JOURNAL OF APPROXIMATION THEORY 58, 151–163 (1989)

Idempotent One-Sided Approximation of Median Smoothers

C. H. ROHWER

Institute for Maritime Technology,
P.O. Box 181, Simon's Town 7995, Republic of South Africa

Communicated by G. Meinardus

Received November 12, 1986

There has been considerable interest recently in nonlinear smoothing algorithms for handling outliers in time series. The simpler algorithms are predominantly chosen to be rank-based selectors, and concatenations of these. Analytical investigation lags and is considered difficult, so that theoretical understanding seems inadequate. It may be that a lattice is a useful structure for investigating and comparing smoothers, and a few related simple selectros are defined and investigated for this purpose. © 1989 Academic Press, Inc.

INTRODUCTION

Progressively more publications in the last decade have reported on the use and on investigations into the behaviour of some popular smoothers. Examples are [1–8]. Mallows [3] lists some of these, including the 5-point and the 3-point running medians, as well as some iterates and/or combinations of these. The main motivation is the treatment of "impulsive" noise or "outliers" in time series, which cannot be adequately treated by linear filters. (A definition of impulsive noise is provisionally avoided.)

Mallows has proposed a framework for studying and comparing smoothers, but concludes that the behaviour of the smoothers is far from clear and theoretical development is slow and analytical results hardly forthcoming. A different approach can augment this framework and may clarify some ideas. The basic framework is developed as follows.

Let x be a doubly infinite sequence of real numbers of X, the vector space of all such sequences with the obvious addition and scalar multiplication. A smoother can be defined as any operator/algorithm mapping X into X, satisfying a convenient set of axioms. Mallows chose the following:

A1. S is stationary, $S(Ex) = E(Sx)$, for E the shift operator.

A2. S is location invariant, $S(x + c) = S(x) + c$, c a constant.

640 58·2-3

0021-9045/89 $3.00

A3. S is centered, $S0 = 0$, where 0 is the null-sequence.

A4. S is local; the span of S, sp(S), if finite.

A5. Sx has finite variance at i.

A further axiom to make S scale independent is given as:

A6. $S(\gamma x) = \gamma Sx$, for any constant γ.

This axiom is sensible but unduly restrictive since it need only be required for $\gamma > 0$.

The basic results of Mallows are very interesting [4]. Under some very simple assumptions on the series, every smoother has a unique "linear component," whose coefficients can be calculated very easily in the case of rank-based selectors. Further results suggest an approach to the design of a nonlinear smoother when the process to be modelled is a Gaussian process with "additive noise". Roughly speaking, some selector can be chosen to remove outliers, and then this is followed by a choice of linear filter to augment the "linear component" of the selector for the removal of "better behaved" noise.

Popular smoothers (often concatenated with linear smoothers) are (using the Tukey–Mallows notation temporarily), the 5-point running median "5," the 3-point running median "3" and powers/combinations of these such as "53" ("5" followed by "3") and "3R," the limit of "3" iterated to convergence [2, 3]. The usual assumption is that "3R" exists since a sequence usually converges after a few iterations. The sequence $x(i) = (-1)^i$ is, however, only given a phase shift by the running median, with no convergence, and even sequences that are finitely square summable may give arbitrarily slow convergence.

An intuitive feeling arises that these smoothers and other selectors can be studied and compared using *partial ordering* of operators in a lattice. For this purpose *unsymmetric* idempotent operators may be useful. There seems to be no generally known investigation of this idea.

The Basic Unsymmetric Smoothers

Going back to the basic problem, a sequence that is constant with an occasional *upward* noise pulse can be considered. A primitive procedure for removal would be to apply a running minimum to the sequence. [An element is replaced by the minimum of its $2n+1$ nearest neighbours.] This algorithm will certainly remove the pulses except in the event of more than $2n$ adjacent pulses. The method seems to be computationally efficient and well defined, but it is unsatisfactory as it will widen an occasional downward pulse and an upward trend in a sequence will be retarded. This shortcoming can be overcome by following the procedure by a running maximum. This simple idea leads to the following development.

Let a partial ordering in X be defined in the usual pointwise way. Define a partial ordering on the set of operators from X to X in the usual way.

DEFINITION. $Q \geqslant S$ if $Qx \geqslant Sx$, for each $x \in X$.

DEFINITION. S is syntone if $x \geqslant y$ implies that $Sx \geqslant Sy$.

