published in Journal of the Royal Statistical Society B

Andrew Gelman's contribution to the discussion of "Statistical exploration of the manifold hypothesis" by Whiteley et al.

Andrew Gelman, Professor, Department of Statistics and Department of Political Science, Columbia University, New York, ag389@columbia.edu

Whiteley et al. (2025) discuss scenarios in which high-dimensional data can be well described by an underlying sparse structure, providing a theoretical explanation for an important idea in applied statistics.

In an article entitled "In praise of sparsity and convexity," Tibshirani (2014) wrote, "Hastie et al. (2001) coined the informal 'Bet on Sparsity' principle. The ℓ_1 methods assume that the truth is sparse, in some basis. If the assumption holds true, then the parameters can be efficiently estimated using ℓ_1 penalties. If the assumption does not hold—so that the truth is dense—then no method will be able to recover the underlying model without a large amount of data per parameter."

This is an appealing argument along the lines of Pascal's wager: bet on sparsity because, in the absence of sparsity, the truth cannot be discovered anyway.

But how do we think about the "bet on sparsity" principle in a world where the truth is dense? I'm thinking here of social science, where no effects are clean and no coefficient is zero (see Gelman, 2011, p. 960, for some discussion of this point), where every contrast is meaningful—but some of these contrasts might be lost in the noise with any realistic finite amount of data.

I think there is a way out here, which is that in a dense setting we are not actually interested in "recovering the underlying model." The underlying model, such as it is, is a continuous mix of effects. If there's no discrete thing to recover, there's no reason to worry that we can't recover it!

I'm sure things are different in a field such as chemistry, where you can try to identify the key compounds that make up some substance, or image reconstruction, in which it should be possible to identify cars on a street or stars in the sky or organs within a body.

I do think it can often make sense to consider the decision-analytic reasons why it can make sense to go for sparsity: sparse models can be faster to compute, easier to understand, and yield more stable inferences. (Sometimes people say that a sparse model is less likely to overfit but I don't think that's quite right, as you can also get rid of overfitting by using a strong regularizer. But I think it is fair to say that a sparse model can yield more stable inferences, in that the inferences for the more complex model can be sensitive to the details of the regularizer or the prior distribution.)

References

Gelman, A. (2011). Causality and statistical learning. *American Journal of Sociology* 117, 955-966.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer.

Tibshirani, R. J. (2014). In praise of sparsity and convexity. In *Past, Present, and Future of Statistical Science*, ed. X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott, and J. L. Wang, 497-505. New York: Wiley.

Whiteley, N., Gray, A., and Rubin-Delanchy, P. (2025). Statistical exploration of the manifold hypothesis. Journal of the Royal Statistical Society B.