Approximation Complexity for Piecewise Monotone Functions and Real Data

M. Brooks
Institute for Information Technology
M50 Montreal Road, Ottawa, Ontario, K1A 0R6 Canada
brooks@iit.nrc.ca

(Received January 1993; revised and accepted May 1993)

Abstract—This paper investigates the relationship between approximation error and complexity. A variety of complexity measures are used, including: the number of alternating strictly monotone segments; computation time; and for piecewise linear approximations, the number of linear segments. The results apply to piecewise monotone functions and to finite maps from reals to reals, i.e., real data. We provide a theoretical framework expressing the exact relationship between approximation error and the number of alternating strictly monotone segments. We provide a linear-time algorithm taking an error bound as input and returning a minimal segmentation of the approximated function’s domain such that there exists an approximation, alternatingly strictly monotone on the segments, with error less than the given bound. For real data, we provide a suboptimal tradeoff between approximation error and number of linear segments in piecewise linear approximation. The results are obtained by extending the theory of best piecewise monotone approximation to piecewise monotone functions, and by application of a new concept, scale-dependent monotonicity.

1. INTRODUCTION

Approximation complexity concerns the tradeoff between an approximating function’s error versus its complexity. When seeking an approximation $g$ to a function $f$, we attempt to minimize the expression $\sup \{ |g(x) - f(x)| + A \text{ complexity } (g)$, where $A$ is a positive constant. Complexity measures for $g$ include polynomial degree, the number of alternating strictly monotone segments, the number of linear segments, and computational complexity. Other error metrics may also be appropriate.

This paper provides two approximation complexity results applying to both piecewise monotone functions on a closed real interval and finite maps from reals to reals; we call the latter finite real functions or real data:

- A theoretical framework, suitable for practical computation, exactly describes the relationship between an approximating function’s error and its number of alternating strictly monotone segments.
- A linear-time algorithm takes an error bound as input and returns a minimal segmentation of the approximated function’s domain such that there exists an approximation, alternatingly strictly monotone on the segments, with error less than the given bound.

For real data, we have an additional approximation complexity result:

- A construction for continuous piecewise linear approximation gives an explicit, although suboptimal, relationship between approximation error and number of linear segments.

The connection between piecewise monotone functions and real data is derived from an extension of recent results in piecewise monotone approximation [1].
Given a piecewise monotone function \( f \) and an error bound \( \varepsilon \), let \( m \) be the minimal number of alternating strictly monotone segments required for piecewise monotone approximation to \( f \) with error less than \( \varepsilon \). Then, there exist best piecewise monotone approximations having \( m \) alternating strictly monotone segments.

- The monotonicity structure and approximation error of these best approximations depends only on the monotonicity structure of the approximated function.

We introduce scale-dependent monotonicity, the notion that monotonicity can be defined relative to measurement scale. Measurement scale is quantified by a positive real number \( \delta \). A finite real function is strictly \( \delta \)-monotone on a segment of its domain if either its values rise by \( \delta \), but do not drop by \( \delta \), or vice versa. A \( \delta \)-structure for finite real function is a partition of its domain into alternating strictly \( \delta \)-monotone segments. The results of this paper are derived using \( \delta \)-structures as a framework for approximation.

The sequel is divided into four main sections. Section 2 presents new results in best piecewise monotone approximation, providing a connection between piecewise monotone functions and real data. Section 3 develops the theory of scale-dependent monotonicity. Sections 4 and 5 provide results in approximation complexity, obtained by application of scale-dependent monotonicity to piecewise monotone and piecewise linear approximation.

The following conventions hold throughout:

- \( I = [a,b] \) is a closed real interval, with \( a < b \).
- Lower case function symbols \( f, g \) map \( I \to \mathbb{R} \).
- \( D \) is a finite set of real numbers, \( |D| \geq 2 \).
- Upper case function symbols \( F, G \) map \( D \to \mathbb{R} \).
- \( \delta, \delta_1, \delta_2, \ldots \) are positive real numbers.

For \( f : I \to \mathbb{R} \), we use the norm \( ||f||_{\infty} = ||f|| = \sup\{|fx| \mid x \in I\} \). For real numbers \( x, y \) and sets \( X, Y \), we write expressions \( y < X, X < Y, x < y \in X \) having the obvious meaning. For a function \( F \) and a set \( X \), \( FX \) denotes the image of \( X \). The restriction of \( F \) to a subset of its domain is written \( F|_X \).

A nonempty subset \( X \) of \( D \) is an interval in \( D \) if \( \forall x < y < z \in D \), \( z \in X \Rightarrow y \in X \). When we write \( D \supset X \), it will be assumed that \( X \) is a nonempty interval unless explicitly stated otherwise. A partition of \( D \) is a sequence of nonempty intervals \( X_1, \ldots, X_m \) whose union is \( D \), with \( X_1 < \cdots < X_m \).

### 2. APPROXIMATION OF PIECEWISE MONOTONE FUNCTIONS AND REAL DATA

In this section, we provide a close connection between approximation of piecewise monotone functions on a real interval and approximation of finite real functions. We extend the results of [1], showing that best piecewise monotone approximations exist for piecewise monotone functions; furthermore, we show that their monotonicity structure and approximation error depend only on a certain finite sample of the function. Conversely, we let a finite real function describe a family of piecewise monotone functions on a real interval, and we use best piecewise monotone approximations to the latter as approximations to the finite real function.

#### 2.1. Best Piecewise Monotone Approximation of Piecewise Monotone Functions

Piecewise monotone approximation of continuous real functions on a real interval has recently been investigated in [1]. The main result is that for any \( m > 0 \), there exist continuous and \( C^\infty \) best piecewise monotone approximations having \( m \) or fewer monotone segments. The analysis is based on knot vectors: sequences of points defining the intervals upon which the approximating function is to be monotone. Best knot vectors define the monotone behavior of approximations having minimal error; however, continuous approximations cannot necessarily be chosen. This is
improved by a procedure transforming any best knot vector into another best knot vector having the additional property of being “alternant local extremal points”; these support continuous best piecewise monotone approximations.

In [1, Theorem 2.1], continuity of \( f \) is shown to be a sufficient condition to ensure the existence of a best piecewise monotone approximation to \( f \). (The example in the proof of [1, Proposition 2.1] constructs a discontinuous \( f \) for which a best piecewise monotone approximation does not exist). We show that an alternative condition—piecewise monotonicity—is also sufficient to show the existence of a best approximation.


A knot vector of size \( n \) on \( I = [a,b] \) is defined in [1] as \( p = (p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1} \) with \( a = p_0 < p_1 < \cdots < p_n = b \); a proper knot vector fulfills \( p_0 < p_1, \ldots, < p_n \). A knot vector \( q = (q_0, q_1, \ldots, q_m) \) on \( I \) is a subvector of \( p = (p_0, p_1, \ldots, p_n) \) if there are distinct indices \( i_1, \ldots, i_{m-1} \) such that \( q = (p_{i_0}, p_{i_1}, \ldots, p_{i_{m-1}}, p_n) \).

