

## Research Article

# A Generalized Cauchy Distribution Framework for Problems Requiring Robust Behavior

**Rafael E. Carrillo, Tuncer C. Aysal (EURASIP Member), and Kenneth E. Barner**

*Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA*

Correspondence should be addressed to Rafael E. Carrillo, carrillo@ee.udel.edu

Received 8 February 2010; Revised 27 May 2010; Accepted 7 August 2010

Academic Editor: Igor Djurović

Copyright © 2010 Rafael E. Carrillo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Statistical modeling is at the heart of many engineering problems. The importance of statistical modeling emanates not only from the desire to accurately characterize stochastic events, but also from the fact that distributions are the central models utilized to derive sample processing theories and methods. The generalized Cauchy distribution (GCD) family has a closed-form pdf expression across the whole family as well as algebraic tails, which makes it suitable for modeling many real-life impulsive processes. This paper develops a GCD theory-based approach that allows challenging problems to be formulated in a robust fashion. Notably, the proposed framework subsumes generalized Gaussian distribution (GGD) family-based developments, thereby guaranteeing performance improvements over traditional GCD-based problem formulation techniques. This robust framework can be adapted to a variety of applications in signal processing. As examples, we formulate four practical applications under this framework: (1) filtering for power line communications, (2) estimation in sensor networks with noisy channels, (3) reconstruction methods for compressed sensing, and (4) fuzzy clustering.

## 1. Introduction

Traditional signal processing and communications methods are dominated by three simplifying assumptions: (1) the systems under consideration are linear; the signal and noise processes are (2) stationary and (3) Gaussian distributed. Although these assumptions are valid in some applications and have significantly reduced the complexity of techniques developed, over the last three decades practitioners in various branches of statistics, signal processing, and communications have become increasingly aware of the limitations these assumptions pose in addressing many real-world applications. In particular, it has been observed that the Gaussian distribution is too light-tailed to model signals and noise that exhibits impulsive and nonsymmetric characteristics [1]. A broad spectrum of applications exists in which such processes emerge, including wireless communications, teletraffic, hydrology, geology, atmospheric noise compensation, economics, and image and video processing (see [2, 3] and references therein). The need to describe impulsive data, coupled with computational advances that enable processing

of models more complicated than the Gaussian distribution, has thus led to the recent dynamic interest in heavy-tailed models.

Robust statistics—the stability theory of statistical procedures—systematically investigates deviation from modeling assumption affects [4]. Maximum likelihood (ML) type estimators (or more generally,  $M$ -estimators) developed in the theory of robust statistics are of great importance in *robust signal processing techniques* [5].  $M$ -estimators can be described by a cost function-defined optimization problem or by its first derivative, the latter yielding an implicit equation (or set of equations) that is proportional to the influence function. In the location estimation case, properties of the influence function describe the estimator robustness [4]. Notably, ML location estimation forms a special case of  $M$ -estimation, with the observations taken to be independent and identically distributed and the cost function set proportional to the logarithm of the common density function.

To address as wide an array of problems as possible, modeling and processing theories tend to be based on density families that exhibit a broad range of characteristics.

Signal processing methods derived from the generalized Gaussian distribution (GGD), for instance, are popular in the literature and include works addressing heavy-tailed process [2, 3, 6–8]. The GGD is a family of closed form densities, with varying tail parameter, that effectively characterizes many signal environments. Moreover, the closed form nature of the GGD yields a rich set of distribution optimal error norms ( $L_1$ ,  $L_2$ , and  $L_p$ ), and estimation and filtering theories, for example, linear filtering, weighted median filtering, fractional low order moment (FLOM) operators, and so forth. [3, 6, 9–11]. However, a limitation of the GGD model is the tail decay rate—GGD distribution tails decay exponentially rather than algebraically. Such light tails do not accurately model the prevalence of outliers and impulsive samples common in many of today’s most challenging statistical signal processing and communications problems [3, 12, 13].

As an alternative to the GGD, the  $\alpha$ -stable density family has gained recent popularity in addressing heavy-tailed problems. Indeed, symmetric  $\alpha$ -stable processes exhibit algebraic tails and, in some cases, can be justified from first principles (Generalized Central Limit Theorem) [14–16]. The index of stability parameter,  $\alpha \in (0, 2]$ , provides flexibility in impulsiveness modeling, with distributions ranging from light-tailed Gaussian ( $\alpha = 2$ ) to extremely impulsive ( $\alpha \rightarrow 0$ ). With the exception of the limiting Gaussian case,  $\alpha$ -stable distributions are heavy-tailed with infinite variance and algebraic tails. Unfortunately, the Cauchy distribution ( $\alpha = 1$ ) is the only algebraic-tailed  $\alpha$ -stable distribution that possesses a closed form expression, limiting the flexibility and performance of methods derived from this family of distributions. That is, the single distribution Cauchy methods (Lorentzian norm, weighted myriad) are the most commonly employed  $\alpha$ -stable family operators [12, 17–19].

The Cauchy distribution, while intersecting the  $\alpha$ -stable family at a single point, is generalized by the introduction of a varying tail parameter, thereby forming the Generalized Cauchy density (GCD) family. The GCD has a closed form pdf across the whole family, as well as algebraic tails that make it suitable for modeling real-life impulsive processes [20, 21]. Thus the GCD combines the advantages of the GGD and  $\alpha$ -stable distributions in that it possesses (1) heavy, algebraic tails (like  $\alpha$ -stable distributions) and (2) closed form expressions (like the GGD) across a flexible family of densities defined by a tail parameter,  $p \in (0, 2]$ . Previous GCD family development focused on the particular  $p = 2$  (Cauchy distribution) and  $p = 1$  (meridian distribution) cases, which lead to the myriad and meridian [13, 22] estimators, respectively. (It should be noted that the original authors derived the myriad filter starting from  $\alpha$ -stable distributions, noting that there are only two closed-form expressions for  $\alpha$ -stable distributions [12, 17, 18].) These estimators provide a robust framework for heavy-tail signal processing problems.

In yet another approach, the generalized- $t$  model is shown to provide excellent fits to different types of atmospheric noise [23]. Indeed, Hall introduced the family of generalized- $t$  distributions in 1966 as an empirical model for atmospheric radio noise [24]. The distribution possesses

algebraic tails and a closed form pdf. Like the  $\alpha$ -stable family, the generalized- $t$  model contains the Gaussian and the Cauchy distributions as special cases, depending on the degrees of freedom parameter. It is shown in [18] that the myriad estimator is also optimal for the generalized- $t$  family of distributions. Thus we focus on the GCD family of operators, as their performance also subsumes that of generalized- $t$  approaches.

In this paper, we develop a GCD-based theoretical approach that allows challenging problems to be formulated in a robust fashion. Within this framework, we establish a statistical relationship between the GGD and GCD families. The proposed framework subsumes GGD-based developments (e.g., least squares, least absolute deviation, FLOM,  $L_p$  norms,  $k$ -means clustering, etc.), thereby guaranteeing performance improvements over traditional problem formulation techniques. The developed theoretical framework includes robust estimation and filtering methods, as well as robust error metrics. A wide array of applications can be addressed through the proposed framework, including, 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. As illustrative and evaluation examples, we formulate four particular applications under this framework: (1) filtering for power line communications, (2) estimation in sensor networks with noisy channels, (3) reconstruction methods for compressed sensing, and (4) fuzzy clustering.

The organization of the paper is as follows. In Section 2, we present a brief review of  $M$ -estimation theory and the generalized Gaussian and generalized Cauchy density families. A statistical relationship between the GGD and GCD is established, and the ML location estimate from GCD statistics is derived. An  $M$ -type estimator, coined  $M$ -GC estimator, is derived in Section 3 from the cost function emerging in GCD-based ML estimation. Properties of the proposed estimator are analyzed, and a weighted filter structure is developed. Numerical algorithms for multiparameter estimation are also presented. A family of robust metrics derived from the GCD are detailed in Section 4, and their properties are analyzed. Four illustrative applications of the proposed framework are presented in Section 5. Finally, we conclude in Section 6 with closing thoughts and future directions.

## 2. Distributions, Optimal Filtering, and $M$ -Estimation

This section presents  $M$ -estimates, a generalization of maximum likelihood (ML) estimates, and discusses optimal filtering from an ML perspective. Specifically, it discusses statistical models of observed samples obeying generalized Gaussian statistics and relates the filtering problem to maximum likelihood estimation. Then, we present the generalized Cauchy distribution, and a relation between GGD and GCD random variables is introduced. The ML estimators for GCD statistics are also derived.

**2.1. *M*-Estimation.** In the *M*-estimation theory the objective is to estimate a deterministic but unknown parameter  $\theta \in \mathbb{R}$  (or set of parameters) of a real-valued signal  $s(i; \theta)$  corrupted by additive noise. Suppose that we have  $N$  observations yielding the following parametric signal model:

$$x(i) = s(i; \theta) + n(i) \quad (1)$$

for  $i = 1, 2, \dots, N$ , where  $\{x(i)\}_{i=1}^N$  and  $\{n(i)\}_{i=1}^N$  denote the observations and noise components, respectively. Let  $\hat{\theta}$  be an estimate of  $\theta$ , then any estimate that solves the minimization problem of the form

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \rho(x(i); \theta) \quad (2)$$

or by an implicit equation

$$\sum_{i=1}^N \psi(x(i); \hat{\theta}) = 0 \quad (3)$$

is called an *M*-estimate (or maximum likelihood type estimate). Here  $\rho(x; \theta)$  is an arbitrary cost function to be designed, and  $\psi(x; \theta) = (\partial/\partial\theta)\rho(x; \theta)$ . Note that *ML*-estimators are a special case of *M*-estimators with  $\rho(x; \theta) = -\log f(x; \theta)$ , where  $f(\cdot)$  is the probability density function of the observations. In general, *M*-estimators do not necessarily relate to probability density functions.

In the following we focus on the location estimation problem. This is well founded, as location estimators have been successfully employed as moving window type filters [3, 5, 9]. In this case, the signal model in (1) becomes  $x(i) = \theta + n(i)$  and the minimization problem in (2) becomes

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \rho(x(i) - \theta) \quad (4)$$

or

$$\sum_{i=1}^N \psi(x(i) - \hat{\theta}) = 0. \quad (5)$$

For *M*-estimates it can be shown that the influence function is proportional to  $\psi(x)$  [4, 25], meaning that we can derive the robustness properties of an *M*-estimator, namely, efficiency and bias in the presence of outliers, if  $\psi$  is known.

**2.2. Generalized Gaussian Distribution.** The statistical behavior of a wide range of processes can be modeled by the GGD, such as DCT and wavelets coefficients and pixels difference [2, 3]. The GGD pdf is given by

$$f(x) = \frac{k\alpha}{2\Gamma(1/k)} \exp(-\alpha|x - \theta|^k), \quad (6)$$

where  $\Gamma(\cdot)$  is the gamma function  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ ,  $\theta$  is the location parameter, and  $\alpha$  is a constant related to the standard deviation  $\sigma$ , defined as  $\alpha = \sigma^{-1} \sqrt{\Gamma(3/k)(\Gamma(1/k))^{-1}}$ .

In this form,  $\alpha$  is an inverse scale parameter, and  $k > 0$ , sometimes called the shade parameter, controls the tail decay rate. The GGD model contains the Laplacian and Gaussian distributions as special cases, that is, for  $k = 1$  and  $k = 2$ , respectively. Conceptually, the lower the value of  $k$  is the more impulsive the distribution is. The ML location estimate for GGD statistics is reviewed in the following. Detailed derivations of these results are given in [3].

Consider a set of  $N$  independent observations each obeying the GGD with common location parameter, common shape parameter  $k$ , and different scale parameter  $\sigma_i$ . The ML estimate of location is given by

$$\hat{\theta} = \arg \min_{\theta} \left[ \sum_{i=1}^N \frac{1}{\sigma_i^k} |x(i) - \theta|^k \right]. \quad (7)$$

There are two special cases of the GGD family that are well studied: the Gaussian ( $k = 2$ ) and the Laplacian ( $k = 1$ ) distributions, which yield the well known *weighted mean* and *weighted median* estimators, respectively. When all samples are identically distributed for the special cases, the *mean* and *median* estimators are the resulting operators. These estimators are formally defined in the following.

**Definition 1.** Consider a set of  $N$  independent observations each obeying the Gaussian distribution with different variance  $\sigma_i^2$ . The ML estimate of location is given by

$$\hat{\theta} = \frac{\sum_{i=1}^N h_i x(i)}{\sum_{i=1}^N h_i} \triangleq \text{mean} \{ h_i \cdot x(i) |_{i=1}^N \}, \quad (8)$$

where  $h_i = 1/\sigma_i^2$  and  $\cdot$  denotes the (multiplicative) weighting operation.

