

Isotonic, Convex and Related Splines¹

by

Ian W. Wright

and

Edward J. Wegman

Department of Statistics
University of North Carolina

¹This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2840. Dr. Wright's work was done while on leave from Department of Mathematics, Papua New Guinea University of Technology, Lae, Papua New Guinea.

Abstract. In this paper, we consider the estimation of isotonic, convex or related functions by means of splines. It is shown that certain classes of isotone or convex functions can be represented as a positive cone embedded in a Hilbert space. Using this representation, we give an existence and characterization theorem for isotonic or convex splines. Two special cases are examined showing the existence of a globally monotone cubic smoothing spline and a globally convex quintic smoothing spline. Finally, we examine a regression problem and show that the isotonic-type of spline provides a strongly consistent solution. We also point out several other statistical applications.

Key Words and Phrases: isotone, isotonic, convex, partial order, spline, isotonic spline, smoothing spline, interpolating spline, regression, strong consistency.

AMS Classification Nos: 62G05, 62M15, 65D05, 65D10, 06A10

1. Introduction: A good deal of literature concerning statistical inference under order restrictions has appeared in the last 15 years. Barlow et al (1972) presents a rather convenient source book for this material. The main body of this work can be conveniently divided into two areas: inferences concerning functions on a finite or countable set and inferences concerning functions on the real line or intervals in the real line. It is with this latter group of inferences that we are concerned.

Examples of functions of interest would include regression functions, probability distribution and density functions, failure rate functions and functions related to spectral analysis such as spectral densities, gain and transfer functions and so on. It is often known that these functions are *isotone* (i.e. order-preserving) and, hence, should be estimated with a function that preserves the order. See Robertson (1967), Wegman (1970a,b), Barlow, Marshall and Proschan (1963) and Marshall and Proschan (1965) for examples of such estimation problems. A characteristic feature of these isotonic estimators is that they are step functions. In most of these situations, smoothness is frequently just as desirable as isotonicity so that while the step function may be isotone, its lack of continuity prevents it from being widely accepted as a satisfactory estimator.

More recently, there has appeared a good deal of literature concerning the use of splines in statistical estimation problems. See Wright (1977) for a review of some of these efforts. While a spline fit satisfies the requisite smoothness properties, it may not be isotone as desired. In this present paper, we present a combined approach. For technical convenience, the class of partial orders is restricted to those compatible with the additive group of real functions. Our framework is general enough to encompass estimation problems for monotone, or convex or positive functions and other related families of functions.

2. Partial Orders and Isotonic Splines: In this paper we shall be dealing only with abelian groups of real functions which can be added pointwise.

Definition 2.1. The relation, \gg , on the group, G , of functions is required to satisfy the following conditions:

- i. $f \gg f$
- ii. $f \gg g, g \gg h$ implies $f \gg h$
- iii. $f \gg g$ implies $f + h \gg g + h$ for all $h \in G$.
- iv. $f \gg 0, g \gg 0$ implies $f + g \gg 0$
- v. $f \gg g$ and $g \gg f$ implies $f = g$.

A group satisfying i to v is called a *partially ordered group*. If condition v is dropped, G is called a *pre-ordered group*. The set, $P = \{g \in G: g \gg 0\}$ is called the *positive cone* of the order, \gg . The set, P , defines and is defined by the partial order, \gg .

The partial order, \gg , defined here is essentially distinct from the partial order which usually occurs in papers on isotonic methods - (cf. Barlow et al (1972, chapter 7)). In the standard treatment, a partial order, say, \succ , is induced on the real line. A function f is said to be *isotone* if it preserves the order. That is, f is isotone iff $f(x_1) \geq f(x_2)$ whenever $x_1 \succ x_2$. Two functions may both be isotone and yet not comparable. What is true, however, is that the set of isotone functions forms a positive cone as just defined. We have in mind an essentially different use of the partial order, \gg .

