MAT 429 Notes

Max Chien

Fall 2025

Contents

Introduction

Chapter 1

Preliminaries

Definition 1.1

$C^m(\mathbb{R}^n)$ denotes the class of functions $f : \mathbb{R}^n \to \mathbb{R}$ which have m continuous derivatives on $\mathbb{R}^n$, and moreover are *bounded*. Accompanying this, a norm is defined by

$$\|f\|_{C^m(\mathbb{R}^n)} = \max_{|\alpha|\le m} \sup_{\mathbb{R}^n} |\partial^\alpha f|$$

Hence $(x \mapsto x) \notin C^m(\mathbb{R})$ for any m. Also, the norm is equivalent to taking the sum of such suprema. The problem at hand, then, is to consider a function $f : E \to \mathbb{R}$, where $E \subseteq \mathbb{R}^n$ is compact, and ask how we might extend it to a $C^m(\mathbb{R}^n)$ function on all of $\mathbb{R}^n$. Moreover, supposing that such a function exists, we ask for bounds on the $C^m(\mathbb{R}^n)$ norm of such an extension. We also seek formulas for the extension in terms of f, and for bounds on th ederivatives at points close to E.

We will also consider the particular case where E is a finite set. Using bump functions, the first question is trivial; we can easily create extensions that are in $C^\infty(\mathbb{R}^n)$. But this is not a particularly reasonable or useful choice for interpolation or extrapolation. Instead, we will attempt to find extensions which minimize the $C^m(\mathbb{R}^n)$ norm over solutions to the problem, up to some constant factor which depends only on m, n.

For the finite case, we will also attempt to find algorithms which can search for these minimizing solutions, and which minimize compute time and memory resources.

Theorem 1.1

Let $E \subseteq \mathbb{R}^2$ be finite with $|E| \geq 6$, and consider $C^2(\mathbb{R}^2)$. Let $f : E \to \mathbb{R}$. For any 6 distinct points $G \subseteq E$, let

$$g(G) = \inf_{\substack{F \in C^m(\mathbb{R}^n) \\ F|_G = f|_G}} \|F\|_{C^m(\mathbb{R}^n)}$$

Then there is a constant C, not depend on f, such that

$$\max_{\substack{G \subseteq E \\ |G| = 6}} g(G) \leq C \inf_{\substack{F \in C^m(\mathbb{R}^n) \\ F|_E = f|_E}} \|F\|_{C^m(\mathbb{R}^n)}$$

The above is not true for 5 instead of 6.

Another problem, which leads to the Whitney extension theorem, asks if and how we may extend functions which have not just their values but also their derivatives prescribed on an initial set.

To illustrate why the n dimensional case is significantly harder than the 1 dimensional case, suppose E consists of many points on the x axis, and a single point which is offset in the y direction by $+\varepsilon$. Without this point, we have no information about y direction derivatives at other points nearby on the x axis. However, if we do have that point, we can use it to extrapolate bounds on the y derivative at all of the other points. Nevertheless, our algorithm needs to understand how to identify that this data must be obtained from that point.

Similarly, let P be a third degree polynomial in x, y, and let E be its zero set, with a single point lying just off of the curve. Once again we can use this point to estimate the gradient of our function nearby on E.

For another example, consider $C^4(\mathbb{R}^3)$, and let G be an algebraic surface $\{x : P = 0\}$, and consider a curve in G $\{x : P = Q = 0\}$. If all the points in E lie in or near G, and moreover there are points which are close to our curve, the interpolation algorithm should identify these patterns just from the input data E.

We can also make this useful for experiments by considering data which is prescribed with errors. Concretely, consider $E \subseteq \mathbb{R}^n$ finite for m, n fixed, and for $x \in E$ suppose we are given $f(x)$ and $\sigma(x) > 0$. Then we want to find functions $F \in C^m(\mathbb{R}^n)$ such that

$$|F(x) - f(x)| < \sigma(x) \qquad \forall x \in E$$

and such that $\|F\|_{C^m(\mathbb{R}^n)}$ is as small as possible up to a constant factor C. There exist algorithms which can compute a solution for $|E| = N$ in $O(N \log N)$ time, print queries in $O(\log N)$ time, and use $O(N)$ memory.

Example 1.1

Homework: We are given points $x_1, x_2, \ldots, x_N \in \mathbb{R}$ and are working in $C^2(\mathbb{R})$, with initial data $f(x_k) = y_k$. We want an algorithm which computes extensions $F \in C^2(\mathbb{R})$

such that $\sup|F''|$ is optimal up to a universal constant C, in the sense that if $\tilde{F}$ is also an extension, then $\sup|F''| \leq C \sup\left|\tilde{F}''\right|$. Similarly we can look for computations of extension such that $\max_{k=0,1,2} \sup\left|F^{(k)}\right|$ is optimal to a constant. The algorithm should compute F with $O(N \log N)$ overhead, and computation of specific values $F(x)$ should occur in $O(\log N)$ time.

If we add error bars $\sigma(x_k) > 0$ to the above problem, it becomes an open problem, although it can be solved if a $1+\varepsilon$ factor is introduced and the constants allowed to depend on ε.

To extend this problem to arbitrary metric spaces (X, d), with $F : X \to \mathbb{R}^D$, we replace smooth functions with Lipschitz functions. Supposing for each $x \in X$ we are given a convex set $K(x) \subseteq \mathbb{R}^D$, we want to calculate F such that $F(x) \in K(x)$ for all x, and moreover the Lipschitz constant for F is as small as possible, up to a constant. This may be solved using a similar strategy as the previous 6-point method, with 2^D test points.

