ROBUST DETECTION OF MONOTONIC TREND OF A DATA SEQUENCE IN THE PRESENCE OF IMPULSIVE NOISE

Nail Çadallı

Coordinated Science Laboratory and
Dept. of Electrical and Computer Engr.,
University of Illinois at Urbana-Champaign,
Urbana, IL, 61801, U. S. A.
cadalli@dsp.csl.uiuc.edu

Sandip Bose

Schlumberger-Doll Research Center
Old Quarry Road,
Ridgefield, CT, 06877,
U. S. A.
bose@ridgefield.sdr.slb.com

ABSTRACT

Robust detection of monotonic trend of a data sequence, when the data is subject to gross errors, is investigated. The method involves detection of the outliers by using the statistics of the available data and eliminating or estimating them for better line fitting where the slope of the fitted line indicates the trend of the sequence. Examples demonstrate the performance of the method via Monte Carlo simulations.

1. INTRODUCTION

In many practical environments, noise in the acquired data contains a small amplitude component representing random measurement noise and a large amplitude component representing large but infrequent (gross) measurement errors. This kind of disturbance in the data is commonly referred to as impulsive noise [1]. In statistical literature, the data points having the first component are called *inliers* and those having the second component are called *outliers*.

In this study, we investigate the problem of detecting monotonic trend of a data sequence which contains outliers. The data, for instance, could be the output of a diagnostic instrument which is required to periodically take small number of data samples and flag a trend and which is subject to occasional acquisition error. Gross errors may also be the result of preprocessing the raw data which has not been subject to large errors at the time of acquisition.

The problem is formulated from a hypothesis testing viewpoint using the likelihood ratio test between the constant trend and one of either decreasing or increasing trend, whichever is physically possible. Constant trend is indicated by the mean of the data while the slope of a line fitted to the data points in least squares sense gives us the trend of the data sequence.

Since least squares approximation is sensitive to outliers, this formulation is not robust and it can degrade catastrophically unless outliers are detected and either censored or replaced with corresponding inlier estimates. We present herein an algorithm for the robust solution of the above problem in the absence of any knowledge of the error statistics except that the small error component is modeled as a Gaussian distribution with unknown variance. Weak assumptions are made on the gross error statistics, namely that the probability of the gross error contamination is smaller than one third.

In the next section we present the problem formulation. Difficulties with the solution in case of outliers is addressed and robust statistics are proposed in Section 3. The robustness of the algorithm is demonstrated by examples in Section 4.1. Comparison of the Monte Carlo simulations with theoretical results obtained using a simplified model is presented in Section 4.2.

2. PROBLEM FORMULATION

Consider a set of acquired data values $\{y_i\}_{i=1}^{M}$ with corresponding coordinates $\{x_i\}_{i=1}^{M}$. Assume the trend sought is a decreasing trend. In that case, we can formulate our trend detection problem as the binary hypothesis testing problem of choosing between a constant trend and a decreasing linear trend. Even though a trend may not be a linear one, a decreasing trend would be better fitted by a negative slope line than by a constant and this is often a good first order approximation. We can therefore write our problem as follows: Let the two hypotheses be

$$\begin{aligned} \mathbf{H_0} &: \quad y_i = c + n_i \\ \mathbf{H_1} &: \quad y_i = ax_i + b + n_i, \quad a < 0, \quad \forall i \in \mathcal{I}\ , \end{aligned} \tag{1}$$

where $\mathcal{I}$ is the index set $\{1, 2, \ldots, M\}$. $\mathbf{H_0}$ is the hypothesis for no trend (i.e. constant trend, c) and $\mathbf{H_1}$ is the hypothesis for decreasing trend (i.e linear trend, $ax+b$) of the data sequence. Noise n_i is a sample of a random variable N which is a mixture of distributions whose probability density function can be written as:

