GENERALIZED CAUCHY DISTRIBUTION BASED ROBUST ESTIMATION

Rafael E. Carrillo**, Tuncer C. Aysal*† *and Kenneth E. Barner**

*University of Delaware,†McGill University

ABSTRACT

The generalized Cauchy distribution (GCD) family has the property that its pdf has closed form for the whole family and also has algebraic tails which makes it suitable to model many impulsive processes in real life. In this paper we propose a robust M-type estimator based on the pdf of the GCD family. Robustness and properties of the new statistics are analyzed and it is noticed that this estimator provide desired characteristics in robust signal processing applications involving non-Gaussian heavy-tailed models. Simulations of the filtering method are performed to evaluate and compare the proposed filtering structure performance to other classic and robust estimators.

***Index Terms*—** maximum likelihood estimation, nonlinear filters

1. INTRODUCTION

Robust statistics is the stability theory of statistical procedures. It systematically investigates the effects of deviations from modelling assumptions on unknown procedures and, if necessary, develops new, better procedures. Robust nonlinear estimators are critical for applications in real situations involving impulsive processes (e.g. ocean acoustic noise, atmospheric interference in LF and VLF communications and multiple access interference in wireless system communications), where heavy-tailed non-Gaussian distributions model the signal [1].

M-estimators, which were developed in the theory of robust statistics [2], have been of great importance in the development of *robust signal processing techniques* [3]. M-estimators can be described by a cost function $\rho(u)$ (posing an optimization problem) or by its first derivative, $\psi(u)$ (yielding an (set of) implicit equation(s)), which is proportional to the influence function. In the location case properties of ψ describe how robust the estimator is. Maximum likelihood location estimates form a special case of M-estimators, with the observations being independent and identically distributed and $\rho(u) = -\log f(u)$, where $f(u)$ is the common density of the samples.

The α-Stable density family has gained recent popularity in addressing heavytailed problems. Unfortunately, the Cauchy distribution is the only algebraictailed α-Stable distribution that possesses a closed form expression, limiting the flexibility and performance of methods derived from this family of distributions. In this paper the maximum likelihood (ML) estimate of location is derived for the GCD family and then extending the associated norm as an M-type estimator. The fact that M–GC estimator is likelihood–based guarantees that the estimate is, at least asymptotically, unbiased consistent and efficient in GCD statistics [2]. Robustness and properties of the cost function are analyzed and it is noticed that this estimator provide desired characteristics in robust signal processing applications involving non-Gaussian heavy-tailed processes.

This work was supported in part by NSF under grant 0728904.

2. M–ESTIMATION AND GCD

The generalized Cauchy distribution family was proposed by Miller and Thomas in 1972 and has been used in several studies of impulsive radio noise [1]. The PDF of the GCD is given by

$$f(x) = a\sigma(\sigma^p + |x|^p)^{-2/p}$$

with $a = p\Gamma(2/p)/2(\Gamma(1/p))^2$. In this representation, σ is the scale parameter and p is the tail constant. The GCD family contains the Meridian [4] and Cauchy distributions as special cases with $p = 1$ and $p = 2$ respectively. For $p < 2$, the tail of the PDF decays slower than in the Cauchy distribution, resulting in a heavier-tailed PDF.

In the M-estimation theory we want to estimate a deterministic but unknown parameter θ of a real-valued signal $s(i;\theta)$ corrupted by additive noise from a set of noisy observations $\{x(i)\}_{i=1}^N$. M–estimate is given by the solution to an optimization problem

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{N} \rho(x(i) - s(i;\theta))$$

or by an implicit equation

$$\sum_{i=1}^{N} \psi(x(i) - s(i;\hat{\theta})) = 0$$

where ρ is an arbitrary cost function to be designed, Θ is the solution space, and $\psi(x) = (\partial/\partial\theta)\rho(x)$.

Since we are also interested in filtering of corrupted signals, and noting that location estimators have been used successfully as moving window based filters (yielding significant performance improvements over traditional robust methods under algebraic–tailed environments [1, 4]), we focus to the location estimation problem in the following.

Theorem 1 *Consider a set of N i.i.d. observations each obeying the GCD with scale parameter σ and tail constant p. The ML estimate of location is given by*

$$\hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{N} \log\{\sigma^p + |x(i) - \theta|^p\} \right]. \tag{1}$$

Since the performance of the estimator depends on the objective function derived from the PDF in the following the properties of the objective function are analyzed.