1.1 Well Separated Pairs Decomposition

Let $E \subseteq \mathbb{R}^n$ and $|E| = N$. Given $f : E \to \mathbb{R}$, compute the Lipschitz constant

$$L(f) = \max_{\substack{x,y \in E \\ x \neq y}} \frac{|f(x) - f(y)|}{|x - y|}$$

Clearly the constant may be computed in $O(N^2)$ time by using direct comparison of each pair. If we need to compute $L(f)$ merely within 1% error, we can do so in $O(N \log N)$ operations using **well separated pairs decomposition**.

Suppose we have to sets E', E'', and

$$\mathrm{dist}(E', E'') > A(\mathrm{diam}(E') + \mathrm{diam}(E''))$$

with A large, say 10^3. Then we may save a number of operations by computing

$$\max_{\substack{x' \in E' \\ x'' \in E''}} \frac{|f(x') - f(x'')|}{|x' - x''|}$$

The distances between the points may be bounded below easily by the separation assumption. We can also bound the numerator by computing the maximum and minimum values of f over points E', E''. The first step takes $O(1)$ operations, and the second takes $O(N)$ operations. To produce these sets, we will need more operations, but not $O(N^2)$.

Theorem 1.2: Well Separated Pairs Decomposition

Let $E \subseteq \mathbb{R}^n$ and $|E| = N$. Let $\mathcal{E} = E \times E \setminus \text{(diagonal)}$. Then $\mathcal{E}$ may be partitioned into $\nu_{\max}$ rectangular sets $E'_\nu \times E''_\nu, \nu = 1, \ldots, \nu_{\max}$ such that:

- For each ν, $\operatorname{dist}(E'_\nu, E''_\nu) \geq A(\operatorname{diam}(E'_\nu) + \operatorname{diam}(E''_\nu))$.
- $\nu_{\max} \leq CN$, where C depends on A and n.

Proof. Assume that $A \geq 10$ and $A \in \mathbb{N}$, say. We proceed using dyadic cubes. We'll use the convention that cubes are Cartesian products of half open intervals; that is, we begin with the unit cube (say) $[0,1)^n$ and subdivide further. Note that the problem is clearly invariant under scale, so we can assume $E \subseteq [0,1)^n$.

Choose any two dyadic cubes of side length $2^{-(k+A)}$ such that the distance between their centers is between $2^{-(k+1)}$ and 2^{-k}, and each contains a point in E. Call them Q'_ν, Q''_ν. Then set $E'_\nu = Q'_\nu \cap E, E''_\nu = Q''_\nu \cap E$, and take all such ν. This construction suffices as a partition, because any two points lie in exactly one such pair of dyadic cubes.

Let $\mathcal{Q}$ be the collection of all such (Q'_ν, Q''_ν), so that $\nu_{\max} = |\mathcal{Q}|$. Consider the subclass $\hat{\mathcal{Q}} \subseteq \mathcal{Q}$ which consists of those pairs (Q'_ν, Q''_ν) where Q'_ν, Q''_ν have sidelength $2^{-(k+A)}$, and they lie in the same dyadic cube of sidelength $2^{-(k-A)}$. In general the cube of sidelength $2^{-(k-A)}$ is much larger than the distance between the two cubes, so that nearly all pairs in $\mathcal{Q}$ should lie in $\hat{\mathcal{Q}}$.

For pairs which do not, Q''_ν lies in the margin of width $2^{-(k-2)}$ around the $2^{-(k-A)}$ dyadic square containing Q'_ν. If we shift the entire dyadic grid by $1/3$ in one of the dimensions, and repeat the procedure, we will capture some of the pairs in this margin. Repeating in each dimension, and each subset of the possible dimensions (so 2^n of them), we will capture all of the problematic pairs. We may overcount the pairs, and we also may create new bad pairs, but the point is that if one shows that the set of good pairs is bounded by CN for any one of these shifts, then the entire set of pairs is bounded by $2^n CN$. So it suffices to consider the good pairs.

Consider the tree of dyadic cubes beginning with $[0,1)^n$ and including all children which have at least one point of E. For any children which have at least two points of E, we include their children as well, and so on. Note that any path will terminate once a cube is reached which contains exactly one point of E. Call any node which has at least two children a "branching node." For instance, for any pair $(Q'_\nu, Q''_\nu) \in \hat{\mathcal{Q}}$, there is a branching node at most $2A$ generations above the pair (since at worst, the $2^{-(k-A)}$ cube containing them both will be a branching node). The association from (Q'_ν, Q''_ν) to their first mutual ancestor may be many-to-one, but it is bounded-to-one, since the ancestor has a bounded number of descendant pairs at most $2A$ generations deep. Thus we just need to count the number of branching nodes in order to count $\hat{\mathcal{Q}}$.

Note that the number of branching nodes in a tree with N leaves is at most $N-1$. $\square$

Having chosen a well separated pairs decomposition, we can pick a representative point

$(x'_\nu, x''_\nu) \in E'_\nu \times E''_\nu$ for each ν. Then (say $A = 10^6$), $L(f)$ is approximated by

$$\max_\nu \frac{|f(x'_\nu) - f(x''_\nu)|}{|x'_\nu - x''_\nu|} \le \max_{\substack{x,y \in E \\ x \neq y}} \frac{|f(x) - f(y)|}{|x-y|} \le (1 + 10^{-3}) \max_\nu \frac{|f(x'_\nu) - f(x''_\nu)|}{|x'_\nu - x''_\nu|}$$

The first inequality is clear. To prove the second, observe the following:

Proposition 1.3

For $A = 10^6$, if $|f(x'_\nu) - f(x''_\nu)| \le |x'_\nu - x''_\nu|$ for all representatives, then it follows that for any $x, y \in E$, $|f(x) - f(y)| \le (1 + 10^{-3})|x - y|$.