DEFINITION. Let $x \in X$ and $X(s, t) = \{x(i); i \in [s, t]\}$. Then

$$Lx = Lnx = \{y(i) = \max\{\min X(i-n, i), ..., \min X(i, i+n)\}\}.$$

$$Ux = Unx = \{y(i) = \min\{\max X(i-n, i), ..., \max X(i, i+n)\}\}.$$

$$Mx = Mnx = \{y(i) = \text{median}\{X(i-n, i+n)\}\}.$$

A few simple theorems will clarify some aspects of the behaviour of these selectors.

THEOREM 1. *L, U, and M are syntone.*

Proof. The proof is trivial but is included as an introduction to the type of reasoning. Suppose $x \leqslant y$ but $(Mx)(i) > (My)(i)$. Then $n+1$ elements of $X(i-n, i+n)$ are each larger than all of $n+1$ elements of $Y(i-n, i+n)$ which in turn are not smaller than the corresponding $n+1$ elements of x. This is impossible as there are only $2n+1$ elements in the set $X(i-n, i+n)$.

For U and L the proof is even more trivial.

THEOREM. *If $m > n$ then $Um \geqslant Un$ and $Lm \leqslant Ln$.*

Proof. Suppose

$$\begin{aligned} Lmx(i) &= \max\{\min X(i-m, i), ..., \min X(i, i+m)\} \\ &> \max\{\min X(i-n, i), ..., \min X(i, i+n)\} = Lnx(i). \end{aligned}$$

Then one of the sets, say $X(i-m+j, i+j)$, has a minimum which is larger than all the minima of the sets $X(i-n, i), ..., X(i, i+n)$. This is a contradiction since one of these sets is contained in $X(i-m+j, i+j)$. A similar proof holds for U.

The following lemmas are useful for simplifying further proofs.

LEMMA 1. *If* $\gamma x = \{\gamma x(i)\}$ *defines the usual scalar multiplication, then*

(i) $Lx = -U(-x)$,

(ii) $\gamma Lx = L(\gamma x)$ *if* $\gamma \geqslant 0$.

The proof is easy.

LEMMA 2. *If* $q \in [s, t]$ *and* $|s-t| \leqslant n+1$ *then*

$$Lx(s) < Lx(q) \quad \textit{implies } Lx(t) \geqslant Lx(q)$$

$$Ux(s) > Ux(q) \quad \textit{implies } Ux(t) \leqslant Ux(q).$$

Proof. Suppose $Lx(s) < Lx(q)$ and $Lx(t) < Lx(q)$. Then for each $j \in [s, t+n]$,

$$\max\{\min X(q-n, q), \ldots, \min X(q, q+n)\} > \min X(j-n, j).$$

Since $[q, q+n]$ is a subset of $[s, t+n]$ this is a contradiction. The rest follows similarly or by lemma 1.

LEMMA 3.

(i) *If* $\min X(j-n, j) \geqslant x(i)$ *for some* $j \in [i, i+n]$, *then* $Lx(i) = x(i)$.

(ii) *If* $\max X(j-n, j) \leqslant x(i)$ *for some* $j \in [i, i+n]$, *then* $Ux(i) = x(i)$.

Proof.

$$Lx(i) = \max\{\min X(i-n, i), \ldots, \min X(j-n, j), \ldots, \min X(i+n)\}$$

$= x(i)$ since one of the minima is $x(i)$ and the others are not larger than $x(i)$.

The rest of the proof follows equally simply.

LEMMA 4. $L \leqslant I \leqslant U$, *where* I *is the identity operator.*

Proof. $Ux(i) = \min\{\max X(i-n, i), \ldots, \max X(i, i+n)\} \geqslant x(i)$ since each of the maxima is not less than $x(i)$. (Or $I = U = L$ if $n = 0$ in Theorem 2.)

The lemmas establish the basic behaviour of the operators L and U. They show that many portions of a sequence are preserved whereas sharp upward peaks are removed by L, and sharp downward peaks are removed by U. The next theorem shows that L and U are "trend-preserving."

DEFINITION. A sequence x is called n-monotone if any n adjacent elements of the sequence are monotone.

THEOREM 3. *$Lx = Ux = x$ if and only if x is $(n+1)$-monotone.*

Proof. Let $Lx = Ux = x$. Suppose $x(q)$ is the first element of $X(i-n, i+1)$ to differ from $x(i-n)$, and suppose $x(q) > x(i-n)$.

