ANNALS OF
MATHEMATICS

Fitting a C^m-smooth function to data, III

By Charles Fefferman

SECOND SERIES, VOL. 170, NO. 1

July, 2009

ANMAAH

Annals of Mathematics, **170** (2009), 427–441

Fitting a C^m-smooth function to data, III

By Charles Fefferman

Abstract

Fix $m, n \geq 1$. Given an N-point set $E \subset \mathbb{R}^n$, we exhibit a list of $O(N)$ subsets $S_1, S_2, \ldots, S_L \subset E$, each containing $O(1)$ points, such that the following holds: Let $f : E \to \mathbb{R}^n$. Suppose that, for each $\ell = 1, \ldots, L$, there exists $F_\ell \in C^m(\mathbb{R}^n)$ with norm ≤ 1, agreeing with f on S_ℓ. Then there exists $F \in C^m(\mathbb{R}^n)$, with norm $O(1)$, agreeing with f on E.

We give an application to the problem of discarding outliers from the set E.

0. **Introduction**

Let $m, n \geq 1$, and suppose we are given N points in $\mathbb{R}^{n+1}$. Throughout this paper, we regard m, n as fixed, and N as arbitrarily large. How can we find a function $F \in C^m(\mathbb{R}^n)$, whose graph passes through (or close to) the given points, with the C^m-norm of F having the smallest possible order of magnitude? How small can we take that order of magnitude?

In [4], [5], we studied these questions from the viewpoint of theoretical computer science. We exhibited algorithms to compute F, and we estimated the resources required by an (idealized) computer to carry out those algorithms. Here, we prove a theorem announced in [4], and apply it to study what happens when we are allowed to discard some of the N given points as outliers. We want to know which points to discard, and how much we can reduce the order of magnitude of the C^m norm of an optimal F, as a function of the number of points discarded.

To state our results precisely, we introduce some notation and definitions. Let $X, Y \geq 0$ be real numbers determined by m, n and other data (e.g., N given points in $\mathbb{R}^{n+1}$). Then we say that X and Y have the "same order of magnitude", and we write "$X \sim Y$", provided we have $cX \leq Y \leq CX$, with c and C depending only on m and n. (More generally, throughout this paper, c, C, C', etc., denote constants depending only on m and n. These constants may change from one occurrence to

Partially supported by NSF Grant #DMS-0245242.

the next.) To "compute the order of magnitude" of X is to compute some Y such that $X \sim Y$.

For any finite set S, we write $\#(S)$ to denote the number of elements of S. If S is infinite, then we define $\#(S) = \infty$.

As usual $C^m(\mathbb{R}^n)$ denotes the space of all m-times continuously differentiable $F : \mathbb{R}^n \longrightarrow \mathbb{R}$, for which the norm

$$\|F\| = \max_{|\alpha| \le m} \sup_{x \in \mathbb{R}^n} |\partial^\alpha F(x)|$$

is finite.

Next, suppose we are given a finite set $E \subset \mathbb{R}^n$, and functions $f : E \longrightarrow \mathbb{R}$, $\sigma : E \longrightarrow [0, \infty)$. We want to find $F \in C^m(\mathbb{R}^n)$ and $M \in (0, \infty)$ such that

(0.1) $\|F\| \le M$, and $|F(x) - f(x)| \le M\sigma(x)$ for all $x \in E$.

We define $\|f\|_{(E,\sigma)}$ as the infimum of all $M \in (0, \infty)$ for which there exists $F \in C^m(\mathbb{R}^n)$ that satisfies (0.1).

The function σ serves as a "tolerance". It gives a precise meaning to our demand that the graph of F pass "close to" N given points. Taking $\sigma \equiv 0$ amounts to demanding that the graph pass through the given points.

Let E, f, σ be as above, and let $S \subset E$. For simplicity, we write $\|f\|_{(S,\sigma)}$ to denote $\|f\|_{(S,\sigma|_S)}$, where (as usual) $\sigma|_S$ is the restriction of σ to S. Clearly, $\|f\|_{(S,\sigma)} \le \|f\|_{(E,\sigma)}$.

We will exhibit algorithms, to be run on an idealized computer with standard von Neumann architecture [12], able to deal with exact real numbers. The resources required to carry out an algorithm are the "number of operations" (or "work"), and the "storage". Here, an "operation" means, e.g., an addition or multiplication of two given real numbers, or an application of the "greatest integer" function. See [5] for a more careful discussion, including a model of computation with finite-precision numbers. The "storage" is simply the number of reals that can be held in memory.

The main theorem of this paper is the following result, announced in [4].

THEOREM 1. *Let $E \subset \mathbb{R}^n$, with $\#(E) = N < \infty$, and let $\sigma : E \longrightarrow [0, \infty)$. Then there exists a list of subsets $S_1, S_2, \dots, S_L \subset E$, with the following properties.*

(A) *Each S_ℓ has at most C elements.*

(B) *The number of subsets S_ℓ is $L \le CN$.*

(C) *For any $f : E \longrightarrow \mathbb{R}$, we have*

$$\|f\|_{(E,\sigma)} \sim \max_{\ell=1,\dots,L} \|f\|_{(S_\ell,\sigma)}.$$

(D) *The list of subsets $S_1, S_2, \dots, S_L$ may be computed from E, σ, m, n using at most $CN \log N$ operations, and using storage at most CN.*

In view of (A), the order of magnitude of any given $\| f \|_{(S_\ell,\sigma)}$ may be easily computed by standard linear algebra, using at most C' operations. (We spell out the details in Section 1.)

Hence, Theorem 1 allows us to preprocess E, σ, after which we can compute the order of magnitude of $\| f \|_{(E,\sigma)}$ for any given f, using at most CN operations. The preprocessing takes at most $CN \log N$ operations, using storage CN.

Theorem 1 is a refinement of the following result, proven in [8].

THEOREM 1′. *Let $E \subset \mathbb{R}^n$ be finite, and let $f : E \longrightarrow \mathbb{R}$, $\sigma : E \longrightarrow [0, \infty)$. Then*

$$\| f \|_{(E,\sigma)} \sim \max\{\| f \|_{(S,\sigma)} : S \subset E,\ \#(S) \leq k\},$$

where k depends only on m and n.

