Interpolation and approximation

Lectures by Professor Charles Fefferman
Scribe: Holden Lee

October 4, 2018

Contents

Chapter 1

Interpolation

1 Interpolation

Interpolation means various things. In the simplest iteration, I give you values (or approximate values) of a function at finitely many points and ask you to guess the funciton. The function can be anything so we have to make restrictions. Take some function space norm, like C^m norm or Sobolev norm. Find the function with norm as small as possible or almost as small as possible. There's no reason to think the correct answer is the function with smallest norm.

Is there a function that agrees with the data whose norm is bounded by the given bound, exactly or approximately? If no, the data is not consistent with the smoothness assumption. If yes, we would like to find it.

We are interested in theorems and algorithms, implemented on a computer.

There is a theory for interpolating functions of 1 variable. Fit a polynomial to adjacent points; fudge the functions so that they match perfectly. On the line, the points come in order. Given sets in the plane, there's no order; the geometry is much more subtle.

For example, in $\mathbb{R}^2$, suppose we are given a bunch of points on the x-axis, and 2 points off the axis. Is it possible to interpolate the data with $\|F\|_{C^2} \lesssim 1$? If the values of the y derivative are sufficiently close, we can interpolate, otherwise no. We had better not just examine nearby points; information can come from far away. Any algorithm that constructs an interpolate by looking at the 100 points closest to it is bound to fail. If the set looks of this, we have to recognize this and respond to it.

Let E be the set and $N = \#E$. There are algorithms that take time $O(N \log N)$.

I'm only interested in algorithms that always work, not those that require geometric conditions on the points.

Suppose we want $\|F\|_{C^4} \lesssim 1$. Draw the zero set of a polynomial of degree 3. Suppose all but a few points lie on that set. This is really the same problem as before. Given F, I will not see a difference with $F + GP$ except on the points that are off the zero-set. To test this hypothesis, you had better recognize this is going on, and that the clusters with points are outside have the important information and respond to it. Nobody tells you the zero set is around; you are only given the finite set of points. You need to recognize the curve is sitting there. Already, this subject has input from real algebraic geometry.

This is not the worst possible case for C^m. Consider $C^4(\mathbb{R}^3)$. Imagine there's some algebraic surface which is the zero set of a polynomial of low degree and on the surface there is some curve, which is the intersection of the sufrace with another. All but a few lie really close to the surface, and of the ones close to the surface, all but a few lie close to the curve. That affects where the information is coming from. (Real algebraic geometry was for a long time disjoint from anything else, but it's starting to make connections. We won't take textbook theorems from real algebraic geometry and apply them.)

Interpolation with exact values are well understood in $C^m(\mathbb{R}^n)$. They're also well understood in Sobolev spaces $W^{m,p}(\mathbb{R}^n)$. We require $p \geq 1$. If $p > \frac{n}{m}$, then $W^{m,p}(\mathbb{R}^n) \subset C^0$. The problem makes sense in this regime. A lot is known when $p > n$. For $\frac{n}{m} < p < n$, not much is known.

Let X be a Banach space of continuous functions on $\mathbb{R}^n$. Assume $X = C^m(\mathbb{R}^n)$. (There are obvious modifications for other spaces.) Let $E \subset \mathbb{R}^n$ be finite with $\#E = N$. Consider $f : E \to \mathbb{R}$. Define

$$\|f\|_{X(E)} = \inf \left\{ \|F\|_X : F \in X \text{ such that } F = f \text{ on } E \right\}. \tag{1.1}$$

Often this inf is not a min. Say that F is an A-optimal extension of f if $F = f$ on E and $\|F\|_X \leq A \|f\|_{X(E)}$.

The two main problems are

1. Compute the order of magnitude of $\|f\|_{X(E)}$. (We say 2 real numbers are the same order of magnitude if the ratio is bounded above and below by constants independent of the data. We want to compute a number guaranteed to have the same order of magnitude.)

2. Compute a A-optimal extension of f.

Computing a function is a more delicate thing because a computer will only deal with a finite number of values, while a function has infinitely many. Sit at a terminal, enter the data. The computer then displays "Please wait, I'm thinking about it." The computer then executes an algorithm. "Thank you for your patience, I now understand the interpolant F and will respond to your queries." Give a point, the computer will respond with at least F at that point. We can also demand the derivatives up to order m at that point.

This is an extremely demanding notion of computing a function. Take a Bessel function, or even $\sin x$. Depending on what the computer does—suppose it does basic arithmetic. No amount of computation will yield $\sin x$ exactly. It computes $\sin x$ to a given accuracy. But I want *the* value.

I imagine a computer with standard von Neumann architecture. Let's say the computer can deal with exact real numbers. (My coauthors and I have taken into account round-off error into account rigorously; let's not deal with that!) It has RAM, flow of control. Given two real numbers, it can add, subtract, multiply, or divide them; we assume no round-off error. The computer can take one number in a register and put it in RAM, fetch from RAM, has input and output.

For some results, I assume the computer can compute 2^m, and take logarithms.

I'm only interested in efficient algorithms, those that make minimal use of resources of the computer. There are 2 relevant resources: number of computer operations (multiply, fetch, etc.), and the size of the RAM (how many real numbers to store).

Let's make some trivial lower bounds. We had better read the problem, which takes time N. One could imagine an online version which throws away data as it arrives, but it's reasonable to think that the memory required is also N.

For problem 2, we also have a lower bound of N. We had better store the problem, e.g., if I query a point I gave the computer, it needs to remember what the value was. A lower bound for the query work is 1.

For problem 1, there are algorithms that solve this problem where the work is $O(N \log N)$ and the memory is $O(N)$. I believe this is sharp.

For problem 2, there is an A-algorithm, for which the one-time work is $O(N \log N)$, storage is $O(N)$, and query work is $O(\log N)$. Again I believe this is sharp. There are 3 kinds of resources which one would like to minimize: one-time work, storage, and query work. One can optimize all of them at the same time.

There's a reason this solves no practical problems; the constant A is too big. It depends only on the choice of the function space. It is large because of one particular lemma deep inside the machine of the proof. I continue to hope that one can remove the lemma and replace it by something else.

We use Whitney's extension problem and Whitney's extension theorem, and something from computer science called the well-separated pairs decomposition.

