ON THE PROBLEMS OF SMOOTHING
AND NEAR-INTERPOLATION

SCOTT N. KERSEY

ABSTRACT. In the first part of this paper we apply a saddle point theorem from convex analysis to show that various constrained minimization problems are equivalent to the problem of smoothing by spline functions. In particular, we show that near-interpolants are smoothing splines with weights that arise as Lagrange multipliers corresponding to the constraints in the problem of near-interpolation. In the second part of this paper we apply certain fixed point iterations to compute these weights. A similar iteration is applied to the computation of the smoothing parameter in the problem of smoothing.

1. INTRODUCTION

Let $X := L_2^{(m)}([a, b] \longrightarrow \mathbb{R})$ denote the Sobolev space of functions $f : [a, b] \longrightarrow \mathbb{R}$ whose derivatives $D^{j-1}f$ are in the Lebesgue spaces $L_2([a, b] \longrightarrow \mathbb{R})$ for $j=1{:}m$, $m \geq 1$. Let t be a sequence of *data sites* $a = t_1 \leq t_2 \leq \cdots \leq t_n = b$ with multiplicity at most m, i.e., $m_i := \#\{t_k : k \leq i \text{ and } t_k = t_i\} \leq m$ for $i=1{:}n$. For these fixed data sites, let Λ be the *data map* $\Lambda : f \longmapsto \Lambda f = (\lambda_i f)$ with $\lambda_i : f \longmapsto D^{m_i-1}f(t_i)$. In particular, $\lambda_i f = f(t_i)$ when $m_i = 1$. Given a sequence of *data* $z \in \mathbb{R}^n$ and nonnegative *weights* $w \in \mathbb{R}_{\geq 0}^n$, we define the functionals

$$J : f \longmapsto \int_a^b |D^m f(u)|^2 \, \mathrm{d}u,$$
$$E_i : f \longmapsto |\lambda_i f - z_i|,$$
$$E_w : f \longmapsto \Big(\sum_{i=1}^n w_i \, E_i(f)^2 \Big)^{1/2}.$$

Let $\rho > 0$ (the *smoothing parameter*), $\varepsilon \in \mathbb{R}_{>0}^n$ (a sequence of *tolerances* ε_i), $S > 0$, and let q and r in $\mathbb{R}^n$ be such that $q_i < r_i$ for all i. We are interested in the following problems:

(A) $\quad\quad \underset{f \in X}{\text{minimize}} \ \ J(f) + \rho \, E_w(f)^2,$

(B) $\quad\quad \underset{f \in X}{\text{minimize}} \ \{J(f) : E_i(f) \leq \varepsilon_i, \ i=1{:}n\},$

(C) $\quad\quad \underset{f \in X}{\text{minimize}} \ \{J(f) : E_w(f) \leq S\},$

(D) $\quad\quad \underset{f \in X}{\text{minimize}} \ \{J(f) : q_i \leq \lambda_i f \leq r_i, \ i=1{:}n\}.$

(A) is the problem of "smoothing," which has been studied extensively in the literature, e.g., [1], [13] and [16]. It is generalized here to include the smoothing of derivatives when there is repetition of data sites t_i in the sequence t. Solutions to (A) are the so-called "smoothing splines." (B) is the problem of "near-interpolation," which has been the focus of this author's work for some time (see [8] and [9]). The goal is to find the smoothest curves that pass near the data to within tolerances ε_i. One of the goals of this paper is to show that problem (B) is equivalent to problem (A) under certain conditions. In (C) the bound is on the cumulative error, i.e., the "weighted sum of the error" is bounded. This is in contrast to (B), where the "local" error is bounded. In [13] it was shown that problem (C) is equivalent to problem (A) for a certain ρ depending on S, with the functionals λ_i being point evaluation. We reprove this result in the setup here. A similar result was shown in [14], where problem (A) was shown to be equivalent to the problem

$$\operatorname*{minimize}_{f} \{E_w(f) : J(f) \leq S\}.$$

These results motivated this author to conjecture that problems (A) and (B) are equivalent, as we show here. The inequality-constrained spline problem (D) has been studied in more or less generality in [2], [5], [10] and [11], and is closely connected to problem (B). Indeed, they are equivalent when $z_i = (q_i + r_i)/2$ and $\varepsilon_i = (r_i - q_i)/2$. Problem (D) does not, however, generalize to vector-valued data in the same way as problem (B) (see [9] for details on near-interpolation by curves).

The solutions to each of these problems are piecewise polynomials, as is easy to observe. Indeed, suppose that g solves one of these problems. Then g necessarily solves the problem

$$\operatorname*{minimize}_{f \in X} \{J(f) : \lambda_i f = y_i, \ i=1{:}n\}$$

of "best interpolation" to the data $y_i := \lambda_i g$. It is well-known (see [4]) that this variational problem characterizes polynomial spline interpolation.

In Section 2 of this paper we show that problems (B), (C) and (D) are equivalent to problem (A) (i.e., they have the same solution set) for certain ε depending on w, S depending on ρ, and a and b depending on w, respectively. The proofs are based on *duality* theory in convex analysis, whereby the solutions to (constrained) convex problems correspond to saddle points of particular (unconstrained) dual problems. In particular, the objective functional in (A) resembles the "Lagrangian" in the dual formulation of problem (B), and the "Lagrange multipliers" corresponding to the constraints in (B) turn out to be the weights in (A), with $\rho = 1$.

Section 3 concerns the estimation of the weights and smoothing parameter in problem (A) and the computation of approximate solutions to problems (B) and (C) based on certain fixed point iterations, similar to an iteration given in [15]. In particular, in Section 3.1 the iterations generate a sequence of weights (Lagrange multipliers) for problem (A). The fixed points of these iterations correspond to solutions to problem (B). In Section 3.2 a similar iteration is used to estimate the smoothing parameter $\rho = \rho(S)$ in (A) for a prescribed S in problem (C). The iteration converges linearly, but in practice not too many iterations are required, and it has the advantage of being simple to program. We do not achieve quadratic convergence as the authors of [13] and [6] do. In Section 3.3, optimal curve fits are computed for various examples.