We let L_2 represent the set of functions on $[0,1]$ which are Lebesgue measurable and square integrable with the usual Banach space norm. We let $W_m, m \geq 1$, be the set of functions on $[0,1]$ for which $f^{(j)}, j = 0,1,\dots,m-1$ are absolutely continuous and $f^{(m)}$ is in L_2 . This is a Hilbert space with inner product $\langle f, g \rangle = \sum_{j=0}^m \int_0^1 f^{(j)}(t)g^{(j)}(t)dt$. Let $C^k, k = 1,2,\dots,\infty$, be the set of all

functions on $[0,1]$ which are k -times continuously differentiable, and finally we let D be the differentiation operator.

In order to define a suitable positive cone, we let F be a continuous linear map of W_{m-1} into L_2 which commutes with the differentiation operator, i.e. $D(Ff) = F(Df)$ for all $f \in W_m \subset W_{m-1}$. We define a partial order, \gg , on W_m by $f \gg 0$ if and only if $(Ff)(t) \geq 0$ for every $t \in [0,1]$.

There are several operators, F , of particular interest.

Example 2.1. If F is the identity map, the set, $P = \{g \in W_m : g \gg 0\}$ is just the set of positive functions in W_m .

Example 2.2. If $F = D$, P is just the set of monotone non-decreasing functions in W_m . Similarly, if $F = -D$, P is the set of non-increasing functions.

Example 2.3. If $F = D^2$, P is the set of convex functions in W_m while $F = -D^2$ yields the set of concave functions.

Example 2.4. If F is defined by

$$Ff(t) = \begin{cases} Df(t) & 0 \leq t < m \\ -Df(t) & m \leq t \leq 1, \end{cases}$$

then P is the set of unimodal functions with mode m .

Example 2.5. If F is defined by

$$Ff(t) = \begin{cases} -D^2f(t) & t_1 \leq t \leq t_2 \\ D^2f(t) & 0 \leq t \leq t_1 \quad \text{or} \quad t_2 \leq t \leq 1, \end{cases}$$

then P is the set of functions which are concave on $[t_1, t_2]$ and convex elsewhere. The applications of these sample characterizations should be abundantly clear.

In principle we desire to find an estimating function which belongs to P . Clearly, in general, there will be many possibilities. In order to ensure that our estimate is also as smooth as reasonably possible, however, we select a function satisfying the following criterion:

$$\text{Minimize } \int_0^1 (f^{(m)}(t))^2 dt$$

subject to

2.1

$$(a) \quad f \in W_m, \quad f(t) \geq 0 \text{ for every } t \in [0,1]$$

$$(b) \quad \alpha_i \leq f(t_i) \leq \beta_i, \quad i = 1, 2, \dots, n$$

Constraints 2.1a are, of course, simply that $f \in P$. The second constraint 2.1b deserves a bit more comment.

Suppose our data points are $\{(t_i, y_i); i = 1, \dots, n\}$. A function f in W_m which coincides exactly with a polynomial of degree $2m-1$ on each interval $[t_i, t_{i+1}]$ and which minimizes $\int_0^1 (f^{(m)}(t))^2 dt$ is called a *polynomial spline* of degree $2m-1$ and the $t_i, i = 1, 2, \dots, n$ are the *knots* of this spline.

A well-studied problem in approximation theory is to find a solution of the following optimization problem:

$$\text{Minimize } \int_0^1 (f^{(m)}(t))^2 dt \text{ subject to}$$

2.2

$$(a) \quad f \in W_m$$

$$(b) \quad f(t_i) = y_i, \quad i = 1, 2, \dots, n.$$

The solution f , called an *interpolating spline* is a spline as just defined.

In contrast the problem:

$$\text{Minimize } \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int_0^1 (f^{(m)}(t))^2 dt$$

2.3

$$\text{with } \lambda > 0 \text{ fixed and subject to (a) } f \in W_m$$

also has a spline function solution called a *smoothing spline*. See Kimeldorf and Wahba (1970) or Cogburn and Davis (1974) for details.

A smoothing spline intermediate between 2.2 and 2.3 solves the following problem:

$$2.4 \quad \text{Minimize } \int_0^1 (f^{(m)}(t))^2 dt \quad \text{subject to}$$

$$\alpha_i \leq f(t_i) \leq \beta_i, \quad i = 1, 2, \dots, n.$$

See Attéia (1968) for more details.