1.1 Whitney extension

Problem 1.1.1 (Whitney, 1934)**:** Given $W \subset \mathbb{R}^n$ compact, $m \geq 1$, $f : E \to \mathbb{R}$, does there exist $F \in C^m(\mathbb{R}^n)$ such that $F = f$ on E? If so, how small can we take its norm? What can we say about its derivatives of F (up to order m) at a given point? Can we take F to depend linearly on f? (Define $X(E) = \{f : E \to \mathbb{R} : \exists F \in X \text{ such that } F = f \text{ on } E\}$, $\|f\|_{X(E)} = \inf \{\|F\|_X : F \in X, F = f \text{ on } E\}$. Does there exist $T : X(E) \to X$ bounded linear map such that $Tf|_E = f$ for all $f \in X(E)$?)

Whitney solved this in 1 dimension (blah blah blah blah part 1; part 2 never appeared).

In addition to this he proved the very important Whitney extension theorem. For $F \in C^m(\mathbb{R}^n)$, $x \in \mathbb{R}^n$, define the Taylor expansion

$$J_x(F) : y \mapsto \sum_{|\alpha| \leq m} \frac{1}{\alpha!} (\partial^\alpha F(x))(y - x)^\alpha. \tag{1.2}$$

Whitney's question is the following: Suppose we are given at every point of $E \subset \mathbb{R}^n$ a Taylor polynomial, $\vec{P} = (P^x \in \mathcal{P})_{x \in E}$, where $\mathcal{P}$ is the set of polynomials of degree $\leq m$. This is called a **Whitney field**. How can we tell whether there exists $F = C^m(\mathbb{R}^n)$ such that $J_x(F) = P^x$ for all $x \in E$? What are necessary conditions? This is a simpler question; we need to understand it first.

Suppose I gave you 30 minutes; you are not allowed to leave the room unless you come up with necessary conditions. Those obviously necessary conditions you come up with will be sufficient; there is a procedure to construct the function.

Here are necessary conditions.

1. Denote $\partial^\alpha P^x(x) = (\partial_z^\alpha P^x(z))|_{z=x}$. Then $|\partial^\alpha P^x(x)| \le M$ for $x \in E$, $|\alpha| \le m$.

2. $|\partial^\alpha (P^x - P^y)(x)| \le M|x-y|^{m-|\alpha|}$. for $x, y \in E$, $|\alpha| \le m$.

   (If 2 points x, y are close then P^x and P^y are close; Taylor's theorem with remainder tells us how close.)

   (Let's declare $0^0 = 0$ for this condition.)

   (We haven't extracted all the omph yet. We haven't use the fact that they're continuous. Using the modulus of continuity we get the following.)

3. $\frac{|\partial^\alpha(P^x - P^y)(x)|}{|x-y|^{m-|\alpha|}}$ as $|x-y| \to 0^+$, $x, y \in E$.

Theorem 1.1.2 (Whitney's extension theorem , 1934)**.** *These conditions imply that there exist $F \in C^m(\mathbb{R}^n)$ with C^m-norm $\le CM$ such that $J_X(F) = P^x$ for all $x \in E$, where C depends only on m, n.*

This depends on a very fundamental idea that had a huge influence on analysis.

1.2 Well-separated pairs

Given $E \subset \mathbb{R}^n$, $\#E = N$, $f : E \to \mathbb{R}$, compute $\|f\|_{\text{Lip}} = \max_{x,y \in E, x \ne y} \frac{|f(x)-f(y)|}{|x-y|}$. This is trivial; a high school student learning programming can do it. Looping over all x and y takes $O(N^2)$ operations because you look at every x and y. Is that the best you can do?

Suppose you allow yourself wiggle room; you compute it to a small error, say 10^{-3}. Using the well-separated pairs decomposition theorem, we can compute $\|f\|_{\text{Lip}}$ to within a factor 1 ± 10^{-3} using $O(N \log N)$ operations. The constant depends on the accuracy required and the dimension n.

Suppose we just tried to beat N^2, how can we do it? Can we compute many of the quotients at the same time? Suppose

$$E \times E \backslash \text{Diag} \supset E' \times E''. \tag{1.3}$$

We compute $\max_{x \in E', y \in E''} \frac{|f(x)-f(y)|}{|x-y|}$ in far fewer steps than $\#E' \times \#E''$.

Suppose E', E'' have good geometry—they are well-separated: the distance between E' and E'' is large compared to the diameter of E', E''. Then I can take representative points

$\overline{x} \in E'$, $\overline{y} \in E''$ and comute

$$\frac{1}{|\overline{x}-\overline{y}|} \max_{x\in E', y\in E''} |f(x)-f(y)|. \tag{1.4}$$

Take $f(x)$ as small as possible and $f(y)$ as large as possible, or vice versa. The number of operations has decreased from the product of $\#E'$ and $\#E''$ to the sum.

We aim to partition $E \times E$ into many products of this type. We actually don't need to compute all of the numbers (1.4), we just need the maximum. (We can look at $\frac{1}{|\overline{x}-\overline{y}|}|f(\overline{x}) - f(\overline{y})|$.)

9-20

1.3 Proof of Whitney's extension theorem

We work in $C^m(\mathbb{R}^n)$. We have compact $E \subset \mathbb{R}^n$. For each $x \in E$, we are given $P^x \in \mathcal{P}$, polynomials of degree $\le m$. We would like to know whether there exists a function in $C^m(\mathbb{R}^n)$ with these prescribed Taylor polynomials.

We sketch the interesting parts of the proof of Whitney's extension theorem.

- There is some geometry, an argument with a fundamental idea in analysis.
- Construct a partition of unity.
- Construct F and check that it works.

Proof. Step 1: Geometry

Take a enormous cube which contains E, say the middle $\frac{1}{1}0$ contains E. We are not happy with this cube, so we bisect it in each dimension, to get 2^n cubes. Look at each of the pieces and ask, are we happy? For each cube we are unhappy with, bisect again, and repeat.

What does it mean to be happy? Whitney's rule is simple. Given a cube Q, consider Q^*, the cube with the same center but 5 times the side length. We are happy with Q if Q^* is disjoint from E. This generates a decomposition of the big cubes minus E into infinitely many subcubes.

Every point not in E is in one of the Whitney cubes (if your point belongs to E, you will never be happy); every point in E is not contained in a Whitney cube (there is a small enough neighborhood of it not intersecting E. Any 2 Whitney cubes are disjoint.

Each cube Q is comparable to E. $d(E, Q)$ is greater than the side length δ_Q of Q. Let Q^+ be the parent. Q^+ dilated by 5 intersects E.

