Quaderni di Statistica
Vol. 8, 2006

Robust object detection from laser scanner data

Carlo Grillenzoni
Dipartimento di Pianificazione
Università IUAV di Venezia
E-mail: carlog@iuav.it

Summary: Motivated by the need of interpolating spatial data containing strong discontinuities, this paper develops robust nonparametric smoothers. Classical methods (kernel and local regression) are flexible tools, but they are not able to track sudden changes in the surfaces. Their robust version, with non-monotone score functions, are better in this respect because they enhance local properties. On the other hand, statistical analysis shows that such estimators may be non-consistent even in situations of continuous and smooth regions.

Keywords: Discontinuous surfaces, Kernel M-estimates, Redescending scores.

1. Introduction

Interpolation of $spatial$ data is present in many empirical contexts, such as mining surveys, laser scanning reliefs, environmental monitoring, and so on. In all of these cases, the underlying surfaces may present discontinuities and jumps that are difficult to "track" with conventional smoothers.

The method of $kernel$ regression (Härdle, 1991) is suitable for reconstructing complex surfaces, because is flexible and reduces the assumptions at the minimum. Moreover, its $robust$ version (Hall and Jones, 1990) is useful for discontinuities because treats observations beyond jumps as outliers, and therefore local properties of estimates are enhanced.

Robust kernel smoothers are mostly used in image processing, where

they provide the basis for jump detection tests and edge preserving denoising filters (Chu *et al.*, 1998; Polzehl *et al.*, 2000; Qiu, 2002; Hwang, 2004; and references therein). In this context, they exploit the fact that data are available on regular lattices, which simplifies computation and analysis. The proposed filters have worked well; however, they are subject to the strong competition of numerical methods of computer vision engineering (Starck *et al.*, 2003).

In the case of point spatial data, there are only few alternative to non-parametric methods, including kriging, wavelets and their robust versions. However, not all of the robust approaches are useful in tracking jumps, because only $non\text{-}monotone$ score functions enables adaptivity. This excludes, for example, the Huber's preferred solution, which generally enables the consistency to estimators in smooth regions. The trade-off between adaptivity and consistency is well known in time-varying coefficient methods (Grillenzoni, 1994); in the present context, it is the price to be paid for allowing the jump tracking ability to smoothers.

The aim of this article is to adapt robust smoothers of image processing to the interpolation of stochastic 3D point data. This involves modifications of estimators due to the fact that the second problem deals with missing data. Further, there are methodological aspects related to bandwidth design and conditions of convergence that have not been adequately treated in aforementioned papers. We develop these points starting from basic aspects of robust smoothing and with the support of a running example.

The plan of the work is as follows: Section 2 introduces kernel methods and provide a heuristic robust modification. Section 3 discusses kernel M-type smoothers and develops iterative weighted algorithms. Throughout these sections, an extended numerical application on laser scanner data is carried out to test and compare the methods. Finally, Section 4 discusses some statistical properties of kernel M-estimators.

1.2 Aerial Laser Scanner Data

Light detection and ranging (LDR), namely airborne laser scanning, is a relatively new technology for obtaining 3D spatial data with high

density and high positional accuracy of the earth surface. The system is placed on an aircraft and includes a laser scanner, a GPS receiver and a computer. It works by computing the distance to the target point by emitting a laser pulse and measuring the round trip time. Contrary to passive sensors (such as the optical one), the system can work by night, it is not sensitive to shadows and is able to provide the buildings height. Resulting data have a $punctual$ nature, and for each i-th point, the time t, the latitude y, the longitude x, the height z, and the reflectance w, are available: $\{ t_i, \boldsymbol{s}_i, w_i \}$.

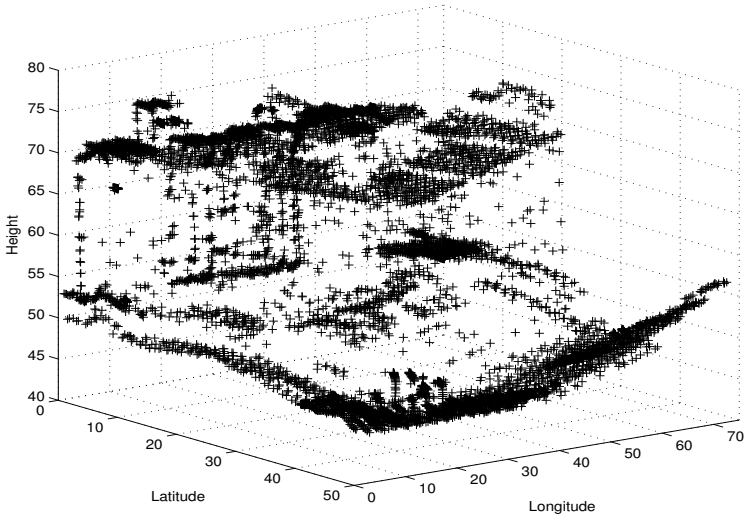

Figure 1. Representation of LDR data; sample size N=8369.

Focusing on spatial coordinates $\boldsymbol{s}_i = [\, x_i, y_i, Z_i \,]'$ LDR data can be used to obtain 3D representation of houses, cities and landscapes. To this end a common problem is interpolation, because observations are not available on regular grids, they are subject to several random effects and the resolution required for the output may be very accurate. As an example, we consider a subset of the data studied in Wang and Tseng (2004) concerning the city of Hsinchu (Taiwan) and generated by Leica ALS40 instrument. Original data cover a square area of 1/2 Km^2, while our subset regards an area of 70×50 m, containing N=8369 points. Figure 1

provides 3D representation: one can note that data are very dense and spatial coordinates are subject to random fluctuations. This data-set will be used in the paper to test and compare the various interpolation methods. These consist of linear kernel regression, robust kernel smoothing, nonlinear M-estimation and adaptive weighted regression.