\( f : I \rightarrow \mathbb{R} \) is properly \( n \)-monotone if there exists a proper knot vector \( p = (p_0, p_1, \ldots, p_n) \) such that \( f \) is strictly monotone on each interval \([p_i, p_{i+1}]\) in alternating directions; we say \( p \) is a minimal knot vector for \( f \). An extremal knot vector for \( f \) is any subvector of any minimal knot vector. Properly \( n \)-monotone functions are bounded. For \( n \neq m \), the intersection of the proper \( n \)-monotone and proper \( m \)-monotone functions is empty. \( n \) is the monotone piece count of a proper \( n \)-monotone function.

**Theorem 1.** Let \( f : I \rightarrow \mathbb{R} \) be properly \( n \)-monotone and let \( \varepsilon > 0 \). Let \( m \leq n \) be the least integer such that there exists a proper \( m \)-monotone function \( g : I \rightarrow \mathbb{R} \) with \( \| f - g \| < \varepsilon \). Then, there exists such a \( g \) having \( \| f - g \| \leq \| f - h \| \) for all proper \( m \)-monotone functions \( h : I \rightarrow \mathbb{R} \).

\( g \) is a best piecewise monotone \( \varepsilon \)-approximation to \( f \); \( \| f - g \| \) is the best piecewise monotone \( \varepsilon \)-approximation error.

**Theorem 2.** Let \( \Phi \) be any collection of proper \( n \)-monotone functions on \( I \) having a common minimal knot vector, and let \( \varepsilon > 0 \). By Theorem 1, each \( f \in \Phi \) has a nonempty set \( \Theta_f \) of best piecewise monotone \( \varepsilon \)-approximations. Then, additionally the monotone piece count of all functions in all the \( \Theta_f \) is the same, say \( m \); the best piecewise monotone \( \varepsilon \)-approximation error is the same for each \( f \in \Phi \); and there exists a proper \( m \)-monotone knot vector that is a minimal knot vector for at least one approximation in each \( \Theta_f \).

Theorem 2 says that the monotonicity structure and approximation error of best piecewise monotone \( \varepsilon \)-approximations to \( f \) is determined only by \( f \)'s monotonicity structure.

In [1], continuous functions on real intervals are shown to have continuous best piecewise monotone approximations. Theorem 1 cannot be extended to provide continuous best piecewise monotone approximations if no constraints on \( f \) are assumed. In [1], the continuity of \( f \) is used to show the existence of best piecewise monotone approximations; continuity of the approximation follows from application of a certain procedure. Our proof of Theorem 1 utilizes a similar procedure to show the existence of best piecewise monotone approximations.

The example in [1] of a discontinuous function \( f \) for which a best piecewise monotone approximation does not exist is \( f : [-1,1] \rightarrow \mathbb{R}, f(x) = 0 \) for \( x \in [-1,0] \), and \( f(x) = 1 - x \) for \( x \in (0,1] \); it is shown that there is no best proper 2-monotone approximation. Note that \( f \) is not properly \( m \)-monotone for any \( m \). The obstacle is the one-sided discontinuity; two-sided discontinuities do not cause such problems. Note also that a continuous function on a closed interval is not necessarily \( m \)-monotone for any \( m \), e.g., \((1/x^2) \sin(1/x)\) on \([0,1] \).

The remainder of this section is proof of Theorems 1 and 2. We make extensive use of concepts from [1,5,6].
DEFINITION 1. (Derived from [1,5]). Let \( f : I \to \mathbb{R} \), bounded, and let \( J \) be a closed subinterval of \( I \). \( f \)'s monotone increasing breakdown measure on \( J \) is

\[
\tilde{\mu}(J) = \sup\{fx - f y \mid x \leq y \in J\},
\]

and \( f \)'s monotone decreasing breakdown measure on \( J \) is

\[
\tilde{\mu}(J) = \sup\{fy - fx \mid x \leq y \in J\}.
\]

These measure the degree to which \( f \) fails to be monotone increasing and decreasing, respectively, on \( J \); they are nonnegative. Let \( p = (p_0, p_1, \ldots, p_n) \) be a knot vector on \( I \); then, \( f \)'s initially increasing piecewise monotone breakdown measure on \( p \) is:

\[
\mu^+(p) = \max\{\text{if } i \text{ odd then } \tilde{\mu}([p_{i-1}, p_i]) \text{ else } \tilde{\mu}([p_{i-1}, p_i]) \mid i = 1, \ldots, n\},
\]

and \( f \)'s initially decreasing piecewise monotone breakdown measure on \( p \) is:

\[
\mu^-(p) = \max\{\text{if } i \text{ odd then } \tilde{\mu}([p_{i-1}, p_i]) \text{ else } \tilde{\mu}([p_{i-1}, p_i]) \mid i = 1, \ldots, n\}.
\]

These measure the degree to which \( f \) fails to fit the monotone structure described by \( p \).

We now use results from [6]: let \( f : I \to \mathbb{R} \), bounded, and let \( p = (p_0, p_1, \ldots, p_n) \) be any proper knot vector. \( f\lvert_{[p_{i-1}, p_i]} \) is approximated by a nondecreasing function \( g_i^+ : [p_{i-1}, p_i] \to \mathbb{R} \) defined as the pointwise average of the least nondecreasing function dominating \( f\lvert_{[p_{i-1}, p_i]} \) and the greatest nondecreasing function dominated by \( f\lvert_{[p_{i-1}, p_i]} \). A nonincreasing function \( g_i^- : [p_{i-1}, p_i] \to \mathbb{R} \) is defined similarly. Then, \( \|f\lvert_{[p_{i-1}, p_i]} - g_i^+\| = \tilde{\mu}([p_{i-1}, p_i])/2 \) and \( \|f\lvert_{[p_{i-1}, p_i]} - g_i^-\| = \mu([p_{i-1}, p_i])/2 \). These approximation errors are optimal for all nondecreasing and nonincreasing functions on \( [p_{i-1}, p_i] \). Letting \( z_i^+ = \sup\{fx \mid x \in [p_{i-1}, p_i]\} \) and \( z_i^- = \inf\{fx \mid x \in [p_{i-1}, p_i]\} \), it is easily derived from the construction that \( g_i^+ \) is constant iff \( f(p_{i-1}) = z_i^+ \) and \( f(p_i) = z_i^- \), in which case \( g_i^+ x = (z_i^+ - z_i^-)/2 \), and similarly for \( g_i^- \).

Functions \( h^+, h^- : I \to \mathbb{R} \) are constructed by alternating the monotone segments \( g_i^+ \) and \( g_i^- \), starting with \( g_1^+ \) for \( h^+ \) and \( g_1^- \) for \( h^- \), choosing between the two values at the knots \( p_i, i = 1, \ldots, n \), as follows. Between \( g_{i-1}^+ \) and \( g_i^- \), choose the larger value, and between \( g_{i-1}^- \) and \( g_i^+ \), choose the smaller value. Then, \( \|f - h^+\| = \mu^+(p)/2 \) and \( \|f - h^-\| = \mu^-(p)/2 \). These functions may fail to be properly \( n \)-monotone only if some of the monotone segments are constant. Using notation \( h_{p_1}^+, h_{p_n}^- \) to show the dependence on \( p \), this is remedied by the following lemma.