Then $Lx(q) > Lx(i-n)$ and by Lemma 2 it follows that $Lx(q+1) \geqslant Lx(q)$, and therefore $x(q+1) \geqslant x(q)$. Since now $x(q+1) > x(i-n)$, repetition of the argument proves that each element of $X(i-n, i+1)$ is not less than its predecessor. If the first element to differ from $x(i-n)$ is less than $x(i-n)$ a similar argument proves that $X(i-n, i+1)$ is monotone decreasing. Conversely, let x be $(n+2)$-monotone. If $X(i-n, i+n)$ is monotone the proof is simple. Suppose therefore that $X(j-1, j+n)$ and $X(j, j+n+1)$ are two subsets with one monotone increasing and the other decreasing. Then $X(j, j+n)$ is a constant set containing $x(i)$ and $Ux(i) = Lx(i) = x(i)$ by Lemma 3.

THEOREM 4. *If $m \geqslant n$ then $UnUm = Um$ and $LnLm = Lm$.*

Proof. Assume $LnLmx(j) < Lmx(j)$. If $K(i) = \min\{Lm(i), ..., Lm(i+n)\}$ then $LnLmx(j) = \max\{K(j-n), ..., K(j)\}$ and therefore

$$K(j-n), ..., K(j) < Lm(j).$$

But then j is in an interval $[i, k]$ such that $|i-k| \leqslant n$ and $Lnx(i), Lnx(k) < Lmx(j)$, and by Theorem 2 it follows that $Lmx(i), Lmx(k) < Lmx(j)$. This contradicts Lemma 2.

COROLLARY. *L and U are idempotent.*

At this stage a comparison with the well-known median smoother is interesting. The following lemma is taken as self-evident.

LEMMA 5. *Let P be any set containing $2n+1$ elements. Then*

$$\text{Median } P = \min\{\max H;\ H \textit{ any subset of } n+1 \textit{ elements of } P\}$$
$$= \max\{\min H;\ H \textit{ any subset of } n+1 \textit{ elements of } P\}.$$

THEOREM 5. *$Ln \leqslant Mm \leqslant Un$ for $m \leqslant n$.*

Proof.

$$Mmx(i) = \max\{\min H;\ H \text{ any subset of } m+1 \text{ elements of } X(i-m, i+m)\}.$$

However, $Lmx(i) = \max\{\min X(i-m, i), ..., \min X(i, i+m)\}$ is the maximum of a subset which contains the minima of only *some* of the subsets containing $m+1$ elements, namely $X(j-m, j)$ with $i \leqslant j \leqslant i+m$. This implies that $Mmx(i) \geqslant Lmx(i)$.

Since $Lnx(i) \leqslant Lmx(i)$, by Theorem 2, the first half of the proof is completed and the second follows similarly, or by Lemma 1.

COROLLARY. *Any power of Mm will also be bounded by Un and Ln, since all the operators are syntone and Un and Ln are idempotent.*

FURTHER DEVELOPMENT

Roughly speaking the previous theorem tells us that any outlier in the upward direction will be removed by L if it is removed by the median of the same support, and any pulse in the downward direction will be removed as well by U if the median of the same span removes it. The computational effort evidently seems to favour L and Un.

There is a serious defect, however, in that L removes only pulses in the upward direction and U only those in the downward direction. Even if the outliers are expected on one side, there two obvious problems, in that additional "Gaussian noise" will "pull down" the average of L compared to the original sequence, and if an occasional outlier in the wrong direction occurs the usual problem still arises. The obvious step is to concatenate the operators U and L. Since they are not commutative there will be two operators to study, namely UL and LU.

