OUTPUT DISTRIBUTIONAL INFLUENCE FUNCTION

Sari Peltonen, Pauli Kuosmanen and Jaakko Astola

Signal Processing Laboratory, Digital Media Institute, Tampere University of Technology
P.O. Box 553, FIN-33101 TAMPERE, FINLAND
e-mail: sari@cs.tut.fi, pqo@cs.tut.fi and jta@cs.tut.fi

ABSTRACT

When selecting a filter for an application, it is often essential to know the behaviour of the filter in presence of contamination. This robustness of a filter is traditionally explored by means of influence function (IF) and change-of-variance function (CVF). However, as these are asymptotic measures there is uncertainty of the applicability of the obtained results to the finite length filters used in the real world applications. This paper disperses this uncertainty by presenting a new method, called output distributional influence function (ODIF), for examining the robustness of the finite length filters. The method gives extensive information about the robustness of any filter with known output distribution function. As examples the ODIFs for distribution function, density function, expectation, and variance are given for the mean and the median filters and interpreted in detail.

1. INFLUENCE FUNCTION

Influence function (IF) is a useful heuristic tool of robust statistics introduced by Hampel [2, 3] under the name influence curve (IC) for studying the performance of filters under noisy conditions.

Definition 1. The IF of estimator T at underlying probability distribution F is given by

$$\mathrm{IF}(x;T,F) = \lim_{t\to 0^+} \frac{T\left((1-t)F + t\Delta_x\right) - T(F)}{t}$$

for those x where this limit exists.

In this definition Δ_x is the probability measure which puts mass 1 at the point x. The IF gives the effect that an infinitesimal contamination at point x has on the estimator T when divided by the mass of the contamination. So the IF gives asymptotic bias caused by the contamination and thus characterizes properties of the estimator as the number of observations approaches infinity.

We denote by Φ and ϕ the distribution and the density functions of the standard normal distribution. The influence functions for the mean and the median are shown in Figure 1 where the underlying distribution $F = \Phi$. For the mean the gross error sensitivity, i.e., the worst influence which a small amount of contamination of fixed size can have on the value of the estimator, equals infinity and for the median it is finite and equals $\sqrt{\frac{\pi}{2}} \approx 1.253$. So for the mean single outlier can carry the estimate over all bounds but for the median an outlier has a fixed influence.

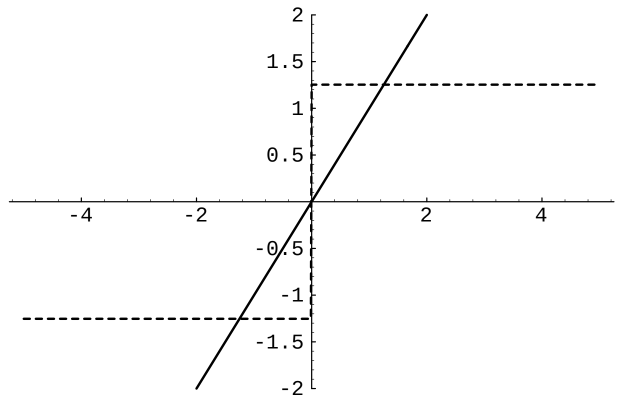

Figure 1: The IF of the mean (–) and the median (- -) at $F = \Phi$.

2. CHANGE-OF-VARIANCE FUNCTION

The IF gives only one aspect of robustness of an estimator, namely local robustness of the asymptotic value of the estimator. Another important aspect is the local robustness of asymptotic variance. The asymptotic variance of estimator T at F denoted by $V(T,F)$ is defined to be the variance of $\sqrt{N}\,[T(F_N) - T(F)]$ as $N \to \infty$, where F_N is the empirical distribution of sample $(X_1, X_2, \ldots, X_N)$. Local robustness of the asymptotic variance can be characterized by the change-of-variance function (CVF) defined as follows, [4].

Definition 2. The CVF of estimator T at F is defined as

$$\mathrm{CVF}(x;T,F) = \lim_{t\to 0^+} \frac{V(T,(1-t)F + t\Delta_x) - V(T,F)}{t}$$

for those x where this limit exists.

