Edge-Preserving in Generalized Smoothing of Signals and Images

V. Mottl, A. Kostin, A. Kopylov
Tula State University
Lenin Ave. 92, 300600 Tula, Russia
mottl@atm.tsu.tula.ru

Abstract

Under the generalized smoothing of a signal or image, a wider class of operations is understood than the plain suppression of noise. A variational approach to the generalized edge-preserving smoothing is considered on the basis of treating the sought-for result of processing as the minimum point of an objective function.

1. Introduction

An edge-preserving smoothing algorithm is inevitably nonlinear, even if the basic smoothing procedure is linear. Many nonlinear smoothing techniques are known to date [1], among which the leading place occupy, perhaps, median filters and new techniques based on competitive Kalman smoothing [2].

Despite the great number of edge-preserving smoothing techniques, all of them are underlaid by the tacit assumption that any two neighbouring edges to be revealed are spaced at a distance not smaller than the minimal effective width of the window sufficient for achieving the desired smoothing degree.

However, it is often required to distinguish between closely-spaced peculiarities of the original signal or image without loss in smoothing degree around them. The principal idea of the approach discussed here consists in finding break-points one by one and incorporating them into the model as they are found, so that at each step the most detectable of not yet legitimated peculiarities is sought for.

We will consider a signal or image as a numerical function y_t of one $t \in T$ or two variables $t = (t_1, t_2) \in T$ that takes values from an appropriate set $y \in \mathcal{Y}$ and is defined on the discrete axis of the variable forming the signal $T = \{t\} = \{1, \ldots, N\}$, usually, time, or on the image pixel grid $T = \{t = (t_1, t_2) \colon t_1 = 1, \ldots, N_1,\, t_2 = 1, \ldots, N_2\}$.

In many practical problems of signal and image analysis, it is expedient to treat the result of a processing of the original array $Y = (y_t, t \in T)$ more widely than the result of smoothing in the ordinary sense, namely, to interpret it as an assembly of estimates $\hat{x}_t$ on the current vector parameter $x_t \in \mathcal{X}$ at point t of an appropriate local data model of general kind. In this case, the entire totality of results should be considered as array of mutually coordinated estimates $\hat{X} = (\hat{x}_t,\, t \in T)$, smoothed in a generalized sense with preserving a relatively small number of abrupt jumps.

Just as in [3], we use here the variational approach to the generalized smoothing problem. The essence of the approach consists in that the algorithm of data analysis is created as that of minimization of a suitable separable objective function whose structure is determined by the undirected graph $G \subset T \times T$ of immediate neighborhood of data array elements. The objective function is sum of two kinds of elementary functions, namely, the node and edge ones:

$$J(X|Y) = \sum_{t \in T} \psi(x_t|Y) + \sum_{(t', t'') \in G} \gamma_{t't''}(x_{t'}, x_{t''}). \quad (1)$$

The data-dependent node functions $\psi(x_t|Y)$ are to be chosen with respect to the essence of the particular data processing problem so that each of them would take the greater value the more evident is the contradiction between the hypothetic value $x_t \in \mathcal{X}$ and a vicinity of point t in the data array.

Each of the model-based edge functions $\gamma_{t't''}(x_{t'}, x_{t''})$ is meant to impose an individual penalty upon the distinction of values in the respective pair of adjacent variables, i.e. unsmoothness of the resulting array at the edge (t', t'') of the neighborhood graph. The more rapidly the edge function grows as the discrepancy between $x_{t'}$ and $x_{t''}$ increases, the smoother will be the resulting process or field x_t at neighboring points (t', t''). In particular, if $\gamma_{t't''}(x_{t'} - x_{t''}) \equiv 0$ in some direction in $\mathcal{X}$, it means that the respective kind of jumps is legitimated by the model and will be completely preserved in the course of data processing.

The problem of finding the global minimum point $\hat{X}(Y)$ of a separable objective function allows for an especially simple algorithmical solution when the graph of variable adjacency has no cycles, i.e. is a tree.

Proceedings of the 14th International Conference on Pattern Recognition. Brisbane, Australia, August 16-20, 1998. Volume II, pp. 1579-1581.

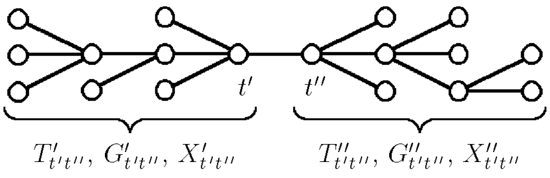

Figure 1. Two subtrees adjacent to an edge of a tree.

The global optimization procedure is based on a recurrent decomposition of the original objective function of many variables into a succession of intervening ones, called Bellman functions, each of only one variable [3], and in case of chain-like graph is nothing else than the classical dynamic programming procedure.

