SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 6, NO. 1
ISSN: 2054 -7420

Outlier Resistant Time Series Operations via Qualitative Robustness and Saddle-Point Game Formalizations- A Review: Filtering and Smoothing

P. Papantoni-Kazakos and A.T. Burrell
University of Colorado Denver, Electrical Engineering Department, Denver, Colorado, USA
Oklahoma State University, Computer Science Department, Stillwater, Oklahoma
titsa.papantoni@ucdenver.edu; tburrell@okstate.edu

ABSTRACT

Time series operations are sought in numerous applications, while the observations used in such operations are generally contaminated by data outliers. The objective is thus to design outlier resistant or "robust" time series operations whose performance is characterized by stability in the presence versus the absence of data outliers. Such a design is guided by the theory of qualitative robustness and is completed by saddle-point game formalizations. The approach is used for the development of outlier resistant filtering and smoothing operations.

Keywords: Time Series Analysis; Qualitative Robustness; Data Outliers; Filtering; smoothing.

1 Introduction

The fundamental desirable characteristic of outlier resistant or "robust: time series operations is performance stability; that is, a robust statistical procedure should guarantee small performance deviations for small perturbations in the data generating stochastic process. Thus, statistical robustness may be qualitatively defined along the latter lines, where for an analytical definition, the use of appropriate stochastic distance measures is essential. This qualitative definition is developed by the theory of *qualitative robustness*, while it also intimately related to the *robust saddle-point game theoretic formalizations.* The theory of qualitative robustness provides necessary conditions to be satisfied by robust operations, while the robust saddle-point game theoretic formalizations provide specific solutions within the qualitatively robust class of operations. In this paper, we will review this composite construction of statistically robust operations. We will then present solutions for outlier resistant or robust filtering and smoothing.

The definition of qualitative robustness was first given by Hampel (1971, who considered only memoryless data processes. The definition was extended to include processes with memory, first by Papantoni-Kazakos and Gray (1979) and then by Cox (1978), Bustos et al (1984) and Papantoni-Kazakos (1984a, 1984b, 1987). Solutions for outlier resistant prediction, filtering and smoothing were first developed by Tsaknakis et al (1988, 1986), while an overview of the theory can be found in Kazakos et al (1990). Extensions of the theory of qualitative robustness to include robust block encoders and quantizers were

DOI: 10.14738/tnc.61.2480
Publication Date: 05th March 2018
URL: http://dx.doi.org/10.14738/tnc.61.2480

developed by Papantoni- Kazakos (1981a, 1981b). Finally, a stochastic neural network was developed by Kogiantis et al (1997) and Burrell et al (1997), for implementation of robust prediction, and has been applied by Burrell et al (2012) for predictive model mapping.

The organization of the paper is as follows: In Section 2, we present the outline of the qualitative robustness theory and its relationship to robust saddle-point game theoretic formalizations. In Section 3, we describe the process for developing robust filtering operations. In Section 4, we draw from the derivations in Section 3, to develop non-causal filtering or smoothing operations, when the nominal information and noise processes are both Gaussian. In Section 5, we focus on robust causal filtering solutions for nominally Gaussian information and noise processes. In Section 6, we include concluding remarks.

2 Qualitative Robustness And Robust Saddle-Point Game Theoretic Formalizations

As discussed in the introduction, qualitative robustness corresponds to small performance deviations for small perturbations in the data generating processes. Alternatively, qualitative robustness is a continuity property defined on the space of stochastic processes via appropriate stochastic measures. In particular, let x^n and y^n denote n-dimensional data sequences, generated respectively by two non-identical n-dimensional probability density functions f_0^n and f^n. Let g(·) denote some function or operation on n-dimensional data sequences, where g(·) could be, for example, a test function in hypothesis testing or a parameter estimate. Let h_{0g} and h_g denote respectively the density function of the random variables $g(X^n)$ and $g(Y^n)$ (where X^n is generated by f_0^n, and where Y^n is generated by f^n), and let $\mathrm{d}_1(f_0^n, f^n)$ and $\mathrm{d}_2(h_{0g}, h_g)$ be two stochastic distance measures respectively between the densities f_0^n and f^n, and the densities h_{0g} and h_g. Then we can present the following definition,

Definition 1: The operation $g(\cdot)$ is qualitatively robust at the density function f_0^n, in stochastic distance measures $\mathrm{d}_1(\cdot,\cdot)$ and $\mathrm{d}_2(\cdot,\cdot)$, iff :

Given $\varepsilon > 0$, there exists $\delta > 0$ such that if f^n is such that $\mathrm{d}_1(f_0^n, f^n) < \delta$, then h_g is such that $\mathrm{d}_2(h_{0g}, h_g) < \varepsilon$.

From the above definition, we conclude that qualitative robustness is a local (around f_0^n) stability property, parallel to the continuity property of real function. The specific analytical properties of a qualitatively robust data operation $\mathrm{g}(\cdot)$ depend on the choice of the stochastic distance measures and $\mathrm{d}_1(\cdot,\cdot)$ and $\mathrm{d}_2(\cdot,\cdot)$. The latter stochastic distances are initially selected to best reflect the desired stability properties of the qualitatively robust data operation, where the weaker the distance $\mathrm{d}_1(\cdot,\cdot)$ and the stronger distance $\mathrm{d}_2(\cdot,\cdot)$, then the stronger the qualitative robustness properties. The main issue arising here is the relationship of the qualitative robustness to the robust saddle-point formalizations, and the choice of the stochastic distance measures $\mathrm{d}_1(\cdot,\cdot)$. We will first address the relationship to the robust saddle-point game-theory formalizations.