2. Kernel regression estimates

Nonparametric smoothers (Härdle, 1991) can be very useful to deal with LDR data because the underlying surfaces are very complex and they do not assume specific mathematical forms for them. Moreover, they can be easily extended to multiple dimensions; consider the bivariate regression model

$$\mathrm{Z}_i = g(\mathrm{x}_i, \mathrm{y}_i) + \varepsilon_i\,, \quad \varepsilon_i \sim \mathrm{IID}(0, \sigma^2_\varepsilon)\,; \qquad i = 1, 2 \ldots N$$

where $g(\cdot)$ is a nonlinear determinist function and ε is a noise process, a typical tool for estimating $g(\cdot)$ is the kernel (K) regression

$$\begin{aligned} \hat{g}_{\mathrm{K}}(x, y) &= \sum_{i=1}^{N} v_i(x, y)\, \mathrm{Z}_i \qquad (1) \\ v_i(x, y) &= \frac{K_1\Big[(\mathrm{x}_i - x)/\lambda_1\Big]\, K_2\Big[(\mathrm{y}_i - y)/\lambda_2\Big]}{\sum_{i=1}^{N}\; K_1\Big[(\mathrm{x}_i - x)/\lambda_1\Big]\, K_2\Big[(\mathrm{y}_i - y)/\lambda_2\Big]} \end{aligned}$$

where $(x, y) \in \Re^2$ are real variables, $\{\mathrm{x}_i, \mathrm{y}_i, \mathrm{Z}_i\}$ are observations, $K_h(\cdot),\ h = 1, 2$ are symmetric density functions and $0 < \lambda_h < \infty,\ h = 1, 2$ are smoothing coefficients.

Optimal design of such coefficients can be obtained with cross-validation techniques, by minimizing the loss function

$$Q_N(\lambda_2,\ \lambda_2) = \sum_{j=1}^{N} \Big[\, \mathrm{Z}_j - \hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j) \Big]^2$$

where $\hat{g}_{\mathrm{K}-j}(\cdot)$ are estimates as in (1) obtained by omitting the j-th observation. Applying this method to the data of Figure 1, under the choice of

Gaussian kernels, provided $\hat{\lambda}_1$=0.05 and $\hat{\lambda}_2$=1.32. However, these coefficients produce estimates which tend to follow the "stripes" of the aircraft. On the contrary, constraining on λ_1=λ_2=λ, yields $\hat{\lambda}$=0.6, which generates regular estimates in Figure 2.

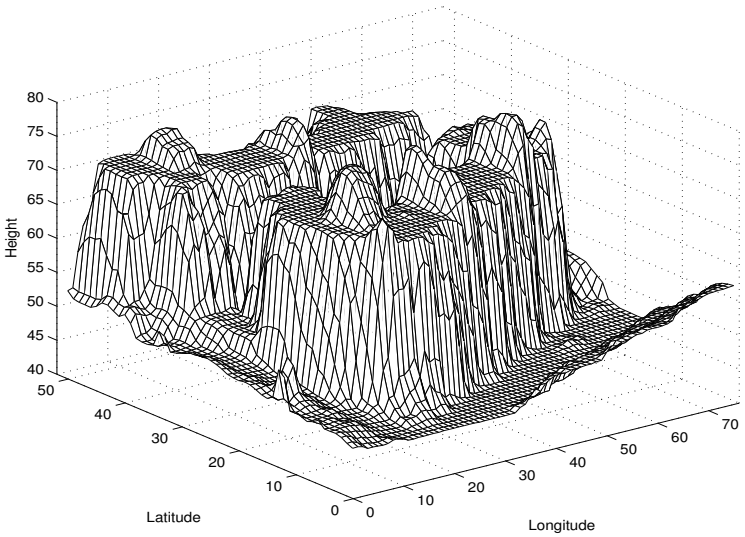

Figure 2. Kernel regression estimates of the data in Figure 1, obtained with algorithm (1), Gaussian kernels, and $\lambda_1 = \lambda_2 = 0.6$.

Despite of their flexibility, classical nonparametric estimators have major problems when surfaces present discontinuities, just because they smooth the data. This can be checked in Figure 2 by noting that building walls are not as sharp as should be expected (Figure 1), and large disturbances occur in correspondence of the jumps. A better $visual$ performance is allowed by a smaller bandwidth, but this slightly improves the situation and cannot be obtained with minimax approach. In fact $\min_\lambda \max_j \left|\mathrm{Z}_j - \hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j)\right|$ yields $\hat{\lambda}$=1.05.

Analysis of the disturbances is an important diagnostic step. Here, one should distinguish prediction errors $\hat{\varepsilon}_j = \left[\mathrm{Z}_j - \hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j)\right]$ from residuals of regression $\check{\varepsilon}_i = \left[\mathrm{Z}_i - \hat{g}_{\mathrm{K}}(\mathrm{x}_i, \mathrm{y}_i)\right]$. In the case of linear models,

it can be shown that the estimator $\tilde{\sigma}_\varepsilon^2 = N^{-1} \sum_{l=1}^{N} (\hat{\varepsilon}_l \check{\varepsilon}_l)$ is unbiased (Grillenzoni, 1994), but is sensitive to outliers. A robust alternative is the median (med) absolute deviation (MAD)

$$\hat{\sigma}_\varepsilon^* = \mathrm{med}_i \Big\{ | \hat{\varepsilon}_i - \mathrm{med}_j(\hat{\varepsilon}_j) | \Big\} / 0.6745$$

where the constant 0.6745 allows consistency in the Gaussian case.

On the basis of these formulas we obtained the estimates $\hat{\sigma}_\varepsilon$=3.24, $\check{\sigma}_\varepsilon$=2.57, $\tilde{\sigma}_\varepsilon$=2.91 and $\hat{\sigma}_\varepsilon^*$=0.25, which are somewhat different from the others and reveal the presence of non-normality. Kernel densities $\hat{f}_{\mathrm{K}}(\varepsilon)$ of prediction errors and residuals confirm this hypothesis. They were computed with the sub-optimal bandwidth $\hat{\sigma}_\varepsilon / N^{1/5}$=0.53 (Härdle, 1991, p.91), which is very similar to the cross-validation estimate $\hat{\lambda}$=0.6. These densities are displayed in Figure 3, showing very heavy tails; these are produced by gross errors in correspondence of jumps.

