right) = \frac{1}{\sigma_\gamma^2}e^{-\frac{\sigma^2}{\sigma_\gamma^2}}, \quad (29)$$

and the cumulative estimator from (28) reduces to:

$$\sigma_{X\_CE,\gamma=1}^2(y) = -\sigma_\gamma^2\ln\left(1-P_{\mathcal{M}}(y)\right). \quad (30)$$

As γ tends to 2, the Gaussian Transform of the GGD tends asymptotically to a Dirac function centered at σ_γ^2 (15), and the estimator (28) to the classical Wiener filter (21). In order to asses the performances of the proposed estimators we measure the accuracy of the estimate according to the L_2 norm:

$$e(x,\hat{x}) = (x-\hat{x})^2.$$

The global distortion induced by an estimator is therefore:

$$D = \iint e(x,\hat{x})p_X(x)p_Z(z)dxdz. \quad (31)$$

We compare our estimator in terms of distortion with the estimators known as shrinkage functions from [7]. The idea of shrinking originates from [8], but [7] provides both improved results and statistical interpretation of the functions. In fact, the shrinkage functions from [7] are the result of direct MAP estimation, which in the case of GGD data and Gaussian noise yields $\hat{x}$ as the solution of the equation in x:

$$y = x - \sigma_Z^2\sigma_\gamma^{-\gamma}\gamma\eta(\gamma)x^{\gamma-1}. \quad (32)$$

Equation (32) has analytical solutions for $\gamma=0.5$ (cubic), $\gamma=1 \text{ or } 2$ (linear) and $\gamma=1.5$ (quadratic). The resulting shrinkage functions have thresholding characteristics for

$\gamma \le 1$ and yield the Wiener filter for $\gamma = 2$. The equivalent IMG CE functions do not present thresholding characteristics for any γ, behaving in continuous manner. Graphic illustration of the shrinkage functions and of the equivalent IMG CE functions obtained via (28) are presented in Fig. 5.

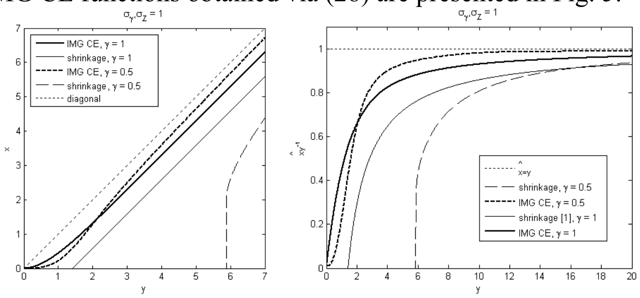

Figure 5. Shrinkage functions and IMG CE estimation curves

One can also compute the lower bound on the distortion induced by an estimation procedure based on signal decomposition into an IMG. Assuming that the Gaussian components of the IMG are separable, and that the variance estimate is constant over each separate component, the equation (21) is the MMSE estimate for each component. The bound is attained when the estimator "knows" the Gaussian component of the mixture that has generated the estimated sample. The ensuing distortion lower bound is:

$$D_{\text{lower bound}} = \int G_X\left(\sigma^2\right) \frac{\sigma^2 \sigma_Z^2}{\sigma^2 + \sigma_Z^2} d\sigma^2 , \qquad (33)$$

where the ratio $\frac{\sigma^2 \sigma_Z^2}{\sigma^2 + \sigma_Z^2}$ is the theoretical distortion of the MMSE filter (21). We also compare our results with the MMSE distortion bound, given by the MMSE estimator (22), which can be rewritten as follows:

$$\hat{x}_{\text{MMSE}} = \int_x x p\left(x \mid y\right) dx = \int_x x \left(\int_{\sigma^2} p\left(x \mid y, \sigma^2\right) p\left(\sigma^2 \mid y\right) d\sigma^2 \right) dx .$$

Using Bayes rule and switching integrals, one obtains:

$$\hat{x}_{\text{MMSE}} = \int_{\sigma^2} \left(\int_x x p\left(x \mid y, \sigma^2\right) dx \right) \frac{p\left(y \mid \sigma^2\right) p\left(\sigma^2\right)}{p\left(y\right)} d\sigma^2 .$$