**Definition 2.** Consider a set of  $N$  independent observations each obeying the Laplacian distribution with common location and different scale parameter  $\sigma_i$ . The ML estimate of location is given by

$$\hat{\theta} = \text{median} \{ h_i \diamond x(i) |_{i=1}^N \}, \quad (9)$$

where  $h_i = 1/\sigma_i$  and  $\diamond$  denotes the *replication* operator defined as

$$h_i \diamond x(i) = \overbrace{x(i), x(i), \dots, x(i)}^{h_i \text{ times}}. \quad (10)$$

Through arguments similar to those above, the  $k \neq 1, 2$  cases yield the fractional lower order moment (FLOM) estimation framework [9]. For  $k < 1$ , the resulting estimators are selection type. A drawback of FLOM estimators for  $1 < k < 2$  is that their computation is, in general, nontrivial, although suboptimal (for  $k > 1$ ) selection-type FLOM estimators have been introduced to reduce computational costs [6].

**2.3. Generalized Cauchy Distribution.** The GCD family was proposed by Rider in 1957 [20], rediscovered by Miller and Thomas in 1972 with a different parametrization [21], and

has been used in several studies of impulsive radio noise [3, 12, 17, 21, 22]. The GCD pdf is given by

$$f_{\text{GCD}}(z) = a\sigma(\sigma^p + |z - \theta|^p)^{-2/p} \quad (11)$$

with  $a = p\Gamma(2/p)/2(\Gamma(1/p))^2$ . In this representation,  $\theta$  is the location parameter,  $\sigma$  is the scale parameter, and  $p$  is the tail constant. The GCD family contains the Meridian [13] and Cauchy distributions as special cases, that is, for  $p = 1$  and  $p = 2$ , respectively. For  $p < 2$ , the tail of the pdf decays slower than in the Cauchy distribution case, resulting in a heavier-tailed distribution.

The flexibility and closed-form nature of the GCD make it an ideal family from which to derive robust estimation and filtering techniques. As such, we consider the location estimation problem that, as in the previous case, is approached from an ML estimation framework. Thus consider a set of  $N$  i.i.d. GCD distributed samples with common scale parameter  $\sigma$  and tail constant  $p$ . The ML estimate of location is given by

$$\hat{\theta} = \arg \min_{\theta} \left[ \sum_{i=1}^N \log\{\sigma^p + |x(i) - \theta|^p\} \right]. \quad (12)$$

Next, consider a set of  $N$  independent observations each obeying the GCD with common tail constant  $p$ , but possessing unique scale parameter  $\nu_i$ . The ML estimate is formulated as  $\hat{\theta} = \arg \max_{\theta} \prod_{i=1}^N f_{\text{GCD}}(x(i); \nu_i)$ . Inserting the GCD distribution for each sample, taking the natural log, and utilizing basic properties of the argmax and log functions yield

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \log \left[ \prod_{i=1}^N a\nu_i (\nu_i^p + |x(i) - \theta|^p)^{-2/p} \right] \\ &= \arg \max_{\theta} \sum_{i=1}^N -\frac{2}{p} \log\{\nu_i^p + |x(i) - \theta|^p\} \\ &= \arg \min_{\theta} \sum_{i=1}^N \log \left\{ 1 + \frac{|x(i) - \theta|^p}{\nu_i^p} \right\} \\ &= \arg \min_{\theta} \sum_{i=1}^N \log\{\sigma^p + h_i |x(i) - \theta|^p\} \end{aligned} \quad (13)$$

with  $h_i = (\sigma/\nu_i)^p$ .

Since the estimator defined in (12) is a special case of that defined in (13), we only provide a detailed derivation for the latter. The estimator defined in (13) can be used to extend the GCD-based estimator to a robust weighted filter structure. Furthermore, the derived filter can be extended to admit real-valued weights using the sign-coupling approach [8].

**2.4. Statistical Relationship between the Generalized Cauchy and Gaussian Distributions.** Before closing this section, we bring to light an interesting relationship between the Generalized Cauchy and Generalized Gaussian distributions. It is wellknown that a Cauchy distributed random variable (GCD  $p = 2$ ) is generated by the ratio of two independent Gaussian

distributed random variables (GGD  $k = 2$ ). Recently, Aysal and Barner showed that this relationship also holds for the Laplacian and Meridian distributions [13], that is, the ratio of two independent Laplacian (GGD  $k = 1$ ) random variables yields a Meridian (GCD  $p = 1$ ) random variable. In the following, we extend this finding to the complete set of GGD and GCD families.

**Lemma 1.** *The random variable formed as the ratio of two independent zero-mean GGD distributed random variables  $U$  and  $V$ , with tail constant  $\beta$  and scale parameters  $\alpha_U$  and  $\alpha_V$ , respectively, is a GCD random variable with tail parameter  $\lambda = \beta$  and scale parameter  $\nu = \alpha_U/\alpha_V$ .*

*Proof.* See Appendix A.  $\square$

### 3. Generalized Cauchy-Based Robust Estimation and Filtering

In this section we use the GCD ML location estimate cost function to define an  $M$ -type estimator. First, robustness and properties of the derived estimator are analyzed, and the filtering problem is then related to  $M$ -estimation. The proposed estimator is extended to a weighted filtering structure. Finally, practical algorithms for the multiparameter case are developed.

**3.1. Generalized Cauchy-Based  $M$ -Estimation.** The cost function associated with the GCD ML estimate of location derived in the previous section is given by

$$\rho(x) = \log\{\sigma^p + |x|^p\}, \quad \sigma > 0, \quad 0 < p \leq 2. \quad (14)$$

The flexibility of this cost function, provided by parameters  $\sigma$  and  $p$ , and robust characteristics make it well-suited to define an  $M$ -type estimator, which we coin the  $M$ -GC estimator. To define the form of this estimator, denote  $\mathbf{x} = [x(1), \dots, x(N)]$  as a vector of observations and  $\theta$  as the common location parameter of the observations.

**Definition 3.** The  $M$ -GC estimate is defined as

$$\hat{\theta} = \arg \min_{\theta} \left[ \sum_{i=1}^N \log\{\sigma^p + |x(i) - \theta|^p\} \right]. \quad (15)$$

The special  $p = 2$  and  $p = 1$  cases yield the *myriad* [18] and *meridian* [13] estimators, respectively. The generalization of the  $M$ -GC estimator, for  $0 < p \leq 2$ , is analogous to the GGD-based FLOM estimators and thereby provides a rich and robust framework for signal processing applications.

As the performance of an estimator depends on the defining objective function, the properties of the objective function at hand are analyzed in the following.

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

(1)  $Q(\theta)$  is strictly decreasing for  $\theta < x_{[1]}$  and strictly increasing for  $\theta > x_{[N]}$ .

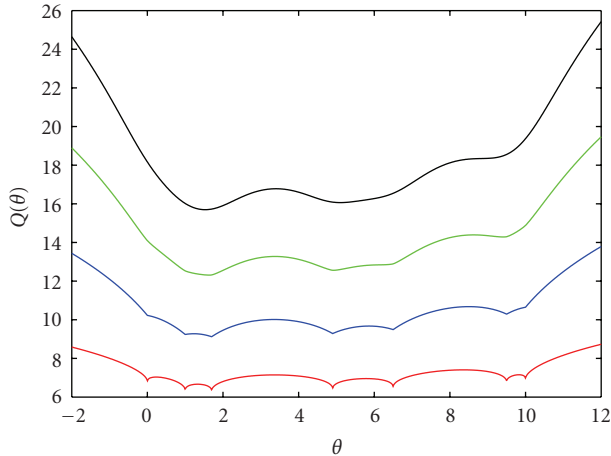


FIGURE 1: Typical M-GC objective functions for different values of  $p \in \{0.5, 1, 1.5, 2\}$  (from bottom to top respectively). Input samples are  $x = [4.9, 0, 6.5, 10.0, 9.5, 1.7, 1]$  and  $\sigma = 1$ .

(2) All local extrema of  $Q(\theta)$  lie in the interval  $[x_{[1]}, x_{[N]}]$ .

(3) If  $0 < p \leq 1$ , the solution is one of the input samples (selection type filter).

(4) 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.

*Proof.* See Appendix B.  $\square$

The M-GC estimator has two adjustable parameters,  $\sigma$  and  $p$ . The tail constant,  $p$ , depends on the heaviness of the underlying distribution. Notably, when  $p \leq 1$  the estimator behaves as a selection type filter, and, as  $p \rightarrow 0$ , it becomes increasingly robust to outlier samples. For  $p > 1$ , the location estimate is in the range of the input samples and is readily computed. Figure 1 shows a typical sketch of the M-GC objective function, in this case for  $p \in \{0.5, 1, 1.5, 2\}$  and  $\sigma = 1$ .

The following properties detail the M-GC estimator behavior as  $\sigma$  goes to either 0 or  $\infty$ . Importantly, the results show that the M-GC estimator subsumes other classical estimator families.

*Property 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  $\sigma \rightarrow \infty$ :

$$\lim_{\sigma \rightarrow \infty} \hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N |x(i) - \theta|^p. \quad (16)$$

*Proof.* See Appendix C.  $\square$

Intuitively, this result is explained by the fact that  $|x(i) - \theta|^p / \sigma^p$  becomes negligible as  $\sigma$  grows large compared to 1. This, combined with the fact that  $\log(1 + x) \approx x$  when  $x \ll 1$ , which is an equality in the limit, yields the resulting cost function behavior. The importance of this result is that M-GC estimators include  $M$ -estimators with  $L_p$  norm ( $0 < p \leq 2$ ) cost functions. Thus M-GC (GCD-based) estimators

should be at least as powerful as GGD-based estimators (linear FIR, median, FLOM) in light-tailed applications, while the untapped algebraic tail potential of GCD methods should allow them to substantially outperform in heavy-tailed applications.

In contrast to the equivalence with  $L_p$  norm approaches for  $\sigma$  large, M-GC estimators become more resistant to impulsive noise as  $\sigma$  decreases. In fact, as  $\sigma \rightarrow 0$  the M-GC yields a mode type estimator with particularly strong impulse rejection.

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

$$\lim_{\sigma \rightarrow 0} \hat{\theta} = \arg \min_{x(j) \in \mathcal{M}} \left[ \prod_{i, x(i) \neq x(j)} |x(i) - x(j)| \right], \quad (17)$$

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

*Proof.* See Appendix D.  $\square$

This mode-type estimator treats every observation as a possible outlier, assigning greater influence to the most repeated values in the observations set. This property makes the M-GC a suitable framework for applications such as image processing, where selection-type filters yield good results [7, 13, 18].

**3.2. Robustness and Analysis of M-GC Estimators.** To formally evaluate the robustness of M-GC estimators, we consider the influence function, which, if it exists, is proportional to  $\psi(x)$  and determines the effect of contamination of the estimator. For the M-GC estimator

$$\psi(x) = \frac{p|x|^{p-1} \text{sgn}(x)}{\sigma^p + |x|^p}, \quad (18)$$

where  $\text{sgn}(\cdot)$  denotes the sign operator. Figure 2 shows the M-GC estimator influence function for  $p \in \{0.5, 1, 1.5, 2\}$ .

To further characterize  $M$ -estimates, it is useful to list the desirable features of a robust influence function [4, 25].

- (i) *B-Robustness.* An estimator is  $B$ -robust if the supremum of the absolute value of the influence function is finite.
- (ii) *Rejection Point.* The rejection point, defined as the distance from the center of the influence function to the point where the influence function becomes negligible, should be finite. Rejection point measures whether the estimator rejects outliers and, if so, at what distance.

The M-GC estimate is  $B$ -robust and has a finite rejection point that depends on the scale parameter  $\sigma$  and the tail parameter  $p$ . As  $p \rightarrow 0$ , the influence function has higher decay rate, that is, as  $p \rightarrow 0$  the M-GC estimator becomes more robust to outliers. Also of note is that  $\lim_{x \rightarrow \pm\infty} \psi(x) = 0$ , that is, the influence function

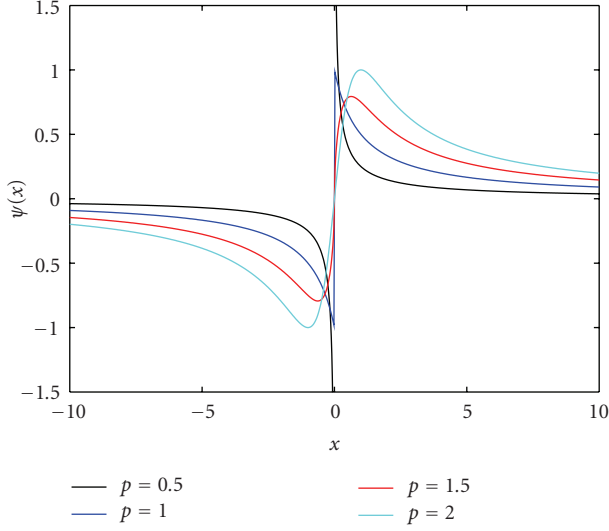


FIGURE 2: Influence functions of the M-GC estimator for different values of  $P$ . (Black:)  $P = .5$ , (blue:)  $P = 1$ , (red:)  $P = 1.5$ , and (cyan:)  $P = 2$ .

is asymptotically redescending, and the effect of outliers monotonically decreases with an increase in magnitude [25].

The M-GC also possesses the followings important properties.