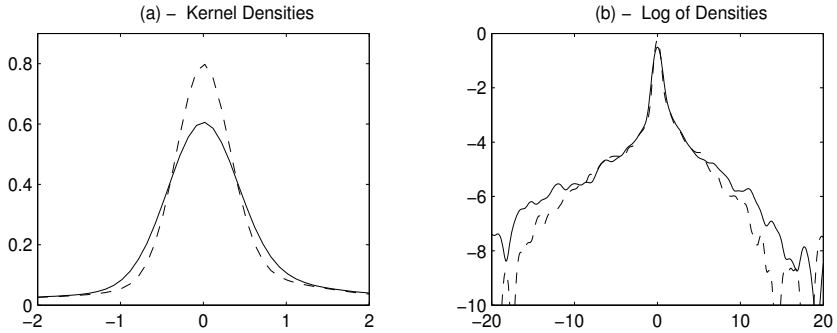

Figure 3. Kernel densities of prediction errors (solid line) and residuals (dashed line) of the kernel regression in Figure 2.

2.1 Robust kernel regression

A possible solution to the poor fitting of the kernel regression can be obtained by filtering large residuals. Indeed, these are the effect and the cause of surface oversmoothing at the jumps. The matter is typically pursued in $robust$ estimation (Huber, 1981), where residuals are controlled by modifying the loss function. Now, because the estimator (1) minimizes the "local" quadratic functional $\sum_{i=1}^{N} v_i(x,y)\,(\mathrm{Z}_i - g)^2$, its "robustification" can be achieved by introducing local weights even for the

Z variable, by means of a suitable kernel function $K(\cdot)$. The loss to be minimized thus becomes

$$P_N(g) = \frac{1}{N} \sum_{i=1}^{N} v_i(x,y)\, K\Big(\mathrm{Z}_i - g\Big)\,(\mathrm{Z}_i - g)^2 \tag{2}$$

where the metric $\big[\, K(\varepsilon)\cdot\varepsilon^2 \,\big]$, which replaces the quadratic distance $(Z-g)^2$, is uniformly bounded; therefore, it belongs to the class of robust loss functions (Hampel *et al.*, 1986).

Minimization of (2) does not lead to an $explicit$ (closed form) solution; however, using the previous estimate $\hat{g}_{\mathrm{K}}(\cdot)$ inside $K(\cdot)$, one can obtain the robust (R) algorithm

$$\hat{g}_{\mathrm{R}}(x,y) = \sum_{i=1}^{N} w_i(x,y)\,\mathrm{Z}_i \tag{3}$$

$$w_i(x,y) = \frac{K_1\Big[(\mathrm{x}_i - x)/\lambda_1\Big]\, K_2\Big[(\mathrm{y}_i - y)/\lambda_2\Big]\, K_3\Big[\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{K}}(x,y)\Big)/\lambda_3\Big]}{\sum_{i=1}^{N} K_1\Big[(\mathrm{x}_i - x)/\lambda_1\Big]\, K_2\Big[(\mathrm{y}_i - y)/\lambda_2\Big]\, K_3\Big[\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{K}}(x,y)\Big)/\lambda_3\Big]}$$

This two-step estimator has a loose connection with the $sigma$ filter of image denoising (Chu *et al.*, 1998). Its performance can be improved through iteration, namely by replacing $\hat{g}_{\mathrm{K}}(\cdot)$ with $\hat{g}_{\mathrm{R}}(\cdot)$, within $K_3(\cdot)$, in the third step, and so on.

In the presence of large amount of data, it is useful to combine the iteration process with data $subsampling$ and $averaging$ of estimates. In practice, the original data-set is split into $m \geq 10$ disjoint random subsets of size $n = N/m$ (rounded). The k-th subset is used in the k-th iteration of the smoother (3), and the resulting estimate $\hat{g}_{\mathrm{R}}^{(k)}(\cdot)$ is averaged to the previous ones, as $\bar{g}_{\mathrm{R}}^{(k)} = k^{-1} \sum_{h=1}^{k} \hat{g}_{\mathrm{R}}^{(h)}$. Finally, this mean value is used in the (k+1)-th iteration within $K_3(\cdot)$, for all $k = 1, \ldots m$.

Design of smoothing coefficients could be carried out with the cross-validation method, by minimizing the functional:

$$Q_N(\lambda_1, \lambda_2, \lambda_3) = \textstyle\sum_{j=1}^{N} [\, \mathrm{Z}_j - \hat{g}_{\mathrm{R}-j}(\mathrm{x}_j, \mathrm{y}_j)\,]^2,$$

where:

$$\hat{g}_{\mathrm{R}-j}(\mathrm{x}_j, \mathrm{y}_j) \propto \sum_{i \neq j}^{N} K_1\left(\frac{\mathrm{x}_i - \mathrm{x}_j}{\lambda_1}\right) K_2\left(\frac{\mathrm{y}_i - \mathrm{y}_j}{\lambda_2}\right) K_3\left(\frac{\mathrm{Z}_i - \hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j)}{\lambda_3}\right) \mathrm{Z}_i$$

Optimization of $Q_N(\cdot)$ is computationally demanding, and attempts to simplify it by substituting $\hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j)$ with Z_j would create numerical problems. Even the solution λ_3=σ_ε, which is sensible in the case of $K_3(\cdot)$ Gaussian, is problematic because previous kernel estimates of σ_ε ranged between (0.25, 2.91).