Proof. Suppose not. Then $|f(x) - f(y)| > (1 + 10^{-3})|x - y|$ for some $x, y \in E$. Pick such x, y where $|x - y|$ is minimzed among pairs of points satisfying this. In particular $x \neq y$, so $(x, y) \in E'_\nu \times E''_\nu$, and we take the representative pair (x'_ν, x''_ν). We have

$$\begin{aligned} |x - x'_\nu| &\le \operatorname{diam}(E'_\nu) \\ |y - x''_\nu| &\le \operatorname{diam}(E''_\nu) \\ \implies |x - y| &\ge \operatorname{dist}(E'_\nu, E''_\nu) \ge 10^6(\operatorname{diam}(E'_\nu) + \operatorname{diam}(E''_\nu)) \ge |x - x'_\nu| + |y - x''_\nu| \end{aligned}$$

In particular, since (x, y) were the minimal counterexample, (x, x'_ν) and (y, x''_ν) are not counterexamples. Hence

$$\begin{aligned} |f(x) - f(x'_\nu)| &\ge (1 + 10^{-3})|x - x'_\nu| \\ |f(y) - f(x''_\nu)| &\ge (1 + 10^{-3})|y - x''_\nu| \end{aligned}$$

Also by assumption,

$$|f(x'_\nu) - f(x''_\nu)| \le |x'_\nu - x''_\nu|$$

so

$$|f(x) - f(y)| \le |x'_\nu - x''_\nu| + (1 + 10^{-3})(|x - x'_\nu| + |y - x''_\nu|) \le (1 + 10^{-3})|x - y|$$

But this contradicts the assumption that (x, y) is a counterexample. □

1.2 Whitney Extension Theorem

Definition 1.2

Let $f \in C^m(\mathbb{R}^n)$, and let $x \in \mathbb{R}^n$. The **jet** of f at x i the function defined by

$$J_x F(y) = \sum_{|\alpha| \le m} \frac{1}{\alpha!} \partial^\alpha f(x)(y - x)^\alpha$$

Suppose we are given $E \subseteq \mathbb{R}^n$ compact, and for each $x \in E$ we are given an m degree polynomial P^x on $\mathbb{R}^n$. We want to know what conditions allow us to find a function

$f \in C^m(\mathbb{R}^n)$ such that $J_x(f) = P^x$ on E. This problem gives rise to the Whitney extension theorem.

Clearly a necessary condition is that the derivatives

$$\left\{ \left| \partial_y^\alpha P^x \Big|_{y=x} \right| : x \in E, |\alpha| \le m \right\}$$

are uniformly bounded, since any $C^m(\mathbb{R}^n)$ function must satisfy this. Of course, any individual P^x has uniformly bounded derivatives since E is compact, the point is that the family is uniformly bounded over $x \in E$.

Definition 1.3

Let $f : E \to \mathbb{R}$ be continuous and $E \subseteq \mathbb{R}^n$ compact. Then $\omega : \mathbb{R} \to \mathbb{R}$ is said to be a **modulus of continuity** if for all x, y,

$$|f(x) - f(y)| \le \omega(|x - y|)$$

Then for $|\alpha|, |\beta| \le m$, and $f \in C^m(\mathbb{R}^n)$, by Taylor's Theorem,

$$\left| \partial^\beta f(y) - \sum_{|\alpha| \le m - |\beta|} \frac{1}{\alpha!} \partial^{\alpha+\beta} f(x)(y-x)^\alpha \right| \le \omega(|x-y|)|y-x|^{m-|\beta|}$$

The same must be true, then, for the P^x, except that the modulus of continuity may be replaced with a uniform bound M, since the derivatives are bounded. Lastly, we have a kind of compatibility condition on the imposed derivatives:

$$\frac{\left| \partial^\beta P^y(y) - \sum_{|\alpha| \le m-|\beta|} \frac{1}{\alpha!} \partial^{\alpha+\beta} P^x(x)(y-x)^\alpha \right|}{|y-x|^{m-|\beta|}} \xrightarrow[x,y \in E]{|y-x| \to 0} 0$$

Proposition 1.4: Moving the Base Point

Let P be a real polynomial on $\mathbb{R}^n$ of degree at most m. Suppose there is K such that $|\partial^\alpha P(0)| \le K$ for all $|\alpha| \le m$. Then by expanding powers of $x = (x-y) + y$, there is C, C' only dependent on m, n such that $|\partial^\alpha P(y)| \le C'K$ whenever $|y| < C$.

Suppose $|\partial^\alpha P(0)| \le K\delta^{-|\alpha|}$ for all $|\alpha| \le m$. Then using the same C, C', $|\partial^\alpha P(y)| \le C'K\delta^{-|\alpha|}$ for all $|\alpha| \le m$, $|y| < C\delta$.

As an immediate consequence, if $|\partial^\alpha P(x)| \le K\delta^{-|\alpha|}$, then for $|x-y| < C\delta$, $|\partial^\alpha P(y)| \le C'K\delta^{-|\alpha|}$.

Theorem 1.5: Whitney Extension Theorem

Suppose $E \subseteq \mathbb{R}^n$ compact, and let $\{P^x : x \in E\}$ be a family of m degree polynomials on $\mathbb{R}^n$. Then there exists $F \in C^m(\mathbb{R}^n)$ such that $J_x f = P^x$ for all $x \in E$ if and only if there exists $M < \infty$ such that

1. $\left|\partial^\beta P^x(x)\right| \leq M$ for all $x \in E, |\beta| \leq m$,
2. $\left|\partial^\beta \left(P^y - P^x\right)(y)\right| \leq M|y-x|^{m-|\beta|}$,
3. $\dfrac{\left|\partial^\beta \left(P^y - P^x\right)(y)\right|}{|y-x|^{m-|\beta|}} \xrightarrow[x,y\in E]{|y-x|\to 0} 0.$

Moreover, if M is a constant such that $1, 2, 3$ hold, then F may be chosen such that $\|f\|_{C^m(\mathbb{R}^n)} \leq CM$, where C depends only on m, n.