2. Duality

2.1. Problems (A) and (B). Let

$$(2.1) \qquad \mathcal{L}_B(f, w) := J(f) + \rho \sum_i w_i (E_i(f)^2 - \varepsilon_i^2)$$

denote the **Lagrangian** associated to (B) (see [3], pp. 175–177) and let

$$F_w : f \longmapsto J(f) + \rho\, E_w(f)^2.$$

Then,

$$F_w(f) = \mathcal{L}_B(f, w) + \rho \sum_i w_i \varepsilon_i^2,$$

and problem (A) can be restated:

$$\underset{f \in X}{\text{minimize}} \ \ F_w(f).$$

Assumption 2.1. There exists $f \in X$ such that $E_i(f) < \varepsilon_i$, $i=1{:}n$.

This "constraint qualification" is known as the "Slater condition" for this problem. It holds iff $\varepsilon_i > 0$ for $i=1{:}n$. The following saddle point result is adapted from [3], Theorem 1.2, page 175.

Lemma 2.2. *Let t, z and ε be given. Then $f_B \in X$ solves (B) if (and "only if" when Assumption 2.1 holds) (f_B, w) is a saddle point of $\mathcal{L}_B$ for some $w \geq 0$ (with $\rho = 1$), meaning that for all $f \in X$ and $v \in \mathbb{R}_{\geq 0}^n$*

$$\mathcal{L}_B(f_B, v) \leq \mathcal{L}_B(f_B, w) \leq \mathcal{L}_B(f, w).$$

Proposition 2.3. *Let t and z be given.*

(i) Let $\varepsilon > 0$. Suppose that Assumption 2.1 holds, and that f_B solves (B). Then f_B solves (A) with weights $w \geq 0$ such that (f_B, w) is a saddle point for $\mathcal{L}_B$ (with $\rho = 1$).

(ii) Let $w \geq 0$ and $\rho = 1$. Suppose that f_A solves (A). Let $\varepsilon_i := E_i(f_A)$ if $E_i(f_A) > 0$ and $w_i > 0$, and let $0 < \varepsilon_i \geq E_i(f_A)$ otherwise. Then f_A solves (B) for this ε.

Proof. For (i), it follows by Lemma 2.2 that for all $f \in X$

$$F_w(f_B) = \mathcal{L}_B(f_B, w) + \sum_i w_i\, \varepsilon_i^2 \leq \mathcal{L}_B(f, w) + \sum_i w_i\, \varepsilon_i^2 = F_w(f).$$

For (ii), let $\hat{w}_i := w_i$ if $E_i(f_A) > 0$ and let $\hat{w}_i := 0$ otherwise. By the choice of ε_i, it follows that $E_i(f_A)^2 - \varepsilon_i^2 \leq 0$ and $\hat{w}_i(E_i(f_A)^2 - \varepsilon_i^2) = 0$ for all i. Hence, for all $v \in \mathbb{R}_{\geq 0}^n$

$$\mathcal{L}_B(f_A, v) = J(f_A) + \sum_i v_i\, (E_i(f_A)^2 - \varepsilon_i^2)$$

$$\leq J(f_A) + \sum_i \hat{w}_i\, (E_i(f_A)^2 - \varepsilon_i^2) = \mathcal{L}_B(f_A, \hat{w}).$$

Moreover, if $E_i(f_A) = 0$ for some i, then f_A solves problem (A) with the corresponding point removed (i.e., with w_i set equal to 0 for this i). So $f_A \in \operatorname{argmin}_f F_{\hat{w}}$, and thus

$$\mathcal{L}_B(f_A, \hat{w}) = F_{\hat{w}}(f_A) - \sum_i \hat{w}_i \varepsilon_i^2 \leq F_{\hat{w}}(f) - \sum_i \hat{w}_i \varepsilon_i^2 = \mathcal{L}_B(f, \hat{w})$$

for all $f \in X$. By Lemma 2.2, f_A solves (B). $\square$

2.2. **Problems (A) and (C).** Let

$$\mathcal{L}_C(f, \rho) := J(f) + \rho \left(E_w(f)^2 - S^2 \right)$$

denote the Lagrangian associated to problem (C) and let

$$F_\rho : f \longmapsto J(f) + \rho \, E_w(f)^2.$$

Then,

$$F_\rho(f) = \mathcal{L}_C(f, \rho) + \rho \, S^2,$$

and problem (A) can be restated:

$$\underset{f \in X}{\text{minimize}} \ \ F_\rho(f).$$

Assumption 2.4. There exists $f \in X$ such that $E_w(f) < S$.

Assumption 2.4 is Slater's condition for problem (C). It holds iff $S > 0$.

Lemma 2.5. *Let t, z and S be given. Then $f_B \in X$ solves (C) if (and "only if" when Assumption 2.4 holds) (f_C, ρ) is a saddle point for $\mathcal{L}_C$ for some $\rho \geq 0$, meaning that for all $f \in X$ and $q \geq 0$*

$$\mathcal{L}_C(f_C, q) \leq \mathcal{L}_C(f_C, \rho) \leq \mathcal{L}_C(f, \rho).$$

Proposition 2.6. *Let t, z and w be given.*
  (i) Let $S > 0$. Suppose that Assumption 2.4 holds, and that f_C solves (C). Then f_C solves (A) for $\rho \geq 0$ such that (f_C, ρ) is a saddle point for $\mathcal{L}_C$.
  (ii) Let $\rho \geq 0$. Suppose that f_A solves (A). Let $S := E_w(f_A)$ if $E_w(f_A) > 0$ and $\rho > 0$, and let $0 < S \geq E_w(f_A)$ otherwise. Then f_A solves (C) for this S.