Let us consider a saddle-point game with payoff function f(x,y), where the function $f(\cdot,\cdot)$ and its arguments x and y are all real and scalar, and where x and y take values respectively in the subsets A and B of the real line R. Consider the metric d (*u* , *v*) = | *u* – *v* | on the real line, and let the subsets A and B both be convex with respect to that metric. Let at least one of those two subsets also be compact with respect to the metric $d(\cdot,\cdot)$, and let the payoff function f (x , y) be convex in x, concave in y, and continuous in x and y, with respect to the same metric. Then, the existence of a saddle-point solution (x* ,y*) such that $f(x^*, y) \le f(x^*, y^*) \le f(x, y^*)\,;\ \forall\ x \in A$ and $\forall\ y \in \mathrm{B}$ is guaranteed and it is unique. If, on the hand, the function *f* (*x* , *y*) is not continuous in x and y, then the existence of a saddle-point solution is not generally guaranteed. The continuity of the payoff function is thus an essential property for the guaranteed existence of a saddle-point solution. The same is true when instead of x and y, we have density functions f^n and h_g as in Definition 1. In the latter case, the metric | *u* – *v* | on the real line is replaced by the stochastic distance measure $d_1(\cdot,\cdot)$ for the data generating densities f^n, and by the stochastic distance measure $d_2(\cdot,\cdot)$, for densities h_g induced by some f^n and some data operation *g*. Therefore, qualitative robustness is essential for the guaranteed solutions of the robust saddle-point game-theory formalization.

Let us now turn to the choice of the distances $d_1(\cdot,\cdot)$ and $d_2(\cdot,\cdot)$ in Definition 1. As we already pointed out, to make the qualitative robustness property strong, we need a weak distance $d_1(\cdot,\cdot)$ and a strong distance $d_2(\cdot,\cdot)$. A weak distance that also represents closeness in data sequences and best reflects the outlier model as well is the Prohorov distance [10], with data distortion measure $\rho_n(x^n, y^n)$ as follows.

$$\rho_n(x^n, y^n) = \begin{cases} n^{-1}\sum_{i=1}^{n}|x_i - y_i| = \gamma_n(x_1^n, y_1^n) & \text{if } n \text{ given and finite} \\ \\ \inf\{\alpha : n^{-1}[\#i : \gamma_m(x_{i+1}^{i+m}, y_{i+1}^{i+m}) > \alpha] \le \alpha\} & \\ \qquad \text{if } n > n_0, \text{where } m \text{ and } n \text{ are postive integers} & \end{cases} \tag{1}$$

The Prohorov distance with data distortion measure as in (1) is a metric; that is, it satisfies the triangular property. For classes of memoryless processes, the distance is identical to the Prohorov distance with dtata distortion measure $\rho_1(x\ ,y\ ) = |x - y|$. Regarding the choice of the distance $d_2(\cdot,\cdot)$, the Vasershtein or Rho-Bar distances [10] are appropriate. Indeed, those two distances are strong and they both bound difference in expected error performance. The choice of the data distortion measure within the latter distances depends on the particular application, where a popular and useful such choice is the difference squared distortion measure $\rho^*(x\ ,y\ ) = (x - y)^2$. The Rho-Bar distance is used for closeness in stochastic processes. Given some data sequences $y_1^{N+n} = \{y_1, \ldots, y_{N+n}\}$ and some scalar operation $g(\cdot)$, let $g(y_i^{i+n})$ estimate the datum x_k of some process whose arbitrary dimensionality density function is f_2 and whose data sequence are …, x_{-1}, x_0, x_1 …. If the sequence y_1^{N+n} is generated by the density function f_0^{N+n}, let h_{0g} denote the arbitrary dimensionality density induced by f_0^{N+n} and the data operation g (·). Let h_g denote the arbitrary dimensionality density induced by g (·) and some other data density function f^{N+n}. Then, h_{0a} and h_a both estimate f_2. Given some data distortion measure $\rho(\cdot,\cdot)$, the goodness of those two estimates is respectively measured by the Rho-Bar distances $\overline{\rho}(f_2, h_{og})$ and $\overline{\rho}(f_2, h_g)$. If $\rho(u,v) = |u-v|$, then $\left|\overline{\rho}(f_2, h_{0g}) - \overline{\rho}(f_2, h_g)\right| \le \overline{\rho}(h_{0g}, h_g)$; thus, the Rho-Bar distance $\overline{\rho}(h_{0g}, h_g)$ measures how closely h_{0g} fits f_2, as compared to the fitness of h_g to f_2. A similar conclusion is drawn, when the data distortion measure is the difference squared, $\rho^*(u,v) = (u-v)^2$ where then $\left|[\overline{\rho}^*(f_2, h_{0g})]^{1/2} - [\overline{\rho}^*(f_2, h_g)]^{1/2}\right| \le [\overline{\rho}^*(h_{0g}, h_g)]^{1/2}$.

The definition of qualitative robustness, in conjunction with the Prohorov and Rho-Bar or Vasershtein distances lead to constructive sufficient conditions that data operations should satisfy [2], [6], [7] and [10]. These conditions are included in Theorem 1 below, whose proof can be found in [2].