Proof. Let Q^0 be a cube containing E. For any cube Q, let Q^* be the cube with the same center as Q but 5 times (say) the sidelength. We say that Q is "good" if $Q^* \cap E = \varnothing$. Clearly Q^0 is bad. Then we bisect Q^0 into 2^n subcubes. For each subcube, if the subcube is good, we call it a "Whitney cube" derived from Q^0 (note that good cubes are defined without reference to Q^0, but Whitney cubes must be obtained from this process). If it is bad, we bisect it and consider the children.

Let $x \notin E$. Then there is a neighborhood around x contained in $Q^0 \setminus E$, so the chain of cubes containing x will eventually terminate at a Whitney cube. On the other hand, if $x \in E$ then clearly this chain never terminates. So the Whitney cubes form an almost pairwise disjoint family of cubes whose union is precisely $Q^0 \setminus E$.

Consider any Whitney cube Q, and let Q^+ be the dyadic cube which is its direct ancestor (we know it has one since Q^0 is not a Whitney cube). Then $Q^* \cap E = \varnothing$ since Q is good, but $(Q^+)^* \cap E \neq \varnothing$ because otherwise Q^+ would be a Whitney cube. So $\operatorname{dist}(Q, E) \sim_n \operatorname{len}(E)$.

Moreover, if Q, Q' are two Whitney cubes with $\overline{Q} \cap \overline{Q'} \neq \varnothing$, then their sidelengths differ by at most a factor of 2. Indeed, suppose not. Then taking Q to have the longer sidelenght, it must have at least 4 times the sidelength of Q'. Then $10Q'$ lies within $5Q = Q^*$, which does not intersect E, so the parent of Q' should have been chosen as a Whitney cube. Also, any point x may lie in at most $A_{m,n}$ slightly dilated Whitney cubes $(1.01)Q$.

Now, we attempt to construct F. For any point $x \in E$, we are forced to define $F(x) = P^x(x)$. On the other hand, let x_Q, Q be a point in E and cube such that $\operatorname{dist}(x_Q, Q) \leq CS_Q$, where S_Q is the sidelength. Since Q is close to E, it is reasonable to choose to define $F_Q = P^{x_Q}$ on Q. However, this can give discontinuities when cubes touch each other. Therefore, we instead define F_Q on $(1.01)Q$, and use a partition of unity to patch these together.

Let Q_0 be a reference cube with sidelength 1, and define θ_0 to be a C^∞ function that is supported on $(1.01)Q_0$ and equal to 1 on Q_0. For any Q, define $\tilde{\theta}_Q = \theta_0(x)\left(\frac{x - z_Q}{S_Q}\right)$ where z_Q is the center of Q. Let C_α be defined for any multiindex such that

$$|\partial^\alpha \theta_0(x)| \leq C_\alpha$$

Then

$$\left|\partial^\alpha \tilde{\theta}_Q(x)\right| \leq C_\alpha S_Q^{-|\alpha|}$$

Then define

$$\theta_Q(x) = \begin{cases} \frac{\tilde{\theta}_Q(x)}{\sum_{Q'} \tilde{\theta}_{Q'}(x)}, & x \in E^c \\ 0, \quad x \in E \end{cases}$$

There are no issues of convergence in the denominator since only finitely many terms are supported. Also, we have

$$\left|\partial^\alpha \sum_{Q'} \tilde{\theta}_{Q'}(x)\right| \leq C_\alpha A_{m,n} 2^{|\alpha|} S_Q^{-|\alpha|}$$

for $x \in (1.01)Q$. So for such x, the derivatives $|\alpha| \leq m$ of the numerator and denominator are bounded and the denominator is at least 1, so

$$|\partial^\alpha \theta_Q(x)| \leq A'_{m,n} S_Q^{-|\alpha|}$$

We now return to constructing F. As before, for $x \in E$, we have $F(x) = P^x(x)$. For $x \notin E$, we instead use the following: let x_Q, Q be a point in E and cube as chosen above. Then we define

$$F(x) = \begin{cases} P^x(x), & x \in E \\ \sum_Q \theta_Q(x) P^{x_Q}(x), & x \notin E \end{cases}$$

It now remains to verify that F is C^m and that its norm is bounded. The fact that this is true on E follows from the assumptions. Fix some cube $\hat{Q}$. Then

$$F(x) = P^{x_{\hat{Q}}} + \sum_Q \theta_Q \cdot (P^{x_Q} - P^{x_{\hat{Q}}})$$

So using the first and second assumptions,

$$\left|\partial^\alpha P^{x_{\hat{Q}}}(x_{\hat{Q}})\right| \leq M$$

$$\left|\partial^\alpha \sum_Q \theta_Q \cdot (P^{x_Q} - P^{x_{\hat{Q}}})(x_{\hat{Q}})\right| \leq \left|\sum_Q \sum_{|\beta| \leq |\alpha|} \operatorname{coeff}(\alpha, \beta) \partial^\beta \theta_Q(x) \partial^{\alpha-\beta}(P^{x_Q} - P^{x_{\hat{Q}}})(x)\right|$$
$$\leq A_{m,n} \sum_{|\beta| \leq |\alpha|} |\operatorname{coeff}(\alpha, \beta)| A'_{m,n} S_Q^{-|\beta|} C M |x-y|^{m-(|\alpha|-|\beta|)} \leq A''_{m,n} M S_Q^{m-|\alpha|}$$