Theorem 1′ in turn overlaps with the earlier results of Y. Brudnyi and P. Shvartsman [2]. They conjectured a "finiteness principle" analogous to Theorem 1′, but with $\sigma \equiv 0$, and with $C^m(\mathbb{R}^n)$ replaced by more general function spaces. For certain function spaces, including $C^2(\mathbb{R}^n)$, they proved their finiteness principle, with an optimal constant k, by the method of "Lipschitz selection".

Recent works of Bierstone-Milman [1] and of Shvartsman [2] have made significant progress toward finding the best constant k in Theorem 1′ for general m, n. Their work applies also to the constant C in Theorem 1(A).

More broadly, Theorems 1 and 1′ pertain to "Whitney's extension problem". We refer the reader to works of Whitney, Glaeser, Brudnyi-Shvartsman, Zobin, Bierstone-Milman-Pawlucki, as cited in [2], [4], [5], [8], [11].

We apply Theorem 1 to the problem of finding and discarding outliers. Our result is as follows.

THEOREM 2. *Suppose we are given the following data:*

- *A finite set $E \subset \mathbb{R}^n$, with $\#(E) = N \geq 2$;*
- *Functions $f : E \longrightarrow \mathbb{R}$, $\sigma : E \longrightarrow [0, \infty)$;*
- *An integer $Z \geq 0$.*

Then, in $C + CZN \log N$ operations, using storage CN, we can compute a subset $S^* \subset E$, with the following properties:

(α) $\#(S^*) \leq CZ$; *and*

(β) *For any $S \subset E$ with $\#(S) \leq Z$, we have*

$$\| f \|_{(E \setminus S^*,\sigma)} \leq C \| f \|_{(E \setminus S,\sigma)}.$$

Thus, if we are allowed to discard $\sim Z$ points, then, essentially, we can do no better than to discard S^*.

It is easy to deduce Theorem 2 from Theorem 1. We explain heuristically the algorithm to produce S^*, and say a few words about why it works. A careful discussion is given in Section 2.

Let E, f, σ, Z be as in Theorem 2. We have to find a subset $S^* \subset E$ satisfying (α) and (β). To do so, we first compute the list of subsets $S_1, S_2, \dots, S_L$ from Theorem 1, and then compute the order of magnitude of each $\|f\|_{(S_\ell,\sigma)}$, $\ell = 1, \dots, L$. This takes at most $CN \log N$ operations and storage CN.

Next, we discard from E the points of S_{ℓ_*}, with ℓ_* picked to maximize the order of magnitude of $\|f\|_{(S_{\ell_*},\sigma)}$.

To see why we do so, suppose $S \subset E$ with $\|f\|_{(E \setminus S,\sigma)} \ll \|f\|_{(E,\sigma)}$. Then S_{ℓ_*} must contain at least one point of S. In fact, otherwise,

$$\|f\|_{(S_{\ell_*},\sigma)} \leq \|f\|_{(E \setminus S,\sigma)} \ll \|f\|_{(E,\sigma)} \leq C \cdot \max_{\ell} \|f\|_{(S_\ell,\sigma)}$$

by Theorem 1(C), contradicting the fact that $\|f\|_{(S_{\ell_*},\sigma)}$ has the maximal order of magnitude among the $\|f\|_{(S_\ell,\sigma)}$.

Thus, with work at most $CN \log N$, we have found a set $S_{\ell_*} \subset E$ with $\#(S_{\ell_*}) \leq C$, guaranteed to contain at least one point of any "bad" set $S \subset E$.

Once we have discarded the points of S_{ℓ_*}, we find ourselves in the same situation as before, but with $E \setminus S_{\ell_*}$ and $Z - 1$ in place of E and Z, respectively. Proceeding recursively until we reach the trivial case $Z = 0$, we discard at most CZ points from E; and the set S^* of discarded points will satisfy the conclusions (α), (β) of Theorem 2. Again, we refer the reader to Section 2 for a rigorous discussion of the ideas sketched above.

We have no reason to think that the work $C + CZN \log N$ in Theorem 2 is best possible, and we look forward to future improvements.

We turn our attention to the proof of Theorem 1. There are two main ingredients.

The first ingredient is a construction of an $F \in C^m(\mathbb{R}^n)$, satisfying (0.1) with the order of magnitude of M as small as possible. That F has the form $F = Tf$ for a linear operator T. This construction is described in [7], [4], [5].

For any $x \in \mathbb{R}^n$ and $|\alpha| \leq m$, we therefore have

(0.2) $\partial^\alpha(Tf)(x) = \sum_{y \in E} \lambda_\alpha(x,y)\, f(y)$, with coefficients $\lambda_\alpha(x,y)$ independent of f. Moreover, for the operator T constructed in the references just mentioned, the coefficients $\lambda_\alpha(x,y)$ are "sparse." More precisely, $\lambda_\alpha(x,y)$ vanishes for all $y \in E$ outside a subset $S(x) \subset E$, with

(0.3) $\#(S(x)) \leq C$. Thus, (0.2) may be rewritten in the form

(0.4) $\partial^\alpha(Tf)(x) = \sum_{y \in S(x)} \lambda_\alpha(x,y)\, f(y)$ for $|\alpha| \leq m$, $x \in \mathbb{R}^n$, and we have

(0.5) $\|Tf\| \leq M, \ |(Tf)(x) - f(x)| \leq M\sigma(x)$ for all $x \in E$, with

(0.6) $M \leq C\|f\|_{(E,\sigma)}$, since $F = Tf$ satisfies (0.1) with the order of magnitude of M as small as possible.

The existence of T, $S(x)$, $\lambda_\alpha(x, y)$ satisfying (0.3), ..., (0.6) is proven in the discussion of "operators of finite depth" in [7].

Furthermore it was shown in [4], [5] how to compute $S(x)$ and the $\lambda_\alpha(x, y)$ $(y \in S(x), \ |\alpha| \leq m)$ for any given x. In particular, after preprocessing E, σ, we can compute $S(x)$ for any given $x \in \mathbb{R}^n$ using at most $C \log N$ operations. The preprocessing takes at most $CN \log N$ operations, with storage CN. (Here, as usual, $N = \#(E) < \infty$.) See Theorem 6 in [5].

