Low-Dimensional Models for Dimensionality Reduction and Signal Recovery: A Geometric Perspective

Richard G. Baraniuk, Volkan Cevher, Michael B. Wakin

***Abstract*—We compare and contrast from a geometric perspective a number of low-dimensional signal models that support stable information-preserving dimensionality reduction. We consider sparse and compressible signal models for deterministic and random signals, structured sparse and compressible signal models, point clouds, and manifold signal models. Each model has a particular geometrical structure that enables signal information to be stably preserved via a simple linear and nonadaptive projection to a much lower dimensional space; in each case the projection dimension is independent of the signal's ambient dimension at best or grows logarithmically with it at worst. As a bonus, we point out a common misconception related to probabilistic compressible signal models, namely, by showing that the oft-used generalized Gaussian and Laplacian models do not support stable linear dimensionality reduction.**

***Index Terms*—dimensionality reduction, stable embedding, sparsity, compression, point cloud, manifold, compressive sensing.**

I. Introduction

A. Dimensionality reduction

Myriad applications in data analysis and processing—from deconvolution to data mining and from compression to compressive sensing—involve a linear projection of data points into a lower-dimensional space via

$$y = \Phi x + n. \tag{1}$$

In this *dimensionality reduction*, the *signal* $x \in \mathbb{R}^N$ and the *measurements* $y \in \mathbb{R}^M$ with $M < N$; Φ is an $M \times N$ matrix; and n accounts for any noise incurred.

Such a projection process loses signal information in general, since Φ has a nontrivial null space $\mathcal{N}(\Phi)$ whenever $M < N$. For any fixed signal x, then, $\Phi x = \Phi(x + z)$ for any $z \in \mathcal{N}(\Phi)$. Hence there has been significant interest over the last few decades in finding dimensionality reductions that preserve as much information as possible in the measurements y about *certain* signals x. In this paper, we consider the application of Φ not to arbitrary signals $x \in \mathbb{R}^N$ but rather to some subset of $\mathbb{R}^N$, which we will call a *low-dimensional signal model* or a *model* for short. For signals from a model we are concerned with ensuring that Φ is *information preserving*, by which we mean that Φ provides a *stable embedding* that approximately preserves distances between all pairs of signals in the model. Such projections certainly seem useful. For instance, in some cases we should be able to recover x from its measurements y; in others we should be able to process the much smaller y to effect processing on the much larger x in order to beat the curse of dimensionality.

Consider a very simple low-dimensional signal model: a K-dimensional linear subspace spanned by K of the canonical basis vectors in $\mathbb{R}^N$, where we assume $K < M < N$. Signals conforming to this model have K nonzero and $(N-K)$ zero coefficients, with the locations of the nonzero coefficients (called the *support* of x) determined by the subspace. In this case, we can write $\Phi x = \Phi_K x_K$, where x_K consists of the K coefficients in the support of x and Φ_K consists of the corresponding columns of Φ. Two key properties follow immediately. First, Φ_K is an $M \times K$ matrix, and so as long as its K columns are linearly independent, its nullspace contains only the zero-vector. Second, Φ_K can be designed to preserve distances between any two signals in the subspace (and thus be a perfectly stable embedding) merely by making those same columns orthonormal.

B. Low-dimensional models

The problem with applying this classical linear algebra theory is that in many applications such a simple linear subspace model does not apply. Nonetheless, there are a handful of alternative low-dimensional models appropriate for wide varieties of real-world signals. In this paper we will focus on the following salient examples.

RGB and VC are with Rice University, Department of Electrical and Computer Engineering. MBW is with the Colorado School of Mines, Division of Engineering. Email: {richb, volkan}@rice.edu, mwakin@mines.edu; Web: dsp.rice.edu/cs. This work was supported by the grants NSF CCF–0431150, CCF–0728867, CCF–0830320, and DMS–0603606; DARPA HR0011–08–1–0078; DARPA/ONR N66001–08–1–2065; ONR N00014–07–1–0936 and N00014–08–1–1112; AFOSR FA9550-07-1-0301; ARO MURI W311NF-07-1-0185; ARO MURI W911NF-09-1-0383; and the Texas Instruments Leadership University Program.

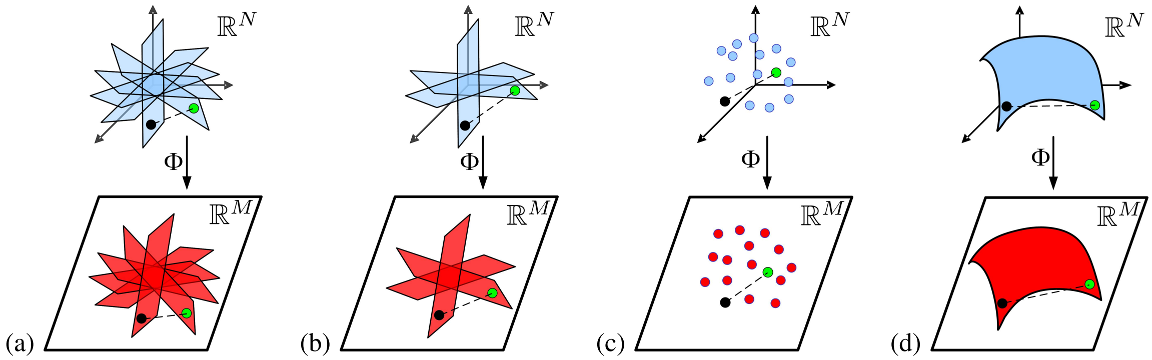

Fig. 1. *A menagerie of low-dimensional signal models in $\mathbb{R}^N$: (a) K-sparse signals, the union of K-dimensional subspaces Σ_K from Section III-A; (b) K-model sparse signals, a reduced union of K-dimensional subspaces from Section V-A; (c) a point cloud from Section VI-A; and (d) a smooth K-dimensional manifold from Section VI-B. All of these signal models support an information-preserving, stable embedding into a much lower-dimensional space $\mathbb{R}^M$.*

To begin, *sparse signal models* generalize the linear subspace model by again considering signals x containing just $K < N$ nonzero coefficients but allowing the locations of those nonzeros to be arbitrary. Moreover, sparse signal models generally allow us to consider arbitrary bases; that is, letting Ψ be an orthonormal basis for $\mathbb{R}^N$ represented as a unitary $N \times N$ matrix, we have $x = \Psi\alpha$ where α has just K nonzeros. Geometrically, the set of K-sparse signals in the basis Ψ consists of the union of all possible K-dimensional subspaces in $\mathbb{R}^N$ spanned by K basis vectors from Ψ. This is a highly nonlinear set (see Figure 1(a)).