*Property 3* (outlier rejection). For  $\sigma < \infty$ ,

$$\lim_{x(N) \rightarrow \pm\infty} \hat{\theta}(x(1), \dots, x(N)) = \hat{\theta}(x(1), \dots, x(N-1)). \quad (19)$$

*Property 4* (no undershoot/overshoot). The output of the M-GC estimator is always bounded by

$$x_{[1]} < \hat{\theta} < x_{[N]}, \quad (20)$$

where  $x_{[1]} = \min\{x(i)\}_{i=1}^N$  and  $x_{[N]} = \max\{x(i)\}_{i=1}^N$ .

According to Property 3, large errors are efficiently eliminated by an M-GC estimator with finite  $\sigma$ . Note that this property can be applied recursively, indicating that M-GC estimators eliminate multiple outliers. The proof of this statement follows the same steps used in the proof of the meridian estimator Property 9 [13] and is thus omitted. Property 4 states that the M-GC estimator is BIBO stable, that is, the output is bounded for bounded inputs. Proof of Property 4 follows directly from Propositions 1 and 2 and is thus omitted.

Since M-GC estimates are  $M$ -estimates, they have desirable asymptotic behavior, as noted in the following property and discussion.

*Property 5* (asymptotic consistency). Suppose that the samples  $\{x(i)\}_{i=1}^N$  are independent and symmetrically distributed around  $\theta$  (location parameter). Then, the M-GC estimate  $\hat{\theta}_N$  converges to  $\theta$  in probability, that is,

$$\hat{\theta}_N \xrightarrow{P} \theta \text{ as } N \rightarrow \infty. \quad (21)$$

Proof of Property 5 follows from the fact that the M-GC estimator influence function is odd, bounded, and continuous (except at the origin, which is a set of measure zero); argument details parallel those in [4].

Notably,  $M$ -estimators have asymptotic normal behavior [4]. In fact, it can be shown that

$$\sqrt{N}(\hat{\theta}_N - \theta) \rightarrow Z \quad (22)$$

in distribution, where  $Z \sim \mathcal{N}(0, \nu)$  and

$$\nu = \frac{E_F \psi^2(X - \theta)}{(E_F \psi'(X - \theta))^2}. \quad (23)$$

The expectation is taken with respect to  $F$ , the underlying distribution of the data. The last expression is the asymptotic variance of the estimator. Hence, the variance of  $\hat{\theta}_N$  decreases as  $N$  increases, meaning that M-GC estimates are asymptotically efficient.

*3.3. Weighted M-GC Estimators.* A filtering framework cannot be considered complete until an appropriate weighting operation is defined. Filter weights, or coefficients, are extremely important for applications in which signal correlations are to be exploited. Using the ML estimator under independent, but non identically distributed, GCD statistics (expression (13)), the M-GC estimator is extended to include weights. Let  $\mathbf{h} = [h_1, \dots, h_N]$  denote a vector of nonnegative weights. The weighted M-GC (WM-GC) estimate is defined as

$$\hat{\theta} = \arg \min_{\theta} \left[ \sum_{i=1}^N \log \{ \sigma^p + h_i |x(i) - \theta|^p \} \right]. \quad (24)$$

The filtering structure defined in (24) is an M-smoother estimator, which is in essence a low-pass-type filter. Utilizing the sign coupling technique [8], the M-GC estimator can be extended to accept real-valued weights. This yields the general structure detailed in the following definition.

*Definition 4.* The weighted M-GC (WM-GC) estimate is defined as

$$\hat{\theta} = \arg \min_{\theta} \left[ \sum_{i=1}^N \log \{ \sigma^p + |h_i| |\operatorname{sgn}(h_i)x(i) - \theta|^p \} \right], \quad (25)$$

where  $\mathbf{h} = [h_1, \dots, h_N]$  denotes a vector of real-valued weights.

The WM-GC estimators inherit all the robustness and convergence properties of the unweighted M-GC estimators. Thus as in the unweighted case, WM-GC estimators subsume GGD-based (weighted) estimators, indicating that WM-GC estimators are at least as powerful as GGD-based estimators (linear FIR, weighted median, weighted FLOM) in light-tailed environments, while WM-GC estimator characteristics enable them to substantially outperform in heavy-tailed impulsive environments.

**Require:** Data set  $\{x(i)\}_i^N = 1$  and tolerances  $\epsilon_1, \epsilon_2, \epsilon_3$ .

- (1) Initialize  $\sigma^{(0)}$  and  $\theta^{(0)}$ .
- (2) **while**  $|\hat{\theta}^{(m)} - \hat{\theta}^{(m-1)}| > \epsilon_1, |\hat{\sigma}^{(m)} - \hat{\sigma}^{(m-1)}| > \epsilon_2$  and  $|\hat{p}^{(m)} - \hat{p}^{(m-1)}| > \epsilon_3$  **do**
- (3) Estimate  $\hat{p}^{(m)}$  as the solution of (30).
- (4) Estimate  $\hat{\theta}^{(m)}$  as the solution of (28).
- (5) Estimate  $\hat{\sigma}^{(m)}$  as the solution of (29).
- (6) **end while**
- (7) **return**  $\hat{\theta}, \hat{\sigma}$  and  $\hat{p}$ .

ALGORITHM 1: Multiparameter estimation algorithm.

**3.4. Multiparameter Estimation.** The location estimation problem defined by the M-GC filter depends on the parameters  $\sigma$  and  $p$ . Thus to solve the optimal filtering problem, we consider multiparameter  $M$ -estimates [26]. The applied approach utilizes a small set of signal samples to estimate  $\sigma$  and  $p$  and then uses these values in the filtering process (although a fully adaptive filter can also be implemented using this scheme).

Let  $\{x(i)\}_{i=1}^N$  be a set of independent observations from a common GCD with deterministic but unknown parameters  $\theta, \sigma$ , and  $p$ . The joint estimates are the solutions to the following maximization problem:

$$(\hat{\theta}, \hat{\sigma}, \hat{p}) = \arg \max_{\theta, \sigma, p} g(\mathbf{x}; \theta, \sigma, p), \quad (26)$$

where

$$g(\mathbf{x}; \theta, \sigma, p) = \prod_{i=1}^N a\sigma(\sigma^p + |x(i) - \theta|^p)^{-2/p}, \quad (27)$$

$a = p\Gamma(2/p)/2(\Gamma(1/p))^2$ . The solution to this optimization problem is obtained by solving a set of simultaneous equations given by first-order optimality conditions. Differentiating the log-likelihood function,  $g(\mathbf{x}; \theta, \sigma, p)$ , with respect to  $\theta, \sigma$ , and  $p$  and performing some algebraic manipulations yields the following set of simultaneous equations:

$$\frac{\partial g}{\partial \theta} = \sum_{i=1}^N \frac{-p|x(i) - \theta|^{p-1} \operatorname{sgn}(x(i) - \theta)}{\sigma^p + |x(i) - \theta|^p} = 0, \quad (28)$$

$$\frac{\partial g}{\partial \sigma} = \sum_{i=1}^N \frac{\sigma^p - |x(i) - \theta|^p}{\sigma^p + |x(i) - \theta|^p} = 0, \quad (29)$$

$$\begin{aligned} \frac{\partial g}{\partial p} = \sum_{i=1}^N \left[ \frac{1}{2p} - \frac{\sigma^p \log \sigma - |x(i) - \theta|^p \log |x(i) - \theta|}{p(\sigma^p + |x(i) - \theta|^p)} \right. \\ \left. - \frac{\log\{\sigma^p + |x(i) - \theta|^p\}}{p^2} \right. \\ \left. - \frac{1}{p^2} \Psi\left(\frac{2}{p}\right) + \frac{1}{p^2} \Psi\left(\frac{1}{p}\right) \right] = 0, \quad (30) \end{aligned}$$

where  $g \equiv g(\mathbf{x}; \theta, \sigma, p)$  and  $\Psi(x)$  is the digamma function. (The digamma function is defined as  $\Psi(x) = (d/dx)\Gamma(x)$ ,

where  $\Gamma(x)$  is the Gamma function.) It can be noticed that (28) is the implicit equation for the M-GC estimator with  $\psi$  as defined in (18), implying that the location estimate has the same properties derived above.

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

**Multiparameter Estimation Algorithm.** For a given set of data  $\{x(i)\}_{i=1}^N$ , we propose to find the optimal joint parameter estimates by the iterative algorithm details in Algorithm 1, with the superscript denoting iteration number.

The algorithm is essentially an iterated conditional mode (ICM) algorithm [27]. Additionally, it resembles the expectation maximization (EM) algorithm [28] in the sense that, instead of optimizing all parameters at once, it finds the optimal value of one parameter given that the other two are fixed; it then iterates. While the algorithm converges to a local minimum, experimental results show that initializing  $\theta$  as the sample median and  $\sigma$  as the median absolute deviation (MAD), and then computing  $p$  as a solution to (30), accelerates the convergence and most often yields globally optimal results. In the classical literature-fixed-point algorithms are successfully used in the computation of  $M$ -estimates [3, 4]. Hence, in the following, we solve items 3–5 in Algorithm 1 using fixed-point search routines.

**Fixed-Point Search Algorithms.** Recall that when  $0 < p \leq 1$ , the solution is the input sample that minimizes the objective function. We solve (28) for the  $1 < p \leq 2$  case using the fixed-point recursion, which can be written as

$$\hat{\theta}_{(j+1)} = \frac{\sum_{i=1}^N w_i(\hat{\theta}_{(j)}) x(i)}{\sum_{i=1}^N w_i(\hat{\theta}_{(j)})} \quad (31)$$

with  $w_i(\hat{\theta}_{(j)}) = p|x(i) - \hat{\theta}_{(j)}|^{p-2}/(\sigma^p + |x(i) - \hat{\theta}_{(j)}|^p)$  and where the subscript denotes the iteration number. The algorithm is taken as convergent when  $|\hat{\theta}_{(j+1)} - \hat{\theta}_{(j)}| < \delta_1$ , where  $\delta_1$  is a small positive value. The median is used as the initial estimate, which typically results in convergence to a (local) minima within a few iterations.

TABLE 1: Multiparameter Estimation Results for GCD Process with length  $N$  and  $(\theta, \sigma, p) = (0, 1, 2)$ .

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

Similarly, for (29) the recursion can be written as

$$\hat{\sigma}_{(j+1)} = \left( \frac{\sum_{i=1}^N b_i(\hat{\sigma}_{(j)}) x(i)}{\sum_{i=1}^N b_i(\hat{\sigma}_{(j)})} \right)^{1/p} \quad (32)$$

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

The parameter  $p$  recursion is given by

$$\hat{p}_{(j+1)} = \frac{2}{N} \sum_{i=1}^N \left[ \Psi \left( \frac{2}{\hat{p}_{(j)}} \right) - \Psi \left( \frac{1}{\hat{p}_{(j)}} \right) + \log \left\{ \sigma^{\hat{p}_{(j)}} + |x(i) - \theta|^{\hat{p}_{(j)}} \right\} + \frac{\hat{p}_{(j)} \left( \sigma^{\hat{p}_{(j)}} \log \sigma - |x(i) - \theta|^{\hat{p}_{(j)}} \log |x(i) - \theta| \right)}{\sigma^{\hat{p}_{(j)}} - |x(i) - \theta|^{\hat{p}_{(j)}}} \right]. \quad (33)$$

Noting that the search space is the interval  $I = (0, 2]$ , the function  $g$  (27) can be evaluated for a finite set of points  $\mathcal{P} \in I$ , keeping the value that maximizes  $g$ , setting it as the initial point for the search.

As an example, simulations illustrating the developed multiparameter estimation algorithm are summarized in Table 1, for  $p = 2$ ,  $\theta = 0$ , and  $\sigma = 1$  (standard Cauchy distribution). Results are shown for varying sample lengths: 10, 100, and 1000. The experiments were run 1000 times for each block length, with the presented results the average on the trials. Mean final  $\theta, \sigma$ , and  $p$  estimates are reported as well as the resulting MSE. To illustrate that the algorithm converges in a few iterations, given the proposed initialization, consider an experiment utilizing data drawn from a GCD  $\theta = 0$ ,  $\sigma = 1$ , and  $p = 1.5$  distribution. Figure 3 reports  $\theta, \sigma, p$  estimate MSE curves. As in the previous case, 100 trials are averaged. Only the first five iteration points are shown, as the algorithms are convergent at that point.