LEMMA 6. *$LnUmLn = UmLn$ and $UnLmUn = LmUn$.*

Proof. Let $LnUmLnx(i) < UmLnx(i)$, for some sequence x. Then $UmLnx(s)$, $UmLnx(t) < UmLnx(i)$ for some $[s, t]$ containing i with $|s-t| \leqslant n$. But then, with $K(s-m) = \max\{Ln(s-m), ..., Ln(s)\}$

$$Lnx(q),\ Lnx(r) > \min\{K(s-m), ..., K(s)\},\ \min\{K(t-m), ..., K(t)\}$$
$$\text{for some } [q, r] \text{ containing } i \text{ with } |q-r| \leqslant m.$$

But this means that

$$Lnx(q),\ Lnx(r) > \{Lnx(w-m), ..., Lnx(w)\},\ \{Lnx(v-m), ..., Lnx(v)\}$$
$$\text{for some } w \in [s, s+m] \text{ and } v \in [t, t+m].$$

$Lnx(s)$ and $Lnx(t)$, respectively, in the two RHS sets, and $|q-p| \leqslant m$ together imply that $[q, r]$ is contained in $[s, t]$. This is impossible since then $Lnx(q) > Lnx(s)$, $Lnx(t)$. This contradicts Lemma 2. Therefore $LnUmLn \geqslant UmLn$. But $Ln(UmLn) \leqslant UmLn$, since $Ln \leqslant I$. The rest follows similarly or by Lemma 1.

THEOREM 6. *UmLn and LmUn are idempotent.*

Proof.

$$\begin{aligned}(UmLn)\,UmLn &= Um(LnUmLn)\\ &= Um(UmLn), \qquad \text{by Lemma 6.}\\ &= (UmUm)\,Ln = UmLn, \qquad \text{since } Um \text{ is idempotent.}\end{aligned}$$

THEOREM 7. $LnUm \geqslant UmLn$.

Proof.

$$\begin{aligned}LnUm &\geqslant Ln(UmLn), \qquad \text{since } Um \geqslant UmLn\\ &= UmLn, \qquad \text{by Lemma 6.}\end{aligned}$$

THEOREM 8. $LnUn \geqslant Mn \geqslant UnLn$.

Proof. Assume that $LnUnx(i) < Mnx(i)$ for some sequence x.

Then $Unx(s)$, $Unx(t) < Mnx(i)$ for some $[s, t]$ containing i, with $|s-t| \leqslant n$. But this means that

$$\max\{x(j-n), ..., x(j)\}, \qquad \max\{x(q-n), ..., x(q)\} < Mn(i)$$
$$\text{for some } j \in [s, s+n] \text{ and } q \in [t, t+n].$$

Of the set $\{x(j-n), ..., x(q)\}$, which includes $x(s)$ and $x(t)$, at least $n+1$ are in the set $X(i-n, i+n)$. But no more than n elements of this set can be less than the median, $Mnx(i)$. This is a contradiction. [The other half of the proof is similar or can be proved from the first part using Lemma 1.]

Remark. Unlike the case with the inclusion theorem for Un and Ln, the statement $LnUn \geqslant Mm \geqslant UnLn$ is not true, for general $m < n$.

At this stage it is prudent to pause and take stock of what has been achieved. A brief excursion into a philosophical basis of nonlinear smoothers should be allowed.

Ignoring ordinary Gaussian noise provisionally, a constant signal can be considered. This constant signal should preferably be passed unaltered by any smoother [Axiom A2]. If occasionally a narrow pulse of unreasonably large amplitude can be expected, any method of automated removal using only order can only distinguish such impulsive noise from an impulsive upward signal by reason of its briefness. The method must remain computationally simple, fast, and predictable. The axioms listed previously sum up reasonable properties. The two smoothers *LnUn* and *UnLn* remove impulsive noise in both directions, are idempotent, and bound the median *Mn*. The output of these smoothers can be vastly different, however, as will be illustrated by later examples. Mn^2 seems generally to be considered better than Mn and, despite the computational burden, the median is sometimes repeated till no futher change is evident. Why should Mn^2 be better than Mn? Clearly the "linear" component of the smoothers will improve the filtering of Gaussian noise with repetition, but that was not the purpose of the smoother. Is there a natural systematic method of designing a smoother for a given requirement, as is done in the design of linear filters? Why does the simplest and most logical requirement on a smoother lead almost inevitably to an unsymmetric pair? This should be more than a mere quirk of logic, and it seems that there is a fundamental underlying ambiguity.

The fundamental problem that cannot easily be circumvented can be illustrated in its simplest form as the following. Suppose the signal x is sampled at integer values of some time scale to give a row $\{x_i\}$. Suppose further that a square pulse of noise, not exceeding a width of n samplings, is expected, yet square pulse signals of width $n+1$ should pass unaltered. This can for instance be accomplished by the smoothers $UnLn$, $LnUn$, and Mn. But what should happen to a signal consisting of two upward pulses of duration n, separated by a gap of n, as represented by the signal in Fig. 1? Is this to be interpreted as two upward noise pulses of width n, or as a signal pulse of width $3n$ with a downward noise pulse of width n superimposed? This question is fundamentally undecidable on consideration of relative order alone.