It will be seen that Problem 2.1 is the isotonic version of Problem 2.4. Our results show that a solution to 2.1 exists and is a polynomial spline of degree $2m-1$. The Hilbert space methods used to solve 2.1 cannot be used to solve the isotonic version of 2.3 because the objective function is not a Hilbert space norm squared. Using convex programming one of the authors has solved the isotonic version of 2.3 for $F = D$, but the general problem remains unsolved.

3. A Restricted Isotonic Spline. We consider first a restricted problem. Let F be a continuous linear map of W_{m-1} into L_2 which commutes with differentiation. Denote the partial order defined by F on W_m by \gg , and assume there exists a function $g \in W_m$ satisfying $\alpha_i < g(t_i) < \beta_i$ for $i = 1, 2, \dots, n$ with $(Fg)(t) \geq \epsilon > 0$ for every $t \in [0, 1]$. The continuity of DF implies there exists a $d > 0$ such that

$$\|DFf\|_{L_2} < d \|D^m f\|_{L_2} \quad \text{for every } f \in W_m.$$

Take $N \geq d^2 \|D^m g\|_{L_2}^2 / \epsilon^2$. In what follows, we shall typically assume $n \geq m$.

Theorem 3.1. Under the conditions just given, the problem:

$$\begin{aligned}
 & \text{Minimize } \int_0^1 (f^{(m)}(t))^2 dt \quad \text{subject to} \\
 & \text{a. } f \in W_m \\
 & \text{b. } f \gg 0 \\
 & \text{c. } \alpha_i \leq f(t_i) \leq \beta_i, \quad i = 1, 2, \dots, n \\
 & \text{d. } Ff(j/N) \geq \epsilon, \quad j = 0, 1, 2, \dots, N,
 \end{aligned}$$

3.1

has a solution which is a polynomial spline of degree $2m-1$ with possible knots at t_i , $i = 1, 2, \dots, n$ and at j/N , $j = 0, 1, \dots, N$.

In order to establish Theorem 3.1, we need to establish some preliminary results. The straightforward proof of Lemma 3.2 is omitted.

Lemma 3.2. $\|D^m f\|_{L_2} \leq \|D^m g\|_{L_2}$ and $Ff(j/N) \geq \epsilon$, $j = 0, 1, \dots, N$ together imply that $f \gg 0$.

We next define

$$\Lambda_1 = \{u_i \in W_m, i = 1, 2, \dots, n: \langle u_i, f \rangle = -f(t_i) \text{ for every } f \in W_m\}$$

$$\Lambda_2 = \{v_i \in W_m, i = 1, 2, \dots, n: \langle v_i, f \rangle = f(t_i) \text{ for every } f \in W_m\}$$

and finally

$$\Lambda_3 = \{w_i \in W_m, i = 0, 1, \dots, N: \langle w_i, f \rangle = -F f(j/N) \text{ for every } f \in W_m\}.$$

Take $L = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$ and define $p: L \rightarrow R$ by

$$\begin{aligned} p(u_i) &= -\alpha_i & u_i &\in \Lambda_1 \\ p(v_i) &= \beta_i & v_i &\in \Lambda_2 \\ p(w_i) &= -\varepsilon & w_i &\in \Lambda_3 . \end{aligned}$$

We define a convex subset C of W_m by

$$C = \{f \in W_m : \text{for every } \ell \in L, \langle \ell, f \rangle \leq p(\ell)\} .$$

That is to say, C is just the set of functions in W_m which satisfy the constraints.

Let T be a continuous linear map of W_m onto L_2 and suppose $\text{Ker } T$ is of finite dimension. (In the case of the application we have in mind, $T = D^m$ so that both conditions are trivially satisfied.) We thus have in mind finding $s \in C$ satisfying $\|Ts\| = \min_{f \in C} \|Tf\|$. Such an element s will be called the generalized spline function relative to T in the set C .

Next let $\tilde{C} = \{f \in W_m : \langle \ell, f \rangle \leq 0 \text{ for every } \ell \in L\}$. We posit four structure axioms and, in turn state two theorems.