If $F = \Phi$, the CVF of the mean is $x^2 - 1$ which is displayed in Figure 2. In the same figure is also shown the CVF of the median at $F = \Phi$. It has a constant value $\sqrt{\frac{\pi}{2}} \approx 1.253$ elsewhere but at zero, where the graph has a negative delta function. If the CVF is negative, the asymptotic variance of the estimator has decreased, and if positive, the asymptotic variance has increased. So for the mean the asymptotic variance decreases if the contamination is in the interval $(-1, 1)$. The further the contamination is from this interval the more the variance is increased and a single outlier can carry the asymptotic variance over all bounds. For the median contamination at the origin reduces the asymptotic variance significantly and the contamination anywhere else causes only a constant increase to the variance of the median. So the median is robust also in this sense.

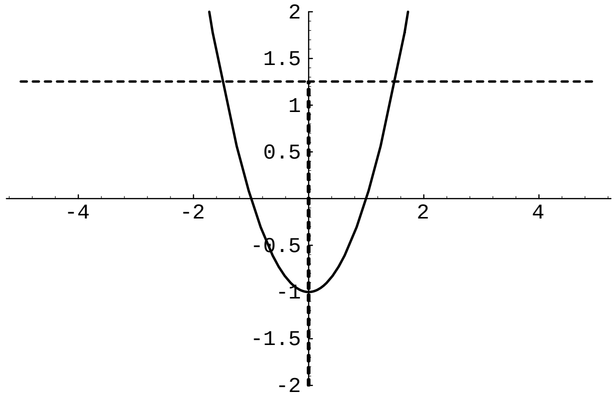

Figure 2: The change-of-variance function of the mean (–) and the median (- -) at $F = \Phi$.

3. OUTPUT DISTRIBUTIONAL INFLUENCE FUNCTION

Since the IF is an asymptotic measure, it describes properties of infinite length filters which may differ from those of finite length filters used in the real world filtering applications. It would be more useful and more interesting to examine properties of these finite length filters rather than the asymptotic properties. For this purpose some finite sample versions of the IF, such as empirical IF, sensitivity curve (SC), and a version using jackknife have been proposed (see e.g. [4] and the references therein). For these either a real sample $(X_1, X_2, \ldots, X_N)$ or an artificial sample generated from the distribution F of the input samples is needed and this sample itself or the way it is derived from the distribution F affects the result. What we would like to have is a general method which uses the distribution function F of the input samples itself and not any artificial sample derived from F. In the case where the output distribution of a filter can be expressed in a closed form as a function of the distribution functions of the input samples we propose this method to be output distributional influence function (ODIF) introduced in this paper.

We assume here that the input samples are independent and identically distributed (i.i.d.) random variables. First we need a way to denote the output distribution function of a filter when a fraction ε of the input samples has different distribution than the rest of the samples. We denote by $H_{(1-\varepsilon)F+\varepsilon G_y}(\cdot)$ the output distribution $H_F(\cdot)$ of the filter where every occurrence of the common distribution function F of the input samples is replaced by $(1-\varepsilon)F+\varepsilon G_y$ and G_y can be any distribution function with mean y. As usual, we define $h_{(1-\varepsilon)F+\varepsilon G_y}(x) = \frac{d}{dx}H_{(1-\varepsilon)F+\varepsilon G_y}(x)$. Now the following definition gives the ODIF for the distribution function.

Definition 3. Let the output distribution function of a filter be $H_F(\cdot)$ where $F(\cdot)$ is the common distribution function of the input samples and let $G_y(\cdot)$ be a distribution function having mean y. Then the ODIF for the distribution function $\Omega(\cdot)$ is

$$\Omega(x,y) = \lim_{\varepsilon\to 0^+} \frac{H_{(1-\varepsilon)F+\varepsilon G_y}(x) - H_F(x)}{\varepsilon}$$

for those x and y where this limit exists.