Theorem 1 : Consider a scalar real operation $g(x^n)$ on data sequences x^n of length n. Let $g(x^n)$ be bounded, and such that :

i. If n is finite, then $g(x^n)$ is pointwise continuous as a function of the data. That is, given ε > 0, there exists δ > 0, such that $n^{-1}\sum_i |x_i - y_i| < \delta$ implies $\left|g(x^n) - g(y^n)\right| < \varepsilon$.

ii. If n is asymptotically large, and given some data generating density function f_0, then $g(x^n)$ is pointwise asymptotically continuous at f_0. That is, given ε > 0 and η > 0, there exist δ > 0, positive integers m and n_0, and for each $n > n_0$ some set A^n ε R^n, such that $\Pr(x^n \in A^n \mid f_0^n) > 1-\eta$ and $x^n \in A^n$ and $\inf\{\alpha : n^{-1}[\#i : \gamma_m(x_{i+1}^{i+m}, y_{i+1}^{i+m}) > \alpha] \le \alpha\} < \delta$ implies $|g(x^n) - g(y^n)| < \varepsilon \ \forall\, n > n_0$, where $\gamma_m(x_{i+1}^{i+m}, y_{i+1}^{i+m}) = m^{-1}\sum_{j=i+1}^{i+m} |x_j - y_j|$. Then the operation $g(\cdot)$ is qualitatively robust at the density function f_0^n, where in Definition 12.1.1, $d_1(\cdot,\cdot)$ is replaced by the Prohorov distance with data distortion measure as in (1) and $d_2(\cdot,\cdot)$ is replaced by either the Vasershtein or the Rho-Bar distances with distortion measure $\rho(u,v)$ equal either to |u – v| or some continuous function of | u – v|.

From Theorem 1, we conclude that to be qualitatively robust, it suffices that a data operation be bounded and continuous. For data sequences of finite length continuity is defined in the usual functional sense. For asymptotically large data sequences, continuity is defined as follows at some data generating density function: If some sequence x^n is representative of the latter density function, in the sense that it belongs to a high-probability set A^n, and if the majority of the elements of another sequence y^n are close to the corresponding elements of the sequence x^n, then the values $g(x^n)$ and $g(y^n)$ of the data operating are close as well. Due to the above results, we conclude that linear operations are not qualitatively robust. This is so because such operations are not bounded, and because closeness between the majority of corresponding elements of two sequences does not guarantee closeness in the values of those operations.

Qualitative robustness is a property that does not induce uniqueness. That is, given a specific problem, and some data generating density function f_0, there generally exists a whole class ϑ of data operations that are qualitatively robust at f_0 . Additional performance criteria are thus needed, to evaluate and compare different data operations in class ϑ. Such performance criteria are the break-down point and the sensitivity, both defined asymptotically ($n \to \infty$) and at the density function f_0. Given f_0 and given some operation $g(\cdot)$ in class ϑ , consider the density functions f that are included in the Prohorov ball $\prod_{n,\rho_n}(f_0, f) \leq \varepsilon$, where ρ_n is as in (1). Let h_{0g} and h_g be the density functions induced by the data operation $g(\cdot)$ and the densities f_0 and f respectively. Given some scalar data distortion measure $\rho(\cdot,\cdot)$, consider the Rho-Bar distance $\overline{\rho}(h_{0g}, h_g)$. Then, the *breakdown point* ε^* , of the operation $g(\cdot)$ at f_0 is the largest value ε, such that, if f is some density in the ball $\lim_{n\to\infty} \prod_{n,\rho_n}(f_0, f) \leq \varepsilon$, then the distance $\overline{\rho}(h_{0g}, h_g)$ is a function of ε. The *sensitivity* of the operation $g(\cdot)$ at the density f_0 is defined as

$$\lim_{\substack{n\to\infty \\ \varepsilon\to 0}} \frac{\overline{\rho}(h_{0g}, h_g)}{\prod_{n,\rho_n}(f_0, f)}$$

It can be found that if bounded sensitivity at f_0 is required (parallel to bounded derivative) then the qualitatively robust operation $g(\cdot)$ should also be differentiable almost everywhere as a real function of the data, and for asymptotically large sequences it should be such that

$$| g(x^n) - g(y^n) | \leq \mathrm{c} \inf\{\alpha : n^{-1}[\# i : \gamma_m(x_{i+1}^{i+m}, y_{i+1}^{i+m}) > \alpha] \leq \alpha\}$$

where c is some bounded constant, and where $x^n \in A^n$ for A^n as in part ii of Theorem 1 [see Papantoni-Kazakos (1984b)].

As may be deduced from the presentation in this section, qualitative robustness is a performance stability property and its time series applications include prediction, interpolation and filtering or smoothing. Solutions for the later time series operations require the marriage of qualitative robustness with the theory of saddle-point game theoretic formalizations. In this paper, we present such solutions for non-causal filtering or smoothing as well as for causal filtering.

3 Robust Filtering

The objective of either non-causal or causal filtering is the extraction of information carrying data from noisy observations. That is, the outcomes generated by an information process are estimated, when distorted by interferences from a noise process. We will assume that the relationship between the information and noise processes is additive. In the robust filtering problem, the information and noise processes are modeled by two disjoint classes, $\mathcal{F}_S$ and $\mathcal{F}_N$, respectively. Arbitrary dimensionality probability density functions in classes $\mathcal{F}_S$ and $\mathcal{F}_N$ are respectively denoted f_s and f_N.