Many transform-based data compression and data processing algorithms rely on sparse models, computing the transform coefficients $\alpha = \Psi^T x$ in order to exactly or approximately "compact" the signal energy into just K coefficients. Sparse signal models are appropriate in the time domain for modeling intermittent spike trains; in the frequency domain for modeling wideband signals comprised of just a few arbitrary sinusoids; and in the wavelet domain for modeling piecewise smooth signals with intermittent discontinuities. An important extension to sparse models are *compressible signal models* in which signals x have coefficients $\alpha = \Psi^T x$ whose sorted magnitudes decay rapidly to zero with a power law. Such signals are well-approximated as K-sparse. Images, for example, are frequently modeled as compressible in the 2-D wavelet domain.

As an additional layer of structure on top of these models, many advanced data compression and Bayesian estimation algorithms involve modeling the deterministic structure or probabilistic correlations among the nonzero entries in a coefficient vector α. For example, the locations of the nonzeros may tend to cluster in groups, or in wavelet-sparse signals, the nonzeros may concentrate along a connected tree. In cases such as these, the resulting *structured sparsity models* consist of a reduced union of subspaces in $\mathbb{R}^N$ (see Figure 1(b)).

In statistics, pattern recognition, and learning theory, we have frequently have access to only a finite number of signals. *Point cloud models* (see Figure 1(c)) are appropriate in settings such as these where a finite—but possibly large—database of signals will be considered but there is no additional structure (such as sparsity) assumed for these signals.

Finally, in many problems such as object recognition, *manifold signal models* may arise when a family of signals $\mathcal{M} = \{x_\theta \in \mathbb{R}^N : \theta \in \Theta\}$ is smoothly parameterized by a K-dimensional parameter vector θ. Example parameterizations include: local signal parameters such as the orientation of an edge in a small image segment, or global parameters such as the position of a camera or microphone recording a scene or the relative placement of objects in a scene. Loosely speaking, the manifold $\mathcal{M}$ corresponds to a nonlinear K-dimensional surface in $\mathbb{R}^N$ (see Figure 1(d)).[1] Manifolds have also been proposed as approximate models for nonparametric signal classes such as images of handwritten digits [1]. In many cases the manifold dimension for a signal family may be much lower than the sparsity level afforded by any basis Ψ.

[1] A linear K-dimensional hyperplane is a very simple example of a manifold. However, the set of K-sparse signals in a fixed basis does not meet the technical criteria for a manifold due to the locations at which the K-planes intersect.

C. Paper organization

The three main goals of this paper are to: (*i*) Provide a unified treatment of the above low-dimensional, nonlinear signal models in both deterministic and random settings. As we will see, each model has a particular geometrical structure that enables information in $\mathbb{R}^N$ to be stably preserved via a simple linear and nonadaptive projection to a much lower dimensional space $\mathbb{R}^M$. Moreover, M either is independent of the ambient dimension N at best or grows logarithmically with N at worst. (*ii*) Point out a somewhat surprising misconception regarding low-dimensional random signal models. In particular, the oft-used generalized Gaussian and Laplacian random models are in fact *not* valid low-dimensional models in that they do not support a stable embedding as N grows. (*iii*) Indicate some fruitful emerging research directions that are inspired by the geometric viewpoint.

The results of this paper should be useful to readers interested in understanding sparse signals models and in particular to those interested in how the concept of sparsity generalizes from deterministic to random models and from unions of subspaces to point clouds and manifolds. Though many interesting applications are enabled by low-dimensional stable embeddings, we focus primarily in this paper on the task of recovering a signal x from its compressive measurements $y = \Phi x + n$. For a given number of measurements M, this recovery is generally best accomplished by employing the lowest-dimensional model for x that supports a stable embedding.

This paper is organized as follows. After defining stable embeddings in Section II, we consider deterministic sparse models in Section III and random sparse models in Section IV. In Section V we discuss structured sparse signals, including models for signal ensembles. We discuss how similar results hold for apparently quite different models in Section VI, where we consider point clouds and manifolds. We conclude with a discussion of future areas for research in Section VII.

II. Dimensionality Reduction and Stable Embedding

In this paper, we will study several classes of low-dimensional models for which the dimensionality reduction process (1) is *stable*, meaning that we have not only the information preservation guarantee that $\Phi x_1 \neq \Phi x_2$ holds for all signal pairs x_1, x_2 belonging to the model set but also the guarantee that if x_1 and x_2 are far apart in $\mathbb{R}^N$ then their respective projections Φx_1 and Φx_2 are also far apart in $\mathbb{R}^M$. This latter guarantee ensures robustness of the dimensionality reduction process to the noise n.

A requirement on the matrix Φ that combines both the information preservation and stability properties for a signal model is the so-called *ϵ-stable embedding*

$$(1-\epsilon)\|x_1-x_2\|_2^2 \leq \|\Phi x_1-\Phi x_2\|_2^2 \leq (1+\epsilon)\|x_1-x_2\|_2^2 \tag{2}$$

which must hold for all x_1, x_2 in the model set. The interpretation is simple: *a stable embedding approximately preserves the Euclidean distances between all points in a signal model.* This concept is illustrated in Figure 1 for the several models that we consider.

Given a signal model, two central questions arise: (*i*) how to find a stable $M \times N$ embedding matrix Φ, and (*ii*) how large must M be? In general, these are very difficult questions that have been attacked for decades using various different approaches for different kinds of models [2–7]. The models we consider in this paper are united by the relatively recently proved facts that their sparse geometrical structures can all be stably embedded with high probability by a *random* projection matrix Φ. To generate such a Φ, we draw realizations from an independent and identically distributed (iid) sub-Gaussian probability distribution.

In this sense, randomized compressive measurements provide a *universal* mechanism for dimensionality reduction in which the operative signal model need not be known at the time of measurement. Indeed a single randomly generated Φ can provide a stable embedding of an exponential number of candidate models simultaneously [8], allowing for later signal recovery, discrimination, or classification [9].

We begin our exposé by introducing deterministic sparse and compressible signal models.

III. Deterministic Sparse Models

A. Sparse deterministic signals

A great many natural and manmade signals have an exact or approximate *sparse representation* in an orthonormal basis $\{\psi_i\}$; that is

$$x = \sum_{i=1}^{N} \alpha_i \, \psi_i \tag{3}$$

as usual but only $K \ll N$ of the coefficients α_i are nonzero. Stacking the $\{\psi_i\}$ as columns into the $N \times N$ matrix Ψ, we can write $x = \Psi\alpha$ and $\alpha = \Psi^T x$, with α a K-sparse vector. Following a standard convention, we define the ℓ_0 "norm" $\|\alpha\|_0$ to be the number of nonzero entries of α (though this is not technically a valid norm).

Geometrical structure: The set $\Sigma_K = \Sigma_K(\Psi)$ of all K-sparse signals in the basis Ψ corresponds to the union of all K-dimensional subspaces in $\mathbb{R}^N$ spanned by all possible combinations of K basis vectors from Ψ. This is the low-dimensional, highly non-convex set depicted in Figure 1(a).