**Lemma 1.1.** For nonconstant bounded \( f : I \to \mathbb{R} \) and proper knot vector \( p \) of size \( n \) on \( I \), a subvector \( q \) of size \( m \leq n \) can be chosen such that \( \min\{\|f - h_q^+\|, \|f - h_q^-\|\} \leq \min\{\|f - h_{p_1}^+\|, \|f - h_{p_n}^-\|\} \) and the better approximation among \( h_q^+, h_q^- \) is properly \( m \)-monotone.

**Proof.** Note that \( h_q^+ \) and \( h_q^- \) cannot both be constant, and if \( h_q^+ \) is constant then, \( \|f - h_q^+\| < \|f - h_q^-\| \). So suppose \( h_q^+ \) is nonconstant and \( \|f - h_q^+\| < \|f - h_q^-\| \); it is sufficient to prove the following three cases:

1. \( h_q^+ \) has monotone segments \( g_{i-1}^+, g_i^+, g_{i+1}^+ \), with \( g_i^- \) constant and \( g_{i-1}^-(q_{i-1}) \leq g_i^-(q_{i-1}) = g_{i+1}^-(q_{i+1}) \); and
2. \( h_q^+ \) has beginning segments \( g_1^+, g_2^- \), with \( g_2^- \) constant and \( g_1^-(q_1) \geq g_2^-(q_1) \); and
3. \( h_q^+ \) has ending segments \( g_{n-1}^+, g_n^- \), with \( g_n^- \) constant and \( g_{n-1}^+(q_{n-1}) \leq g_n^+(q_n) \).

We prove only the first case; the third is similar; the only difference with the second case is that the resulting function will be initially decreasing. For the first case, we claim \( \mu^+(\{q_{i-2}, q_{i+1}\}) \leq \max\{\mu([q_{i-2}, q_{i-1}]), \mu([q_{i+1}, q_i]), \mu([q_i, q_{i+1}])\} \), so \( q \)'s subvector \( q' \) obtained by deleting \( q_{i-1} \) and \( q_i \) has \( \mu^+(q') \leq \mu^+(q) \), i.e., \( \|f - h_{q'}^+\| \leq \|f - h_q^+\| \). If one of \( g_{i-1}^+, g_{i+1}^- \) is nonconstant, then the \( i \)-th segment of \( h_q^+ \) cannot be constant. The claim is proved using the notation \( z_i^+ \), \( z_i^- \) above, noting
that \( g_{i-1}^+(q_{i-1}) \leq g_i^+(q_{i-1}) \leq g_i^{+\pm}(q_i) \) implies \( z_i^+ - z_i^- \leq \mu^+(q), z_i^+ - z_{i+1}^- \leq \mu^+(q) \) and \( z_i^+ - z_{i+1}^- \leq \mu^+(q) \), so \( \mu^+(q') \leq \mu^+(q) \). Iterating this process yields the desired subvector and approximating function.

**Lemma 1.2.** For properly \( n \)-monotone \( f: I \to \mathbb{R} \) and proper knot vector \( q \) of size \( m \leq n \) on \( I \), an extremal subvector \( q' \) of size \( m' \leq m \) can be chosen such that \( \min\{ ||f - h_{q'}^+||, ||f - h_{q'}^-|| \} \) < \( \min\{ ||f - h_q^+||, ||f - h_q^-|| \} \) and the better approximation among \( h_q^+ \), \( h_q^- \) is properly \( m' \)-monotone.

**Proof.** Let \( p = (p_0, p_1, \ldots, p_m) \) be a minimal knot vector for \( f \); suppose \( q = (q_0, q_1, \ldots, q_m) \). By Lemma 1.1, we may assume \( h_q^+ \) is properly \( m \)-monotone and \( ||f - h_q^+|| \leq ||f - h_q^-|| \); this accounts for the possible reduction in size from \( m \) to \( m' \) in the statement of the theorem. \( h_q^+ \) properly \( m \)-monotone implies there are at most two \( q \)-knots in any \( [p_{i-1}, p_i] \), and if \( q_{i-1}, q_i \in [p_{i-1}, p_i] \), then the \( j \)th segment of \( h_q^+ \) is increasing iff \( f \) is increasing on \( [p_{i-1}, p_i] \). We construct \( q' \) in \( m - 1 \) steps from \( q \); \( q_0 \) is defined as \( q \), the result of Step 1 is \( q(1) \), and \( q' = q(m-1) \). At commencement of Step \( i < m \), we have \( q(i-1) = (p_0, p_1, \ldots, p_{i-1}, q_i, q_{i+1}, \ldots, q_m) \), where the \( p_j \) are knots from \( p \); the prime indicates reindexing; for \( i = 1 \), the sequence \( p_1, \ldots, p_{i-1} \) denotes the empty sequence. Let the index \( j \leq n \) be such that \( q_j \in [p_{i-1}, p_i] \). We claim \( j > i \) by induction. We show the case that \( f \) is increasing on \( [p_{i-1}, p_i] \). If the \( i \)th segment of \( h_q^+ \) is increasing, then \( q(i) = (p_0, p_1, \ldots, p_{i-1}, p_i, q_{i+1}, \ldots, q_m) \); otherwise, the claim can be strengthened to \( j > i, \) and \( q(i) = (p_0, p_1, \ldots, p_{i-1}, p_i-1, q_{i+1}, \ldots, q_m) \). In both cases, \( \mu^+(q(i)) \leq \mu^+(q(i-1)) \) and \( h_q^+ \) is properly \( m \)-monotone.

**Proof of Theorem 1.** By Lemma 1.2, we can restrict our attention to extremal knot vectors for \( f \). Exhaustive search yields the shortest extremal knot vector, \( q \), of length \( m \), such that \( \min\{ ||f - h_q^+||, ||f - h_q^-|| \} < \varepsilon \), such that \( \min\{ ||f - h_q^+||, ||f - h_q^-|| \} \) is minimal for all extremal knot vectors of length \( m \), and such that the better approximation among \( h_q^+ \), \( h_q^- \) is properly \( m \)-monotone.

The proof allows recovery of continuity.

**Corollary 1.1.** Let \( f : I \to \mathbb{R} \) be properly \( n \)-monotone and continuous, and let \( \varepsilon > 0 \). Then, there exist continuous best piecewise monotone \( \varepsilon \)-approximations to \( f \).

