A Theoretical Framework for Problems Requiring Robust Behavior

Rafael E. Carrillo, Tuncer C. Aysal and Kenneth E. Barner
Department of Electrical and Computer Engineering
University of Delaware, Newark, DE 19716
Email: {carrillo,aysal,barner}@ee.udel.edu

***Abstract*—This paper develops a generalized Cauchy density (GCD) based theoretical approach that allows the formulation of challenging problems in a robust fashion. The proposed framework subsumes the generalized Gaussian distribution (GGD) family based developments, thereby guaranteeing performance improvements over traditional problem formulation techniques. This robust framework can be adapted to a variety of applications in signal processing. We formulate two particular applications under this framework in this paper: 1) Robust reconstruction methods for compressed sensing and 2) robust estimation in sensor networks with noisy channels.**

***Index Terms*—Maximum likelihood estimation, nonlinear filtering and estimation, impulsive noise, compressed sensing, sensor networks.**

I. Introduction

Robust statistics, the stability theory of statistical procedures, systematically investigates the effects of deviations from modeling assumptions on unknown procedures and, if necessary, develops more effective procedures [1]. Maximum likelihood (ML) type estimators (M-estimators), which were developed in the theory of robust statistics, have been of great importance in the development of *robust signal processing techniques* [2]. M-estimators can be described by a cost function $\rho(z)$ (posing an optimization problem) or by its first derivative, $\psi(z)$ (yielding an (set of) implicit equation(s)), which is proportional to the influence function. In the location estimation case, properties of ψ describe how robust the estimator is [1]. ML location estimates form a special case of M-estimators, with the observations being independent and identically distributed and $\rho(z) = -\log f(z)$, where $f(z)$ is the common density of the samples.

Signal processing methods derived from the generalized Gaussian distribution (GGD) are very popular in the literature to address heavy-tailed process [3]. However, a constraint of the GGD model is that the distribution tails decay exponentially rather than algebraically. The α-stable density family has gained recent popularity in addressing heavy-tailed problems. Unfortunately, the Cauchy distribution is the only algebraic-tailed α-stable distribution that possesses a closed form expression, limiting the flexibility and performance of methods derived from this family of distributions. The GCD combines the advantages of the GGD and α-stable distributions in that it possesses heavy algebraic tails (like α-stable distributions) and closed form expressions (like the GGD) across a flexible family of densities.

In this paper, we develop a GCD based theoretical approach that allows the formulation of challenging problems in a robust fashion. The proposed framework, thanks to its flexibility with its parameters (that can be adaptively calculated from the input data [4]), subsumes the GGD based developments (*e.g*, least squares, least absolute deviation, fractional low order moments (FLOM), L_p norms, k-means clustering, etc.), thereby guaranteeing performance improvements over traditional problem formulation techniques. Applications that can be formulated using this framework include among others: robust regression, robust detection and estimation, clustering in impulsive environments, spectrum sensing when signals are corrupted by heavy-tailed noise and robust compressed sensing (CS) and reconstruction methods. Particulary, in this paper, we formulate two applications under this framework: 1) robust CS reconstruction methods and 2) robust estimation over noisy channels for sensor networks.

II. Robust Parameter Estimation

The GCD family was proposed by Rider in 1957 [5] and rediscovered by Miller and Thomas in 1972 with a different parametrization [6] and has been used in several studies of impulsive radio noise [3]. The PDF of the GCD is given by

$$f(z) = a\sigma(\sigma^p + |z|^p)^{-\frac{2}{p}} \tag{1}$$

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 [7] 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.

A. Location Estimation

The ML estimate of *location* for GCD samples and the properties of the resulting operator, referred to as the M-GC estimator, as a robust estimator are recently derived [4]. Since the M-GC estimator is the ML estimator for the GCD family, it belongs to the class of M-estimators, defining the cost function $\rho(x) = \log\{\sigma^p + |x|^p\}$. We adopt the following notation throughout the section: $x = [x_1, \ldots, x_n]$ is a vector of observations , θ is the common location parameter of the observations, and $h = [h_1, \ldots, h_n]$ a vector of real valued weights. The M-GC estimator can be extended to accept real-valued weights yielding the following formulation.

978-1-4244-5180-7/09/$26.00 ©2009 IEEE

Definition 1 *The weighted M-GC estimate is defined as*

$$\hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{n} \log\{\sigma^p + |h_i||sgn(h_i)x_i - \theta|^p\} \right] \quad (2)$$

where $sgn(\cdot)$ denotes the sign operator.