Proof. The proof is similar to that for Proposition 2.3. Part (i) follows since

$$F_\rho(f_C) = \mathcal{L}_C(f_C, \rho) + \rho \, S^2 \leq \mathcal{L}_C(f, \rho) + \rho \, S^2 = F_\rho(f)$$

for all $f \in X$. For part (ii),

$$\mathcal{L}_C(f_A, q) = J(f_A) + q \left(E_w(f_A)^2 - S^2 \right) \leq J(f_A) + \rho \left(E_w(f_A)^2 - S^2 \right) = \mathcal{L}_C(f_A, \rho),$$

since $\rho(E_w(f_A) - S) = 0$ (without loss of generality we assume that $\rho = 0$ when $E_w(f_A) = 0$), and

$$\mathcal{L}_C(f_A, \rho) = F_\rho(f_A) - \rho \, S^2 \leq F_\rho(f) - \rho \, S^2 = \mathcal{L}_C(f, \rho).$$

By Lemma 2.5, f_A solves (C). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

2.3. **Problems (A) and (D).** Let

$$\mathcal{L}_D(f, u, v) := J(f) + \sum_i \left(u_i(\lambda_i f - q_i) - v_i(\lambda_i f - r_i) \right)$$

denote the Lagrangian associated to problem (D) with $u_i \geq 0$ and $v_i \geq 0$. The constraint qualification for this problem is

Assumption 2.7. There exists $f \in X$ such that $q_i < \lambda_i f < r_i$ for $i=1{:}n$.

Assumption 2.7 holds iff $q < r$, i.e., iff $q_i < r_i$ for all i.

Lemma 2.8. *Let t, q and r be given. Then $f_D \in X$ solves (D) if (and "only if" when Assumption 2.7 holds) (f_D, u, v) is a saddle point for $\mathcal{L}_D$, meaning that for all $f \in X$ and $\hat{u}, \hat{v} \in \mathbb{R}^n_{\geq 0}$*

$$\mathcal{L}_D(f_D, \hat{u}, \hat{v}) \leq \mathcal{L}_D(f_D, u, v) \leq \mathcal{L}_D(f, u, v).$$

It follows by Lemma 2.8 (as is standard in optimization) that, for a solution f_D to (D) with multipliers u and v, $u_i = 0$ when $\lambda_i f_D \neq q_i$ and $v_i = 0$ when $\lambda_i f_D \neq r_i$. Then, with $z_i := (q_i + r_i)/2$, $\varepsilon_i := (r_i - q_i)/2$, and $w_i := (u_i + v_i)/(r_i - q_i)$, it follows that

$$\mathcal{L}_B(f_D, w) = \mathcal{L}_D(f_D, u, v)$$

with (f_D, w) a saddle point for $\mathcal{L}_B$ and (f_D, u, v) a saddle point for $\mathcal{L}_D$, with $\mathcal{L}_B$ as in (2.1).

Proposition 2.9. *Let t be given.*

(i) Suppose that $q < r$. Suppose that Assumption 2.7 holds and that f_D solves (D). Suppose that (f_D, u, v) is a saddle point for $\mathcal{L}_D$. Then f_D solves (A) with $z_i := (q_i + r_i)/2$, $w_i := (u_i + v_i)/(r_i - q_i)$ and $\rho = 1$.

(ii) Suppose that $w \geq 0$ and $\rho = 1$, and that f_A solves (A). Let $\varepsilon_i := E_i(f_A)$ if $E_i(f_A) > 0$ and $w_i > 0$, and let $0 < \varepsilon_i \geq E_i(f_A)$ otherwise. Let $q_i := z_i - \varepsilon_i$ and $r_i := z_i + \varepsilon_i$. Then f_A solves (D) for this q and r.

Proof. For (i), it follows by the particular choice of ε and z that problems (B) and (D) are equivalent. Therefore, f_D solves (B) and Assumption 2.1 holds (since Assumption 2.7 holds). As stated above, $\mathcal{L}_B(f_D, w) = \mathcal{L}_D(f_D, u, v)$ for this choice of w, and so (f_D, w) is a saddle point for $\mathcal{L}_B$ (since (f_D, u, v) is a saddle point for $\mathcal{L}_D$). Therefore, (i) follows by Proposition 2.3 (i).

For (ii), it follows by Proposition 2.3 (ii) that f_A solves problem (B), which is equivalent to problem (D) by the choice of ε, q and r. $\qquad\square$

3. Computation

3.1. **Choosing the weights in (A) to solve (B).** In this section, iterations are given for determining the weights for problem (A). The iterations are similar to one given in [15]. In particular, we show that the fixed points of the iterations are the Lagrange multipliers corresponding to the constraints in the solutions to problem (B). Hence, the resulting curves are an approximate solution to problem (B).

We begin by representing solutions to problem (A) in a finite dimensional basis. In particular, since the solutions to problems (A) through (D) are piecewise polynomial, we can choose such a basis. Here, we choose the basis-map V dual to the data map Λ that was defined in the introduction. Hence, $\Lambda V = I$, the identity matrix in $\mathbb{R}^n$, and $f = V\alpha$ when $\alpha_i = \lambda_i f$. In particular, when each distinct data site t_i occurs exactly m times in the sequence t, then Λ is a Hermite data-map (it evaluates all derivatives up to order $m - 1$), and V is the "piecewise Hermite basis-map". In this basis

$$J(f) = \int_a^b |D^m f|^2 = \int_a^b |D^m V\alpha|^2 = \alpha^T \Big(\int_a^b (D^m V)^T (D^m V) \Big) \alpha =: \alpha^T H \alpha$$

and

$$E_w(f)^2 = (\alpha - z)^T W(\alpha - z)$$

with W the diagonal matrix with entries w_i, and Problem (A) can be restated:

$$\underset{\alpha \in \mathbb{R}^n}{\text{minimize}} \ \alpha^T H \alpha + \rho(\alpha - z)^T W(\alpha - z).$$

Since H and W are symmetric, the stationary points satisfy the system of equations

$$2H\alpha + 2\rho W(\alpha - z) = 0;$$

hence

(3.1) $$(H + \rho W)\Lambda f = \rho W z.$$

Since

$$F_w(f) = \mathcal{L}_B(f, w) + \rho \sum_i w_i \varepsilon_i^2$$

with $\mathcal{L}_B$ the Lagrangian for (B), it follows that (3.1) is also a necessary condition for the solutions to problem (B).