We will need to compute $S(x)$ only for $x \in E$, and we will not need to compute the $\lambda_\alpha(x, y)$. According to the above remarks, we can compute all the $S(x)$ $(x \in E)$ with at most $CN \log N$ operations and storage CN.

The second main ingredient in the proof of Theorem 1 is the work of Callahan and Kosaraju [3] from computational geometry. We state here only the subset of their results needed for our proof.

Let $\kappa \in (0, 1)$ be a small constant. We write c_κ, C_κ, C'_κ, etc., to denote constants depending only on κ, and on the dimension n. As usual, let $E \subset \mathbb{R}^n$, with $\#(E) = N < \infty$. According to [3], we can partition the set of pairs $\{(x, y) \in E \times E : x \neq y\}$ into subsets

(0.7) $E'_1 \times E''_1, E'_2 \times E''_2, \dots, E'_L \times E''_L$ with the following properties.

(0.8) In (0.7), we have $L \leq C_\kappa N$.

(0.9) For each $\ell = 1, \dots, L$, we have $\mathrm{diam}(E'_\ell), \mathrm{diam}(E''_\ell) \leq \kappa \, \mathrm{dist}(E'_\ell, E''_\ell)$, where, as usual,

$$\mathrm{diam}(A) = \max\{|x - y| : x, y \in A\} \text{ and}$$

$$\mathrm{dist}(A, B) = \min\{|x - y| : x \in A, \ y \in B\} \text{ for finite } A, B \subset \mathbb{R}^n.$$

Moreover, we can pick "representatives"

(0.10) $x'_\ell \in E'_\ell$ and $x''_\ell \in E''_\ell$ for each $\ell = 1, \dots, L$, in such a way that

(0.11) The x'_ℓ, x''_ℓ for $\ell = 1, \dots, L$ can all be computed, using at most $C_\kappa N \log N$ operations and storage $C_\kappa N$.

See [3], and also Har-Peled and Mendel [9] for further results. We call (0.7) the "Callahan-Kosaraju decomposition". (In the computer science literature, it is called the "well-separated pairs decomposition".)

We now indicate how our two main ingredients (0.3) ... (0.6) and (0.7) ... (0.11) are used in the proof of Theorem 1. Let E, σ, N be as in the hypotheses of Theorem 1. We have to find a list of subsets $S_1, \dots, S_L \subset E$ satisfying conclusions (A) ... (D).

Taking κ to be a small enough constant depending only on m and n, we produce the Callahan-Kosaraju decomposition (0.7). Let x'_ℓ, x''_ℓ ($\ell = 1, \dots, L$) be the "representatives", as in (0.10) and (0.11). Also, let $S(x)(x \in E)$ be as in (0.3) ... (0.6). As our list of subsets $S_1, \dots, S_L$, we take

(0.12) $S_\ell = S(x'_\ell) \cup S(x''_\ell) \cup \{x'_\ell, x''_\ell\}$ for $\ell = 1, \dots, L$.

In view of (0.3), we have $\#(S_\ell) \leq C'$ for each ℓ. Also, from (0.8), we have $L \leq CN$, since κ depends only on m, n. Recall that with $CN \log N$ operations using storage CN, we can compute all the x'_ℓ, x''_ℓ and all the $S(x)$ $(x \in E)$. A glance at (0.12) shows that all the $S_\ell (1 \leq \ell \leq L)$ can therefore be computed in $CN \log N$ operations using storage CN. Thus, conclusions (A), (B) and (D) are obvious for our $S_1, \dots, S_L$. It remain to establish (C). This comes down to finding an $F \in C^m(\mathbb{R}^n)$ satisfying (0.1), with $M \sim \max_\ell \|f\|_{(S_\ell, \sigma)}$.

The proof of (C) occupies Section 4 below.

This concludes our introductory remarks on the proof of Theorem 1.

Theorem 1 can be generalized, as in [5], by bringing "Whitney t-convex sets" into the statement of the problem. (See also [6].) Our proofs below are easily adapted to this case, so we omit further discussion.

It is a pleasure to thank Bo'az Klartag for many valuable discussions, and Gerree Pecht, for LaTeXing this paper to lofty "Gerree standards".

1. Small subsets of $\mathbb{R}^n$

In this section, we show how to compute the order of magnitude of $\|f\|_{(S,\sigma)}$, under the assumption

(1.1) $\#(S) \leq C$.

We start by setting up notation, to be used throughout this paper. For $F \in C^m(\mathbb{R}^n)$, and for $x \in \mathbb{R}^n$, we write $J_x(F)$ to denote the $(m-1)^{\text{rst}}$ degree Taylor polynomial of F at x. Thus, $J_x(F)$ belongs to $\mathcal{P}$, the vector space of all (real) $(m-1)^{\text{rst}}$ degree polynomials on $\mathbb{R}^n$.

Next, let $S \subset \mathbb{R}^n$, $f : S \longrightarrow \mathbb{R}$, $\sigma : S \longrightarrow [0, \infty)$, with S finite. Suppose that $F \in C^m(\mathbb{R}^n)$ and $M > 0$ satisfy

(1.2) $\|F\| \leq M$, and $|F(x) - f(x)| \leq M\sigma(x)$ for all $x \in S$.

Define a collection of polynomials

(1.3) $\vec{P} = (P_x)_{x \in S} \in \sum_{x \in S} \oplus \mathcal{P}$,

by setting $P_x = J_x(F)$ for each $x \in S$. From (1.2) and Taylor's theorem, we have:

(1.4) $|P_x(x) - f(x)| \leq M\sigma(x)$ for $x \in S$;

(1.5) $|(\partial^\alpha P_x)(x)| \le M$ for $x \in S, |\alpha| \le m-1$; and

(1.6) $|\partial^\alpha (P_x - P_y)(y)| \le CM|x-y|^{m-|\alpha|}$ for $x, y \in S, |\alpha| \le m-1$.