For any $Q, \hat{Q}$ where θ_Q is nonzero on $\hat{Q}$, then the distance between x_Q and $x_{\hat{Q}}$ is comparable to the sidelength. So by shifting the base point, we can shift the bounds back to be evaluated at x, and we get

$$|\partial^\alpha F(x)| \leq C_{m,n} M$$

So F is C^m off E, and the norm is bounded there. Fix $\overline{x} \in E$, $\varepsilon > 0$, and suppose $x \notin E$ but $|x - \overline{x}| < \eta$, with η small and dependent on ε. We will show that

$$\left|\partial^\alpha F(x) - \partial^\alpha P^{\overline{x}}(x)\right| \leq \varepsilon |x - \overline{x}|^{m-|\alpha|}$$

for all $|\alpha| \leq m$. (Note we only have to assume $x \notin E$ since this is true when $x \in E$ by assumption). We compute

$$\partial^\alpha \left(F - P^{\overline{x}}\right)(x) = \partial^\alpha \left(P^{x_{\hat{Q}}} - P^{\overline{x}}\right)(x) + \sum_Q \sum_{\beta \leq \alpha} \mathrm{coeff}(\alpha, \beta) \partial^{\alpha-\beta} \theta_Q(x) \partial^\beta \left(P^{x_Q} - P^{x_{\hat{Q}}}\right)(x)$$

We have $|\overline{x} - x| \geq c\delta_{\hat{Q}}$, and

$$\left|\overline{x} - x_{\hat{Q}}\right| \leq |\overline{x} - x| + \left|x - x_{\hat{Q}}\right| \leq |\overline{x} - x| + C\delta_{\hat{Q}} \leq C'\left|\overline{x} - x_{\hat{Q}}\right|$$

Since $\overline{x}, x_{\hat{Q}}$ are close, we have

$$\left|\partial^\alpha \left(P^{x_{\hat{Q}}} - P^{\overline{x}}\right)\right| \leq \varepsilon \left|x_{\hat{Q}} - \overline{x}\right|^{m-|\alpha|} \leq C\varepsilon |x - \overline{x}|^{m-|\alpha|}$$

For the second term, we have

$$\left|\partial^{\alpha-\beta} \theta_Q(x)\right| \leq C\delta_Q^{-(|\alpha|-|\beta|)}$$

We also have

$$\left|\partial^\beta (P^{x_Q} - P^{x_{\hat{Q}}})(x)\right| \leq \varepsilon C \delta_{\hat{Q}}^{m-|\beta|} \leq C'\varepsilon |\overline{x} - x|^{m-|\beta|}$$

Once again we may shift the basepoint. So in total we have shown that

$$\left|\partial^\alpha \left(F - P^{\overline{x}}\right)(x)\right| \leq \varepsilon |x - \overline{x}|^{m-|\alpha|}$$

and the desired conclusions follow. □

1.3 Extension on Finite Sets

We return now to our problem of optimally extending functions on finite sets. Fix $m, n \geq 1$, and let $E \subseteq \mathbb{R}^n$ be finite with N points. For $x \in E$, suppose we are given P^x a polynomial in $\mathbb{R}^n$ of degree at most m.

Let M be the least positive value such that

$$|\partial^\alpha P^x(x)| \leq M$$
$$\frac{|\partial^\alpha (P^x - P^y)(y)|}{|x - y|^{m-|\alpha|}} \leq M$$

Let M^* be

$$M^* := \inf \left\{ \|F\|_{C^\infty : F \in C^m, J_x F = P^x \forall x \in E} \right\}$$

Then there are c, C dependent only on m, n such that

$$cM \leq M^* \leq CM$$

The upper bound is immediate from Whitney's extension theorem. On the other hand, if there is an extension F, we can take the Taylor polynomials as P^x, and thus $M \leq \|F\|_{C^m(\mathbb{R}^n)}$.

Calculating M reduces to a linear programming problem. Consider the vector space of tuples $((P^x)_{x\in E}, M) \in \mathbb{R}^{N(M+1)+1}$. The set of tuples for which M is an upper bound and $P^x(x) = f(x)$ is defined by a finite list of inequalities and equalities, linear in the coefficients of P^x, so it is a convex polytope. We want to minimize M in this polytope, which is a linear program.

1.4 Minimization on $C^2(\mathbb{R})$

We seek to solve the problem of extend function values on a finite set, such that the second derivative is minimal up to a constant. Let $E = \{x_1, \ldots, x_N\} \subseteq \mathbb{R}$ and $f(x_1), \ldots, f(x_N) \in \mathbb{R}$ be given. We need to find $F \in C^2(\mathbb{R})$ which agrees with f such that for any other extension $\tilde{F}$, $\sup|F''| \leq C \sup\left|\tilde{F}''\right|$.

Suppose without loss of generality that $x_1 < \ldots < x_N$. Let F be an extension. Fix some ν and define $I_{\nu-1} = [x_{\nu-1}, x_\nu], I_\nu = [x_\nu, x_{\nu+1}], I_{\nu_1} = [x_{\nu+1}, x_{\nu+2}]$. Define

$$m_\nu = \frac{f(x_{\nu+1}) - f(x_\nu)}{|I_\nu|}$$

By the mean value theorem, there is a point $y_\nu \in I_\nu$ such that $F'(y_\nu) = m_\nu$. Applying the mean value theorem again, there is a point $z_\nu \in I_\nu \cup I_{\nu+1}$ such that

$$F''(z_\nu) = \frac{m_{\nu+1} - m_\nu}{y_{\nu+1} - y_\nu}$$

so

$$\sup|F''| \geq \max_\nu \frac{|m_{\nu+1} - m_\nu|}{|I_{\nu+1}| + |I_\nu|}$$

Define

$$\tilde{m}_\nu = \begin{cases} m_\nu, & |I_\nu| < |I_{\nu-1}| \\ m_{\nu-1}, & |I_\nu| > |I_{\nu-1}| \end{cases}$$