Our goal is to estimate the weights w_i so that the solutions to problem (A) approximate a solution to problem (B) for prescribed ε_i. From (3.1) we have that

$$H\Lambda f = \rho W (z - \Lambda f).$$

On the right side we enforce the errors $|z_i - \lambda_i f| \approx \varepsilon_i$; on the left we suppose that f interpolates z, i.e., $\Lambda f = z$. Then, we have the estimate

(3.2) $$w_i := |Hz|_i/(\rho \varepsilon_i).$$

This gives a reasonable first choice for the weights. In an attempt to improve this, consider a sequence of weights (w^k) such that

(3.3) $$w_i^{(k+1)} := w_i^{(k)} E_i(f_A^{(k)})/\varepsilon_i.$$

This is similar to an iteration given in [15]. The idea is that a weight is increased if $E_i(f_A^{(k)}) > \varepsilon_i$ and decreased if $E_i(f_A^{(k)}) < \varepsilon_i$. By (3.1), iteration (3.3) is equivalent to

$$w_i^{(k+1)} := |(H\Lambda f_A^{(k)})_i|/(\rho \varepsilon_i)$$

with $f_A^{(k)}$ the solution to problem (A) corresponding to weights $w^{(k)}$.

In Proposition 3.3 we will show that the fixed points of this iteration solve problem (B). The proof requires the following two lemmas. For these, let $\ker D^m$ denote the space of polynomials of degree less than m, i.e., those functions whose m-th derivative is identically zero.

Lemma 3.1. *Assume that the weights w_i are nonnegative, and that $\rho = 1$. If $\ker D^m \cap \ker w\Lambda = \{0\}$, then the matrix $H + W$ is positive definite, hence invertible.*

Proof. Let α be an arbitrary point in $\mathbb{R}^n$, and let f be such that $f = V\alpha$. Then,

$$\alpha^T(H + W)\alpha = \alpha^T H \alpha + \alpha^T W \alpha = J(f) + \alpha^T W \alpha,$$

a sum of two positive terms. Therefore, $H + W$ is positive semi-definite. If $\alpha^T(H + W)\alpha = 0$, then $f = V\alpha \in \ker J = \ker D^m = 0$ and $w\Lambda f = W\alpha = 0$, and so $f = 0$ when $\ker D^m \cap \ker w\Lambda = \{0\}$. In this case $H + W$ is positive definite. $\square$

Lemma 3.2. *Suppose that $w^{(k)} \longrightarrow w$ in $\mathbb{R}_{\geq 0}^n$. For large enough k,*

$$\ker D^m \cap \ker w \Lambda = \{0\} \quad \Longrightarrow \quad \ker D^m \cap \ker w^{(k)} \Lambda = \{0\}.$$

Proof. To make this precise, let $\epsilon_k := \max_i |w_i^{(k)} - w_i|$. Then $\epsilon_k \longrightarrow 0$. Choose k_0 such that $\epsilon_k < w_i/2$ for all $k \geq k_0$ and i such that $w_i > 0$. Therefore, if $w_i \neq 0$, then $w_i^{(k)} \neq 0$ for all $k > k_0$, and so $\lambda_i f = 0$ when $w_i^{(k)} \lambda_i f = 0$ for any $k > k_0$. In particular, $w_i \lambda_i f = 0$. On the other hand, if $w_i = 0$, then trivially $w_i \lambda_i f = 0$. So we have that $w_i \lambda_i f = 0$ when $w_i^{(k)} \lambda_i f = 0$ for any $k > k_0$. Therefore, $\ker w^{(k)} \Lambda \subset \ker w \Lambda$ for $k > k_0$, and the claim follows. $\square$

Proposition 3.3. *Assume that $\varepsilon > 0$. Suppose that $(w^{(k)})$ is a sequence of weights such that*

$$(3.4) \qquad w_i^{(k+1)} := \begin{cases} 1, & \text{if } w_i^{(k)} = 0 \text{ and } E_i(f_A^{(k)}) > \varepsilon_i, \\ w_i^{(k)} E_i(f_A^{(k)})/\varepsilon_i, & \text{otherwise,} \end{cases}$$

with $f_A^{(k)}$ a solution to (A) corresponding to $w^{(k)}$ (and with $\rho = 1$). Then $(w^{(k)})$ has convergent subsequences. Suppose moreover that $w^{(k)}$ converges to w. If $\ker D^m \cap \ker w\Lambda = \{0\}$, then f_A^k converges to a solution f_A to problem (B) with weights w.

Proof. Let s be any function in X that satisfies the conditions $\Lambda s = z$, such as the spline interpolant. Since $f_A^{(k)}$ solves problem (A) for weights $w^{(k)}$, it follows that

$$w_i^{(k)} E_i(f_A^{(k)})^2 \le F_{w^{(k)}}(f_A^{(k)}) \le F_{w^{(k)}}(s) = J(s)$$

for each k. If also $E_i(f_A^{(k)}) > \varepsilon_i$, then $w_i^{(k)} E_i(f_A^{(k)}) \le J(s)/\varepsilon_i$, and so either $w_i^{(k+1)} = 1$ (when $w_i^{(k)} = 0$ and $E_i(f_A^{(k)}) > \varepsilon_i$), or

$$w_i^{(k+1)} = w_i^{(k)} E_i(f_A^{(k)})/\varepsilon_i \le J(s)/\varepsilon_i^2.$$

On the other hand, if $E_i(f_A^{(k)}) \le \varepsilon_i$, then

$$w_i^{(k+1)} = w_i^{(k)} E_i(f_A^{(k)})/\varepsilon_i \le w_i^{(k)},$$

and so $w_i^{(k+1)} \le w_i^{(k)}$. Therefore,

$$w_i^{(k)} \le \max\{w_i^{(1)}, 1, J(s)/\varepsilon_i^2\}$$

for all i, and so $(w^{(k)})$ is a bounded sequence in $\mathbb{R}^n$; hence, it has convergent subsequences.