Application of the cross-validation method to the smoother (3) converged only under the constraint of single bandwidth, providing $\hat{\lambda} = 0.56$. Instability $\hat{\lambda}_3 \to \infty$ originates from the $flatness$ of Q_N in the direction of λ_3, which reveals a problem of parametric identifiability with respect to this coefficient (Hall and Jones, 1990, p.1717). In practice, increasing the sharpness of estimates in correspondence of the jumps, which is determined by $K_3(\cdot)$, only has a weak correspondence in terms of reduction of the prediction errors variance. This is due to the fact that discontinuity edges have area zero in the plane, and the number of observations in correspondence of the jumps is too small to influence Q_N. Therefore, we decided to keep fixed the (stable) kernel estimate $\hat{\lambda}_1 = \lambda_2$=0.6 and investigating the graphical/numerical effects of varying λ_3. Now, for $\lambda_3 < 0.3$ the estimates $\hat{g}_{\mathrm{R}}(\cdot)$ became unstable, whereas for $\lambda_3 > 0.9$ they became smooth like the kernel ones. As a consequence, we selected the midpoint λ_3=0.6 and turn to investigate the effects of data subsampling under this design.

Here, we found that n (the subsample size) outside the range (100,500) produces similar effects as those of λ_3 outside (0.3, 0.9). Thus, we tentatively selected n=300, which implies $N/n \approx 28$ iterations. Smoothers implemented with sequential processing of sub-samples enjoy a better jump-preserving ability because, at each iteration, the probability to have multiple observations within the same "pixels" is drastically reduced. This is particularly useful for the pixels placed on the edges because it avoids averaging data placed on different heights, which causes smoothing of jumps. Resulting estimates, obtained with Gaussian kernels, are displayed in Figure 4. Compared with Figure 2, they enjoy better jump

tracking properties, both in situations of large and small scale variability.

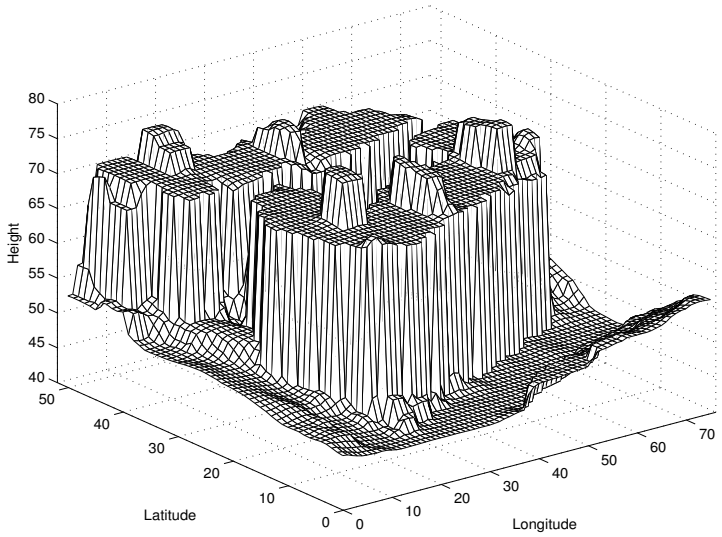

Figure 4. Robust regression estimates generated with the smoother (3), with Gaussian kernels, and the coefficient $\lambda_1 = \lambda_2 = \lambda_3 = 0.6$.

3. Robust nonlinear smoothers

In this section we discuss robust nonparametric smoothers by linking kernel regression and M-estimation methods. This approach was developed by Härdle and Gasser (1984) and Hall and Jones (1990), with regard to one-dimensional problems with fixed and random designs respectively. Usefulness of robust methods in dealing with discontinuous surfaces raises from the fact that they treat data beyond the jumps as outliers. Such data are then implicitly ignored by robust smoothers and local properties of estimates result enhanced.

We refer to the same model $Z = g(x, y) + \varepsilon$ in Section 2, but with the specification that the regression function $g(\cdot)$ is discontinuous deterministic, with jumps located at unknown points. For example, one may

have

$$g(x,y) = \gamma(x,y) + \delta_2 \cdot I_2\Big\{(x,y):\ y \geq \big[\,\varphi(x) + \delta_1 \cdot I_1(x \geq \mathrm{x}_0)\,\big]\Big\} \quad (4)$$

where $\gamma(\cdot)$ is a continuous function, $\delta_{1,2}$ are jumps and $I_{1,2}(\cdot)$ are indicator functions. Notice that in the above scheme, the discontinuity edge of $g(\cdot)$ follows the relationship $y = \varphi(x)$, which also has a jump at the point x_0.

3.1 M-type Estimation

Given the data $\{\mathrm{x}_i, \mathrm{y}_i, \mathrm{Z}_i\}_{i=1}^N$, the kernel M-estimator (KM) is the solution of the locally weighted maximum likelihood type problem

$$\hat{g}_{\mathrm{M}}(x,y) = \arg\min_g \left[P_N(g) = \frac{1}{N}\sum_{i=1}^N v_i(x,y)\,\rho\Big(\mathrm{Z}_i - g\Big)\right] \quad (5)$$

where the weights $\{v_i\}$ are defined as in (1), and $\rho(\cdot)$ is a loss function which controls the robustness (namely the resistance to outliers).

To this end the loss $\rho(z)$ must not grow too rapidly as $|z| \to \infty$; or, more precisely, the score $\psi(z) = \partial\rho(z)/\partial z$ must be uniformly bounded. In this context there are two philosophies: Huber (1981) states that $\psi(z)$ must be monotone and must achieve its maximum value asymptotically, because outliers may contain useful information. On the contrary, Hampel *et al.* (1986) claim that it must be $redescending$, namely $|\psi(z)| \to 0$, because outliers are usually extraneous to the models. These approaches have opposite consequences in terms of convergence and adaptivity of the estimates.