Let f_{0S} and f_{0N} be two nominal well known, stationary density functions, such that $f_{0s} \in \mathcal{F}_S$ and $f_{0N} \in \mathcal{F}_N$. Let us assume that some density function f_s from class $\mathcal{F}_S$ is a priori selected by the system designer to represent the information process throughout the over all observation interval, and let us denote by …, X_{-1} , X_0, X_1, … a random data sequence generated by f_s . We initially assume that the class $\mathcal{F}_S$, consists of f_{0s} only.

Let us denote by …, W_{-1} , W_0, W_1, … random noise data sequences, and let …, Z_{-1} , Z_0, Z_1, … be data sequences from the nominal noise density function f_{0N} . Given some number ε_N in (0,1), let the class $\mathcal{F}_N$ of noise processes then be such that

$$W_n = (1-\varepsilon_N) Z_n + \varepsilon_N V_n \tag{2}$$

where …, V_{-1} , V_0, V_1, … is a random sequence generated by any arbitrary dimensionality stationary density function. The noise model in (2) represents the occurrence of outliers, with probability ε_N per datum. Given f_s in $\mathcal{F}_S$ and f_N in $\mathcal{F}_N$, we assume that the data sequences from f_s and f_N. are additive and that f_s and f_N. are mutually independent. Then, if …, Y_{-1} , Y_0, Y_1, … denote random observation sequences, we have,

$$Y_n = X_n + W_n \quad \forall n \tag{3}$$

where X_n is generated by f_s , W_n is generated by f_N [as in (2)], and the sequences …, X_{-1} , X_0, X_1, … and …, W_{-1} , W_0, W_1, … are mutually independent. Let $g_{n+l,F}(y_{-n}^{l-1})$ denote a filtering operation, estimating the information datum X_0 , via the observation sequence y_{-n}^{l-1} . Let $e_F(g_{n+l,F,}\, f_s, f_N)$ denote the mean-squared error induced by the operation $g_{n+l,F}(y_{-n}^{l-1})$+ at the density functions $f_s \in \mathcal{F}_S$ and $f_N \in \mathcal{F}_N$. That is,

$$e_F(g_{n+l,F,}\, f_s, f_N) = E\left\{\left[X_0 - g_{n+l,F}(Y_{-n}^{-1})\right]^2 \mid f_s, f_N\right\} \tag{4}$$

Consider then the following saddle-point game. Search for the triple $(g_{n+l,F}^*, f_s^*, f_N^*)$ such that $f_s^* \in \mathcal{F}_S$ and $f_N^* \in \mathcal{F}_N$ and

$$\forall f_s \in \mathcal{F}_S\ ,\ f_N \in \mathcal{F}_N,\ \ e_F(g_{n+l,F}^*, f_s, f_N) \le e_F(g_{n+l,F}^*, f_s^*, f_N^*) \le e_F(g_{n+l,F}, f_s^*, f_N^*) \quad \forall\, g_{n+1,F} \tag{5}$$

The right part of (5) is satisfied for $g_{n+1,F}^*(y_{-n}^{l-1})$ being the conditional expectation of X_0 at f_s^* and f_N^* . That is

$$g_{n+1,F}^*(y_{-n}^{l-1}) = E\left\{X_0 \mid y_{-n}^{l-1}, f_s^*, f_N^*\right\} \tag{6}$$

The game in (5) reduces then to the following search. Find the pair (f_s^* , f_N^*) such that $f_s^* \in \mathcal{F}_S$ and $f_N^* \in \mathcal{F}_N$, and

$$E\left\{\left[X_0 - E\left\{X_0 \mid y_{-n}^{l-1}, f_s^*, f_N^*\right\}\right]^2 \mid f_s^*, f_N^*\right\} =$$

$$= \sup_{\substack{f_s \in F_s \\ f_N \in F_N}} E\left\{\left[X_0 - E\left\{X_0 \mid y_{-n}^{l-1}, f_s, f_N\right\}\right]^2 \mid f_s, f_N\right\} \tag{7}$$

and select $g_{n+l,F}^{*}(y_{-n}^{l-1})$ as in (6).

Given $f_s \in \mathcal{F}_s$ and $f_N \in \mathcal{F}_N$, and due to their additivity and mutual independence, the induced observation density f equals the convolution $f_s * f_N$, between the densities f_s and f_N. If μ_s and σ_s^2 denote respectively the mean and variance of the density f_s and defining then

$$\alpha(Y_{-n}^{l-1}) = \int_{R^{n-1}} dx_{-n}^{l-1}\, x_0\, f_s(y_{-n}^{l-1}) f_N(y_{-n}^{l-1}, x_{-n}^{l-1}) \tag{8}$$

we easily find, for $f = f_s * f_N$

$$E\left\{\left[X_0 - E\left\{X_0 \mid y_{-n}^{l-1}, f_s, f_N\right\}\right]^2 \mid f_s, f_N\right\} =$$

$$= \sigma_s^2 - \int_{R^{n+1}} dy_{-n}^{l-1} \frac{\left[\alpha(y_{-n}^{l-1}) - \mu_s f(y_{-n}^{l-1})\right]^2}{f(y_{-n}^{l-1})} \tag{9}$$