In summary, $Q^\circ \backslash E$ is partitioned into Whitney cubes, and

- $Q \in$Wh implies $d(Q, E) \sim \delta_Q$. ($\delta_Q \le d(Q,E) \le 10\delta_Q$.)
- Good geometry: Neighboring cubes are about the same size. $Q^{\text{closure}} \cap Q'^{\text{closure}} \ne \phi$ then $\frac{1}{2} \le \frac{\delta_Q}{\delta_{Q'}} \le 2$.

  (If Q were $< \frac{1}{2}$ the size, then its parent would be contained in Q' and hence also be good, and wouldn't have been cut.)

Idea of making a decomposition was used by Calderon, Zygmund, 1954, on some function in L^1. You are happy if the average over the cube is $> \alpha$. The fact that you got the cube from cutting up something you were not happy with, gives you a lot of control.

Step 2: Partition of unity

Let $\varphi^0 = 1$ on Q^0 (with side length 1), $\varphi^0 = 0$ outside $1.01Q^0$, with φ^0 smooth and $0 \le \varphi^0 \le 1$ everywhere. Suppose $|\partial^\alpha \varphi_Q(x)| \le C\delta_Q^{-|\alpha|}$ for $|\alpha| \le m+5$.

We take the picture and translate it and dilate it.

For Q with side length δ_Q and center Q, define $\varphi_Q(x) = \varphi^0\left(\frac{x-x_Q}{\delta_Q}\right)$.

These functions don't sum to 1, so we define

$$\theta_Q(x) = \frac{\varphi_Q(x)}{\sum_{Q'} \varphi_{Q'}(x)} \tag{1.5}$$

for $x \in Q^{\text{ closure}} \backslash E$.

Only Q' that abuts $\hat{Q}$ has a chance of entering into the computation for $\theta_{\hat{Q}}(x)$; they are about the same size as $\hat{Q}$; the others are shielded. There are a bounded number of cubes that enter in the sum, so

$$1 \le \sum_{Q'} \varphi_{Q'}(x) \le C \tag{1.6}$$

$$|\partial^\alpha \sum_{Q'} \varphi_{Q'}(x)| \le C\delta_{\hat{Q}}^{-|\alpha|} \tag{1.7}$$

$$\left|\partial^\alpha \left(\frac{1}{\sum_{Q'} \varphi_{Q'}(x)}\right)\right| \le C\delta_{\hat{Q}}^{-|\alpha|}. \tag{1.8}$$

To see this, induct by how many times we differentiate. Differentiating gives something of the form

$$\frac{\prod_{i=1}^{s-1} \partial^{\alpha_j}(\sum_{Q'} \varphi_{Q'})}{(\sum_{Q'} \varphi_{Q'})^s}, \tag{1.9}$$

where $\alpha_1 + \cdots + \alpha_S = \alpha$. This has to be dimensionally correct.

The partition of unity reflects the geometry of the cubes and satisfies

- $\theta_Q \ge 0$, $\mathrm{Supp}(\theta_Q) \subset (1.01)Q$.
- $$|\partial^\alpha \theta_Q(x)| \le C\delta_Q^{-|\alpha|} \tag{1.10}$$

  for $|\alpha| \le m+5$.
- $\sum \theta_Q(x) = \begin{cases} 1, & x \in E \\ 0, & x \notin E. \end{cases}$

Step 3: Construct the function.

Let's look at one of the Whitney cubes. Make a guess to what function should look like on this cube. P^{x_Q} is the best guess to how f should behave on Q. But if you have different cubes, there will be jumps at boundaries where they meet. Instead of using the sharp indicator function, we want to use the partition of unity to patch together the functions.

Let

$$F(x) = \sum_{Q\in\mathrm{Wh}} \theta_Q(x)P^{x_Q}(x), \quad x \notin E \tag{1.11}$$

$$F(x) = (P^x)(x) \text{ if } x \in E. \tag{1.12}$$

Note F depends linearly on the data. Given x, there are only a bounded number of Q that determine what F is doing. We call a formula like this "bounded depth".

F is supposed to be a C^m function. The Taylor polynomial is equal to P^x at the point x.

We hope that for $|\alpha| \le m$,

$$\partial^\alpha F(x) = \begin{cases} \sum_{Q\in\mathrm{Wh}} \partial^\alpha(\theta_Q(x)P^{x_Q}(x)) \text{ if } x \notin E \\ (\partial^\alpha P^x)(x) \text{ if } x \in E. \end{cases} \tag{1.13}$$

We verify that it satisfies the definition of C^m function; these quantities are the derivatives of F.

What happens on points close to E? We worry that the derivatives might blow up as we approach E.

$$\partial^\alpha F(x) = \sum_{\beta+\gamma=\alpha} \mathrm{coeff}(\beta,\gamma)\underbrace{\partial^\beta\theta_Q(x)}_{C\delta_Q^{-|\beta|}}\partial^\gamma P^{x_Q}(x) \tag{1.14}$$

If δ_Q is small, we are in bad shape. There is a very clever trick, Whitney 1934 which gets around this. There's a reason that gets into trouble, we haven't used the hypotheses!

For $x \in \hat{Q}$, let

$$F = \sum \theta_Q(P^{x_Q} - P^{x_{\hat{Q}}}) + P^{x_{\hat{Q}}} \tag{1.15}$$

$$\partial^\alpha F(x) = \sum_{\beta+\gamma=\alpha} \mathrm{coeff}(\beta,\gamma)\partial^\beta\theta_Q(x)\cdot\partial^\gamma(P^{x_\alpha} - P^{x_{\hat{Q}}})(x) + (\partial^\alpha P^{x_{\widehat{Q}}}(x)). \tag{1.16}$$

By assumption $\partial^\alpha P^{x_{\widehat{Q}}}(x)$ is bounded. Again $|\partial^\beta\theta_Q(x)| \le C\delta_Q^{-|\beta|}$ might be large. But the difference $\partial^\gamma(P^{x_\alpha} - P^{x_{\hat{Q}}})(x)$ is small.

We have $d(x_Q, x_{\hat{Q}}) \le C\delta_Q$, $\delta_{\widehat{Q}} \sim \delta_Q$. One hypothesis is that

$$|\partial^\gamma(P^x - P^y)(x)| \le C|x-y|^{m-|\gamma|}. \tag{1.17}$$

We use this useful fact about polynomials: If $|\partial^\alpha P(x_0)| \le A\delta^{-|\alpha|}$ for $|\alpha| \le \deg p$, and $|x_0 - y_0| \le C\delta$, then

$$|\partial^\alpha P(y_0)| \le C'A\delta^{-|\alpha|} \tag{1.18}$$

for $|\alpha| \le \deg p$.