Conversely, suppose $S \subset \mathbb{R}^n$, $f : S \longrightarrow \mathbb{R}$, $\sigma : S \longrightarrow [0,\infty)$ with S finite, and suppose $\vec{P} = (P_x)_{x\in S}$ satisfies (1.3) ... (1.6). Then, by the classical Whitney extension theorem [10], there exists $F \in C^m(\mathbb{R}^n)$, such that

(1.7) $\|F\| \le C'M$, and $J_x(F) = P_x$ for each $x \in S$.

In particular, (1.4) and (1.7) give

(1.8) $|F(x) - f(x)| \le M\sigma(x)$ for all $x \in S$.

Consequently, $\|f\|_{(S,\sigma)}$ has the same order of magnitude as the infimum of all $M > 0$ for which there exists $\vec{P} = (P_x)_{x\in S}$ satisfying (1.3) ... (1.6). Under assumption (1.1), the order of magnitude of this infimum is easily computed by linear algebra, as follows. Let H be the affine space of all $\vec{P} = (P_x)_{x\in S} \in \sum_{x\in S} \oplus \mathcal{P}$ such that $P_x(x) = f(x)$ whenever $\sigma(x) = 0$. On H, we define a quadratic function $\mathfrak{Q}(\vec{P}) = \mathfrak{Q}_1(\vec{P}) + \mathfrak{Q}_2(\vec{P}) + \mathfrak{Q}_3(\vec{P})$, by setting

$$\mathfrak{Q}_1(\vec{P}) = \sum_{\substack{x\in S\\ \sigma(x)\neq 0}} \left[\frac{P_x(x) - f(x)}{\sigma(x)}\right]^2,$$

$$\mathfrak{Q}_2(\vec{P}) = \sum_{x\in S} \sum_{|\alpha|\le m-1} [(\partial^\alpha P_x)(x)]^2,$$

$$\mathfrak{Q}_3(\vec{P}) = \sum_{\substack{x,y\in S\\ (x\neq y)}} \sum_{|\alpha|\le m-1} \left[\frac{\partial^\alpha (P_x - P_y)(y)}{|x-y|^{m-|\alpha|}}\right]^2.$$

Suppose S satisfies (1.1). Then, clearly, the desired infimum (of all $M > 0$ for which (1.3) ... (1.6) can be satisfied) has the same order of magnitude as the square root of the minimum of $\mathfrak{Q}(\vec{P})$ over all $\vec{P} \in H$.

Since $\mathfrak{Q}$ is a quadratic function and H is an affine space of dimension at most C', we can compute $\min\{\mathfrak{Q}(\vec{P}) : \vec{P} \in H\}$ by linear algebra, using at most C'' operations.

Hence, if f, S, σ are given, with $\#(S) \le C$, then we can compute the order of magnitude of $\|f\|_{(S,\sigma)}$ with work at most C''.

2. **Removing outliers**

In this section, we prove Theorem 2, assuming Theorem 1. The algorithm sketched in the introduction in connection with Theorem 2 is as follows.

PROCEDURE. Outliers(E, f, σ, Z).

/* Defined for $E \subset \mathbb{R}^n$ finite, $f : E \longrightarrow \mathbb{R}$, $\sigma : E \longrightarrow [0, \infty)$, $Z \geq 0$ an integer. Prints out a finite list of points of E. The set S^*, consisting of all the points printed out, will later be shown to have the properties asserted in Theorem 2.
*/

LINE 1 {while ($Z \neq 0$ and $E \neq \varnothing$)

LINE 2 {Compute the list of sets $S_1, \dots, S_L \subset E$ associated to (E, σ), as in Theorem 1.

LINE 3 For each $\ell = 1, \dots, L$, compute $X_\ell \sim \|f\|_{(S_\ell, \sigma)}$, as in Section 1.

LINE 4 Find ℓ_* ($1 \leq \ell_* \leq L$) such that $X_{\ell_*} \geq X_\ell$ for $\ell = 1, \dots, L$.

LINE 5 Print out the elements of S_{ℓ_*}.

LINE 6 Replace (E, f, σ, Z) by $(E \setminus S_{\ell_*}, f|_{E \setminus S_{\ell*}}, \sigma|_{E \setminus S_{\ell*}}, Z - 1)$.

LINE 7 }

LINE 8 }

The loop body here consists of Lines $2 \dots 7$. Every time we execute the loop body, Z decreases by one. Once we reach $Z = 0$, the procedure terminates. (Perhaps the procedure terminates before we reach $Z = 0$.) Consequently, execution of Outliers(E, f, σ, Z) always terminates, and the loop body is executed at most Z times.

Let $\#(E) \leq N$, with $N \geq 2$. From Theorem 1 and Section 1, we see that execution of LINE 2 requires work $\leq CN \log N$ and storage $\leq CN$, while execution of LINE 3 requires work and storage $\leq CN$. We also need storage CN, simply to hold (E, f, σ, Z). Inspection of LINES 2 ... 7 now shows that we can execute the loop body with work $\leq CN \log N$ and storage $\leq CN$. Since the loop body is executed at most Z times, we see that

(2.1) Outliers(E, f, σ, Z) requires work $\leq C + CZN \log N$, and storage $\leq CN$.

Here, the extra work C enters because we must execute LINE 1, even if $Z = 0$.

Next, note that each S_ℓ computed in LINE 2 satisfies $\#(S_\ell) \leq C$, by Theorem 1(A). Consequently, each time we execute the loop body, we print out at most C points. Since the loop body is executed at most Z times, Outliers (E, f, σ, Z) prints out at most CZ points. Let us write $S^* = S^*(E, f, \sigma, Z)$ for the set of points printed out by Outliers(E, f, σ, Z). Thus,

(2.2) $S^* \subset E$ and $\#(S^*) \leq CZ$.

To complete the proof of Theorem 2, we have to demonstrate this:

(2.3) Let $S \subset E$ with $\#(S) \leq Z$. Then $\|f\|_{(E \setminus S^*, \sigma)} \leq C_1 \|f\|_{(E \setminus S, \sigma)}$.