This verifies the first claim. Now, suppose that $w^{(k)} \longrightarrow w$ in $\mathbb{R}^n$. Since $\ker D^m \cap \ker w\Lambda = \{0\}$, it follows by Lemmas 3.1 and 3.2 that $H + W$ and $H + W^k$ for large k are invertible. Therefore,

$$\Lambda f_A^{(k)} = (H + W^{(k)})^{-1} W^{(k)} z \longrightarrow (H + W)^{-1} W z = \Lambda f_A.$$

Hence, $\Lambda f_A^{(k)}$ converges to Λf_A in $\mathbb{R}^n$, and so $f_A^{(k)}$ converges to f_A in the n-dimensional spline space. Moreover, $E_i(f_A^{(k)}) \longrightarrow E_i(f_A)$, and so w is a fixed point of (3.4). Since the first case in (3.4) is impossible at a fixed point, it follows that $w_i = w_i E_i(f_A)/\varepsilon_i$ for all i, in which case either $\varepsilon_i = E_i(f_A) > 0$ when $w_i > 0$, or $0 < \varepsilon_i \ge E_i(f_A)$ when $w_i = 0$. By Proposition 2.3 (ii), f_A solves problem (B). $\square$

In computing solutions to problem (B) it is of particular interest to identify the inactive constraints, i.e., those constraints with a zero Lagrange multiplier (weight) at a minimizer. We would like to use iteration (3.4) for the weights, but it is not particularly effective in identifying these inactive constraints. The reason for this is that the iteration attempts to satisfy $E_i(f) = \varepsilon_i$ when updating the weights, and therefore is not well-suited to handle the inactive constraints where $E_i(f) < \varepsilon_i$. With this in mind, the following iteration accelerates the demise of constraints that are clearly inactive.

Let k denote the k-th iteration, beginning at $k = 1$, and let $f_A^{(k)}$ be the smoothing spline (solution to problem (A)) corresponding to weights $w^{(k)}$, with $\rho := 1$. With

$\beta_k := \frac{k-1}{k}$, let

(3.5)

$$w_i^{(k+1)} := \begin{cases} 1, & \text{if } w_i^{(k)} = 0 \text{ and } E_i(f_A^{(k)}) > \varepsilon_i. \\ w_i^{(k)}(E_i(f_A^{(k)})/\varepsilon_i)^k, & \text{if } E_i(f_A^{(k)}) < \beta_k\,\varepsilon_i \text{ and } E_i(f_A^{(k-1)}) < \beta_k\,\varepsilon_i. \\ w_i^{(k)}E_i(f_A^{(k)})/\varepsilon_i, & \text{otherwise.} \end{cases}$$

For $k = 1$ (the first iterate) $\beta_k = 0$; hence the second case is automatically by-passed and we revert to the third case, as in iteration (3.4). If $k = 2$, then $\beta_k = .5$, and the iteration will pass to the second case if $E_i(f_A^{(2)}) < \varepsilon_i/2$, assuming also that $w_i \neq 0$ and $E_i(f_A^{(1)}) < \beta_k\,\varepsilon_i$ at the previous iteration. The new weight is then $w_i^{(2)} = w_i^{(1)}(E_i(f_A^{(1)})/\varepsilon_i)^2$. Since $E_i(f_A^{(1)})/\varepsilon_i < \beta_k < 1$, $(E_i(f_A^{(1)})/\varepsilon_i)^2$ is potentially very small. Therefore, $w_i^{(2)}$ would be much smaller using (3.5) than it would have been by (3.4). As the iteration progresses, $\beta_k \longrightarrow 1$ (from below) and $k \longrightarrow \infty$, so that in the limit all inactive constraints default to the second case, and the corresponding weights are forced smaller at increasing orders. The third example in Section 3.3 compares the convergence of the weights using (3.4) versus (3.5).

3.2. Choosing the smoothing parameter in (A) to solve (C).

In this section we derive a fixed point iteration for computing the smoothing spline parameter ρ in problem (A) that produces solutions to problem (C). That is, we compute $\rho = \rho(S)$ for prescribed bounds S in (C). The iteration is similar to iteration (3.4) given in Section 3.1. As proved in the next proposition and as is verified in Example 3.8 in the next section, the iteration can be expected to achieve linear convergence. However, due to a relatively small growth constant in the examples tested, we obtain good results without too many iterations, typically 10-30. It is also easy to program, which makes it appealing. However, the iteration does not match the performance of the iterations given in [13] and [6], both of which achieve quadratic convergence.

Proposition 3.4. *Let $S > 0$, and let $w \geq 0$ in $\mathbb{R}^n$ be such that $\ker D^m \cap \ker w\Lambda = \{0\}$. Choose $\rho^{(1)} > 0$, and let*

(3.6) $$\rho^{(k+1)} : l = \begin{cases} 1, & \text{if } \rho^{(k)} = 0 \text{ and } E_w(f_A^{(k)}) > S, \\ \rho^{(k)}E_w(f_A^{(k)})/S, & \text{otherwise,} \end{cases}$$

with $f_A^{(k)}$ a solution to (A) corresponding to $\rho^{(k)}$. Then $(\rho^{(k)})$ converges monotonically, and can be expected to achieve linear convergence. Let ρ be the limit point and f_A a corresponding solution to (A). Then, f_A solves problem (C).

Proof. Let s be any interpolant to the data z, such as an interpolating spline. Since $f_A^{(k)}$ solves (A) for weights $w^{(k)}$, then

$$\rho^{(k)}E_w(f_A^{(k)})^2 \leq J(f_A^{(k)}) + \rho^{(k)}E_w(f_A^{(k)})^2 \leq J(s) + \rho^{(k)}E_w(s)^2 = J(s).$$

If $E_w(f_A^{(k)}) > S$, then $\rho^{(k)}E_w(f_A^{(k)}) \leq J(s)/S$, and so $\rho^{(k+1)} = \rho^{(k)}E_w(f_A^{(k)})/S \leq J(s)/S^2$; if $E_w(f_A^{(k)}) \leq S$, then $\rho^{(k+1)} = \rho^{(k)}E_w(f_A^{(k)})/S \leq \rho^{(k)}$. Therefore,

$$\rho^{(k)} \leq \max\{\rho^{(1)}, 1, J(s)/S^2\}.$$

In particular, $(\rho^{(k)})$ is a bounded sequence.