We use this to move the basepoint from $x_{\hat{Q}}$ to x_Q.

$$|\partial^\gamma (P^{x_Q} - P^{x_{\widehat{Q}}})(x)| \le C\delta_Q^{m-|\gamma|} \tag{1.19}$$

$$|\partial^\beta \theta_Q(x)| \le C\delta_Q^{-|\beta|} \tag{1.20}$$

□

1.4 Using the well-separated pairs decomposition to compute the Lipschitz constant

We want to compute $\|f\|_{\text{Lip}} = \max_{x,y\in E, x\ne y} \frac{|f(x)-f(y)|}{|x-y|}$ efficiently, to within a constant (e.g. 1.01) factor.

$E \times E \backslash \text{Diag}$ can be partitioned into $E'_\nu \times E''_\nu$ for $\nu = 1, \ldots, \nu_{\max}$ with the following good properties.

- $\nu_{\max} \le CN$.
- $d(E'_\nu, E''_\nu) > 10^5[\operatorname{diam} E'_\nu + \operatorname{diam} E''_\nu]$
- The decomposition can be computed in $O(N\log N)$ steps. (To compute $E'_\nu \times E''_\nu$ we exhibit one point $(x'_\nu, x''_\nu) \in \mathbb{E}'_\nu \times \mathbb{E}''_\nu$.)

Assume this is true; I show how to compute the Lipschitz constant. Then I show the mathematical part by showing the first two bullet points. The punch line of the math discussion is that it's true thanks to the decomposition of a set into Whitney cubes. Define

$$|||f||| = \max_{\nu=1,\ldots,\nu_{\max}} \frac{f(x'_\nu) - f(x''_\nu)}{|x'_\nu - x''_\nu|}. \tag{1.21}$$

We claim

$$|||f||| \le \|f\|_{\text{Lip}} \le 1.001|||f||| \tag{1.22}$$

The left inequality is clear.

Assume $|f(x'_\nu) - f(x''_\nu)| \le |x'_\nu - x''_\nu|$ for each ν. We must prove

$$|f(x') - f(x'')| \le (1.01)|x' - x''|. \tag{1.23}$$

for all x', x'' distinct. Suppose not. Pick x', x'' with $|x' - x''|$ as small as possible.

$$|f(x') - f(x'')| > 1.01|x' - x''| \tag{1.24}$$

for $(x', x'') \in E'_\nu \times E''_\nu$ for some ν. Fix that ν. $(x'_\nu, x''_\nu) \in E'_\nu \times E''_\nu$. Then $x', x'_\nu \in E'_\nu$ implies $|x' - x'_\nu| \le \operatorname{diam} E'_\nu$ and $x'', x''_\nu \in E''_\nu$ implies $|x'' - x''_\nu| \le \operatorname{diam} E''_\nu$, and

$$|x' - x'_\nu| + |x'' - x''_\nu| \le 10^{-5}|x' - x''|. \tag{1.25}$$

Then

$$|f(x') - f(x'')| \le \underbrace{|f(x') - f(x'_\nu)|}_{\le 1.01|x'-x'_\nu|} + \underbrace{|f(x'_\nu) - f(x''_\nu)|}_{\le |x'_\nu - x''_\nu|} + \underbrace{|f(x''_\nu) - f(x'')|}_{\le 1.01|x''-x''_\nu|} \tag{1.26}$$
$$\le 2.01[|x' - x'_\nu| + |x'' - x''_\nu|] + |x' - x''| \tag{1.27}$$
$$\le (1 + 2.01 \cdot 10^{-5})|x' - x''|. \tag{1.28}$$

9-25

1.5 Well-separated pairs decomposition, proof

The idea is simple. Suppose that $E \subseteq Q_\nu \subset \mathbb{R}^n$. Let's look at $Q^\circ \times Q^\circ$. Make a Whitney decomposition of the complement of the diagonal. Each is comparable to its distance from the diagonal. We make these cubes $Q_\nu = Q'_\nu \times Q''_\nu$ such that

$$d(Q_\nu, \text{Diag}) \sim 10^5 \delta_{Q_\nu}. \tag{1.29}$$

I take $E'_\nu \times E''_\nu = E \times E \cap Q'_\nu \times Q''_\nu$, by setting

$$E'_\nu = E \cap Q'_\nu \tag{1.30}$$
$$E''_\nu = E \cap Q''_\nu \tag{1.31}$$

Then

$$d(E'_\nu, E''_\nu) \ge d(Q'_\nu, Q''_\nu) \tag{1.32}$$
$$\sim 10^5(\operatorname{diam} Q'_\nu + \operatorname{diam} Q''_\nu) \tag{1.33}$$
$$\ge 10^5(\operatorname{diam} E'_\nu + \operatorname{diam} E''_\nu) \tag{1.34}$$

There's a slight embarrassment that we need $\nu_{\max} \le CN$, whereas the number of Whitney cubes are infinitely. Fortunately, most of these E'_ν, E''_ν are empty.

Lemma 1.1.3. *The number of nonempty $E'_\nu \times E''_\nu$ is at most CN.*

Proof. Suppose Q is a dyadic cube with Q'_ν and Q''_ν are contained in Q ($Q'_\nu, Q''_\nu \subset Q$) and

$$\delta_Q < 2^{20}\delta_{Q'_\nu} + 2^{20}\delta_{Q''_\nu} \tag{1.35}$$
$$E \cap Q'_\nu, E \cap Q''_\nu \ne \phi. \tag{1.36}$$

Then we say that Q accounts for Q'_ν and Q''_ν.

Unfortunately, it could be that no cube accounts for the pair. Consider 2 tiny intervals, one slightly $< \frac{1}{2}$ and the other slightly $> \frac{1}{2}$, so that the distance between them is 10^6 times the side length. The smallest dyadic interval that contains them is the whole unit interval

This seems an unusual situation; in a typical case it will not happen.

We first count the pairs that are accounted by something, and then fix it up.

$$\sum \# \{Q'_\nu \times Q''_\nu : \text{some dyadic } Q \text{ accounts for } Q'_\nu, Q''_\nu\} \le CN. \tag{1.37}$$

I'll look at all dyadic Q that contain points of E. Under inclusion, such cubes form a tree T; stop cutting when Q contains a single point of E.

If there is a cube Q that accounts for Q'_ν, Q''_ν, then after a bounded number of cuts, there is a branch. (By definition, Q is at most a constant times larger than Q'_ν, Q''_ν.) The number of dyadic cubes that account for something is $\le$ a constant times the number of nodes in the graph where the graph branches.