We prove (2.3) by induction on Z. For $Z = 0$, we have $S^* = \varnothing$ (see LINE 1), and therefore (2.3) simply asserts that $\|f\|_{(E,\sigma)} \le C_1\|f\|_{(E,\sigma)}$. Thus, (2.3) holds trivially for $Z = 0$.

For the induction step, we fix $Z \ge 1$, and assume the analogue of (2.3) with Z replaced by $Z-1$. We will prove (2.3) for our given Z.

If $E = \varnothing$, then again (2.3) holds trivially. (By definition, $\|f\|_{(E,\sigma)} = 0$ when $E = \varnothing$.) Hence, we may suppose $E \ne \varnothing$.

Consequently, we have initially that ($Z \ne 0$ and $E \ne \varnothing$), as in LINE 1, and therefore the loop body is executed at least once. Let $S_1, \dots, S_L$, $X_1, \dots, X_L$, and ℓ_* be as they are computed the first time we execute the loop body.

Also, let $S^{**} = S^*(E \setminus S_{\ell_*}, f|_{E\setminus S_{\ell_*}}, \sigma|_{E\setminus S_{\ell_*}}, Z-1)$.

That is, S^{**} is the analogue of S^*, when we take as input $(E\setminus S_{\ell_*}, f|_{E\setminus S_{\ell_*}}, \sigma|_{E\setminus S_{\ell_*}}, Z-1)$ in place of (E, f, σ, Z).

Inspection of LINES 5 and 6 shows that

(2.4) $S^* = S_{\ell_*} \cup S^{**}$.

On the other hand, our induction hypothesis (namely, (2.3) for $Z-1$) tells us the following.

(2.5) Let $\hat{S} \subset E\setminus S_{\ell_*}$, with $\#(\hat{S}) \le Z-1$. Then

$$\|f\|_{((E\setminus S_{\ell_*})\setminus S^{**},\sigma)} \le C_1\|f\|_{((E\setminus S_{\ell_*})\setminus \hat{S},\sigma)}.$$

Now let $S \subset E$, with $\#(S) \le Z$. We show that (2.3) holds for S. We consider separately two cases.

Case 1: $S \cap S_{\ell_*} = \varnothing$. Then Theorem 1(C) gives

$$\begin{aligned}\|f\|_{(E\setminus S^*,\sigma)} &\le \|f\|_{(E,\sigma)} \le C \max_{\ell=1,\dots,L} \|f\|_{(S_\ell,\sigma)} \\ &\le C' \max_{\ell=1,\dots,L} X_\ell = C' X_{\ell_*} \le C''\|f\|_{(S_{\ell_*},\sigma)} \le C''\|f\|_{(E\setminus S,\sigma)},\end{aligned}$$

since $S_{\ell_*} \subset E\setminus S$ and $X_\ell \sim \|f\|_{(S_\ell,\sigma)}$ for each ℓ. Thus, (2.3) holds in Case 1, for C_1 large enough.

Case 2: $S \cap S_{\ell_*} \ne \varnothing$. Set $\hat{S} = S\setminus S_{\ell_*}$. Then $\#(\hat{S}) \le Z-1$, and therefore (2.5) applies. Since also $(E\setminus S_{\ell_*})\setminus \hat{S} \subset E\setminus S$, we learn from (2.4) and (2.5) that

$$\begin{aligned}\|f\|_{(E\setminus S^*,\sigma)} &= \|f\|_{((E\setminus S_{\ell_*})\setminus S^{**},\sigma)} \\ &\le C_1\|f\|_{((E\setminus S_{\ell_*})\setminus \hat{S},\sigma)} \le C_1\|f\|_{(E\setminus S,\sigma)}.\end{aligned}$$

Thus, (2.3) holds also in Case 2. This completes our induction on Z, proving (2.3).

The conclusions of Theorem 2 are our results (2.1), (2.2), (2.3). Thus, we have established Theorem 2, assuming Theorem 1.

As a conclusion to this section, we compare Outliers(E, f, σ, Z) with Outliers(E, f, σ, Z') for $Z' < Z$. These two procedures print out the same points, in the same order, until Outliers(E, f, σ, Z') terminates, after which Outliers (E, f, σ, Z) may print out additional points. Consequently, the list of points $x_1, x_2, \dots, x_{k_{\max}}$ printed out by Outliers(E, f, σ, Z) has the following property, strengthening Theorem 2:

(2.6) Suppose $S \subset E$, with $\#(S) = Z' \le Z$.

Let $S^* = \{x_k : 1 \le k \le \min(CZ', k_{\max})\}$. Then

$$\| f \|_{(E \setminus S^*, \sigma)} \le C \| f \|_{(E \setminus S, \sigma)}.$$

Detailed verifications are left to the reader.

3. **Comparing polynomials at representative points**

In this section, we suppose we are given the following data:

- a finite set $E \subset \mathbb{R}^n$, with $\#(E) = N \ge 2$;
- a constant $\kappa \in (0, 1)$;
- a Callahan-Kosaraju decomposition $E_1' \times E_1'', \dots, E_L' \times E_L''$; and representatives $(x_\ell', x_\ell'') \in E_\ell' \times E_\ell''(\ell = 1, \dots, L)$, as in (0.7) ... (0.11);
- a polynomial $P_x \in \mathcal{P}$, for each $x \in E$; and
- a number $M \in (0, \infty)$.

(Recall that $\mathcal{P}$ is the vector space of $(m-1)^{\text{rst}}$ degree polynomials on $\mathbb{R}^n$.) Motivated by the classical Whitney extension theorem as in Section 1, we want to know whether

$$|\partial^\alpha (P_x - P_y)(y)| \le M |x - y|^{m - |\alpha|} \text{ for } x, y \in E, \ |\alpha| \le m - 1.$$

The next result shows that it is enough to look at the case $x = x_\ell'$, $y = x_\ell''$ for $\ell = 1, \dots, L$.