Proposition 1 *Let $Q(\theta) = \sum_{i=1}^{N} \log\{\sigma^p + |x(i) - \theta|^p\}$ denote the objective function (for fixed σ and p) and $\{x[i]\}_{i=1}^{N}$ the order statistics of $\mathbf{x}$. Then the following statements hold.*

1. *$Q(\theta)$ is strictly increasing for $\theta > x[N]$ and strictly decreasing for $\theta < x[1]$.*
2. *All local extrema of $Q(\theta)$ lie in the interval $[x[1], x[N]]$.*
3. *If $0 < p \leq 1$, then the local minimas are the input samples. If $1 < p \leq 2$, then the objective function has at most $2N - 1$ local extrema points and therefore a finite set of local minima.*
4. *If $0 < p \leq 1$, the solution is one of the input samples(selection type filter). If $1 < p \leq 2$, the solution is one of the local extrema.*

The M–GC estimator has two adjustable parameters, σ and p. The tail constant, p, depends on the heaviness of the tails of the underlying distribution. When $p \leq 1$, the estimator behaves as a selection type filter and as $p \to 0$ it is more robust to heavier tailed distributions. When $p > 1$ the location estimate is in the range of the input samples and can be easily computed. There are two particular cases of the GCD that have recently been studied, which are the estimators for the Cauchy and the Meridian distributions that led to the Myriad and Meridian estimators, respectively [1, 4]. The myriad and meridian estimators have tunable parameters as it's the case in M–GC estimators. The tunable parameter controls the behavior of the objective function and therefore the properties of the estimate. The following corollaries show the behavior of the M–GC estimator when the tunable parameter goes to either 0 or ∞ and more importantly show that it subsumes other classical families of estimators.

Corollary 1 *Given a set of input samples $\{x(i)\}_{i=1}^{N}$, the M–GC estimate converges to the ML GGD estimate (L_p norm as cost function) as $\sigma \to \infty$.*

$$\lim_{\sigma \to \infty} \hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} |x(i) - \theta|^p. \tag{2}$$

Corollary 2 *Given a set of input samples $\{x(i)\}_{i=1}^{N}$, the M–GC estimate converges to a mode type estimator as $\sigma \to 0$. This is*

$$\lim_{\sigma \to 0} \hat{\theta} = \arg \min_{x(j) \in \mathcal{M}} \left[\prod_{i, x(i) \neq x(j)} |x(i) - x(j)| \right] \tag{3}$$

where $\mathcal{M}$ is the set of most repeated values.

The proofs of these corollaries are not included due to space constraints but can be found in [5]. In the case when $p = 1$ and $p = 2$ these results were already presented in [4] and [1] for the meridian and myriad filters. The importance of this limiting behavior is that M–GC estimators include M–estimators with the L_p norm ($0 < p \leq 2$) as cost function which are optimal for the GGD family and also include mode–type estimators which are robust to outlier rejection, providing a wide range of optimality for different distributions.

The M–GC estimator was derived as an ML estimator for the GCD density and thus belong to the class of M–estimators as a more general one, defining the cost function $\rho(x) = \log\{\sigma^p + |x|^p\}$. In M–estimators, the influence function is proportional to $\psi(x)$ if it exists and determines the effect of contamination of the estimator. For the M–GC estimator

$$\psi(x) = \frac{p|x|^{p-1} sgn(x)}{\sigma^p + |x|^p}. \tag{4}$$

It can be noticed that $\lim_{x \to \pm\infty} \psi(x) = 0$, meaning that is asymptotically redescending, i.e., the effect of the errors monotonically decreases as the error increase. Redescendency is a necessary and sufficient condition for outlier rejection [2].

Since M–GC estimates are M–estimates and $\psi(x)$ is odd and bounded, then $\hat{\theta} \to \theta$ as $N \to \infty$ in probability, and are consistent. The variance of $\hat{\theta}$ decreases with increasing N.

Theorem 2 *Consider a set of N i.i.d. observations each obeying the GCD with tail constant p and varying scale parameter $\nu(i) = \sigma/(h(i))^{1/p}$. The ML estimate of location is given by*

$$\hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{N} \log\{\sigma^p + h(i)|x(i) - \theta|^p\} \right]. \tag{5}$$

Theorem 2 can be used to extend the M–GC estimator to a weighted filter structure keeping all the robust properties of the M–GC estimator, the weighted M–GC filter. The M–GC filter can be extended to admit real–valued weights using the sign–coupling approach [4].