To conclude this section, we consider the computational complexity of the proposed multiparameter estimation algorithm. The algorithm in total has a higher computational complexity than the FLOM, median, meridian, and myriad operators, since Algorithm 1 requires initial estimates of the location and the scale parameters. However, it should be noted that the proposed method estimates *all* the parameters

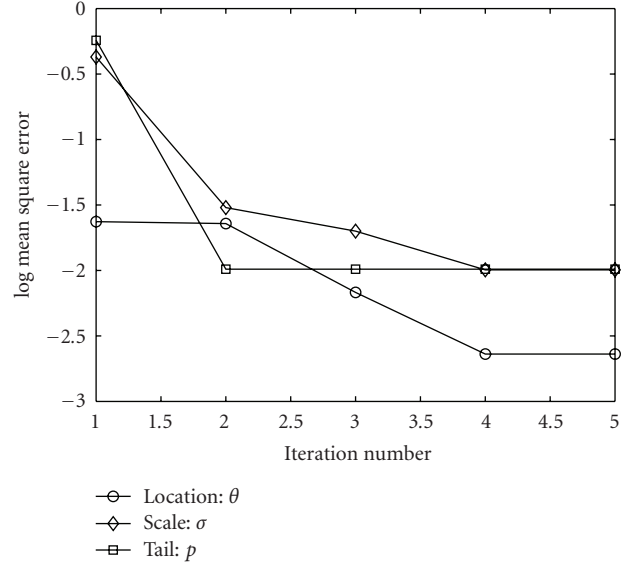


FIGURE 3: Multiparameter estimation MSE iteration evolution for a GCD process with  $(\theta, \sigma, P) = (0, 1, 1.5)$ .

of the model, thus providing advantage over the aforementioned methods that require *a priori* parameter tuning. It is straightforward to show that the computational complexity of the proposed method is  $\mathcal{O}(N^2)$ , assuming the practical case in which the number of fixed-point iterations is  $\ll N$ . The dominating  $N^2$  term is the cost of selecting the input sample that minimizes the objective function, that is, the cost of evaluating the objective function  $N$  times. However, if faster methods that avoid evaluation of the objective function for all samples (*e.g.*, subsampling methods) are employed, the computational cost is lowered.

#### 4. Robust Distance Metrics

This section presents a family of robust GCD-based error metrics. Specifically, the cost function of the M-GC estimator defined in Section 3.1 is extended to define a quasinorm over  $\mathbb{R}^m$  and a semimetric for the same space—the development is analogous to  $L_p$  norms emanating from the GGD family. We denote these semimetrics as the  $\log-L_p$  ( $LL_p$ ) norms. (Note that for the  $\sigma = 1$  and  $p = 1$  case, this metric defines the  $\log-L$  space in Banach space theory.)

*Definition 5.* Let  $u \in \mathbb{R}^m$ , then the  $LL_p$  norm of  $u$  is defined as

$$\|u\|_{LL_p, \sigma} = \sum_{i=1}^m \log \left\{ 1 + \frac{|u_i|^p}{\sigma^p} \right\}, \quad \sigma > 0. \quad (34)$$

The  $LL_p$  norm is not a norm in the strictest sense since it does not meet the positive homogeneity and subadditivity properties. However, it follows the positive definiteness and a scale invariant properties.

**Proposition 2.** Let  $c \in \mathbb{R}$ ,  $u, v \in \mathbb{R}^m$ , 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  $\delta = \sigma/|c|$ ;
- (iii)  $\|u + v\|_{LL_p,\sigma} = \|v + u\|_{LL_p,\sigma}$ ;
- (iv) let  $C_p = 2^{p-1}$ . Then

$$\|u + v\|_{LL_p,\sigma} \leq \begin{cases} \|u\|_{LL_p,\sigma} + \|v\|_{LL_p,\sigma}, & \text{for } 0 < p \leq 1, \\ \|u\|_{LL_p,\sigma} + \|v\|_{LL_p,\sigma} + m \log C_p, & \text{for } p > 1. \end{cases} \quad (35)$$

*Proof.* Statement 1 follows from the fact that  $\log(1 + a) \geq 0$  for all  $a \geq 0$ , with equality if and only if  $a = 0$ . Statement 2 follows from

$$\sum_{i=1}^m \log \left\{ 1 + \frac{|cu_i|^p}{\sigma^p} \right\} = \sum_{i=1}^m \log \left\{ 1 + \frac{|u_i|^p}{(\sigma/|c|)^p} \right\}. \quad (36)$$

Statement 3 follows directly from the definition of the  $LL_p$  norm. Statement 4 follows from the well-known relation  $|a + b|^p \leq C_p(|a|^p + |b|^p)$ ,  $a, b \in \mathbb{R}$ , where  $C_p$  is a constant that depends only on  $p$ . Indeed, for  $0 < p \leq 1$  we have  $C_p = 1$ , whereas for  $p > 1$  we have  $C_p = 2^{p-1}$  (for further details see [29] for example). Using this result and properties of the  $\log$  function we have

$$\begin{aligned} \|u + v\|_{LL_p,\sigma} &= \sum_{i=1}^m \log \left\{ 1 + \frac{|u_i + v_i|^p}{\sigma^p} \right\} \\ &\leq \sum_{i=1}^m \log \left\{ 1 + \frac{C_p(|u_i|^p + |v_i|^p)}{\sigma^p} \right\} \\ &= \sum_{i=1}^m \log C_p + \log \left\{ \frac{1}{C_p} + \frac{(|u_i|^p + |v_i|^p)}{\sigma^p} \right\} \\ &\leq \sum_{i=1}^m \log C_p + \log \left\{ 1 + \frac{(|u_i|^p + |v_i|^p)}{\sigma^p} \right\} \\ &\leq \sum_{i=1}^m \log \left\{ 1 + \frac{|u_i|^p}{\sigma^p} + \frac{|v_i|^p}{\sigma^p} + \frac{|u_i|^p |v_i|^p}{\sigma^{2p}} \right\} \\ &\quad + m \log C_p \\ &= \sum_{i=1}^m \log \left\{ \left( 1 + \frac{|u_i|^p}{\sigma^p} \right) \left( 1 + \frac{|v_i|^p}{\sigma^p} \right) \right\} \\ &\quad + m \log C_p \\ &= \|u\|_{LL_p,\sigma} + \|v\|_{LL_p,\sigma} + m \log C_p. \end{aligned} \quad (37)$$

The  $LL_p$  norm defines a robust metric that does not heavily penalize large deviations, with the robustness depending on the scale parameter  $\sigma$  and the exponent  $p$ . 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:

$$\sigma^p \|u\|_{LL_p,\sigma} \leq \|u\|_p^p \leq \sigma^p m (e^{\|u\|_{LL_p,\sigma}} - 1). \quad (38)$$

*Proof.* The first inequality comes from the relation  $\log(1 + x) \leq x$ , for all  $x \geq 0$ . Setting  $x_i = |u_i|^p/\sigma^p$  and summing over  $i$  yield the result. The second inequality follows from

$$\begin{aligned} \|u\|_{LL_p,\sigma} &= \sum_{i=1}^m \log \left\{ 1 + \frac{|u_i|^p}{\sigma^p} \right\} \\ &\geq \max_i \log \left\{ 1 + \frac{|u_i|^p}{\sigma^p} \right\} = \log \left\{ 1 + \frac{\|u\|_\infty^p}{\sigma^p} \right\}. \end{aligned} \quad (39)$$

Noting that  $\|u\|_\infty \leq \sigma(e^{\|u\|_{LL_p,\sigma}} - 1)^{1/p}$  and  $\|u\|_p^p \leq m\|u\|_\infty^p$  for all  $p > 0$  gives the desired result.  $\square$

The particular case  $p = 2$  yields the well-known Lorentzian norm. The Lorentzian norm has desirable robust error metric properties.

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

Contour plots of select norms are shown in Figure 4 for the two-dimension case. Figures 4(a) and 4(c) show the  $L_2$  and  $L_1$  norms, respectively, while the  $LL_2$  (Lorentzian) and  $LL_1$  norms (for  $\sigma = 1$ ) are shown in Figures 4(b) and 4(d), respectively. It can be seen from Figure 4(b) that the Lorentzian norm tends to behave like the  $L_2$  norm for points within the unitary  $L_2$  ball. Conversely, it gives the same penalization to large sparse deviations as to smaller clustered deviations. In a similar fashion, Figure 4(d) shows that the  $LL_1$  norm behaves like the  $L_1$  norm for points in the unitary  $L_1$  ball.

## 5. Illustrative Application Areas

This section presents four practical problems developed under the proposed framework: (1) robust filtering for power line communications, (2) robust estimation in sensor networks with noisy channels, (3) robust reconstruction methods for compressed sensing, and (4) robust fuzzy clustering. Each problem serves to illustrate the capabilities and performance of the proposed methods.

**5.1. Robust Filtering.** The use of existing power lines for transmitting data and voice has been receiving recent interest [30, 31]. The advantages of power line communications (PLCs) are obvious due to the ubiquity of power lines and power outlets. The potential of power lines to deliver broadband services, such as fast internet access, telephone,

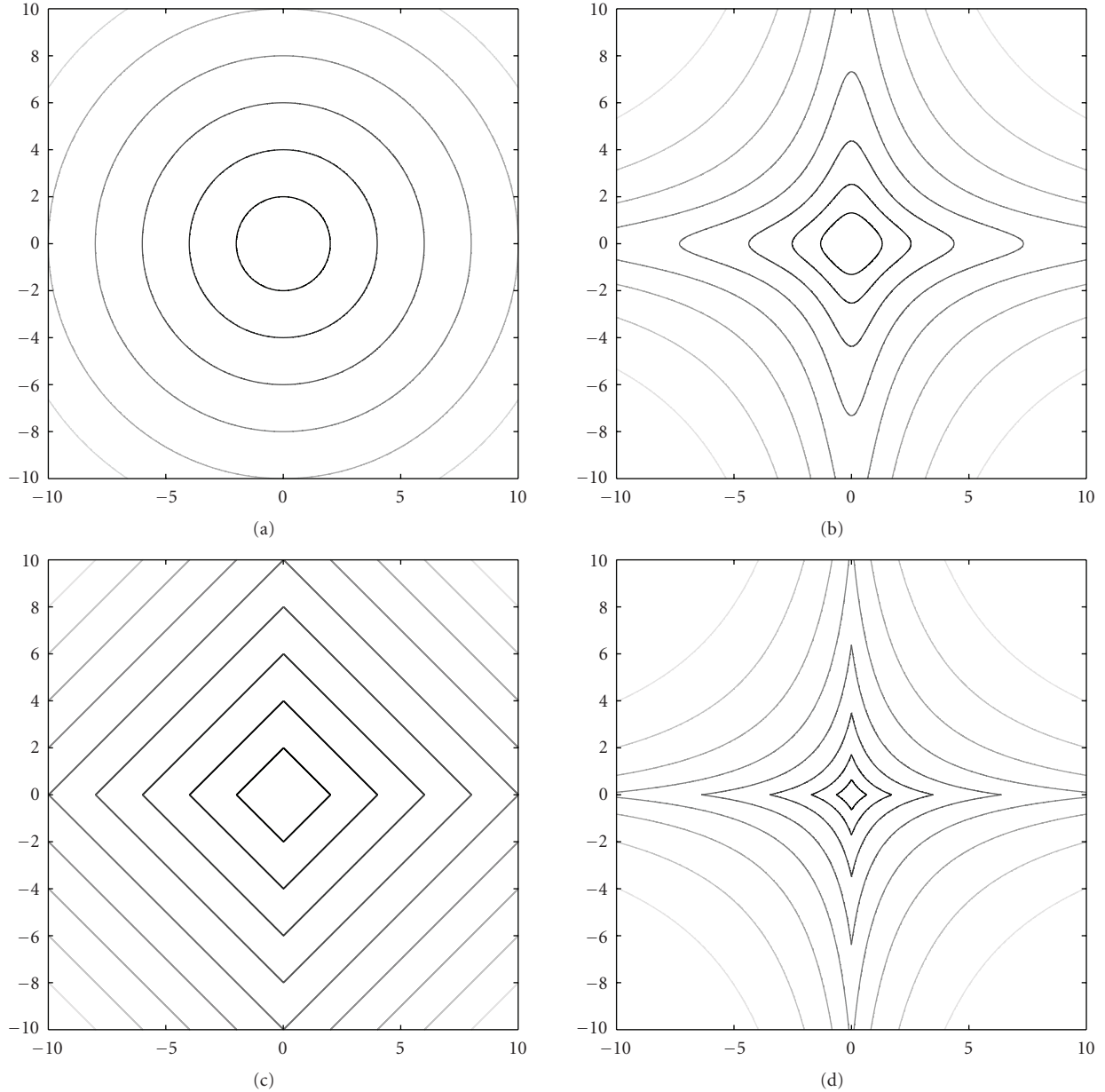


FIGURE 4: Contour plots of different metrics for two dimensions: (a)  $L_2$ , (b)  $LL_2$  (Lorentzian), (c)  $L_1$ , and (d)  $LL_1$  norms.

fax services, and home networking is emerging in new communications industry technology. However, there remain considerable challenges for PLCs, such as communications channels that are hampered by the presence of large amplitude noise superimposed on top of traditional white Gaussian noise. The overall interference is appropriately modeled as an algebraic tailed process, with  $\alpha$ -stable often chosen as the parent distribution [31].