We need to estimate the number of branch points in the tree. Elementary fact about tree: the number of branch points is the number of leaves minus 1. To see this, induct on the size of the graph.

The number of leaves is the number of points of E, so the number of branch points is $\le \#E - 1$.

Consider pairs Q that account for something, and branch points ≤ 11 levels below it. For every Q that accounts for something, there is a branch point ≤ 11 levels below it. For each branch point there are at most 11 dyadic Q above it. Hence the number of dyadic Q that account for some $Q'_\nu \times Q''_\nu$ is less than $11\#E$.

It can happen there are pairs not accounted for. Let $\mathcal{D}_0$ be the set of all dyadic cubes. For $\xi \in \mathbb{R}^n$, let D_ξ be all $Q + \xi$, $Q \in \mathcal{D}_0$. For a cube not accounted for in $\mathcal{D}_0$, consider it in $\mathcal{D}_\xi$. We can talk about whether Q accounts for something with respect to D_ξ. For any fixed ξ, the number of $Q' \times Q''$ accounted for by D_ξ is $\le C\#E$.

Picking ξ at random, what is the probability that two fixed cubes Q'_ν, Q''_ν lie on different squares of the 2^{11}-times-larger grid and hence aren't accounted for? Unlikely. We have Q'_ν, Q''_ν is accounted for by some $Q = \mathcal{D}_\xi$ with probability $> \frac{1}{2}$.

We estimate the number of pairs $(Q'_\nu \times Q''_\nu, \xi)$ such that Q accounts for $Q'_\nu \times Q''_\nu$ in $\mathcal{D}_\xi$ in 2 ways.

$$\mathbb{E}_\xi(\cdot) = \sum_{Q'_\nu \times Q''_\nu} \mathbb{P}(Q'_\nu \times Q''_\nu \text{ is accounted for by some } Q \in \mathcal{D}_\xi) \ge \frac{1}{2}\#\{Q'_\nu \times Q''_\nu\}. \tag{1.38}$$

$$\mathbb{E}_\xi(\cdot) \le C\#E. \tag{1.39}$$

□

Note in each of the Cartesian products $E'_\nu \times E''_\nu$, each of E'_ν, E''_ν is the intersection of E with a cube.

One convenient way to write down the well-separated pairs decomposition theorem is to write down the cube.

How to compute it efficiently? It can be; the number of steps is $O(N \log N)$. We write down all the relevant cubes; for each $Q'_\nu \times Q''_\nu$, we exhibit one particular point.

Let $E \subset \mathbb{R}^n$. For all $x \in E$ given $P^x \in \mathcal{P}$, does there exist $F \in C^m$ such that $J_x F = P^x$ for all $x \in E$? If so, how small can we take the C^m norm of F?

First do one-time work, $O(N \log N)$. Then we can answer queries.

Let $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$, $\#E = N$. We want $F \in C^m(\mathbb{R}^n)$ such that $F = f$ on E, with norm of F'' as small as possible.

We look for $[(P^x)_{x\in E}, M]$, satisfying the following constraints with M as small as possible.

$$(P^x)(x) = f(x) \qquad \forall x \in E \tag{1.40}$$

$$|\partial^\alpha (P^x)(x)| \le M \qquad \forall x \in E, |\alpha| \le m \tag{1.41}$$

$$|\partial^\alpha (P^x - P^y)(x)| \le M|x-y|^{m-|\alpha|} \qquad \forall x, y \in E \text{ distinct}, |\alpha| \le m. \tag{1.42}$$

For finite E, the $o(\cdot)$ condition is vacuous.

This problem can be reduced to linear programming. This is a big LP: there are N constraints, N constraints, and N^2 constraints in the three sets. The well-separated pairs decomposition reduces this to $O(N)$ constraints.

In the WSPD, we get $E'_\nu \times E''_\nu$, $\nu = 1, \ldots, \nu_{\max}$, $\nu_{\max} \le CN$. Pick $(x'_\nu, x''_\nu) \in E'_\nu \times E''_\nu$ for each ν. We can replace the third constraints by

$$|\partial^\alpha (P^{x'_\nu} - P^{x''_\nu})(x'_\nu)| \le M|x'_\nu - x''_\nu|^{m-|\alpha|} \qquad \forall x'_\nu, x''_\nu, 1 \le \nu', \nu'' \le \nu_{\max} \text{ distinct}, |\alpha| \le m. \tag{1.43}$$

This is similar to the proof that when you estimate the Lipschitz constant, you can just estimate it over the representatives. We have

$$\operatorname{diam} E'_\nu + \operatorname{diam} E''_\nu \le ad(E'_\nu, E''_\nu), \tag{1.44}$$

and P^x for each $x \in E$. Given (1.43), we will prove that

$$|\partial^\alpha (P^{x'} - P^{x''})(x')| \le 1.01|x' - x''|^{m-|\alpha|} \tag{1.45}$$

for any $x', x'' \in E$, $|\alpha| < m$. Suppose not. Pick a counterexample (x', x'', α_0) with $|x' - x''|$ as small as possible.

$$(x', x'') \in E \times E \backslash \mathrm{Diag} \tag{1.46}$$

$$(x', x''), (x'_\nu, x''_\nu) \in E'_\nu \times E''_\nu \tag{1.47}$$

By minimality, $|x' - x'_\nu|, |x'' - x''_\nu| \le a|x' - x''|$.

$$|\partial^\alpha (P^{x'} - P^{x'_\nu})(x'_\nu)| \le |x' - x'_\nu|^{m-|\alpha|} \le a|x'_\nu - x''_\nu|^{m-|\alpha|} \tag{1.48}$$

$$|\partial^\alpha (P^{x''} - P^{x''_\nu})(x''_\nu)| \le |x'' - x''_\nu|^{m-|\alpha|} \le \alpha|x'_\nu - x''_\nu|^{m-|\alpha|} \tag{1.49}$$

$$|\partial^\alpha (P^{x'_\nu} - P^{x''_\nu})(x'_\nu)| \le |x'_\nu - x''_\nu|^{m-|\alpha|} \tag{1.50}$$

We can move the point at which we evaluate the polynomial. We get

$$|\partial^\alpha (P^{x''} - P^{x''_\nu})(x'_\nu)| \le Ca|x'_\nu - x''_\nu|^{m-|\alpha|} \tag{1.51}$$

$$|\partial^\alpha (P^{x'} - P^{x''})(x'_\nu)| \le (L + Ca)|x'_\nu - x''_\nu|^{m-|\alpha|} \tag{1.52}$$

for $|\alpha| \le m$. (We actually need a stronger version of "moving the basepoint".)

