On the Statistical Optimality of Locally Monotonic Regression

Alfredo Restrepo (Palacios) and Alan C. Bovik

Abstract—Locally monotonic regression is a recently proposed technique for the deterministic smoothing of finite-length discrete signals under the smoothing criterion of local monotonicity. Locally monotonic regression falls within a general framework for the processing of signals that may be characterized in three ways: regressions are given by projections that are determined by semi-invariants, the processed signals meet shape constraints that are defined at the local level, and the projections are optimal statistical estimates in the maximum likelihood sense. Here, we explore the relationship between the geometric and deterministic concept of projection onto (generally nonconvex) sets and the statistical concept of likelihood, with the object of characterizing projections under the family of p-semi-metrics as maximum likelihood estimates of signals contaminated with noise from a well-known family of exponential densities.

I. INTRODUCTION

We discuss here a statistical aspect of the concept of projection onto (generally nonconvex) sets of signals, or regression, as it was recently proposed [1], [2], for the processing of finite-length discrete signals. One may argue, particularly with optical images, that some signals carry their information explicitly as shape. A shape constraint is a property defined in the natural domain of the signal, e.g., time or space rather than frequency; we consider shape constraints that are defined at the local level rather than at the global level, e.g., local monotonicity versus (global) monotonicity.

The projections of a signal on a set of signals are defined as the signals in the set that are closest to the signal being projected, under a semi-metric for the space. Local monotonicity [3] is shape constraint for one-dimensional signals that provides a measure of the smoothness of a signal; it sets a limit on the roughness of a signal by limiting how often the signal may have a change of trend (increasing to decreasing or vice versa). In a sense of the word, it limits the frequency of the oscillations that a signal may have without making restrictions on the magnitude of the changes of the signal from each coordinate to the following one. Piecewise constancy, piecewise linearity, and local convex/concavity are other, similar shape constraints that have been explored [4]. Algorithms for the computation of the locally monotonic regression of a finite-length signal of length \( n \) have a complexity that is exponential [1]; however, much faster algorithms have been developed that employ regression only over a moving window (hence, computation is linear with signal length) [5], and that compute a fuzzy approximation to the true locally monotonic regression using a generalized deterministic annealing algorithm [6].

II. SHAPE CONSTRAINTS

An integer interval \([a, b]\), where \(a\) and \(b\) are integer numbers, is defined as the subset of the integers that are larger than or equal to \(a\) and smaller than or equal to \(b\). An \(n\)-point signal (or a discrete signal of length \( n \)) is a real function \( x \) having as domain a nonempty integer interval \( /a, b/\), where \(b - a = n - 1\). An \(n\)-point signal is a point of \( R^n \) and may be expressed as the \(n\)-tuple \(\{x_1, \ldots, x_n\}\) of the values it takes. The origin \(0, \ldots, 0\) of \( R^n \) is denoted as \( \theta \).

The slope skeleton of an \(n\)-point signal \( x = \{x_1, \ldots, x_n\} \) is the \((n-1)\)-point signal \( s = \{s_1, \ldots, s_{n-1}\} \) having components \( s_i = \text{sgn}(x_{i+1} - x_i) \) where \( \text{sgn} \) is the signum function that is, respectively, one, zero, and minus one, when its argument is positive, zero, or negative.

The segments of length \( r \) of a signal \( x: [1, n] \to R^1 \) \( r \leq n \) are the restrictions of \( v \) to integer intervals of length \( r \). For example, \([2, 3, 4]\) is a segment of length 3 of the signal \([1, 2, 3, 4]\).

Constancy and linearity are well-known shape constraints. Less commonly used shape constraints are monotonicity, convexity, concavity, piecewise constancy, piecewise linearity, local monotonicity, and local convex/concavity [4], [5].

