On the Modeling of Small Sample Distributions with Generalized Gaussian Density in a Maximum Likelihood Framework

Sylvain Meignen*
LMC-IMAG Laboratory, University of Grenoble, France
Tel:0033-4-76-51-43-95
FAX:0033-4-76-63-12-63
E-mail: sylvain.meignen@imag.fr

Hubert Meignen
Tel:0033-4-76-51-43-95
E-mail: hubert.meignen@wanadoo.fr

Abstract- The modeling of sample distributions with generalized Gaussian density (GGD) has received a lot of interest. Most papers justify the existence of GGD parameters through the asymptotic behavior of some mathematical expressions (i.e. the sample is supposed to be large). In this paper we show that the computation of GGD parameters on small samples is not the same as on larger ones. In a maximum likelihood framework, we exhibit a necessary and sufficient condition for the existence of the parameters. We derive an algorithm to compute them and then compare it to some existing methods on random images of different sizes.

Keywords- generalized Gaussian density, parameter estimation, maximum likelihood.
EDICS Category:2-MODL

I. Introduction

The modeling of probability density distributions of coefficients produced by the discrete cosine transform [4] [8], by wavelet transform subbands [9] [3] [13] [15] or by steerable pyramid transform algorithms [12] may be efficiently achieved by adaptively adjusting the parameters of a *generalized Gaussian density* (GGD) function. Applications of the modeling of subband coefficients with GGD range from texture analysis [15] [3] [13] and image denoising [10] to video coding [11]. The estimation of GGD parameters may be carried out either by use of the moment method [3] [13], entropy matching [2] or in a maximum likelihood (ML) framework [15]. In all these approaches, the existence and the uniqueness of the parameters are based on asymptotic behavior, that is the sample is supposed to be sufficiently large. However, in signal and image processing, we often deal with small samples for which the existence of the parameters is unknown. In this paper, after recalling the two main approaches to compute GGD parameters in section II (moment method and ML framework), we give a necessary and sufficient condition for the existence of the parameters in a ML framework (section III). Section IV is devoted to the derivation of a new algorithm to compute the parameters in a ML framework. In section V, we compare our method with the moment method (MM) used in [9] [3] [13] and to the ML framework proposed in [15]. The comparison is done on random images of different sizes for which the theoretical GGD parameters are known.

II. GGD Parameters Estimation

We present, in the following, the two main approaches used in image processing for GGD parameters estimation. We assume that a sample $X_L = (x_1, x_2, \cdots, x_L)$ is such that each x_i is a realization of the variable x whose density is given by:

$$P_{\alpha,\beta}(x) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} e^{-(\frac{|x|}{\alpha})^{\beta}} \tag{1}$$

The problem we address is the estimation of (α, β) given the sample X_L.

A. *GGD Parameters Estimation with the Moment Method*

A first estimator of GGD parameters can be computed with the first two moments $m_1 = \int |x| P_{\alpha,\beta}(x)dx$ and $m_2 = \int |x|^2 P_{\alpha,\beta}(x)dx$ through $\alpha = m_1 \frac{\Gamma(\frac{1}{\beta})}{\Gamma(\frac{2}{\beta})}$ and $\beta = F^{-1}\left(\frac{m_1^2}{m_2}\right)$, where $F(x) = \frac{\Gamma^2(\frac{2}{x})}{\Gamma(\frac{3}{x})\Gamma(\frac{1}{x})}$ and $\Gamma(t) = \int_0^{+\infty} e^{-x} x^{t-1} dx$. One often uses $\hat{m}_1 = \frac{1}{L}\sum_i |x_i|$ and $\hat{m}_2 = \frac{1}{L}\sum_i |x_i|^2$ as estimators of m_1 and m_2 respectively leading to $\hat{\alpha}_{MM} = \hat{m}_1 \frac{\Gamma(\frac{1}{\beta})}{\Gamma(\frac{2}{\beta})}$ and $\hat{\beta}_{MM} = F^{-1}(\frac{\hat{m}_1^2}{\hat{m}_2})$ where the index MM stands for the moment method. This approach imposes that $\frac{\hat{m}_1^2}{\hat{m}_2}$ be smaller than $\frac{3}{4}$. Indeed, with the so-called "Euler infinite product" [1], $\log \Gamma(z) = -\gamma z - \log(z) + \sum_{k=1}^{\infty} \left[\frac{z}{k} - \log(1+\frac{z}{k})\right]$, where γ is the Euler constant, we can write $\log F(\frac{1}{x}) = \log(\frac{3}{4}) - \sum_{k=1}^{\infty} \log \frac{(1+2\frac{x}{k})^2}{(1+3\frac{x}{k})(1+\frac{x}{k})} = \log(\frac{3}{4}) - G(x)$. Since $\lim_{x\to 0} G(x) = 0$ and $G(x) > 0$, $F(x)$ is strictly inferior to $\frac{3}{4}$ and $\lim_{x\to+\infty} F(x) = \frac{3}{4}$. The question we ask is: can $\frac{\hat{m}_1^2}{\hat{m}_2}$ be superior to $\frac{3}{4}$ for some L ? We will see that $(\hat{\alpha}_{MM}, \hat{\beta}_{MM})$ may not exist for small L.