The number of steps to solve a LP of size n is poly(n), like n^3. But we will get it down to $n \log n$.

10-2

2 Γ and σ

In the problem we're interested in, we're only given the values, not the Taylor polynomials at the points. It's a linear programming problem; in principle we can do it, but in time N^3. We'll get the time down to $N \log N$. We have to come up with a consistent set of Taylor polynomials. What might it be at one point?

For $m, n > 1$, we are given $E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}$. We want to extend to a function $C^m(\mathbb{R}^n)$. We ask the right question: if we fix one particular point, what might the Taylor polynomial be at that one point? Let

$$\Gamma(x, M) = \{J_x : F = f \text{ on } E, \|F\|_{C^m} \le M\} \subset \mathcal{P}. \tag{1.53}$$

This is a possibly empty convex set. What is the approximate size and shape of $\Gamma(x, M)$?

It would be great if we can find the exact possible C^m norm, but we don't know how; also there are lots of equivalent C^m norms. Think of the inequality up to a constant.

$$\Gamma(x, c_1 M) \subset \Gamma_{\text{computed}}(x, M) \subset \Gamma(x, C_1 M), \tag{1.54}$$
$$\Gamma_{\text{computed}}(x, c_2 M) \subset \Gamma(x, M) \subset \Gamma_{\text{computed}}(x, C_2 M). \tag{1.55}$$

These sets have a lot of information. They're useful and even if we couldn't use them to compute interpolants we would still want to use them. Suppose that I want to interpolate data. They come from some experiment or observations. The f we're trying to find is presumably a smooth function of the data, but it's highly unlikely that the function is the interpolant with smallest norm. What do we believe? We believe it matches the data and the C^m norm is not that large.

Pick a point, what can we say? Suppose we observe where we are at discrete points in time. Pick some particular time, what can I say about position, velocity, acceleration? Given what you believe about your interpolant, we would like to know what kind of uncertainty there is in the interpolant.

If we can decide whether Γ is empty, that tells us approximately the best possible norm of the interpolant.

When we prove theorems and construct interpolants, these Γ's are what we use to construct.

First we look at $\Gamma(x, M)$ at one particular point $x \in E$. We guess Γ_{computed}, and prove correctness by constructing interpolants. We have to not find the jet at one point, but a family of jets that are mutually consistent.

For $x \in E$ and M fixed, we construct

$$\Gamma_\ell(x, M) \supset \Gamma(x, M) \tag{1.56}$$

for $\ell > 0$, convex, possibly empty. Define by induction on ℓ. Let different points in E talk to each other and reduce the size to $\Gamma_{\ell+1}(x, M)$.

We induct on ℓ. Let

$$\Gamma_0(x, M) := \{P \in \mathcal{P} : P(x) = f(x), |\partial^\alpha P(x)| \le M \text{ for } |\alpha| \le M\}. \tag{1.57}$$

This is all the info that comes if you ignore all other points except x.

Suppose we know $\Gamma_\ell(x, M)$ for all $x \in E$. Suppose $\Gamma_\ell(x, M) \supset \Gamma(x, M)$. We will define $\Gamma_{\ell+1}(x, M)$ for all $x \in E$ such that $\Gamma_\ell(x, M) \supset \Gamma_{\ell+1}(x, M) \supset \Gamma(x, M)$.

Let $x \in E$, $P \in \Gamma_\ell(x, M)$. Then let $P \in \Gamma_{\ell+1}(x, M)$ iff for all $y \in E \backslash \{x\}$ there exists $P' \in \Gamma_\ell(y, M)$ such that

$$|\partial^\alpha (P - P')(x)| \le M|x - y|^{m-|\alpha|}. \tag{1.58}$$

How complicated are these sets? They are much too complicated, but we will go ahead and prove math theorems about them. We will not define the Γ_ℓ this way, but differently to retain the key properties.

The Γ_ℓ's are convex polytopes defined by linear constraints. But Γ_ℓ are polytopes defined by growing number of linear constraints. To get to the next steps, ontersect polytopes. The number of constraints defining them will grow very fast. How to cope? Compute them approximately.

Define a blob, a 1-parameter family of growing convex sets. We say two blobs are C-equivalent if they are the same up to a constant C.

Before the Γ's are getting rapidly more complicated, but we can arrange things so they aren't.

The more serious point: every y is talking to every x. It appears this definition requires N^2 steps. Fortunately, there is a clever way to use the well-separated pairs decomposition to compute something enough like them in time $N \log N$.

Theorem 1.2.1. *$\Gamma(x, M) \subset \Gamma_\ell(x, M)$, and for $\ell_* = \ell_*(m, n)$, we have $\Gamma_{\ell_*}(x, M) \subset \Gamma(x, CM)$ for C depending only on m, n.*

[1]

The number of times you iterate to get something comparable to the true Γ is a fixed number.

Facts:

- $\Gamma_\ell(x, M) \subset \mathcal{P}$ is a possibly empty convex set.
- $\Gamma_\ell(x, M) \subset \Gamma_\ell(x, M')$ if $M \le M'$, by induction (conditions grow weaker as M grows).
- Given $x, y \in E$, given $P \in \Gamma_\ell(x, M)$, there exists

$$P' \in \Gamma_{\ell-1}(y, M) \qquad \text{such that } |\partial^\alpha (P - P')(x)| \le M(x, y)^{m-|\alpha|}. \tag{1.59}$$

We talk only about Γ's and their cousin the σ's.

We want $F = C^m(\mathbb{R}^n)$ such that $\|F\|_{C^m} \le CM$ and $J_x(F) \in \Gamma_0(x, M)$ for all $x \in E$.

There are two ways to construct Γ_ℓ's. We show another. Then we introduce the related sets σ.

We use the finiteness and refined finiteness theorems.

[1] So far E has been finite. The analogue is not true for C^m of infinite sets. One needs something more. That is the Glazer refinement. This is what Glazer refinement looks like for finite sets.

Theorem 1.2.2 (Finiteness Theorem)**.** *Let $E \subset \mathbb{R}^n$, $\#E = N < \infty$, $f : E \to \mathbb{R}$. Given $M > 0$, suppose that for every subset $S \subset E$ with at most $k^\#$ points (depending only on m, n), there exists $F^S \in C^m(\mathbb{R}^n)$ of norm $\le M$ such that $F^S = f$ on S. Then there exists $F \in C^m(\mathbb{R}^n)$ of norm $\le CM$ (C depending only on m, n) such that $F = f$ on E.*

To use this we would need to look at $\sim N^{k^\#}$ sets. We need a refinement.