The *best K-sparse approximation* x_K to a signal x in the basis Ψ is obtained via projection onto Σ_K:

$$x_K := \operatorname*{argmin}_{x' \in \Sigma_K} \|x - x'\|_2 . \quad (4)$$

The corresponding coefficient vector $\alpha_K = \Psi^T x_K$ can be computed simply by thresholding the vector α, that is, by setting to zero all except the K largest entries.

Despite the large number $\binom{N}{K}$ of subspaces comprising Σ_K, since each subspace has K-dimensions, the overall model set is amenable to a very low-dimensional embedding. For example, for an $M \times N$ iid Gaussian Φ with $M \geq 2K$, with probability one $\Phi x_1 \neq \Phi x_2$ for all $x_1, x_2 \in \Sigma_K$ [10]. Indeed, since Φ is linear, each K-dimensional subspace from Σ_K is mapped to a unique K-dimensional hyperplane in $\mathbb{R}^M$ (see Figure 1(a)).

Stable embedding: For a stable embedding of the form (2), more than $2K$ measurements (but not many more) are required. It has been shown for a wide range of different random Φ and for any fixed Ψ that the set Σ_K will have a stable embedding in the measurement space $\mathbb{R}^M$ with high probability if $M = O(K \log(N/K))$ [8]. Note that the dimensionality of the measurement space is linear in the sparsity K but only logarithmic in the ambient dimensionality N. This bodes well for many applications involving high-dimensional (large N) but sparse (small K) data.

Stable recovery from compressive measurements: Compressive sensing (CS) is a new field that fuses data acquisition and data compression into a single dimensionality reduction step of the form (1) [11–13]. In the CS literature, the stable embedding property (2) is known as the *Restricted Isometry Property* (RIP) [11] of order $2K$.[2] The centerpiece of CS is the fundamental result that a K-sparse signal x can be stably recovered from the measurements $y = \Phi x$ by searching for the sparsest signal $\widehat{x}$ that agrees with those measurements.

There are many forms of the CS recovery problem; we will sketch just one of them here. Assume that Φ is a stable embedding for $2K$-sparse signals with $\epsilon < \sqrt{2} - 1$ in (2), and assume that the measurements y are corrupted by noise n with $\|n\|_2 \leq \delta$ for some δ (recall (1)). Let $\widehat{x} = \Psi \widehat{\alpha}$, where $\widehat{\alpha}$ is the solution to the following convex optimization problem

$$\widehat{\alpha} = \operatorname*{argmin}_{\alpha' \in \mathbb{R}^N} \|\alpha'\|_1 \text{ subject to } \|y - \Phi\Psi\alpha'\|_2 \leq \delta. \quad (5)$$

Then it has been shown [14] that

$$\|x - \widehat{x}\|_2 = \|\alpha - \widehat{\alpha}\|_2 \leq C_1 K^{-1/2} \|\alpha - \alpha_K\|_1 + C_2 \delta \quad (6)$$

for constants C_1 and C_2. The recovery algorithm (5) is a noise-aware variation on the ℓ_1-minimization procedure known as *basis pursuit*; similar recovery results have also been shown for iterative greedy recovery algorithms [15–17]. Bounds of this type are extremely encouraging for signal processing. From only $M = O(K \log(N/K))$ measurements, it is possible to recover a K-sparse x with zero error if there is no noise and with bounded error $C_2\delta$ if there is noise. In other words, despite the apparent ill-conditioning of the inverse problem, the noise is not dramatically amplified in the recovery process.

Note that optimization (5) searches not directly for the α' that is sparsest (and hence has smallest ℓ_0 quasi-norm) but rather for the α' that has smallest ℓ_1 norm. It has been shown that this relaxation of the highly non-convex ℓ_0 search into a convex ℓ_1 search does not affect the solution to (5) as long as Φ is a stable embedding.

Note also the use of the ℓ_1 norm on the right hand side of (6). Interestingly, it has been shown that replacing this ℓ_1 norm with the more intuitive ℓ_2 norm requires that M be increased from $M = O(K \log(N/K))$ to $M = O(N)$ regardless of K [18]. However, it is possible to obtain an ℓ_2-type bound for sparse signal recovery in the noise-free setting by changing from a deterministic formulation (where Φ satisfies the RIP) to a *probabilistic* one where x is fixed and Φ is randomized [18, 19].

B. Compressible deterministic signals

In practice, natural and manmade signals are not exactly sparse. Fortunately, in many cases they can be closely approximated as sparse. One particularly useful model is the set of *compressible* signals whose coefficients α in the basis Ψ have a rapid power-law decay when sorted:

$$|\alpha_{(i)}| \leq R i^{-1/p}, \quad p \leq 1 \quad (7)$$

where $\alpha_{(i)}$, $i = 1, 2, \ldots$ denote the coefficients of α sorted from largest to smallest. For a given $\alpha \in \mathbb{R}^N$, if we take the smallest R for which (7) is satisfied, then we say that α lives in the *weak ℓ_p ball*, $w\ell_p$, of radius R, and we define $\|\alpha\|_{w\ell_p} = R$.

[2]The CS literature uses $2K$ rather than K to describe the order of the RIP, because Φ acts as an approximate isometry on the difference signal $x_1 - x_2$, which is $2K$-sparse in general. To connect with our linear subspace example from Section I-A, a stable embedding of Σ_K implies that all sets of $2K$ columns from $\Phi\Psi$ are nearly orthonormal.

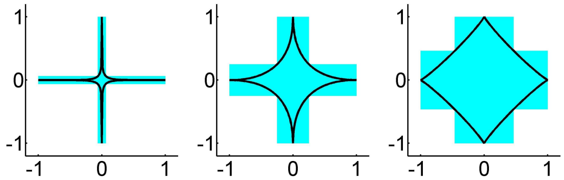

Fig. 2. *The weak ℓ_p balls (blue) and the standard ℓ_p balls (outlined in black) in $\mathbb{R}^2$ of radius $R = 1$ and, from left to right, $p = 0.25$, $p = 0.5$, and $p = 0.9$.*

Geometrical structure: Like the set of K-sparse signals Σ_K, $w\ell_p$ compressible signal models are highly non-convex. Figure 2 depicts three $w\ell_p$ balls in $\mathbb{R}^2$ for various values of p. (As demonstrated in the figure, for any N, the standard ℓ_p ball in $\mathbb{R}^N$ is strictly contained inside the $w\ell_p$ ball; for $N \to \infty$ the fit is very tight, since $\|\alpha\|_p \leq \|\alpha\|_{w\ell_p} \leq \|\alpha\|_{p'}$ for any $p' > p$ [20].) Moreover, the $w\ell_p$ balls are very close to Σ_K in the sense that any compressible signal x whose coefficients α obey (7) can be closely approximated by its best K-sparse approximation [20]:

$$\|\alpha - \alpha_K\|_q \leq RK^{1/q-1/p}, \tag{8}$$

where $q > p$.