Assume first that all the weights w_i are positive. By (3.1),

$$E_i(f_A) = |\lambda_i f_A - z_i| = |(H + \rho W)^{-1} \rho W z - z|_i.$$

Moreover,

$$(H + \rho W)^{-1} \rho W z - z = (H + \rho W)^{-1}(\rho W z - (H + \rho W)z) = -(H + \rho W)^{-1} H z.$$

Therefore,

$$
\begin{aligned}
E_w(f_A)^2 &= \sum_i w_i |\lambda_i f_A - z_i|^2 \\
&= \sum_i w_i |(H + \rho W)^{-1} H z|_i^2 \\
&= z^T H (H + \rho W)^{-1} W (H + \rho W)^{-1} H z \\
&= z^T H W^{-1/2}(W^{-1/2} H W^{-1/2} + \rho I)^{-2} W^{-1/2} H z.
\end{aligned}
$$

Let $T^{-1} \theta T$ be the Schur decomposition of $W^{-1/2} H W^{-1/2}$. The matrix T is orthogonal, and, since $W^{-1/2} H W^{-1/2}$ is symmetric, θ is the diagonal matrix of eigenvalues θ_i. Moreover, since H and $W^{-1/2}$ are positive semi-definite (see Lemma 3.1), it follows that the θ_i are nonnegative. After some simplification we have

$$
\begin{aligned}
E_w(f_A)^2 &= z^T H W^{-1/2} T^{-1}(\theta + \rho I)^{-2} T W^{-1/2} H z \\
&= v^T (\theta + \rho I)^{-2} v = \sum_i v_i^2/(\theta_i + \rho)^2
\end{aligned}
$$

with $v := T W^{-1/2} H z$. By this and (3.6) we have that

$$
\begin{aligned}
\big((\rho^{(k+2)})^2 &- (\rho^{(k+1)})^2\big) S^2 \\
&= (\rho^{(k+1)} E_w(f_A^{(k+1)}))^2 - (\rho^{(k)} E_w(f_A^{(k)}))^2 \\
&= \sum_i v_i^2 \Big(\frac{(\rho^{(k+1)})^2}{(\theta_i + \rho^{(k+1)})^2} - \frac{(\rho^{(k)})^2}{(\theta_i + \rho^{(k)})^2} \Big) \\
&= \sum_i v_i^2 \Big(\frac{\rho^{(k+1)}}{\theta_i + \rho^{(k+1)}} + \frac{\rho^{(k)}}{\theta_i + \rho^{(k)}} \Big) \Big(\frac{\rho^{(k+1)}}{\theta_i + \rho^{(k+1)}} - \frac{\rho^{(k)}}{\theta_i + \rho^{(k)}} \Big) \\
&= (\rho^{(k+1)} - \rho^{(k)}) \sum_i v_i^2 \Big(\frac{\rho^{(k+1)}}{\theta_i + \rho^{(k+1)}} + \frac{\rho^{(k)}}{\theta_i + \rho^{(k)}} \Big) \frac{\theta_i}{(\theta_i + \rho^{(k+1)})(\theta_i + \rho^{(k)})}.
\end{aligned}
$$

Therefore, $\rho^{(k+2)} - \rho^{(k+1)} = C_k (\rho^{(k+1)} - \rho^{(k)})$ with

$$
\begin{aligned}
C_k = \ &\frac{1}{(\rho^{(k+2)} + \rho^{(k+1)}) S^2} \\
&\times \sum_i v_i^2 \Big(\frac{\rho^{(k+1)}}{\theta_i + \rho^{(k+1)}} + \frac{\rho^{(k)}}{\theta_i + \rho^{(k)}} \Big) \frac{\theta_i}{(\theta_i + \rho^{(k+1)})(\theta_i + \rho^{(k)})}.
\end{aligned}
$$

In particular, C_k is positive, and so the sequence $(\rho^{(k)})$ is monotonic. Since it is bounded, as shown above, it converges. Moreover, since C^k is bounded away from zero, one expects linear convergence. In Section 3.3 of this paper the growth constant is estimated experimentally for a particular data set.

For the above estimates it was assumed that the weights w_i are strictly positive, so that the inverse of W is defined. In the case that some weights are zero, one can

simply eliminate the corresponding data from the problem, and carry through the above analysis with the remaining data, whose corresponding weights are positive.

Let ρ be the limit point, i.e., $\rho^{(k)} \longrightarrow \rho$. Note that near the limit point we cannot be in the first case of (3.6), i.e., $\rho^{(k)} \neq 0$ for large k if $E_w(f_A^{(k)}) > S$. As in the proof of Proposition 3.3, $E_i(f_A^{(k)}) \longrightarrow E_i(f_A)$. Therefore $E_w(f_A^{(k)}) \longrightarrow E_w(f_A)$, and on passing to the limit we have $\rho = \rho\, E_w(f_A)/S$. If $\rho > 0$, then $E_w(f_A) = S$, and, by Proposition 2.6 (ii), f_A solves problem (C). $\qquad\square$

3.3. **Examples.** In this section, examples are given to illustrate the performance of the fixed point iterations given in the previous sections. In the first three examples the iterations from Sections 3.1 are used to approximate solutions to problem (B), the problem of near-interpolation. In the fourth example the iteration in Section 3.2 is used to approximate solutions to problem (C). In both cases the smoothing splines (A) are computed for the sequences of weights or smoothing parameters determined by the iterations. In each example the functions are piecewise cubic ($m = 2$), and the data sites are simple (no repetitions). Hence, we are interpolating to function values, not derivatives. In the second example, the formulas in this paper are generalized to parametric curves. The reader is referred to [9] for additional examples and discussion.

Example 3.5. In this example we fit smoothing splines to the "titanium heat data" given in [4]. The tolerances are uniformly taken as $\varepsilon_i = .02$, as illustrated by the vertical hash marks in Figure 3.1. In (a), the weights were computed by (3.2). In (b), the weights were computed by 17 iterations of (3.4), with the first iterate chosen by (3.2). In particular, curve (b) is the (approximate) solution to problem (B) for this data set.

One consequence of choosing the Lagrange multipliers from problem (B) as the weights in problem (A) is that these weights are zero for inactive constraints. For these inactive constraints the corresponding "spline knots" are also inactive; hence they can be removed from the representation of the spline curve. This effectively reduces the amount of data needed to represent the curve. For the curve in (b), 14 of the 49 data points are active. Hence, the spline curve needs roughly 14 knots (plus end knots). This is in contrast to the smoothing spline with strictly positive weights, which requires 49+ knots.

