Discussion of Donoho:

Sparse Signals and Approximate Message Passing

David Banks
Duke University

1. Introduction

David has described an active area of research that, over the last seven years, has revolutionized the conventional thinking in mathematics, statistics and imaging. He makes it sound intuitive, but:

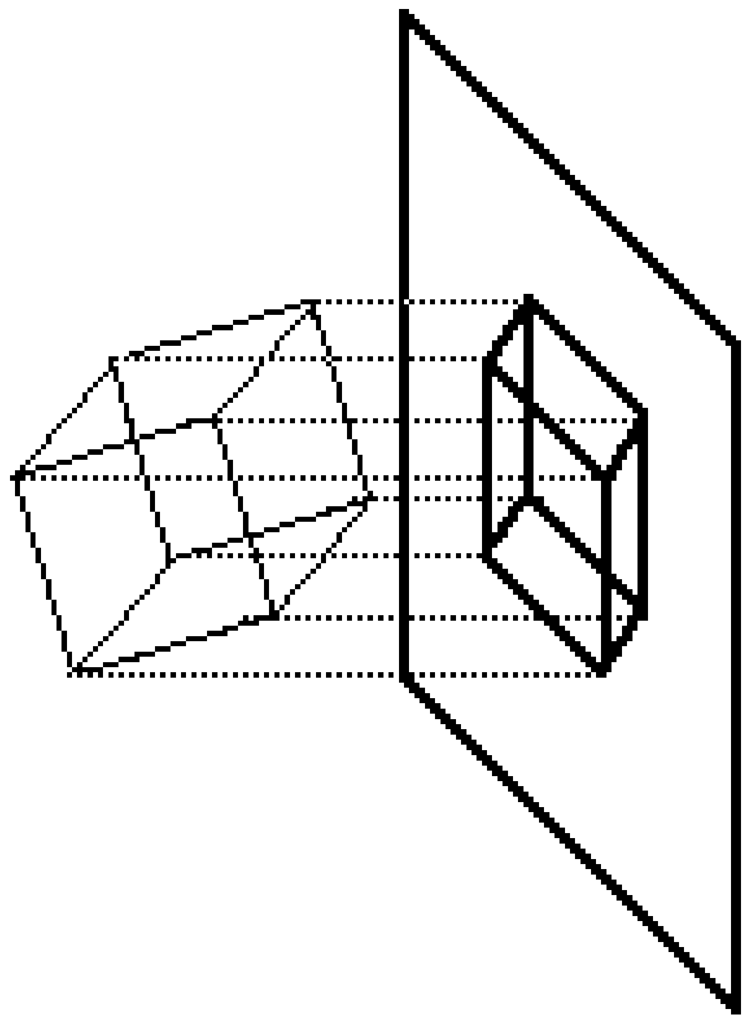
Rubáiyát Quatrain

Myself when young did eagerly frequent
Doctor and Saint, and heard great argument
About it and about: but evermore
Came out by the same door where in I went.

Omar Khayyám

The key steps in David's argument are:

- Often, under-determined linear systems with sparse solutions can be solved by L_1 minimization. (Combinatorial geometry: which vertices of a polytope in $\mathbb{R}^p$ remain vertices after projection onto $\mathbb{R}^k$?)
- There is a phase transition that depends on true sparsity, problem dimension, and undersampling (ill-posedness). This transition distinguishes cases that are almost surely solvable from those that are almost surely insolvable.
- When there is noise in the data, sparsity does not hold. But one can look in a neighborhood of the data, and find the sparse solution that (may) correspond to the denoised data. (This relates to minimaxity theory, harmonic analysis, ℓ^p balls.)
- For large dimensions and many measurements, traditional computation is too slow. But Approximate Message Passing is magic! (The stochastic formulation of the problem satisfies key conditions for the replica trick from spin glass models and enables exponential convergence rates.)



This work has broad applicability: medical imaging, data compression and mining, seismology, radar, genotyping, and so forth.

But our goal is to understand how this can apply to astrostatistics.

- Kirk Borne described the LSST project, which will produce a stadium-full of CDs of data and requiring real-time analysis.
- Hyperspectral measurements of the kind Ann Lee and Kaisey Mandel (and others) have discussed.
- Othman Benomar addressed astro-seismology.
- Alexander Gray directly called for faster algorithms to do heavy lifting in machine learning, and David's work is a swift response.
- I suspect that Thomas Lee's talk will also find this methodology to be pertinent.