Let $\Phi_s(D_{-n},\dots,D_{l-1},)$ and $A(D_{-n},\dots,D_{l-1},)$ denote the characteristic functions (or Fourier transforms) at $\{D_i; -n \le i \le l-i\}$ of respectively the densities $f_s(y_{-n}^{l-1})$, $f_N(y_{-n}^{l-1})$, $f(y_{-n}^{l-1})$ and the function $\alpha(y_{-n}^{l-1})$ in (8), assuming that the former exist. Let us define the operator :

$$P(D_{-n},\dots,D_{l-1}) = \frac{\frac{\partial}{\partial D_0}\Phi_s(D_{-n},\dots,D_{l-1})}{\Phi_s(D_{-n},\dots,D_{l-1})} \tag{10}$$

Then, the supremum in (7) reduces to the search of the infimum below, where $\mathcal{F}$ denotes the class induced by f_{0s} and f_N ; that is, $\mathcal{F} = \{f = f_{0s} * f_N,\ f_N \in \mathcal{F}_N\}$.

$$\inf_{f \in F} \int_{R^{n+1}} dy_{-n}^{l-1} \frac{\left\{P(D_{-n},\dots,D_{l-1})\left[f(y_{-n}^{l-1})\right]\right\}^2}{f(y_{-n}^{l-1})} \tag{11}$$

We consider the class F_N of noise processes, as described by the probability density functions these processes induce and we select this class to be given by expression (12) below.

$$\mathcal{F}_N = \begin{array}{l}\{ f : f = (1-\varepsilon_N) f_{0s} * f_{0N} + \varepsilon_N h \\ h \text{ is any arbitary dimensionality density function} \}\end{array} \tag{12}$$

We then express Theorem 2 below. This theorem and the subsequent Lemma 1 are due to Tsaknakis et. al. (1986).

Theorem 2 : Let the density f_{0s} have a nonzero and analytic characteristic function $\Phi_s(D_{-n},...,D_{l-1},)$ = $\Phi_s(\underline{D})$, that also admits a Taylor series expansion everywhere. Consider then the operator $P(\underline{D})$ = $P(D_{-n},...,D_{l-1},)$ in (10) which also admits then a Taylor series expansion. Consider the class $\mathscr{F}_N$ in (12), and denote

$$f_0 = f_{0s} * f_{0N} \tag{13}$$

Let $d(y_{-n}^{-1})$ be a positive solution of the equation

$$|P(\underline{D})d(y_{-n}^{l-1})| = \lambda\, d(y_{-n}^{l-1}) \quad \lambda > 0 \tag{14}$$

such that $d(y_{-n}^{l-1})$ is integrable over R^{n+1}, it is analytic for all nonzero vectors y_{-n}^{l-1} , and the quantity $P(\underline{D})[d*(y_{-n}^{l-1})]$ exists for all y_{-n}^{l-1} in R^{n+1}, where

$$d*(y_{-n}^{l-1}) = \begin{cases} (1-\varepsilon_N) f_0(y_{-n}^{l-1}) & \text{for } y_{-n}^{l-1} \in A^{n+1} \\ \lambda\, d(y_{-n}^{l-1}) & \text{otherwise} \end{cases} \tag{15}$$

where, A^{n+1} includes all y_{-n}^{l-1} , such that $|P(\underline{D})[f_0(y_{-n}^{l-1})]/ f_0(y_{-n}^{l-1})| \quad \le \lambda$.

Then, the infimum in (11) with substitution of $\mathscr{F}_N$ for $\mathscr{F}$, exists and is attained by the following density f^*

$$f*(y_{-n}^{l-1}) = d^*(y_{-n}^{l-1}) \tag{16}$$

with λ such that

$$\int_{R^{n+1}} f^*(y_{-n}^{l-1})\, d\, y_{-n}^{l-1} = 1\,.$$

Furthermore, the filtering operation $g_{n+1,F}^*(Y_{-n}^{l-1})$ = $E\{X_0 \mid y_{-n}^{l-1}, f^*\}$ that satisfies (5) then the game in (5) on $\mathscr{F}_N$ is

$$g_{n+1,F}^*(y_{-n}^{l-1}) = \begin{cases} \dfrac{P(\underline{D})]f_0(y_{-n}^{l-1})]}{f_0(y_{-n}^{l-1})} & \text{for } y_{-n}^{l-1} \in A^{n+1} \\ \pm\lambda & \text{for } y_{-n}^{l-1} \in [R^{n+1} - A^{n+1}] \end{cases} \tag{17}$$

Lemma 1 below is a consequence of Theorem 2.

Lemma 1 : Let the densities f_{0s} and f_{0N} in Theorem 2 be both zero mean Gaussian, with respective auto-covariance matrices M_{n+1} and N_{n+1} , where the elements of M_{n+1} are denoted $\{m_{i,j}\}$.Then, the density f_0 in (13) is zero mean Gaussian, with auto-covariance matrix A_{n+1}= M_{n+1} + N_{n+1} and the density f^* in (16) and the filtering operator g^* in (17) take then the following special form, where | A_{n+1} | means determinant, T means transpose and (-1) denotes inverse, where it is assumed that Λ_{n+1} is nonsingular, and where $a_{n+1}^T = [m_{0,l-1},...,m_{0,-n}]$, $\operatorname{sgn} x = \{1\,;\, x \ge 0 \text{ and } -1;\, x < 0\}$.