2. The principal idea of edge-preserving processing of a tree-like data array

The computational simplicity of the optimization procedures for objective functions with tree-like separability rests on the fact that any deletion of an edge $(t', t'') \in G$ turns tree G into two isolated ones $G'_{t't''}$ and $G''_{t't''}$ defined on nonintersecting sets of nodes $T'_{t't''}$ and $T''_{t't''}$, so that $T = T'_{t't''} \cup T''_{t't''}$ (Fig. 1). If $J'_{t't''}(X'_{t't''})$ and $J''_{t't''}(X''_{t't''})$ are respective partial objective functions, the full objective function (1) is representable as sum $J(X) = J'_{t't''}(X'_{t't''}) + \gamma_{t't''}(x_{t'}, x_{t''}) + J''_{t't''}(X''_{t't''})$. The functions

$$\tilde{J}'_{t't''}(x_{t'}) = \min J'_{t't''}(X'_{t't''}) \text{ over } x_s,\ s \in T'_{t't''},\ s \neq t',$$
$$\tilde{J}''_{t't''}(x_{t''}) = \min J'_{t't''}(X''_{t't''}) \text{ over } x_s,\ s \in T''_{t't''},\ s \neq t'',$$

defined at the two nodes of an edge, have the same sense as Bellman functions in the classical dynamic programming, and we will call them the left-side and the right-side Bellman function at edge (t', t''). These two kinds of Bellman functions can be recurrently found by the generalized dynamic programming procedure running twice through all the nodes of the tree [3].

The fundamental property of the Bellman functions is the fact that

$$\min_X J(X) = \min_{x_{t'}, x_{t''}} \{\tilde{J}'_{t't''}(x_{t'}) + \gamma_{t't''}(x_{t'}, x_{t''}) + \tilde{J}''_{t't''}(x_{t''})\},$$

so, the optimal values $(\hat{x}_{t'}, \hat{x}_{t''}) = \arg\min\{\tilde{J}'_{t't''}(x_{t'}) + \gamma_{t't''}(x_{t'}, x_{t''}) + \tilde{J}''_{t't''}(x_{t''})\}$ are just the sought-for result of processing. It is clear that the more indifferent the edge function $\gamma_{t't''}(x_{t'}, x_{t''})$ is, i.e. the less it differs from zero with increase in the discrepancy between $x_{t'}$ and $x_{t''}$, the less will be the minimum value of the objective function $\min J(X)$. But the degree of such a sensitivity must be different for different pairs of adjacent variables, and the greater it is, the more reason for the supposition that this pair is one of the not yet found hidden jumps in the data array.

Let $\gamma_{t't''}(x_{t'}, x_{t''})$ be the basic sufficiently severe edge function providing the required normal degree of smoothing at the edge (t', t'') where no jump is legitimated as yet, and $\gamma^*_{t't''}(x_{t'}, x_{t''})$ be a milder one considered as its alternative. The sufficiently distant local maxima of the ratio

$$z_{t't''} = \frac{\min\limits_{x_{t'}, x_{t''}} \left\{\tilde{J}'_{t't''}(x_{t'}) + \gamma_{t't''}(x_{t'}, x_{t''}) + \tilde{J}''_{t't''}(x_{t''})\right\}}{\min\limits_{x_{t'}, x_{t''}} \left\{\tilde{J}'_{t't''}(x_{t'}) + \gamma^*_{t't''}(x_{t'}, x_{t''}) + \tilde{J}''_{t't''}(x_{t''})\right\}}, \quad (2)$$

which exceed a preset threshold $z_{t't''} > h$, should be considered as potential hidden jumps. Let $z_{t't''}$ be assigned the value (2) if the ratio exceeds the threshold and zero otherwise. We will call this non-negative variable the strain excess at edge (t', t'').

Let us, sequentially for all the edges (t', t''), calculate the strain excess $z_{t't''}$. If there are edges with positive strain, those of them are to be pronounced the new legitimate jumps $(t', t'')^*$ at which the strain excess achieves its absolute positive maximum simultaneously in the intervals of a preset length Δ in all the branches of the two subtrees immediately adjacent to this edge (Fig. 1). It means that the edge functions $\gamma_{t't''}(x_{t'}, x_{t''})$ are to be replaced at edges $(t', t'')^*$ by more tolerant functions $\gamma^*_{t't''}(x_{t'}, x_{t''})$.

The procedure starts with the initial edge functions which presuppose no jumps in the data array, and is to be repeated again and again, each time with the renewed edge function set, while there is at least one point with non-zero strain excess. Such a procedure is able to find jump-points located closer to each other than Δ , but it detects them sequentially, at different stages of processing.

3. Edge-preserving signal processing

The neighborhood relation between elements of a signal is completely expressed by a chain-like graph, i.e. tree without branching nodes. We consider here usual smoothing and constrained time-frequency analysis as two examples of edge-preserving signal processing.