$$f_N(t) = p_l f_{N_l}(t) + p_c f_{N_c}(t) + p_r f_{N_r}(t) \tag{2}$$

where N_l, N_c and N_r, stand for random variables with left, center and right distributions, f_{N_l}, f_{N_c} and f_{N_r} respectively. The left and right distributions, which have means $\mu_l \leq 0$ and $0 \leq \mu_r$ and variances σ_l^2 and σ_r^2, model the negative and positive gross errors due to outliers. The center distribution models the small error component due to inliers. It is assumed to be a zero-mean Gaussian distribution with variance σ_c^2. A sample is drawn from left or right distributions with prior probabilities p_l or p_r, sum of which, $(1 - p_c)$, is smaller than one third corresponding to the assumption that the number of outliers in the data set is not more than about one third of the total number of data points.

Let us define

$$\mathbf{y}_m = (y_{k_1}, y_{k_2}, \ldots, y_{k_m}) \tag{3}$$
$$\mathbf{x}_m = (x_{k_1}, x_{k_2}, \ldots, x_{k_m}) \tag{4}$$

where $\mathcal{I}_m = \{k_1, k_2, \ldots, k_m\}$ is a subset of the index set $\mathcal{I}$. Note that $m \leq M$ because, as we shall see in the next section, we may not use all of the available data points for line fitting but rather use the ones that are identified as inliers corresponding to $\mathcal{I}_m$. We can write the generalized likelihood ratio test (GLRT) statistics under the Gaussian noise assumption (for inliers only) as

$$L(\mathbf{x}_m, \mathbf{y}_m) = \frac{\sum_{i=1}^{m} (y_{k_i} - \bar{y})^2}{\sum_{i=1}^{m} (y_{k_i} - \hat{y}_{k_i})^2} \ . \tag{5}$$

The numerator of this ratio is the sum of the squared error between the data points $\{y_{k_i}\}_{i=1}^{m}$ and the mean of these points

$$\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_{k_i} \ . \tag{6}$$

The denominator is the sum of the squared error between the data points and the line fitted to these points which is given by

$$\hat{y}_{k_i} = \hat{a} x_{k_i} + \hat{b}, \quad \hat{a} \leq 0, \quad 1 \leq i \leq m \tag{7}$$

where $\hat{a}$ and $\hat{b}$ are obtained by least squares estimation. The condition on $\hat{a}$ ensures that only a negative slope is fitted. However, this condition is not imposed as a constraint in least squares estimation but rather taken care of by the detection rule. When the slope is of opposite sign (positive, if we assume the only possible trend other than a constant one is a decreasing trend), $\mathbf{H_0}$ is decided. The reason is that when the actual trend is constant, falsely fitting a positive slope line is more probable than fitting a positive slope line to an actual decreasing trend. The likelihood ratio is simply the measure of how well our approximation is fitted to a constant or a line with slope and offset given as in (7). Then the decision rule is

$$\delta(\mathbf{x}_m, \mathbf{y}_m) = \begin{cases} 0 & \text{if } L(\mathbf{x}_m, \mathbf{y}_m) \leq \nu \\ 1 & \text{if } L(\mathbf{x}_m, \mathbf{y}_m) > \nu \end{cases} \tag{8}$$

where threshold ν can be found given a false alarm rate [2]. Note that this has the desirable constant false alarm rate (CFAR) property.

3. ROBUST STATISTICS

In the above procedure, least-squares method is used for line fitting in (7). Least-squares method finds an approximate line that minimizes the l_2 norm, that is,

$$[\hat{a}\ \hat{b}] = \arg\min_{a,\ b} \ \sum_{i=1}^{m} (y_{k_i} - a x_{k_i} - b)^2 \ . \tag{9}$$

When the error components are independent, identically Gaussian distributed with zero mean, least square estimator is the same as maximum likelihood estimator [3]. However, if the data contains outliers, least square estimate is no longer preferable because large residues dominate the sum of the squares, resulting in unreliable estimates. In order to eliminate the effect of outliers prior to line fitting we introduce a novel *editing filter* for censoring the outliers in the data sequence and compare its performance with the more standard median filter which substitutes for the outliers.