Stable embedding: $w\ell_p$ balls occupy too much of $\mathbb{R}^N$ to qualify as low-dimensional and support a stable embedding for compressible signals of the form (2). However, since the $w\ell_p$ balls live near Σ_K, near stable embeddings exist in the following sense. Given two compressible signals x_1 and x_2 from a $w\ell_p$ ball, then while their projections Φx_1 and Φx_2 will not satisfy (2), the projections of their respective best K-sparse approximations $\Phi x_{1,K}$ and $\Phi x_{2,K}$ will satisfy (2).

Stable recovery from compressive measurements: Returning to (6), then, we see that the ℓ_2 error incurred in recovering a compressible signal x from the noisy measurements $y = \Phi x + n$ decays as $K^{1/2-1/p}$, which is the same rate as the best K-sparse approximation to x [21].

IV. Probabilistic Sparse Models

A. Bayesian signal recovery

In parallel with the methods of the previous section, a substantial literature has developed around probabilistic, Bayesian methods for recovering a signal x from its noisy measurements $y = \Phi x + n$ from (1) [22–25]. Under the Bayesian framework, we assume that x is a realization of a random process X with joint probability distribution function (pdf) $f(x)$, which we call the *prior*. The information about x that is carried in the measurements y is expressed using the conditional pdf $f(y|x)$; when the noise n is iid Gaussian with mean zero and variance τ, then we can easily compute that $f(y|x)$ is also iid Gaussian with mean Φx and the same variance. To estimate x, we use Bayes rule to form the *posterior* pdf $f(x|y) \propto f(y|x)f(x)$ and then compute some kind of summary. Here, we are particularly interested in $\arg\max_x f(x|y)$, which yields the *maximum a posteriori* or *MAP estimate* of x given y.

Developing $f(x|y)$ for the iid Gaussian noise case, we can write the MAP optimization as

$$\widehat{x}_{\text{MAP}} = \underset{x'}{\operatorname{argmin}} \|y - \Phi x'\|_2^2 - 2\tau \log f(x'). \tag{9}$$

Different choices of the prior $f(x)$ reflecting different knowledge about the signal yield different optimizations. We are interested here in pdfs whose realizations have a stable embedding into a lower dimensional space. In these cases, we can hope to be able to recover accurate signal estimates from noisy compressive measurements.

B. Sparse random signals

To keep our notation simple, in this section, we take Ψ to be the canonical basis; hence $x = \alpha$. The simplest sparse signal prior is an iid, two-state *mixture model* in which independently with probability K/N the signal coefficient x_i is distributed according to a distribution $f_1(x)$ and with probability $1-K/N$ the signal coefficient $x_i = 0$. The resulting prior pdf can be written as

$$f(x) = \frac{K}{N} f_1(x) + \left(1 - \frac{K}{N}\right) \delta(x), \tag{10}$$

where δ is the Dirac delta function.

Geometrical structure and stable embedding: Realizations of this mixture model will be K-sparse on average, and hence all of the theory of Section III-A applies. Namely, on average, realizations of the model can be stably embedded into $\mathbb{R}^M$ and then stably recovered as long as $M = O(K \log(N/K))$.

C. Compressible random signals

Of particular interest to the Bayesian community for dimensionality reduction problems has been the iid zero-mean generalized Gaussian distribution (GGD) prior

$$f_{\text{GGD}}(x) = \frac{q\, e^{-\left|\frac{x}{\sigma}\right|^q}}{2\Gamma\left(1/q\right)}, \tag{11}$$

where q is the order, σ is the shape parameter, and Γ is the gamma function. When $q = 1$, this is known as the *Laplacian* distribution. The reason for the lure is that

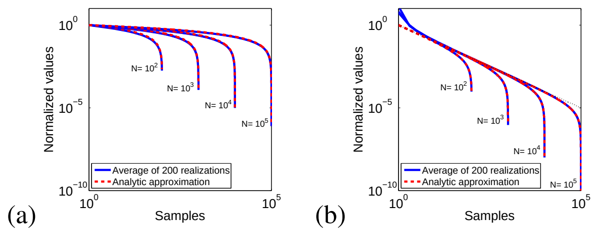

Fig. 3. *Numerical illustration of the $w\ell_p$ exponents of two different pdfs. We generated iid realizations of each distribution of various lengths N, sorted the values from largest to smallest, and then plotted the results on a log-log scale. To live in $w\ell_p$ with $0 < p \leq 1$, the slope of the resulting curve must be ≤ -1. We also overlay the analytical expressions for the distributions derived in Appendix A and B. (a) iid Laplacian distribution (generalized Gasusian distribution, GGD with $q = 1$) slope is $\gg -1$ and moreover grows with N. (b) In contrast, iid generalized Pareto distribution (GPD) with $q = 1$ exhibits a constant slope < -1 that is is independent of N.*

when (11) is substituted into the MAP equation (9), we stumble upon

$$\widehat{x}_{\text{MAP}} = \underset{x'}{\operatorname{argmin}} \|y - \Phi x'\|_2^2 + \lambda \|x'\|_q^q, \qquad (12)$$

with $\lambda = 2\tau\sigma^{-q}$. For $q = 1$ (Laplacian), this optimization is known widely as *basis pursuit denoising* (BPDN) [26] which is simply the scalarization of (5); a related formulation is known as the *LASSO* [27]. For $q < 1$ this corresponds to the non-convex recovery algorithm of [28]. Thus, we see a potentially striking correspondence between deterministic and probabilistic sparse models.

Unfortunately, there is an important and heretofore apparently unrecognized bug in this argument, because realizations of the iid GGD are in general *not* compressible. Indeed, as we calculate in Appendix A, with high probability, a GGD realization with parameter q lives in the $w\ell_p$ ball with $p = q \log \frac{N}{q}$. Recall from (7) that we need $p \leq 1$ for compressibility. Thus, no matter what GGD parameter q is specified, as the signal length N grows, eventually we will have $p > 1$ and will lose compressibility. Figure 3(a) illustrates via a numerical simulation both the non-compressibility of the Laplacian and the fit of our analytical expression derived in Appendix A. In this log-log plot, sorted realizations of a Laplacian distribution need to decay with a slope faster than -1, which is clearly not the case.

What is going on here? There is nothing particularly wicked with the GGD prior, since clearly the resulting MAP recovery optimization coincides with classical signal recovery tools like BPDN and the LASSO. What is wrong is that while it appears natural to use a GGD prior for signal recovery, we have no assurance that we can actually recover a signal generated from the prior density. Hence the GGD puts us in a Bayesian Catch-22 [29].

As an alternative to the GGD, we suggest the iid zero-mean *generalized Pareto distribution* (GPD):

$$f_{\text{GPD}}(x) = \frac{1}{2\sigma}\left(1 + \frac{|x|}{q\sigma}\right)^{-(q+1)} \qquad (13)$$

with order q and shape parameter σ. As we show analytically in Appendix B and numerically in Figure 3(b), realizations of the GPD do, in fact, live in $w\ell_p$ balls with $p \leq 1$ and hence are compressible. The compressibility of this pdf can also be used to prove the compressibility of the *Student's t-distribution* [29], which is widely employed in relevance vector machines [25].