The M-GC is simply a special case of the weighted M-GC estimator when all weights are set to unity. The special $p = 1$ and $p = 2$ cases yield the *meridian* [7] and *myriad* [3] estimators respectively. These estimators provide robust frameworks for signal processing in heavy-tailed environments.

Importantly, weighted M-GC estimators converge to weighted L_p estimators (ML estimates for GGD distributed samples) as $\sigma \to \infty$.

Lemma 1 *The weighted M-GC estimator converges to the weighted L_p estimator as $\sigma \to \infty$,* i.e.*,*

$$\lim_{\sigma\to\infty} \hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{n} |h_i||sgn(h_i)x_i - \theta|^p \right]. \quad (3)$$

The proof follows very similar to the proof of the median property of the weighted meridian estimator (property 1 in [7]) and is omitted for brevity. In particular, when $p = 2$ and when $p = 1$, we obtain the weighted myriad (WMy) and the weighted meridian (WMe) estimators:

Corollary 1 *[3] The WMy estimator converges to a normalized linear estimator as $\sigma \to \infty$,* i.e.*,*

$$\lim_{\sigma\to\infty} \hat{\theta} = \left(\sum_{i=1}^{n} |h_i| \right)^{-1} \sum_{i=1}^{n} h_i x_i. \quad (4)$$

Corollary 2 *[7] The WMe estimator converges to the weighted median estimator as $\sigma \to \infty$,* i.e.*,*

$$\lim_{\sigma\to\infty} \hat{\theta} = \textit{median}(|h_i| \diamond sgn(h_i)x_i|_{i=1}^{n}) \quad (5)$$

where $\diamond$ denotes the replication *operator.*

B. Multiparameter Estimation

The location estimation problem defined by the M-GC depends on the parameters σ and p which determine the properties of the estimator. In order to find the optimal tunable parameters (if the model considered is GCD) we detail multiparameter M-estimates [4].

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

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

The solution to this multiparameter optimization problem is obtained by solving a set of simultaneous equations for each component using a flip-flop algorithm coupled with fix point search [4].

III. Robust Distance Metrics

The cost function of the M-GC estimator can be extended to define a quasi-norm over $\mathbb{R}^m$ and a semimetric for the same space similar to L_p norms derived from the GGD family.

Definition 2 *Let $u \in \mathbb{R}^m$, then the LL_p norm of u is defined as*

$$\|u\|_{LL_p,\sigma} = \sum_{i=1}^{m} \log\left\{1 + \frac{|u_i|^p}{\sigma^p}\right\}, \ \sigma > 0. \quad (7)$$

The LL_p norm is not a norm in the strict sense since it does not meet the positive homogeneity and subadditivity properties. However, it follows the positive definiteness property and a scale invariant property. Proofs of the statements presented in this section can be found in [8].

Proposition 1 *For all $c \in \mathbb{R}$, $u \in \mathbb{R}^m$, and $p, \sigma > 0$ the following statements hold*

i) $\|u\|_{LL_p,\sigma} \geq 0$*, with* $\|u\|_{LL_p,\sigma} = 0$ *if and only if* $u = 0$*.*
ii) $\|cu\|_{LL_p,\sigma} = \|u\|_{LL_p,\delta}$ *where* $\delta = \sigma/|c|$*.*

The LL_p norm defines a robust metric that does not penalize heavily large deviations unlike the L_p norm, with the robustness depending on the scale parameter σ and p. The following Lemma constructs a relationship between the L_p norms and the LL_p norms.

Lemma 2 *For every $u \in \mathbb{R}^m$, $0 < p \leq 2$ and $\sigma > 0$ the following relations hold:*

$$\sigma^p \|u\|_{LL_p,\sigma} \leq \|u\|_p^p \leq \sigma^p m (e^{\|u\|_{LL_p,\sigma}} - 1). \quad (8)$$

The LL_p norms can be used to define robust regressors for statistical estimation. Let $v \in \mathbb{R}^t$, $\beta_0 \in \mathbb{R}^n$ and $A \in \mathbb{R}^{t\times n}$, $t > n$, with the observation model $v = A\beta_0 + r$ where r is additive white noise (possibly impulsive). The regression problem is defined as

$$\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^n} \|v - A\beta\|_{LL_p,\sigma}. \quad (9)$$

The particular case when $p = 2$ is known as the Lorentzian norm. The Lorentzian norm has desirable properties as a robust error metric in regression problems:

- It is an everywhere continuous function.
- It is convex near the origin ($0 \leq u \leq \sigma$), behaving as an L_2 cost function for small variations.
- Large deviations are not heavily penalized as in the case of L_1 or L_2 norms leading to a more robust error metric when the deviations contain gross errors.

