

# $\ell_1$ -magic

By Barry A. Cipra

*You will surely be a-hurtin',  
If you throw it all away.*—Bob Dylan

Data compression is one of the mainstays of imaging science. Standards like jpeg and mpeg make possible the efficient storage and transmission of images ranging from family snapshots and home movies to the tidal bore of Internet pornography. Digital cameras routinely reduce their multi-megapixel capture of light to a few hundred kilobytes, tossing aside superfluous ones and zeros. Dylan had it right, except for the emphasis: You don't want to throw it *all* away.

Conventional wisdom has it that compression needs to start with uncompressed data. But conventional wisdom, especially when based on quirks of language, is notoriously unreliable. A new theory of “compressive sampling” has crept onto the scene, showing how an image of interest—or structured signals generally—can be reconstructed, often exactly, from a surprisingly small set of direct measurements. Emmanuel Candès of Caltech gave an over-view of the subject, also known as “compressed sensing,” in an invited presentation at the 2006 SIAM Conference on Imaging Science, held in May at the Institute for Mathematics and its Applications.\* A series of minisymposia fleshed out the nascent theory and its applications, including what might be the epitome of compressed sensing: a one-pixel digital camera, developed by a group of engineers at Rice University.

Compressibility is based on the observation that, for a well-chosen basis set, such as a Fourier or wavelet basis, most signals of interest are concentrated on small subsets of the basis, meaning that they are closely approximated by vectors with a small number of nonzero coefficients. The meanings of “well-chosen” and “of interest” are a bit hazy and slightly circular: A basis is well-chosen if it succinctly describes a signal of interest; likewise, a signal is of interest if it can be described with just a handful of basis elements. Uninteresting signals have another name: noise.

The problem is, different signals are concentrated on different small subsets, so there would seem to be no alternative to processing each signal in its entirety—without, of course, opting for a Gallup-poll-like random sample, with its guaranteed uncertainties. Overlooking an important component of a signal seems almost inevitable if the whole haystack isn't thoroughly pitchforked over.

Enter  $\ell_1$ -magic.

The name refers to the normed vector space pronounced “little ell one.” In functional analysis, “Big Ell One,” usually written  $L^1$ , is a vector space of Lebesgue-integrable functions. When necessary, the domain of integration is specified, as in  $L^1(R)$  or  $L^1([0,1])$ .  $L^1$ 's little brother is the space of absolutely convergent infinite series. In signal processing, infinity is reduced to a large finite number  $N$ ; the trick, as always, is to find algorithms that remain practical as  $N$  tends technologically toward infinity.

A number of researchers, including Candès, Terrence Tao of UCLA, and Justin Romberg of Georgia Tech, have done just that. Candès and Tao, for example, have shown that  $\ell_1$ -magic boils down to a problem in linear programming. The basic idea is to minimize an  $\ell_1$  norm,  $\|x\|_1 = \sum |x_i|$ , subject to agreement with an appropriately random set of measurements. The nonlinear kink of the absolute value function is ironed out through expansion of the problem to one of minimizing  $\sum u_i$  with the additional inequalities  $-u_i \leq x_i \leq u_i$ . Under very general conditions, the researchers' “Dantzig selector” (named, they note, in honor of the late inventor of the simplex method) is about as good a solution as one could hope to get, even if one knew where in the haystack the signal was hidden.

