

**2021-03-25**

## 1 Introduction

Our goal in the coming three weeks largely involves approximating functions  $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  by some “simple” family of functions (particularly polynomials and combinations of translates of a radial basis function). If we want to prove something, we will usually assume  $\Omega$  is compact and  $f$  has some degree of smoothness or regularity.

Approximation of continuous functions  $f : [a, b] \rightarrow \mathbb{R}$  plays a key role in most introductions to analysis, whether classical or numerical. On the classical side, we have the Weierstrass theorem, stating that every such function can be uniformly approximated by a polynomial; equivalently,

$$\forall \epsilon > 0, \exists p \in \mathcal{P} : \|f - p\|_\infty \leq \epsilon,$$

where in this setting  $\|u\|_\infty = \max_{[a,b]} |u(x)|$ . Work by Jackson and Bernstein made Weierstrass’s results more precise, connecting the degree of smoothness to the polynomial degree required to achieve a given accuracy. These types of results are generally referred to as constructive function theory, and are well described in now-classic books by Akhieser and by Rivlin.

In numerical analysis, most students are first exposed to function approximation in discussions of polynomial interpolation: given function values  $y_0, \dots, y_d$  at nodes  $x_0, \dots, x_d$ , find a polynomial  $p \in \mathcal{P}_d$  (i.e. a polynomial with degree at most  $d$ ) such that  $p(x_j) = y_j$  for  $j = 0, \dots, d$ . The study of polynomial interpolation stands on two legs: on the computational side, one needs a representation of the interpolating polynomial that can be computed (and evaluated at new points) quickly and without numerical instability; and on the theoretical side, one needs an understanding of how well polynomial interpolation approximates a target function  $y = f(x)$  when  $f$  has some known smoothness. The theory and computational practice come together in the design of adaptive approximation algorithms; the Chebfun system is a modern instance of tying these threads thoroughly together.

Alas, approximating univariate functions by polynomials is not enough for our purposes, and things immediately become more complicated when we move to higher-dimensional spaces. When we go beyond one space dimension, we can no longer always uniquely define an interpolant from an  $n+1$ -dimensional linear space of functions with samples at  $n+1$  data points;

we need an additional condition in order to ensure the problem is well-posed (the famous Mairhuber-Curtis theorem). We can get around this issue by constraining the location of the sample points or by choosing a method that goes beyond taking approximants from a fixed linear space. At a computational level, high-dimensional approximation suffers from the “curse of dimensionality”: if all we have is a fixed amount of smoothness, the number of data points required to reach a target accuracy tends to grow exponentially in the dimension of the space. Therefore, we tend to seek methods that effectively lower the problem dimension, as we discuss in the coming week.

## 2 A concrete error bound

Some of this lecture will be rather abstract. In order to ground ourselves, let's start with a very concrete problem. Consider  $n + 1$  points  $x_0, \dots, x_n \in \mathbb{R}^n$  with associated function values  $y_j = f(x_j) \in \mathbb{R}$ . Assuming  $f$  has bounded second derivatives, how can we bound  $|f(x) - p(x)|$  where  $p(x) = c^T x + d$  is the linear interpolant through the data points?

Even in this simple case, we need an extra hypothesis to make sense of the problem; the system of equations

$$\begin{bmatrix} x_0^T & 1 \\ x_1^T & 1 \\ \vdots & \\ x_n^T & 1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = y$$

should not be singular. Partial Gaussian elimination gives that this is equivalent to the condition that  $x_1 - x_0, \dots, x_n - x_0$  be a basis for  $\mathbb{R}^n$ .

Assuming that this system is nonsingular, we can evaluate  $p(x)$  at a given target point  $x$  by either solving the linear system for the coefficients  $c$  and  $d$  or solving the system

$$\begin{bmatrix} x_0 & x_1 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{bmatrix} w = \begin{bmatrix} x \\ 1 \end{bmatrix}$$

and then evaluating  $p(x) = w^T y$ . If  $x$  is in the convex hull of the  $x_i$ , then the weights  $w$  are all non-negative.

This latter form is convenient for error analysis, together with the trick of writing each of the function evaluations in terms of a Taylor expansion

about  $x$ ; if we let  $z_j = x_j - x$ , then

$$f(x_j) = f(x) + f'(x)z_j + \frac{1}{2}z_j^T f''(\xi_j)z_j.$$

where  $\xi_j$  is some point on the line segment connecting  $x$  and  $x_j$ . Substituting in, we have

$$p(x) = \sum_j w_j f(x) + \sum_j w_j f'(x)z_j + \frac{1}{2} \sum_j w_j z_j^T f''(\xi_j)z_j;$$

and from the defining equations for  $w$ , we have that

$$\sum_j w_j = 1, \quad \sum_j w_j z_j = 0,$$

so that we are left with

$$p(x) - f(x) = \frac{1}{2} \sum_j w_j z_j^T f''(\xi_j)z_j$$

Therefore, if  $\|f''\| \leq M$  uniformly, we have

$$|p(x) - f(x)| \leq \frac{1}{2} \left( \sum_j |w_j| \right) \left( \max_j M \|z_j\|^2 \right).$$

If  $x$  is in the convex hull of the data locations, we have that the weights are all positive and sum to one, and the distances  $\|z_j\|$  are bounded by the maximum edge length  $d = \max_{k,\ell} \|x_k - x_\ell\|$ , so that

$$|p(x) - f(x)| \leq \frac{1}{2} M d^2$$

where  $d$  is the maximum edge length  $\|x_i - x_j\|$ . In fact, we can tighten the constant somewhat, though not enormously.

