SCALABLE INFERENCE AND RECOVERY FROM COMPRESSIVE MEASUREMENTS

Mark A. Davenport and Richard G. Baraniuk

Electrical and Computer Engineering
Rice University
{md,richb}@rice.edu

Michael B. Wakin

Applied and Computational Mathematics
California Institute of Technology
wakin@acm.caltech.edu

ABSTRACT

Despite the apparent need for adaptive, nonlinear techniques for dimensionality reduction, random linear projections have proven to be extremely effective at capturing signal structure in cases where the signal obeys a low-dimensional model. Similarly, random projections are a useful tool for solving problems where the ultimate question of interest about the data requires a small amount of information compared to the dimensionality of the data itself. The success of random projections in both of these arenas can be traced to an elementary concentration of measure property, which allows us to extend the utility of random projections to a variety of new signal models and applications.

1. INTRODUCTION

Over the past several years, sensors and signal processing algorithms and hardware have been under increasing pressure to efficiently acquire, store, and process ever larger and higher-dimensional data sets. In some cases, dimensionality reduction techniques can help to reduce (the dimension of) this burden by extracting key low-dimensional information about the high-dimensional signals from which we can later infer the key properties of the original data. This low-dimensional information can often be more efficiently acquired, stored, and/or processed than the original high-dimensional data.

The success of dimensionality reduction often derives from one of two sources: **(1) Low-dimensional signal models:** The signals of interest may have "few degrees of freedom" relative to their size N. Examples include sparse and compressible signals, manifolds, etc. **(2) Low-complexity inference:** The problem we wish to solve may have "low-complexity" in that its solution requires a small amount of information relative to the dimension N. Examples include function estimation, signal detection, classification, etc.

Standard techniques for dimensionality reduction often attempt to discover this low-dimensional structure from a collection of training data and then *adaptively* construct a *nonlinear* mapping that preserves the key information. However, during the last decade, a number of communities have discovered that a surprisingly effective technique for dimensionality reduction is simply to collect a reduced set of *random*, *nonadaptive*, *linear* measurements of the data.

In the computer science community, random measurements have been proposed for solving nearest neighbor and clustering problems, estimating database statistics, etc. [1,2]. More recently, in the mathematics community, random measurements have been proposed as a means for the compressive sampling of sparse signals: obtaining a reduced set of measurements from which a sparse signal can be recovered [3, 4]. In each of these cases, random projections capture the substantive information about a signal (or group of signals) without regard to a priori knowledge about the data itself or to any structure within the data that might aid acquisition. This aspect is very attractive for practical applications, as there is no need for preprocessing on the data to discover its structure. From an engineering perspective, this has spurred research into developing physical devices for directly acquiring low-rate digital compressive measurements of high-bandwidth analog signals [5].

We will discuss a number of possible extensions of random projections in signal processing, with applications much more broad than have previously been explored by the above communities. This work is inspired by the fact that a solid mathematical theory actually connects much of the previous work. In particular, we have recently shown that the Johnson-Lindenstrauss (JL) Lemma [6], which ensures an isometric embedding of a finite point cloud under a random dimensionality-reducing projection and is fundamental to several results in computer science, also underlies the CS theory via a concentration of measure inequality [7]. This simple inequality, stating that the norm of a signal is well-preserved under a random dimensionality-reducing projection, allows us to show that in many settings the distinguishing characteristics of a signal can be encoded in a few random measurements. We will discuss a number of implications of this result in signal processing applications, both for processing signals obeying a low-dimensional signal model and for solving low-complexity inference problems. The following sections give a brief overview of some of these topics.

2. LOW-DIMENSIONAL SIGNAL MODELS

In many cases, one may have a model for the signals of interest that carries a notion *conciseness*; for example, one may believe that a signal $x \in \mathbb{R}^N$ has "few degrees of freedom" relative to its size N. Letting $\mathcal{F} \subset \mathbb{R}^N$ denote the class of signals of interest under a given model, many concise models correspond to a *low-dimensional geometric structure* for $\mathcal{F}$, which suggests the possibility for and gives new insight into dimensionality reduction techniques.