Theorem 1.2.3 (Refined finiteness theorem)**.** *Fix m, n. Let $E \subset \mathbb{R}^n$, $\#E = N$. Then there exist $S_1, \ldots, S_L \subset E$ with the following properties*

- *$\#(S_\ell) \le k^\#$ for each ℓ,*
- *$L \le CN$,*
- *Let $f : E \to \mathbb{R}$ and let $M > 0$. Suppose that for each $\ell = 1, \ldots, L$ there exists $F_\ell \in C^m(\mathbb{R}^n)$ with norm $\le M$ such that $F_\ell = f$ on S_ℓ. Then there exists $F \in C^m(\mathbb{R}^n)$ of norm $\le CM$ such that $F = f$ on E.*
- *The $S_1, \ldots, S_L$ can be computed from E in $O(N \log N)$ computer operations.*

where $k^\#$ and C depend only on m, n

Digression on outliers: Suppose we collect data from a physics experiment, and the machine malfunctions or the technician falls asleep. We get data points that are completely wrong and should be discarded. You discover that the smallest possible norm is enormous, and would like to discard some data. If you are allowed to discard a few to bring the interpolant way down, which should you ignore? As a consequence of the refined finiteness theorem, there is a theorem with an algorithm attached.

Theorem 1.2.4. *Fix m, n. Given $f : E \to \mathbb{R}$, $\#E = N$, there exists an enumeration of E, $E = \{x_1, \ldots, x_N\}$ such that the following holds.*

Let $S \subset E$ and suppose there exists an interpolant of norm $\le M$ for $f|_{E \setminus S}$, with $\#S = Z$. Then $f|_{E\setminus\{x_1,\ldots,x_{CZ}\}}$ has an interpolant with norm $\le CM$.

The points $x_1, \ldots, x_N$ can be computed using $\le CN \log N$ operations per point.

Suppose we have a contest against God to remove outliers. God instantaneously know the very best set of Z outliers to remove. We are at a disadvantage. Let's cheat to give us a chance to prevail anyway. We can throw away 50 times as many points, and declare victory if our norm is within 50 times. Then we can win anyway.

It's remarkable this can be done at all; I don't think this is optimal.

10-4

We will cover

- σ's
- connection of finiteness theorem to Γ's

- Infinite E's (new ingredients)
- The key properties of Γ's and σ's.
- Outliers

Let $f : E \to \mathbb{R}$, $E \subset \mathbb{R}^n$ finite, $m \geq 1$ fixed. Let $\Gamma(x, M) = \{J_x(F) : F \in C^m$ with norm $\leq M, F = f$ or
For $x \in E$, we defined $\Gamma_\ell(x, M)$ by

$$\Gamma_0(x,M) := \{P \in \mathcal{P} : P(x) = f(x), |\partial^\alpha P(x)| \leq M \text{ for } |\alpha| \leq M\} \supset \Gamma(x,M) \tag{1.60}$$

$$\Gamma_{\ell+1}(x,M) := \left\{P \in \Gamma_\ell(x,M) : \forall y \in E\backslash\{x\}, \exists P' \in \Gamma_\ell(y,M), |\partial^\alpha(P-P')(x)| \leq M|x-y|^{m-|\alpha|}, |\alpha| \leq m\right\}. \tag{1.61}$$

If I have 2 interpolants for the same function, then their difference is an interpolant for 0. So the interpolants for 0 tell us how much arbitrariness there is in the interpolants.

WLOG consider $M = 1$. Let

$$\sigma(x) = \{J_x(F) : F \in \mathcal{C}^m, \text{ norm } \leq 1, F = 0 \text{ on } E\} \tag{1.62}$$

$$\sigma_0(x) = \{P : \forall |\alpha| \leq m, |\partial^\alpha P(x)| \leq 1, P(x) = 0\} \tag{1.63}$$

$$\sigma_{\ell+1}(x) = \left\{P \in \sigma_\ell(x) : \forall y \in E\backslash\{x\}, \exists P' \in \sigma_\ell(y) \text{ such that } \forall |\alpha| \leq m, |\partial^\alpha(P-P')(x)| \leq |x-y|^{m-|\alpha|}\right\} \tag{1.64}$$

For $P, P' \in \Gamma(x,M)$, $P - P' \in 2M \cdot \sigma(x)$. For $P \in \Gamma(x,M)$, $Q \in 2M\sigma(x)$, $P+Q \in \Gamma(x,3M)$. If $\Gamma(x, M_0) \ni P_0$ is nonempty, then for all $M \geq 5M_0$,

$$P_0 + M\sigma(x) \subset \Gamma(x,M) \subset P_0 + 10M\sigma(x) \tag{1.65}$$

To understand the Γ's, we need to find the magnitude of the smallest M_0 such that $\Gamma(x, M_0)$ is nonempty, and produce one element.

For $P, Q \in \mathcal{P}$, the pointwise product $P \odot_x Q = J_x(PQ)$, and $J_x(FG) = J_x(F) \odot_x J_x(G)$. For $P(x) = \sum_{|\alpha| \leq m} a_\alpha x^\alpha$ and $Q(x) = \sum_{|\alpha| \leq m} a_\beta x^\beta$, $PQ(x) = \sum_{|\alpha|,|\beta| \leq m} a_\alpha b_\beta x^{\alpha+\beta}$, $P \odot_0 Q = \sum_{|\alpha|+|\beta| \leq m} a_\alpha b_\beta x^{\alpha+\beta}$.

For $\|F\|_{C^m} \leq 1$, $F = 0$ on E, $J_x(F) = P$, Apply this with the Whitney extension theorem. Interpolants work locally on different scales, and patch them together. The cutoff functions will behave quite badly. Consider $x = 0$. Let $P = J_0(F)$. Suppose $|\partial^\alpha(0)| \leq \delta^{m-|\alpha|}$, $|\partial^\alpha F(0)| \leq \delta^{m-|\alpha|}$, $|\partial^\alpha Q(0)| \leq \delta^{-|\alpha|}$, $|\partial^\alpha F(0)| \leq \delta^{m-|\alpha|}$ for $\delta \leq 1$. If $P \in \sigma(0)$ then $Q \odot_0 P \in C\sigma(0)$, for C depending only on m, n.

There exists $\theta \in C_0^\infty(\mathbb{R}^n)$ such that $\operatorname{Supp}\theta \in B(0,\delta)$, $J_0(\theta) = Q$, $|\partial^\alpha \theta| \leq C\delta^{-|\alpha|}$ everywhere, for $|\alpha| \leq m$.