H1. $\tilde{C} \cap (\text{Ker } T) = \{0\}$

H2. The subset $L \subset W_m$ is weakly compact and the map p is a continuous map of L into R .

H3. $C \cap (\text{Ker } T)$ is empty.

H4. $I = \{f \in W_m : \text{for every } \ell, \langle \ell, f \rangle < p(\ell)\}$ is not empty.

Under H1, Attéia (1968) proves Theorem 3.3.

Theorem 3.3. Under H1, there is at least one spline function relative to T in the convex set C.

We refer the reader to Atteia for proof. If, in addition, we consider the remaining axioms, we have the following result due to Laurent (1969).

Theorem 3.4. Under H1-4, the element $s \in C$ is a spline function (relative to T in C) if and only if

$$-T^*Ts \in \overline{CC}(F_s)$$

where

$$F_s = \{\ell \in L: \langle \ell, s \rangle = p(\ell)\}$$

and $\overline{CC}(F_s)$ is the smallest closed convex cone with vertex 0 containing F_s and T^* is the adjoint of T.

We are now in a position to apply Lemma 3.2 and Theorems 3.3 and 3.4 to the present case.

Proof of Theorem 3.1: At the optimum s , $\|D^m s\|_{L_2} \leq \|D^m g\|_{L_2}$ and so by Lemma 3.2, the spline function must satisfy all the constraints of L, and hence $s \gg 0$. We need only to verify axioms H1-4 to guarantee the existence of s via Atteia's existence theorem and its characterization via Laurent's characterization theorem.

In our case, $T = D^m$, so that $\text{Ker } T$ is the set of polynomials of degree $m-1$. Every function in \tilde{C} is zero at all points t_i , $i = 1, 2, \dots, n$. Hence as soon as $n > m-1$, $\tilde{C} \cap \text{Ker } T = \{0\}$. Hence H1 holds for $n \geq m$. For the L described, L is finite hence compact in any topology. This implies p is continuous so that H2 is satisfied.

Next we consider H3. C is the set of functions which satisfy the constraints, while $\text{Ker } T$ is the set of polynomials of degree $m-1$. Hence if $C \cap (\text{Ker } T)$ is not empty, then the optimal solution, s , independent of the Theorems of Atteia and Laurent, will be a polynomial of degree $m-1$. A polynomial of degree $m-1$ is trivially a polynomial spline of degree $2m-1$. Hence, we may assume $C \cap (\text{Ker } T)$ is empty.

Finally, $I = \{f \in W_m : \langle \ell, f \rangle < p(\ell) \text{ for every } \ell\}$ contains the function g of the conditions to Theorem 3.1 so that I is not empty. Thus provided $C \cap \text{Ker } T$ is empty, the characterizing Theorem 3.4 applies. Thus, $-T^*Ts$ belongs to the closed convex cone generated by all $\ell \in L$ corresponding to active constraints. In other words,

$$T^*Ts = - \sum_{\ell \in L} d_\ell \ell$$

with $d_\ell \geq 0$ and $d_\ell = 0$ when ℓ is not an active constraint. Recall that all $\ell \in L$ are point evaluation functionals. Thus $T^*T s$ is zero at all points except those corresponding to active constraints (which become knots). Between knot points, $T^*Ts = 0$, i.e. $D^{2m}s = 0$. Hence, s is a polynomial of degree $2m-1$ between knot points. □

4. The General Isotonic Spline. We have already mentioned the general isotonic spline with possibly an infinite number of knots able to occur anywhere. The following result makes the details precise.

Theorem 4.1. Let \gg be the partial order defined by $F: W_{m-1} \rightarrow L_2$. If there exists $g \in W_m$ satisfying

$$Fg(t) > 0 \quad t \in [0,1]$$

and

$$\alpha_i < g(t_i) < \beta_i \quad i = 1, 2, \dots, n,$$

then the problem:

$$\text{Minimize } \int_0^1 (f^{(m)}(t))^2 dt$$

4.1

subject to a) $f >> 0$

$$b) \alpha_i \leq f(t_i) \leq \beta_i \quad i = 1, 2, \dots, n$$

has a solution which is a polynomial spline of degree $2m-1$. Knots are located (potentially) at data points, t_i , and, in exceptional cases, at a countable number of points elsewhere.