Sparsity. In a sparse signal model, every signal from the class $\mathcal{F}$ can again be represented (either exactly or approximately) using a K-term representation from some basis Ψ, but the relevant set of basis elements may change from signal to signal. Transform coding algorithms (which form the heart of many modern signal and image compression standards such as JPEG and JPEG-2000 [10]) exploit this "conciseness". With few exceptions, such approaches tend to be nonlinear (owing to the nonlinear structure of the signal class $\mathcal{F}$) and adaptive (requiring a search for the few relevant dictionary vectors for each signal of interest).

In the last two years, however, a radically different technique known as *Compressed Sensing* (CS) [3, 4] has emerged that relies only on *nonadaptive*, *linear*, *random* projections for dimensionality reduction. The CS theory states that with high

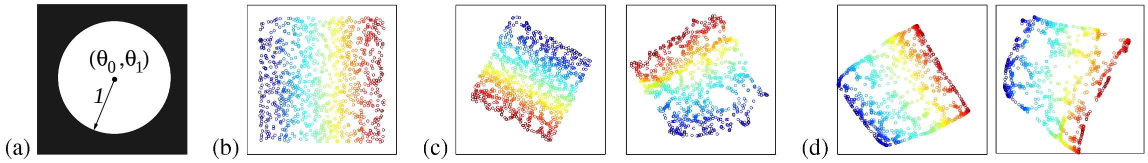

Fig. 1. *Manifold learning from compressive measurements. (a) Model for data. We generate 1000 images of a shifted disk, each of size $N = 64 \times 64 = 4096$. (b) True θ_0 and θ_1 parameter values for original data in $\mathbb{R}^N$. (c) ISOMAP [8] embedding learned from (left) original data in $\mathbb{R}^N$ and (right) a random projection of the data to $\mathbb{R}^{15}$. (d) Laplacian eigenmaps [9] embedding learned from (left) original data in $\mathbb{R}^N$ and (right) a random projection of the data to $\mathbb{R}^{15}$.*

probability, every K-sparse signal x can be recovered from just $M = O(K \log(N/K))$ linear projections onto random vectors in $\mathbb{R}^N$ [3, 4]. CS decoding involves recovering the signal $x \in \mathbb{R}^N$ from its measurements $y = \Phi x$, where $y \in \mathbb{R}^M$ and Φ is a random $M \times N$ matrix. Although such inverse problems are generally ill-posed whenever $M < N$, CS recovery algorithms exploit the additional assumption of *sparsity* in the basis Ψ to identify the correct signal x from an uncountable number of possibilities. In a sense, CS allows us to directly acquire signals is a compressed form, opening new possibilities in signal acquisition, imaging, and sensor networks [5, 11–17].

Manifolds. Manifold models for the signal class $\mathcal{F}$ generalize the notion of concise signal structure beyond the framework of sparse representations. These models arise in more broad cases where (i) a K-dimensional parameter θ can be identified that carries the relevant information about a signal and (ii) the signal $x_\theta \in \mathbb{R}^N$ changes as a continuous (typically nonlinear) function of these parameters. Low-dimensional manifolds have also been proposed as approximate models for nonparametric signal classes such as images of human faces or handwritten digits [18, 19].

Similar to the case of the sparse models described above, the low-dimensional geometry of manifold signal models makes them amenable to dimensionality reduction in general and random projections in particular. In parallel with the CS theory, we have established that from a sufficient number $M = O(K \log(N))$ of random measurements, with high probability, all pairwise distances between points on a manifold $\mathcal{F} \subset \mathbb{R}^N$ are well-preserved under the mapping Φ to $\mathbb{R}^M$ [20]. We will discuss the possible applications of this fact in CS recovery, which can be extended beyond sparse signals to include manifold-modeled signals.

3. LOW-COMPLEXITY INFERENCE

Even in the more general case where we do not have a low-dimensional signal model, there is still hope that dimensionality reduction may be possible when the problem that we ultimately wish to solve is of low-complexity in that it requires only a small amount of information to solve compared to the dimension N. There is a spectrum of such problems, ranging from the estimation of arbitrary functions of the data to classification and detection.

Estimation. Consider a signal $x \in \mathbb{R}^N$, and suppose that we wish to estimate some function $f(x)$ but only observe the measurements $y = \Phi x$, where Φ is again an $M \times N$ matrix. The *data streaming* community – which is concerned with processing large streams of data using efficient algorithms – has previously analyzed this problem for many common functions (such as linear functions, ℓ_p norms, and histograms). These estimates are often based on so-called *sketches*, which can be thought of as random projections. For a concise review of results from this community, see [2].