If the output distribution function $H_F(\cdot)$ can be expressed as a simple function of the input distribution F and thus does not contain any derivative of F, then $H_{(1-\varepsilon)F+\varepsilon G_y}(\cdot)$ is the first-order von Mises expansion of $H_F(\cdot)$ at F evaluated in $(1-\varepsilon)F(\cdot) + \varepsilon G_y(\cdot)$ and is given by

$$\begin{aligned} H_{(1-\varepsilon)F+\varepsilon G_y}(x) &= H_F(x) + \frac{h_F(x)}{f(x)}\varepsilon\left(G_y(x) - F(x)\right) \\ &\quad + \sum_{k=2}^{\infty} \frac{1}{k!}\frac{d^k H_F(x)}{(dF(x))^k}\varepsilon^k \left(G_y(x) - F(x)\right)^k . \end{aligned}$$

For this subclass of output distribution functions the ODIF for the distribution function $\Omega(\cdot)$ can be expressed as

$$\Omega(x,y) = \frac{h_F(x)}{f(x)}\left(G_y(x) - F(x)\right). \tag{1}$$

The notations used for the expectation μ, the variance σ^2, the r^{th} central moment μ_r, and the r^{th} moment about the origin α_r of different distributions are discriminated by giving the distribution function as subindex, e.g., $\mu_{H_F,r}$ is the r^{th} central moment of the distribution H_F. We can define the ODIF in the same way as for distribution function in Definition 3 also for the density function and moments. The ODIFs for the density function, the r^{th} central moment, and the r^{th} moment about the origin are

$$\begin{aligned} \omega(x,y) &= \lim_{\varepsilon\to 0^+} \frac{h_{(1-\varepsilon)F+\varepsilon G_y}(x) - h_F(x)}{\varepsilon}, \\ \omega_{\mu_r}(y) &= \lim_{\varepsilon\to 0^+} \frac{\mu_{H_{(1-\varepsilon)F+\varepsilon G_y},r} - \mu_{H_F,r}}{\varepsilon}, \end{aligned}$$

and

$$\omega_{\alpha_r}(y) = \lim_{\varepsilon\to 0^+} \frac{\alpha_{H_{(1-\varepsilon)F+\varepsilon G_y},r} - \alpha_{H_F,r}}{\varepsilon},$$

respectively for those x and y where the limits exist. They can also be derived to be

$$\omega(x,y) = \frac{d}{dx}\Omega(x,y), \tag{2}$$

$$\omega_{\alpha_r}(y) = \int_{-\infty}^{\infty} x^r \omega(x,y)dx, \tag{3}$$

and

$$\begin{aligned} \omega_{\mu_r}(y) &= \sum_{k=0}^{r-2}\binom{r}{k}(-1)^k \mu_{H_F}^k \omega_{\alpha_{r-k}}(y) \\ &\quad + \omega_\mu(y)\sum_{k=1}^{r-1} k\binom{r}{k}(-1)^k \mu_{H_F}^{k-1}\alpha_{H_F,r-k}. \end{aligned} \tag{4}$$

4. ODIFS FOR MEAN

Let X be the sum of i.i.d. random variables $X_1, X_2, \ldots, X_N$ having common density function $f(\cdot)$. Since the density of the sum of N independent random variables is the convolution of the individual densities (see e.g. [1]), the density function of X is

$$f_X(x) = \overbrace{f(x) * f(x) * \ldots * f(x)}^{N \text{ times}}.$$

For the sake of clarity of representation we introduce a new notation for convolution power

$$f(x)^{*N} = \overbrace{f(x) * f(x) * \ldots * f(x)}^{N \text{ times}}.$$

By using this notation the above density function of X obtains the form $f_X(x) = f(x)^{*N}$. Now by the density function method, [1], we obtain that the density function of the mean $\frac{X}{N}$ is

$$h_F(x) = N^N f(Nx)^{*N}$$

and thus the distribution function is

$$H_F(x) = N^{N-1} f(Nx)^{*(N-1)} * F(Nx).$$

The ODIFs for the distribution and the density functions $\Omega(\cdot)$ and $\omega(\cdot)$ of the mean of N samples can be derived by Definition 3 and equation (2) to be

$$\Omega(x,y) = N^N f(Nx)^{*(N-1)} * (G_y(Nx) - F(Nx))$$

and

$$\omega(x,y) = N^{N+1} f(Nx)^{*(N-1)} * (g_y(Nx) - f(Nx)). \quad (5)$$

For $F = \Phi$ and $G_y = \Delta_y$ we obtain

$$\omega(x,y) = N^N f(Nx - y)^{*(N-1)} - N^{N+1} f(Nx)^{*N},$$

where the first term is N times the density function of the mean of length $N-1$ shifted to the location $x = \frac{y}{N}$ and the second term is $(-N)$ times the density function of the mean of length N. So the contamination decreases the density function in the area of the original density function and increases it around the location $x = \frac{y}{N}$. Figure 3 shows the ODIF for the density function of the mean of length 5 for eight different values of y as function of x. The graph where the maximum value is smallest is the one with the smallest y. When $y = 0$, the contamination is placed to the mean of Φ and it increases the height of the density peak and decreases the tails of the density. For larger values of y the contamination attracts the peak of the density and simultaneously decreases the density in the area of original density.