Proof: The proof follows the lines of argument for Theorem 3.1. We take $T = D^m$ and define

$$\Lambda_3 = \{w_t \in W_m, t \in [0,1]: \langle w_t, f \rangle = \int_0^1 f(t) dt \text{ for every } f \in W_m\} .$$

Let $p(w) = 0$ for all $w \in \Lambda_3$. The key requirement that Λ_3 be weakly compact holds in this case and p is continuous. The remainder follows as before. \square

5. Two examples. Passow (1974) and Passow and Roulier (1977) presented some initial results concerning monotone and convex splines. If (t_i, y_i) , $i = 1, 2, \dots, n$ represents the data points, they were interested in obtaining piecewise monotone or convex interpolating splines. Theorems 3.1 and 4.1 may be used to extend these results. We assume $t_1 < t_2 < \dots < t_n$.

Theorem 5.1. If $y_1 < y_2 < \dots < y_n$. With arbitrarily small error bounds, there is a globally monotone cubic smoothing spline of the form of Theorems 3.1 and 4.1 minimizing $\int_0^1 (f''(t))^2 dt$.

Proof: We need establish that there exists a suitable g satisfying the assumptions of Theorem 3.1. Let

$$\phi(t) = \begin{cases} \exp(1/(t^2 - 1)) & |t| < 1 \\ 0 & |t| \geq 1 \end{cases}$$

$\phi(t)$ is known to be infinitely differentiable on the real line and hence $\phi \in W_m$, $m \geq 1$. Let $\phi_1(t) = \int_{-\infty}^t \phi(u) du$. ϕ_1 is in C^∞ and it is clear that

$$\phi_1(t) = \begin{cases} 0 & t \leq -1 \\ k = \int_{-1}^1 \phi(u) du & t \geq 1 \end{cases}$$

and, of course, $D\phi_1(t) \geq 0$ for all t . When the conditions of this proposition hold, adding together suitably scaled translates of ϕ_1 will give a function $f_1 \in C^\infty$ with $Df_1(t) \geq 0$ for every t which satisfies $f_1(t_i) = y_i$, $i = 1, 2, \dots, n$. If ϵ is the error bound on y_i , the function $g_1(t) = f_1(t) + \epsilon t$, $t \in [0, 1]$ satisfies the hypotheses of Theorem 3.1 and 4.1 with $F = D$ and $m = 2$. \square

Theorem 5.2. If $\frac{y_{i+1}-y_i}{t_{i+1}-t_i} < \frac{y_{i+2}-y_{i+1}}{t_{i+2}-t_{i+1}}$ for $i = 1, 2, \dots, n-2$, and the y_i have arbitrarily small error bounds, then there exists a globally convex quintic smoothing spline of the form of Theorems 3.1 and 4.1 minimizing $\int_0^1 (f''''(t))^2 dt$.

Proof: From the function $\phi(t)$ of Theorem 5.1, we define

$$\phi_2(t) = \int_{-\infty}^t \left[\int_{-\infty}^x \phi(u) du \right] dx .$$

When the conditions of the theorem hold, adding together suitably scaled translates of $\phi_2(t)$ will give a function $f_2 \in C^\infty$ with $D^2 f_2(t) \geq 0$ for all t which satisfies $f_2(t_i) = y_i$, $i = 1, 2, \dots, n$. If ϵ is the maximum error bound on the y_i , the function $g_2(t) = f_2(t) + \frac{\epsilon t^2}{2}$, $t \in [0, 1]$ meets the assumptions of Theorems 3.1 and 4.1 with $F = D^2$ and $m = 3$. \square

6. Statistical Interpretation. In the foregoing sections, we describe the isotonic spline purely in terms of its mathematical character. We now turn to its statistical nature. We consider a statistical model of the form

$$6.1 \quad Y(t) = f(t) + e(t), \quad t \in [0, 1],$$

where $Y(t)$ is an observation at location t , $f(t)$ is a function to be estimated and $e(t)$ is an error. Of course, $[0, 1]$ can be replaced with any finite interval. We have in mind to discuss several possible error structures.