IV. Illustrative Application Areas

In this section, we consider two possible application venues for the robust problem formulation framework developed in this work: 1) Robust reconstruction methods for compressed sensing and 2) robust blind decentralized estimation over noisy channels for sensor networks.

A. Robust Reconstruction Methods

Compressed sensing is a recently introduced novel framework that goes against the traditional data acquisition paradigm. Consider a set of m sensors making observations of a signal $x_0 \in \mathbb{R}^n$. Suppose that the signal x_0 is s-sparse in some orthogonal basis Ψ and let $\{\phi_i\}_{i=1}^m$ be a set of measurements vectors that are incoherent with the sparsity basis. Each sensor takes measurements projecting x_0 onto $\{\phi_i\}_{i=1}^m$ and communicates its observation to the fusion center over a noisy channel. Then the measurement process can be modeled as $y = \Phi x_0 + z$, where Φ is an $m \times n$ matrix with vectors ϕ_i as rows and z is white additive noise (with possibly impulsive behavior).

The problem is how to estimate x_0 from the noisy measurements y. The reconstruction strategies need to be robust and stable in the sense that small variations in the noiseless samples should yield small variations in the reconstructed signal, even when a fraction of the measurements are corrupted by gross errors. Most of current CS reconstruction algorithms use the L_2 norm as the metric for the residual error but, as detailed in [8], the L_2 norm is not an appropriate metric when the measurements are corrupted by outliers. Using these arguments, we propose to use a robust metric to penalize the residual and address the impulsive sampling noise problem.

To estimate x_0 from y we propose the following non-linear optimization problem:

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s. t. } \|y - \Phi x\|_{LL_2,\gamma} \leq \epsilon. \tag{10}$$

The following result presents an upper bound for the reconstruction error of the proposed estimator and is based on restricted isometry properties (RIP) of the matrix Φ (see [8] and references therein for more details on RIP).

Proposition 2 *[8] Assume the matrix Φ meets a RIP, then for any s-sparse signal x_0 and observation noise z with $\|z\|_{LL_2,\gamma} \leq \epsilon$, the solution to (10), denoted as x^*, obeys*

$$\|x^* - x_0\|_2 \leq C_s \cdot 2\gamma \cdot \sqrt{m(e^\epsilon - 1)}, \tag{11}$$

where C_s is a small constant.

Notably, γ controls the robustness of the employed norm and ϵ the radius of the feasibility set LL_2 ball. Let Z be a Cauchy random variable with scale parameter σ and zero location parameter. Assuming a Cauchy model for the noise vector yields $\mathbb{E}\|z\|_{LL_2,\gamma} = m\mathbb{E}\log\{1+\gamma^{-2}Z^2\} = 2m\log(1+\gamma^{-1}\sigma)$. We use this value for ϵ and set γ as MAD(y).

Debiasing is achieved through robust regression on a subset of indexes of $\hat{x}$ using the Lorentzian norm. The subset is defined as $I = \{i : |\hat{x}_i| > \alpha\}$, $\alpha = \lambda \max_i |\hat{x}_i|$, where $0 < \lambda < 1$. Thus $\tilde{x}_I \in \mathbb{R}^d$ is defined as

$$\tilde{x}_I = \arg\min_{x \in \mathbb{R}^d} \|y - \Phi_I x\|_{LL_2,\sigma} \tag{12}$$

where $d = |I|$. The final reconstruction, after the regression ($\tilde{x}$) is defined as $\tilde{x}_I$ for indexes in the subset I and zero outside I. This algorithm is referred to as Lorentzian BP [8].