Considering Gaussian noise the first parenthesis yields the MMSE estimate (21), and by integrating the Gaussian Transform and developing $p\left(y \mid \sigma^2\right)$ and $p\left(y\right)$ one finds:

$$\hat{x}_{\text{MMSE}} = \frac{\int_{\sigma^2} \frac{\sigma^2}{\sigma^2 + \sigma_Z^2} \mathcal{N}\left(y \mid \sigma^2 + \sigma_Z^2\right) G_X\left(\sigma^2\right) d\sigma^2}{\int_{\sigma^2} \mathcal{N}\left(y \mid \sigma^2 + \sigma_Z^2\right) G_X\left(\sigma^2\right) d\sigma^2} y . \qquad (34)$$

The MMSE distortion bound is then computed by plugging the result (34) into (31). The IMG lower bound (33) and the MMSE bound (34) are superposed in Fig. 6, which displays the distortion of the estimation (31) as a function of the signal to noise ratio $\text{SNR} = 10 \log_{10} \frac{\sigma_\gamma^2}{\sigma_Z^2}$ (the variance of the GGD family is given by σ_γ^2, see the moments method from [6] and equation (18)). One may notice that the IMG lower bound is lower than the MMSE bound, a result due to the assumption of separable IMG components. In fact, for a given *y*, the MMSE estimator is unique (injective estimation), which is not the case for the ideal estimator that would lead to the proposed IMG lower bound, and that would yield, for the same value of *y*, different estimates for each isolated Gaussian component.

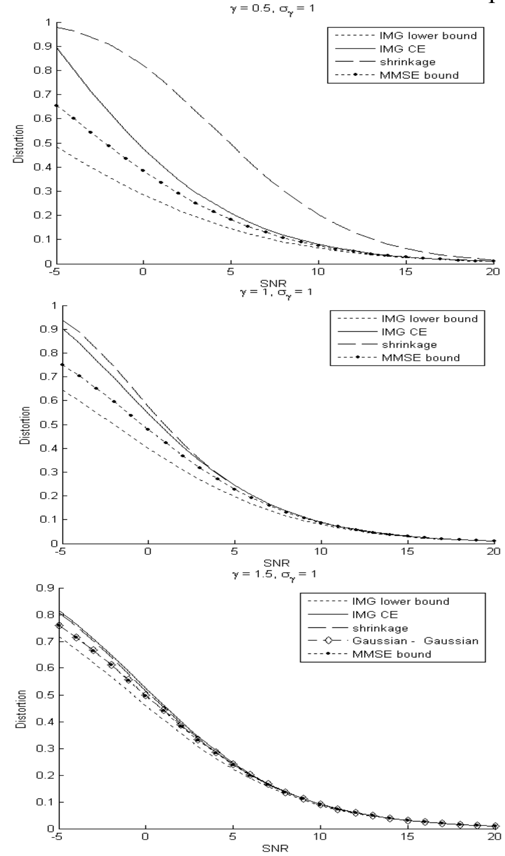

Figure 6. Distortion for GGD data as a function of SNR (dB)

A considerable difference in performance in the favor of IMG CE with respect to the shrinkage functions is observed for $\gamma = 0.5$. The difference tends to decrease for larger γ, and inverts for $\gamma = 1.5$, albeit with a very small relative module. However, the IMG CE estimator being unique for a given y, the IMG CE distortion curve necessarily remains above the MMSE bound, as for any other injective estimator. All curves tend to the distortion of the Wiener filter when γ approaches 2, plotted dashed-dotted in Fig.2 on the graph corresponding to $\gamma = 1.5$.

C. IMG ML and MAP variance estimation

The simplest way of obtaining the variances in a statistical framework is through *Maximum Likelihood* (ML) estimation:

$$\hat{\sigma}_{X\_ML}^2 = \arg\max_{\sigma^2} \left[p\left(y \mid \sigma^2\right) \right], \qquad (35)$$

which for zero-mean independent Gaussian source and noise data yields the solution:

$$\hat{\sigma}_{X\_ML}^2 = \max\left(y^2 - \sigma_Z^2, 0\right). \qquad (36)$$

If the variance distribution were known, one could improve (36) through MAP estimation. Using the Gaussian Transform the MAP variance estimate is:

$$\hat{\sigma}_{X\_MAP}^2 = \arg\max_{\sigma^2} \left[G_X\left(\sigma^2\right) p\left(y \mid \sigma^2\right) \right]. \qquad (37)$$