In our first type of error structure, we consider a data set $(t_i, Y(t_i))$, $i = 1, 2, \dots$ such that the t_i are dense on the interval $[0, 1]$ and such that $e(t_i)$ form an i.i.d. sequence of independent errors. We assume, first of all, that the support of the common density of $e(t)$ is a finite interval, say

$[-e_1, e_2]$ containing 0. There is no a priori need to consider the density to be symmetric or with mean 0. We note that $Y(t_i) + e_1 = \beta_i$ and $Y(t_i) - e_2 = \alpha_i$ forms a 100% confidence interval for $f(t_i)$. We have the following lemma.

Lemma 6.1. Suppose the model 6.1 holds with $e(t_i)$, $i = 1, 2, \dots$ a sequence i.i.d. random variables with support as described above. Suppose further that $\{t_i : i = 1, 2, \dots\}$ is dense in $[0, 1]$. Then for $\eta > 0$ and any interval (t, t') , there is a t_i and t_j in (t, t') such that

$$\beta_i - f(t_i) < \eta \text{ almost surely}$$

and

$$f(t_j) - \alpha_j < \eta \text{ almost surely.}$$

Proof: We consider the subsequence of t_i 's falling in the interval (t, t') . We relabel this sequence t_i and we note that the subsequence, $e(t_i)$, of random errors corresponding to the $t_i \in (t, t')$ again forms an i.i.d. sequence. Consider the interval $(-e_1, -e_1 + \eta)$. Since the support of $e(t)$ is $(-e_1, e_2)$, the set $(-e_1, -e_1 + \eta)$ has positive probability. Hence with probability one, for sufficiently large i , $e(t_i) \in (-e_1, -e_1 + \eta)$. Thus

$$\beta_i = Y(t_i) + e_1 = f(t_i) + e(t_i) + e_1 < f(t_i) - e_1 + \eta + e_1 = f(t_i) + \eta.$$

Similarly for α_j . □

We may now state a consistency theorem.

Theorem 6.2. Let $F = D$ be the operator, so that we have a monotone non-decreasing spline. Suppose further that α_i, β_i are as defined above and that $\{t_i\}$ are dense in $[0,1]$. Finally let $f \in W_m$ be non-decreasing, then s_n , the isotonic spline based on $(t_1, Y(t_1)), \dots, (t_n, Y(t_n))$, exists and

$$s_n \xrightarrow{\text{almost uniformly}} f \text{ with probability one.}$$

Proof: Since $\alpha_i < f(t_i) < \beta_i$ w.p. 1, f is the g needed to guarantee the existence of s_n in Theorems 3.1 and 4.1.

Let $\epsilon > 0$ be given. Since f is continuous on $[0,1]$, f is uniformly continuous. Corresponding to $\epsilon/3$ there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/3$ whenever $|x-y| \leq \delta$. We divide $[0,1]$ into consecutive intervals, I_j , of length δ except for possibly the last which may be shorter than δ .

Consider the interval I_j and let its endpoints be $x_j < x_{j+1}$. Since both f and s_n are non-decreasing on $[x_j, x_{j+1}]$,

$$|f(t) - s_n(t)| \leq \max\{f(x_{j+1}) - s_n(x_j), s_n(x_{j+1}) - f(x_j)\}, \quad t \in I_j.$$

Consider first $f(x_{j+1}) - s_n(x_j)$.

$$\begin{aligned} f(x_{j+1}) - s_n(x_j) &\leq f(x_{j+1}) - f(x_j) + f(x_j) - s_n(x_j) \leq \frac{\epsilon}{3} \\ &\quad + f(x_j) - s_n(x_j), \end{aligned}$$

since $|x_{j+1} - x_j| = \delta$. Now $s_n(x_j) \geq s_n(t_i) \geq \alpha_i$ for every $t_i \leq x_j$. For the interval, $(x_{j-1}, x_j]$, we may choose an i such that $t_i \in (x_{j-1}, x_j]$ and $f(t_i) - \alpha_i < \epsilon/3$. This happens with probability one by Lemma 6.1. Thus since $s_n(x_j) \geq s_n(t_i) \geq \alpha_i$

$$f(x_j) - s_n(x_j) \leq f(x_j) - s_n(t_i) \leq f(x_j) - \alpha_i \leq f(x_j) - f(t_i) \\ + f(t_i) - \alpha_i .$$