The ODIF for the expectation gives the effect that the infinitesimal contamination in the input has on the expectation of the output of the filter when divided by the mass of this contamination. When the ODIF for the expectation is negative, the contamination G_y has decreased the expectation of the filter, and when it is positive, the expectation has increased. So the ODIF for the expectation actually is similar to the IF but it is defined for the finite length filters and the distribution function of the contamination is not limited to be Δ_y but can be any distribution function G_y.

After substituting $\omega(\cdot)$ from equation (5) into equation (3) and simplifying the expression, the ODIF for the expectation $\omega_\mu(y)$ is

$$\omega_\mu(y) = y - \mu_F.$$

So for the mean filter the ODIF for the expectation is the same for all lengths N and depends only on the expectations of distributions G_y and F. If $\mu_F = 0$, then the ODIF for the expectation is the same as the IF of the mean for $F = \Phi$ in Figure 1. More generally, for any distributions F and G_y the IF and the ODIF for the expectation are equal. So the finite length mean filters have exactly the same robustness properties as the asymptotic case given by the IF, i.e., the further the contamination is from the origin the larger is the influence it has on the expectation.

The ODIF for the variance $\omega_{\sigma^2}(\cdot)$ of the mean of N samples can be derived by using equation (4) to be

$$\omega_{\sigma^2}(y) = \frac{\sigma^2_{G_y} + y^2 - 2\mu_F y + \mu_F^2 - \sigma_F^2}{N}.$$

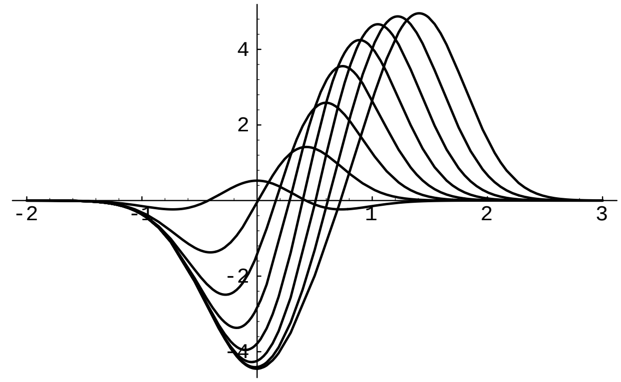

Figure 3: The ODIFs for the density function of the mean of length 5 for $y = 0, 1, \ldots, 7$ at $F = \Phi$ and $G_y = \Delta_y$. The graph where the maximum value is smallest is the one with the smallest y.

Now $N\omega_{\sigma^2}(\cdot)$ is the same for any filter length N and depends only on the means and the variances of the distributions F and G_y. If $F = \Phi$ and $G_y = \Delta_y$, then $N\omega_{\sigma^2}(y)$ is $y^2 - 1$ which is the same as the CVF for $F = \Phi$ displayed in Figure 2. The only difference between $N\omega_{\sigma^2}(\cdot)$ and the CVF of the mean for any distributions F and G_y is that the former has an additional term $\sigma^2_{G_y}$ which for $G_y = \Delta_y$ equals zero.

5. ODIFS FOR MEDIAN

Since the median of $N = 2n+1$ samples is the $(n+1)^{th}$ order statistic, the output distribution and density functions of the median of N samples are

$$H_F(x) = \sum_{k=n+1}^{N} \binom{N}{k} F(x)^k (1 - F(x))^{N-k}$$

and

$$h_F(x) = \frac{N!}{(n!)^2} F(x)^n (1 - F(x))^n f(x).$$

The ODIF for the distribution and the density functions $\Omega(\cdot)$ and $\omega(\cdot)$ of the median of N samples are by equations (1) and (2)

$$\Omega(x,y) = \frac{N!}{(n!)^2} F(x)^n (1 - F(x))^n (G_y(x) - F(x))$$

and

$$\begin{aligned} \omega(x,y) \quad = \quad & \frac{N!}{(n!)^2} F(x)^{n-1} (1 - F(x))^{n-1} \\ & \times \Big[n f(x) (1 - 2F(x)) (G_y(x) - F(x)) \\ & + F(x) (1 - F(x)) (g_y(x) - f(x)) \Big]. \end{aligned}$$