$$f^*(y_{-n}^{l-1}) = \begin{cases} (1-\varepsilon_N)(2\pi)^{-(n-l)/2}\left|\Lambda_{n+1}\right|^{1/2} \exp\{-2^{-1}(y_{-n}^{l-1})^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}\} & \text{for } y_{-n}^{l-1}: \left|a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1} \le \lambda\right| \\ (1-\varepsilon_N)(2\pi)^{-(n-l)/2}\left|\Lambda_{n+1}\right|^{1/2} \exp\{-2^{-1}(y_{-n}^{l-1})^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}\} + \dfrac{[\lambda - \left|a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}\right|]^2}{2a_{n+1}^T \Lambda_{n+1}^{-1} a_{n+1}} & y_{-n}^{l-1}: \left|a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}\right| > \lambda \end{cases} \tag{18}$$

$$g_{n+1,F}^*(y_{-n}^{l-1}) = \begin{cases} a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1} & \text{for } y_{-n}^{l-1}: \left|a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}\right| \le \lambda \\ \lambda \operatorname{sgn}(a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}) & \text{for } y_{-n}^{l-1}: \left|a_{n+1}^T \Lambda_{n+1}^{-1} y_{-n}^{l-1}\right| > \lambda \end{cases} \tag{19}$$

where denoting c = $\lambda\,[a_{n+1}^T \Lambda_{n+1}^{-1} a_{n+1}]^{1/2}$, and for $\phi(x)$ and $\Phi(x)$ denoting respectively the density at *x* and the cumulative distribution at *x* of the zero mean, unit variance Gaussian random variable, the constant λ is such that,

$$\Phi(c) + c^{-1}\phi(c) = 2^{-1}[1 + (1-\varepsilon_N)^{-1}] \tag{20}$$

Given ε_N , *n* and *l*, the constant λ is positive and unique. Given *n* and *l*, λ decreases monotonically with increasing ε_N. For $\varepsilon_N = 0$, λ equals infinity, and the filtering operation in (19) becomes then identical to the optimal at the Gaussian noise, linear, mean-squared filter.

Denoting, $I(f\ ) = \int_{R^{n+1}} d\, y_{-n}^{l-1}\ f^{-1}(y_{-n}^{l-1})\{P(\underline{D})[f(y_{-n}^{l-1})]\}^2$, for the operator, $P(\underline{D})$, in (10), we also find $I(f^*)$ for density f^* in (18), where *c* is as in (20).

$$I(f^*) = 2(1-\varepsilon_N)\, a_{n+1}^T \Lambda_{n+1}^{-1} a_{n+1} [\Phi(c) - 2^{-1}] \tag{21}$$

We observe that the filtering operation is (19) is a truncated linear function of the data; it is thus bounded and continuous in the sense of part i in Theorem 1, but it is not asymptotically continuous in the sense of part ii in the same theorem. The latter operation is therefore qualitatively robust for finite data dimensionalities *n+l* only. We will extend the operation in (19), to create a filtering operation that is both asymptotically and non-asymptotically robust. We distinguish between casual and non-casual filtering, and we present then two different extensions.

4 Robust Non-Causal Filtering or Smoothing for Nominally Gaussian Information and Noise Processes

Consider the Gaussian densities f_{0s} and f_{0N} in Lemma 1. We then select some ε_N and some finite non-negative integer *m*. Let { ..., X_{-1} , X_0, X_1, ... } and { ..., W_{-1} , W_0, W_1, ... } denote sequences of random variables that are respectively generated by f_{0s} and f_{0N} . Given some integer *k* and some non-negative integer *n*, let

$N_{2n+1,k}$ and $M_{2n+1,k}$ denote respectively the auto-covariance matrices $E\{W_{k-n}^{k+n}(W_{k-n}^{k+n})^T \mid f_{0N}\}$ and $E\{X_{k-n}^{k+n}(X_{k-n}^{k+n})^T \mid f_{0s}\}$. Let $a_{2n+1,k}^T$ denote the (n+1) th row of the matrix $M_{2n+1,k}$, let $\Lambda_{2n+1,k} = M_{2n+1,k} + N_{2n+1,k}$, and let $g_{kl}^0(x_{k-n}^{k-l}, x_{k+l}^{k+n});\, n \geq l$, denote the optimal mean-squared interpolation operation at the Gaussian density f_{0s} for the datum x_k, given x_{k-n}^{k-l} and x_{k-l}^{k+n}. Let us then define the sets $\{d_{k,n,l,j};\, k-n \leq j \leq k-l,\, k+l \leq j \leq k+n\}$ and $\{b_{k,n,j};\, k-n \leq j \leq k+n\}$ of coefficients as follows, where $\Lambda_{2n+1,k}$ is assumed non-singular.

$$\{d_{k,n,l,j}\}: g_{kl}^0(x_{k-n}^{k-l}, x_{k+l}^{k+n}) = \sum_{j=k-n}^{k-l} d_{k,n,l,j}\, x_j + \sum_{j=k+l}^{k+n} d_{k,n,l,j}\, x_j \tag{22}$$

$$[b_{k,n,k-n}, \ldots, b_{k,n,k+n}] = a_{2n+1,k}^T\, \Lambda_{2n+1,k}^{-1}$$

Let us now define

$$g_n^s(x) = \begin{cases} x & \text{if } |\mathrm{x}| \leq \lambda_n \\ \lambda_n\, \mathrm{sgn}(x) & \text{otherwise} \end{cases} \tag{23}$$

where $c = \lambda [a_{2n+1,k}^T\, \Lambda_{2n+1,k}^{-1}\, a_{2n+1,k}]^{-1/2}$ is such that

$$\Phi(c) + c^{-1}\phi(c) = 2^{-1}[1 + (1-\varepsilon_N)^{-1}] \tag{24}$$

Let $\hat{x}_{k,n}^{s}$ denote the estimate of the signal datum x_k from the observation vector y_{k-n}^{k+n}.