Generally, the expression (37) requires numerical maximization, but an analytical form exists for $\gamma = 1$ and Gaussian noise (using (29) and differentiation of (37)):

$$\hat{\sigma}^2_{X\_MAP,\gamma=1} = \max\left(\frac{2y^2}{1+\sqrt{1+\frac{8y^2}{\sigma_\gamma^2}}} - \sigma_Z^2, 0 \right). \quad (38)$$

One would expect the IMG MAP estimator to behave similarly to the original shrinkage functions of Moulin and Liu, as it applies the MAP principle two times (37), (23) to obtain an indirect estimate, whereas the shrinkage functions are the result of direct MAP estimation (32). This is confirmed in Fig. 7, which compares the equivalent shrinkage functions of the IMG ML and IMG MAP estimators with the IMG CE curve and with the Moulin and Liu curve for $\gamma = 1$.

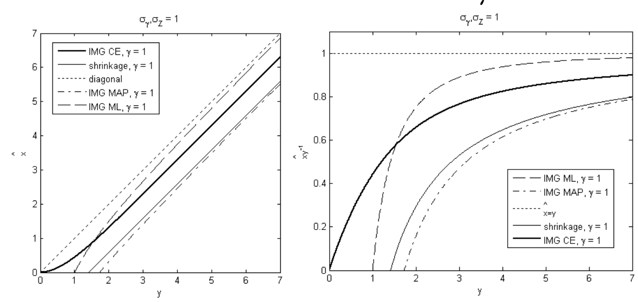

Figure 7. IMG ML and MAP equivalent shrinkage functions

Unlike the IMG CE estimator, the two proposed IMG ML and MAP estimators do present thresholding characteristics, and, as expected, the equivalent IMG MAP curve behaves very similarly to the original shrinkage curve.

We tested empirically the MAP (38) estimator by generating 1D Generalized Gaussian data of length N=10^6 with $\gamma = 0.5$ and $\gamma = 1$, adding independent white Gaussian noise, and performing denoising according to (21). The source and noise probability densities were assumed to be known, and the corresponding distortion was computed as the mean square error (Fig. 8). We also tested the cumulative estimator (28) and the shrinkage function (32).

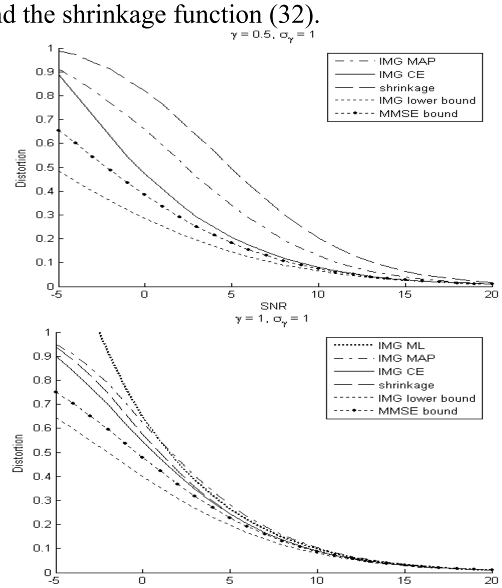

Figure 8. Empirical distortion curves

As expected, empirical results reproduce the theoretical results obtained for the scalar estimators (Fig. 6). The IMG MAP results are close, but above the direct MAP estimation of the shrinkage functions for $\gamma = 1$ (this result was predictable from the curves in Fig.7), however they notably improve on the Moulin and Liu results for $\gamma = 0.5$. We expect the IMG MAP curves to remain above the Moulin and Liu curves for $\gamma \geq 1$, both falling slowly down to the Gaussian-Gaussian curve in Fig.6 as the IMG MAP estimator tends to the Wiener filter for $\gamma = 2$.

While the IMG MAP estimator does not outperform the cumulative estimator, nor the shrinkage estimator for $\gamma \geq 1$, and while the IMG CE estimator does not outperform the classical MMSE estimate, the linear nature of the estimator (23), coupled with the simple analytical forms (38) and (30) for $\gamma = 1$, allowed the successful use of the IMG MAP and IMG CE estimators in the highly underdetermined EEG (electroencephalogram) inverse problem [15] with Laplacian prior, where a direct application of the MAP principle would lead to an underdetermined quadratic programming problem, and where a direct application of the MMSE principle would lead to highly expensive computations.