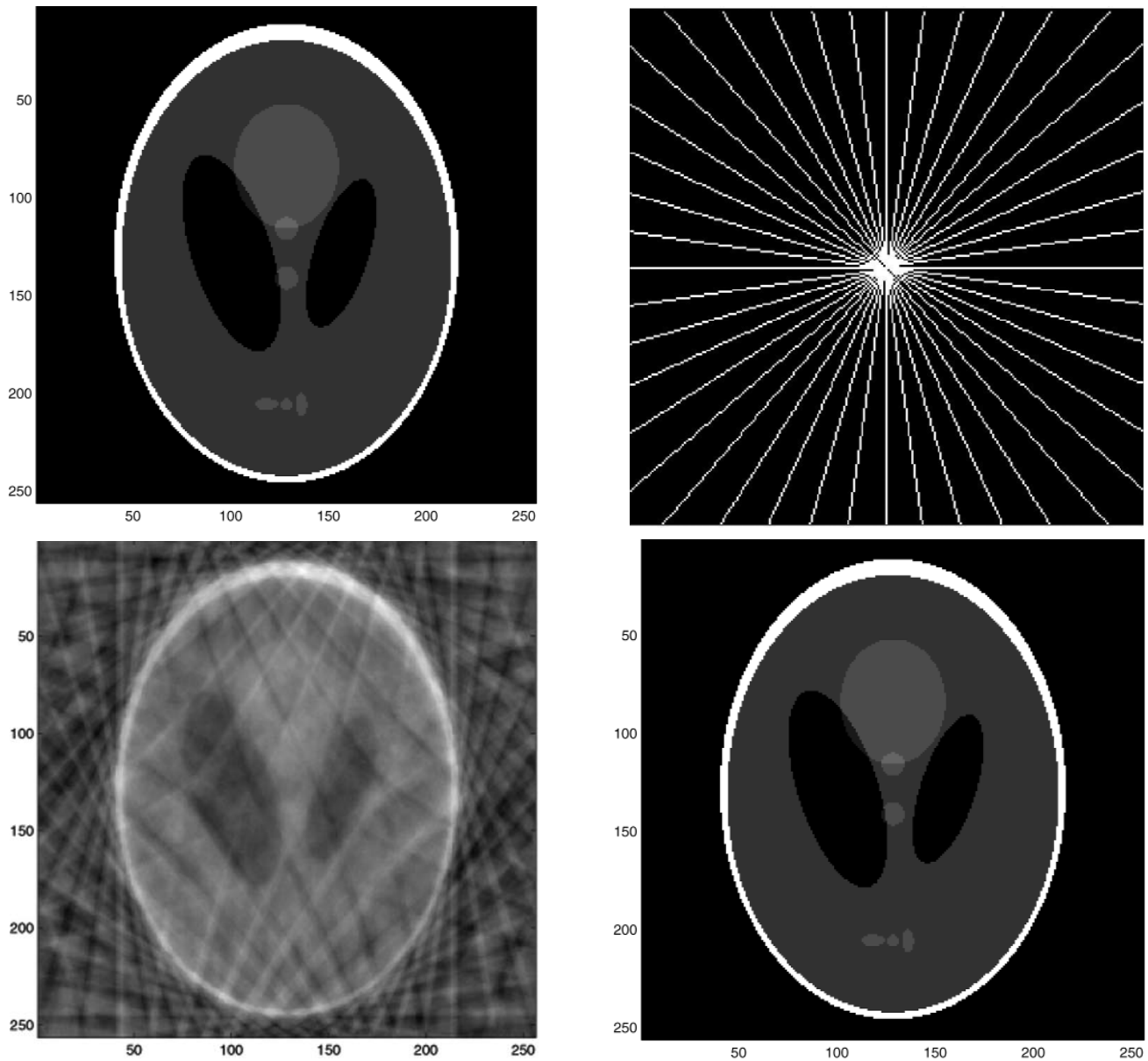
In effect,  $\ell_1$  is a happy medium between the least-squares norm  $\ell_2$ , which sums the squares of the coefficients, and the combinatorial metric  $\ell_0$ , which simply counts the number of nonzero coefficients (i.e., it sums the zeroth power of the coefficients). It shares  $\ell_0$ 's sensitivity to sparse vectors and  $\ell_2$ 's property of having a convex unit ball. (The unit ball for  $\ell_1$  is a generalized octahedron, which, Candès observes, is “pointy near sparse vectors.”) Minimizing the  $\ell_0$  norm would be ideal if that computational task weren't NP-complete; minimizing the  $\ell_2$  norm is much easier (that's why the Kalman filter has been so successful), but it requires too many measurements to guarantee an accurate picture. The stunning “magic” of  $\ell_1$  is that it combines the parsimony of  $\ell_0$  and the computational efficiency of  $\ell_2$ .

The key lies in the randomness of the measurements. Candès and colleagues have defined a notion of “uniform uncertainty” that guarantees, with arbitrarily high probability, an exact solution when the signal is sparse and a good approximation when it's noisy. Their uniform uncertainty principle is satisfied by many of the familiar approaches, including Gaussian random matrices and random row selections from the Fourier and general orthogonal measurement ensembles.

The notion that sparse signals—meaning signals with a small number of nonzero coefficients for a given basis (and no noise)—can (with high probability) be reconstructed exactly via  $\ell_1$ -minimization is not exactly new. The idea was first floated in 1986, by Fadil Santosa and William Symes, in a paper in *SIAM Journal on Scientific and Statistical Computing*. But the full extent of the theory, including the robustness of the reconstruction procedure, is only now coming into focus. One of the leaders is David Donoho, a statistician at Stanford University, who coined the term “compressed sensing” to emphasize the fact that  $\ell_1$ -magic is not just a new way of massaging a “complete” set of measurements into

---

\*Organizing committee co-chair Fadil Santosa's overview of the conference appeared in the September issue of *SIAM News*; [www.siam.org/news](http://www.siam.org/news). See Barry Cipra's report in this issue on work presented at the conference by Alexander Katsevich.