**Proof.** This is proved in [1], but in our case a direct proof is simple: let \( g \) be a best piecewise monotone \( \varepsilon \)-approximation to \( f \) having a minimal knot vector \( q = (q_0, q_1, \ldots, q_m) \) that is also one of \( f \)'s extremal knot vectors; assume \( g \) is constructed using monotone segments \( g_+ \), \( g_- \) as described in the text. By construction, the \( g_+ \) and \( g_- \) are continuous when \( f \) is continuous. We repair discontinuities at the knots as follows: if \( g_+^+(q_i) \neq g_-^-(q_i) \), then \( g_+^+(q_i) > g_-^-(q_i) = f(q_i) \); in fact, \( g_{i+1} = f([q_i, q_{i+1}]) \). Define a linear segment \( h \) on \([q_i, q_{i+1}]\) by \( h(q_i) = g_+^+(q_i) - f(q_i) \) and \( h(q_{i+1}) = 0 \). Replacing \( g_{i+1} \) by \( g_{i+1} + h \) irons out the discontinuity without increasing the approximation error.

**Proof of Theorem 2.** We show that choice of minimal knot vector \( p \) in the proof of Lemma 1.2 is irrelevant. This proof used a deterministic construction to obtain an extremal knot vector \( q' \) for \( f \) of size \( m' \leq m \) such that \( \min\{ ||f - h_{q'}^+||, ||f - h_{q'}^-|| \} \leq \min\{ ||f - h_q^+||, ||f - h_q^-|| \} \). We need to show that for a different choice \( p' \) of minimal knot vector, the resulting extremal knot vector \( q'' \) of size \( m'' \) would have \( m'' = m' \) and \( \min\{ ||f - h_{q''}^+||, ||f - h_{q''}^-|| \} \leq \min\{ ||f - h_q^+||, ||f - h_q^-|| \} \). That \( m'' = m' \) is obvious. The second part follows from the observation that the method of construction of \( h_q^+ \) implies \( ||f - h_q^+|| = \sup ||f(x) - h_q^+(x)|| \) is attained at one of \( q \)'s knots, and for each knot \( p_i \) of \( q \) the value of \( ||f(p_i) - h_q^+(p_i)|| \) is insensitive to changes in \( p_i \) along intervals where \( f \) is constant.

### 2.2. Approximation of Real Data by Functions on a Real Interval

Intuitively, real-data \( F : D \to \mathbb{R} \) may be approximated by \( g : [\min D, \max D] \to \mathbb{R} \). In some applications, \( g \) may be interpreted as a model for \( F \). In other applications, \( F \) may be a sample of
an unknown \( f \cdot [\min D \max D] \rightarrow \mathbb{R} \); the approximation \( g \) is an attempt to recover \( f \). Technically, we need to agree on where to measure approximation error. One approach is to restrict \( g \) to \( D \); however, this leaves "most" of \( g\)'s behavior unconstrained. The approach considered here is to extend \( F \) to some \( f : [\min D \max D] \rightarrow \mathbb{R} \) and approximate \( f \). The challenge is to identify such \( f \) having meaningful structural similarity to \( F \).

We use the monotonicity structure of the function \( F : D \rightarrow \mathbb{R} \) as a constraint on extensions to \( f : [\min D \max D] \rightarrow \mathbb{R} \). The definitions of knot vectors and piecewise monotonicity apply naturally to finite real functions. Suppose \( F \) is properly \( n \)-monotone. Let \( \Phi_0 \) be the set of all proper \( n \)-monotone \( f : [\min D \max D] \rightarrow \mathbb{R} \) such that \( f_{|D} = F \), and \( f \) shares a minimal knot vector with \( F \). Theorem 2 implies that best piecewise monotone \( \varepsilon \)-approximation to functions in \( \Phi_0 \) all have the common knot vectors and equal approximation errors; i.e., each \( f \in \Phi_0 \) is strongly characterized by \( F \), and this characterization is uniform over \( \Phi_0 \).

Note that although the monotonicity structure and approximation error of best piecewise monotone \( \varepsilon \)-approximations to \( f \in \Phi_0 \) depend only on \( f \)'s monotonicity structure, the best piecewise monotone \( \varepsilon \)-approximations themselves depend on all of \( f \). Thus, different extensions of \( F : D \rightarrow \mathbb{R} \) to \( [\min D \max D] \) generate different best piecewise monotone \( \varepsilon \)-approximations. Choice of a most appropriate extension will be application dependent.

Best piecewise monotone \( \varepsilon \)-approximations to \( f \in \Phi_0 \) optimally reduce the number of alternating strictly monotone segments; this is a form of smoothing. In this sense, all \( f \in \Phi_0 \) are "equally smooth" as \( F \) and they all "smooth out" equivalently by taking best piecewise monotone \( \varepsilon \)-approximations. Functions \( f : [\min D \max D] \rightarrow \mathbb{R} \) having greater monotone complexity than \( F \) may also be characterized by \( F \); these are "noisier" than \( F \). If this noise has limited "amplitude," then for sufficiently large \( \varepsilon \), these noisy functions "smooth out" equivalently to functions in \( \Phi_0 \) by taking best piecewise monotone \( \varepsilon \)-approximations. Suppose \( F \) is properly \( n \)-monotone, let \( \varepsilon > 0 \), and let \( \Phi_{\varepsilon} \) be all proper \( m \)-monotone \( f : [\min D \max D] \rightarrow \mathbb{R} \), for all \( m > n \), with \( f_{|D} = F \), such that one of \( F \)'s minimal knot vectors \( p = (p_0, p_1, \ldots, p_n) \) is a subvector of one of \( f \)'s minimal knot vectors \( q = (q_0, q_1, \ldots, q_m) \) with the additional attribute that if \( F \) is increasing on \([p_{i-1}, p_i] \), then \( \mu F((p_{i-1}, p_i)) \leq 2\varepsilon \), else \( \mu F((p_{i-1}, p_i)) \leq 2\varepsilon \). Then, if \( \varepsilon' \) is large enough that the best piecewise monotone \( \varepsilon' \)-approximation error to \( F \) is \( \varepsilon^* > \varepsilon \), then best piecewise monotone \( \varepsilon' \)-approximations to functions in \( \Phi_{\varepsilon} \) have the same approximation errors and share minimal knot vectors with functions in \( \Phi_0 \).

### 3. SCALE-DEPENDENT MONOTONICITY

In this section, we investigate scale-dependent monotonicity, the notion that monotonicity can be defined relative to measurement scale, obtaining results used for approximation complexity in Sections 4 and 5.

**Definition 2.** \( x < y \in D \) is a \( \delta \)-pair for \( F \) if \(|Fy - Fx| \geq \delta \) and \( \forall z \in D \ x < z < y \Rightarrow |Fz - Fx| < \delta \) and \(|Fy - Fx| < \delta \). \([x y]_{\delta,F} \) denotes that \( x < y \) is a \( \delta \)-pair for \( F \); we write \([x y]_{\delta} \) suppressing \( F \). A \( \delta \)-pair's direction \( d[x y]_{\delta} = +1 \) else \(-1 \). We write \([x y]_{\delta} \in X \) to mean \( x < y \in X \) and \([x y]_{\delta} \).