While the M-GC filter is optimal for GCD noise, it is also robust in general impulsive environments. To compare the robustness of the M-GC filter with other robust filtering schemes, experiments for symmetric  $\alpha$ -stable noise corrupted PLCs are presented. Specifically, signal enhancement for the power line communication problem with a 4-ASK

signaling, and equiprobable alphabet  $\nu = \{-2, -1, 1, 2\}$ , is considered. The noise is taken to be white, zero location,  $\alpha$ -stable distributed with  $\gamma = 1$  and  $\alpha$  ranging from 0.2 to 2 (very impulsive to Gaussian noise). The filtering process employed utilizes length nine sliding windows to remove the noise and enhance the signal. The M-GC parameters were determined using the multiparameter estimation algorithm described in Section 3.4. This optimization was applied to the first 50 samples, yielding  $p = 0.756$  and  $\sigma = 0.896$ . The M-GC filter is compared to the FLOM, median, myriad, and meridian operators. The meridian tunable parameter was also set using the multiparameter optimization procedure, but without estimating  $p$ . The myriad filter tuning parameter was set according to the  $\alpha - k$  curve established in [18].

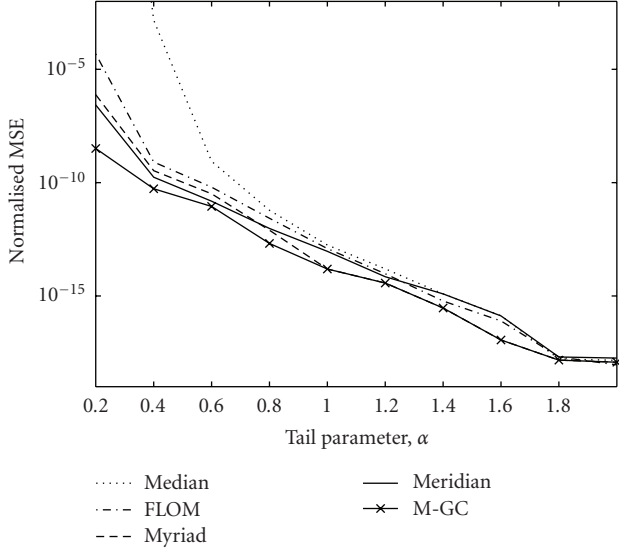


FIGURE 5: Power line communication enhancement. MSE for different filtering structures as function of the tail parameter  $\alpha$ .

The normalized MSE values for the outputs of the different filtering structures are plotted, as a function of  $\alpha$ , in Figure 5. The results show that the various methods perform somewhat similarly in the less demanding light-tailed noise environments, but that the more robust methods, in particular the M-CG approach, significantly outperform in the heavy-tailed, impulsive environments. The time-domain results are presented in Figure 6, which clearly show that the M-GC is more robust than the other operators, yielding a cleaner signal with fewer outliers and well-preserved signal (symbol) transitions. The M-GC filter benefits from the optimization of the scale and tail parameters and therefore perform at least as good as the myriad and meridian filters. Similarly, the M-GC filter performs better than the FLOM filter, which is widely used for processing stable processes [9].

**5.2. Robust Blind Decentralized Estimation.** Consider next a set of  $K$  distributed sensors, each making observations of a deterministic source signal  $\theta$ . The observations are quantized with one bit (binary observations), and then these binary observations are transmitted through a noisy channel to a fusion center where  $\theta$  is estimated (see [32, 33] and references therein). The observations are modeled as  $x = \theta + n$ , where  $n$  are sensor noise samples assumed to be zero-mean, spatially uncorrelated, independent, and identically distributed. Thus the quantized binary observations are

$$b_k = \mathbf{1}\{x_k \in (\tau, +\infty)\} \quad (40)$$

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

$$y = (2b - 1) + w = m + w, \quad (41)$$

where  $w$  are zero-mean independent channel noise samples and the transformation  $m_k = 2b_k - 1$  is made to adopt a binary phase shift keying (BPSK) scheme.

The channel noise density function is denoted by  $w_k \sim f_w(u)$ . When this noise is impulsive (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 [33], modeling the channel corruption as GCD noise and deriving a robust estimation method for impulsive channel noise scenarios. The sensor noise,  $n$ , is modeled as zero-mean additive white Gaussian noise with variance  $\sigma_n^2$ , while the channel noise,  $w$ , is modeled as zero-location additive white GCD noise with scale parameter  $\sigma_w$  and tail constant  $p$ . A realistic approach to the estimation problem in sensor networks assumes that the noise pdf is known but that the values of some parameters are unknown [33]. In the following, we consider the estimation problem when the sensor noise parameter  $\sigma_n$  is known and the channel noise tail constant  $p$  and scale parameter  $\sigma_w$  are unknown.

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

$$\psi(\theta) \triangleq \Pr\{b_k = +1\} = 1 - F_n(\tau - \theta), \quad (42)$$

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

$$f_y(y) = \psi(\theta)f_w(y - 1) + [1 - \psi(\theta)]f_w(y + 1). \quad (43)$$

Note that the resulting pdf is a GCD mixture with mixing parameters  $\psi$  and  $[1 - \psi]$ . To simplify the problem, we first estimate  $\psi = \psi(\theta)$  and then utilize the invariance of the ML estimate to determine  $\theta$  using (42).

Using the log-likelihood function, the ML estimate of  $\psi \in (0, 1)$  reduces to

$$\hat{\psi} = \arg \max_{\psi} \sum_{k=1}^K \log\{\psi f_w(y_k - 1) + [1 - \psi]f_w(y_k + 1)\}. \quad (44)$$

The unknown parameter set for the estimation problem is  $\{\psi, \sigma_w, p\}$ . We address this problem utilizing the well known EM algorithm [28] and a variation of Algorithm 1 in Section 3.4. The followings are the  $E$ - and  $M$ -steps for 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

$$q_k = \frac{\hat{\psi}^{(j)} f_w(y_k - 1 | \Gamma^{(j)})}{\hat{\psi}^{(j)} f_w(y_k - 1 | \Gamma^{(j)}) + [1 - \hat{\psi}^{(j)}] f_w(y_k + 1 | \Gamma^{(j)})}. \quad (45)$$

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

$$\hat{\psi}^{(j+1)} = \frac{1}{K} \sum_{k=1}^K q_k, \quad \Gamma^{(j+1)} = \arg \max_{\Gamma} \Lambda(\Gamma), \quad (46)$$

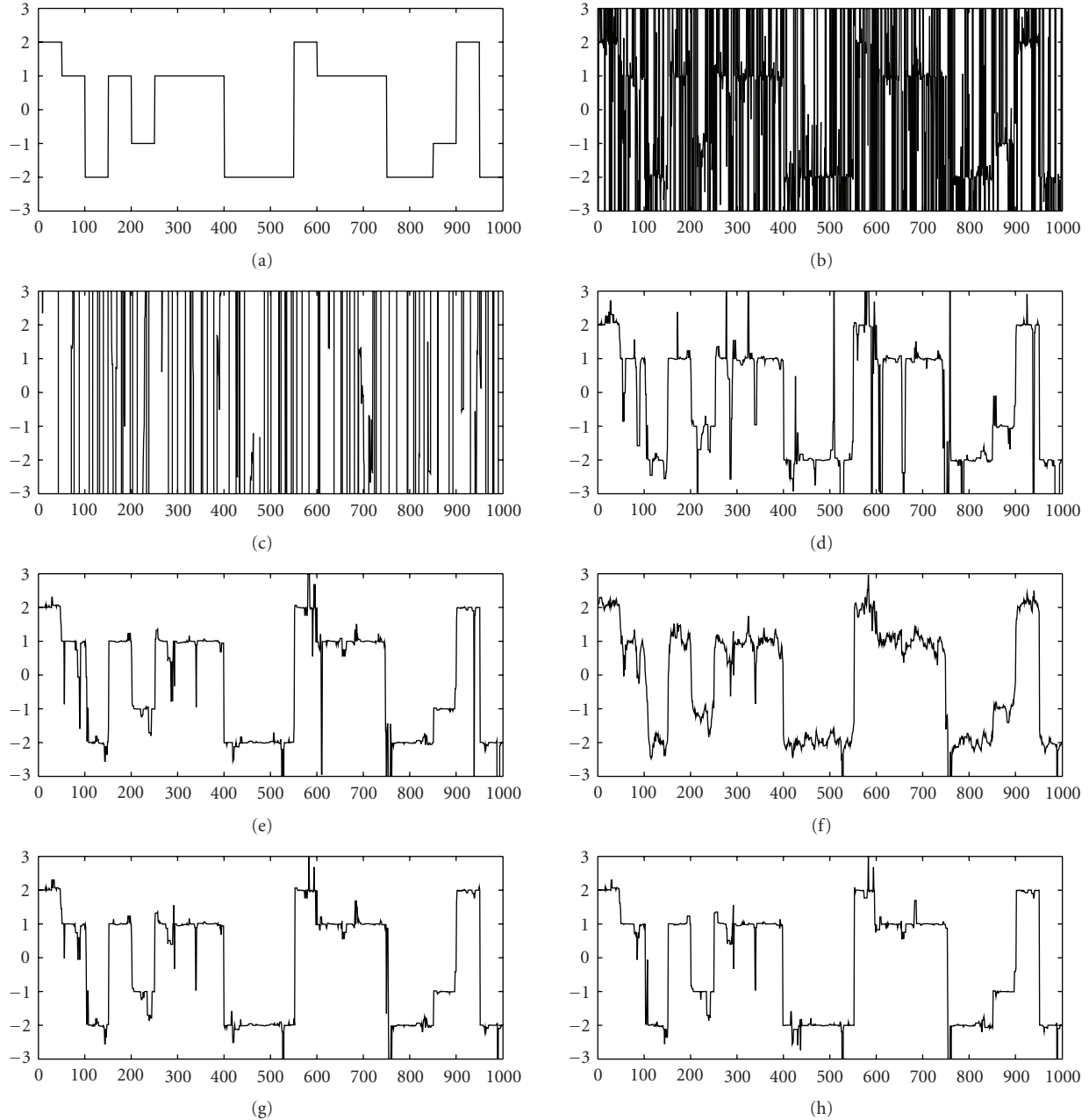


FIGURE 6: Power line communication enhancement. (a) Transmitted signal, (b) Received signal corrupted by  $\alpha$ -stable noise  $\alpha = 0.4$ . Filtering results with: (c) Mean, (d) Median, (e) FLOM  $P = .25$ , (f) Myriad, (g) Meridian, (h) M-GC.

where

$$\Lambda(\Gamma) = \sum_{k=1}^K q_k \Upsilon(y_k - 1; \Gamma) + (1 - q_k) \Upsilon(y_k + 1; \Gamma), \quad (47)$$

where  $\Upsilon(u; \Gamma) = \log a(p) + \log \sigma_w - 2p^{-1} \log(\sigma_w^p + |u|^p)$  and  $a(p) = p\Gamma(2/p)/2(\Gamma(1/p))^2$ . We use a suboptimal estimate of  $p$  in this case, choosing the value from  $P = \{0.5, 1, 1.5, 2\}$  that maximizes (46).

Numerical results comparing the derived GCD method, coined maximum likelihood with unknown generalized Cauchy channel parameters (MLUGC), with the Gaussian

channel-based method derived in [33], referred to as maximum likelihood with unknown Gaussian channel parameter (MLUG), are presented in Figure 7. The MSE is used as a 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  $K = 1000$ , and the results are averaged for 200 independent realizations. For the channel noise we use two models: contaminated  $p$ -Gaussian and  $\alpha$ -stable distributions. Figure 7(a) shows results for contaminated  $p$ -Gaussian noise with the variance set as

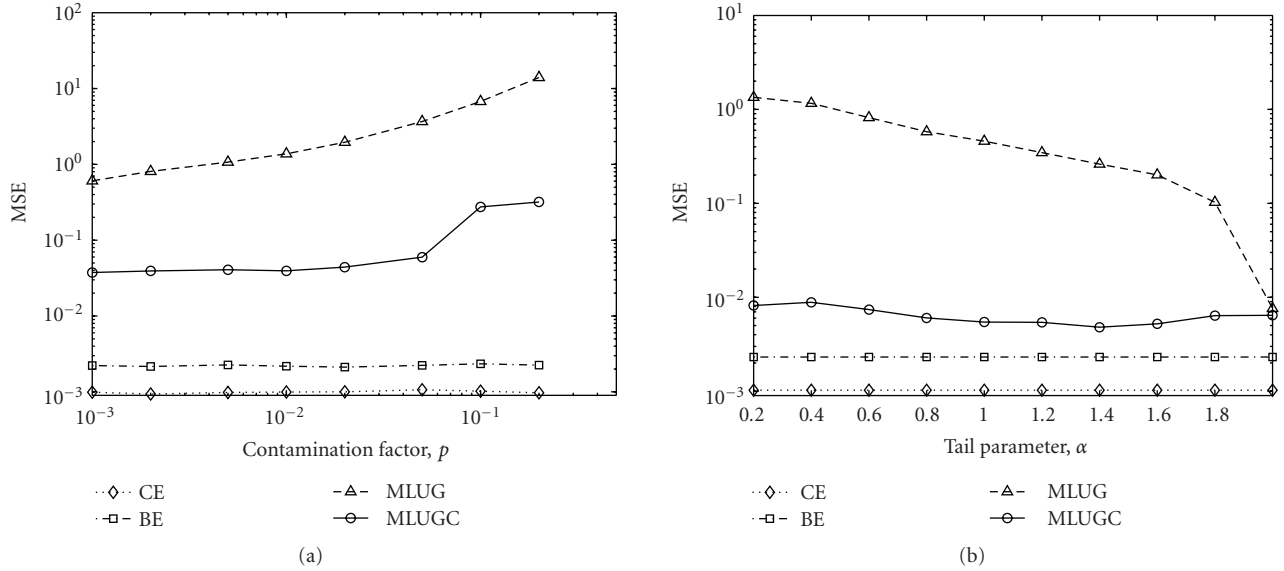


FIGURE 7: Sensor network example with parameters:  $\theta = 1$ ,  $\tau = 0$ ,  $\sigma_n = 1$ , and  $K = 1000$ . Comparison of MLUGC, MLUG, BE, and CE. (a) Channel noise contaminated  $p$ -Gaussian distributed with  $\sigma_w^2 = 0.5$ . MSE as function of the of the contamination parameter,  $p$ . (b) Channel noise  $\alpha$ -stable distributed with  $\sigma_w = 0.5$ . MSE as function of the tail parameter,  $\alpha$ .