Most common robust loss functions are given as follows:

$$\begin{array}{lrcll}
a) & \rho_a(z) & = & z^2/2\,, & |z| \leq \alpha \\
 & & = & \alpha\,|z| - \alpha^2/2\,, & |z| > \alpha \\
b) & \rho_b(z) & = & z^2/2\,, \quad |z| \leq \alpha & \\
 & & = & \alpha^2/2\,, \quad |z| > \alpha & \\
c) & \rho_c(z) & = & -\exp(-z^2/2) & \\
d) & \rho_d(z) & = & z^2\,\exp(-z^2/2) &
\end{array} \quad (6)$$

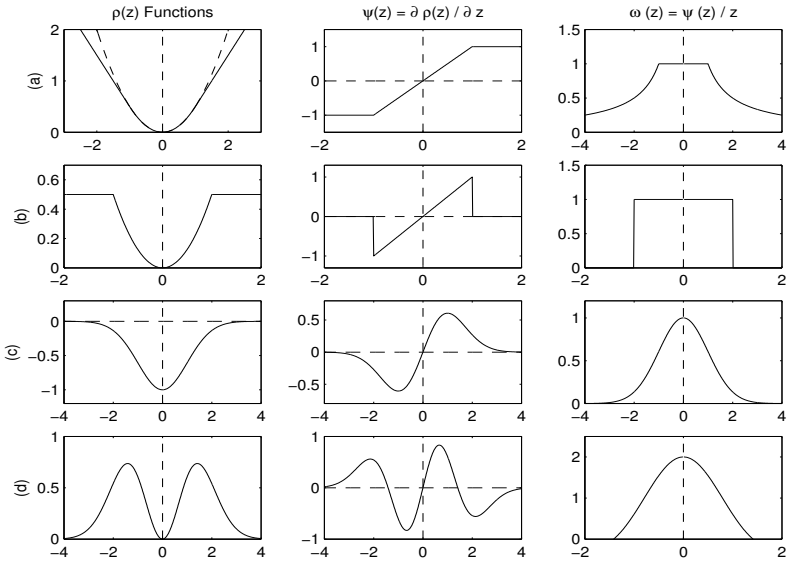

Figure 5. Display of the robust functions in (6) with $\alpha = 1$.

where $0 < \alpha < \infty$ is a design constant that must be selected according to the rate of outlier contamination. Typically, under the assumption of Gaussian disturbances one can choose $\alpha = 2\,\sigma_{\varepsilon}$. Loss function (a) is the preferred one of Huber, while (b) corresponds to the $trimmed$ method. The other two are smoothed solutions and provide redescending ψ-functions; the last, in particular, combines (a) and (c) and coincides with the heuristic method in (2). Graphical behavior of the criteria in (6), and of their corresponding ψ-functions, is shown in Figure 5.

Usefulness of the KM-smoother (5) in estimating discontinuous regression functions, raises from the fact that it has better $local$ properties with respect to classical nonparametric estimators. This is particularly true in the case of redescending ψ-functions, because they create local weighting also in the direction of the dependent variable Z. To appreciate this feature, consider the solution (6,c), namely assume that $-\rho(z)$ is Gaussian. In this case, the minimization problem (5) does coincide with

the maximization of the kernel density function

$$\hat{f}_{\mathrm{K}}(x,y,Z) = \frac{1}{\lambda_1\lambda_2\lambda_3\, N}\sum_{i=1}^{N} K_1\left(\frac{\mathrm{x}_i - x}{\lambda_1}\right) K_2\left(\frac{\mathrm{y}_i - y}{\lambda_2}\right) K_3\left(\frac{\mathrm{Z}_i - Z}{\lambda_3}\right) \tag{7}$$

for any pair (x,y), the estimate $\hat{g}_{\mathrm{M}}(\cdot)$ is the highest value of $\hat{f}_{\mathrm{K}}(x,y,Z)$ in the Z-direction. This relationship demonstrates the close connection between KM-smoothing and the $modal$ regression approach discussed in Scott (1992, p.157).

Because kernel densities have multiple maxima, redescending KM-estimators usually have difficulty to converge to the global optimum point. However, it is the very nature of jumping between the local maxima of $\hat{f}_{\mathrm{K}}(\cdot)$ that determines the jump tracking ability of $\hat{g}_{\mathrm{M}}(\cdot)$ with respect to discontinuous surfaces. The computation of (5), for every point (x,y), typically proceeds by nonlinear algorithms, such as the steepest descent one

$$\hat{g}_{\mathrm{M}}^{(k+1)}(x,y) = \hat{g}_{\mathrm{M}}^{(k)}(x,y) + \alpha_k \sum_{i=1}^{N} v_i(x,y)\,\psi\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{M}}^{(k)}(x,y)\Big)$$

where $\{\alpha_k\}$ is a positive and summable sequence that controls the convergence and the initial value may be selected as $\hat{g}_{\mathrm{M}}^{(0)}(\cdot) = \hat{g}_{\mathrm{K}}(\cdot)$. However, the direct minimization of (5) is computationally demanding, and is suitable only if the grid of values for (x,y) and/or the sample size N, are small.

3.2 W-type estimation

An alternative approach, which leads to a "quasi" linear solution of (5), may be obtained from the $weighted$ form of M-estimates (Hampel *et al.*, 1986, p.115). This follows by the definition of the residual weight function $\omega(z) = \psi(z)/z$, which implies

$$\sum_{i=1}^{N} v_i(x,y)\,\psi\big(\mathrm{Z}_i - g\big) = \sum_{i=1}^{N} v_i(x,y)\,\omega\big(\mathrm{Z}_i - g\big)\,\big(\mathrm{Z}_i - g\big) = 0$$

thus, solving for g, in iterative form, provides the weighted (W) local algorithm

$$\hat{g}_{\mathrm{W}}^{(k+1)}(x,y)=\left[\sum_{i=1}^{N} v_i\,\omega\Big(\mathrm{Z}_i-\hat{g}_{\mathrm{W}}^{(k)}(x,y)\Big)\right]^{-1}\sum_{i=1}^{N} v_i\,\omega\Big(\mathrm{Z}_i-\hat{g}_{\mathrm{W}}^{(k)}(x,y)\Big)\,\mathrm{Z}_i$$

To appreciate the advantage of this structure consider the solution (6,c). It implies that $\omega(z)=\exp(-z^2/2)$, which is a Gaussian kernel. With this, the smoother $\hat{g}_{\mathrm{W}}(\cdot)$ becomes equivalent to the heuristic robust estimator $\hat{g}_{\mathrm{R}}(\cdot)$ in (3). The equivalence is also confirmed by the closeness of the ω-functions (c,d) in Figure 5. Therefore, resorting to weighted form of M-estimators provides robust $pseudolinear$ smoothers which can be managed with non-gradient procedures of Section 2.