Minimaly of \( \delta \)-pairs implies that if \( [x y]_{\delta} \neq [u v]_{\delta} \), then \( y \leq u \) or \( u \leq x \).

**Definition 3.** Let \( D \supset X \). \( X \)'s left extension \( \overline{X} \) is: if \( \min D \in X \), then \( \overline{X} = X \), else \( \overline{X} = X \cup \{\max(x \in D|x < X\})\).

**Definition 4.** Let \( D \supset X \). \( F \) is \( \delta \)-monotone on \( X \) if \( \forall[u v]_{\delta},[x y]_{\delta} \in \overline{X} d[u v]_{\delta} = d[x y]_{\delta} \). If \( F \) is \( \delta \)-monotone on \( X \), then:

- \( F \) is \( \delta \)-increasing on \( X \), written \( d_{\delta,F} X = d_{\delta} X = +1 \), if \( \exists[x y]_{\delta} \in \overline{X} d[x y]_{\delta} = +1 \).
- \( F \) is \( \delta \)-decreasing on \( X \), written \( d_{\delta,F} X = d_{\delta} X = -1 \), if \( \exists[x y]_{\delta} \in \overline{X} d[x y]_{\delta} = -1 \).
- \( F \) is strictly \( \delta \)-monotone on \( X \) if it is \( \delta \)-increasing or \( \delta \)-decreasing on \( X \).
**DEFINITION 5.** A δ-structure for $F$ is a partition $X_1, \ldots, X_m$ of $D$ such that $F$ is strictly δ-monotone on each $X_i$, and $d_δ X_i \neq d_δ X_i+1$, for $i = 1, \ldots, m - 1$.

If $\max FD - \min FD \geq δ$, then clearly $F$ has at least one δ-pair; we show the existence of δ-structures for $F$ by constructing one. Suppose $[x_1, y_1]_δ$ is the δ-pair having minimal $x_1$. Moving through $D$ in increasing order, $[x_1, y_1]_δ$ may be followed by zero or more possibly overlapping δ-pairs with the same direction before any δ-pairs having the opposite direction are encountered; let $[x_2, z_2]_δ$ be the first δ-pair with $d[x_2, z_2]_δ \neq d[x_1, y_1]_δ$ if such δ-pairs exist. Continuing until $F$'s δ-pairs are exhausted, we have points $x_1, \ldots, x_m$. Define a partition $P = X_1, \ldots, X_m$ of $D$ by $\max X_i = x_{i+1}$, for $i = 1, \ldots, m - 1$. Then, $P$ is a δ-structure for $F$.

The insight here is to look at $F$'s behavior as we move through $D$ in increasing order, observing a run of possibly overlapping δ-pairs of one direction followed by a run of δ-pairs in the opposite direction, and so on. Let $[x y]_δ$ be the last δ-pair of the $i$th run; then the points $x_i, y_i$ enclose the $i$th run. For convenience, we define $M_i$ as all points of $D$ between $y_i$ and $x_{i+1}$, inclusive, for $i = 1, \ldots, m - 1$.

This δ-structure constructed above is not necessarily unique; a δ-structure will result when each $\max X_i$ is independently chosen from $\mathbb{N}$. In fact, the boundaries of the $X_i$ may lie outside $M_i$, but in this case they cannot be chosen independently.

If $[x y]_δ$ and $[w z]_δ$ are δ-pairs in the same run appearing in that order, then $z > y$ or $z < y$ according to $d[x y]_δ = +1$ or $d[x y]_δ = -1$. Let $v_i \in M_i$ be such that $Fv_i$ is maximal if $d[x y]_δ = +1$ or $Fv_i$ is minimal if $d[x y]_δ = -1$; we say $v_i$ is extremal in $M_i$. The $v_i$ are where $F$ takes extreme values in its δ-monotone rising and falling. A δ-structure with $\max X_i = v_i$, for $i = 1, \ldots, m - 1$, is an extremal δ-structure.

From the preceding discussion, it is clear that all δ-structures for $F$ have the same size. Also, they are all initially increasing or all initially decreasing; we say $F$ is initially δ-increasing or initially δ-decreasing. This gives the following definition.

**DEFINITION 6.** If $\max FD - \min FD > δ$, then $F$'s δ-measure, denoted $m_δ F$, is the unique partition size for any δ-structure for $F$. For $δ$ such that $\max FD - \min FD < δ$, we define $m_δ F = 0$.

The function symbol $m$ stands for monotone; $m_δ F$ says how many times $F$ wiggles up and down at measurement scale $δ$. In this sense, it is a smoothness measure.

If $δ_1 > δ_2$, then each of $F$'s $δ_1$-pairs must contain a $δ_2$-pair having the same direction; this implies $m_δ F \leq m_δ F$, i.e., δ-measure is monotone.

The monotone breakdown measures $μ^+, μ^-$ apply to a finite real function $F : D \to \mathbb{R}$ as follows: let $P = X_1, \ldots, X_m$ be a partition of $D$. Then:

$$μ^+(P) = \max \{\text{if } i \text{ odd then } μ (X_i) \text{ else } μ \left(\frac{i}{X_i}\right) \mid i = 1, \ldots, m\},$$

$$μ^-(P) = \max \{\text{if } i \text{ odd then } μ \left(\frac{i}{X_i}\right) \text{ else } μ (X_i) \mid i = 1, \ldots, m\}.$$}

**THEOREM 3.** Let $m_δ F > 0$ and let $P = X_1, \ldots, X_m$ be any partition with $m \leq m_δ F$. Then, for every extremal δ-structure $Q$ for $F$, $\min\{μ^+(Q), μ^-(Q)\} \leq \min\{μ^+(P), μ^-(P)\}$.

Let $m_δ F > 0$ and let the points $x_i, y_i$ and sets $M_i$ be defined as previously in the text. Let $P = X_1, \ldots, X_n$ be a partition of $D$. We say that $P$ is a regular partition for $δ$ if no $X_i$ is contained in any $M_i$, no $X_i$ is contained $y_{i-1} \leq X_i \leq x_{i+1}$ with $i$ and $j$ of unequal parity, at most one $X_i$ has $X_i \leq x_2$, and at most one $X_i$ has $y_{n-1} \leq X_i$. Regularity implies $n \leq m_δ F$.

Every δ-structure is regular. A regular partition $P$ is extremal if each $\max X_i$ is extremal in some $M_j$. If $P$ is extremal and $n = m_δ F$, then $P$ is an extremal δ-structure.

**LEMMA 3.1.** Let $m_δ F > 0$ and let $P = X_1, \ldots, X_m$ be a regular partition for $δ$ with $μ^+(P) < δ$. If $F$ is initially δ-increasing, then there is an extremal partition $Q = Y_1, \ldots, Y_m$ with $μ^+(Q) < μ^+(P)$. A similar statement holds if $F$ is initially δ-decreasing.