LEMMA 3.1. *Suppose κ is less than a small enough constant c_1, and assume we have*

(3.1) $|\partial^\alpha (P_{x_\ell'} - P_{x_\ell''})(x_\ell'')| \le M |x_\ell' - x_\ell''|^{m - |\alpha|}$ *for* $|\alpha| \le m - 1, \ell = 1, \dots, L.$

Then we have

$$|\partial^\alpha (P_x - P_y)(y)| \le CM |x - y|^{m - |\alpha|} \textit{ for all } x, y \in E, \ |\alpha| \le m - 1.$$

Proof. Let A be a large enough constant, to be picked in (3.15) below. We will prove by contradiction that

(3.2) $|\partial^\alpha (P_x - P_y)(y)| \le AM |x - y|^{m - |\alpha|}$ for $|\alpha| \le m - 1, \ x, y \in E.$

In fact, suppose (3.2) fails. Since E is finite, we can pick $\bar{x}, \bar{y} \in E$ and $\bar{\alpha}$ ($|\bar{\alpha}| \le m-1$) violating (3.2), with $|\bar{x} - \bar{y}|$ as small as possible. Thus,

(3.3) $|\partial^{\bar{\alpha}}(P_{\bar{x}} - P_{\bar{y}})(\bar{y})| > AM|\bar{x} - \bar{y}|^{m-|\bar{\alpha}|}$, and

(3.4) $\bar{x}, \bar{y} \in E$ and $|\bar{\alpha}| \le m-1$, but

(3.5) $|\partial^{\alpha}(P_x - P_y)(y)| \le AM|x - y|^{m-|\alpha|}$ for $|\alpha| \le m-1$, $x, y \in E$, $|x - y| < |\bar{x} - \bar{y}|$.

Note that $\bar{x} \ne \bar{y}$, as we see at once from (3.3). Since the $E'_\ell \times E''_\ell$ ($\ell = 1, \dots, L$) form a Callahan-Kosaraju decomposition, there exists $\ell (1 \le \ell \le L)$ such that

(3.6) $\bar{x} \in E'_\ell$ and $\bar{y} \in E''_\ell$.

We fix such an ℓ for the rest of the proof of Lemma 3.1.

By (0.10), the representatives x'_ℓ, x''_ℓ satisfy

(3.7) $x'_\ell \in E'_\ell$ and $x''_\ell \in E''_\ell$.

From (3.6), (3.7) and (0.9), we conclude that

(3.8) $|\bar{x} - x'_\ell|, |\bar{y} - x''_\ell| \le \kappa|\bar{x} - \bar{y}|$, and therefore (3.5) applies to the pairs

$$(x, y) = (x'_\ell, \bar{x}) \text{ and } (x, y) = (x''_\ell, \bar{y}).$$

Thus,

(3.9) $|\partial^{\alpha}(P_{x'_\ell} - P_{\bar{x}})(\bar{x})| \le AM|x'_\ell - \bar{x}|^{m-|\alpha|} \le \kappa AM|\bar{x} - \bar{y}|^{m-|\alpha|}$ for $|\alpha| \le m-1$,

and

(3.10) $|\partial^{\alpha}(P_{x''_\ell} - P_{\bar{y}})(\bar{y})| \le AM|x''_\ell - \bar{y}|^{m-|\alpha|} \le \kappa AM|\bar{x} - \bar{y}|^{m-|\alpha|}$ for $|\alpha| \le m-1$.

We will combine (3.1), (3.9) and (3.10) to estimate $P_{\bar{x}} - P_{\bar{y}}$. To do so, we must first move the base point in (3.9) from $\bar{x}$ to $\bar{y}$, and similarly for (3.1).

For $|\alpha| \le m-1$, (3.9) gives

$$\text{(3.11)}\quad \begin{aligned} |\partial^{\alpha}(P_{x'_\ell} - P_{\bar{x}})(\bar{y})| &= \left| \sum_{|\beta| \le m-1-|\alpha|} \frac{1}{\beta!} (\partial^{\beta+\alpha}(P_{x'_\ell} - P_{\bar{x}})(\bar{x}))(\bar{y} - \bar{x})^{\beta} \right| \\ &\le C \sum_{|\beta| \le m-1-|\alpha|} (\kappa AM|\bar{x} - \bar{y}|^{m-(|\alpha|+|\beta|)})|\bar{x} - \bar{y}|^{|\beta|} \\ &\le C'\kappa AM|\bar{x} - \bar{y}|^{m-|\alpha|}. \end{aligned}$$

Regarding (3.1), we first note that (3.8) yields

$$|x'_\ell - x''_\ell| \le |x'_\ell - \bar{x}| + |\bar{x} - \bar{y}| + |\bar{y} - x''_\ell| \le 3|\bar{x} - \bar{y}| \text{ and } |x''_\ell - \bar{y}| \le |\bar{x} - \bar{y}|.$$

Therefore, for $|\alpha| \le m-1$, (3.1) gives

$$(3.12)\quad |\partial^\alpha(P_{x'_\ell} - P_{x''_\ell})(\bar{y})| = \left| \sum_{|\beta| \le m-1-|\alpha|} \frac{1}{\beta!}(\partial^{\beta+\alpha}(P_{x'_\ell} - P_{x''_\ell})(x''_\ell))(\bar{y} - x''_\ell)^\beta \right|$$
$$\le C \sum_{|\beta| \le m-1-|\alpha|} M|x'_\ell - x''_\ell|^{m-(|\alpha|+|\beta|)} \cdot |\bar{y} - x''_\ell|^{|\beta|}$$
$$\le C'M|\bar{x} - \bar{y}|^{m-|\alpha|}.$$

Now, combining (3.10), (3.11) and (3.12), we obtain

$$(3.13)\quad |\partial^\alpha(P_{\bar{x}} - P_{\bar{y}})(\bar{y})| \le C''(1 + \kappa A)M|\bar{x} - \bar{y}|^{m-|\alpha|} \text{ for } |\alpha| \le m-1.$$

We now take

$$(3.14)\quad \kappa < 1/(2C'')$$

and

$$(3.15)\quad A = 2C'',$$

with C'' as in (3.13). Thus,

$$C''(1 + \kappa A) = C'' + (C''\kappa)A < C'' + \tfrac{1}{2}A = A,$$

and (3.13) implies

$$(3.16)\quad |\partial^\alpha(P_{\bar{x}} - P_{\bar{y}})(\bar{y})| \le AM|\bar{x} - \bar{y}|^{m-|\alpha|} \text{ for } |\alpha| \le m-1.$$

Taking $\alpha = \bar{\alpha}$ in (3.16), we obtain a contradiction to (3.3) and (3.4).