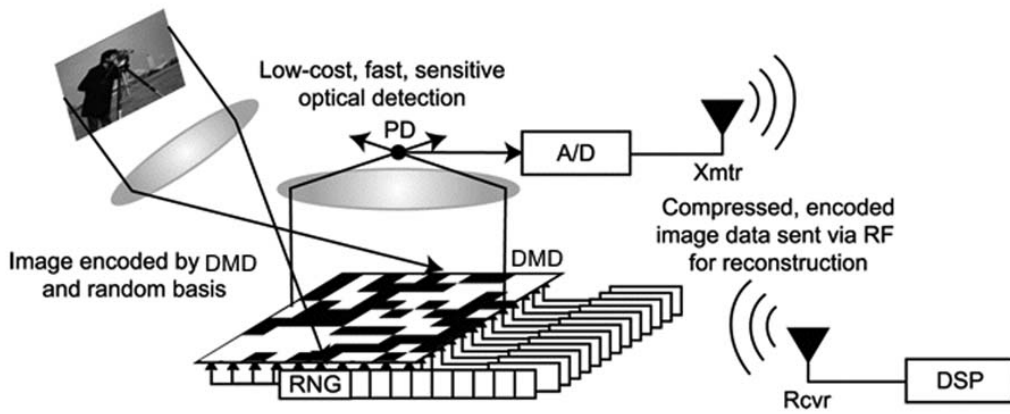
**Figure 1.** When Fourier coefficients of a testbed medical image known as the Logan–Shepp phantom (top left) are sampled along 22 radial lines in the frequency domain (top right), a naive, “minimal energy” reconstruction setting unobserved Fourier coefficients to 0 is marred by artifacts (bottom left). Surprisingly,  $l_1$ -magic reconstruction (bottom right) is not just better—it’s perfect!

a compact form, but rather a new way of thinking about how to measure things in the first place.

This new way of thinking has profoundly practical implications. Making measurements can be expensive, in terms of time, money, or (in the case of, say, x-rays) damage done to the object being imaged. Compressed sensing has the potential to provide substantial cost savings without sacrificing accuracy. In one impressive numerical experiment, Candès, Romberg, and Tao showed that a  $512 \times 512$ -pixel test image, known as the Logan–Shepp phantom, can be reconstructed exactly from 512 Fourier coefficients sampled along 22 radial lines—with, in other words, more than 95% of the ostensibly relevant data missing (see Figure 1).

The trick, Candès points out, will be to incorporate the appropriate randomness into the sensing hardware. In a minisymposium on random sensing at the SIAM conference, Kevin Kelly of Rice University described one such trick: a digital camera with a single photon detector. The heart of the camera, developed in conjunction with Richard Baraniuk and the Digital Signal Processing group at Rice ([dsp.rice.edu/cs/cscamera](http://dsp.rice.edu/cs/cscamera)), is a digital micromirror device (DMD) manufactured by Texas Instruments. The DMD consists of a  $1024 \times 768$  array of hinged mirrors. Each mirror has two positions—one in which it reflects light toward the detector, and one in which it reflects light away—and is capable of switching from one to the other thousands of times per second.

The detector measures the total intensity of light for a given arrangement of the micromirrors—in essence, it registers the average grayscale for the subset of pixels corresponding to the micromirrors that are “turned on.” By rapidly running the DMD through a predetermined set of



**Figure 2.** Mixed signals. A "one-pixel" camera developed at Rice University uses a random number generator (RNG) to control the orientations of mirrors on a digital micromirror device (DMD), thereby steering light from random portions of a scene onto a single photodiode (PD). The PD's voltage output for a sequence of random orientations is enough to reconstruct the scene.

random (or pseudorandom) settings, the camera outputs a string of bits that feed into a reconstruction algorithm (see Figure 2). The Rice researchers' prototype shows that it's not necessary to throw data away—it suffices not to collect it in the first place.

Candès sees an unbounded future for the  $\ell_1$  norm, with applications ranging from statistical design to error-correcting codes. Its impact, he says, may have been described most succinctly by a colleague he spoke with recently. "I explained my work to him and he said this: 'The 20th century was the century of least squares, and the 21st may very well be that of  $\ell_1$ .'"

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