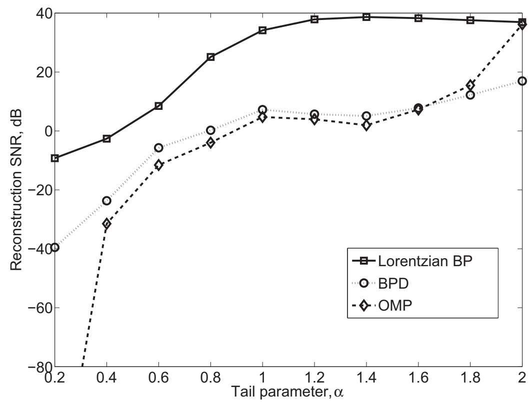

Fig. 1. Comparison of Lorentzian BP with BPD and OMP for impulsive contaminated samples. α-stable noise, $\sigma = 0.1$. R-SNR as a function of the tail parameter, α.

Results comparing Lorentzian BP with basis pursuit denoising (BPD) and orthogonal matching pursuit (OMP) (see [8] and references for details) are presented in Fig. 1. The signals are synthetic s-sparse signals with $s = 10$ and length $n = 1024$. The number of measurements is $m = 128$ and the sampling noise is α-stable distributed with scale parameter $\sigma = 0.1$ and the tail parameter, α, is varied from 0.2 to 2.

B. Robust Blind Decentralized Estimation

Consider a set of K distributed sensors, each making observations of a deterministic source signal θ. The observations are quantized with one bit (binary observations) and then transmitted trough a noisy channel to the fusion center where θ is to be estimated (see [9] and references therein). The observations are modeled as $x = \theta + n$, where n are sensor noise samples assumed zero-mean, spatially uncorrelated, independent and identically distributed.

Due to bandwidth limitations the observations are quantized as binary observations,

$$b_k = \mathbf{1}\{x_k \in (\tau, +\infty)\} \tag{13}$$

for $k = 1, 2, \ldots, K$, where τ is a real valued constant and $\mathbf{1}\{\cdot\}$ is the indicator function. The observations received at the fusion center are modeled by

$$y = (2b - 1) + w = m + w \tag{14}$$

where w are assumed to be zero-location independent channel noise samples, for the sake of simplicity, and the transformation $m_k = 2b_k - 1$ is made to adopt a binary phase shift keying (BPSK) scheme. The channel density function is denoted by $w_k \sim f_w(u)$. When the channel noise is of impulsive behavior (*e.g.* atmospheric noise or underwater acoustic noise) traditional Gaussian-based methods (*e.g.* least squares) do not perform well. We extend the blind decentralized estimation method proposed in [9] modeling the channel corruption as GCD noise and deriving a robust estimation method for impulsive channel noise scenarios. The sensor noise is modeled as zero-mean additive white Gaussian noise with variance σ_n^2.

The channel noise is modeled as zero-location additive white GCD noise with scale parameter σ_w and tail constant p.

A more realistic approach to the estimation problem in sensor networks is to assume that the noise PDF is known but that the values of some parameters are unknown [9]. In the following, we consider the estimation problem when the sensor noise parameter σ_n is known and the channel noise tail constant p and scale parameter σ_w are unknown.

Instrumental to the scheme presented is the fact that b_k is a Bernoulli random variable with parameter

$$\psi(\theta) \triangleq \Pr\{b_k = +1\} = 1 - F_n(\tau - \theta) \qquad (15)$$

where $F_n(\cdot)$ is the cumulative distribution function of n_k. The PDF of the noisy observations received at the fusion center is given by

$$f_y(y) = \psi(\theta) f_w(y-1) + [1 - \psi(\theta)] f_w(y+1). \qquad (16)$$

Note that the resulting PDF is a mixture of GCD with mixing parameters ψ and $[1-\psi]$. Given the PDF of the observations, the ML estimate of θ can be formulated. To simplify the problem, we first estimate $\psi = \psi(\theta)$ and then utilize the invariance of the ML estimate to determine θ using (15). Using the log-likelihood function, the ML estimate of $\psi \in (0,1)$ reduces to

$$\hat{\psi} = \arg\max_{\psi} \sum_{k=1}^{K} \log\{\psi f_w(y_k - 1) + [1-\psi] f_w(y_k+1)\}. \quad (17)$$

The unknown parameter set for the estimation problem is $\{\psi, \sigma_w, p\}$. We address this problem utilizing the well known EM algorithm. Each iteration of the EM algorithm consists of two steps: the expectation step (E-step) and the maximization step (M-step). The followings are the E- and M- steps for the unknown parameter set estimation in the considered sensor network application.