A signal is constant if its slope skeleton is null: \( s = \theta \). A signal is said to be nondecreasing if the components of its slope skeleton are nonnegative and nonincreasing if they are nonpositive. A signal is monotonic if it is either nonincreasing or nondecreasing. A signal is said to be convex if its slope skeleton is nondecreasing and concave if its slope skeleton is nonincreasing. These are global shape constraints.

A signal is locally monotonic of degree \( \alpha \) (\alpha-monotone) if each of its segments of length \( \alpha \) is monotonic. A signal is locally convex/concave of degree \( \alpha \) (\alpha-convex/concave) if each of its segments of length \( \alpha \) has a monotonic slope skeleton. These are shape constraints defined at the local level.

A signal is linear if its slope skeleton is constant. The algebraic span of the constant signal \([1, 1, \ldots, 1]\) and the linear signal \([1, 2, \ldots, n]\) is the collection of the linear signals of length \( n \).

A signal may be segmented into longest constant segments in a unique way. For example, the longest constant segments in \([1, 3, 3, 2, 2, 2, 4, 5, 5]\) are \([1, 3, 3]\), \([2, 2, 2]\), \([4, 5, 5]\), and \([5, 5]\). A signal is said to be piecewise constant of degree \( \alpha \) (\alpha-piecewise) if, besides the first and last segments, the shortest segments of its segmentation into constant segments have length at least \( \alpha \).

Each coordinate point \( i \) of a signal \( s \) such that \( s_i - 2s_{i-1} + s_{i+1} \) is nonzero is a point where the slope of \( s \) changes and is called a hinge of \( s \). In addition, the first and last coordinates of a signal are called hinges; accordingly, a linear signal of length larger than one has exactly two hinges. A signal is said to be piecewise linear of degree \( \alpha \) (\alpha-piecewise) if the difference between each two consecutive hinges is at least \( \alpha \); any signal is \( \alpha \)-piecewise. An \( n \)-point signal may have as few as two hinges and as many as \( n \) hinges.

For example, the signal in Fig. 1(a) is \( \alpha \)-piecewise 4, \( \alpha \)-piecewise 7, \( \alpha \)-piecewise 3, and \( \alpha \)-piecewise 1. The signal in Fig. 1(b) is \( \alpha \)-piecewise 2, \( \alpha \)-piecewise 5, \( \alpha \)-piecewise 1, and \( \alpha \)-piecewise 3. All of these shape constraints are defined locally and are novel measures of the smoothness of a signal; the larger the degree \( \alpha \), the smoother the signal.

III. SEMI-METRICS AND LIKELIHOOD FUNCTIONS

A semi-metric differs from a metric in that it may lack the triangle inequality property. A semi-metric measures, in a coordinatewise way, the similarity of a pair of signals. A semi-metric for \( R^n \) is a function \( d: R^n \times R^n \to [0, \infty) \) that is positive definite and symmetric.
Fig. 1. Two discrete signals that satisfy some shape constraints.

that is,

1) \( \forall x, y \in \mathbb{R}^n, d(x, y) = 0 \Rightarrow x = y \)  
(positive definiteness)

2) \( \forall x, y \in \mathbb{R}^n, d(x, y) = d(y, x) \)  
(symmetry).

A large class of functions are semi-metrics for \( \mathbb{R}^n \); for measuring the similarity between signals, it is convenient to have a semi-metric that is translation-invariant:

3) \( \forall x, y, z \in \mathbb{R}^n, d(x + z, y + z) = d(x, y) \)  
(translation invariance).

To prove the existence of projections, it is also necessary that \( d \) be continuous (in the standard topologies of \( \mathbb{R}^n \) and \( \mathbb{R}^1 \)). It is also natural to require that the semi-metric be nondecreasing: for each signal \( y \) with nonnegative components and for each \( n \)-point signal \( x, d(\theta, x) \leq d(\theta, x + y) \). Many semi-metrics are positive homogeneous as well:

4) \( \forall x, y \in \mathbb{R}^n, \forall \theta \in \mathbb{R}^1, d(\gamma x, \gamma y) = |\gamma|d(x, y) \)  
(positive homogeneity).