Substituting the GPD prior into the MAP estimation formula (9) we obtain the non-convex optimization

$$\widehat{x}_{\text{MAP}} = \underset{x'}{\operatorname{argmin}} \|y - \Phi x'\|_2^2 + \lambda \log \prod_{i=1}^{N} \left(1 + \frac{|x_i'|}{q\sigma}\right) \qquad (14)$$

with $\lambda = 2\tau(1+q)$. Using the inequality of the arithmetic and geometric means, we can relax the GPD penalty term to

$$\begin{aligned} \log \prod_{i=1}^{N}\left(1 + \frac{|x_i'|}{q\sigma}\right) &\leq N \log\left(1 + \frac{\|x'\|_1}{Nq\sigma}\right) \\ &\leq (q\sigma)^{-1}\|x'\|_1, \end{aligned} \qquad (15)$$

which coincides with (12) with $p = 1$ (BPDN) and $\lambda = 2\tau\sigma^{-1}\left(1 + q^{-1}\right)$. To summarize, a GPD prior both generates compressible signal realizations *and* supports stable recovery of those signals from compressive measurements via the same convex linear program employed for deterministic signals.

V. Structured Sparse Models

The deterministic and probabilistic sparse and compressible signal models we have discussed above are powerful yet simplistic, because they do not aim to capture any of the inter-dependencies or correlations between the various signal coefficients. In contrast, state-of-the-art compression algorithms for signal, image, video, and other data (which have dimensionality reduction at their heart) significantly improve over more naïve methods by codifying the inter-dependency structure among the signal coefficients. In this section, we will review related recent progress on structured sparsity models.

A. Deterministic structured sparsity

Recall from Section III that K-sparse signals live in Σ_K, the union of $\binom{N}{K}$, K-dimensional subspaces in $\mathbb{R}^N$. But what if we know more about a signal than merely that it has K dominant coefficients? For example, what if we know that the significant coefficients of the signal occur in pairs? It is easy to see that this structure disqualifies many of the subspaces in Σ_K, since $\binom{N}{K/2} \ll \binom{N}{K}$, which significantly lowers the already low-dimensionality of the model (see Figure 1(b)). As a result, we should expect to be able to reduce the size M of the dimensionality reduction matrix Φ from (1) while still preserving its information preservation and stable embedding properties.

Geometrical structure: As an extension of K-sparse signals, we define a *reduced union-of-subspaces* (RUS) signal model as a subset of signals from Σ_K [30, 31] (see Figure 1(b)). We call signals from a RUS model *K-model sparse*.

There are two broad classes of RUS models. (*i*) In many applications, the significant coefficients *cluster*, meaning, loosely speaking, that the significant coefficients have adjacent indices. A general model is the (K, C)-sparse model, which constrains the K-sparse signal coefficients to be contained within at most C-clusters [32]. A specific instance of this model is *block sparsity*, which assumes prior knowledge of the locations and sizes of the clusters [33–36]. Clustered sparsity is a natural model for a variety of applications, including group sparse regression problems, DNA microarrays, MIMO channel equalization, source localization in sensor networks, and magnetoencephalography [10, 32–40]. A related clustered model has been developed for wavelet transforms of piecewise smooth signal and image data, which in addition to their sparse primary structure have a strong secondary structure where the significant wavelet coefficients tend to cluster along the branches of the wavelet tree [41–45]. (*ii*) In many other applications, the coefficients are *dispersed* in that each significant coefficient must be separated from all others by some number Δ of samples. Such a model has applications in neuroscience [46]. Still more general union-of-subspace models can also be learned from sampled data [47].

Stable embedding: Blumensath and Davies have quantified the number of measurements M necessary for a sub-Gaussian random matrix Φ to be a stable embedding in the sense of (2) for signals from a particular structured sparse model [30]. Applying their result to the above models, it can be shown that (K, C)-sparse signals require $M = O(K + C \log(N/C))$ measurements; tree-sparse signals require $M = O(K)$ measurements; and dispersed signal models require $M = O(K \log(N/K - \Delta))$ measurements. All of these are significantly smaller than the $M = O(K \log(N/K))$ required for the unstructured sparse signals of previous sections.

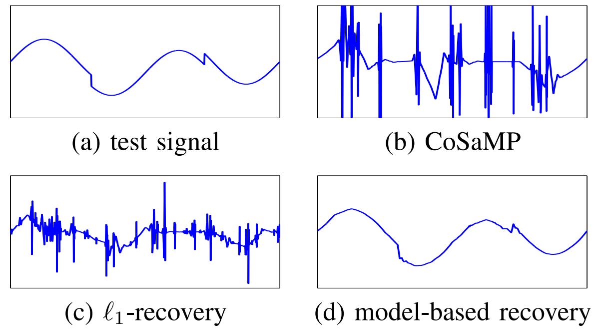

Fig. 4. *Example of model-based signal recovery. (a) Piecewise-smooth* HeaviSine *test signal of length $N = 1024$. This signal is compressible under a connected wavelet tree model. Signal recovered from $M = 80$ random Gaussian measurements using (b) the iterative recovery algorithm CoSaMP, (c) standard ℓ_1 basis pursuit, and (d) the wavelet tree-based CoSaMP algorithm.*

Stable recovery from compressive measurements: Signals from the structured sparse models described above can be recovered efficiently with provable performance guarantees using a variety of algorithms [33, 35, 36], including model-aware iterative greedy algorithms that simply replace the best K-sparse approximation (4) in each iteration with a best K-model sparse approximation [33]. Algorithms that have been adapted to date include CoSaMP [16] and iterative thresholding [17]. Figure 4 demonstrates the potential gains for the clustered tree model, for example.

B. Deterministic structured compressibility

Geometrical structure: Just as compressible signals are "nearly K-sparse" and thus live close to the union of subspaces Σ_K in $\mathbb{R}^N$, model-compressible signals are "nearly K-model sparse" and live (fairly) close to the restricted union of subspaces defined by the K-model sparse model. This concept can be made rigorous by defining a model-compressible signal as one that can be approximated by a K-model sparse signal with an approximation error that decays according to a power law as in (8). The only difference is that now α_K stands for the best K-model sparse approximation.

Stable embedding: A key result of Section III-B was that when Φ is a stable embedding in the sense of (2) for sparse signals, then it is a near stable embedding for compressible signals. Unfortunately, this is not the case for model-compressible signals: when Φ is a stable embedding for model sparse signals, then it is not necessarily a near stable embedding for model-compressible signals. The reason is that the set of model-compressible signals live fairly close, but not close enough to the set of model-sparse signals. A sufficient condition to guarantee that Φ is a near-stable embedding for model-compressible signals has been put forth in [33]; the new condition is called the *restricted amplification property* (RAmP). Fortunately the RAmP is not much more restrictive than the usual stable embedding condition (RIP); for the three models described in Section V-A (tree, (K, C) clustered, block sparse), the number of measurements M necessary for a sub-Gaussian random matrix Φ to have the RAmP is of the same order as the number necessary to have the RIP [33].