3.1. Editing Filter

The editing filter, as its name suggests, edits the data sequence and eliminates the outliers from the sequence. To decide which data points are outliers, first a line is fitted to data points in least absolute value (l_1) sense which is given by

$$\tilde{y}_i = \tilde{a} x_i + \tilde{b}, \quad \tilde{a} \leq 0, \quad 1 \leq i \leq M \ , \tag{10}$$

where $\tilde{a}$ and $\tilde{b}$ are the minimizer of the total absolute error (l_1 norm), that is,

$$[\tilde{a}\ \tilde{b}] = \arg\min_{a,\ b} \ \sum_{i=1}^{M} |y_i - a x_i - b| \ . \tag{11}$$

Least absolute value method is not as sensitive to outliers as the least squares method is and it is known as a robust estimator in statistics [4, 5]. Intuitively, since the l_2 sum contains squares of the residues, the sum is dominated by the sum of square of the gross errors. This is not so severe in the case of l_1 where the absolute value is used in the summand. Another interpretation can be given in terms of influence function of the theory of M-estimators [6, 7]. Influence function of the l_2 approximation is such that influence of a datum on the estimate increases linearly with the size of its error. For l_1, influence function is the sign function, which is bounded, hence the effect of large errors is not as much.

Since there is an approximate line to the given data points which minimizes the l_1 norm and which passes through two points of the data set [5, Proposition 2.2], it is simple to find the optimum l_1 line. There also are computationally efficient algorithms [4, 5].

Notice that in (10) all M of the data points are used including the outliers, which have not been identified at this step. Let us define the residue of each data point due to the l_1 line as

$$e_i = |y_i - \tilde{y}_i|, \quad 1 \le i \le M \ , \tag{12}$$

and sort them in ascending order of magnitude resulting in

$$\{e_{k_1}, e_{k_2}, \ldots, e_{k_M}\}, \quad k_i \in \mathcal{I}, \quad 1 \le i \le M \ . \tag{13}$$

Since we have the assumption that the probability of having an outlier is less than about one third, we can consider two thirds of the data points with less error as the inliers. Consider the quantity

$$\sigma = \frac{1}{\sqrt{l}} \| [e_{k_1}, e_{k_2}, \ldots e_{k_l}] \|_2 \tag{14}$$

with $l \le M$. Then assigning

$$l = \lfloor \frac{2M}{3} \rfloor \tag{15}$$

yields that σ is the censored inlier standard deviation proportional to the true value σ_c. This now allows us to test for the remaining points for outliers. For Gaussian small error assumption, these could for instance be done at 4σ. Actually, data points $\{y_{k_i}\}_{i=l+1}^{M}$ may have greater deviations but this does not make them outliers automatically. This is because the actual trend may not be linear. It is, therefore, necessary to allow certain mismatch between the linear and actual trend and for this purpose in practice, we can choose a larger threshold for rejecting outliers. The robust points are then used in the likelihood ratio test described above. After this step, we have $l \le m \le M$ and this value of m is to be used in the likelihood ratio test (5).

3.2. Median Filter

Median filter is a sliding window whose l-th output component is the median of the windowed samples that is centered at coordinate l. For a median filter $g(\cdot)$ of length $2K+1$,

$$g(l) = \mathrm{median}(\{y_i\}_{i=l-K}^{l+K}) \quad 1 \le l \le M \ . \tag{16}$$

The beginning and the end of the data sequence may be thought as being padded with zeros to enable the filter operate at both ends of the sequence. Among the desirable properties of the median filter is that it preserves monotonic trends but eliminates impulses which makes it a natural choice for our problem [8, 9, 10]. During filtering, each outlier is replaced by one of the values inside the filtering window. It is not of high probability that outlier itself is the median value for that position of the window because it requires at least $K+1$ outliers present inside the window of length $2K+1$. Also assuming only a few outliers can be present in one median window, we can say that outliers are replaced by neighboring inlier values. This estimate is not the optimum one but it eliminates the effect due to the gross errors in line fitting. After filtering, likelihood ratio test (5) is used with $m = M$, since none of the data points is discarded.