Example 3.6. The results in this paper extend to vector-valued data, as shown in [9]. In this case, the functions are vector-valued curves $f : [a, b] \longrightarrow \mathbb{R}^2$ or $\mathbb{R}^3$. In Figure 3.2, curve (a) solves problem (A) with the weights

$$w_i := \|H\Lambda f\|_2 / \varepsilon_i$$

(a generalization of (3.2)), and curve (b) solves problem (A) with the weights determined by 24 iterations of (3.4), here with $E_i(f) := \|\lambda_i f - z_i\|_2$. In particular, the second constraint is inactive; hence the corresponding weight, w_2, is zero. The weights for this problem are computed in Example 3.7. In the plots, the allowable tolerances of $\varepsilon_i = 0.1$ correspond to the radii of the circles displayed.

Example 3.7. In this third example, iteration (3.4) is compared to (3.5) for Example 3.6. The results are displayed in Table 3.1. Of particular interest is the weight w_2 corresponding to the second data point. Using (3.5), this weight tends to zero extremely fast, much faster than by (3.4). Hence, the inactive constraint
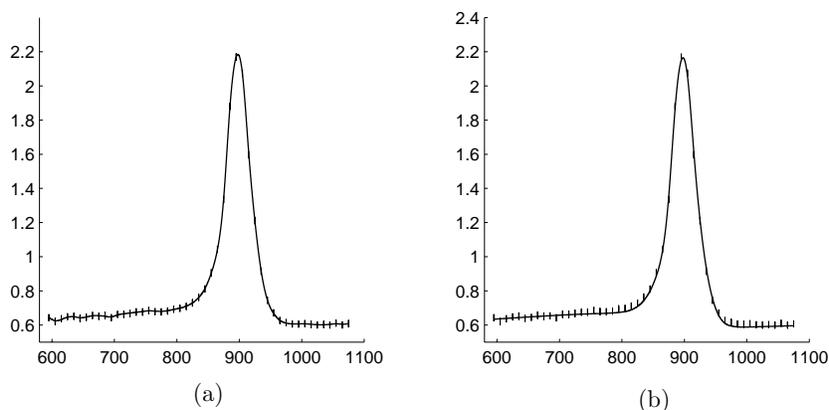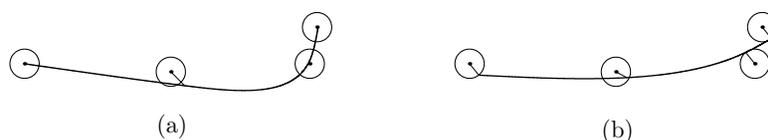
FIGURE 3.1. Titanium heat data, Example 3.5.

FIGURE 3.2. Vector-valued data, Example 3.6.

is identified much quicker by (3.5). In the table, the bold-faced weights are those whose convergence to zero was accelerated by (3.5). In particular, the second constraint has been identified as inactive after the ninth iteration. In practice, this constraint can be removed once it is clearly seen as inactive. Note that, using (3.5), the first constraint is falsely identified as inactive for one iteration, but this is quickly corrected.

TABLE 3.1. Weight iterations, Example 3.7.

	Iteration (3.4)				Iteration (3.5)			
#	w_1	w_2	w_3	w_4	w_1	w_2	w_3	w_4
1.	791.9	$1.7e+02$	14122	13477	791.9	$1.7e+02$	14122	13477
2.	6.4	$2.5e+02$	1349	1064	6.4	$2.5e+02$	1349	1064
3.	3.0	$1.5e+02$	825.4	643.3	**1.4**	$1.5e+02$	825.4	643.3
4.	3.8	$1.1e+02$	624.2	483.6	2.0	$1.1e+02$	625.4	483.3
5.	6.1	$8.0e+01$	517.7	400.4	3.5	$8.5e+01$	520.1	400.4
6.	9.7	$6.0e+01$	451.8	350.4	6.5	**2.5e+01**	455.0	350.6
7.	14.0	$4.4e+01$	406.8	317.7	16.0	$3.3e+01$	383.0	300.9
8.	17.9	$3.1e+01$	374.3	294.7	19.9	$2.4e+01$	356.0	281.6
9.	20.7	$2.1e+01$	350.2	277.5	22.1	**1.0e+00**	336.5	267.6
10.	22.6	$1.4e+01$	332.1	264.4	27.4	**3.3e-01**	311.4	251.7
11.	23.8	$9.4e+00$	318.3	254.3	27.0	**8.9e-03**	300.3	242.3
12.	24.5	$6.1e+00$	307.8	246.4	26.3	**8.7e-05**	293.2	236.1
13.	24.9	$3.9e+00$	299.9	240.4	25.8	**3.6e-07**	288.5	232.0
14.	25.0	$2.5e+00$	294.0	235.8	25.5	**6.9e-10**	285.3	229.1
15.	25.1	$1.6e+00$	289.7	232.3	25.2	**6.4e-13**	283.1	227.2
16.	25.1	$9.5e-01$	286.5	229.7	25.1	**3.1e-16**	281.6	225.8
17.	25.0	$5.8e-01$	284.1	227.8	25.0	**7.9e-20**	280.5	224.9
18.	25.0	$3.5e-01$	282.4	226.4	24.9	**1.1e-23**	279.8	224.2
19.	24.9	$2.1e-01$	281.2	225.3	24.8	**8.2e-28**	279.3	223.8
20.	24.9	$1.2e-01$	280.3	224.6	24.8	**3.5e-32**	278.9	223.5