We consider as an example also for the median the case where $F = \Phi$, $G_y = \Delta_y$ and $N = 5$. Now the ODIF for the density function follows the solid line in Figure 4 when $x < y$ and the dashed line when $x > y$. At point $x = y$ there is a delta function having coefficient $30\Phi(y)^5(1-\Phi(y))^5$. The plot of this coefficient is bell shaped having highest value at $y = 0$ and going practically

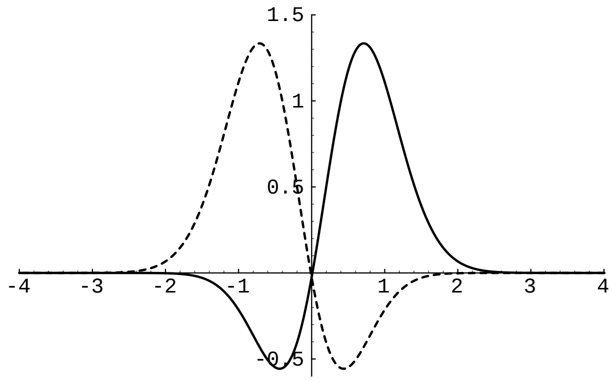

Figure 4: Graphs at $F = \Phi$ and $G_y = \Delta_y$ that the ODIF for the density function follows when $x < y$ (–) and $x > y$ (- -).

to zero in the same area as the graphs in Figure 4 go. This figure together with the knowledge that the coefficient of delta function is practically zero when $|y| > 2.5$ predicts that contamination anywhere in this area has the same fixed influence predicting bounded ODIFs for the expectation and the variance and robustness of the filter. This will be confirmed in the next section.

5.1. Effect of the Filter Length upon Robustness

In this subsection we examine the effect of the filter length upon the robustness of the median filter by using ODIF for the expectation and the variance. Often when the nonlinear filters are used, the lengths are short and the asymptotic behaviour given by the IF and the CVF can be far from the actual one and even include behaviour that does not exist for the small filter lengths. For this problem the ODIF provides a solution by giving the graphs for the specified filter length.

When $F = \Phi$ and $G_y = \Delta_y$, the following form is obtained for the ODIF for the expectation $\omega_\mu(\cdot)$ of the median of N samples from equation (3)

$$\begin{aligned}
\omega_\mu(y) \quad = \quad & \frac{N!}{(n!)^2}\Bigg[\int_{-\infty}^{y} xnF(x)^n(1-F(x))^{n-1}f(x) \\
& \times(2F(x)-1)dx + \int_{y}^{\infty} xnF(x)^{n-1} \\
& \times(1-F(x))^n f(x)(1-2F(x))dx \\
& + yF(y)^n\,(1-F(y))^n \\
& - \int_{-\infty}^{\infty} xF(x)^n\,(1-F(x))^n\,f(x)dx\Bigg].
\end{aligned}$$

In Figure 5 are graphs of the ODIFs for the expectation of the median filter for three different filter lengths and the IF of the median from Figure 1. The ODIFs for the expectation provide similar quantitative information and similar quantities can be derived from them as from the IF but now for finite length filters. The ODIFs for the expectation for different filter lengths differ from each other and also from the IF of the median. However, as the length of the filter increases, the ODIF for the expectation approaches to the IF of the median. As can be observed from Figure 5 the supremum of the absolute value of the median is limited also for finite length filters but has higher value for smaller filter length. Since the IF

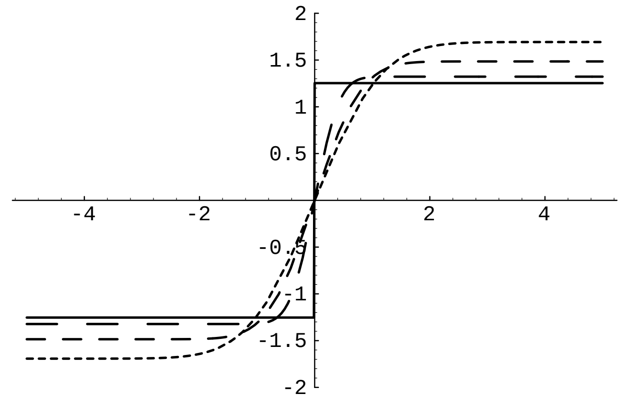

Figure 5: The IF of the median (solid line) and the ODIFs for the expectation of the median filters of lengths 3 (short dashes), 5 (medium dashes), and 15 (long dashes) at $F = \Phi$ and $G_y = \Delta_y$.