Let us now apply this approach to the trimmed (T) specification (6,b). Based on the initial kernel estimator $\hat{g}_{\mathrm{K}}(\cdot)$, the two-step version becomes similar to (3), with $K_3(\cdot)$ replaced by the indicator function $I(\cdot)$=$\omega(\cdot)$. Moreover, the iterative version, integrated with data subsampling and sequential averaging, becomes

$$\bar{g}_{\mathrm{T}}^{(k)}(x,y)=\frac{k-1}{k}\,\bar{g}_{\mathrm{T}}^{(k-1)}(x,y)+\frac{1}{k}\,\hat{g}_{\mathrm{T}}^{(k)}(x,y) \tag{8}$$

$$\hat{g}_{\mathrm{T}}^{(k+1)}(x,y)\propto\sum_{i=1}^{n_k} K_1\left(\frac{\mathrm{x}_{ki}-x}{\lambda_1}\right)K_2\left(\frac{\mathrm{y}_{ki}-y}{\lambda_2}\right)I\left(\left|\mathrm{Z}_{ki}-\bar{g}_{\mathrm{T}}^{(k)}(x,y)\right|<\lambda_3\right)\mathrm{Z}_{ki}$$

where $\{\mathrm{x}_{ki},\mathrm{y}_{ki},\mathrm{Z}_{ki}\}$ are random subsamples of size n_k, and $\sum_k n_k=N$. The formula (8) has a recursive nature, in that the average values are updated with the estimate obtained with the current subsample.

Design of smoothing coefficients of (8) was pursued as in Section 2. Namely, conditioned on the value $\hat{\lambda}_1=\lambda_2=0.6$, of the kernel estimation, we evaluated the effect of λ_3 on the graphical behavior of $\hat{g}_{\mathrm{T}}(\cdot)$. Now, values of $\lambda_3<4$ generate numerical instability, whereas $\lambda_3>6$ produce smoothing effect; therefore, optimal value may be λ_3=5, which is very similar to $2\,\sigma_\varepsilon$ of the kernel estimation. Iterative estimates of $\hat{g}_{\mathrm{T}}(\cdot)$, computed with moderate subsampling $n=N/20$, are reported in Figure 6. With respect to Figure 4, one may note that details on the building roofs are much less clear.

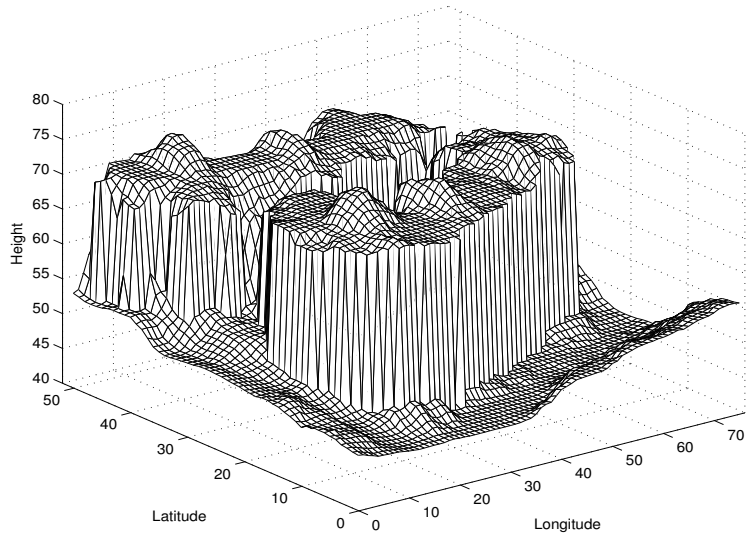

Figure 6. Robust regression estimates generated with the algorithm (8) and the coefficients $\lambda_1 = \lambda_2 = 0.6$ *and* $\lambda_3 = 5$.

As a conclusion, one can state that redescending KM-smoothers are preferable to those with truncated score function. Further experiments have also shown that the Huber solution (6,a) provides estimates which are just intermediate to those in Figure 2 and 6; this means that monotone ψ-functions have weak tracking ability. Finally, it is possible to show that KM-estimates obtained by nonlinear minimization of (5) are very similar to the pseudolinear ones in Figure 4 and 6.

3.3 S-type estimation

Although previous estimators are satisfactory in tracking big jumps, they exhibit some weakness in small variability contexts, such as on the ground or on the building roofs. A way of improving this aspect, consists of simplifying the algorithms, starting from their underlying objective functions. For instance, in (2) one can drop the local weights $v_i(\cdot)$, by

including the kernel estimates in $K(\cdot)$

$$P_N(g) \simeq \frac{1}{N} \sum_{i=1}^{N} K_\lambda \Big[\mathrm{Z}_i - \hat{g}_\mathrm{K}(x, y) \Big] \, (\mathrm{Z}_i - g)^2$$

in fact, $\hat{g}_\mathrm{K}(\cdot)$ already incorporates the weights themselves. Furthermore, by noting that $\mathrm{Z}_i = g(\mathrm{x}_i, \mathrm{y}_i) + \varepsilon_i$, one could also approximate $\mathrm{Z}_i \approx \hat{\mathrm{Z}}_i = \hat{g}_\mathrm{K}(\mathrm{x}_i, \mathrm{y}_i)$, obtaining

$$P_N(g) \approx \frac{1}{N} \sum_{i=1}^{N} K_\lambda \Big[\hat{g}_\mathrm{K}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_\mathrm{K}(x, y) \Big] \, (\mathrm{Z}_i - g)^2$$

Finally, minimizing the above with respect to g, and iterating the resulting solution, one can obtain the simplified (S) algorithm

$$\hat{g}_\mathrm{S}^{(k+1)}(x, y) = \frac{\sum_{i=1}^{N} K \Big[\Big(\hat{g}_\mathrm{K}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_\mathrm{S}^{(k)}(x, y) \Big) / \lambda \Big] \, \mathrm{Z}_i}{\sum_{i=1}^{N} K \Big[\Big(\hat{g}_\mathrm{K}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_\mathrm{S}^{(k)}(x, y) \Big) / \lambda \Big]} \tag{9}$$

Fundamentally, this is a classical (non-local) weighted algorithm, where the weights $\omega(z)$ are specifically suitable in the case of piecewise constant surfaces. Indeed, by modeling $K(\cdot)$ as the indicator function $I(\cdot)$, (9) just provides local sample means of the data Z_i. Moreover, as the value $\Big| \hat{g}_\mathrm{K}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_\mathrm{S}(x, y) \Big|$ becomes large compared to λ, then the two points are almost classified in different regions.