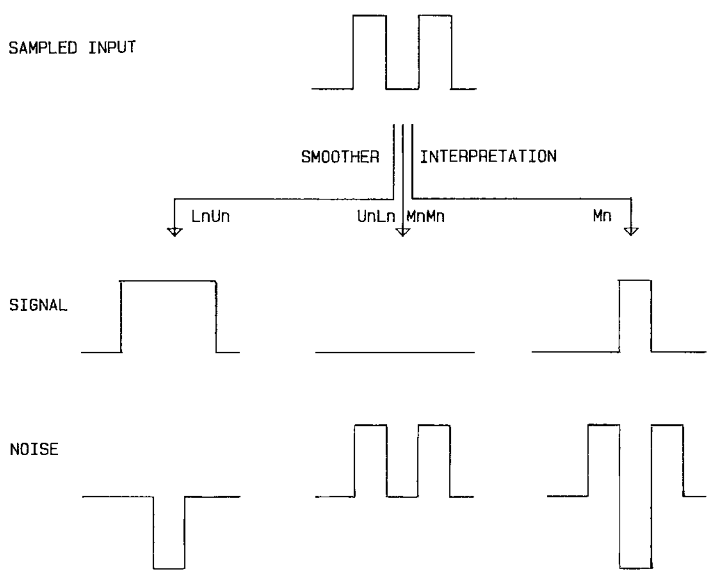

FIG. 1. Two interpretations of the same input as smoothed by Mn, Mn^2, $UnLn$, and $LnUn$.

UL and L will choose the first interpretation and LU and U the second option, thus precisely bounding this logically undecidable interval. The median smoother will, in the face of the paradox, falter and choose neither of the two simplest options. If the median is applied twice it will agree with the first interpretation, thus rejecting its own initial interpretation.

A heuristic reason for further preferring the interpretations of LU and UL is that, in the calculation of the smoothed value at i, the median operator Mn discards all information contained in the index order of the set $X(i-n, i+n)$. (The median selects the minimum of the maxima of *all* subsets of $n+1$ elements, whereas Un and Ln select minima and maxima from the sets of $n+1$ adjoining elements only.)

Commutation with Powers of Medians

The operators LU and UL do exactly what they were intended to do, and their lack of symmetry should be seen as natural, indicating an interval of fundamental uncertainty associated with a concept of impulsive noise. The difference between LU and UL can also be expected to indicate the amount of Gaussian noise present in the signal, and their avarage should be an unbiased estimator of the signal. The simplest median smoother for the same purpose, however, is inconsistent and *cannot do better than LU and UL, in the precise sense that Mn does not modify any sequence that has been smoothed* by $UnLn$ or $LnUn$, as is shown in the following theorem.

THEOREM 9. *$MmUn = LmUn$ and $MmLn = UmLn$ for $n \geqslant m$.*

Proof. $MmUn \leqslant (LmUm)\,Un$, by the previous theorem. But then, $MmUn \leqslant (LmUm)\,Un = Lm(UmUn) = LmUn$, by Theorem 4.

By Theorem 5, however, we get $MmUn \geqslant LmUn$, which concludes the first half of the proof. The other half is similar or follows from Lemma 1, as usual.

COROLLARY. *$MmMmUn = MmUn$ and $MmMmLn = MmLn$.*

The behaviour of the median smoother Mn *after* $UnLn$ or $LnUn$ differs from the behaviour *before* smoothing by $UnLn$ or $LnUn$. The reason for this is the type of sequence in the Range of the operators $UnLn$ and $LnUn$. [The singular is used as the two operators clearly have the same Range.] The following theorem indicates to what degree higher powers of Mn are "better" smoothers than Mn itself.