B. *GGD Parameters Estimator in a ML Framework*

An alternative approach is to consider the log-likelihood (LL) function under independence hypothesis:

$$\mathcal{L}(X_L, \alpha, \beta) = \sum_{i=1}^{L} \log\left(P_{\alpha,\beta}(x_i)\right) \tag{2}$$

and to solve the associated Euler-Lagrange (EL) equations (the derivatives of $\mathcal{L}$ with respect to α and β equal 0) to find the estimators [15]. Given $\hat{\beta}$, this defines a unique estimator $\hat{\alpha} = \left(\frac{\hat{\beta}}{L}\sum_i |x_i|^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}$ (using the derivative with respect to α) while $\hat{\beta}$ satisfies:

$$g(\hat{\beta}) = 1 + \frac{\Psi(\frac{1}{\hat{\beta}})}{\hat{\beta}} - \frac{\sum_{i=1}^{L} |x_i|^{\hat{\beta}} \log|x_i|}{\sum_{i=1}^{L} |x_i|^{\hat{\beta}}} + \frac{\log\left(\frac{\hat{\beta}}{L}\sum_{i=1}^{L} |x_i|^{\hat{\beta}}\right)}{\hat{\beta}} = 0 \tag{3}$$

where [7]:

$$\Psi(x) = \frac{d\log(\Gamma(x))}{dx} = -\gamma - \frac{1}{x} + \sum_{k=1}^{+\infty}\left(\frac{1}{k} - \frac{1}{k+x}\right). \tag{4}$$

This equation is shown to have a unique root in probability (when L tends to infinity) [14] which corresponds to the maximum of the LL function. Therefore for large L, solving the EL equations is the same as finding the maximum of the LL function, at least in probability. However, when L is finite, we show that g, defined in (3), has either no root or at least two roots. In other words, solving the EL equations is no longer equivalent to finding the maximum of the LL function. Indeed, the mathematical study of g leads to the following theorem:

Theorem 1: for any sample $(x_1, \cdots, x_L)$, g satisfies $\lim_{\beta\to 0} g(\beta) = \frac{1}{2}$ and $\lim_{\beta\to+\infty} g(\beta) = 0^+$. Therefore, g has no root or at least two roots.

The proof is given in Appendix A. This brings up the important issue: is it possible to find a ML estimator? We answer this question in the following way: we find a necessary and sufficient condition for the existence of a ML estimator.

III. On the Existence of a ML Estimator

A. A Preliminary Result

Note that we can write:

$$\begin{aligned} \beta g(\beta) &= \left(\beta + \Psi(\frac{1}{\beta}) + \log(\beta)\right) - \left(\frac{\sum_{i=1}^{L} |x_i|^\beta \log |x_i|^\beta}{\sum_{i=1}^{L} |x_i|^\beta} - \log(\frac{1}{L}\sum_{i=1}^{L} |x_i|^\beta) \right) \\ &= \mu(\beta) - \lambda(\beta). \end{aligned} \tag{5}$$

Theorem 2: μ is strictly increasing and maps $]0,+\infty[$ to $]0,+\infty[$ while λ is strictly increasing and maps $]0,+\infty[$ to $]0,\log(\frac{L}{\#I})[$, where I is the set of indices such that $|x_i| = \max_{j=1,\ldots,L} |x_j|$, for $i \in I$ and $\#I$ is the cardinal of I.

The proof is given in Appendix B.

B. Necessary and Sufficient Condition for the Existence of a ML Estimator

The LL function (2) with fixed β has a unique maximum at $\hat{\alpha}_{ML}(\beta) = \left(\frac{\beta}{L}\sum_{i=1}^{L}|x_i|^\beta\right)^{\frac{1}{\beta}}$. Therefore, we study

$$h(\beta) = \mathcal{L}\left(X_L, \hat{\alpha}_{ML}(\beta), \beta\right) + L\log(2) = Lu(\beta), \tag{6}$$

where the link between u and g is $u'(\beta) = \frac{g(\beta)}{\beta}$, to prove:

Theorem 3: a ML estimator exists if and only if there exists β such that $u(\beta) = -\log(M)$ where M is the maximum of the absolute value of the x_i.

The proof is given in Appendix C.

C. Practical Determination of the Existence of a ML Estimator

The result given in Theorem 3 needs to be reinterpreted to be exploited: we show that it is equivalent to the convergence of a specific sequence. Let us consider $f_1(\beta) = (1-\frac{1}{\beta})\log(\beta) - \log\left(\Gamma(\frac{1}{\beta})\right) - \frac{1}{\beta}$ and $f_2(\beta) = \frac{1}{\beta}\log\left(\frac{1}{L}\sum_{i=1}^{L}|x_i|^\beta\right) - \log(M)$ where $f_1(\beta) - f_2(\beta) = u(\beta) + \log(M)$. The condition given in Theorem 3 amounts to solving $f_1(\beta) = f_2(\beta)$ which is equivalent to $\beta = f_1^{-1}(f_2(\beta))$ since $f_1'(\beta) = \frac{1}{\beta^2}\mu(\beta) > 0$ (cf Theorem 2). We can now reformulate Theorem 3.