of the median has a jump at zero, a wiggling phenomenon can occur when there are small fluctuations in the observations. As the graphs of the ODIFs for the expectation in Figure 5 do not have this jump, the wiggling phenomenon is not a problem in the small finite length median filters but becomes such as the filter length gets longer.

In the same case as above, where $F = \Phi$ and $G_y = \Delta_y$, the ODIF for the variance $\omega_{\sigma^2}(\cdot)$ of the median is obtained by using equation (4) and noticing that in this case $\mu_{H_F} = 0$. Thus

$$\begin{aligned}
\omega_{\sigma^2}(y) \quad = \quad & \frac{N!}{(n!)^2}\Bigg[\int_{-\infty}^{y} x^2 nF(x)^n\,(1-F(x))^{n-1}\,f(x) \\
& \times(2F(x)-1)dx + \int_{y}^{\infty} x^2 nF(x)^{n-1} \\
& \times(1-F(x))^n\,f(x)\,(1-2F(x))\,dx \\
& + y^2F(y)^n\,(1-F(y))^n \\
& - \int_{-\infty}^{\infty} x^2F(x)^n\,(1-F(x))^n\,f(x)dx\Bigg].
\end{aligned}$$

The graph of function $N\omega_{\sigma^2}(y)$ is shown in Figure 6 for the same three lengths as in the previous figure. There is also shown for comparison the CVF from Figure 2. The graphs of the ODIFs for the variance multiplied by N can be interpreted in a similar manner as the graphs of the CVF. So we can see that the finite length median is robust in this sense since the graphs are bounded above. As N increases, the positive constant part gets smaller and simultaneously the negative spike in the origin gets narrower and deeper. So the form of the graph of ODIF for variance multiplied by N approaches the CVF as N approaches infinity.

5.2. Analysis of Different Noise Distributions G_y

In the IF quite unrealistic noise model, the delta distribution, is always used. In our ODIFs we can use more realistic noise distributions and thus make use of the knowledge we have of the possible noise sources in specific applications. So the ODIF is much more flexible method than the IF and the CVF and it can be adjusted to varying conditions by adjusting the distribution G_y.

As an example we examine the median filter of length 5 when $F = \Phi$ and G_y is delta, Laplace, normal, and uniform distribution.

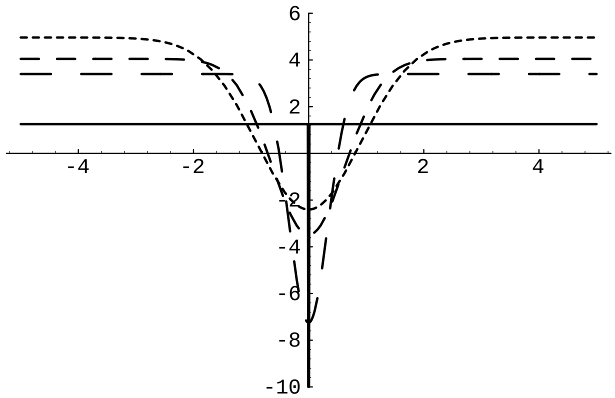

Figure 6: The change-of-variance function of the median (solid line) and the ODIFs for the variance of the median filters of lengths 3 (short dashes), 5 (medium dashes), and 15 (long dashes) multiplied by the filter length N at $F = \Phi$ and $G_y = \Delta_y$.