4. PERFORMANCE

4.1. Examples

As our first example, we applied the scheme to data simulated with a trend as shown in Figure 1. Note that the trend

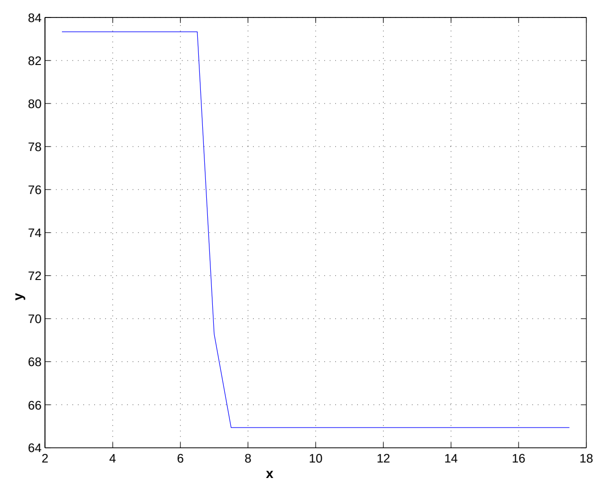

Figure 1: Data sequence $\{(x_i, y_i)\}_{i=1}^{M}$ without noise.

is monotonic but not linear. Noise was modeled as a mixture of small and large noise components where the former was a zero mean Gaussian distribution with standard deviation $\sigma_c = 2$ and the latter was modeled as two Gaussian distributions with means $\mu_r = 100$, $\mu_l = -100$ and standard deviations $\sigma_l = \sigma_r = 20$. The probability of contamination by gross error was 0.1, that is, $p_l = p_r = 0.05$. Length of the median filter was 11. We demonstrate the performance of the scheme by plotting the receiver operating characteristics (ROC) curves which show the probability of detection versus false alarm probability [2]. For this case, ROC curves are shown in Figure 2. These were generated via Monte Carlo simulations with 5000 realizations by passing noise only and noisy data separately through the detector and computing the detection rate for a given false alarm probability. As can be noticed from the curves, median filtering increases the performance of the detector and the editing filter performs far more better.

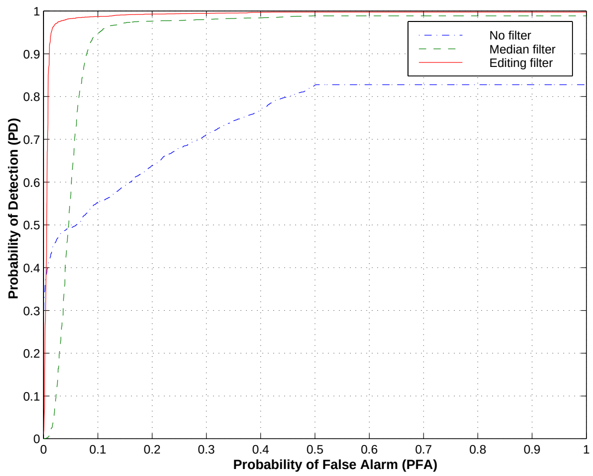

Figure 2: ROC curves with different filtering methods for the trend shown in Figure 1. Curves are for no filter (dash-dotted), median filter (dashed) and editing filter (solid). Parameters: $\sigma_c = 2$, $\sigma_r = \sigma_l = 20$, $\mu_r = -\mu_l = 100$, $p_l = p_r = 0.05$, length of the median filter is 11.