THEOREM 10. *$UmMnMn \leqslant MnUnMn$ and $LmMnMn \geqslant MnLnMn$, if $m \leqslant n$.*

Proof.

$$\begin{aligned} Um(Mn)\,Mn &\leqslant Um(LnUn)\,Mn \\ &= LnUnMn, \qquad \text{since } m \leqslant n \\ &= MnUnMn, \qquad \text{by the previous theorem.} \end{aligned}$$

COROLLARY. *$UnMnMn \leqslant UnMn$ and $LnMnMn \geqslant LnMn$.*

COROLLARY. *$UnMn \geqslant (UnMn)\,UnMn \geqslant UnMnMn$, $LnMn \leqslant (LnMn)\,LnMn \leqslant LnMnMn$.*

The smoother *UnMn* is not idempotent. However, its third power is equal to its second power, ensuring convergence of powers of *UnMn* and *LnMn*.

THEOREM 11. *Let O be any power of Mn. Then UnO and LnO are idempotent on the range of UnO.*

Proof.

$$\begin{aligned} Un(OUn)O &= Un(LnUn)O, \qquad \text{by induction and Theorem 9} \\ &= (UnLnUn)0 = LnUnO, \qquad \text{by Lemma 6} \\ (UnO)(UnOUnO) &= (UnO)(LnUnO) \\ &= Un(OLnUn)O \\ &= Un(LnUn)O, \qquad \text{by Theorem 9} \\ &= (UnLnUn)O \\ &= LnUnO = MnUnO \\ &= (UnO)\,UnO = (UnO)^2. \end{aligned}$$

Although it is clear from the last two theorems that there are pairs of unsymmetric smoothers providing a narrower "interval of ambiguity" in which the signal should be estimated from the measurements, justification for any such algorithm is needed. Since a smoothed signal is generally followed by some linear filter, the quality of the "linear component" of a smoother should be of secondary importance compared to the exact specification of the smoother task.

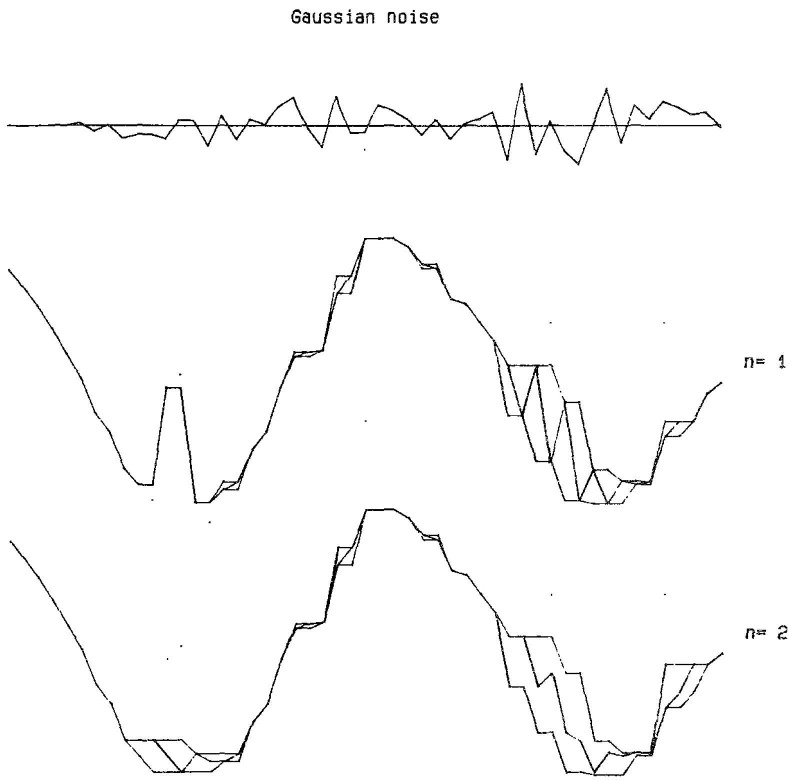

FIG. 2. The basic smoothers applied to a signal with impulsive and additive noise. The signal is sinusoidal with an increasing amount of Gaussian noise, and representative impulsive noise pulses.