Theorem 4: The existence of β such that $u(\beta) = -\log(M)$ is equivalent to the convergence of

$$\begin{aligned} \beta_0 &= f_1^{-1}\left(\frac{1}{L}\sum_{i=1}^{L}\log|x_i| - \log(M)\right) \\ \beta_{n+1} &= f_1^{-1}\left(f_2(\beta_n)\right). \end{aligned} \tag{7}$$

PROOF: We first prove that β_n exists and is increasing. First, note that f_1 maps $]0,+\infty[$ to $]-\infty,0[$, using the limits computed in Appendix C. One then shows that $f_2'(\beta) = \frac{1}{\beta^2}\lambda(\beta) > 0$ (cf Theorem 2) and, using Appendix C, we see that f_2 maps $]0,+\infty[$ to $]\frac{1}{L}\sum_{i=1}^{L}\log|x_i| - \log(M), 0[$. From this, we deduce that β_n, $n \geq 0$, exists and $\beta_1 = f_1^{-1}(f_2(\beta_0)) > f_1^{-1}(\frac{1}{L}\sum_{i=1}^{L}\log|x_i| - \log(M)) = \beta_0$. Then, by induction and since $f_1^{-1} o f_2$ is strictly increasing, $\beta_{n+1} > \beta_n$ for all $n \geq 0$. If β_n converges the limit l satisfies $l = f_1^{-1} o f_2(l) \Leftrightarrow f_1(l) = f_2(l)$ that is $u(l) = -\log(M)$, i.e. the LL function has a global maximum. Conversely, if there exists β such that $u(\beta) = -\log(M)$, i.e. $f_1(\beta) = f_2(\beta)$, one shows by induction that $\beta > \beta_n$ for all $n \geq 0$, which leads to the convergence of β_n (β_n is increasing and smaller than β) $\blacksquare$.

D. *Lower and Upper Bounds for ML estimators*

We determine lower and upper bounds for a ML estimator $\hat{\beta}_{ML}$ when it exists. First, note that when β_n, defined in (7), converges, its limit is the smallest β, denoted β_{min}, satisfying $u(\beta) = -\log(M)$. Now, we determine an upper bound for $\hat{\beta}_{ML}$. As $\hat{\beta}_{ML}$ is a root of the EL equation (3), $\mu(\hat{\beta}_{ML}) = \lambda(\hat{\beta}_{ML}) < \log(\frac{L}{\#I})$ (cf Theorem 2) and, as μ is strictly increasing, $\hat{\beta}_{ML} < \mu^{-1}(\log(\frac{L}{\#I})) = \beta_0'$. β_0' is such that $g(\beta_0') > 0$ because otherwise, since $\lim_{\beta \to +\infty} g(\beta) = 0^+$, there is no root to (3) in $[\beta_0'; +\infty[$. We also note that $g(\beta) > 0$ on $[\beta_0'; +\infty[$, giving us that $u(\beta) + \log(M)$ increases to 0 on $[\beta_0', +\infty[$. In particular, $u(\beta_0') < -\log(M)$ which can be written $f_1(\beta_0') < f_2(\beta_0')$. If we assume that there exists β such that $u(\beta) = -\log(M)$, we can write $\beta_0' > \beta \Rightarrow f_1(\beta_0') > f_1(\beta) = f_2(\beta) \Rightarrow f_1(\beta_0') \in f_2(]0,+\infty[)$. We can then define $\beta_1' = f_2^{-1} o f_1(\beta_0') < \beta_0'$ and, by induction, $\beta_{n+1}' = f_2^{-1} o f_1(\beta_n')$ which is convergent (decreasing and positive). One remarks that $\beta_n' > \beta_{max}$, where β_{max} is the greatest value of β such that $u(\beta) = -\log(M)$ and that β_n' converges to β_{max}. β_n and β_n' enable to define a union of intervals $[0, \beta_{min}[\bigcup]\beta_{max}, +\infty[$, on which $u(\beta) < -\log(M)$, therefore the maximum of u is inside $[\beta_{min}, \beta_{max}]$ (note that δ and Δ used in the proof of Theorem 3 correspond to β_{min} and β_{max} respectively).

IV. ALGORITHMS FOR THE EXISTENCE AND THE COMPUTATION OF $\hat{\beta}_{ML}$

A. *Existence of $\hat{\beta}_{ML}$, Computation of β_{min} and β_{max}*

When β_n is convergent, its limit β_{min} is the smallest value such that $f_1(\beta) = f_2(\beta)$. On the contrary, the sequence is divergent when $\beta_n > \beta_0'$ for some n (see section III. D for the definition of β_0') as $\hat{\beta}_{ML} < \beta_0'$ when it exists. The algorithm for the existence of $\hat{\beta}_{ML}$ and then for the computation of β_{min}

is the following:

$n = 0;$ %*we initialize with* β_0

while $(f_2(\beta_n + \epsilon) > f_1(\beta_n + \epsilon)$ and $\beta_n < \beta'_0)$

$\quad \beta_{n+1} = f_1^{-1} o f_2(\beta_n); \quad n = n + 1$

end

if $(\beta_n > \beta'_0)$

$\quad \beta_n$ is divergent , $\hat{\beta}_{ML}$ does not exist

else

$\quad \beta_n$ is convergent , $\beta_{min} = \beta_n$

end

The algorithm to compute β_{max} is similar except that when $\hat{\beta}_{ML}$ exists, $\beta'_{n+1} = f_2^{-1} o f_1(\beta'_n)$ is known to be convergent (see section III.D).