3. MULTIPARAMETER ESTIMATION

The location estimate problem defined by the M–GC depends on the parameters σ and p which determine the properties of the location estimate. To solve this problem in a filtering scenario and find the optimal tunable parameters (if the model considered is GCD) we consider multiparameter M–estimates. The idea behind this scheme is to use a small set of samples of the signal to be filtered to estimate σ and p and then use these values through the filtering process.

Let $\{x(i)\}_{i=1}^N$ be a set of independent observations with common GCD with deterministic but unknown parameters θ, σ and p. Let $g(\mathbf{x};\theta,\sigma,p) = \log[f(\mathbf{x};\theta,\sigma,p)]$, then joint estimates are the solutions to the following maximization problem

$$(\hat{\theta},\hat{\sigma},\hat{p}) = \arg\max_{\theta,\sigma,p} g(\mathbf{x};\theta,\sigma,p). \tag{6}$$

The solution to this optimization problem is obtained by solving a set of simultaneous equations. Differentiating the log–likelihood and doing some algebraic manipulations the simultaneous equations are rewritten as

$$\frac{\partial g}{\partial \theta} = \sum_{i=1}^{N} \frac{-p|x(i)-\theta|^{p-1} sgn(x(i)-\theta)}{\sigma^p + |x(i)-\theta|^p} = 0 \tag{7}$$

$$\frac{\partial g}{\partial \sigma} = \sum_{i=1}^{N} \frac{\sigma^p - |x(i)-\theta|^p}{\sigma^p + |x(i)-\theta|^p} = 0 \tag{8}$$

and

$$\frac{\partial g}{\partial p} = \sum_{i=1}^{N} \left[\frac{1}{2p} - \frac{\sigma^p \log\sigma - |x(i)-\theta|^p \log|x(i)-\theta|}{p(\sigma^p - |x(i)-\theta|^p)} \right. \tag{9}$$
$$\left. -\frac{\log\{\sigma^p + |x(i)-\theta|^p\}}{p^2} - \frac{1}{p^2}\Psi\left(\frac{2}{p}\right) + \frac{1}{p^2}\Psi\left(\frac{1}{p}\right)\right] = 0.$$

where $g \equiv g(\mathbf{x};\theta,\sigma,p)$ and $\Psi(x)$ is the digamma function. It can be noticed that (7) is the implicit equation for the M–GC estimator with ψ as defined in (4) so the location estimate has the same properties.

Of note is that $\log f(\mathbf{x};\theta,\sigma,p)$ has a unique maximum in σ for fixed θ and p, and also has a unique maximum for $p \in (0,2]$ for fixed θ and σ [5]. In the following, we provide an algorithm to simultaneously solve the above set of equations.

Flip Flop Algorithm: For a given set of data $\{x(i)\}_{i=1}^N$, we propose to find the optimal joint parameter estimates by the following algorithm with the superscript denoting iteration number.

1. Initialize $\sigma^{(0)}$ and $\theta^{(0)}$.
2. Estimate $\hat{p}^{(m)}$ as the solution of (9).
3. Estimate $\hat{\theta}^{(m)}$ as the solution of (7).
4. Estimate $\hat{\sigma}^{(m)}$ as the solution of (8).
5. Repeat steps 2-4 until $|\hat{\theta}^{(m)} - \hat{\theta}^{(m-1)}| < \epsilon_1$, $|\hat{\sigma}^{(m)} - \hat{\sigma}^{(m-1)}| < \epsilon_2$ and $|\hat{p}^{(m)} - \hat{p}^{(m-1)}| < \epsilon_3$, where $\epsilon_1, \epsilon_2, \epsilon_3$ are small positive numbers.

The algorithm will converge to a local minimum. Experimental results have shown that initializing θ as the median and σ as the MAD and then computing p as a solution for (9) will yield most of the time to the global solution and accelerate the convergence. In the classical literature fixed point algorithms have been successfully used in the computation of M-estimates [1, 2]. Hence, in the following, we solve items 2–4 of the flip–flop algorithm using fixed point algorithms.