A metric is a semi-metric that has the triangle-inequality property:

3) \( \forall x, y, z \in \mathbb{R}^n, d(x, z) \leq d(x, y) + d(y, z) \)  
(triangle inequality).

The distance between a point \( x \) and a set \( S \) is given by

\[
D(x, S) = \inf \{d(x, z) : z \in S\}.
\]

A. A Family of Semi-Metrics

A well-known collection of positive homogeneous and translation-invariant semi-metrics on \( \mathbb{R}^n \) that is indexed by the parameter \( p \in (0, \infty) \) is defined as

\[
d_p(x, y) = \left( \sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}, \quad p \in (0, \infty).
\]

For \( p \in [1, \infty), d_p \) is a metric; for \( p \in (0, 1), d_p \) is a semi-metric, but it is not a metric. This family includes the Euclidean metric \( d_2 \) and the square metric \( d_1 \). These \( p \)-semi-metrics are translation-invariant, nondecreasing, and continuous.

Assume that a shape constraint has been specified. Let \( Q \) be the truth-value function that is true on signals meeting the constraint and false otherwise, and let \( A = \{x \in \mathbb{R}^n : Q(x)\} \) be the set of signals having the required shape. Given a signal \( x \) that is not in \( A \). If \( A \) is a nonempty proper subset of \( \mathbb{R}^n \) and \( x \in (\mathbb{R}^n - A) \), the set

\[
P_A(x) = \{a \in A : d(x, a) = D_p(x, A)\}
\]

gives the set of the projections of \( x \) on \( A \) (or of regressions of \( x \) with respect to \( A \)) under the semi-metric \( d \). The existence and multiplicity of projections have been characterized in [1]. The distance between \( x \) and \( A \) is given by the smallest radius \( \rho \) for which the boundary of \( A \) \( \cap \) \( B(\rho, x) \) contains \( x \) (\( \rho \) stands for the operator of Minkowski set addition [1]). In Fig. 2, a signal and a corresponding locally monotonically increasing or locally convex/concave may be found in [4], [6].

B. A Family of Densities

We make use of a family of generalized exponential probability density functions that have been used extensively in robust statistics [8] and in the design of order statistic filters [9], [10]. It is defined as follows:

\[
f_p(x) = \gamma \exp(-\zeta|x|^p), \quad p \in (0, \infty), x \in \mathbb{R}^n
\]

where \( \gamma \) and \( \zeta \) are positive constants (that depend on \( p \)) that determine the variance of the random variable and ensure that each density \( f_p \) integrates to one. The Gaussian and Laplacian densities are in the family.

C. Likelihood Functions

Assume that a system outputs (deterministic) signals of length \( n \) that are characterized by a shape constraint \( Q \), and that any signal meeting the shape constraint may be an output of the system, with uniform probability on the set \( A = \{x \in \mathbb{R}^n : Q(x)\} \); signals not in \( A \) are outputs of the system with probability zero. Also, before the output signal \( t \) can be observed, it becomes contaminated with a random signal \( e \). Given an observed noisy signal \( z = t + e \), it is desired to obtain an estimate \( \hat{a} \) of \( t \). The likelihood \( L(z; a) \) of \( z \) being equal to a signal \( a \) in \( A \) plus a sample \( e \) from \( z \) depends both on \( a \) and \( z \). Given the distribution of the components of \( e \), the set...
regression [13] and isotonic regression [14, 15] are perhaps the most commonly used types of regression in statistics. Conditions of monotonicity arise naturally in certain classes of problems; for example, consider the collection of attention times of the elements of a queue [16]. Projections on the set of monotonic signals are called monotonic regressions. Isotonic regression and related topics have been considered previously for image processing [17].