Then the estimate $\hat{x}_{k,n}^{s}$ is designed as in (25) below, where it can be shown that it is qualitatively robust both non-asymptotically and asymptotically.

$$\hat{x}_{k,n}^{s} = \begin{cases} g_n^s(a_{2n+1,k}^T\, \Lambda_{2n+1,k}^{-1}\, y_{k-n}^{k+n}) & \text{if } n \leq \mathrm{m} \\ g_{kl}^0(\hat{x}_{k-n}^{k-1}, \hat{x}_{k+1}^{k+n}) & n > m \end{cases} \tag{25}$$

where $\hat{x}_j^i = [\hat{x}_{j,m}^s, \ldots, \hat{x}_{i,m}^s];\, i > j$ and $g_{kI}^0(\cdot)$ is as in (22).

Let us define

$$r_n^s(n) = a_{2n+1,k}^T\, \Lambda_{2n+1,k}^{-1}\, a_{2n+1,k} \tag{26}$$

Then, $r_k^s(n)$ represents a variance gain in estimating the signal datum x_k from the observation vector y_{k-n}^{k+n} at the zero mean Gaussian noise density whose auto-covariance matrix is as in (22). Therefore, $r_k^s(n)$ is monotonically non-decreasing with increasing n. Given ε_N, the same monotonicity characterizes the truncation constant λ_n in (23), whose maximum value λ_∞ equals $c \lim_{n\to\infty}[r_k^s(n)]^{1/2}$, where c is the solution of (24). If the densities f_{0s} and f_{0N} are both stationary, with respective power spectral densities, $p_{0s}(\lambda)$ and $p_{0N}(\lambda)$; $\lambda \in [-\pi, \pi]$ and if $m \to \infty$, then directly from (19) we obtain

$$\lambda_\infty = c[E\{X_0^2 \mid f_{0s}\} - e_F(p_{0s}, p_{0N}]^{1/2}$$

$$= c\{(2\pi)^{-1}\int_{-\pi}^{\pi} p_{0s}(\lambda)\,d\lambda - (2\pi)^{-1}\int_{-\pi}^{\pi} p_{0s}(\lambda)[p_{0s}(\lambda)+p_{0N}(\lambda)]^{-1} p_{0N}(\lambda)d\lambda\}^{1/2}$$

$$= c\{(2\pi)^{-1}\int_{-\pi}^{\pi} p_{0s}^2(\lambda)[p_{0s}(\lambda)+p_{0N}(\lambda)]^{-1}d\lambda\}^{1/2}$$

5 Robust Causal Filtering for Nominally Gaussian Information and Noise Processes

Given the Gaussian densities f_{0s} and f_{0N} in Lemma 1 and the sequences { …, X_{-1} , X_0, X_1, … } and { …, W_{-1} , W_0, W_1, … }of random variables as in the non-causal filtering, let $M_{n,k}$ and $N_{n,k}$ denote respectively the auto-covariance matrices $E\{X_{k-n+1}^k (X_{k-n+1}^k)^T | f_{0s}\}$ and $E\{W_{k-n+1}^k (W_{k-n+1}^k)^T \mid f_{0N}\}$, where n ≥ 0 . Let then $a_{n,k}^T$ denote the first row of the matrix $M_{n,k}$, and let $\Lambda_{n,k} = M_{n.k} + N_{n,k}$. Let $g_{kl}^0(x_{k-n+1}^{k-l}), n-1 \geq l$, denote the optimal mean-squared prediction operation at the Gaussian density f_{0s} for the datum x_k , given x_{k-n+1}^{k-l} . Assuming that $\Lambda_{n,k}$ is nonsingular, let us then define the sets $\{c_{k,n-1,l,j}\,; k-n+1 \leq j \leq k-l\}$ and $\{h_{k,n,j}\,; k-n+1 \leq j \leq k\}$ of coefficients as

$$\{c_{k,n-1,l,j}\} : g_{kl}^0(x_{k-n+1}^{k-l}) = \sum_{j=k-n+1}^{k-l} c_{k,n-1,l,j}\, x_j \tag{27}$$

$$[h_{k,n,k-n+1}, \ldots, h_{k,n,k}] = a_{n,k}^T\, \Lambda_{n,k}^{-1}$$

Let us now define

$$g_n^c(x) = \begin{cases} x & \text{if } |x| \leq \mu_n \\ \mu_n\, \mathrm{sgn}(x) & \text{otherwise} \end{cases} \tag{28}$$

where $c = \mu\,[a_{n,k}^T\, \Lambda_{n,k}^{-1}\, a_{n,k}]^{-1/2}$ is such that

$$\Phi(c) + c^{-1}\phi(c) = 2^{-1}[1 + (1-\varepsilon_N)^{-1}] \tag{29}$$

Let $\hat{x}_k^c(n)$ denote the estimate of the signal datum x_k from the observation vector y_{k-n+1}^k . Then, the estimate $\hat{x}_k^c(n)$ is designed as follows, where ε_N and *m* are a priori selected.