**PROOF.** We construct $Q$ in $m - 1$ steps from $P$; Step $i$ defines $Y_i$ and a regular partition $P^{(i)}$, with $P^{(0)} = P$ and $P^{(m-1)} = Q$. Let $X'_i = X_i$ and $X'_{i+1} = X_{i+1} - Y_i$ for $i = 1, \ldots, m - 1$;
then, \( p^{(i-1)} = Y_1, \ldots, Y_{i-1}, X'_i, X_{i+1}, \ldots, X_m \). At commencement of Step \( i \), if \( \max X_i \in M_j \), for some \( j \), then \( Y_i = X'_i \). Otherwise, let the index \( j \) be maximal such that \( x_j \in X_i \). If \( i \) and \( j \) have equal parity, then choose any extremal \( z \in M_j \) and define \( Y_i = X'_i \cup \{ X \in X'_i \mid x \leq z \} \). If \( i \) and \( j \) have unequal parity, then choose any extremal \( z \in M_{j-1} \) and define \( Y_i = X'_i \setminus \{ x \in X'_i \mid x > z \} \). In all cases, regularity and an induction argument imply \( Y_i \) and \( X_i \) are nonempty and that \( P^{(i)} = Y_1, \ldots, Y_i, X_{i+1}, \ldots, X_m \) is a regular partition for \( \delta \). It is clear that \( P^{(m-1)} \) is an extremal partition. It remains to show that \( \mu^+ (P^{(i)}) \leq \mu^+ (P^{(i+1)}) \). Consider the case where \( i \) and \( j \) have equal parity, say odd, so that \( F \) is \( S \)-increasing between \( x_j \) and \( z \). For any \( u, v \in D, u < x_j \leq v \leq z \), we have \( F_u - F_v \leq F_u - F_x_j \). For any \( u, v \in X_{i+1}, u < v \leq z \) with \( F_u > F_v \), we have \( F_x - F_v \geq F_u - F_v \). Thus, \( \mu (Y_i) \leq \mu (X'_i) \) and \( \mu (X'_{i+1}) \leq \mu (X_{i+1}) \), i.e., \( \mu^+ (P^{(i)}) \leq \mu^+ (P^{(i-1)}) \).

**Proof of Theorem 3.** If \( P \) is regular and \( m < m_\delta F \), then it follows from the lemma that \( \mu^+(P) \geq \delta \) and \( \mu^-(P) \geq \delta \). Suppose \( m \leq m_\delta F \), \( P \) is not regular, and \( \min\{\mu^+(P), \mu^-(P)\} > \delta \). Then, a regular but smaller partition \( P' \) with \( \min\{\mu^+(P'), \mu^-(P')\} < \delta \) can be derived by taking unions of the appropriate \( X_i \). But this contradicts the first statement, so \( P \) is not regular, then \( \mu^+(P) \geq \delta \) and \( \mu^-(P) \geq \delta \). If \( P \) is regular and \( m = m_\delta F \), then the lemma gives an extremal \( \delta \)-structure \( Q \) with \( \min\{\mu^+(Q), \mu^-(Q)\} \leq \min\{\mu^+(P), \mu^-(P)\} \). Finally, note that all extremal \( \delta \)-structures have equal monotone breakdown measures.

Finally, we define \( \delta \)-variation, a scale-dependent version of total variation, that we use in Section 5 when characterizing the number of segments required in piecewise linear approximations.

**Definition 7.** Let \( D \supset X \). Then, \( F \)'s \( \delta \)-variation on \( X \) is \( \delta \)-variation \( \delta \)-structure \( X_1, \ldots, X_m, \mu \) defines a minimal knot vector \( (p_0, \max X_1, \ldots, \max X_m) \) for a best piecewise monotone \( \varepsilon \)-approximation. Define \( S = \{ Q \mid Q \) is a \( \delta \)-structure for \( F \), for some \( \delta \) such that \( \min\{\mu^+(Q), \mu^-(Q)\} < 2\varepsilon, \) and then finding a partition \( Q \) minimizing \( \{\mu^+(Q), \mu^-(Q)\} \) over all partitions of \( D \) of this size. By Lemma 3.1, no partition smaller than a \( 2\varepsilon \)-structure for \( F \) has sufficiently small piecewise monotone breakdown measures, proving (1). By Theorem 3, the piecewise monotone breakdown measures are minimized by extremal \( 2\varepsilon \)-structures, proving (2).

**Proof.** By Lemma 1.2, there exists a best piecewise monotone \( \varepsilon \)-approximation with minimal knot vector that is a subvector of \( p \), so the problem of determining such a knot vector reduces to finding the minimal size for a partition \( P \) of \( D \) such that \( \min\{\mu^+(P), \mu^-(P)\} < 2\varepsilon \), and then finding a partition \( Q \) minimizing \( \{\mu^+(Q), \mu^-(Q)\} \) over all partitions of \( D \) of this size. By Lemma 3.1, no partition smaller than a \( 2\varepsilon \)-structure for \( F \) has sufficiently small piecewise monotone breakdown measures, proving (1). By Theorem 3, the piecewise monotone breakdown measures are minimized by extremal \( \varepsilon \)-structures, proving (2).

(3) is proved as follows: define \( S = \{ Q \mid Q \) is a \( \delta \)-structure for \( F \), for some \( \delta \) such that \( m_\delta F = m_\delta F \}. Any \( \delta \)-structure \( Q \in S \) is also a \( \delta' \)-structure for all \( \delta' \) such that \( \delta' > \min\{\mu^+(Q), \mu^-(Q)\} \). Now let \( \delta^* = \min\{\mu^+(Q), \mu^-(Q)\} \mid Q \in S \}. If \( m_\delta F = m_\delta F \), then \( \delta^* = 0 \), else \( m_\delta F > m_\delta F \) and \( m_\delta^* \alpha F = m_\delta F \) for all sufficiently small \( \alpha > 0 \), so if \( m_\delta F = m_\delta F \), then \( \delta^* = \delta_{k+1} \).
THEOREM 5. Let $f : I \rightarrow \mathbb{R}$ be properly $n$-monotone and let $\varepsilon > 0$. Given any minimal knot vector for $f$, we can compute in $O(n)$ time the best piecewise monotone $\varepsilon$-approximation error and a minimal knot vector for a best piecewise monotone $\varepsilon$-approximation.

PROOF. Let $p = (p_0, p_1, \ldots, p_n)$ be a minimal knot vector for $f$. Define $D = \{p_0, p_1, \ldots, p_n\}$ and $F : D \rightarrow \mathbb{R}$ by $F = f|_D$. Suppose $m_\delta F > 0$ and consider the following nondeterministic algorithm constructing a partition of $D$.

**δ-structure algorithm**

Input: $D, F, \delta, \varepsilon$, and $Z_0$, a strictly $\delta$-monotone initial segment of $D$.

Let $Z = Z_0$ and $i = 0$.

For $y \in D - Z_0$, in ascending order:

- If $Z \cup \{y\}$ is $\delta$-monotone then Set $Z = Z \cup \{y\}$.
- else Split $Z \cup \{y\}$ into strictly $\delta$-monotone sets $X_{i+1} < Z'$.
  