Stable recovery from compressive measurements: Given a Φ that satisfies both the RIP and the RAmP, signals from one of the models discussed above can be recovered from their compressive measurements by the same algorithms described in Section V-A with a performance guarantee like that in (6) (up to a log constant that depends on N [33]).

C. Probabilistic structured sparsity

Signal models that capture not just coefficient sparsity but also correlation structure have been widely developed and applied in Bayesian circles. Markov and hidden Markov models are ubiquitous in speech processing [48], image processing [49], and wavelet-domain processing [41–43]. However, these techniques have been applied only sparsely to dimensionality reduction and stable signal recovery, and few theoretical performance guarantees are available. We cite two examples. A variational Bayes recovery algorithm has been developed [50] for a tree-structured, conjugate-exponential family sparse model for wavelet coefficients. The classical Ising Markov random field model capturing how the significant image pixels cluster together has been applied to background subtracted image recovery [51].

D. Multi-signal sparse models

Thus far, our attention has focused on the sparsity structure of a single signal. However, many applications involve a plurality of J signals $x_1, \ldots, x_J \in \mathbb{R}^N$. When the signal ensemble shares a common sparsity structure, then we can consider joint dimensionality reduction and stable recovery schemes that require fewer than $O(J \cdot K \log(N/K))$ total measurements.

Geometrical structure: A number of sparsity models have been developed to capture the intra- and inter-signal correlations among signals in an ensemble [10, 34, 52–59]. For the sake of brevity, we consider just one here: the *joint sparsity model* that we dub JSM-2 [10, 34]. Under this model, the K locations of the significant coefficients are the same in each signal's coefficient vector α_j (there is no constraint on any of the coefficient values across the signals). Stacking the coefficient vectors as rows into a $J \times N$ matrix, we obtain a matrix containing only K nonzero columns; vectorizing this matrix column-by-column into the super vector $\underline{x}$, we obtain a special case of the block sparse signal model [33, 60] from Section V-A. The geometrical structure of multi-signal sparse models is thus closely related to that of single-signal structured sparse models.

Stable embedding: An added wrinkle that distinguishes signal ensembles from individual signals is that many applications dictate that we apply a separate dimensionality reduction Φ_j independently to each signal x_j to obtain the measurements $y_j = \Phi_j x_j$. Such a distributed approach to dimensionality reduction is natural for sensor network scenarios, where each signal x_j is acquired by a different geographically dispersed sensor. For the JSM-2 / block sparse model discussed above, applying Φ_j independently to each x_j is equivalent to multiplying the super-vector $\underline{x}$ by a block diagonal matrix $\underline{\Phi}$ composed of the individual Φ_j. In [10], we call this process *distributed compressive sensing* (DCS) and show that $\underline{\Phi}$ can be a very efficient stable embedding. By efficient we mean that for JSM-2 ensembles of K-sparse signals, $\underline{\Phi}$ can be as small as $JK \times JN$, which means that only $M_j = K$ measurements are needed per K-sparse signal. This is clearly best-possible, since at least K measurements are needed per signal to characterize the values of their respective significant coefficients.

Stable recovery from compressive measurements: Various algorithms have been proposed to jointly recover the ensemble of signals $\{x_j\}$ from the collection of independent measurements $\{y_j\}$, including convex relaxations [10, 52–56, 58, 60, 61], greedy algorithms [10, 34, 62], and statistical formulations [57, 59].

VI. Other Low-Dimensional Geometric Models

In this section we look beyond unions of subspaces and $w\ell_p$ balls to other low-dimensional geometric models that support similar stable embeddings.

A. Point clouds

A low-dimensional (actually zero-dimensional) set of great importance in statistics, pattern recognition, and learning theory is a *point cloud* consisting of a finite number Q of points at arbitrary positions in $\mathbb{R}^N$. In the context of this paper, each point corresponds to one of the Q signals in the set.

Stable embedding: The classical Johnson-Lindenstrauss (JL) lemma [6, 7] states that with high probability, a randomly generated $M \times N$ matrix Φ with $M = O(\log(Q)\epsilon^{-2})$ yields an ϵ-stable embedding of the point cloud. An illustration is shown in Figure 1(c).

The implications of the JL lemma are numerous. For example, by storing only a set of compressive measurements of a database of signals, we can reduce both the memory storage requirements and the computational complexity of solving problems such as nearest neighbor search [63]. In addition, the JL lemma can be used to prove that certain random probability distributions generate Φ with the stable embedding property [8].

B. Manifold models

As discussed in the introduction, nonlinear K-dimensional manifold models arise when a family of signals $\mathcal{M} = \{x_\theta \in \mathbb{R}^N : \theta \in \Theta\}$ is smoothly parameterized by a K-dimensional parameter vector θ. Manifolds are also suitable as approximate models for certain nonparametric signal classes such as images of handwritten digits [1].

Stable embedding: The theory of manifold embeddings in low-dimensional Euclidean space has a rich history. A proof of Whitney's embedding theorem [2], for example, demonstrates that with probability one, a random linear projection Φ with $M = 2K + 1$ will ensure that $\Phi x_1 \neq \Phi x_2$ for all $x_1, x_2 \in \mathcal{M}$. (Note the parallel with Section III, in which the same can be guaranteed for all $x_1, x_2 \in \Sigma_K$ if $M = 2K$.) However, this projection carries no guarantee of stability. Nash [3], in contrast, considered embeddings of generic manifolds not necessarily beginning in $\mathbb{R}^N$; his objective was to preserve only intrinsic metric structure (geodesic distances).

Building on the JL lemma, we have shown [64] that with high probability, a randomly generated $M \times N$ matrix Φ with $M = O(K \log(CN)\epsilon^{-2})$ yields an ϵ-stable embedding of a K-dimensional manifold $\mathcal{M}$. Note the strong parallel with all of the above models (M grows linearly in K and logarithmically in N), except that now K is the dimensionality of the manifold, not the signal sparsity. The constant C depends on factors such as the curvature and volume of the manifold; precise technical details are given in [64]. An illustration of this stable embedding is shown in Figure 1(d). As a corollary to this result, properties such as the dimension, topology, and geodesic distances along the manifold are also preserved exactly or approximately. Under slightly different assumptions on $\mathcal{M}$, a number of similar embedding results involving random projections have been established [65–67].

Stable recovery from compressive measurements: The stable embedding of manifold signal models tempts us to consider new kinds of signal recovery problems. For example, given $y = \Phi x + n$, suppose that the signal x belongs exactly or approximately to a manifold $\mathcal{M} \subset \mathbb{R}^N$. By positing the manifold model for x (instead of a sparse model like Σ_K), we can formulate a manifold-based CS recovery scheme. If we define x^* to be the optimal "nearest neighbor" to x on $\mathcal{M}$, that is,

$$x^* = \arg\min_{x' \in \mathcal{M}} \|x - x'\|_2 \,, \tag{16}$$

then we consider x to be well-modeled by $\mathcal{M}$ if $\|x - x^*\|_2$ is small. The hope is that such signals x can be recovered with high accuracy using an appropriate recovery scheme.