$$\hat{x}^{c}_{k,n} = \begin{cases} g^{c}_{n}(a^{T}_{n,k}\,\Lambda^{-1}_{n,k}\,y^{k}_{k-n+1}) & \text{if } n \le \text{m} \\ \sum_{j=k-n+1}^{k-m} c_{k,n-1,m,j}\,\hat{x}^{c}_{j,\,j+n-k} & \\ + g^{c}_{m}\left(\sum_{\text{j=k-m+1}}^{\text{k}} h_{k,n,j}[y_j - g^{0}_{jp}(\hat{x}^{\,j-m}_{k-n+1})]\right) \;; & \text{if } n > m \end{cases} \tag{30}$$

Where $g^{0}_{jp}(\cdot)$ is as in (27) , and where $\hat{x}^{\,i}_{j} = [\hat{x}^{\,c}_{j,j+n-k},\ldots,\hat{x}^{\,c}_{i,i+n-k}]$.

Let us define,

$$r^{c}_{k}(n) = a^{T}_{n,k}\,\Lambda^{-1}_{n,k}\,a_{n,k} \tag{31}$$

Then, $r^{c}_{k}(n)$ represents the variance gain in estimating the datum x_k from the observation vector y^{k}_{k-n+1} at the zero mean Gaussian density, whose auto-covariance matrix is as in (27). Thus, $r^{\,c}_{\,k}(n)$ is monotonically non-decreasing with increasing n, and so is then the truncation constant μ_n in (27), where ε_N remains fixed.

It can be shown [Tsaknakis (1986)] that the operations in (30) are qualitatively robust, in both the asymptotic and the non-asymptotic sense. In the later operation, the integer *m* and ε_N represent a tradeoff between optimality at the Gaussian noise f_{0N} density robustness, and they are both system parameters. As *m* increases and ε_N decreases, the filtering operation in (30) tends to the optimal at the Gaussian density f_{0N}, linear data operation.

6 Conclusions

We have examined outlier resistant time series operations in the light of the theory of qualitative robustness. The resulting operations are continuous, both in a pointwise and an asymptotic sense, as well as bounded. Their performance is controlled by two parameters, one of which represents outlier contamination level. Special attention has been given to causal and non-causal filtering.

REFERRENCES

[1] Hampel, F.R. (1971). A General Qualitative Definition of Robustnes. *Ann. Math. Statist.* 42, 1887-1895.

[2] Papantoni-Kazakos, P. and R.M. Gray (1979). Robustness of Estimators on Stationary Observations. *Ann. Prob.* 7, 989-1002.

[3] Cox, D (1978). Metrics on Stochastic Processes and Qualitative Robustness. *Tech. Report No. 23.* Dept. of Staistics, Univ. of Wash., Seattle.

[4] Bustos, O., R. Fraiman and V. Yohai (1984). Asymptotic Behavior of the Estimates Based on Residual Autocovariances for ARMA Models. Robust and Nonlinear Time Series Analysis in *Lecture Notes in Statistics.* Springer-Verlag, New York, 26-49.

[5] Papantoni-Kazakos, P. (1984a). A Game Theoretic Approach to Robust Filtering. *Information and Control* 60, Nos. 1-3, 168-191.

[6] Papantoni-Kazakos, P. (1984b). Some Aspects of Qualitative Robustness in Time Series. Robust and Nonlinear Time Series Analysis in *Lecture Notes in Statistics.* Springer-Verlag, New York, 218-230.

[7] Papantoni-Kazakos, P. (1987). Qualitative Robustness in Time Series. *Information and Computation* 72, No. 3, 239-269.

[8] Tsaknakis, H. and P. Papantoni-Kazakos (1988). Outlier Resistant Filtering and Smoothing. *Information and Computation* 79, 163-192.

[9] Tsaknakis, H., D. Kazakos and P. Papantoni-Kazakos (1986). Robust Prediction and Interpolation for Vector Stationary Processes. *Prob. Th. And Related Fields* 72, Springer-Verlag, New York, 589-602.

[10] Kazakos, D. and P. Papantoni-Kazakos (1990). Detection and Estimation. *Computer Science Press,* New York.

[11] Papantoni-Kazakos, P. (1981a). Sliding Block Encoders that are Rho-Bar Continuous Functions of Their Input. *IEEE Trans. Inf. Theory,* IT-27, 372-376.

[12] Papantoni-Kazakos, P. (1981b). Stochastic Quantization for Performance Stability. *Information and Control* 49, No. 3, 171-198.

[13] Kogiantis, A. and P. Papantoni-Kazakos (1997). Operations and Learning in Neural Networks for Robust Prediction. *IEEE Trans. Systems, Man and Cybernetics* 27, No. 3, 402-411.

[14] Burrell, A.T., A. Kogiantis and P. Papantoni-Kazakos (1997). Detecting Changes in Acting Stochastic Models and Model Implementation via Stochastic Neural Networks. *Statistical Methods in Control and Signal Processing*. Editors: S. Sugimoto and T. Katayama. Marcel Dekker.

[15] Burrell, A.T. and P. Papantoni-Kazakos (2012). Stochastic Binary Neural Networks for Qualitatively Robust Predictive Model Mapping. *International Journal of Communications Network and System Sciences (IJCNS).* Special Issue on Models and Algorithms for Applications, September, Vol. 5, 603-608.