But things will not be straightforward.

One issue is that the current theory applies to linear regression with Gaussian noise. There are obvious ways to extend this to nonlinear models and (less obviously) non-Gaussian situations. But I expect astrostatisticians will always want to use tools that are just a little beyond the domain in which strong theoretical justification is available.

In particular, some of the issues raised by Brandon Kelly and David Ruppert on measurement error models will not be confidently addressed in this framework.

For example, I believe one would want to put a small ℓ^p ball B_ϵ around $\boldsymbol{y}^*$ and then find the most sparse solution $\boldsymbol{x}$ that

$$\exists y \in B_\epsilon \textbf{ such that } y = A\boldsymbol{x}.$$

David's result for the noisy case looks a little different—it seems to be more of a bound for an errors-in-variables situation.

In some cases, such as hyperspectral data cube analysis, one will want to apply the compressed sensing methodology to multiple, but similar, problems. For example, it might be reasonable to assume that all of the separate solutions share about the same innate sparsity. How could one use that to jointly improve all of the estimates?

Heavy-tailed noise will be a significant challenge (i.e., blunders). One would like to find the most sparse solution that fits, say, 95% of the observations. (This is similar to the outlier problem in Ann Lee's diffusion metrics.)

In some cases, it would be nice to be able to plot the mean squared error of the best solution with sparsity k against k. As with the LASSO and LARS, this kind of trace plot might be informative.

Nonetheless, this is an amazing key for the LSST, SDSS, Pan-STARRS, VISTA, and other high-profile, multivariate, data-intensive collection effors that are ongoing or soon to start.

Along with his papers, David generously sent me a crib sheet for my discussion. He suggested I make the following points:

- Statisticians have been doing this sort of thing for generations, it's called model selection.
- Bayesians have slab and spike models for doing the same thing.
- This shows once again the power of prior information—in this case sparsity.
- This shows once again the power of theory—it enables practitioners to do what they would want to do (undersample)—but with a secure foundation.

All these points are exactly correct. But I want to emphasize that David and his collaborators and colleagues have kicked things up a notch.

It is also notable that these results draw upon diverse and sometimes obscure mathematical results.

I'd like to extend David's suggestion a bit more broadly, in response to many of the talks I've heard. There are many kinds of sparsity, and some seem clearly important to astrostatistics.

- The assumption that only a small number of covariates or basis elements have non-zero coefficients. This is the case David has addressed.
- Locally low-dimensional structure. Here all of the variables may be important some of the time, but in a given region, only a few are relevant (c.f. Ann Lee's diffusion metric problem, Chris Genovese's work on foams and filaments).
- Mixture models: one observes a superposition of data arising from many simpler models (as in the L_2 estimation via mixutres of normals that Tamas Budavari discussed, but it applies far more generally, e.g., different evolutionary tracks through the Hertzsprung-Russell diagram).

One can ask more of sparsity. In a mixture, a small number of components may fit well, but several are close together; from the suprise detection standpoint that Kirk Borne advocated, it is better to find a slightly poorer fit that has widely dispersed components. Or, in the context of bases, a good fit with 10 nearby wavelets may be less useful that a slightly poorer fit with distinct elements.

Plug for a SAMSI Program

In 2006, the Statistical and Applied Mathematical Sciences Institute at the Research Triangle, NC, sponsored a semester-long program on astrostatistics. It was a great success; Jogesh Babu, Tom Loredo, David van Dyk, Chris Genovese, and many other people here here played a large role and deserve great credit.

In the academic year of 2012/13, there will be a new SAMSI program on massive data sets. One-third of that program will be devoted to astrostatistics. The program co-leader is Tamas Budavari. I hope you can join us in North Carolina for at least some portion of the year.

All the presentations so far, and certainly David's, have underscored the need for a new and more ambitious vehicle for collaborative research in this field. I think the SAMSI program offers a unique opportunity to achieve this.

When I heard the Learn'd Astronomer

When I heard the learn'd astronomer;
When the proofs, the figures, were ranged in columns before me;
When I was shown the charts and the diagrams, to add, divide, and measure them;
When I, sitting, heard the astronomer, where he lectured with much applause in the lecture-room,
How soon, unaccountable, I became tired and sick;
Till rising and gliding out, I wander'd off by myself,
In the mystical moist night-air, and from time to time,
Look'd up in perfect silence at the stars.

Walt Whitman