$\sigma_w^2 = 0.5$  and varying  $p$  (percentage of contamination) from  $10^{-3}$  to 0.2. The results show a gain of at least an order of magnitude over the Gaussian-derived method. Results for  $\alpha$ -stable distributed noise are shown in Figure 7(b), with scale parameter  $\sigma_w = 0.5$  and the tail parameter,  $\alpha$ , varying from 0.2 to 2 (very impulsive to Gaussian noise). It can be observed that the GCD-derived method has a gain of at least an order of magnitude for all  $\alpha$ . Furthermore, the MLUGC method has a nearly constant MSE for the entire range. It is of note that the MSE of the MLUGC method is comparable to that obtained by the MLUG (Gaussian-derived) for the especial case when  $\alpha = 2$  (Gaussian case), meaning that the GCD-derived method is robust under heavy-tailed and light-tailed environments.

### 5.3. Robust Reconstruction Methods for Compressed Sensing.

As a third example, consider compressed sensing, which is a recently introduced novel framework that goes against the traditional data acquisition paradigm [34]. Take a set of  $m$  sensors making observations of a signal  $x_0 \in \mathbb{R}^n$ . Suppose that  $x_0$  is  $s$ -sparse in some orthogonal basis  $\Psi$ , 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. The measurement process can be modeled as  $y = \Phi x_0 + z$ , where  $\Phi$  is an  $m \times n$  matrix with vectors  $\phi_i$  as rows and  $z$  is white additive noise (with possibly impulsive behavior). The problem is how to estimate  $x_0$  from the noisy measurements  $y$ .

A range of different algorithms and methods have been developed that enable approximate reconstruction of sparse signals from noisy compressive measurements [35–39]. Most such algorithms provide bounds for the  $L_2$  reconstruction error based on the assumption that the corrupting noise is

bounded, Gaussian, or, at a minimum, has finite variance. Recent works have begun to address the reconstruction of sparse signals from measurements corrupted by outliers, for example, due to missing data in the measurement process or transmission problems [40, 41]. These works are based on the sparsity of the measurement error pattern to first estimate the error and then estimate the true signal, in an iterative process. A drawback of this approach is that the reconstruction relies on the error sparsity to first estimate the error, but if the sparsity condition is not met, the performance of the algorithm degrades.

Using the arguments above, we propose to use a robust metric derived in Section 4 to penalize the residual and address the impulsive sampling noise problem. Utilizing the strong theoretical guarantees of basis pursuit (BP)  $L_1$  minimization, for sparse recovery of underdetermined systems of equations (see [34]), we propose the following nonlinear optimization problem to estimate  $x_0$  from  $y$ :

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ subject to } \|y - \Phi x\|_{LL_2, \gamma} \leq \epsilon. \quad (48)$$

The following result presents an upper bound for the reconstruction error of the proposed estimator and is based on restricted isometry properties (RIPs) of the matrix  $\Phi$  (see [34, 42] and references therein for more details on RIPs).

**Theorem 1** (see [42]). *Assume the matrix  $\Phi$  meets an RIP, then for any  $s$ -sparse signal  $x_0$  and observation noise  $z$  with  $\|z\|_{LL_2, \gamma} \leq \epsilon$ , the solution to (48), denoted as  $x^*$ , obeys*

$$\|x^* - x_0\|_2 \leq C_s \cdot 2\gamma \cdot \sqrt{m(e^\epsilon - 1)}, \quad (49)$$

where  $C_s$  is a small constant.

Notably,  $\gamma$  controls the robustness of the employed norm and  $\epsilon$  the radius of the feasibility set  $LL_2$  ball. Let  $Z$  be a

Cauchy random variable with scale parameter  $\sigma$  and location parameter zero. 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  $\epsilon$  and set  $\gamma$  as  $\text{MAD}(y)$ .

Debiasing is achieved through robust regression on a subset of  $\hat{x}$  indexes using the Lorentzian norm. The subset is set 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

$$\tilde{x}_I = \arg \min_{x \in \mathbb{R}^d} \|y - \Phi_I x\|_{LL_2, \sigma}, \quad (50)$$

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  $I$ . The reconstruction algorithm composed of solving (48) followed by the debiasing step is referred to as Lorentzian basis pursuit (BP) [42].

Experiments evaluating the robustness of Lorentzian BP in different impulsive sampling noises are presented, comparing performance with traditional CS reconstruction algorithms orthogonal matching pursuit (OMP) [38] and basis pursuit denoising (BPD) [34]. The signals are synthetic  $s$ -sparse signals with  $s = 10$  and length  $n = 1024$ . The number of measurements is  $m = 128$ . For OMP and BPD, the noise bound is set as  $\epsilon = m\sigma^2$ , where  $\sigma$  is the scale parameter of the corrupting distributions. The results are averaged over 200 independent realizations.

For the first scenario, we consider contaminated  $p$ -Gaussian as the model for the sampling noise, with  $\sigma^2 = 10^{-2}$ , resulting in an SNR of 18.9 dB when no contamination is present ( $p = 0$ ). The amplitude of the outliers is set as  $\delta = 10^3$ , and  $p$  is varied from  $10^{-3}$  to 0.5. The results are shown in Figure 8(a), which demonstrates that Lorentzian BP significantly outperforms BPD and OMP. Moreover, the Lorentzian BP results are stable over a range of contamination factors  $p$ , up to 5% of the measurements, making it a desirable method when measurements are lost or erased.

The second experiment explores the behavior of Lorentzian BP in  $\alpha$ -stable environments. The  $\alpha$ -stable noise scale parameter is set as  $\sigma = 0.1$  ( $\gamma$  in the traditional characterization) for all cases, and the tail parameter,  $\alpha$ , is varied from 0.2 to 2, that is, very impulsive to the Gaussian case. The results are summarized in Figure 8(b), which shows that all methods perform poorly for small values of  $\alpha$ , with Lorentzian BP yielding the most acceptable results. Beyond  $\alpha = 0.8$ , Lorentzian BP produces faithful reconstructions with an SNR greater than 20 dB, and often 30 dB greater than BPD and OMP results. Also of importance is that when  $\alpha = 2$  (Gaussian case), the performance of Lorentzian BP is comparable with that of BPD and OMP, which are Gaussian-derived methods. This result shows the robustness of Lorentzian BP under a broad range of noise models, from very impulsive heavy-tailed to light-tailed environments.

**5.4. Robust Clustering.** As a final example, we present a robust fuzzy clustering procedure based on the  $LL_p$  metrics defined in Section 4, which is suitable for clustering data points involving heavy-tailed nonGaussian processes. Dave proposed the *noise clustering* (NC) algorithm to address

noisy data in [43, 44]. The NC approach is successful in improving the robustness of a variety of prototype-based clustering methods. This method considers the noise as a separate class and represents it by a prototype that has a constant distance  $\delta$ .

Let  $\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^N$ ,  $\mathbf{x}_j \in \mathbb{R}^n$ , be a finite data set and  $C$  the given number of clusters. NC partitions the data set by minimizing the following function proposed in [43]:

$$J(\mathbf{Z}) = \sum_{i=1}^C \sum_{j=1}^N (u_{ij})^m d(\mathbf{x}_j, \mathbf{z}_i) + \sum_{j=1}^N \delta \left(1 - \sum_{i=1}^C u_{ij}\right)^m, \quad (51)$$

where  $\mathbf{Z} = [\mathbf{z}_1; \dots; \mathbf{z}_C]$  is a matrix whose rows are the cluster centers,  $m \in (1, \infty)$  is a weighting exponent, and  $d(\mathbf{x}_j, \mathbf{z}_i)$  is the squared  $L_2$  distance from a data point  $\mathbf{x}_j$  to the center  $\mathbf{z}_i$ .  $\mathbf{U} = [u_{ij}]$  is a  $C \times N$  matrix, called a constraint fuzzy partition of  $\mathcal{X}$ , which satisfies [43]

$$\begin{aligned} u_{ij} &\in [0, 1] \quad \forall i, j, \\ 0 &< \sum_{j=1}^N u_{ij} < N \quad \forall i, \\ \sum_{i=1}^C u_{ij} &< 1 \quad \forall j. \end{aligned} \quad (52)$$

The  $u_{ij}$  weight represents the membership of the  $i$ -th sample to the  $j$ -th cluster. Minimization of the objective function with respect to  $\mathbf{U}$ , subject to the constraints in (52), gives [43]

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left[ \frac{d(\mathbf{x}_j, \mathbf{z}_i)}{d(\mathbf{x}_j, \mathbf{z}_k)} \right]^{1/(m-1)} + \left[ \frac{d(\mathbf{x}_j, \mathbf{z}_i)}{\delta} \right]^{1/(m-1)}}. \quad (53)$$

Compared with the basic fuzzy  $C$ -means (FCM), the membership constraint is relaxed to  $\sum_{i=1}^C u_{ij} < 1$ . The second term in the denominator of (53) becomes large for outliers, thus yielding small membership values and improving robustness of prototype-based clustering algorithms.

To further improve robustness, we propose the application of  $LL_p$  metrics in the NC approach. Substituting the  $LL_p$  norm for  $d$  in (51) yields the objective function

$$J(\mathbf{Z}) = \sum_{i=1}^C \sum_{j=1}^N (u_{ij})^m \|\mathbf{x}_j - \mathbf{z}_i\|_{LL_p, \sigma} + \sum_{j=1}^N \delta \left(1 - \sum_{i=1}^C u_{ij}\right)^m. \quad (54)$$

Given the objective function  $J(\mathbf{Z})$ , a set of vectors  $\{\mathbf{z}_i\}_{i=1}^N$  that minimize  $J(\mathbf{Z})$  must be determined. As in FCM, fixed-point iterations are utilized to obtain the solution. We use a variation of the fixed point recursion proposed in Section 3.4 to achieve this goal. Differentiating  $J(\mathbf{Z})$  with respect to each dimension  $l$  of  $\mathbf{z}_s$ , treating the  $u_{ij}$  terms as constants, and setting it to zero yield the fixed point function. Thus the recursion algorithm can be written as

$$z_{sl}(t+1) = \frac{\sum_{j=1}^N w_j(t) x_{jl}}{\sum_{j=1}^N w_j(t)} \quad (55)$$

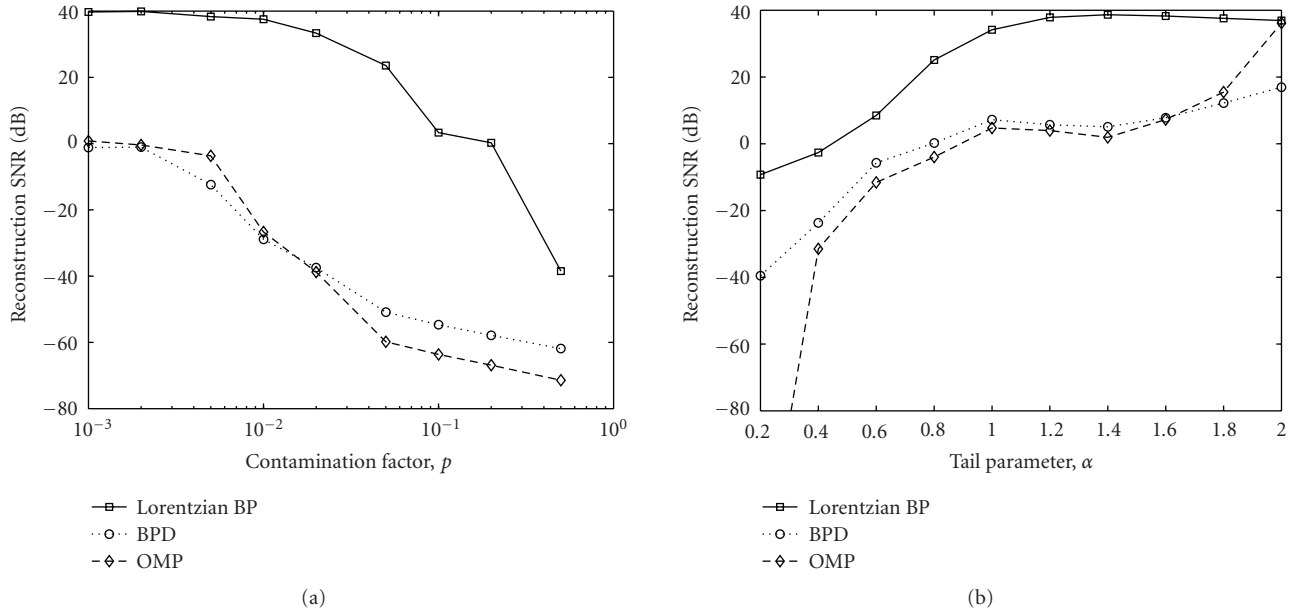


FIGURE 8: Comparison of Lorentzian BP with BPD and OMP for impulsive contaminated samples. (a) Contaminated  $p$ -Gaussian,  $\sigma^2 = 0.01$ . R-SNR as a function of the contamination parameter,  $p$ . (b)  $\alpha$ -stable noise,  $\sigma = 0.1$ . R-SNR as a function of the tail parameter,  $\alpha$ .