  Set $Z = Z'$ and $i = i + 1$.

If $i = 0$ return $Z$ else return $X_1, \ldots, X_i, Z$.

We prove that the $\delta$-structure algorithm computes a $\delta$-structure for $F$: let $Z_0$ be any such initial segment; we show by induction on $|D - Z_0|$ that the $\delta$-structure algorithm returns a $\delta$-structure for $F$ when run with inputs $D, F, \delta, \varepsilon$, and $Z_0$. If $|D - Z_0| = 0$, then the algorithm returns the single-element sequence $Z_0$, which, in this case, is a $\delta$-structure for $F$. Suppose $|D - Z_0| > 0$. We show that upon completing each iteration of the loop, if $i = 0$, then $Z$ is a $\delta$-structure for $F|_Z$ else $X_1, \ldots, X_i, Z$ is a $\delta$-structure for $F|_{X_1 \cup \ldots \cup X_i \cup Z}$. Let $Z_*$ denote the value of $Z$ before commencing each iteration of the loop. The induction hypothesis is: if $i = 0$, then $Z_*$ is a $\delta$-structure for $F|_{Z_*}$ else $X_1, \ldots, X_i, Z_*$ is a $\delta$-structure for $F|_{X_1 \cup \ldots \cup X_i \cup Z_*}$. Note that this implies that $Z_*$ is strictly $\delta$-monotone. If $Z_* \cup \{y\}$ is $\delta$-monotone, then $Z = Z_* \cup \{y\}$ and the result is immediate, by induction. Otherwise, it suffices to show two things:

(a) that $Z_* \cup \{y\}$ can be split into strictly $\delta$-monotone sets $X_{i+1} < Z'$; and

(b) that any choice of split $X_{i+1} < Z'$ gives $d_\delta Z' \neq d_\delta X_{i+1} = d_\delta Z_*$.

We show (a) as follows: since $Z_* \cup \{y\}$ is not $\delta$-monotone, there exists $[x, y]_\delta \in Z_* \cup \{y\}$ with $d([x, y]_\delta) = d_\delta Z_*$. For any $[v, u]_\delta \in Z_*$, we have $v \leq x$. Thus, we can choose $X_{i+1} = \{z \in Z : z \leq x\}$ and $Z' = \{z \in Z : z > x\} \cup \{y\}$. Now we show (b) using the same notation: let $X_{i+1} < Z'$ be any split; then $[x, y]_\delta \in Z'$, since otherwise $Z'$ would not be strictly $\delta$-monotone. Thus, $d_\delta Z' \neq d_\delta X_{i+1} = d_\delta Z_*$.

Splitting $Z \cup \{y\}$ into strictly $\delta$-monotone sets $X_{i+1} < Z'$ is the nondeterministic kernel of the $\delta$-structure algorithm. Distinct $\delta$-structures result from different splits. Extremal $\delta$-structures are created by picking $w = \max X_{i+1}$ such that: if $d_\delta Z = +1$, then $Fw = \max FZ$ else $Fw = \min FZ$.

A strictly $\delta$-monotone initial segment of $D$ can be found in linear time; we will not go into the details. By remembering appropriate maximal and minimal values as the $\delta$-structure algorithm iterates through $D$, we can split $Z \cup \{y\}$ to get an extremal $\delta$-structure in constant time, and we can keep track of the maximum monotone breakdown measure with constant overhead per iteration.

5. SUBOPTIMAL PIECEWISE LINEAR APPROXIMATION

THEOREM 6. Suppose $F' : D \rightarrow \mathbb{R}$ and $m_\delta F' > 0$. Then, $F'$ has a continuous proper $m_\delta F'$-monotone piecewise linear approximation with approximation error less than $\delta$ and no more than $[2u_\delta F'] + m_\delta F'$ linear segments.

Theorem 6 provides one point on the tradeoff curve between approximation accuracy and complexity. It is an open question as to what additional constraints would sharpen this result.

A result similar to Theorem 6 can also be proved: $F$ has a (discontinuous) proper $m_\delta F$-monotone piecewise constant approximation with error less than $\delta$ having $m_\delta F$ monotone segments and no more than $[u_\delta F] + m_\delta F$ constant segments.
Our piecewise linear construction for the proof of Theorem 6 will be based on an extremal \( \delta \)-structure. The approximation error will not exceed \( \delta \); this is clearly suboptimal. The number of linear segments, \( [2\nu_b F] + m_b F \), may also be suboptimal, even for this large error bound.

Inspection of the proof of Theorem 6 shows that if \( F \) is extended to \( f : I \rightarrow \mathbb{R} \) by “connecting the dots” of \( F \)’s graph, i.e., linear interpolation, then the constructed approximation to \( F \) is also an approximation to \( f \) having error less that \( \delta \).

We construct piecewise linear functions as follows: let \( a < b, r, s \in \mathbb{R} \). The linear segment \( U : [a b] \rightarrow [r s] \) is the linear function determined by \( Ua = r \) and \( Ub = s \). \( U \) is increasing if \( r < s \), and decreasing if \( r > s \). Constant segments \( U : [a b] \rightarrow [r r] \) and \( V : [b c] \rightarrow [s t] \) are abbreviated \( U : [a b] \rightarrow r \) and \( V : [b c] \rightarrow [s t] \) be linear segments. Their concatenation is the continuous piecewise linear function \( U \oplus V : [a c] \rightarrow [r t] \) defined in the obvious way. Concatenation is extended to multiple segments in the obvious way.

Let \( P = X_1, \ldots, X_m \) be an extremal \( \delta \)-structure for \( F \). The piecewise linear construction will focus on the individual \( X_i \) using the following definition.

**Definition 8.** Suppose \( X \) is strictly \( \delta \)-monotone with \( d_X \delta = +1 \). Let \( p = [\nu_b X + 1] \). \( F \)’s \( \delta \)-blocks on \( X \) are sets \( R_i \) and \( S_i \), defined using intermediate values \( \lambda_i, P_i, \) and \( Q_i, \) as follows:

\[
\lambda_i = \min F X + i \delta \\
P_i = \{ x \in X \mid z \leq x \in X \Rightarrow Fz < \lambda_i \} \quad \text{for } i = 0, \ldots, p \quad (\lambda \text{ stands for “level”}) \\
Q_i = \{ x \in X \mid z \geq x \in X \Rightarrow Fz \geq \lambda_{i-1} \} \quad \text{for } i = 1, \ldots, p \\
R_i = \text{if } P_i \cap Q_i \neq \emptyset \text{ then } P_i \cap Q_i \text{ else } \{ z \} \quad \text{for any } z \in \mathbb{R} \text{ such that } P_i < z < Q_i \\
S_i = \left((P_{i+1} - P_i) \cap (Q_i - Q_{i+1})\right) \cup \{ z \} \quad \text{for any } z \in \mathbb{R} \text{ such that } R_i < z < R_{i+1} \\
\text{for } i = 1, \ldots, p-1.
\]

The definition of \( F \)’s \( \delta \)-blocks when \( d_X \delta = -1 \) is similar, but with reversed inequalities and with \( \lambda_i = \max F X - i \delta \).