B. Computation of $\hat{\beta}_{ML}$

Once the existence of $\hat{\beta}_{ML}$ has been proved, assuming the uniqueness of the ML estimator (which we do not discuss here), one can compute:

$$\hat{\beta}_{ML} = \underset{\beta \in [\beta_{min}, \beta_{max}]}{argmax} \; u(\beta). \tag{8}$$

The estimation of $\hat{\beta}_{ML}$ with formula (8) is computationally expensive for two main reasons: first, we numerically notice that β_{max} is increasing with L which makes the algorithm time-consuming if we require a fixed precision on $\hat{\beta}_{ML}$ and, second, the computation of β_{max} is itself time-consuming. We develop an algorithm to compute $\hat{\beta}_{ML}$ that avoids the computation of β_{max}. The study of the distribution function associated with the probability density function $P_{\alpha,\beta}$ (see (1)), with fixed α, shows a fast convergence to the uniform distribution function on $[-\alpha, \alpha]$. The interval for β on which the law differs significantly from the uniform distribution is $[0, 4]$. In equation (8), no hypothesis is made on the localization of the global maximum of u. We numerically notice (see below for the details on simulations) that, for the relevant range for β, the smallest value β_1 such that $g(\beta) = 0$ is such that $u(\beta_1)$ is the global maximum of u. Note that we already know it is a local maximum since $\lim_{\beta \to 0} g(\beta) = \frac{1}{2}$ implies $g'(\beta_1) < 0$ and consequently $u''(\beta_1) = \frac{1}{\beta_1} g'(\beta_1) - \frac{g(\beta_1)}{\beta_1^2} < 0$. For the simulations, we proceed this way: given α and β, we build the corresponding distribution function, then we generate a $N \times N$ bidimensional sample with parameters (α, β) by applying the inverse of the distribution function to a $N \times N$ bidimensional uniform noise. We compute β_1 by Newton iterations on g (with precision ϵ_{β_1} to stop the iterations) and starting from β_{min}. We also compute $\beta_2 = \underset{\beta\in[\beta_{min},\beta_{max}]}{argmax} \; u(\beta)$ where we sample $[\beta_{min}, \beta_{max}]$ at sampling frequency ϵ_{β_2}. This implies that if $|\beta_1 - \beta_2| < \epsilon_{\beta_1} + \epsilon_{\beta_2}$ then β_1 and β_2 should be considered equal. We numerically found that, for $\beta \in \{0.5, 1, 2, 3, 4\}$, for $\alpha \in \{1, 10, 100\}$, for $N \in \{8, 16, 32\}$, and for 200 samples for each (α, β, N), when $\hat{\beta}_{ML}$ exists, β_1 and β_2 are equal (we took $\epsilon_{\beta_1} = 10^{-3}, \epsilon_{\beta_2} = 10^{-2}$, and we found $|\beta_1 - \beta_2| < 6.10^{-3}$ in every case). This leads to the simple algorithm to compute $\hat{\beta}_{ML}$:

$\beta = \beta_{min}$

while $(g(\beta + \epsilon_\beta) > 0)$

$\quad \beta = \beta - \frac{g(\beta)}{g'(\beta)}$;

end

$\hat{\beta}_{ML} = \beta$.

The method we developed for the existence and the computation of $\hat{\beta}_{ML}$ is called ML_2 in the following.

V. Results

A. On the Existence of a ML Estimator and Sample Size

We now compare ML_2 to the method based on ML estimation proposed in [15] (ML_1) which assumes the existence of the ML estimator. We focus on the estimation of $\hat{\beta}_{ML}$ since, when this parameter exists, $\hat{\alpha}_{ML}$ is uniquely defined. ML_1 is based on Newton iterations on g starting from $\hat{\beta}_{MM} = F^{-1}(\frac{\hat{m}_1^2}{\hat{m}_2})$ (see section II.A for details) and computes $\hat{\beta}_{ML_1}$ which is not necessarily equal to $\hat{\beta}_{ML}$ since it is based on the study of g, not u (see [15] for details). Second, this algorithm fails to provide a solution in the following cases: $\hat{\beta}_{MM}$ does not exist, $g(\beta) > 0$ for all β and $\hat{\beta}_{MM}$ exists but g' changes signs in its neighborhood (the Newton iterations do not converge). We display, in Table I, the percentage of occurrences where each algorithm (MM, ML_1 and ML_2) fails to provide a solution. For each (α, β, N), the percentage is computed over 200 $N \times N$ samples built in the same way as in section IV (we choose a dyadic N because images of dyadic sizes are very often used in image processing). Let us first say that for $N \geq 16$ and $\beta \leq 2$, any of the three method provides a solution. The results deteriorate for ML_1 on small samples and for large β since ML_1 first requires that MM provides a solution which explains why the results are worse for ML_1 than for MM and second since even when MM provides a solution, g may not have a root or the Newton iterations may not converge. For $N = 8$, We report instances where the three methods do not provide a solution. The main difference between MM and ML_2 on the one hand and ML_1 on the other is that ML_1 computes the solution without knowing its existence. For $N = 8$ and $\beta \geq 2$ and also for $N = 16$ and large β, the use of ML_1 should be prohibited for that reason.