As an example, in the case where f is a *linear* function, one can show that the estimation error (relative to the norms of x and f) can be bounded by a constant determined by M. We have recently demonstrated that this result, originally proven in [21], holds for a wide class of random matrices, and can be viewed as a straightforward consequence of the same concentration of measure inequality that has proven useful for CS and in proving the JL Lemma [22].

Detection. As opposed to estimation where there is a continuum of possible values, in detection one simply wishes to answer the question, is a (known) signal present in the observed signal? To solve this problem, it is sufficient to estimate the relevant *sufficient statistic*. In [22] we have shown, again using the concentration of measure inequality, that we can estimate the sufficient statistic for such a detection problem from random projections, where the quality of this estimate depends on the SNR. We make no assumptions on the signal of interest s, and hence we can build systems capable of detecting s even when s is not known in advance. Thus we can use random projections for dimensionality-reduction in the detection setting without knowing the relevant structure.

Classification. Similarly, random projections have long been used for a variety of classification and clustering problems. The JL Lemma is often exploited in this setting to compute approximate nearest neighbors, which is naturally related to classification [1]. The key result that random projections result in an isometric embedding allows us to generalize this work to several new classification algorithms and settings. See [22] for some initial steps in this direction.

4. LOW-DIMENSIONAL MODELS MEET LOW-COMPLEXITY INFERENCE

While random projections have frequently been exploited to take advantage of low-dimensional models or to solve low-complexity inference problems, we believe that these techniques may be most useful when these areas begin to overlap.

For example, in [23] we study the case of sparse signal detection in the presence of inference/noise. We propose a greedy algorithm for solving this problem and demonstrate that the number of measurements and computations necessary for successful detection is significantly lower than what would be necessary for successful reconstruction. Simulations show this algorithm is very resilient to strong interference, additive noise, and measurement quantization.

Random projections can also be used to solve low-complexity inference problems in the context of manifold signal models. The utility of random projections in such settings comes in the form of several corollaries of the distance-preserving result mentioned in Sec. 2: many key properties of a manifold are in fact preserved under a random projection to lower-dimensional space, including dimension, topology, geodesic distances, and curvature. This opens several new application areas for random projections in manifold-based learning, recognition, classification. We will present several promising experiments in these areas (see Fig. 1 for one example).

5. REFERENCES

[1] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *Proc. 30th STOC*, 1998, pp. 604–613.

[2] S. Muthukrishnan, *Data Streams: Algorithms and Applications*, now, 2005.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," 2004, Preprint.

[5] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K. F. Kelly, and R. G. Baraniuk, "A compressed sensing camera: New theory and an implementation using digital micromirrors," in *Proc. Computational Imaging IV at SPIE Electronic Imaging*, San Jose, January 2006, SPIE.

[6] W. B Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. in Modern Analysis and Probability*, 1984, pp. 189–206.

[7] R. G. Baraniuk, M. A. Davenport, R. A. DeVore, and M. B. Wakin, "The Johnson-Lindenstrauss lemma meets compressed sensing," 2006, Preprint.

[8] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, December 2000.

[9] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, June 2003.

[10] D. S. Taubman and M. W. Marcellin, "JPEG 2000: Image compression fundamentals, standards and practice," *Kluwer*, 2001.

[11] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[12] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," 2005, Preprint.

[13] E. Candès and T. Tao, "Error correction via linear programming," *Found. of Comp. Math.*, 2005, Preprint.

[14] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," 2004, Preprint.

[15] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[16] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005, Preprint.

[17] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Atlanta, GA, 2006.

[18] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[19] D. S. Broomhead and M. J. Kirby, "The Whitney Reduction Network: A method for computing autoassociative graphs," *Neural Computation*, vol. 13, pp. 2595–2616, 2001.

[20] R. G. Baraniuk and M. B. Wakin, "Random projections of signal manifolds," 2006, Preprint.

[21] N. Alon, P. Gibbons, Y. Matias, and M. Szegedy, "Tracking join and self-join sizes in limited storage," in *Proc. Symp. Principles of Database Systems (PODS)*, Philadelphia, PA, 1999.

[22] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "The compressive matched filter," Tech. Rep. TREE 0610, Rice University, Dept. of Elec. and Comp. Engineering, 2006, See http://www.ece.rice.edu/~md.

[23] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2006, Toulouse, France.