Note that \( Q_1 = P_p = X \) and for \( i = 1, \ldots, p-1 : P_i \neq \emptyset, Q_i \neq \emptyset, P_{i+1} \supset P_i, \) and \( Q_{i+1} \supset Q_i \). The arbitrary element in \( R_i \) and \( S_i \) keeps them nonempty, simplifying a number of arguments. Note that \( Q_p \) may be empty; this results in \( R_p > \max X \).

The following four lemmas use the notation of the preceding definition and assume \( d_X \delta = +1 \). The case \( d_X \delta = -1 \) would be handled similarly: the statements of Lemmas 6.1 and 6.2 are the same for \( d_X \delta = -1 \); Lemmas 6.3 and 6.4 would have reversed inequalities.

**Lemma 6.1.** \( R_i < S_i < R_{i+1} \) for \( i = 1, \ldots, p-1 \).

**Proof.** First, we show \( R_i < S_i \):

For \( y \in S_i \), if \( y \in (P_{i+1} - P_i) \cap (Q_i - Q_{i+1}) \)

then for \( x \in R_i \), if \( x \in P_i \cap Q_i \)

then \( y \in P_{i+1} - P_i \Rightarrow y > P_i \Rightarrow y > x \)

else \( x < Q_i \), so \( y \in Q_i - Q_{i+1} \Rightarrow y \in Q_i \Rightarrow y > x \)

else for \( x \in R_i \), \( y > x \) by definition.

And similarly \( S_i < R_{i+1} \):

For \( y \in S_i \), if \( y \in (P_{i+1} - P_i) \cap (Q_i - Q_{i+1}) \)

then for \( x \in R_{i+1} \), if \( x \in P_{i+1} \cap Q_{i+1} \)

then \( y \in Q_i - Q_{i+1} \Rightarrow y < Q_{i+1} \Rightarrow y < x \)

else \( x > P_{i+1} \), so \( y \in P_{i+1} - P_i \Rightarrow y \in P_{i+1} \Rightarrow y < x \)

else for \( x \in R_i + 1 \), \( y < x \) by definition.
Lemma 6.2. \( R_1 \cup S_1 \cup \cdots \cup S_{p-1} \cup R_p \supset X \).

Proof. Every \( x \in X \) is contained in \( P_p \) and \( Q_1 \). For a given \( x \in X \), choose the smallest \( i \) and largest \( j \) such that \( x \in P_i \cap Q_j \). Clearly, \( i \geq j \). Thus, if \( i = 1 \) or \( j = p \), then \( i = j \), so \( x \in P_i \cap Q_i = R_i \). Consider the case where \( i > 1 \) and \( j < p \). For \( x \in P_j \cap Q_j = R_j \); or \( i = j + 1 \), so \( x \in P_{j+1} \cap Q_j \) and so \( x \in (P_{j+1} \setminus P_j) \cap (Q_j \setminus Q_{j+1}) = S_j \), since \( x \in P_j \setminus P_{j-1} = P_{j+1} \setminus P_j \) and \( x \in Q_j \setminus Q_{j+1} \).

Lemma 6.3. For \( i = 1, \ldots, p-1 \), if \( S_i \cap D \neq \emptyset \), then \( |S_i \cap D| \geq 2 \) and \( F \min S_i \cap D > \lambda_i \) and \( F(\max S_i \cap D) < \lambda_i \).

Proof. Otherwise, \( \min S_i \cap D \) would be in \( R_i \) or \( \max S_i \cap D \) would be in \( R_{i+1} \).

Lemma 6.4. For \( i = 1, \ldots, p-1 \), if \( x \in S_i \cap D \), then \( \lambda_{i-1} < Fx < \lambda_{i+1} \), i.e., \( |Fx - \lambda_i| < \delta \), and for \( i = 1, \ldots, p \), if \( x \in R_i \cap D \), then \( \lambda_{i-1} \leq Fx < \lambda_i \).

Proof. Clearly, if \( x \in S_i \), then \( \lambda_{i-1} \leq Fx < \lambda_{i+1} \), however, \( Fx \neq \lambda_{i-1} \) because \( d_x X = +1 \) and \( F(\min S_i \cap D) > \lambda_i \). The rest is obvious.

Proof of Theorem 6. Let \( P = X_1, \ldots, X_m \) be an extremal \( \delta \)-structure for \( F \). We construct a monotone piecewise linear approximation to \( F \) on each \( X_j \) having error less than \( \delta \). Concatenating these approximations gives the desired piecewise linear approximation to \( F \).

Choose an \( X_j \), let \( p = [v_0 X_j + 1] \), and let \( R_i, i = 1, \ldots, p \), and \( S_i, i = 1, \ldots, p-1 \), be \( F \)'s \( \delta \)-blocks on \( X_j \). Define linear and constant segments for the case \( d_x X = +1 \) as follows:

\[
U^1 : [\min R_1 \min S_1] \to [\lambda_0 \lambda_1] \\
W^1 : [\min S_1 \max S_1] \to \lambda_1,
\]

For \( i = 2, \ldots, p-1 \):

\[
U^i : [\max S_{i-1} \min S_i] \to [\lambda_{i-1} \lambda_i] \\
W^i : [\min S_i \max S_i] \to \lambda_i.
\]

If \( j < m \) then \( U^p : [\max S_{p-1} \max R_p] \to [\lambda_{i-1} F(\max R_p)] \)

else if \( R_p > \max X_j \), then \( U^p \) is undefined

else \( U^p : [\max S_{p-1} \max D] \to \lambda_{p-1} \).

The fuss with \( U^p \) works since \( P \) being an extremal \( \delta \)-structure for \( F \) implies \( \max R_p = \max X_j \) for \( j < m \).

By Lemma 6.4, the \( W^1 \) have approximating error less than \( \delta \). The \( U^i \) have approximation error less than \( \delta \), since if \( x \in R_i \cap D \), for any \( 1 \leq i \leq p \), then \( \lambda_{i-1} \leq U^i x < \lambda_i \). Concatenation of all the segments gives the desired approximation, \( G_j \), to \( F \) on \( X_j \).

The definition of \( G_j \) for the case \( d_x X_j = -1 \) is similar. We create the final approximation by concatenating the \( G_j \). We need to show continuity, i.e., for \( j < m \), \( G_j \max X_j = G_{j+1} \max X_j \). This follows since \( P \) is an extremal \( \delta \)-structure.

We complete the proof of Theorem 6 by counting the segments. For each \( X_j \), there are at most \( 2[v_0 X_j + 1] - 1 \) segments. Summing these we get:

\[
\sum_{i=1}^{m} (2[v_0 X_j + 1] - 1) \leq \sum_{i=1}^{m} (2v_0 X_j + 1) = 2v_0 F + m_0 F \leq [2v_0 F] + m_0 F.
\]
REFERENCES