A last point we investigate is the relation between $\hat{\beta}_{ML_1}$ and $\hat{\beta}_{ML}$ when they exist. For each (α, β, N), we compute the maximum of $D = |\hat{\beta}_{ML_1} - \hat{\beta}_{ML}|$ (over 200 samples in each case). As we use the same precision ϵ for the computation of $\hat{\beta}_{ML}$ and $\hat{\beta}_{ML_1}$, they are considered equal if $D < 2\epsilon$. The results are as follows: if $N = 8$ and $\beta \leq 1$ or if $N = 16$ and $\beta \leq 3$, ML_1 and ML_2 both provide a solution and these solutions are equal. When these conditions are not fulfilled, one had rather check the existence of the solution before computing it.

			$\beta = 0.5$	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$
$\alpha = 1$	$N = 8$	MM	0	0	0	6.5	13.5
		ML_1	0	0	0.5	8.5	20.5
		ML_2	0	0	0.5	4	11.5
	$N = 16$	MM	0	0	0	0	1
		ML_1	0	0	0	0	2.5
		ML_2	0	0	0	0	0
$\alpha = 10$	$N = 8$	MM	0	0	0	7.5	17.5
		ML_1	0	0	0.5	10.5	24.5
		ML_2	0	0	0.5	6.5	7
	$N = 16$	MM	0	0	0	0	0.5
		ML_1	0	0	0	0	2
		ML_2	0	0	0	0	0
$\alpha = 100$	$N = 8$	MM	0	0	0.5	5.5	17.5
		ML_1	0	0	1.5	10.5	18.50
		ML_2	0	0	1.5	3.5	13.5
	$N = 16$	MM	0	0	0	0	1
		ML_1	0	0	0	0	4
		ML_2	0	0	0	0	0

TABLE I

COMPUTATION OF THE PERCENTAGE OF OCCURRENCES WHERE EACH ALGORITHM FAILS TO PROVIDE A SOLUTION

B. Results on the Computational Cost

We now study the computational cost of the algorithms: for ML_1, we calculate the average number of Newton iterations to compute $\hat{\beta}_{ML_1}$ (see I_{ML_1} in Table II), while, for ML_2, we compute the average number of iterations to calculate β_{min} (see $I^1_{ML_2}$ in Table II), the average number of iterations to prove the divergence of β_n (see $I^2_{ML_2}$ in Table II) and, finally, the average number of Newton iterations to compute $\hat{\beta}_{ML}$ (see $I^3_{ML_2}$ in Table II). The computation involves 200 samples for each (α, β, N), with fixed N, which corresponds to a total of 2400 samples (NB: the results are rounded to the nearest integer). We note that the existence of $\hat{\beta}_{ML}$ is proved with very few iterations but when it does not exist, finding n such that $\beta_n > \beta'_0$ may be expensive since if $N = 8$, $\beta'_0 = 112$ (i.e. $\mu^{-1}(2\log(8))$, see the algorithm in section IV). However, for $N \geq 16$, the probability that $\hat{\beta}_{ML}$ does not exist is very low (cf Table I) which implies that the algorithm to prove the divergence of β_n is very seldom applied. When $\hat{\beta}_{ML}$ and $\hat{\beta}_{ML_1}$ exist, the computational cost of ML_1 and ML_2 is of the same order.

	$N = 8$	$N = 16$
$I^1_{ML_1}$	2	2
$I^1_{ML_2}$	3	1
$I^2_{ML_2}$	49	$\times$
$I^3_{ML_2}$	4	4

TABLE II

FIRST ROW: AVERAGE NUMBER OF NEWTON ITERATIONS TO COMPUTE $\hat{\beta}_{ML_1}$,SECOND ROW: AVERAGE NUMBER OF ITERATIONS TO COMPUTE $\beta_{\min}$,THIRD ROW: AVERAGE NUMBER OF ITERATIONS TO PROVE THE DIVERGENCE OF β_n ($\times$ MEANS THE SEQUENCE β_n ALWAYS CONVERGES), FOURTH ROW: AVERAGE NUMBER OF ITERATIONS TO COMPUTE $\hat{\beta}_{ML}$

C. On the Modeling of the Distribution of Subband Coefficients and Sample Size

The modeling of the distribution of subband coefficients with GGD models have been extensively used [3] [13] [15] [9] [12]. The problem of the size of the sample is particularly crucial for orthogonal wavelet transform subbands since the number of subband coefficients is divided by 2 in each direction from one scale to another (see [6] for details on orthogonal wavelet transforms). In [15], it is found that wavelet transform subbands associated with the decomposition of natural images may lead to $2 \leq \beta \leq 3$, which is a typical range for β where the computation must not be carried out on 8×8 samples. Our study enables to say that from the point of view of the existence of the ML parameter $\hat{\beta}_{ML}$, its computation should not be carried out on samples of dyadic size smaller than 16×16. Note that, in texture classification problems, such a subband size was already brought about from the point of view of the robustness both of GGD parameters [15] and of the energy of subband coefficients [5].