Here, we have examined a signal-smoothing paradigm, designed under geometric and deterministic concepts of projection, that is optimal for signal estimation in the maximum likelihood sense. The use of semi-metrics in the projection apparatus, although not common in signal processing, provides estimators of signals contaminated with highly impulsive noise. A family of robust estimators, each optimal for a given density, is obtained. The shape constraints considered here provide criteria of smoothness that emphasize different characteristics of signals. We believe that the use of shape constraints that are defined at the local level, together with the concept of projection provide a powerful tool for the shaping, smoothing, and filtering of signals.

REFERENCES


IV. CONCLUSION

The technique of regression is widely used in statistics, but it is rarely used with the purpose of shaping or filtering signals. Linear
A Smoothing Property of the Median Filter

Alfredo Restrepo (Palacios) and Liliana Chacon

Abstract—The variation of a discrete signal is defined and used as a measure of roughness. It is shown that the median filter does not increase the variation of a signal, supporting the perception of the median filter as a smoother. Signals of both finite and infinite lengths are considered; for signals of finite length, nonpadding median filters are used.

I. INTRODUCTION

The median filter, introduced by Tukey as a running statistic for the analysis of time series [1], is a practical smoothing device, particularly for image processing. It is defined as a local and simple way: it is a sliding-window filter whose output component is the median of the windowed sample that is centered at coordinate i. The median filter is known for its ability to preserve edges and constant neighborhoods [2] and to “throw away isolated wild points” [3]. The root signals of the median filter, also known as invariant signals and as fixed points, have been studied in [4]–[7], for example.

In Section II, the variation of a signal is defined as the sum of the absolute values of the changes of value the signal makes in going from each coordinate point to the next one. Considering the variation of a signal as a measure of its roughness, the median filter is a smoother in the sense that it never increases the variation of a signal, and quite often decreases it. This result is proven in Section III for both finite-length and infinite-length signals. Unlike [2], for finite-length signals, nonpadding median filters are used.

The median filter has been observed to possess low-pass characteristics when filtering white noise [8]–[10]; nevertheless, it is not a convolution filter and cannot be suitably represented in the Fourier frequency domain. On the other hand, since the sample median function is not differentiable, it does not have a Taylor series expansion which makes it difficult to obtain a Voiterra series representation for the filter. With the exception of the results about its root signals, and a few others, the median filter has been characterized locally but not globally. The smoothing property presented here provides a global characterization of the median filter.

II. NOTATION AND DEFINITIONS

Let Z and R respectively denote the set of the integer numbers and the set of the real numbers. Let n be a positive natural number; a signal of length n is a point of R^n; a signal of infinite length is a double-sorted sequence, that is, a point of R^Z. An integer interval [a, b], where a and b are integer numbers, is defined as the set of integers larger than or equal to a and smaller than or equal to b.

Let n be an integer number larger than or equal to two, and let r = [r_1, r_2, ..., r_n] be a point in R^n; let (1), (2), ... (n) denote a permutation of {1, 2, ..., n} that makes the list (r_1), (r_2), ..., (r_n) ordered: (r_1) ≤ (r_2) ≤ ... ≤ (r_n); the (r_i)'s are called the order statistics [11], or ranked values, of r. If n is an even number, (r_{n/2}) and (r_{n/2+1}) are called the central order statistics of r. If n is odd, (r_{(n+1)/2}) is the median of r.

Let u = 2k + 1 where k is an integer number larger than or equal to 1, let m : R^u → R be the function that maps each point r = [r_1, r_2, ..., r_u] of R^u to the median m(r) = r_{(k+1)} of its components; m is known as a sample median function. m is not differentiable, in particular, at the origin of R^u. A median filter of window size w may be described in the following way: if x and y are its input and output signals, respectively, then for each output coordinate i, y_i = m(x_{i-k}, x_{i-k+1}, x_{i-1}, x_i, x_{i+1}, x_{i+k}). If x = [x_i] is of finite length then so is y and y_i is defined for integer i; for infinite-length signals, the median filter is a function from R^Z into R^Z. If x = [x_1, x_2, ..., x_N] is of finite length N then the length of y = [y_1, y_2, ..., y_{N−k}] is N − 2k and y_i is defined for i ∈ [k+1, N−k]/k; thus, for finite length signals, the median filter is a function from R^N into R^{N−2k} and this defines a nonpadding filter. We consider that the pre-padding of the input signal udually (and in many instances unnecessarily) overemphasizes its two extreme points.