**Require:** cluster number  $C$ , weighting parameter  $m$ ,  $\delta$ , maximum number of iterations or terminate parameter  $\epsilon$ .

- (1) Initialize cluster centers.
- (2) **While**  $\|z_s(t+1) - z_s(t)\| > \epsilon$  or a maximum number of iterations is not reached **do**
- (3)   Compute the fuzzy set  $U$  using (53) and
- (4)   Update cluster centers using (55).
- (5) **end while**
- (6) **return** Cluster centroids  $Z = [z_1; \dots; z_C]$ .

ALGORITHM 2:  $LL_p$ -based noise clustering algorithm.

with

$$w_j(t) = \frac{u_{sj}^m p |x_{jl} - z_{sl}(t)|^{p-2}}{\sigma^p + |x_{jl} - z_{sl}(t)|^p}, \quad (56)$$

where  $t$  denotes the iteration number. The recursion is terminated when  $\|z_s(t+1) - z_s(t)\|_2 < \epsilon$  for some given  $\epsilon > 0$ . This method is used to find the update of the cluster centers. Alternation of (53) and (55) gives an algorithm to find the cluster centers that converge to a local minimum of the cost function.

In the NC approach,  $m = 1$  corresponds to crisp memberships, and increasing  $m$  represents increased fuzziness and soft rejection of outliers. When  $m$  is too large, spurious cluster may exist. The choice of the constant distance  $\delta$  also influences the fuzzy membership; if it is too small, then we cannot distinguish good clusters from outliers, and if it is too large, the result diverges from the basic FCM. Based on [43], we set  $\delta = (\lambda/N^2) \sum_{i \neq j}^N \|\mathbf{x}_i - \mathbf{x}_j\|_{LL_p, \sigma}$ , where  $\lambda$  is a scale parameter. In order to reduce the local minimum caused by initialization of the NC approach, we use classical

$k$ -means on a small subset of the data to initialize a set of cluster centers. The proposed algorithm is summarized in Algorithm 2 and is coined the  $LL_p$ -based Noise Clustering ( $LL_p$ -NC) algorithm.

Experimental results show that for multigroup heavy-tailed process, the results of the  $LL_p$  based method generally converges to the global minimum. However, to address the problem of local minima, the clustering algorithm is performed multiple times with different random initializations (subsets randomly sampled) and with a fixed small number of iterations. The best result is selected as the final solution.

Simulations to validate the performance of GCD-based clustering algorithm ( $LL_p$ -NC) in heavy tailed environments are carried out and summarized in Table 2. The experiment uses three synthetic data sets of 400 points each with different distributions and 100 points in each cluster. The locations of the centers for the three sets are  $[-6, 2]$ ,  $[-2, -2]$ ,  $[2, 4]$ , and  $[3, 0]$  for each set. The first set has Cauchy distributed clusters (GCD,  $p = 2$ ) with  $\sigma = 1$  and is shown in Figure 9. The second has the meridian distribution (GCD,  $p = 1$ ), with  $\sigma = 1$ . The meridian is a very impulsive distribution. The third set has a two-dimensional  $\alpha$ -stable distribution with  $\alpha = 0.9$

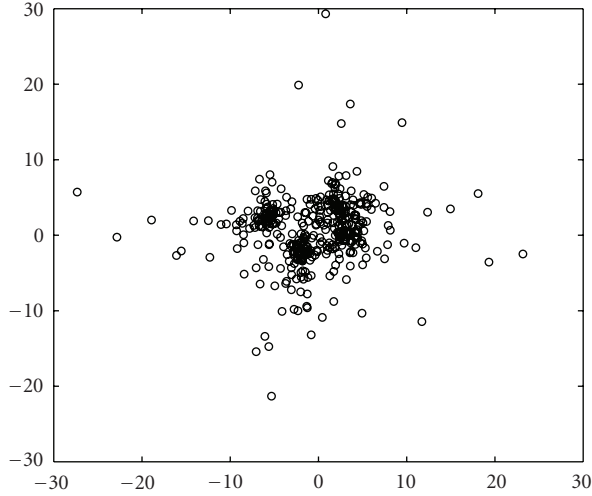


FIGURE 9: Data set for clustering example 1. Cauchy distributed samples with cluster centers  $[-6, 2]$ ,  $[-2, -2]$ ,  $[2, 4]$ , and  $[3, 0]$ .

TABLE 2: Clustering results for GCD processes and  $\alpha$ -stable process.

$N$	MSE	MAD	$LL_p$	Average Distance
$LL_p$ -NC	0.34987	0.62897	0.0968	Cauchy
$L_1$ -NC	1.8186	1.8361	0.1262	15.39
Similarity-based	1.6513	1.136	0.18236	
$LL_p$ -NC	0.85197	0.9283	0.1521	Meridian
$L_1$ -NC	5.887	2.7311	0.5573	50.363
Similarity-based	5.2309	2.4627	1.8416	
$LL_p$ -NC	0.50408	0.73618	0.1896	$\alpha$ -stable
$L_1$ -NC	3.2105	2.7684	0.2174	44.435
Similarity-based	1.7578	1.6322	1.0112	

and  $\gamma = 1$ , which is also a very impulsive case. The algorithm was run 200 times for each set with different initializations, setting the maximum number of iterations to 50,  $\epsilon = 0.0001$ , and  $\lambda = 0.1$ .

To evaluate the results, we calculate the MSE, the mean absolute deviation (MAD), and the  $LL_p$  distance between the solutions and the true cluster centers, averaging the results for 200 trials. The  $LL_p$  NC approach is compared with classical NC employing the  $L_1$  distance and the similarity-based method in [45]. The average  $L_2$  distance between all points in the set (AD) is shown as a reference for each sample set. As the results show, GCD-based clustering outperforms both traditional NC and similarity-based methods in heavy-tailed environments. Of note is the meridian case, which is a very impulsive distribution. The GCD clustering results are significantly more accurate than those obtained by the other approaches.

## 6. Concluding Remarks

This paper presents a GCD-based theoretical approach that allows the formulation of challenging problems in a robust fashion. Within this framework, we establish a statistical

relationship between the GGD and GCD families. The proposed framework, due to its flexibility, subsumes GGD-based developments, thereby guaranteeing performance improvements over the traditional problem formulation techniques. Properties of the derived techniques are analyzed. Four particular applications are developed under this framework: (1) robust filtering for power line communications, (2) robust estimation in sensor networks with noisy channels, (3) robust reconstruction methods for compressed sensing, and (4) robust fuzzy clustering. Results from the applications show that the proposed GCD-derived methods provide a robust framework in impulsive heavy-tailed environments, with performance comparable to existing methods in less demanding light-tailed environments.

## Appendices

### A. Proof of Lemma 1

Let  $X$  be the RV formed as the ratio of two RVs,  $U$  and  $V$ :  $X = U/V$ . In the case where  $U$  and  $V$  are independent, the pdf of the RV  $X$ ,  $f_X(\cdot)$ , is given by [46]

$$f_X(x) = \int_{-\infty}^{\infty} |v| f_U(xv) f_V(v) dv, \quad (\text{A.1})$$

where  $f_U(\cdot)$  and  $f_V(\cdot)$  denote the pdf s of  $U$  and  $V$ , respectively. Replacing the GGD in (A.1) and manipulating the obtained expression yield

$$\begin{aligned} f_X(x) &= \mathcal{C}(\alpha_U, \beta) \mathcal{C}(\alpha_V, \beta) \\ &\times \int_{v=-\infty}^{\infty} |v| \exp\left\{-\left(\frac{|xv|}{\alpha_U}\right)^\beta\right\} \exp\left\{-\left(\frac{|v|}{\alpha_V}\right)^\beta\right\} dv, \end{aligned} \quad (\text{A.2})$$

where  $\mathcal{C}(\alpha, \beta) \triangleq \beta / (2\alpha\Gamma(1/\beta))$ . Noting that  $|ab| = |a||b|$  and dividing the integral give

$$\begin{aligned} f_X(x) &= \mathcal{C}(\alpha_U, \beta) \mathcal{C}(\alpha_V, \beta) \\ &\times \left[ \int_{v>0} v \exp\left\{-v^\beta \mathcal{K}(\alpha_U, \alpha_V, \beta, x)\right\} dv \right. \\ &\quad \left. - \int_{v\leq 0} v \exp\left\{-(-v)^\beta \mathcal{K}(\alpha_U, \alpha_V, \beta, x)\right\} dv \right], \end{aligned} \quad (\text{A.3})$$

where

$$\mathcal{K}(\alpha_U, \alpha_V, \beta, x) \triangleq \frac{|x|^\beta}{\alpha_U^\beta} + \frac{1}{\alpha_V^\beta}. \quad (\text{A.4})$$

Consider first

$$\mathcal{I}_1(v) \triangleq \int_{v>0} v \exp\left\{-v^\beta \mathcal{K}(\alpha_U, \alpha_V, \beta, x)\right\} dv. \quad (\text{A.5})$$

Letting  $z = v^\beta \mathcal{K}(\alpha_U, \alpha_V, \beta, x)$ , after some manipulations, yields

$$\mathcal{I}_1(v) = \frac{1}{\beta \mathcal{K}^{2/\beta}(\alpha_U, \alpha_V, \beta, x)} \int_{z>0} z^{(2/\beta)-1} \exp(-z) dz. \quad (\text{A.6})$$



Noting that

$$\int_{z>0} z^{(2/\beta)-1} \exp(-z) dz = \Gamma\left(\frac{2}{\beta}\right) \quad (\text{A.7})$$

gives

$$\mathcal{J}_1(\nu) = \frac{1}{\beta \mathcal{K}^{2/\beta}(\alpha_U, \alpha_V, \beta, x)} \Gamma\left(\frac{2}{\beta}\right). \quad (\text{A.8})$$

Consider next

$$\mathcal{J}_2(\nu) \triangleq \int_{\nu \leq 0} \nu \exp\{-(-\nu)^\beta \mathcal{K}(\alpha_U, \alpha_V, \beta, x)\} d\nu. \quad (\text{A.9})$$

Letting  $w = -\nu$ , it is easy to see that  $\mathcal{J}_2(\nu) = -\mathcal{J}_1(\nu)$ , thus  $\mathcal{J}_1(\nu) - \mathcal{J}_2(\nu) = 2\mathcal{J}_1(\nu)$ . Thus,

$$f_X(x) = \mathcal{C}(\alpha_U, \beta) \mathcal{C}(\alpha_V, \beta) 2\mathcal{J}_1 \quad (\text{A.10})$$

gives the desired result after substituting the corresponding expressions and letting  $\alpha_U/\alpha_V = \nu$  and  $\beta = \lambda$ .

## B. Proof of Proposition 1

(1) Differentiating  $Q(\theta)$  yields

$$Q'(\theta) = \sum_{i=1}^N \frac{-p|x(i) - \theta|^{p-1} \text{sgn}(x(i) - \theta)}{\sigma^p + |x(i) - \theta|^p}. \quad (\text{B.1})$$

For  $\theta < x_{[1]}$ ,  $\text{sgn}(x(i) - \theta) = 1$  for all  $i$ . Then  $Q'(\theta) < 0$ , which implies that  $Q(\theta)$  is strictly decreasing in that interval. Similarly for  $\theta > x_{[N]}$ ,  $\text{sgn}(x(i) - \theta) = -1$  for all  $i$  and  $Q'(\theta) > 0$ , showing that the function is strictly increasing in that interval.

(2) From (1) we see that  $Q'(\theta) \neq 0$  if  $\theta \notin [x_{[1]}, x_{[N]}]$ , then all local extrema of  $Q(\theta)$  lie in the interval  $[x_{[1]}, x_{[N]}]$ .