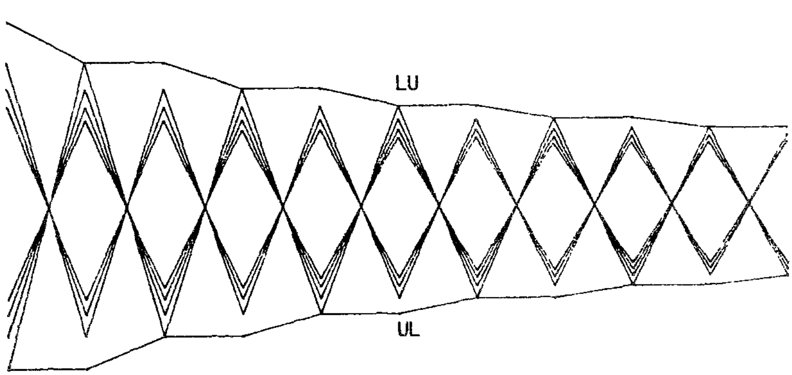

FIG. 3. The basic smoothers applied to the sequence $x(i) = (-1)^i i^{(-1)}$ demonstrating slow convergence of powers of median operators.

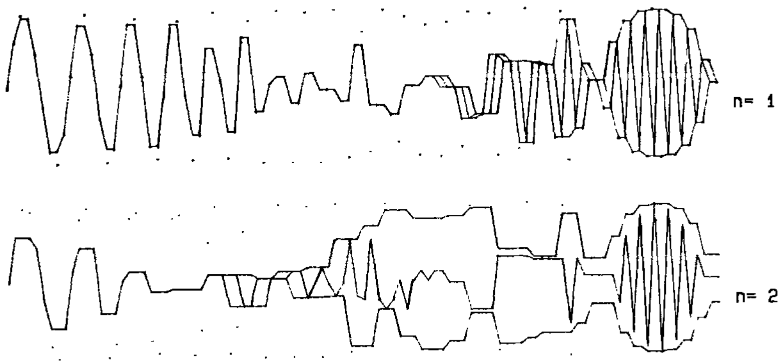

FIG. 4. An oscillation of a frequency increasing linearly to higher than sampling frequency, to illustrate the effect on the basic smoothers.

EXAMPLES

Figures 2–4 illustrate some of the properties of the different basic smoothers. The output of the median is the graph between the graphs of the other two outputs in each case.

CONCLUSION

The defined unsymmetric smoothers have to their advantage, over comparable conventional smoothers, the fact that they perform their task in a prescribed and predictable way. Any unreasonably large pulse of prescribed briefness is removed, and where two such pulses occur too close to each other, the fundamental ambiguity inherent in the prescribed task is neither ignored nor treated with indicision.

Furthermore, the smoothers LU and UL have some obvious and some subtle computational advantages.

1. The operators LU and UL can be calculated by successive application of running maxima and minima, a process that can easily be implemented in dedicated hardware.

2. Vector processors can calculate LU and UL simultaneously by processing x and $-x$ identically, and using the result of Lemma 1; $-UL(-x) = -U(-Ux) = LU(x)$.

3. Vector processors can also exploit Lemma 1 by applying U to both x and to $-Ux$ (with a suitable index lag), in the calculation of $LU(x)$.

A subtle possible advantage is the empirically discovered property that recursive applications of U and L do not alter their outcome. Alternatively $Ux(i) = Ux^*(i)$, where

$$x^*(j) = x(i) \qquad \text{for } j \geqslant i$$
$$= Ux(j) \qquad \text{for } j < i.$$

This surprising property, as well as the possibility of bounding LU and UL by running rank selectors, is being investigated.

Perhaps the most valuable contribution the smoothers LU and UL make in the study of nonlinear smoothers is the light they shed on the behaviour of median and related smoothers. In this respect they are much more than mere approximations of median smoothers.

References

1. L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, Application of a nonlinear smoothing algorithm to speech processing, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-23**, No. 6 (1975).
2. J. W. Tukey, "Exploratory Data Analysis," Addison-Wesley, Reading, MA, 1977.
3. C. L. Mallows, Some theory of nonlinear smoothers, *Ann. Statist.* **8**, No. 4 (1980), 695–715.
4. C. L. Mallows, "Resistant Smoothing," Time Series [O. D. Anderson, Ed.] North-Holland, Amsterdam, 1980.
5. E. Ataman, V. K. Aatre, and K. M. Wong, A fast method for real-time median filtering, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-28**, No. 4 (1980).
6. E. Ataman, V. K. Aatre, and K. M. Wong, Some statistical properties of median filters, *IEEE Trans. Speech Signal Process.* **ASSP-29**, No. 5 (1981).
7. A. R. Butz, A class of rank order smoothers, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-34**, No. 1 (1986).
8. N. C. Callagher, A theoretical analysis of the properties of median filters, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-29**, No. 6 (1981).