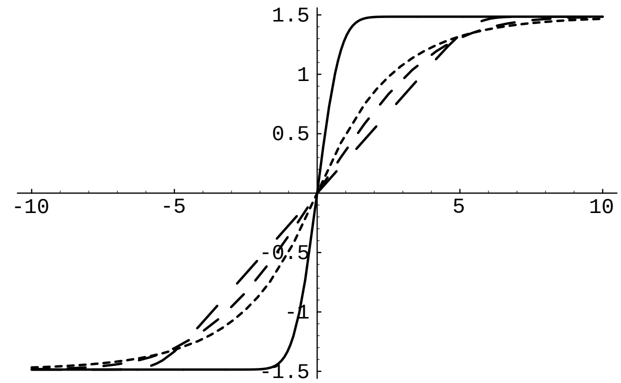

Figure 7: The ODIFs for the expectation of the median filter of length 5 at $F = \Phi$ when G_y is Laplace (short dashes), normal (medium dashes), and uniform distribution (long dashes) with variances 10 and when $G_y = \Delta_y$ (solid line).

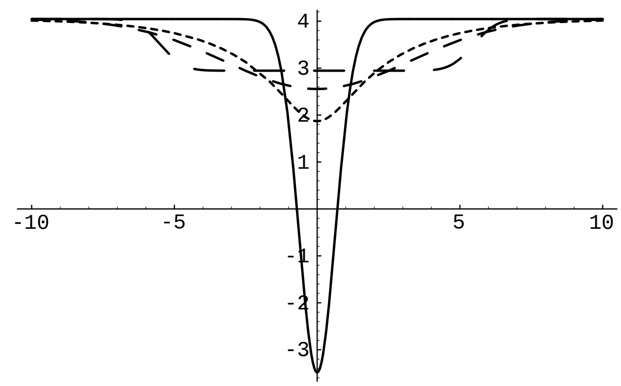

Figure 8: The ODIFs for the variance of the median filter of length 5 multiplied by 5 at $F = \Phi$ when G_y is Laplace (short dashes), normal (medium dashes), and uniform distribution (long dashes) with variances 10 and when $G_y = \Delta_y$ (solid line).

In Figure 7 we show the ODIFs for the expectation for the above median filter for the different noise distributions G_y. We can see from this figure that the constant parts of the graphs are all on the same level and thus the supremum of the absolute value is the same for all the four cases. In the area of small contamination the graphs are different but graphs for other distributions than the delta distribution differ from each other only slightly and much more from the one for the delta distribution. The mean, the variance and the shape of the density of the contamination all have influence on the ODIF for the expectation obtained for the median. When the variance of any of the distributions approaches zero the graphs approach the solid line in Figure 7 obtained for delta distribution. The smaller the variance the nearer the origin the constant level is reached. The shape of the density determines the behaviour of the graphs near the origin. The Laplace distribution is the most peaked of the three and the uniform the flattest. The more peaked the distribution the closer the result is to the graph for delta distribution in the vicinity of the origin. The tails of the contaminating distribution on the other hand determine how soon the constant level is reached. So the graph for the uniform distribution reaches the constant level with smallest y and the graph for the Laplace with highest.

Figure 8 shows the ODIFs for the variance multiplied by N in the same situation as Figure 7 for the expectation. Very similar behaviour due to the shape of the contaminating distribution is noticeable from this figure as was from the previous one. Since the variance of contamination is ten times the variance of distribution Φ, for the Laplace, the normal, and the uniform distributions, the variance never decreases compared to the case with no contamination. By lowering the variances of the contaminations the ODIFs for the variance multiplied by N would approach the solid line and become negative near the origin.

6. CONCLUSIONS

In this paper we have introduced the ODIF, a new useful method for assessing the robustness properties of finite length filters. The ODIF was shown to be a good theoretical analysis tool which at the same time is applicable to the real filtering situations. Nonlinear filters do not have many theoretical analysis tools and the ODIF gives a new possibility to analyse and compare many of these filters. The performance of the method was demonstrated by using the commonly used mean and median filters as examples and the relations of the ODIFs for the expectation and the variance to the traditionally used IF and CVF were discussed. Since in the same way as for the IF and the CVF the interpretation of the obtained curves is very essential, detailed interpretation was given for all the figures. Especially the effect of the filter length and the different noise distributions were considered.

7. REFERENCES

[1] C. Ash, *The Probability Tutoring Book*, IEEE Press, 1993.

[2] F. R. Hampel, Contribution to the Theory of Robust Estimation, Ph.D. Thesis, University of California, Berkeley, 1968.

[3] F. R. Hampel, ”The Influence Curve and Its Role in Robust Estimation”, in *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, June 1974.

[4] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.