But $f(x_j) - f(t_i) \leq \epsilon/3$ since $|x_j - t_i| < |x_j - x_{j-1}| = \delta$ and $f(t_i) - \alpha_i < \epsilon/3$ with probability one by above. Hence $f(x_{j+1}) - s_n(x_j) \leq 3\epsilon/3 = \epsilon$ with probability one. Similarly $s_n(x_{j+1}) - f(x_j) \leq \epsilon$ with probability one so that for n sufficiently large

$$|f(t) - s_n(t)| \leq \epsilon, \quad t \in I_j, \quad \text{w.p. 1.}$$

Clearly this convergence holds for all I_j except possibly the first and last, hence except on a set of measure less than 2δ . The almost uniform convergence is clear. \square

This result may be extended by the following Corollary.

Corollary 6.3. Suppose the operator F in Theorem 6.2 is replaced by any of those in Examples 2.2, 2.3, 2.4 or 2.5, and suppose $Ff(t) \geq 0$. If the remaining conditions of Theorem 6.2 hold, then s_n exists and

$$s_n \xrightarrow{\text{almost uniformly}} f \quad \text{with probability one .}$$

Proof: If $F = -D$, the result follows in an obvious parallel to Theorem 6.2. If $F = D^2$, the result follows since f and s_n will be first non-increasing and then non-decreasing. The results of Theorem 6.2 can be applied to each part individually. Similarly for the remaining cases. \square

In the foregoing results, the convergence is uniform except possibly near the end points 0 and 1. We remark here that by suitably regularizing the behavior of s_n outside the interval $(\min t_i, \max t_i)$ we can extend the uniformity of convergence to the entire interval, $[0,1]$. This can be done in several obvious ways which we shall not detail here.

The regression model with $e(t)$ having bounded support may be viewed with suspicion by those used to conventional models with normal errors. However, much data these days are collected with digital instrumentation which almost implies bounded range on the errors. We also point out that at any given data point, $(t_i, Y(t_i))$, the error bounds α_i, β_i do not change as a function of n , i.e., we are not supposing increasingly accurate bounds.

Of course, the fact that the support is bounded allows us to give 100% error bounds which implies that $f(t_i)$ always falls in (α_i, β_i) , which, in turn, guarantees the existence of s_n . If the support is unbounded, as with normal errors, finite 100% error bounds are impossible. A natural suggestion is to replace the 100% error bounds with $\alpha\%$ bounds for some $\alpha < 100$. The problem is that, with probability one, $f(t_i)$ will fall outside the error bounds for some i , and, hence, we are not even guaranteed the existence of a sequence of splines much less their consistency. Hence, the consistency in the style of Theorem 6.2 for unbounded support is a moot point.

Not all is hopeless, of course, in this case. For large sets of data, about $\alpha\%$ of the intervals are not going to contain $f(t)$ and hence may even be inconsistent with isotonicity. A practical procedure may be to discard a percentage of the bounds (say no more than $\alpha\%$) which prevent us from fitting an isotone spline, and use the remainder to fit the spline. Alternatively, we could borrow a leaf from Barlow et al (1972) and simply "pool adjacent violators". Hence extend the length of interval for a violating interval by "averaging" with an adjacent non-violating interval.

We may also note that the situation where there are many data points is not really the optimal situation for use of the isotonic spline anyway. If there is much noisy data, a conventional smoothing spline is appropriate. And of course in a noise-free data situation, an ordinary interpolating spline is appropriate. It is in the context of a relatively sparse data set that the added knowledge of isotonicity will allow for the relatively largest improvement in efficiency.

