From *SIAM News*, Volume 43, Number 9, November 2010

Norm!

By Barry A. Cipra

Coach: Hey, Norm. What do you know?
Norm: Not enough.
—"Cheers" (1982–93)

From climate modeling to materials science, many of the modern applications of mathematics share an analytic challenge that goes back decades, if not centuries: the need to parlay finite data sets into continuous functions with some semblance of smoothness. Atmospheric pressure, for example, needs at least one derivative to fit sensibly into the Navier–Stokes equations. How best to do this—and, in some settings, whether it can be done at all—has been a long-running question. But as Charles Fefferman of Princeton University explained in an invited presentation at this year's SIAM Annual Meeting in Pittsburgh, there is now a firm theoretical basis for best-possible practice in smooth interpolation. There's only one catch: The theory, for now, produces algorithms that are wildly impractical.

The theoretical basis for interpolation starts with a result known as the Whitney extension theorem. First formulated and proved in the early 1930s by Hassler Whitney, the extension theorem gives conditions under which a function defined on a closed subset of $\mathbb{R}^n$ can be extended to all of $\mathbb{R}^n$ with a specified degree of smoothness. The "degree of smoothness" of a function is the number of continuous mixed partial derivatives it has; it can range from 0 to infinity.

Fefferman described recent work in which he and his colleague Bo'az Klartag of Tel Aviv University extended Whitney's extension theorem. In 2003, Fefferman proved a "sharp" version of the result. Fefferman's theorem applies to all subsets of $\mathbb{R}^n$, not just closed subsets. Moreover, whereas the conditions of Whitney's extension theorem amount to assuming that there is a self-consistent family of Taylor polynomials at all the points of the (closed) subset, Fefferman's version puts conditions only on finite subsets of the given (arbitrary) subset. Furthermore, the size of the finite subsets is predetermined only by the dimension n and the desired degree of smoothness m. In effect, Fefferman's theorem is a one-size-fits-all approach to function extension.

The focus on finite sets puts a practical spin on things. In 2005–06, Fefferman and Klartag wrote a pair of papers on fitting C^m smooth functions to data. (C^m is analysts' notation for continuous functions with m continuous derivatives. The C^m norm, roughly speaking, is the largest value of any of the first m mixed partial derivatives anywhere in $\mathbb{R}^n$.) The central problem is to find a function of minimal C^m norm that fits a finite data set, either exactly or within some specified error. Fefferman and Klartag have found two algorithms, one that computes the order of magnitude of the minimal norm and another that produces the corresponding C^m function. (Actually, compactness being problematic in infinite-dimensional Banach spaces, there may not be a norm-minimizing C^m function; the precise statements involve infimums.)

Fefferman and Klartag's algorithms are theoretically efficient: Computing the order of magnitude of the minimal norm requires at most $CN\log N$ work and CN storage, where N is the number of data points and C is a constant depending only on n and m. (In many applications, the parameters n and m are fairly small; it's the amount of data that's alarmingly large these days.) Computing the corresponding function elsewhere in $\mathbb{R}^n$, which amounts to producing its mth-order Taylor polynomial approximation at a given point, similarly takes at most $CN\log N$ "one-time" work and CN storage, plus at most $C\log N$ work to "answer a query." Subtleties arise here as well, concerning the model of computation with real numbers: In their main paper, which is 266 pages long, Fefferman and Klartag devote 87 pages to the theory of computation with finite-precision numbers—pages that Fefferman wryly describes as "the world's most boring appendix."

In a 2006 follow-up paper, Fefferman addressed the problem of finding and discarding outliers. Tossing out a few data points should allow a fit with smaller C^m norm. The question is, if you budget for discarding, say, k points, which ones give the biggest bang for the buck? Offhand, this would seem to require an amount of computation on the order of N^k. But Fefferman has shown how, with a modicum of overkill, a discard set of size Ck (with C again depending only on n and m) can be produced with a mere $CkN\log N$ computation. Fefferman's enlarged discard set is guaranteed to do at least as well as the optimal discard set of size k. Moreover, with at most $CN^2\log N$ computation (and probably much less), the algorithm produces a rough ordering of the data points $x_1, x_2, \ldots, x_N$ from least to most reliable, so that discarding the first k is essentially optimal.

Despite the algorithmic emphasis, Fefferman and Klartag's results are a long way from practical. The various constants C, for example, depend on n and m in the sort of superexponential manner that only a theorist can love; the algorithms satisfy the technical definition of efficiency, but that's about it. If history is any guide, though, the theory should serve as a guidepost for eventual implementations. As Fefferman modestly puts it, "I would dearly love for this actually to be good for something."

Barry A. Cipra is a mathematician and writer based in Northfield, Minnesota.