E-step: Let the parameters estimated at the j-th iteration be marked by a superscript (j) and $\Gamma^{(j)} = (\hat{\sigma}_w^{(j)}, \hat{p}^{(j)})$. The posterior probabilities are computed as

$$q_k = \frac{\hat{\psi}^{(j)} f_w(y_k - 1|\Gamma^{(j)})}{\hat{\psi}^{(j)} f_w(y_k - 1|\Gamma^{(j)}) + [1 - \hat{\psi}^{(j)}] f_w(y_k + 1|\Gamma^{(j)})}. \qquad (18)$$

M-step: The ML estimates $\{\hat{\psi}^{(j+1)}, \Gamma^{(j+1)}\}$ are given by

$$\hat{\psi}^{(j+1)} = \frac{1}{K}\sum_{k=1}^{K} q_k, \text{ and, } \Gamma^{(j+1)} = \arg\max_{\Gamma} \Lambda(\Gamma) \qquad (19)$$

where

$$\Lambda(\Gamma) = \sum_{k=1}^{K} q_k \Upsilon(y_k - 1; \Gamma) + (1 - q_k)\Upsilon(y_k + 1; \Gamma) \qquad (20)$$

where $\Upsilon(u;\Gamma) = \log a(p) + \log \sigma_w - 2p^{-1}\log(\sigma_w^p + |u|^p)$ and $a(p) = p\Gamma(2/p)/2(\Gamma(1/p))^2$. The maximization in (19) is the ML estimate of scale and tail constant of a mixture of GCD and is addressed using fixed point search algorithms similar to those derived in [4] for multiparameter estimation. Numerical results comparing the GCD method derived in this paper (MLUGC) with the Gaussian channel based method derived in [9] (MLUG) are presented in Fig. 2. The MSE is used as comparison metric. As a reference, the MSE of the binary estimator (BE) and the clairvoyant estimator (CE) (estimators in perfect transmission) are also included. A sensor network with the following parameters is used: $\theta = 1$, $\tau = 0$, $\sigma_n = 1$ and $K = 1000$. The channel noise is α-stable distributed with $\sigma_w = 0.5$ and the tail parameter, α, is varied from 0.2 to 2.

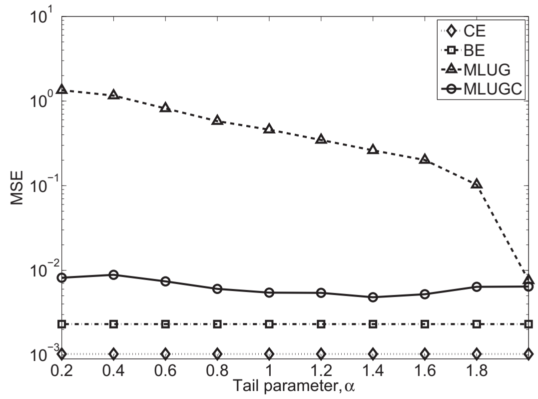

Fig. 2. A sensor network with the following parameters: $\theta = 1$, $\tau = 0$, $\sigma_n = 1$ and $K = 1000$. Channel noise α-stable distributed with $\sigma_w = 0.5$. Output variances of MLUGC, MLUG, BE and CE as function of the tail parameter, α.

V. Concluding Remarks

This paper presents a GCD based theoretical approach that allows the formulation of challenging problems in a robust fashion. The proposed framework, thanks to its flexibility, subsumes GGD based developments, thereby guaranteeing performance improvements over the traditional problem formulation techniques. Properties of the derived techniques are analyzed and two particular applications are developed under this framework: robust reconstruction methods for CS and robust estimation over noisy channels for sensor networks.

References

[1] Huber, *Robust Statistics*. John Wiley & Sons, Inc., 1981.

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

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

[4] R. E. Carrillo, T. C. Aysal, and K. E. Barner, "Generalized cauchy distribution based robust estimation," in *Proceedings, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Apr. 2008.

[5] P. R. Rider, "Generalized cauchy distributions," *Annals of the Institute of Statistical Mathematics*, vol. 9, pp. 215–223, 1957.

[6] J. Miller and J. Thomas, "Detectors for discrete-time signals in non-gaussian noise," *IEEE Transactions on Information Theory*, vol. 8, no. 2, pp. 241–250, Mar. 1972.

[7] 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.

[8] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE Journal of Selected Topics in Signal Processing*, accepted for publication.

[9] T. C. Aysal and K. E. Barner, "Blind decentralized estimation for bandwidth constrained wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1466–1471, May 2008.