For example, we have shown that for any fixed $x \in \mathbb{R}^N$, then with high probability over Φ, letting

$$\widehat{x} = \arg\min_{x' \in \mathcal{M}} \|y - \Phi x'\|_2 \tag{17}$$

yields the recovery error

$$\|\widehat{x} - x\|_2 \leq (1+0.25\epsilon)\,\|x - x^*\|_2 + (2+0.32\epsilon)\,\|n\|_2 + C. \tag{18}$$

Precise technical details are given in [68]. Unfortunately, in contrast to the sparsity-based case, we cannot hope to solve a search such as (17) using a universal convex optimization; in practice one may resort to iterative algorithms such as Newton's method [69]. Nonetheless, we are able to provide a guarantee (18) that is highly analogous to (6) and other guarantees from the CS literature [19] in which the recovery error scales with the distance from the signal to its model. For sparsity-based models, this distance is measured between x and x_K; for manifold models it is measured between x and x^*.

Manifold-based CS also enjoys the same universality properties as sparsity-based CS, since the manifold $\mathcal{M}$ need not be known at the time of data acquisition. Consequently, we may also consider multi-class recognition problems [9], in which the signal is compared to a number of possible candidate manifold models.

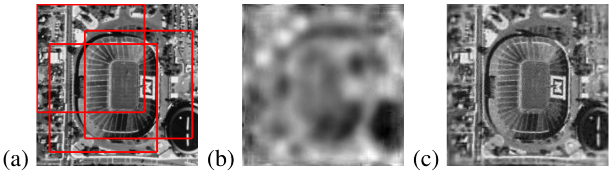

Fig. 5. *Manifold lifting example. (a) Compressive measurements of $J = 200$ overlapping sub-images were obtained from hypothetical satellites. From each image of size $N = 4096$, just $M = 96$ random measurements were obtained. The goal is to reconstruct a high-resolution fused image from the low-dimensional sub-image measurements. (b) Fusion of images recovered independently via sparsity-based CS reconstruction, PSNR=15.4dB. (c) Fusion of images via joint CS reconstruction that exploits a manifold model for the inter-signal structure, PSNR=23.8dB.*

Multi-signal recovery using manifold models: Extending the JSM/DCS framework of of Section V-D, we can hypothesize a manifold model for the inter-signal structure within a signal ensemble $x_1, \ldots, x_J$. For example, each of the J signals x_j might represent an image of a scene obtained from a different perspective. It follows that the signals $x_1, \ldots, x_J$ will live on a common (but unknown) K-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$, where K equals the number of degrees of freedom in the camera position. Supposing that all $\Phi_j = \Phi$ for some $M \times N$ matrix Φ, the measurement vectors $y_j = \Phi x_j$ will live along the image $\Phi\mathcal{M}$ of $\mathcal{M}$ in the measurement space $\mathbb{R}^M$. From these points, it is possible to estimate the underlying parameterization of the data using manifold learning algorithms such as ISOMAP [70] and to develop a "manifold lifting" recovery technique, in which each x_j is restored from $\mathbb{R}^M$ back to $\mathbb{R}^N$, while ensuring that the recovered signals live along a common, plausible manifold. See Figure 5 and [71] for a promising proof of concept.

Alternatively, one can simply stack the signal ensemble $x_1, \ldots, x_J$ into a single super-vector $\underline{x} \in \mathbb{R}^{JN}$. The resulting super-vector still lives on a K-dimensional manifold $\mathcal{M}$ that we call the *joint manifold* [72]. Thus, with high probability, a randomly generated $M \times JN$ matrix Φ with $M = O(K \log(CJN))$ yields a stable embedding of $\mathcal{M}$. This enables solutions to multi-sensor data processing problems where the complexity grows only logarithmically in both the ambient dimensionality N and the number of sensors J.

VII. Discussion and Conclusions

We conclude our tour of low-dimensional signal models for dimensionality reduction with a summary and some perspectives for the future. As we have seen, there are large classes of diverse signal models that support stable dimensionality reduction and thus can be used to effect data acquisition, analysis, or processing more efficiently. (Of course, we have also seen that certain random signal models that are commonly cited as producing low-dimensional, compressible signals—the iid GGD and Laplacian distributions—actually do not.) We expect that the diversity of the available models will inspire the development of new kinds of signal recovery algorithms that go beyond today's more agnostic sparsity-based approaches. In particular, we see significant promise in structured sparsity models (to leverage the decades of experience in data compression) and manifold models (to solve complex object recognition and other reasoning tasks). We also expect that new low-dimensional models will continue to be discovered, which will further broaden and deepen the impact of this rapidly developing field.

Appendix A
Non-Compressibility of the Generalized Gaussian Distribution

In this section, we show that signal realizations x whose entries are generated according to the iid GGD distribution (11), that is, $x_i \sim \text{GGD}(x_i; q, \sigma)$, are not compressible according to the definition in Section III-B.

Order statistics: Let $u_i = |x_i|$. From basic probability, the random variables (RV) u_i are also iid and have pdf $u_i \sim f(u)$ where $f(u) = \text{GGD}(u; q, \sigma) + \text{GGD}(-u; q, \sigma)$. Denote the cumulative distribution function (cdf) of $f(u)$ by $F(u)$.

Arranging the u_i in decreasing order of size

$$u_{(1)} \geq u_{(2)} \geq \ldots \geq u_{(N)}, \tag{19}$$

we call $u_{(i)}$ the i-th magnitude order statistic (OS) of x. Even though the RVs x_i are independent, the RVs $u_{(i)}$ are statistically dependent.

Quantile approximation to the OS: For relatively large sample sizes N, a well-known approximation to the expected OS of a RV is given by the quantile of its cdf [73]:

$$E[u_{(i)}] \approx F^{-1}\left(1 - \frac{i}{N+1}\right). \tag{20}$$

The variance of this approximation is quantified in [29].

Bounds on $F(u)$: The cdf $F(u)$ can be written as

$$F(u) = 1 - \frac{\Gamma\left(1/q, \left(\frac{u}{\sigma}\right)^q\right)}{\Gamma\left(1/q\right)}, \tag{21}$$

where $\Gamma\left(s, z\right) = \int_z^\infty t^{s-1}e^{-t}dt$ is the incomplete Gamma function. It is possible to bound (21) by [74]:

$$\left[1 - e^{-\beta\left(\frac{u}{\sigma}\right)^q}\right]^{1/q} \leq F(u) \leq \left[1 - e^{-\alpha\left(\frac{u}{\sigma}\right)^q}\right]^{1/q} \tag{22}$$

where $\alpha \geq \max\left\{1, [\Gamma(1+1/q)]^{-q}\right\}$ and $0 \leq \beta \leq \min\left\{1, [\Gamma(1+1/q)]^{-q}\right\}$.