(3) Let  $x_{[k]} < \theta < x_{[k+1]}$  for any  $k \in 1, 2, \dots, N-1$ . Then the objective function  $Q(\theta)$  becomes

$$\begin{aligned} Q(\theta) &= \sum_{i=1}^k \log\{\sigma^p + (\theta - x(i))^p\} \\ &+ \sum_{i=k+1}^N \log\{\sigma^p + (x(i) - \theta)^p\}. \end{aligned} \quad (\text{B.2})$$

The second derivative with respect to  $\theta$  is

$$\begin{aligned} Q''(\theta) &= \sum_{i=1}^k \frac{p(p-1)(\theta - x(i))^{p-2} \sigma^p - p(\theta - x(i))^{2p-2}}{(\sigma^p + (\theta - x(i))^p)^2} \\ &+ \sum_{i=k+1}^N \frac{p(p-1)(x(i) - \theta)^{p-2} \sigma^p - p(x(i) - \theta)^{2p-2}}{(\sigma^p + (x(i) - \theta)^p)^2}. \end{aligned} \quad (\text{B.3})$$

From (B.3) it can be seen that if  $0 < p \leq 1$ , then  $Q''(\theta) < 0$  for  $x_{[k]} < \theta < x_{[k+1]}$ , therefore  $Q(\theta)$  is concave in the intervals  $I_k = (x_{[k]}, x_{[k+1]})$ ,

$k \in 1, 2, \dots, N-1$ . If all the extrema points lie in  $[x_{[1]}, x_{[N]}]$ , the function is concave in  $I_k$ , and since the function is not differentiable in the input samples  $\{x(i)\}_{i=1}^N$  (critical points), then the only possible local minimums of the objective function are the input samples.

(4) Consider the  $i$ th term in  $Q(\theta)$  and define

$$q_i(\theta) = \log\{\sigma^p + |x(i) - \theta|^p\}. \quad (\text{B.4})$$

Clearly for each  $q_i(\theta)$  there exists a unique minima in  $\theta = x(i)$ . Also, it can be easily shown that  $q_i(\theta)$  is convex in the interval  $[x(i) - a, x(i) + a]$ , where  $a = \sigma(p-1)^{1/p}$ , and concave outside this interval (for  $1 < p \leq 2$ ). The proof of this statement is divided in two parts. First we consider the case when  $N = 2$  and show that there exist at most  $2N - 1 (= 3)$  local extrema for this case. Then by induction we generalize this result for any  $N$ .

Let  $N = 2$ . If  $|x_{[2]} - x_{[1]}| < a$  the cost function is convex in the interval  $[x_{[1]}, x_{[2]}]$  since it is the sum of two convex functions (in that interval). Thus,  $Q(\theta)$  has a unique minimizer. Now if  $|x_{[2]} - x_{[1]}| \geq a$ , the cost function has at most one inflexion point (local maxima) between  $(x_{[1]}, x_{[2]})$  and at most two local minima in the neighborhood of  $x_{[1]}$  and  $x_{[2]}$  since  $q_i(\theta)$ ,  $i = 1, 2$ , are concave outside the interval  $[x_{[i]} - a, x_{[i]} + a]$ . Then, for  $N = 2$  we have at most  $2N - 1 = 3$  local extrema points.

Suppose that we have  $N = K$  samples. If  $|x_{[K]} - x_{[1]}| < a$ , the cost function is convex in the interval  $[x_{[1]}, x_{[K]}]$  since it is the sum of convex functions (in that interval) and it has only one global minima. Now suppose that  $|x_{[K]} - x_{[1]}| \geq a$ , and also suppose that there are at most  $2K - 1$  local extrema points. Let  $x(K+1)$  be a new sample in the data set, and without loss of generality assume that  $x(K+1) > x_{[K]}$ .

If  $|x(K-1) - x_{[K]}| < a$ , the new sample will not add a new extrema point to the cost function, due to convexity of  $q_{K+1}(\theta)$  for the interval  $[x(K+1) - a, x(K+1) + a]$  and the fact that  $Q(\theta)$  is strictly increasing for  $\theta > x_{[K]}$ . If  $|x(K-1) - x_{[K]}| \geq a$ , the new sample will add at most two local extrema points (one local maxima and one local minima) in the interval  $(x_{[K]}, x(K+1))$ . The local maxima is an inflexion point between  $(x_{[K]}, x(K+1))$ , and the local minima is in the neighborhood of  $x(K+1)$ . Therefore, the total number of extrema points for  $N = K+1$  is at most  $2K - 1 + 2 = 2(K+1) - 1$ , which is the claim of the statement. This concludes the proof.

## C. Proof of Property 1

Using the properties of the argmin function, the M-GC estimator can be expressed as

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \log \left\{ 1 + \frac{|x(i) - \theta|^p}{\sigma^p} \right\}. \quad (\text{C.1})$$

Let  $\delta = \sigma^p$ . Since multiplying by a constant does not affect the result of the arg min operator, we can rewrite (C.1) as

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \delta \log \left\{ 1 + \frac{|x(i) - \theta|^p}{\delta} \right\}. \quad (\text{C.2})$$

Using the fact that  $a \log b = \log b^a$  and taking the limit as  $\delta \rightarrow \infty$  yield

$$\begin{aligned} \lim_{\delta \rightarrow \infty} \hat{\theta} &= \lim_{\delta \rightarrow \infty} \arg \min_{\theta} \sum_{i=1}^N \log \left\{ 1 + \frac{|x(i) - \theta|^p}{\delta} \right\}^{\delta} \\ &= \arg \min_{\theta} \sum_{i=1}^N |x(i) - \theta|^p, \end{aligned} \quad (\text{C.3})$$

where the last step follows since

$$\lim_{\delta \rightarrow \infty} \log \left\{ 1 + \frac{u}{\delta} \right\}^{\delta} = u. \quad (\text{C.4})$$

## D. Proof of Property 2

The M-GC estimator can be expressed as

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \sum_{i=1}^N \log \left\{ 1 + \frac{|x(i) - \theta|^p}{\sigma^p} \right\} \\ &= \arg \min_{\theta} \log \left\{ \prod_{i=1}^N \left[ 1 + \frac{|x(i) - \theta|^p}{\sigma^p} \right] \right\}. \end{aligned} \quad (\text{D.1})$$

Define

$$H_{\sigma}(\theta; \mathbf{x}) = \prod_{i=1}^N \left[ 1 + \frac{|x(i) - \theta|^p}{\sigma^p} \right]. \quad (\text{D.2})$$

Since the log function is monotone nondecreasing, the M-GC estimator can be reformulated as

$$\hat{\theta} = \arg \min_{\theta} H_{\sigma}(\theta; \mathbf{x}). \quad (\text{D.3})$$

It can be checked that when  $\sigma$  is very small

$$H_{\sigma}(\theta; \mathbf{x}) = \mathcal{O} \left( \frac{1}{\sigma^p} \right)^{N-r(\theta)}, \quad (\text{D.4})$$

where  $r(\theta)$  is the number of times the value  $\theta$  is repeated in the sample set and  $\mathcal{O}$  denotes the asymptotic order as  $\sigma \rightarrow 0$ . In the limit the exponent  $N - r(\theta)$  must be minimized for  $H_{\sigma}(\theta; \mathbf{x})$  to be minimum. Therefore,  $\hat{\theta}$  will be one of the most repeated values in the input set. Define  $r = \max_j r(x(j))$ , then for  $x(j) \in \mathcal{M}$ , expanding the product in (D.2) gives

$$H_{\sigma}(x(j); \mathbf{x}) = \left\{ \prod_{i: x(i) \neq x(j)} \frac{|x(i) - \theta|^p}{\sigma^p} \right\} + \mathcal{O} \left( \frac{1}{\sigma^p} \right)^{N-r-1}. \quad (\text{D.5})$$

Since the first term in (D.5) is  $\mathcal{O}(1/\sigma^p)^{N-r}$ , the second term is negligible for small  $\sigma$ . Then, in the limit,  $\hat{\theta}$  can be computed as

$$\begin{aligned} \lim_{\sigma \rightarrow 0} \hat{\theta} &= \arg \min_{x(j) \in \mathcal{M}} [H_{\sigma}(x(j); \mathbf{x})] \\ &= \arg \min_{x(j) \in \mathcal{M}} \left[ \prod_{i: x(i) \neq x(j)} \frac{|x(i) - x(j)|^p}{\sigma^p} \right] \\ &= \arg \min_{x(j) \in \mathcal{M}} \left[ \prod_{i: x(i) \neq x(j)} |x(i) - x(j)| \right]. \end{aligned} \quad (\text{D.6})$$

## Acknowledgment

This paper was supported in part by NSF under Grant no. 0728904.

## References

- [1] E. E. Kuruoglu, "Signal processing with heavy-tailed distributions," *Signal Processing*, vol. 82, no. 12, pp. 1805–1806, 2002.
- [2] K. E. Barner and G. R. Arce, *Nonlinear Signal and Image Processing: Theory, Methods and Applications*, CRC Press, Boca Raton, Fla, USA, 2003.
- [3] G. R. Arce, *Nonlinear Signal Processing: A Statistical Approach*, John Wiley & Sons, New York, NY, USA, 2005.
- [4] P. Huber, *Robust Statistics*, John Wiley & Sons, New York, NY, USA, 1981.
- [5] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.
- [6] J. Astola and Y. Neuvo, "Optimal median type filters for exponential noise distributions," *Signal Processing*, vol. 17, no. 2, pp. 95–104, 1989.
- [7] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: a tutorial," *IEEE Transactions on Circuits and Systems II*, vol. 43, no. 3, pp. 157–192, 1996.
- [8] G. R. Arce, "A general weighted median filter structure admitting negative weights," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3195–3205, 1998.
- [9] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.
- [10] K. E. Barner and T. C. Aysal, "Polynomial weighted median filtering," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 636–650, 2006.
- [11] T. C. Aysal and K. E. Barner, "Hybrid polynomial filters for Gaussian and non-Gaussian noise environments," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4644–4661, 2006.
- [12] J. G. Gonzales, *Robust techniques for wireless communications in nongaussian environments*, Ph.D. dissertation, ECE Department, University of Delaware, 1997.
- [13] 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, 2007.
- [14] V. Zolotarev, *One-Dimensional Stable Distributions*, American Mathematical Society, Providence, RI, USA, 1986.
- [15] J. P. Nolan, *Stable Distributions: Models for Heavy Tailed Data*, Birkhuser, Boston, Mass, USA, 2005.

- [16] R. F. Bricich, D. R. Iskander, and A. M. Zoubir, "The stability test for symmetric alpha-stable distributions," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 977–986, 2005.
- [17] J. G. Gonzalez and G. R. Arce, "Optimality of the myriad filter in practical impulsive-noise environments," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 438–441, 2001.
- [18] J. G. Gonzalez and G. R. Arce, "Statistically-efficient filtering in impulsive environments: weighted myriad filters," *Eurasip Journal on Applied Signal Processing*, vol. 2002, no. 1, pp. 4–20, 2002.
- [19] T. C. Aysal and K. E. Barner, "Myriad-type polynomial filtering," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 747–753, 2007.
- [20] P. R. Rider, "Generalized cauchy distributions," *Annals of the Institute of Statistical Mathematics*, vol. 9, no. 1, pp. 215–223, 1957.
- [21] J. Miller and J. Thomas, "Detectors for discrete- time signals in non- Gaussian noise," *IEEE Transactions on Information Theory*, vol. 18, no. 2, pp. 241–250, 1972.
- [22] T. C. Aysal, *Filtering and estimation theory: first-order, polynomial and decentralized signal processing*, Ph.D. dissertation, ECE Department, University of Delaware, 2007.
- [23] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. 19, no. 3, pp. 106–127, 1977.
- [24] H. M. Hall, "A new model for impulsive phenomena: application to atmospheric-noise communication channels," Tech. Rep. 3412 and 7050-7, Stanford Electronics Laboratories, Stanford University, Stanford, Calif, USA, 1966.
- [25] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, NY, USA, 1986.
- [26] R. E. Carrillo, T. C. Aysal, and K. E. Barner, "Generalized Cauchy distribution based robust estimation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3389–3392, April 2008.
- [27] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B*, vol. 48, no. 3, pp. 259–302, 1986.
- [28] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, NY, USA, 1997.
- [29] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, Mass, USA, 1988.
- [30] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Transactions on Electromagnetic Compatibility*, vol. 44, no. 1, pp. 249–258, 2002.
- [31] Y. H. Ma, P. L. So, and E. Gunawan, "Performance analysis of OFDM systems for broadband power line communications under impulsive noise and multipath effects," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 674–682, 2005.
- [32] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1398–1410, 2008.
- [33] 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, Article ID 4524301, pp. 1466–1471, 2008.
- [34] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling: a sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [35] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [36] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.
- [37] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [38] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [39] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [40] E. J. Candès and P. A. Randall, "Highly robust error correction by convex programming," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2829–2840, 2008.
- [41] B. Popilka, S. Setzer, and G. Steidl, "Signal recovery from incomplete measurements in the presence of outliers," *Inverse Problems and Imaging*, vol. 1, no. 4, pp. 661–672, 2007.
- [42] 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 on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 392–408, 2010.
- [43] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, no. 11, pp. 657–664, 1991.
- [44] R. N. Dave and R. Krishnapuram, "Robust clustering methods: a unified view," *IEEE Transactions on Fuzzy Systems*, vol. 5, no. 2, pp. 270–293, 1997.
- [45] M.-S. Yang and K.-L. Wu, "A similarity-based robust clustering method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 434–448, 2004.
- [46] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 1984.