and define

$$\tilde{L}_\nu(x) = f(x_\nu) + \tilde{m}_\nu(x - x_\nu)$$

which is the linear extrapolation from the left or right, depending on which direction is closer. On I_ν, define

$$F(x) = \theta_\nu(x)\tilde{L}_\nu(x) + (1 - \theta_\nu(x))\tilde{L}_{\nu+1}(x)$$

where θ_ν is a smooth function which is 1 on left third of I_ν and zero on the right third. For this F,

$$\sup|F''| \leq C\tilde{M}$$

where

$$\tilde{M} = \max_\nu \frac{|\tilde{m}_{\nu+1} - \tilde{m}_\nu|}{|I_\nu|}$$

So we have $M \leq \|F\| \leq C\tilde{M}$, and we just need to show $\tilde{M}$ is comparable to m. We claim that $|\tilde{m}_\nu - m_\nu| \leq 2|I_\nu|M$. If $|I_\nu| \leq |I_{\nu-1}|$ then this is true automatically since the left side is zero. Otherwise,

$$\tilde{m}_\nu = m_{\nu-1} \implies |\tilde{m}_\nu - m_\nu| = |m_{\nu-1} - m_\nu| = \frac{|m_{\nu-1} - m_\nu|}{|I_{\nu-1}| + |I_\nu|}(|I_{\nu-1}| + |I_\nu|) \leq 2M|I_\nu|$$

so

$$\tilde{M} = \max_\nu \frac{|\tilde{m}_{\nu+1} - \tilde{m}_\nu|}{|I_\nu|} \leq 2\max_\nu \frac{|\tilde{m}_\nu - m_\nu|}{|I_\nu|} + \max_\nu \frac{|m_{\nu+1} - m_\nu|}{|I_\nu|} \leq 5M$$

To minimize the entire $C^2(\mathbb{R})$ norm rather than just the second derivative, we define the same F but take a modified approach. First we solve the case where $x_{\nu+1} - x_\nu \leq 1$ for all ν. Define m_ν, M, F as before and define

$$M^+ = \max\{|f(x_\nu)|, |m_\nu|, M\}$$

Then $\|F\|_{C^2} \geq M^+$. We showed previously that $\sup|F''| \leq CM \leq CM^+$. To estimate $|F|$, pick some interval I_ν, and write

$$\begin{aligned}\left|\tilde{L}_\nu(x)\right| &= |f(x_\nu) + m_\nu(x - \nu) + (\tilde{m}_\nu - m_\nu)(x - x_\nu)| \\ &\leq \max\{|f(x_\nu)|, |f(x_{\nu+1})|\} + 2M|I_\nu|(x - x_\nu) \\ &\leq 3M^+\end{aligned}$$

so $|F(x)| \leq 3M^+$ as well since it is a convex combination. Lastly,

$$\begin{aligned}|F'(x)| &= \left|\tilde{L}'_\nu(x) + (1 - \theta_\nu(x))[\tilde{L}'_{\nu+1}(x) - \tilde{L}'_\nu(x)] + \theta'_\nu(x)[\tilde{L}_\nu(x) - \tilde{L}_{\nu+1}(x)]\right| \\ &\leq |\tilde{m}_\nu| + |(1 - \theta_\nu(x))[\tilde{m}_{\nu+1} - \tilde{m}_\nu]| + \left|\theta'_\nu(x)[\tilde{L}_\nu(x) - \tilde{L}_{\nu+1}(x)]\right| \\ &\leq M^+ + 2M^+ + 6|\theta'_\nu(x)|\left|\tilde{L}_\nu(x) - \tilde{L}_{\nu+1}(x)\right| \leq 3M^+ + \frac{C}{|I_\nu|}\left|\tilde{L}_{\nu+1}(x) - \tilde{L}_\nu(x)\right|\end{aligned}$$

We have

$$\begin{aligned}\left|\tilde{L}_{\nu+1}(x) - \tilde{L}_\nu(x)\right| &\leq \frac{|f(x_{\nu+1}) - f(x_\nu)|}{|I_\nu|}|I_\nu| + |\tilde{m}_{\nu+1} - \tilde{m}_\nu||x - x_{\nu+1}| + |\tilde{m}_\nu||x_{\nu+1} - x_\nu| \\ &\leq \tilde{M}|I_\nu| + 2M|I_\nu| + M^+|I_\nu| \leq CM^+|I_\nu|\end{aligned}$$

so

$$|F'(x)| \leq CM^+$$

and we are finished.

1.5 Minimization on $C^2(\mathbb{R}^2)$

Now we consider the same problem as before, but minimizing over $\mathbb{R}^2$ rather than $\mathbb{R}$. We consider minimization of

$$\|F\|_{C^2(\mathbb{R}^2)} = \max\left\{\sup|F|, \sup|\nabla F|, \sup\left|\nabla^2 F\right|\right\}$$

or alternatively merely minimizing

$$\sup\left|\nabla^2 F\right|$$

The second problem is much more amenable to scaling arguments, since the three quantities in the first do not scale at the same rate.