Fixed–Point Algorithms: Recall that when $0 < p \leq 1$ the solution can be found as the input sample that minimizes the objective function. We solve (7) for the $1 < p \leq 2$ case in the following. The fixed point recursion for this case can be written as

$$\hat{\theta}_{(j+1)} = \frac{\sum_{i=1}^N w_i(\hat{\theta}_{(j)}) x(i)}{\sum_{i=1}^N w_i(\hat{\theta}_{(j)})} \tag{10}$$

with $w_i(\hat{\theta}_{(j)}) = p|x(i) - \hat{\theta}_{(j)}|^{p-2}/(\sigma^p + |x(i) - \hat{\theta}_{(j)}|^p)$ where the subscript denotes the iteration number. The algorithm converges when $|\hat{\theta}_{(j+1)} - \hat{\theta}_{(j)}| < \delta_1$ where δ_1 is a small number. However since the objective function has more than one extrema point the algorithm can converge to any of these points. In the case when $1 < p \leq 2$, a minima of the objective function, hence, is found using the estimate of the preceding iteration of the flip flop algorithm as initial point for the fixed point search. Since in the first iteration the median was used as first estimate the algorithm converges quickly to a local minima.

Similarly for (8) the recursion can be written as

$$\hat{\sigma}_{(j+1)} = \left(\frac{\sum_{i=1}^N b_i(\hat{\sigma}_{(j)}) x(i)}{\sum_{i=1}^N b_i(\hat{\sigma}_{(j)})}\right)^{\frac{1}{p}} \tag{11}$$

with $b_i(\hat{\sigma}_{(j)}) = 1/(\hat{\sigma}_{(j)}^p + |x(i) - \theta|^p)$. The algorithm converges when $|\hat{\sigma}_{(j+1)} - \hat{\sigma}_{(j)}| < \delta_2$ for δ_2 a small positive number. Since the objective function has only one minimum for fixed θ and p the recursion will converge.

The recursion computing the parameter p is given by

$$\hat{p}_{(j+1)} = \frac{2}{N}\sum_{i=1}^{N}\left[\Psi\left(\frac{2}{\hat{p}_{(j)}}\right) - \Psi\left(\frac{1}{\hat{p}_{(j)}}\right)\right. \tag{12}$$
$$+\log\{\sigma^{\hat{p}_{(j)}} + |x(i)-\theta|^{\hat{p}_{(j)}}\}$$
$$\left.+\frac{\hat{p}_{(j)}(\sigma^{\hat{p}_{(j)}}\log\sigma - |x(i)-\theta|^{\hat{p}_{(j)}}\log|x(i)-\theta|)}{\sigma^{\hat{p}_{(j)}} - |x(i)-\theta|^{\hat{p}_{(j)}}}\right].$$

Noting that the search space is the interval $I = (0,2]$, the function g can be evaluated for a finite set of points $\mathcal{P} \in I$ and look for the value that maximizes g and then take it as the initial point for the search.

4. NUMERICAL RESULTS

Simulations to validate the multiparameter estimation algorithm were carried out and are summarized in Table 1, for $p = 2$, with $\theta = 0$ and $\sigma = 1$ for the original distribution. The experiments used the flip flop algorithm varying the sample length from 10 to 1000. For each block length the experiments were run 1000 times and the presented results are the average on those 1000 trials.

The flip flop algorithm converges in few iterations. Figure 1 depicts the MSE curves for each of the estimates $\hat{\theta}, \hat{\sigma}, \hat{p}$

Table 1. Results for GCD process with $(\theta, \sigma, p) = (0, 1, 2)$.

N	10	100	1000
$\hat{\theta}$	0.0035	-0.0009	-0.0002
MSE	0.0302	2.4889×10^{-3}	1.7812×10^{-4}
$\hat{\sigma}$	0.9563	1.0224	1.0186
MSE	0.0016	1.7663×10^{-5}	1.1911×10^{-6}
$\hat{p}$	1.5816	1.8273	1.9569
MSE	0.0519	0.0109	1.5783×10^{-6}