In edge-preserving smoothing, we keep to the approach depicted in [2]. In our terms, that approach consists in treating the hidden process in the generalized model as vector $x_t = (x_t^0, \ldots, x_t^n)^T \in \mathcal{X} = \mathsf{R}^{n+1}$, whose elements are the hidden signal itself $x_t^0 = x_t$ and its differences $x_t^i = \nabla^i x_t$ up to order n. We used the first-order model $n = 1$ with some more sophisticated variety of sought-for discontinuities than it was concisely described in the previous Section. Namely, the algorithm distinguished between jumps in the signal itself (ruptures) and those in the signal's first difference

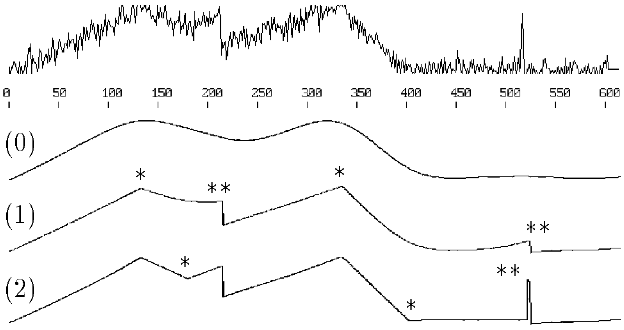

Figure 2. Result of continuous smoothing of a simulated signal (0) and two steps (1) and (2) of finding essential discontinuities: fractures * and ruptures **.

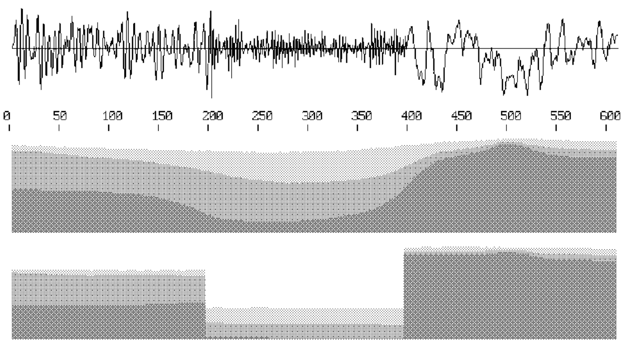

Figure 3. Results of continuous and edge-preserving time-frequency analysis of a simulated signal.

(fractures). The node and edge functions forming the objective function (1) were taken quadratic.

Fig. 2 shows an example of stepwise finding jumps in the hidden process with minimal interval between simultaneously registered jumps $\Delta = 100$. No more than two steps turned out to be required for finding all the discernible discontinuities in the original signal.

The kind of time-frequency analysis illustrated in Fig. 3 is called constrained because the instantaneous spectra are calculated as secondary result of nonstationary edge-preserving autoregression analysis [3]. As in the former example, we used quadratic node and edge functions, but only one kind of jumps was registered by the algorithm. The results of time-frequency analysis of a simulated signal with changing spectral properties are shown in an abridged form as diagrams of partial signal variances estimated in three adjacent frequency bands (low, medium and high). The spectral densities were calculated through the autoregression of the second order. Both jumps in the spectral properties were found on the first run with $\Delta = 100$.

4. Edge-preserving smoothing procedure for images

In contrast to the chain-like graph in the problems of signal analysis, the pixel-neighborhood lattice supporting the objective function of the generalized image processing problem is no tree, and the principal technique of generalized edge-preserving smoothing considered in Section 2 cannot be immediately applied to images.

The question of how a lattice can be replaced by a combination of trees is discussed in [3], but once it is done, the edge-preserving optimization procedure can be applied to each tree. The respective technique is based on the principal idea outlined in Section 2 which requires only a slight adjustment to the case of images.

For the simplicity sake, we consider here only the problem of edge-preserving elementary smoothing of images. An example of a dirty image along with results of its zero-order smoothing and of two steps of edge preserving are shown in Fig. 4. To find several jump points of the image plane at once, we used the half-length of the search interval Δ approximately equal to one-half the side of the bright and dark squares. The vertices of the inclined squares remained almost uncropped despite the fact that their boundaries contain jump-points located much closer to each other than Δ.

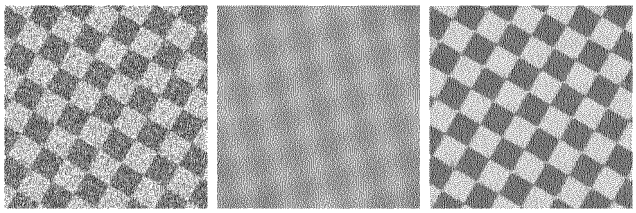

Figure 4. A dirty image and results of usual and edge-preserving smoothing.

References

[1] I. Pitas, and A.N. Venetsanopoulos. *Nonlinear Digital Filters. Principles and Applications.* Boston: Kluwer Academic Publishers, 1990.

[2] M. Niedzwiecki, and W.A. Sethares. Smoothing of discontinuous signals: The competitive approach. *IEEE Trans. on Signal Processing,* Vol. 43, No. 1, January 1995, pp. 1-13.

[3] V. Mottl, A. Blinov, and A. Kostin. Variational methods in signal and image analysis. *14th ICPR,* in this volume.