Consequently, our initial assumption (that (3.2) fails) is absurd, and we have (3.2), with A given by (3.15), provided κ satisfies (3.14).

The proof of Lemma 3.1 is complete. □

We invite the reader to compare Lemma 3.1, and its use below, with the computation of Lipschitz norms using the Callahan-Kosaraju decomposition (see [9], for example).

We close this section with an obvious remark on the Callahan-Kosaraju decomposition.

LEMMA 3.2. *Every $x \in E$ arises as an x'_ℓ for some $\ell (1 \le \ell \le L)$.*

Proof. Let $x \in E$. We pick $y \in E \setminus \{x\}$ to minimize $|x - y|$. Since the $E'_\ell \times E''_\ell (1 \le \ell \le L)$ form a Callahan-Kosaraju decomposition, we have $(x, y) \in E'_\ell \times E''_\ell$ for some ℓ. Fix such an ℓ.

Then we have $x, x'_\ell \in E'_\ell$, and $y, x''_\ell \in E''_\ell$, thanks to (0.10). Consequently, (0.9) gives $|x - x'_\ell| \le \kappa|x - y| < |x - y|$. Since y was picked to satisfy $|x - z| \ge |x - y|$ for all $z \in E \setminus \{x\}$, it follows that x'_ℓ does not belong to $E \setminus \{x\}$. On the other hand, $x'_\ell \in E'_\ell \subset E$, by (0.10). Thus, $x = x'_\ell$, proving the lemma. □

4. **Proof of Theorem** 1

Let E, σ be as in the hypotheses of Theorem 1, with $N = \#(E)$. We take $\kappa \in (0, 1)$ to be a constant, depending only on m and n, and small enough that Lemma 3.1 applies. We then introduce the Callahan-Kosaraju decomposition for κ, as in (0.7) ... (0.11). Also, we introduce the linear operator T, the sets $S(x)$ $(x \in E)$, and the coefficients $\lambda_\alpha(x, y)$ $(|\alpha| \leq m, x, y \in E)$, as in (0.3) ... (0.6).

As promised in Section 0, we take

(4.1) $S_\ell = S(x'_\ell) \cup S(x''_\ell) \cup \{x'_\ell, x''_\ell\}$ for $\ell = 1, \dots, L$.

(See (0.12).) We have already seen that conclusions (A), (B), (D) of Theorem 1 hold for our $S_1, \dots, S_L$. Our task here is to prove conclusion (C), i.e., $\|f\|_{(E,\sigma)} \sim \max\limits_{\ell=1,\dots,L} \|f\|_{(S_\ell,\sigma)}$, for any $f : E \longrightarrow \mathbb{R}$. Since obviously $\|f\|_{(E,\sigma)} \geq \|f\|_{(S_\ell,\sigma)}$ for each ℓ, our task is really to show that

(4.2) $\|f\|_{(E,\sigma)} \leq CM,$ where

(4.3) $M = \max_{\ell=1,\dots,L} \|f\|_{(S_\ell,\sigma)}$.

That is, we must show that there exists $F \in C^m(\mathbb{R}^n)$ satisfying

(4.4) $\|F\| \leq CM$, and $|F(x) - f(x)| \leq CM\sigma(x)$ for all $x \in E$, with M as in (4.3).

To find F, we will define polynomials $P_x \in \mathcal{P}$ for all $x \in E$, and check that they satisfy

(4.5) $|(\partial^\alpha P_x)(x)| \leq CM$ for $|\alpha| \leq m - 1,\ x \in E;$

(4.6) $|\partial^\alpha (P_x - P_y)(y)| \leq CM|x - y|^{m-|\alpha|}$ for $|\alpha| \leq m - 1,\ x, y \in E;$ and

(4.7) $|P_x(x) - f(x)| \leq CM\sigma(x)$ for all $x \in E$.

As in our discussion of (1.3) ... (1.8), the classical Whitney extension theorem then produces an $F \in C^m(\mathbb{R}^n)$, satisfying

$$\|F\| \leq C'M, \text{ and } J_x(F) = P_x \text{ for } x \in E,$$

and consequently satisfying (4.4).

To summarize: We will define $P_x \in \mathcal{P}$ for each $x \in E$, and prove (4.5), (4.6), (4.7), with M as in (4.3). This will complete the proof of Theorem 1.

For $x \in E$, we specify $P_x \in \mathcal{P}$ by stipulating that

(4.8) $(\partial^\alpha P_x)(x) = \sum_{y \in S(x)} \lambda_\alpha(x, y) f(y)$ for $|\alpha| \leq m - 1$.

(Clearly, given $x \in E$, there is one and only one $P_x \in \mathcal{P}$ satisfying (4.8).)

We begin proving (4.5) ... (4.7) for the above P_x. By definition (4.3), we have $\|f\|_{(S_\ell,\sigma)} \leq M$ for $\ell = 1, \dots, L$.

Thus, for each $\ell = 1, \dots, L$, there exists $F_\ell \in C^m(\mathbb{R}^n)$, with

(4.9) $\|F_\ell\| \le 2M$, and $|F_\ell(x) - f(x)| \le 2M\sigma(x)$ for all $x \in S_\ell$.

Fix such F_ℓ. For $\ell = 1, \dots, L$, we define $f_\ell : E \longrightarrow \mathbb{R}$, by setting

(4.10) $f_\ell(x) = f(x)$ for $x \in S_\ell$, $f_\ell(x) = F_\ell(x)$ for $x \in E \setminus S_\ell$.

In particular, (4.1) and (4.10) give

(4.11) $f_\ell = f$ on $S(x'_\ell) \cup S(x''_\ell) \cup \{x'_\ell, x''_\ell\}$, for each $\ell = 1, \dots, L$.