What does this mean when $\delta = 1$? All the coefficients are bounded. There exists some C^∞ function supported on the unit ball whose Taylor polynomial at 0 is Q and whose derivatives are bounded. This is clear by multiplying by a cutoff function. For $\delta \neq 1$, this follows from $\delta = 1$ by rescaling.

Look at $F \cdot \theta$. $\partial^\alpha(F \cdot \theta)$ is a sum of terms

$$|\partial^\beta F(x)\partial^\gamma\theta(x)| \le C\delta^{-|\gamma|} \tag{1.66}$$

where $\beta + \gamma = \alpha$. If $|\partial^\beta F(0)| \le \delta^{m-|\beta|}$ for all $|\beta| \le m$ and $|\partial^{\widehat{\beta}} F(x)| \le 1$ for $|\widehat{\beta}| = m$, then $|\partial^\beta F(x)| \le C\delta^{m-|\beta|}$ for $|\beta| \le m$.

(To check effortlessly something is scale-invariant, assign units.)

We've proven that if $P \in \sigma(0)$, then $Q \odot_0 P \in C\sigma(0)$.

Definition 1.2.5: Let $\sigma \subset \mathcal{P}$ be any convex symmetric set. Let $x \in \mathbb{R}^n$, $C_w > 0$, $\delta_{\max} > 0$.

Then σ is a **Whitney convex** at x, with Whitney constant C_w below length scale $\delta_{\max}$, iff for all $P, Q \in \mathcal{P}$, $0 < \delta \le \delta_{\max}$, if $|\partial^\alpha P(x)| \le \delta^{m-|\alpha|}$ and $|\partial^\alpha Q(x)| \le \delta^{-|\alpha|}$ for all $|\alpha| \le m$, and if also $P \in \sigma$, then $P \odot_x Q \in C_w\sigma$.

If I don't specify the lengths scale, it is 1.

Remark 1.2.6: Given $E \subset \mathbb{R}^n$, construct $\sigma(x)$ from interpolants. Then for all $x \in \mathbb{R}^n$, $\sigma(x)$ is Whitney convex at x, with Whitney constant defined by m, n below length scale $\delta_{\max} = 1$.

The definition of the C^m norm is somewhat unnatural—taking different dimensional quantities (different order derivatives) and taking the max. This unnaturalness is reflected in the length scale. We can also define $\dot{C}^m$ norm which takes a sup over derivatives with order precisely m.

Check by induction on ℓ that $\sigma_\ell(x)$ is also Whitney convex that depends on ℓ, m, n below length scale 1.

Let $x \in E$, $M > 0$, $\ell \ge 0$. Then $\Gamma_\ell(x, M)$ is a (possibly empty) convex subset of $\mathcal{P}$, $\sigma_\ell(x)$ is a convex symmetric set in $\mathcal{P}$. For $M \ll M'$, with a small constant determined by m, n, ℓ, $\Gamma_\ell(x, M) \subset \Gamma_d(x, M')$, $\Gamma_{\ell+1}(x, M) \subset \Gamma_\ell(x, M)$.

If $x, y \in E$, $P \in \Gamma_\ell(x, M)$, $\ell \ge 1$, then there exists $P' \in \Gamma_{\ell-1}(y, CM)$, C, depending only on m, n, ℓ , such that $|\partial^\alpha(P - P')(x)| \le CM|x - y|^{m-|\alpha|}$, for $|\alpha| \le m$.

If $P, P' \in \Gamma_\ell(x, M)$, then $P - P' \in CM\sigma_\ell(x)$. If $P \in \Gamma_\ell(x, M)$, $Q \in M\sigma_\ell(x)$ then $P + Q \in \Gamma_\ell(x, CM)$. $\sigma_\ell(x)$ is Whitney convex at x, with Whitney constant C below length scale 1.

We talk about the finiteness theorem.

Define

$$\Gamma(x, M, S) = \{J_x(F) : F = f \text{ on } S, \|F\|_{C^m} \le M\} \tag{1.67}$$

$$\widehat{\Gamma}_\ell(x, M) = \bigcap_{S \subset E, \# S \le k_\ell} \Gamma(x, M, S) \in \mathcal{P} \tag{1.68}$$

convex. We need to prove this is nonempty. We use the following basic theorem in convex geometry.

Theorem 1.2.7 (Helly's Theorem)**.** *Suppose that $K_1, \ldots, K_N$ are convex (and not necessarily compact) subsets of $\mathbb{R}^D$. Suppose that any $D + 1$ of the K_i have a point in common. Then $K_1 \cap \cdots \cap K_N \ne \phi$.*

If there are infinitely many convex compact sets, then this is still true. If you take infinitely many convex sets, not necessarily compact, then this is not true, ex. $(0, \frac{1}{n}]$.

Proof. Given $L \geq D+1$, we show that if any L of the K's intersect, then also $L+1$ of the K's intersect.

Consider $K_1, \ldots, K_{L+1}$. For each $i = 1, \ldots, L+1$, pick x_i in the intersection of all these K's except K_i. We obtain $x_1, \ldots, x_{L+1}$.

Look for coefficients $\beta_1, \ldots, \beta_{L+1} \in \mathbb{R}$ such that

$$\beta_1 + \cdots + \beta_{L+1} = 0 \tag{1.69}$$

$$\beta_1 x_1 + \cdots + \beta_{L+1} x_{L+1} = 0. \tag{1.70}$$

There are $D+1$ equations, and at least $D+2$ unknowns, so there is a nonzero solution. Put the positive ones on the LHS and the negative ones on the RHS: after possibly relabeling,

$$\lambda_1 x_1 + \cdots + \lambda_a x_a = \mu_1 x_{a+1} + \cdots + \mu_b x_{a+b} \tag{1.71}$$

$$\lambda_1 + \cdots + \lambda_a = \mu_1 + \cdots + \mu_b. \tag{1.72}$$

The λ's and μ's are nonnegative and not all 0; we can rescale so that

$$\lambda_1 + \cdots + \lambda_a = \mu_1 + \cdots + \mu_b = 1. \tag{1.73}$$

We claim in $\lambda_1 x_1 + \cdots + \lambda_a x_a = \mu_1 x_{a+1} + \cdots + \mu_b x_{a+b} \in \bigcap_{i=1}^{L+1} K_i$. The LHS lies in $\bigcap_{i=a+1}^{a+b} K_i$ and the RHS lies in $\bigcap_{i=1}^{a} K_i$. This completes the induction step and proves the theorem. $\square$