On the other hand, under a generic regular change of variables $y = \phi(x)$, then setting $\tilde{F} = F \circ \phi$,

$$\frac{\partial \tilde{F}}{\partial x_j} = \sum_k \frac{\partial F}{\partial y_k} \frac{\partial y_k}{\partial x_j}$$

$$\frac{\partial^2 \tilde{F}}{\partial x_i \partial x_j} = \sum_{k,l} \frac{\partial^2 F}{\partial y_k \partial y_l} \frac{\partial y_k}{\partial x_j} \frac{\partial y_l}{\partial x_i} + \sum_k \frac{\partial F}{\partial y_k} \frac{\partial^2 y_k}{\partial x_i \partial x_j}$$

Under a change of variables, $\|F\|_{C^2(\mathbb{R}^2)}$ only changes by a bounded quantity; the same is not true for $\sup\left|\nabla^2 F\right|$.

We begin by considering some simple cases. When E is contained in the x-axis, the problem reduces to the one dimensional case. When there is a C^2 graph $x_2 = \psi(x_1)$ fit to E with $|\psi'|, |\psi''| < c$ small, we can shift this to the straight line by using the change of variables

$$\begin{cases} \tilde{x}_1 = x_1 \\ \tilde{x}_2 = x_2 - \psi(x_1) \end{cases}$$

This works well for $\|F\|_{C^2(\mathbb{R}^2)}$ but not for $\sup\left|\nabla^2 F\right|$. In that case, we can pick a point z off of E, prescribe the function and gradient $F(z), \nabla F(z)$ there, and minimze. We would then minimize over the prescribed values. Define

$$L(x) = F(z) + \nabla F(z) \cdot (x - z)$$

to be the linear extrapolation from z. Then we need

$$F - L = f - L$$

at points of E, and

$$\partial^2 (F - L) = \partial^2 F$$

But in particular $F - L = 0$ and $\nabla(F - L)$ at z, so by Taylor's theorem, if M bounds the second derivatives of F, then

$$|\nabla F| \leq CM, |F| \leq C' M^2$$

so this is similar to $\|F\|_{C^2(\mathbb{R}^2)}$. This helps us solve the issue of change of variables instability.

Theorem 1.6

Let $Q = [1, 1]^2$ and let $F : Q \to \mathbb{R}$ be C^2. Assume also that $F(0,0) = 0, F_x(0,0) \leq 5$, and $F_y(0,0) \geq A$, and that $|F_{xx}|, |F_{xy}|, |F_{yy}| \leq 1$ on Q. Then the zero set of F can be written as $\{(x, \psi(x)) : x \in [-1, 1]\}$ for some ψ with $\psi(0) = \psi'(0) = 0$, $|\psi''| \leq 100/A$.

When z_1, z_2, z_3 are three points such that they do not lie on a C^2 graph with small derivatives, we are essentially in the case of noncharacteristic initial data for PDEs. If we assume that $\left|\nabla^2 F\right|$ is globally bounded by some M, we can determine $\nabla F(z_i)$ up to some error given just the prescribed function values $F(z_i)$.

To see this, without loss of generality assume that $z_1 = (x_1, y), z_2 = (x_2, y)$ share the same second component. Then

$$\frac{F(z_2) - F(z_1)}{|z_2 - z_1|} = \frac{1}{|x_2 - x_1|} \int_{x_1}^{x_2} F_x(x, y) \, \mathrm{d}x$$

By the assumption on $\left|\nabla^2 F\right|, |F_x(x, y) - F_x(z_1)| \leq M|z_1 - z_2|$. So we get an estimate for the horizontal component of the gradient.

To look at the error in the gradient, let F_1, F_2 be two extensions of the initial data (both with $\left|\nabla^2 \cdot\right| \leq M$), and define $\tilde{F} = (F_2 - F_1)/2M$. Then $\tilde{F}$ is zero at z_1, z_2, z_3, and $\left|\nabla^2 \tilde{F}\right| \leq 1$. Also, by the above argument, $\nabla_x \tilde{F}(z_1) \leq 2$. So by the Theorem, if $\tilde{F}_y(0, 0)$ is large, then z_1, z_2, z_3 would lie on a nice C^2 graph. This is not the case, so we also have a bound on the vertical derivative error.

Thus we can pick any nice extension to estimate the vertical gradient, and in particular we take the linear fit between the three points. This has zero second derivative, so any other function has a difference of ∇F by at most, say, $100M$. At any other point it can change by at most $(100 + \operatorname{diam}(Q))M$ by our bound on $\left|\nabla^2 F\right|$.

When Q is not $[-1, 1]^2$ but instead a square of sidelength δ_Q centered at the origin, we can define $\tilde{F}$ on $Q^0 = [-1, 1]^2$ by

$$\tilde{F}(x, y) = \delta_Q^{-2} F(\delta_Q x, \delta_Q y)$$

so that $\sup\left|\nabla^2 \tilde{F}\right| = \sup\left|\nabla^2 F\right|$. If $y = \psi(x)$ is the zero set of F, then

$$\tilde{\psi}(\tilde{x}) = \delta_Q^{-1} \psi(\delta_Q \tilde{x})$$

If $\left|\tilde{\psi}'\right| \leq 5$ and $\left|\tilde{\psi}''\right| \leq 10^{-2}$, then $|\psi'| \leq 5$ as well and $|\psi''| \leq 10^{-2}/\delta_Q$.

To construct a minimal extension, we will use this strategy, but since it depends on the configuration of the points, we will subdivide into problems with nice configurations and patch them together with a partition of unity. Let Q^0 be some initial square containing E. We subdivide Q^0 into quadrants repeatedly, accepting a square if

$$E \cap 20Q$$

lies in a graph of the form $y = \psi(x)$ (or $x = \psi(y)$) such that $|\psi'| \leq 5, |\psi''| \leq 10^{-2}/\delta_Q$. (This is guaranteed to terminate since it is always true if $E \cap 20Q$ contains exactly 1 or 2 points.) If Q^0 satisfies this property then we can change variables so that E sits on the x-axis, solve the problem in one dimension, and conclude.