VI. Conclusion

The goal of this paper was to study the validity of the *generalized Gaussian density* (GGD) model under small sample situations. We proposed a necessary and sufficient condition for the existence of GGD parameters in a ML framework. We then tested our criterion on random images of different sizes to put forward that GGD models may fail to characterize small sample distributions (8×8 pixels image and $\beta \geq 2$). We showed that proving the existence of the ML estimator avoids algorithmic problem at a reasonable computational cost. We also developed a new method for the estimation of GGD parameters in a ML framework when they exist. We compared our results to those given by the algorithm proposed in [15]: we reported the instances where both algorithms behave similarly and when our method should be preferred. One important theoretical point that remains to be investigated is under which conditions when the ML estimator exists it is unique.

VII. Acknowledgments

This work is partially supported by the European project IHP "Breaking complexity" HPRN-CT-2002-00286 and by the "Région Rhône-Alpes".

Appendix A

Proof of Theorem 1: Let us first remark that $g(\beta)$ has the same roots as $\beta g(\beta)$ ($\beta > 0$) which can be written $\beta g(\beta) = \mu(\beta) - \lambda(\beta)$ as in (5). We then study separately μ and λ. We use, in the following, Landau notations: $o(z)$ means negligible compared to z while $O(z)$ means of the same order as z.

Study of the function μ

In the neighborhood of 0: we use the asymptotic development of Ψ [1] ($z \to +\infty$), $\Psi(z) = \log(z) - \frac{1}{2z} + O(\frac{1}{z^2})$, to deduce that when β tends to 0 $\mu(\beta) = \frac{1}{2}\beta + O(\beta^2)$. So, $\mu(\beta) \underset{0}{\sim} \frac{1}{2}\beta$ and $\lim\limits_{\beta \to 0^+} \mu(\beta) = 0$.

In the neighborhood of $+\infty$: in the neighborhood of 0, we have, using (4), $\Psi(z) = -\frac{1}{z} - \gamma + \sum\limits_{k=1}^{\infty} \frac{z}{k^2}\left(1 - \frac{z}{k} + o(z)\right)$ $= -\frac{1}{z} - \gamma + z\frac{\pi^2}{6} + o(z)$. So, in the neighborhood of $+\infty$, $\mu(\beta) = -\gamma + \log(\beta) + o(1) \underset{+\infty}{\sim} \log(\beta)$ from which we deduce $\lim\limits_{\beta \to +\infty} \mu(\beta) = +\infty$.

Study of the function λ

In the neighborhood of 0: for all i, we have $\lim\limits_{\beta \to 0^+} |x_i|^\beta = 1$ and $\lim\limits_{\beta \to 0^+} \lambda(\beta) = 0$. Indeed, a Taylor expansion at zero leads to $|x_i|^\beta = e^{\beta \log |x_i|} = 1 + \beta \log |x_i| + o(\beta)$. Thus, $\lambda(\beta) = \frac{\beta \sum_{i=1}^{L} \log |x_i| (1 + o(1))}{L + o(1)} - \log\left(\frac{1}{L}(L + \beta \sum\limits_{i=1}^{L} \log |x_i| + o(\beta))\right) = o(\beta)$ which implies $\lim\limits_{\beta \to 0^+} \lambda(\beta) = 0$.

In the neighborhood of $+\infty$: Let $M = \max|x_i|$ and let I be the set of indices such that $|x_i| = M$ for $i \in I$. Let us note $\#I$ the cardinal of I. Then, we can write

$$\begin{aligned}\lambda(\beta) &= \frac{\#I\beta\log(M) + \sum_{i\notin I}\left(\frac{|x_i|}{M}\right)^{\beta}\beta\log|x_i|}{\#I + \sum_{i\notin I}\left(\frac{|x_i|}{M}\right)^{\beta}} - \log\left(\frac{M^\beta}{L}(\#I + \sum_{i\notin I}\left(\frac{|x_i|}{M}\right)^{\beta})\right)\\ &= \log\left(\frac{L}{\#I}\right) + o(1) \underset{\beta\to\infty}{\longrightarrow} \log\left(\frac{L}{\#I}\right).\end{aligned}$$

From this study, we conclude that as $g(\beta) = \frac{1}{\beta}(\mu(\beta) - \lambda(\beta))$, $\lim\limits_{\beta\to 0^+} g(\beta) = \frac{1}{2}$ and as, in the neighborhood of $+\infty$, $g(\beta) = \frac{1}{\beta}\left(\log(\beta) - \log(\frac{L}{\#I}) + o(1)\right) = \frac{\log(\beta)}{\beta} + O\left(\frac{1}{\beta}\right)$, $\lim\limits_{\beta\to+\infty} g(\beta) = 0$ and $g(\beta) > 0$. Consequently, g has either no root or at least two roots.

APPENDIX B

PROOF OF THEOREM 2:

Study of the variations of the function μ: let us consider $\rho(\beta) = \mu(\frac{1}{\beta})$ whose derivative is equal to $\rho'(\beta) = \mu'(\frac{1}{\beta}) \times \left(-\frac{1}{\beta^2}\right)$. Then $\rho'(z) = -\frac{1}{z^2} + \Psi'(z) - \frac{1}{z}$ with $\Psi'(z) = \sum_{k=0}^{+\infty} \frac{1}{(k+z)^2}$ [7]. As $\int_k^{k+1} \frac{dt}{(t+z)^2} = \frac{1}{(k+z)^2} + 2\int_k^{k+1} \frac{t-k-1}{(t+z)^3}dt$ (integration by parts), we can write

$$\Psi'(z) = \sum_{k=0}^{+\infty}\int_k^{k+1}\frac{dt}{(t+z)^2} - 2\sum_{k=0}^{+\infty}\int_k^{k+1}\frac{t-k-1}{(t+z)^3}dt = \frac{1}{z} - 2\sum_{k=0}^{+\infty}\int_k^{k+1}\frac{t-k-1}{(t+z)^3}dt.$$