At the implementation level, even the calculation of the component $\hat{g}_\mathrm{K}(\mathrm{x}_i, \mathrm{y}_i)$ of (9) could be iterated; however, this has shown worsening of the jump tracking performance. The above algorithm is very fast and does not need data subsampling; moreover, it is stable and relatively insensitive to the design of λ. Suitable range for such coefficient now is (0.2, 1), and Figure 7 exhibits estimates (9) obtained with λ=0.6 and k=15 iterations. As one can see, small objects and constant components of the surface are significantly enhanced; on the other hand, smooth regions tend to be segmented as a staircase. This problem cannot be solved by increasing λ, because this only cause erasing of small objects on the roof.

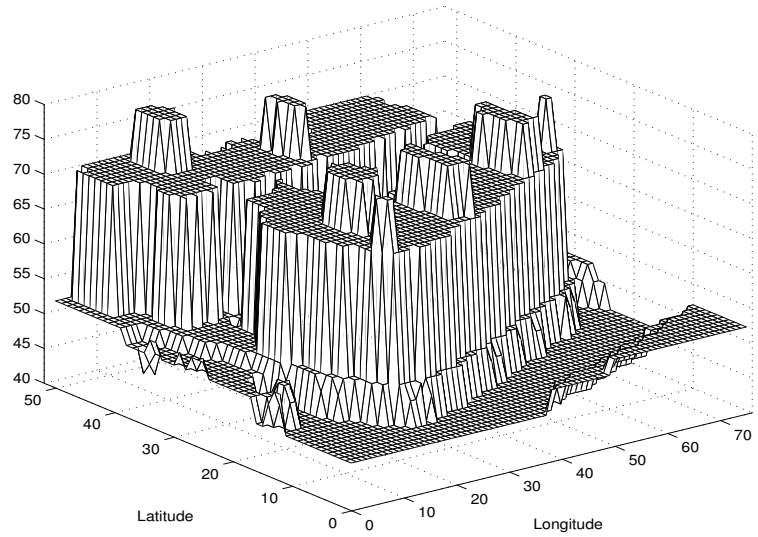

Figure 7. Regression estimates generated with the algorithm (9), with Gaussian kernels and the coefficient $\lambda = 0.6$.

4. Statistical properties

In the previous sections we have discussed several kinds of robust kernel estimators, showing their algebraical connections. In particular, we have proved that nonlinear (5) and pseudolinear (3),(8) smoothers are equivalent if they are fully iterated. The bridge between the two classes of estimators is provided by the weighted average form of M-estimates (Hampel *et al.*, 1986, p. 115). Applications to laser data have also shown that bounded loss functions (6,$b-d$), are suitable for tracking discontinuities in regression surfaces. For smooth regions, however, this feature may rise lack of convergence because they involve non-monotone score functions which involve multiple solutions of the equation $\sum_i v_i(x,y)\,\psi(\mathrm{Z}_i - g) = 0$.

This problem is well known in classical (parametric) M-estimation, where specific assumptions are needed to establish the consistency of algorithms with redescending score functions. For example, Freedman

and Diaconis (1982) show that uniqueness of the global minimum is not enough, and symmetry and monotonicity, on $(-\infty, 0)$, of the underlying probability density are necessary. What emerges in these studies is that the non-monotone nature of $\psi(z)$, by allowing for multiple local solutions, conflicts with the possible multi-modality of $f(Z)$. Now, if $v_i(x, y)$ are strictly positive weights, this conclusion can be extended to KM-smoothers, yielding that their convergence strongly depends on the shape of the noise density. In particular, if $f(\varepsilon)$ has saddle points, then consistency is not guaranteed.

In the following we outline the proof of the consistency of KM-smoothers at the points of continuity of a regression surface. For monotone ψ-functions, such as the Huber solution (6,a), the result was proved by Hall and Jones (1990); they also showed that the asymptotic variance is proportional to $\mathrm{E}[\psi^2(\varepsilon)]/\mathrm{E}^2[\psi'(\varepsilon)]$, as in parametric M-estimation. For redescending score functions, however, more specific assumptions for model (4) and estimator (5) are needed:

*A*1. The $\{\varepsilon_i\}$ are independent and identically distributed, with mean zero and finite variance. The density $f(\varepsilon)$ is strongly unimodal in 0, and has continuous and bounded derivative $f'(\varepsilon)$.

*A*2. The loss function $\rho(z)$ is a negative kernel, namely $\int -\rho(z)\,\mathrm{d}z = 1$, with continuous and bounded derivative $\psi(z)$. The common bandwidth $\lambda \to 0$ in such a way that $N\lambda^3 \to \infty$ as $N \to \infty$.

Under the stated assumptions, $\rho(\varepsilon)$ is of type (6,c) and $-P_N(g)$ in (5) behaves asymptotically like a kernel density estimator for $\{\varepsilon_i\}$. Thus, given the assumption of continuity, as $(\lambda, N^{-1}) \to 0$ the derivative $P'_N(g)$ converges in probability to $f'(\varepsilon)$. More specifically, using the same arguments as Parzen (1962), one can show that the sequence of expectations $\mathrm{E}\big[P'_N(g)\big]$ converges uniformly, namely

$$\sup_g \Big| \mathrm{E}\big[P'_N(g)\big] - f'(Z-g) \Big| \;=\; O\Big(\lambda, (\lambda N)^{-1}\Big)$$

From this it follows the uniform stochastic convergence of $P'_N(g)$

$$\lim_{N\to\infty} \lim_{\lambda\to 0} \mathrm{P}\left(\sup_g \Big| P'_N(g) - f'(Z-g) \Big| < \eta \right) = 1$$

for any constant $\eta > 0$ and for all the points of continuity. Notice that, until now, we have just exploited the equivalence between KM-smoothing and modal regression as defined on the basis of (7).