Notice that ROC curve corresponding to the case with no filtering does not reach to unity for any false alarm probability. This is because of the nonzero probability of having positive slope in line fitting when the actual trend is decreasing. This is more noticeable in the case with no filtering where we have the outliers present in the data. It is much less for the median filter curve and almost none for the editing filter case.

In our second example, we used a linear trend with slope $a = -0.5$ and offset $b = 100$ to obtain synthetic data. Here, we want to compare the performances of using median or editing filter on noisy data in l_2 line fitting, which is an important problem in many engineering areas [11]. Noise characteristics were the same as above except for we had $\sigma_c = 5$ and $p_l = p_r = 0.1$. For each of the 5000 noise realizations, a line was fitted, in l_2 sense, to noisy data and to data filtered using median and editing filters. Mean and variance of the resulting slope and offset estimates $\hat{a}$ and $\hat{b}$ are shown in Table 1. Using editing or median filters on the data before line fitting provided lower variances on the estimates than that of using no filtering.

	No filter	Median Fltr.	Editing Fltr.
mean of $\hat{a}$	-0.5187	-0.4634	-0.5028
variance of $\hat{a}$	0.7767	0.0412	0.0253
mean of $\hat{b}$	100.3397	99.4045	100.0268
variance of $\hat{b}$	276.3910	13.0301	9.4356

Table 1: Mean and variance of 5000 estimates of slope and offset of a line with $a = -0.5$ and $b = 100$.

4.2. Comparison with Theory

It is also possible to characterize the distribution of the test statistic under $\mathbf{H_0}$ with Gaussian errors. Recall that the test statistic L is the ratio of the sum of squared errors of the best fit of the filtered data $(\mathbf{x}_m, \mathbf{y}_m)$ to a constant versus that of the best fit to a decreasing linear trend. Then under $\mathbf{H_0}$ (when the real trend is constant), with probability $1/2$, the best fit linear trend is an increasing one and L takes the value 1. With the remaining probability $1/2$, if we make the assumption that the editing filter perfectly edits out the outliers and retains the rest, both the numerator and the denominator of the test statistic are χ^2 distributions scaled by the variance of the small error. In particular the numerator is the squared norm of the projection of $\mathbf{y}_m$ onto the orthogonal complement of the space spanned by $1_m = [1\,1\,\cdots\,1]$ and so has $m-1$ degrees of freedom. The denominator is the squared norm of the projection of $\mathbf{y}_m$ onto the orthogonal complement of the space spanned by 1_m and $\mathrm{li_m}$ where the latter corresponds to a linear trend. The numerator is therefore the sum of the denominator and an independent χ^2 variable with 1 degree of freedom. In this case, L is of the form $1 + \frac{\chi_1^2}{\chi_{m-2}^2}$ where the ratio is also known as the F-distribution (Fisher's variance-ratio distribution) [12]. Thus L has a distribution which is a $(1/2, 1/2)$ mixture of a point mass at 1 and a shifted F-distribution.

This does not depend on the variance of the error or on the probability of gross error (if the editing filter works perfectly, otherwise depends weakly) and therefore allows a pre-computation of the threshold for fixing the false alarm.

The distribution under $\mathbf{H_1}$ depends on a pair of parameters δ_1 and δ_2. δ_1 is the ratio of the norm of the projection of the actual trend onto a linear one (suitably accounting for the discarded outliers) to the error variance and plays the same role in performance as the SNR in the signal detection problem. δ_2 captures the misfit to the linear trend and has a more complex and weak influence on the performance.

We next display the comparison between the ROC curves generated from the Monte Carlo simulations and the behavior of the modified F statistic described above. For our problem with large errors, we assumed a simplified theoretical model in which the error distribution corresponded to the small error distribution and points were selected at random with the probability of the gross error and discarded. This simulates the action of a perfect editing filter which rejects all the data points with gross error and no other points. Figure 3 shows the above comparison of the ROC curve generated by this simplified theoretical model with the Monte Carlo results. The reasonably close match indicates the efficiency of the editing filter. Moreover it allows us to use the parameters defined above as indicators of detection performance.