TABLE 3.2. Smoothing parameter, Example 3.8.
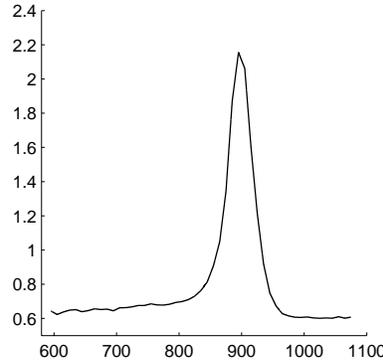
$\rho^{(k)}$	$\frac{\rho^{(k)}-\rho^{(k-1)}}{\rho^{(k)}}$	$\frac{\rho^{(k)}-\rho^{(k-1)}}{\rho^{(k-1)}-\rho^{(k-2)}}$	$\frac{E_w^{(k)}-S}{S}$	$\frac{E_w^{(k)}-S}{E_w^{(k-1)}-S}$
1				
$2.000000e+00$	$5.00e-01$	$-1.00e+00$	$1.88e+03$	$1.88e+03$
$3.770538e+03$	$9.99e-01$	$3.77e+03$	$1.17e+03$	$6.21e-01$
$4.416009e+06$	$9.99e-01$	$1.17e+03$	$6.04e+00$	$5.16e-03$
$3.108553e+07$	$8.58e-01$	$6.04e+00$	$1.69e-01$	$2.80e-02$
$3.634040e+07$	$1.45e-01$	$1.97e-01$	$5.29e-03$	$3.13e-02$
$3.653272e+07$	$5.26e-03$	$3.66e-02$	$1.66e-04$	$3.14e-02$
$3.653879e+07$	$1.66e-04$	$3.16e-02$	$5.22e-06$	$3.14e-02$
$3.653898e+07$	$5.22e-06$	$3.14e-02$	$1.64e-07$	$3.14e-02$
$3.653899e+07$	$1.64e-07$	$3.14e-02$	$5.15e-09$	$3.14e-02$
$3.653899e+07$	$5.15e-09$	$3.14e-02$	$1.61e-10$	$3.13e-02$
$3.653899e+07$	$1.61e-10$	$3.13e-02$	$5.02e-12$	$3.12e-02$

Example 3.8. In this final example, the solution to problem (C) for the titanium heat data is approximated by choosing the smoothing parameter in problem (A) based on iteration 3.6. The weights are uniformly taken as one, $i.e., w_i = 1$ for all i, and $S = 1/2$. Hence, the square root of the sum of the squares of the error from the data is less than $1/2$. The final curve fit is displayed in Table 3.2. The curve reproduces the data faithfully; however, it does not smooth out the oscillations between $t = 600$ and $t = 800$ like the near-interpolant in Figure 3.1 (b). Indeed, for this data with uniform weights it is virtually impossible to accurately reproduce the desired shape while smoothing out high frequencies, for any choice of ρ.

Table 3.2 lists values for the smoothing parameter for several iterations, and, as predicted in Proposition 3.4, convergence is monotonic. Also listed in the table are "relative errors" in the second and fourth columns, and estimated growth constants in the third and fifth columns. In particular, from the third column we have the estimate

$$|\rho^{(k)} - \rho^{(k-1)}| \approx .0314\,|\rho^{(k-1)} - \rho^{(k-2)}|$$

for $\rho^{(k)}$ near the optimal ρ. This follows the prediction in Proposition 3.4 with $C_k = .0314$ in that proof. From the last column,

$$|E_w(f_A^{(k)}) - S| \approx .0314\,|E_w(f_A^{(k-1)}) - S|,$$

which verifies the linear convergence that also was predicted in Proposition 3.4. The fact that the growth constants determined from the third and fifth columns

are nearly identical is not surprising. Indeed,

$$\frac{\rho^{(k)} - \rho^{(k-1)}}{\rho^{(k-1)} - \rho^{(k-2)}} = \frac{\rho^{(k-1)}(E_w^{(k-1)}/S - 1)}{\rho^{(k-2)}(E_w^{(k-2)}/S - 1)} = \frac{\rho^{(k-1)}(E_w^{(k-1)} - S)}{\rho^{(k-2)}(E_w^{(k-2)} - S)} \approx \frac{E_w^{(k-1)} - S}{E_w^{(k-2)} - S}$$

for large k. Although only linear, $\rho^{(k)}$ achieves a rapid rate of convergence to ρ, due to the small growth constant.

References

1. P. M. Anselone and P. J. Laurent, *A general method for the construction of interpolating or smoothing spline-functions*, Numerische Mathematik **12** (1968), 66–82. MR **40:**3145
2. M. Atteia, *Fonctions spline avec contraintes linéaires de type inégalite*, Congrès de l'AFIRO, Nancy, Mai (1967).
3. V. Barbu and Th. Precupanu, *Convexity and optimization in Banach spaces*, 3rd ed., Reidel, Dordrecht (1986). MR **87k:**49045
4. C. de Boor, *A Practical Guide to Splines*, Springer Verlag, New York (1978). MR **80a:**65027
5. P. Copley and L. L. Schumaker, *On pLg-splines*, J. Approx. Theory **23** (1978), 1–28. MR **58:**23258
6. P. Dierckx, *Curve and surface fitting with splines*, The Clarendon Press, Oxford University Press, New York (1993). MR **94m:**65028
7. J. Jerome and L. L. Schumaker, *A note on obtaining natural spline functions by the abstract approach of Atteia and Laurent*, SIAM J. Numer. Anal. **5** (1968), 657–663. MR **40:**6127
8. S. Kersey, *Best near-interpolation by curves: existence*, SIAM J. Numer. Anal. **38** (2000), 1666-1675. MR **2001m:**41034
9. S. Kersey, *Near-interpolation*, Numerische Mathematik (2003) to appear.
10. P. J. Laurent, *Construction of spline functions in a convex set*, Approximation with special emphasis on spline functions, I. J. Schoenberg, ed., Academic Press, New York (1969), 415–446. MR **40:**6147
11. O. L. Mangasarian and L. L. Schumaker, *Splines via optimal control*, Approximation with special emphasis on spline functions, I. J. Schoenberg, ed., Academic Press, New York (1969), 119–155. MR **41:**4073
12. C. H. Reinsch, *Smoothing by spline functions*, Numerische Mathematik **10** (1967), 177–183. MR **45:**4598
13. C. H. Reinsch, *Smoothing by spline functions II*, Numerische Mathematik **16** (1971), 451–454. MR **45:**4598
14. I. J. Schoenberg, *Spline functions and the problem of graduation*, Proc. Nat. Acad. Sci. USA **52** (1964), 947–950. MR **29:**5040
15. E. V. Shikin and A. I. Plis, *Handbook on splines for the user*, CRC Press, Boca Raton, FL (1995). MR **98f:**65015
16. G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics **59**, SIAM, Philadelphia, PA, (1990). MR **91g:**62028

Department of Mathematics, Case Western Reserve University, 10900 Eulcid Avenue, Cleveland, Ohio 44106-7085

E-mail address: snk@po.cwru.edu