V. CONCLUSION AND FUTURE WORK

We introduced in this paper the Gaussian Transform, which allows for the representation of symmetric distributions as an infinite Gaussian mixture. The scope of applicability of the Gaussian Transform is potentially very broad, from denoising and regularization to filtering, coding, compression, watermarking etc. However, an extension of the concept to non-symmetric distribution would be required for some specific applications. Further investigation of the existence conditions, especially non-negativity, is also necessary. Finally, one would need adapted numerical packages (most likely based on existing Laplace and Fourier transform computational packages) for the computation of Gaussian Transforms of both analytically and numerically defined distributions.

We also presented in this paper various denoising algorithms based on the Infinite Mixture of Gaussians model. Their application to the denoising of Generalized Gaussian data showed significant improvements in terms of distortion, both theoretically and empirically, with respect to the shrinkage functions of Moulin&Liu. Also, the lower bound under the proposed IMG estimation framework is lower than the classical MMSE bound.

However the potential of IMG based denoising is not yet fully explored, as the lower bound presented in this paper is still below current results. The estimation methods proposed may reach this lower bound in the case where the data is generated locally from the same Gaussian component, which would allow for exact variance estimation.

ACKNOWLEDGMENT

The authors thank Dr. Oleksiy Koval for careful reading and fruitful discussions. The authors also thank anonymous reviewers for valuable comments and suggestions.

REFERENCES

1. A. D. Poularikas, S. Seely, “Laplace Transforms. The Transforms and Applications Handbook: Second Edition”. Ed. Alexander D. Poularikas Boca Raton: CRC Press LLC, 2000
2. G. Arfken, Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, pp. 481-485, 1985.
3. G. Arfken, "Inverse Laplace Transformation." 15.12 in Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, pp. 853-861, 1985.
4. R. Wilson, “MGMM: Multiresolution Gaussian Mixture Models for Computer Vision”, in IEEE International Conference on Pattern Recognition, Barcelona, Spain, September 2000
5. A. Hjørungnes, J. M. Lervik, and T. A. Ramstad , “Entropy Coding of Composite Sources Modeled by Infinite Mixture Gaussian Distributions”, in Proc. 1996 IEEE Digital Signal Processing Workshop , Loen, Norway, September 1996
6. S.G. Mallat, “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation”, in IEEE Transactions on Pattern Analysis and Machine Intelligence Volume 11 , Issue 7 (July 1989) p. 674-693.
7. P. Moulin and J. Liu, "Analysis of Multiresolution Image Denoising Schemes Using Generalized - Gaussian and Complexity Priors", IEEE Trans. Info. Theory, Special Issue on Multiscale Analysis, Vol. 45, No. 3, pp. 909-919, Apr. 1999
8. D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation Via Wavelet Shrinkage", Biometrika, Vol. 81, pp. 425-455, 1994.
9. A. Bijaoui, "Wavelets, Gaussian mixtures and Wiener filtering", Signal Processing, Vol.82, No.4, April 2002, pp 709-712.
10. Benaroya, L; Bimbot, F; "Wiener-based source separation with HMM/GMM using a single sensor", Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation, April 2003, Nara, Japan.
11. T. M. Cover, J. A. Thomas, "Elements of Information Theory", John Wiley & Sons Inc, 1991
12. M. Rangaswamy, D. Weiner, A. Ozturk, "Non-Gaussian Random Vector Identification Using Spherically Invariant Random Processes", in IEEE Transactions on Aerospace and Electronic Systems, vol. 29, No. 1, January 1993
13. E. Conte and M. Longo, "Characterization of radar clutter as a spherically invariant random process", in IEEE Proceedings, Pt. F (Communication, Radar and Signal Processing), 134, 2 (1987), 191-197
14. M.Bertero, P.Boccacci, "Introduction to Inverse Problems in Imaging", IOP Publishing, Bristol, 1998
15. T. I. Alecu, P. Missionnier, S. Voloshynovskiy, P. Giannakopoulos and T. Pun, "Soft/hard focalization in the EEG inverse problem", in IEEE Workshop on Statistical Signal Processing (SSP'05), Bordeaux, France, July, 17-20 2005.