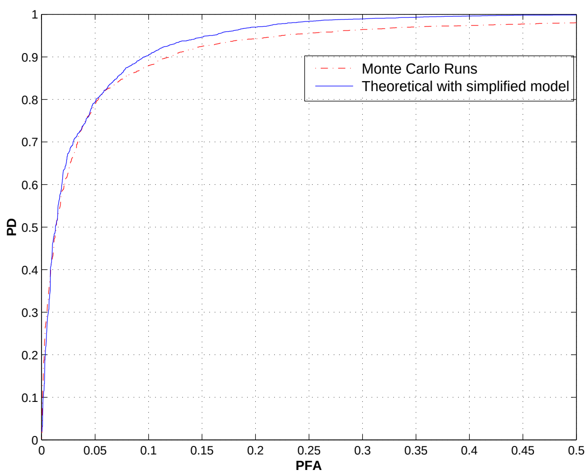

Figure 3: The ROC curves obtained using a simplified theoretical model and Monte Carlo simulations. The small error standard deviation is 2 and the gross error probability is 0.07.

5. CONCLUSION

A robust procedure for detection of monotonic trend of a data sequence, when the data is subject to gross errors, is described. The problem is formulated as one of hypothesis testing with the hypothesis being constant trend or decreasing trend. Robust statistics are obtained by line fitting to data sequence after detecting the outliers and either eliminating them using editing filter or replacing them with proper substitutes using median filter. Performance of the method is demonstrated by using receiver operating characteristic curves where we have shown that using median filter on the raw data enhances the detection probability. Editing filter performs far more better. Also, it has been demonstrated by Monte Carlo simulations that using editing filter improves the performance of line fitting in least squares sense to a linear trend in impulsive noise.

6. REFERENCES

[1] A. J. Efron and H. Jeen, "Detection in impulsive noise based on robust whitening," *IEEE Trans. Signal Processing.*, vol. 42, pp. 1572–1576, June 1994.

[2] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 2nd ed., 1994.

[3] A. A. Giordano and F. M. Hsu, *Least Square Estimation with Application to Digital Signal Processing*. New York: John Wiley & Sons, 1985.

[4] I. Barrodale and F. D. K. Roberts, "An improved algorithm for discrete l_1 linear approximation," *SIAM J. Numerical Analysis*, vol. 10, pp. 839–848, Oct. 1973.

[5] H. Imai, K. Kato, and P. Yamamoto, "A linear-time algorithm for linear l_1 approximation of points," *Algorithmica*, vol. 4, pp. 77–96, 1989.

[6] P. J. Huber, *Robust Statistical Procedures*. Philadelphia, PA: SIAM, 2nd ed., 1996.

[7] W. J. J. Rey, *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin: Springer-Verlag, 1983.

[8] F. Kuhlmann and G. L. Wise, "On the second moment properties of median filtered sequences of independent data," *IEEE Trans. Comm.*, vol. COM-29, pp. 1374–1379, Sept. 1981.

[9] Neal C. Gallagher, Jr. and G. L. Wise, "A theoretical analysis of the properties of median filters," *IEEE Trans. Acous. Speech, Signal Proc.*, vol. ASSP-29, pp. 1136–1141, Dec. 1981.

[10] A. Restrepo (Palacios) and L. Chacon, "A smooting property of the median filter," *IEEE Trans. Signal Processing.*, vol. 42, pp. 1553–1555, June 1994.

[11] G. Danuser and M. Stricker, "Parametric model fitting: From inlier characterization to outlier detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, pp. 263–280, March 1998.

[12] M. G. Kendall, A. Stuart, and J. K. Ord, *Kendall's Advanced Theory of Statistics*, vol. 1. New York: Oxford University Press, 5 ed., 1987.