If x = [x_1, x_2, ..., x_N] is a signal of finite length N ≥ 2, its variation is given by \( \text{var}(x) = \sum_{i=1}^{N} |x_i - x_{i+1}| \).

A signal of length one is said to have zero variation. If x = [x_n] is a signal of infinite length, its variation is given by \( \text{var}(x) = \sum_{n=-\infty}^{\infty} |x_n - x_{n+1}| \) where this series of nonnegative terms is the limit when N tends to infinity of \( \sum_{n=-N}^{N} |x_n - x_{n+1}| \); if the limit is a real number, x is said to be of bounded variation; otherwise, it is said to be of unbounded variation. Note that an infinite-length signal that is in \( L_1 \), has bounded variation but a signal of bounded variation may not be in \( L_1 \).

If x is a signal, its slope skeleton \( s \) is the \{−1, 0, 1\}−valued signal having as components \( s_n = \text{sgn}(x_n - x_{n-1}) \) where \( \text{sgn} \) is the signum function given by \( \text{sgn}(t) = \begin{cases} -1 & \text{if } t < 0; \\ 0 & \text{if } t = 0; \\ 1 & \text{if } t > 0. \end{cases} \) If x is of finite length \( N \), its slope skeleton has length \( N-1 \). If x is of infinite length, so is its slope skeleton.

A signal is said to have alternating slope skeleton if the product of any two consecutive components of its skeleton is negative: \( s_n s_{n+1} = -1 \).

III. THE MEDIAN FILTER DOES NOT INCREASE THE VARIATION OF A SIGNAL

We present the main results of the paper in this section. Lemma 1 is a preliminary result that is used several times throughout the rest of the paper. Theorem 1 states that when a finite-length signal is filtered, its variation may not be increased; Corollary 2 states the corresponding result for infinite-length signals. In Section IV, two particular cases where stronger results hold, are considered.

Lemma 1: Let k be a positive integer, let \( x_1 = [x_1, x_2, ..., x_{2k+1}] \) ∈ R^{2k+1} and \( z = [z_1, z_2, ..., z_{2k+2}] \) ∈ R^{2k+2} be two overlapping samples. Let m be the sample median function of window size \( 2k+1 \) and let y = m(z_1) and u = m(z_2). Then, \( y < u \) implies \( x_1 \leq y \leq u \leq x_{2k+2} \) and \( y > u \) implies \( x_1 \geq y \) and \( y \geq u \). Moreover, if \( y = u \) then no component of x belongs to the open interval \( (\text{min}\{y, u\}, \text{max}\{y, u\}) \).

Proof: Let \( z_1, z_2, y, \) and u be as stated. Let \( x = [x_1, x_2, ..., x_{k+1}, x_{k+2}] \) ∈ R^{2k+2} and consider the order statistics \( x_1, x_2, ..., x_{2k+2} \) of x; x is of even length \( 2k+2 \) and \( x_{k+1} \) and \( x_{k+2} \) are its central-order statistics. First, we show that \( \{x_{1+k}, x_{2+k+1}\} \subset \{y, u\} \). Let s be a subsignal of x of length k + 1 that lacks an element of value \( x_{1+k} \); i.e., \( s \in [1, 2k+2] \). It follows that if \( x_{1+k} \leq x_{k+1} \) then \( m(z) = (x_{k+2}) \) and, if \( x_{1+k} \geq x_{k+1} \) then \( m(z) = x_{k+1} \). Since each \( z_1 \) and \( z_2 \) lack an element of x, then...