There are a few things to take away from this concrete example (apart from the error bound, which is useful in its own right):

- We need a hypothesis on the points at the outset to ensure that we can solve the interpolation system (this property of the points is sometimes called *well-posedness* for interpolation).

- There are multiple ways of formulating the interpolation problem (solve for the coefficients in an affine function, solve for weights associated with a given point).
- The fact that the method exactly reconstructs linear functions played an important (if implicit) role in our error analysis.

Let's now consider how we would tackle something like this in a more abstract setting.

### 3 Approximation from a fixed space

Consider a Banach space  $\mathcal{F}$  (a complete, normed vector space) containing a target function  $f$ , and suppose we seek to approximate  $f$  by some function  $v \in \mathcal{V}$  where  $\mathcal{V} \subset \mathcal{F}$  is a finite-dimensional subspace. We now have two distinct questions:

- Does  $\mathcal{V}$  contain a good approximation to the target function?
- If a good approximation exists, how can it be found?

There are several different ways that we might choose a good approximation, including minimizing a squared error (in the case that  $\mathcal{F}$  is a Hilbert space); including interpolating at a set of points; forcing certain linear functionals of the error to be zero (a *Petrov-Galerkin* approach); minimizing a (weighted) squared error at a larger set of sample points; or minimizing the maximum error over a large set of points (leading to a so-called Chebyshev optimization problem).

For the moment, we consider approximation schemes that are linear in  $f$  and are exact on  $\mathcal{V}$ . This includes all of the schemes mentioned above except the last (Chebyshev optimization leads to a nonlinear scheme). In these methods, we approximate  $f$  by  $Pf$  where  $P$  is a linear operator with range  $\mathcal{V}$  and  $P\mathcal{V} = \mathcal{V}$ . These two conditions imply that the approximation operator  $P$  is a projection, i.e.  $P^2 = P$ . We also have an associated *error projection*  $I - P$ . Now, note that for any  $v \in \mathcal{V}$ , we have

$$\|f - Pf\| = \|(I - P)(f - v)\| \leq \|I - P\| \|f - v\|,$$

and therefore

$$\|f - Pf\| \leq \|I - P\| \inf_{v \in \mathcal{V}} \|f - v\|.$$

Hence, approximation via  $f$  is *quasi-optimal*, yielding an approximation within a factor of  $\|I - P\|$  of the best possible approximation within the space.

Making this a little more concrete, suppose we use the max norm on  $\Omega$ , i.e.

$$\|f\|_\infty = \sup_{x \in \Omega} |f(x)|.$$

Then  $\|I - P\|_\infty \leq 1 + \|P\|_\infty$  and  $\|P\|_\infty$  is the maximum over  $\Omega$  of the so-called *Lebesgue function*

$$\Lambda(x) = \sum_{j=1}^n |v_j(x)|$$

where  $|v_j(x)|$  are the so-called cardinal functions (or Lagrange functions) for which  $v_j(x_i) = \delta_{ij}$ .

## 4 Kolmogorov $n$ -widths and convergence rates

The quasi-optimality framework described above only helps with half of our problem. We want an approximation within a constant factor of the best thing possible; we *also* want the best thing possible to be good! This is an instance of the consistency-stability paradigm common in numerical analysis: stability gives a small quasi-optimality constant, consistency means that the best possible approximation has a small error. Having briefly described the stability piece, let's now talk about accuracy.

Let  $K$  be a (usually compact) subset of  $\mathcal{F}$ . When  $\mathcal{F}$  has a norm (or quasi-norm) that measures smoothness, the set  $K$  may be associated with a unit ball of functions up to a prescribed smoothness. In our concrete case described before, for instance, we might consider the set of all functions where the second derivative was pointwise bounded in norm by some constant  $M$ . The Kolmogorov  $n$ -width describes the *worst-case* approximation error for functions from  $K$  under a *best-case* choice of  $n$ -dimensional approximation spaces of functions; that is,

$$d_n(K) := \inf_{\dim(\mathcal{V}_n)=n} \sup_{f \in K} E(f, \mathcal{V}_n),$$

where  $E(f, \mathcal{V}_n)$  is the minimum error norm for approximating  $f$  by elements of  $\mathcal{V}_n$ . In general, these  $n$ -widths decay as  $O(n^{-\alpha})$  for some  $\alpha$  associated with smoothness of the functions in  $K$ .