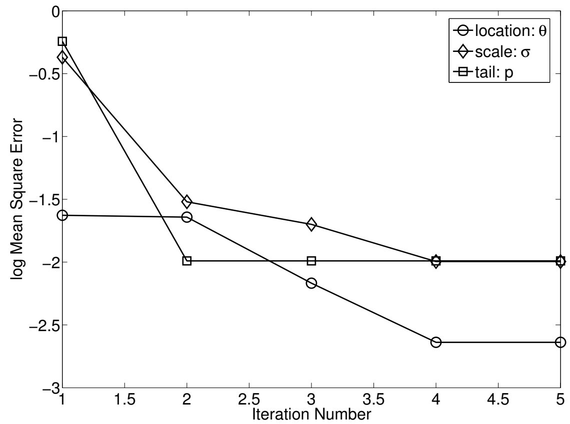

Fig. 1. Evolution of the MSE in 5 iterations of the flip flop algorithm.

for the first five iterations. The experiment was run doing 100 independent realizations of the algorithm with data drawn from the same distribution ($\theta = 0$, $\sigma = 1$ and $p = 1.5$).

The M–GC filter is optimal for GCD noise but is also robust in general impulsive environments. To compare the robustness of the M–GC filter with other robust filtering schemes experiments with the symmetric α–stable distribution were made. The experiment uses a power line communication problem with finite equiprobable alphabet $v = \{-2, -1, 1, 2\}$ and a channel modeled as additive noise. The noise was white, zero–mean, α–stable with $\alpha = 0.4$ and $\gamma = 1$. The filtering process was made using sliding windows of length 9. For the M–GC the resulting p was 0.756 and $\sigma = 0.8963$ using the first 50 samples for training. The M–GC filter was compared to the state–of–the–art robust filters: FLOM, median, myriad and meridian. The results in Fig. 2 show that the M–GCD filter is more robust to impulsive noise than the other filters. The M–GC filter benefits from the selection of the scale and tail parameters and therefore perform at least as good as the myriad and meridian filters in heavy-tailed environments and although is not optimal for α–stable environments it performs as good as the FLOM filter, which it is.

5. CONCLUSIONS

This paper derives the ML location estimate for the Generalized Cauchy distribution and extends its likelihood function as an M–estimator. Noting that the meridian and myriad filters come from ML estimates for the Meridian and Cauchy distributions, a more general filtering structure is proposed based on the maximum likelihood estimate of the GCD. Properties of the cost function of the M–GC estimator are derived and the robustness of the M–GC estimator is analyzed through its influence function. A robust filtering structure based on the M–GC estimator is proposed and its extension to admit real valued weights is presented. Methods to adjust the scale and tail parameters are proposed and evaluated. The M–GC filter offers a robust structure for signal processing in heavy tailed signal models and has the advantage of having the scale and tail parameter to adjust to the signal environment.

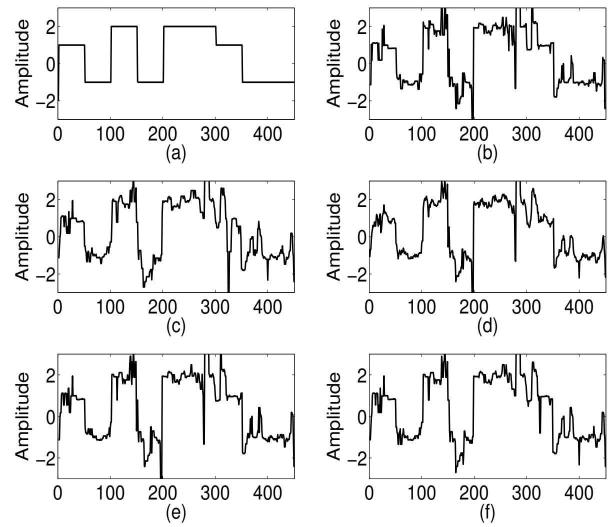

Fig. 2. Power line communication enhancement. Filtering results with: (a) Original, (b) FLOM (c) Median, (d) Myriad, (e) Meridian, (f) M–GC.

6. REFERENCES

[1] G. R. Arce, *Nonlinear Signal Processing: A Statistical Approach*, John Wiley & Sons, Inc., 2005.

[2] P. J. Huber, *Robust Statistics*, John Wiley & Sons, Inc., 1981.

[3] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing," *Proceedings of IEEE*, vol. 73, Mar. 1985.

[4] T. C. Aysal and K. E. Barner, "Meridian filtering for robust signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 3949–3962, Aug. 2007.

[5] T. C. Aysal, R. E. Carrillo, and K. E. Barner, "Generalized cauchy density based estimation for statistical signal processing," 2008, to be submitted.