To apply this to an extension problem, we use a partition of unity on subproblems. First consider the problem where $E \cap (1.1Q)$ lies on a nice curve, there are three points $D = \{d_1, d_2, d_3\}$ which do not lie on a nice curve, and F is prescribed by f on $[E \cap (1.1)Q] \cup D$. Let $L_Q(x,y) = a + b_1 x + b_2 y$ be the linear function which agrees with f on D. Then $\|f - L_q\|_{C^2}$ is comparable to discrete approximations of $f - L_Q$ at one, two, or three consecutive points of E, depending on the order of the derivative. Since L_Q is determined by f on D, these values are determined by the maximum evaluations of functions of values of f at at most 6 points. By looking at the change in the gradient, we have

$$|\nabla F - \nabla L_Q| \leq C \max_{|\alpha|=2} |\partial^\alpha F| \delta_Q$$

and

$$|F - L_Q| \leq C \max_{|\alpha|=2} |\partial^\alpha F| \delta_Q^2$$

so

$$\max_{|\alpha|\leq 2} \delta_Q^{|\alpha|-2} |\partial^\alpha (F - L_Q)| \leq C \max_{|\alpha=2|} |\partial^\alpha F|$$

We can extend this to the general problem by subdividing our cube until we can solve the subproblem, then patch the solutions together with a partition of unity (for the subproblems we reqire $F_Q = \theta_Q f$). For fixed E, the mapping from $f \mapsto F$ is linear. Also, $F(x)$, as well as its Taylor polynomial of order 2, depend on the value of f at a bounded number of points.

If we can show this is optimal up to a constant, then we can show that taking the maximum of six-point evaluations on E gives the optimal norm up to a constant, without having to actually construct a function. We define

$$M = \max \begin{cases} \max\limits_{\substack{|\alpha|\leq 2 \\ x \in 1.1Q \\ Q}} \delta_Q^{|\alpha|-2} |\partial^\alpha (F_q - L_q)| \\ \max\limits_{x \in E} |f(x)| \\ \max\limits_{Q} |\nabla L_Q| \\ \max\limits_{\substack{Q, \tilde{Q} \text{ touching} \\ |\alpha|\leq 1 \\ x \in 1.1Q}} \delta_Q^{|\alpha|-2} \left| \partial^\alpha (L_Q - L_{\tilde{Q}}) \right| \end{cases}$$

Then we will show that $\|F\| \leq CM$ and any other interpolant $F^\#$ satisfies $\left\|F^\#\right\| \geq cM$. M is determined by the maximum of quantities that are calculated by evaluating at at most six points.

First we show that $\left\|F^\#\right\| \geq cM$. We just need to show this for each of the terms inside the overall maximum. Since $F^\# = f$ on E it is true for the second. For the first, we know that $F^\# = L_Q$ on sets D as defined above. Then it follows that on $1.1Q$,

$$\left|\nabla (F^\# - L_Q)\right| \leq C\delta_Q \left\|F^\#\right\|$$
$$\left|F^\# - L_Q\right| \leq C\delta_Q^2 \left\|F^\#\right\|$$

and of course

$$\left|\nabla^2 (F^\# - L_Q)\right| = \left|\nabla^2 F^\#\right| \leq \left\|F^\#\right\|$$

Also it follows from the above that

$$|\nabla L_Q| \leq \left|\nabla F^{\#}\right| + \left|\nabla(F^{\#} - L_Q)\right| \leq C\left\|F^{\#}\right\|$$

since $\delta_Q \leq 1$. For the fourth term, we know that $\delta_Q, \delta_{\tilde{Q}}$ are comparable since they touch, so

$$\left|L_Q - L_{\tilde{Q}}\right| \leq \left|F^{\#} - L_Q\right| + \left|F^{\#} - L_{\tilde{Q}}\right| \leq C\delta_Q^2\left\|F^{\#}\right\|$$

and similarly

$$\left|\nabla(L_q - L_{\tilde{Q}})\right| \leq C\delta_Q\left\|F^{\#}\right\|$$

So $\left\|F^{\#}\right\| \geq cM$. To show the other direction, we have

$$F = \sum_Q \theta_Q L_Q = \sum_Q \theta_Q L_Q + \sum_Q \theta_Q(F_Q - L_Q)$$

We have

$$|L_Q| \leq C \max_{x\in E}|f(x)| \leq CM$$

and

$$|F_Q - L_Q| \leq \delta_Q^2 M \leq M$$

so

$$|F| \leq CM$$

Now fix a cube $\tilde{Q}$. Write

$$F = L_{\tilde{Q}} + \sum_Q \theta_Q(L_Q - L_{\tilde{Q}}) + \sum_Q \theta_Q(F_Q - L_Q)$$

so

$$\partial^\alpha F = \partial^\alpha L_{\tilde{Q}} + \sum_Q \sum_{\beta+\gamma=\alpha} c_{\beta,\gamma}\partial^\beta\theta_Q\partial^\gamma(L_Q - L_{\tilde{Q}}) + \sum_Q \sum_{\beta+\gamma=\alpha} c_{\beta,\gamma}\partial^\beta\theta_Q\partial^\gamma(F_Q - L_Q)$$

The first term is bounded by M since it is in the max. The second sum is actually a sum over Q which touch $\tilde{Q}$, since that is where θ_Q is supported. On these cubes,

$$\partial^\beta\theta_Q\partial^\gamma(L_Q - L_{\tilde{Q}}) = C\delta_Q^{-|\beta|}M\delta^{2-|\gamma|} = C\delta_Q^{2-|\alpha|}M \leq CM$$

The last term proceeds identically. So on $\tilde{Q}$, $|\partial^\alpha F| \leq CM$ for all α, and this is true for all $\tilde{Q}$. So we are done.