Finally, because $f(Z - g)$ has no saddle points and just one local maximum, one can show that the estimator $\hat{g}_{\mathrm{M}}(\cdot)$, solution of the equation $P'_N(g) = 0$, converges in probability to the unique zero of $f'(Z - g)$, namely

$$\lim_{N\to\infty} \lim_{\lambda\to 0} \mathrm{P}\left(\sup_{g} \left| \hat{g}_{\mathrm{M}}(x,y) - g(x,y) \right| < \eta \right) = 1$$

for any point (x, y) where $g(\cdot)$ si continuous, and provided that $\lambda^3 N \to \infty$. It should be noted that the condition of strong unimodality of $f(\varepsilon)$ is necessary; thus, when it is not assumed (as in Chu *et al.*, 1998) consistency is not guaranteed.

From the computational point of view, problems of convergence of redescending KM-smoothers are apparent if one use Newton type algorithms. In this case, the minimization of $P_N(g)$ would involve the Hessian matrix of second derivatives of $\rho(z)$, namely first derivative of $\psi(z)$. Thus, if $\psi(z)$ is non-monotone, its derivative may be negative, or does not exist, and so the numerical convergence fails. To avoid these problems it is advisable to solve (5) with direct search methods, or, iteratively, with pseudolinear algorithms described in Section 2.

Until now we have considered the behavior of KM-smoothers at points of continuity of $g(\cdot)$. At the jumps non-consistency should always be expected. Indeed, Chu *et al.* (1998) have shown that asymptotic bias (AB) and variance (AV) depend on the jump size δ, and resemble the moments of a Bernoulli probability function

$$|\mathrm{AB}| = \pi_\delta \delta\,, \quad \mathrm{AV} = \pi_\delta (1 - \pi_\delta)\, \delta^2\,, \quad \text{with} \quad \pi_\delta = \int_{\delta/2}^{\infty} f(\varepsilon)\, \mathrm{d}\varepsilon$$

where $f(\varepsilon)$ is assumed symmetric and $\psi(z)$ redescending. Therefore, as far as the jump signal is large compared to the noise variance, then statistical quality of KM-estimates improves. This means that consistency cannot be excluded at jump points provided that $f(\varepsilon)$ has a bounded support, with range less than δ.

Finally, let us discuss the property of $robustness$: It is known that monotone KM-estimators are robust against outliers (Härdle and Gasser, 1984), but we have checked they are not jump preserving. Thus, one may wonder if the two properties are incompatible, and conclude that redescending KM-smoothers could be non-robust. In reality, jumps in regression surfaces are typically represented by step-functions, whereas outliers correspond to pulses, and these are difficult to track. Now, if the bandwidth λ_3 is different from zero, then isolated pulses are ignored and robustness holds even for non-monotone ψ-functions. On the other hand, it should be noted that presence of outliers may hinder consistency, because KM-estimators asymptotically $(\lambda_3, N^{-1} \to 0)$ are maximizer for the noise density, but contaminated $f(\varepsilon)$ could not respect the hypothesis of unimodality in $A1$.

References

Chu C.K., Glad I., Godtliebsen F. and Marron J.S. (1998), Edge-Preserving Smoothers for Image Processing, *Journal of American Statistical Association*, 93, 526-541.

Freedman D.A. and Diaconis P. (1982), On Inconsistent M-Estima-tors, *The Annals of Statistics*, 10, 454-461.

Grillenzoni C. (1994), Optimal Recursive Estimation of Dynamic Models, *Journal of American Statistical Association*, 89, 777-787.

Hall P. and Jones M.C. (1990), Adaptive M-Estimation in Nonparametric Regression, *The Annals of Statistics*, 18, 1712-17-28.

Hall P. and Rau C. (2000), Tracking a Smooth Fault Line in a Response Surface, *The Annals of Statistics*, 28, 713-733.

Hampel F., Ronchetti E., Rousseeuw P. and Stahel W. (1986), *Robust Statistics: the Approach Based on Influence Functions*, Wiley, New York.

Härdle W. (1991), Smoothing Techniques: with Implementation in S, Springer, Berlin.

Härdle W. and Gasser T. (1984), Robust Non-parametric Function Fitting, *Journal of Royal Statistical Society, ser. B*, 46, 42-51.

Huber P.J. (1981), *Robust Statistics*, Wiley, New York.

Hwang R.-C. (2004), Local Polynomial M-smoothers in Nonparametric Regression, *Journal of Statistical Planning and Inference*, 126, 55-72. Jurečková

J. and Sen P.K. (1996), *Robust Statistical Procedures*, Wiley, New York.

Parzen E. (1962), On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics*, 33, 1065-1076.

Polzehl J. and Spokoiny V.G. (2000), Adaptive Weights Smoothing with Application to Image Restoration, *Journal of Royal Statistical Society, ser. B*, 62, 335-354.

Qiu P. (2002), A Nonparametric Procedure to Detect Jumps in Regression Surfaces, *Journal of Computational and Graphical Statistics*, 11, 799-822.

Scott D.W. (1992), *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley, New York.

Starck J.L., Murtagh F., Candes E.J. and Donoho D.L. (2003), Image contrast enhancement by curvelet transform, *IEEE Transactions on Image Processing*, 12, 706-717.

Wang M. and Tseng Y.-H. (2004), LDR Data Segmentation and Classification based on Octree Structure, available at: http://www.isprs.org/istanbul2004 /comm3/papers/286.pdf