The regression model, 6.1 of course, is not the only possible statistical application. If $(\alpha_{in}, \beta_{in})$ represent strongly consistent estimates of $f(t_i)$, i.e. $\alpha_{in}, \beta_{in} \rightarrow f(t_i)$ with probability 1 as $n \rightarrow \infty$, then clearly $s_n(t_i) \rightarrow f(t_i)$ w.p.1. Clearly if the t_i 's become dense in $[0,1]$ as n increases and if a suitable sequence of s_n exists, then for Examples 2.2 through 2.5, s_n will be a strongly consistent pointwise estimator of f . Using the methods of the Glivenko-Cantelli theorem, the pointwise consistency can be extended to almost uniform consistency.

Finally, we note in some circumstances it is possible to construct confidence bands (rather than just confidence intervals). The cumulative periodogram, for example, is an isotone function which admits confidence bands (c.f. Box and Jenkins (1970)). A confidence band amounts to an uncountable set of constraints, but for which only a finite number need be active. Clearly, if the confidence bands converges to f as n increases, we again may construct a strongly consistent estimator of f using the appropriate isotonic spline.

REFERENCES

- Attéia, M (1968), "Fonctions (spline) définies sur un ensemble convexe",
Num. Math., 12, 192-210.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972),
Statistical Inference under Order Restrictions, John Wiley and Sons, New
York.
- Barlow, R. E., Marshall, A. W., and Proschan, F. (1963), "Properties of
probability distributions with monotone hazard rate", *Ann. Math. Statist.*,
34, 375-389.
- Box, G. E. P. and Jenkins, G. M. (1970), *Time Series Analysis Forecasting and
Control*, Holden-Day, San Francisco.
- Cogburn, R. and Davis, H. T. (1974), "Periodic splines and spectral estimation",
Ann. Statist., 2, 1108-1126.
- Kimeldorf, G. S. and Wahba, G. (1970), "A correspondence between Bayesian
estimation on stochastic processes and smoothing by splines," *Ann. Math.
Statist.*, 41, 495-502.
- Laurent, P. J. (1969), "Construction of spline functions on a convex set",
Approximation with Special Emphasis on Spline Functions, (ed. I. J.
Schoenberg) 415-446, Academic Press, New York.
- Marshall, A. W. and Proschan, F. (1965), "Maximum likelihood estimation for
distributions with monotone failure rate", *Ann. Math. Statist.*, 36,
69-77.
- Passow, E. (1974), "Piecewise monotone spline interpolation", *J. Approx. Theory*,
12, 240-241.
- Passow, E. and Roulier, J. A. (1977), "Monotone and convex spline interpolation",
SIAM J. Numer. Anal., 14, 904-909.

- Robertson, T. (1967), "On estimating a density which is measurable with respect to a σ -lattice," *Ann. Math. Statist.* 38, 482-493.
- Wegman, E. J. (1970a), "Maximum likelihood estimation of a unimodal density function", *Ann. Math. Statist.*, 41, 457-471.
- Wegman, E. J. (1970b), "Maximum likelihood estimation of a unimodal density, II," *Ann. Math. Statist.*, 41, 2169-2174.
- Wright, I. W. (1977), "Splines in statistics," Institute of Statistics Mimeo Series No. 1146, Chapel Hill, North Carolina.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Isotonic, Convex and Related Splines		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
7. AUTHOR(s) Ian W. Wright and Edward J. Wegman		6. PERFORMING ORG. REPORT NUMBER Mimco Series No. 1145
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514		8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2840
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE November 1977
		13. NUMBER OF PAGES 21 pages
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release: Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) isotone, isotonic, convex, partial order, spline, isotonic spline, smoothing spline, interpolating spline, regression, strong consistency		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In this paper, we consider the estimation of isotonic, convex or related functions by means of splines. It is shown that certain classes of isotone or convex functions can be represented as a positive cone embedded in a Hilbert space. Using this representation, we give an existence and characterization theorem for isotonic or convex splines. Two special cases are examined showing the existence of a globally monotone cubic smoothing spline and a globally convex quintic smoothing spline. Finally, we examine a regression problem and show that the isotonic-type of spline provides a strongly consistent solution.		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

(20.) We also point out several other statistical applications.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)