From (4.9) and (4.10), we obtain the estimates

$$\|F_\ell\| \le 2M, \text{ and } |F_\ell(x) - f_\ell(x)| \le 2M\sigma(x) \text{ for all } x \in E.$$

This shows that

$$\|f_\ell\|_{(E,\sigma)} \le 2M \text{ for } \ell = 1, \dots, L.$$

Therefore, applying (0.5) and (0.6), we learn that the function

(4.12) $\tilde{F}_\ell = Tf_\ell \in C^m(\mathbb{R}^n)$

satisfies the estimates

(4.13) $\|\tilde{F}_\ell\| \le CM$, and $|\tilde{F}_\ell(x) - f_\ell(x)| \le CM\sigma(x)$ for all $x \in E$.

Also, (4.12) and (0.4) yield

(4.14) $\partial^\alpha \tilde{F}_\ell(x'_\ell) = \sum_{y \in S(x'_\ell)} \lambda_\alpha(x'_\ell, y) f_\ell(y)$

and

(4.15) $\partial^\alpha \tilde{F}_\ell(x''_\ell) = \sum_{y \in S(x''_\ell)} \lambda_\alpha(x''_\ell, y) f_\ell(y)$,

for $|\alpha| \le m-1$ and $\ell = 1, \dots, L$.

Thanks to (4.11), equations (4.14), (4.15) may be rewritten in the form

(4.16) $\partial^\alpha \tilde{F}_\ell(x'_\ell) = \sum_{y \in S(x'_\ell)} \lambda_\alpha(x'_\ell, y)\, f(y)$,

and

(4.17) $\partial^\alpha \tilde{F}_\ell(x''_\ell) = \sum_{y \in S(x''_\ell)} \lambda_\alpha(x''_\ell, y)\, f(y)$,

for $|\alpha| \le m-1$ and $\ell = 1, \dots, L$.

Comparing (4.16) and (4.17) with (4.8), we conclude that

(4.18) $J_{x'_\ell}(\tilde{F}_\ell) = P_{x'_\ell}$, and $J_{x''_\ell}(\tilde{F}_\ell) = P_{x''_\ell}$.

Since $\|\tilde{F}_\ell\| \le CM$ by (4.13), it follows from (4.18) and Taylor's theorem that

(4.19) $|\partial^\alpha (P_{x'_\ell} - P_{x''_\ell})(x''_\ell)| \le C'M|x'_\ell - x''_\ell|^{m-|\alpha|}$ for $|\alpha| \le m-1$, $\ell = 1, \dots, L$;

and that

(4.20) $|(\partial^\alpha P_{x'_\ell})(x'_\ell)| \le CM$ for $|\alpha| \le m-1$, $\ell = 1, \dots, L$.

Also, from (4.11), (4.13), and (4.18), we see that

(4.21) $|P_{x'_\ell}(x'_\ell) - f(x'_\ell)| = |\tilde{F}_\ell(x'_\ell) - f_\ell(x'_\ell)| \le CM\sigma(x'_\ell)$ for each $\ell = 1, \dots, L$.

It is now easy to complete the proof of (4.5), (4.6) and (4.7).

In fact, by virtue of Lemma 3.2, our results (4.20) and (4.21) imply the estimates (4.5) and (4.7), respectively. Also, (4.19) implies (4.6), thanks to Lemma 3.1.

Thus, (4.5), (4.6), (4.7) hold for our P_x, with M given by (4.3). The proof of Theorem 1 is complete.

References

[1] E. BIERSTONE and P. D. MILMAN, $\mathscr{C}^m$-norms on finite sets and $\mathscr{C}^m$ extension criteria, *Duke Math. J.* **137** (2007), 1–18. MR 2008f:58009

[2] Y. BRUDNYI and P. SHVARTSMAN, Generalizations of Whitney's extension theorem, *Internat. Math. Res. Notices* **1994** no. 3 (1994), 129–139. MR 95c:58018 Zbl 0845.57022

[3] P. B. CALLAHAN and S. R. KOSARAJU, A decomposition of multidimensional point sets with applications to k-nearest-neighbors and n-body potential fields, *J. Assoc. Comput. Mach.* **42** (1995), 67–90. MR 96i:68033 Zbl 0886.68078

[4] C. FEFFERMAN and B. KLARTAG, Fitting a C^m-smooth function to data, I, *Ann. of Math.* **169** (2009), 315–346.

[5] ———, Fitting a C^m-smooth function to data, II, *Rev. Mat. Iberoamericana* **25** (2009), 49–273.

[6] C. FEFFERMAN, A generalized sharp Whitney theorem for jets, *Rev. Mat. Iberoamericana* **21** (2005), 577–688. MR 2007a:58009 Zbl 1102.58004

[7] ———, Interpolation and extrapolation of smooth functions by linear operators, *Rev. Mat. Iberoamericana* **21** (2005), 313–348. MR 2006h:58009 Zbl 1084.58003

[8] C. L. FEFFERMAN, A sharp form of Whitney's extension theorem, *Ann. of Math.* **161** (2005), 509–577. MR 2006h:58008 Zbl 1102.58005

[9] S. HAR-PELED and M. MENDEL, Fast construction of nets in low-dimensional metrics and their applications, SIAM *J. Comput.* **35** (2006), 1148–1184. MR 2007e:68071 Zbl 1100.68014

[10] B. MALGRANGE, *Ideals of Differentiable Functions*, *Tata Institute of Fundamental Research Studies in Mathematics* **3**, Tata Institute of Fundamental Research, Bombay, 1967. MR 35 #3446

[11] P. SHVARTSMAN, The Whitney extension problem and Lipschitz selections of set-valued mappings in jet-spaces, *Trans. Amer. Math. Soc.* **360** (2008), 5529–5550. MR 2415084

[12] J. VON NEUMANN, First draft of a report on the EDVAC, IEEE *Ann. Hist. Comput.* **15** (1993), 27–75. MR 94j:01007 Zbl 0944.01510

(Received March 9, 2006)

E-mail address: cf@math.princeton.edu
PRINCETON UNIVERSITY, DEPARTMENT OF MATHEMATICS, PRINCETON, NJ 08544, UNITED STATES