This leads to

$$\rho'(z) = -\frac{1}{z^2} - 2\sum_{k=0}^{+\infty}\int_k^{k+1}\frac{t-k-1}{(t+z)^3}dt.$$

Since $\frac{1}{z^2} = \int_0^{+\infty}\frac{2}{(t+z)^3}dt = 2\sum_{k=0}^{+\infty}\int_k^{k+1}\frac{dt}{(t+z)^3}$,

$$\rho'(z) = -2\sum_{k=0}^{+\infty}\left(\int_k^{k+1}\frac{dt}{(t+z)^3} + \int_k^{k+1}\frac{t-k-1}{(t+z)^3}dt\right) = -2\sum_{k=0}^{+\infty}\int_k^{k+1}\frac{t-k}{(t+z)^3}dt.$$

As $\frac{t-k}{(t+z)^3} > 0$ for $t \in]k, k+1]$, we deduce that $\rho'(z) < 0$. This leads to $\mu'(\beta) > 0$ for all $\beta > 0$. Thus, μ is strictly increasing on $]0, +\infty[$, and maps $]0, +\infty[$ to $]0, +\infty[$ as $\mu(\beta) \underset{0}{\sim} \frac{1}{2}\beta$ and $\mu(\beta) \underset{+\infty}{\sim} \log(\beta)$ (cf. Appendix A).

Study of the variations of the function λ: the derivative of λ is

$$\lambda'(\beta) = \frac{\sum_{i=1}^{L}\left(|x_i|^\beta\log|x_i|\log(|x_i|^\beta)\right)\left(\sum_{i=1}^{L}|x_i|^\beta\right) - \left(\sum_{i=1}^{L}|x_i|^\beta\log|x_i|\right)\left(\sum_{i=1}^{L}|x_i|^\beta\log|x_i|^\beta\right)}{\left(\sum_{i=1}^{L}|x_i|^\beta\right)^2}.$$

Using the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^{L}|x_i|^{\beta}\log|x_i|\right)^2=\left(\sum_{i=1}^{L}|x_i|^{\frac{\beta}{2}}(|x_i|^{\frac{\beta}{2}}\log|x_i|)\right)^2\leq\left(\sum_{i=1}^{L}|x_i|^{\beta}\right)\left(\sum_{i=1}^{L}|x_i|^{\beta}(\log|x_i|)^2\right),$$

with equality if all the x_i are equal, we get $\lambda'(\beta)>0$. One can reasonably assume that the x_i are not all equal (we can show that when the x_i are equal the ML estimator does not exist), therefore we conclude that $\lambda(\beta)$ is strictly increasing on $]0,+\infty[$. Finally, since $\lambda(\beta)\underset{0}{\sim}o(\beta)$ and $\lambda(\beta)\underset{+\infty}{\sim}\log(\frac{L}{\#I})$ (cf. Appendix A), λ maps $]0,+\infty[$ on $]0,\log(\frac{L}{\#I})[$.

APPENDIX C

PROOF OF THEOREM 3: We study the limits and the variations of u in the neighborhood of 0 and of $+\infty$ to conclude that there exists an interval on which the ML estimator necessarily is.

Study of the function u

In the neighborhood of 0: let us write $|x_i|^{\beta}=e^{\beta\log|x_i|}=1+\beta\log|x_i|+o(\beta)$.This implies that

$$\frac{1}{\beta}\log(\frac{1}{L}\sum_{i=1}^{L}|x_i|^{\beta})=\frac{1}{\beta}\log\left(1+\frac{\beta}{L}\sum_{i=1}^{L}\log|x_i|+o(\beta)\right)\underset{\beta\to0^+}{\to}\frac{1}{L}\sum_{i=1}^{L}\log|x_i|.$$

Furthermore, the asymptotic development of $\log(\Gamma(z))$ gives us [1] ($z\to+\infty$), $\log(\Gamma(z))-z\log(z)+z+\frac{1}{2}\log z=\frac{1}{2}\log(2\pi)+O(\frac{1}{z})$. Consequently, we have $\log\left(\Gamma(\frac{1}{\beta})\right)=-\frac{1}{\beta}\log(\beta)-\frac{1}{\beta}+\frac{1}{2}\log(\beta)+O(1)$, when $\beta\to0^+$, which leads to

$$u(\beta)=\frac{1}{2}\log(\beta)+O(1)-\frac{1}{\beta}\log\left(\frac{1}{L}\sum_{i=1}^{L}|x_i|^{\beta}\right)\underset{\beta\to0^+}{\to}-\infty.$$

In the neighborhood of $+\infty$: let M and I have the same meaning as in Theorem 3 and 2 respectively. As previously, we can write $\sum_{i=1}^{L}|x_i|^{\beta}=M^{\beta}\left(\#I+\sum_{i\notin I}\left(\frac{|x_i|}{M}\right)^{\beta}\right)$, leading to

$$\frac{1}{\beta}\log\left(\frac{1}{L}\sum_{i=1}^{L}|x_i|^{\beta}\right)=-\frac{1}{\beta}\log(L)+\log M+\frac{1}{\beta}\log\left(\#I+\sum_{i\notin I}\left(\frac{|x_i|}{M}\right)^{\beta}\right)\underset{\beta\to\infty}{\to}\log(M).$$

Let us then recall the "Euler infinite product" [1], $\log\left(\Gamma(z)\right)=-\gamma z-\log(z)+\sum_{k=1}^{+\infty}\left(\frac{z}{k}-\log(1+\frac{z}{k})\right)$, in which the series $\sum_{k=1}^{+\infty}\left(\frac{z}{k}-\log\left(1+\frac{z}{k}\right)\right)$ converges normally which entails the continuity of the sum in $z=0$ and leads to $\lim_{z\to0}\sum_{k=1}^{+\infty}\left(\frac{z}{k}-\log(1+\frac{z}{k})\right)=0$. We then get $\lim_{z\to0}\log(\Gamma(z))+\log(z)=0$ which is equivalent to $\lim_{\beta\to+\infty}\log\left(\Gamma(\frac{1}{\beta})\right)-\log(\beta)=0$ and finally leads to

$$\lim_{\beta\to+\infty}u(\beta)=-\log(M)$$

Variations of u close to 0 and $+\infty$: let us recall that $u'(\beta) = \frac{1}{\beta} g(\beta)$. Using the study of g made in Appendix A, we have that u both increases in the neighborhood of 0 and in the neighborhood of $+\infty$. It follows that u has a global maximum if and only there exists a $\beta > 0$ such that $u(\beta) = -\log(M)$. Indeed, assume that there exists β such that $u(\beta) = -\log(M)$, then as u increases from $-\infty$ in the neighborhood of 0 and decreases from $-\log(M)$ in the neighborhood of $+\infty$ (cf the beginning of the proof), there exist δ and Δ such that $u(\beta) < -\log(M)$ on $]0, \delta[\bigcup]\Delta, +\infty[$. On $[\delta, \Delta]$, the function u is continuous and reaches its maximum, i.e. u has a global maximum. Conversely, assume that u has a global maximum, since $\lim_{\beta \to +\infty} u(\beta) = -\log(M)$, this global maximum is superior or equal to $-\log(M)$, i.e. there exists β such that $u(\beta) = -\log(M)$.

REFERENCES

[1] M. Abramovitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1972.

[2] B. Aiazzi, L. Alparone and S. Baronti, *Estimation Based on entropy Matching for Generalized Gaussian PDF Modeling*, IEEE Signal Processing Letters, vol. 6, pp. 138-140, 1999.

[3] J.F. Aujol, G. Aubert and L. Blanc-Féraud, *Wavelet-Based Level Set Evolution for Classification of Textured Images*, IEEE Trans. Image Processing, vol. 12, pp. 1634-1641, 2003.

[4] K. A. Birney and T. R. Fisher, *On the Modeling of DCT and Subband Image Data for Compression*, IEEE Trans. Image Processing, vol. 4, pp. 186-193, 1995.

[5] T. Chang and C.C.J. Kuo, *Texture Analysis and Classification with Tree Structure Wavelet Transform*, IEEE Trans. Image Processing, vol. 2, pp. 429-441, 1993.

[6] I. Daubechies, *Ten lectures on Wavelets*, SIAM, 1992.

[7] I.S. Gradshteyn and I.M. Ryzhik, *Table of integrals, Series and Products*, Academic Press, A. Jeffrey edition, 1980.

[8] R. L. Joshi and T. R. Fisher, *Comparison of Generalized Gaussian and Laplacian Modeling in DCT Image Coding*, IEEE Signal Processing Letters, vol.2, pp. 81-82, 1995.

[9] S. Mallat, *A Theory for Multiresolution Signal Decomposition: the Wavelet Representation*, IEEE Trans. Patt. Anal. and Machine Intel., vol. 11, pp. 675-693, 1989.

[10] P. Moulin and J. Liu, *Analysis of Multiresolution Image Denoising schemes using Generalized Gaussian and Complexity Prior*, IEEE Trans. Info. Theory, vol. 45, pp. 909-919, 1999.

[11] K. Sharifi and L. Leon-Garcia, *Estimation of Shape Parameter for Generalized Gaussian Density Distributions in Subband Decomposition of Video*, IEEE Trans. Circuits Syst. Video Technol., vol. 5, pp. 52-56,1995.

[12] E.P. Simoncelli and E.H. Adelson, *Noise Removal via Bayesian Wavelet Coring*, in Third Int'l Conf on Image Proc, Lausanne, vol. 1, pp. 379-382, 1996.

[13] G. Van de Wouver, P. Scheunders and D. Van Dyck, *Statistical Texture Characterization from Discrete Wavelet Representation*, IEEE Trans. Image Processing, vol. 8, pp. 592-598, 1999.

[14] M.K. Varanasi and B. Aazhang, *Parametric Generalized Gaussian Density Estimation*, J. Acous. Soc. Amer., vol. 86, pp. 1404-1415,1989.

[15] M. N. Do and M. Vetterli, *Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Liebler Distance*, IEEE Trans. on Image Processing, vol. 11, pp. 146-158, 2002.