Bounds and approximations of $F^{-1}(u)$: Using (22), we can bound the inverse of the cdf $F(u)$ as

$$\sigma\left[-\alpha^{-1}\log(1-u^q)\right]^{1/q} \leq F^{-1}(u) \leq \sigma\left[-\beta^{-1}\log(1-u^q)\right]^{1/q}. \tag{23}$$

Hence, $F^{-1}(u) = \Theta\left(\sigma\left[-\beta^{-1}\log(1-u^q)\right]^{1/q}\right)$.

It is then straightforward to derive the following approximations:

$$F^{-1}\left(1-\frac{i}{N+1}\right) \leq \sigma\left[-\beta^{-1}\log\left(\frac{iq}{N+1}\right)\right]^{1/q} \leq Ri^{-1/p}, \tag{24}$$

where

$$R = \sigma\left[-\beta^{-1}\log\left(\frac{q}{N+1}\right)\right]^{1/q}, \quad p = q\log\frac{N+1}{q}. \tag{25}$$

The non-compressibility then follows from (20).

Appendix B
Compressibility of the Generalized Pareto Distribution

For the GPD (13), our analysis parallels that of Appendix A except with a more positive conclusion. The key departure is that the magnitude order statistics of the iid GPD RVs can be approximated as

$$u_{(i)} = q\sigma\left[\left(\frac{i}{N+1}\right)^{-1/q} - 1\right] < Ri^{-1/p}i^{-1/q}, \tag{26}$$

where

$$R = q\sigma(N+1)^{1/q}, \qquad p = q. \tag{27}$$

Note that the GPD converges to the Laplacian distribution in the limit as $q \to \infty$.

References

[1] G. E. Hinton, P. Dayan, and M. Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Networks*, vol. 8, pp. 65–74, Jan. 1997.

[2] M. W. Hirsch, *Differential Topology*, vol. 33 of *Graduate Texts in Mathematics*. Springer, 1976.

[3] J. Nash, "The imbedding problem for Riemannian manifolds," *Annals of Mathematics, 2nd Series*, vol. 63, pp. 20–63, January 1956.

[4] B. Kashin, "The widths of certain finite-dimensional sets and classes of smooth functions," *Izv. Akad. Nauk SSSR Ser. Mat*, vol. 41, no. 2, pp. 334–351, 1977.

[5] A. Garnaev and E. Gluskin, "On the widths of the Euclidean ball," *Doklady, Akad. Nauk SSSR*, vol. 277, pp. 1048–1052, 1984.

[6] S. Dasgupta and A. Gupta, "An elementary proof of the Johnson-Lindenstrauss lemma," Tech. Rep. TR-99-006, Berkeley, CA, 1999.

[7] D. Achlioptas, "Database-friendly random projections," in *Proc. Symp. on Principles of Database Systems (PODS '01)*, pp. 274–281, ACM Press, 2001.

[8] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, pp. 253–263, December 2008.

[9] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. Computational Imaging V at SPIE Electronic Imaging*, January 2007.

[10] D. Baron, M. B. Wakin, M. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing." Rice University Technical Report TREE-0612, Nov 2006.

[11] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289–1306, Sept. 2006.

[13] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–120, 124, July 2007.

[14] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Series I*, vol. 346, pp. 589–592, May 2008.

[15] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," Dec. 2007. Preprint.

[16] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, June 2008. To be published.

[17] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," July 2008. Preprint.

[18] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," 2006. Preprint.

[19] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an ell-1 decoder," 2008. Preprint.

[20] R. A. DeVore, "Lecture notes on Compressed Sensing," *Rice University ELEC 631 Course Notes*, Spring 2006. Available at http://cnx.org/content/col10458/latest/.

[21] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[22] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[23] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.

[24] P. Schniter, L. C. Potter, and J. Ziniel, "Fast bayesian matching pursuit," in *Information Theory and Applications Workshop*, pp. 326–333, 2008.

[25] M. Tipping, "Sparse bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[26] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol. 20, p. 33, 1998.

[27] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, pp. 267–288, 1996.

[28] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *ICASSP 2008*, pp. 3885–3888, 2008.

[29] V. Cevher, "Learning with compressible priors," in *Neural Information Processing Systems (NIPS)*, (Vancouver, B.C., Canada), 7–12 December 2008.

[30] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Info. Theory*, Dec. 2008. To be published.

[31] Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Processing Mag.*, vol. 25, pp. 41–47, Mar. 2008.

[32] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," in *Sampling Theory and Applications (SAMPTA)*, (Marseilles, France), May 2009.

[33] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," 2008. Preprint. Available at http://dsp.rice.edu/cs.

[34] J. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, Apr. 2006.

[35] Y. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," 2008. Preprint.

[36] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," Mar. 2008. Preprint.

[37] Y. Kim, J. Kim, and Y. Kim, "Blockwise sparse regression," *Statistica Sinica*, vol. 16, no. 2, p. 375, 2006.

[38] L. Meier, S. van de Geer, and P. Buhlmann, "The group lasso for logistic regression," *Journal of Royal Stat. Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 53–71, 2008.

[39] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering Sparse Signals Using Sparse Measurement Matrices in Compressed DNA Microarrays," *IEEE Journal of Selected Topics in Sig. Proc.*, vol. 2, no. 3, pp. 275–285, 2008.

[40] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroenceph. and Clin. Neurophys.*, vol. 95, no. 4, pp. 231–251, 1995.

[41] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov Models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, Apr. 1998.

[42] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain Hidden Markov Models," *IEEE Trans. Image Processing*, vol. 10, pp. 1056–1068, July 2001.

[43] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, Nov. 2003.

[44] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[45] A. Cohen, W. Dahmen, I. Daubechies, and R. A. DeVore, "Tree approximation and optimal encoding," *Applied and Computational Harmonic Analysis*, vol. 11, pp. 192–226, Sept. 2001.

[46] C. Hegde, M. F. Duarte, and V. Cevher, "Compressive sensing recovery of spike trains using a structured sparsity model," in *Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, (Saint-Malo, France), 13–16 April 2009.

[47] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Review*, vol. 50, no. 3, pp. 413–458, 2008.

[48] L. Rabiner *et al.*, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[49] R. Chellappa and A. Jain, "Markov random fields: theory and application," *Academic Press*, 1993.

[50] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," 2008. Preprint. Available at http://people.ee.duke.edu/ lcarin/Papers.html.

[51] V. Cevher, M. F. Duarte, C. Hegde, and R. G. Baraniuk, "Sparse signal recovery using markov random fields," in *Neural Information Processing Systems (NIPS)*, (Vancouver, B.C., Canada), 8–11 December 2008.

[52] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, pp. 589–602, Mar. 2006.

[53] P. Nagesh and B. Li, "Compressive imaging of color images," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, (Taipei, Taiwan), April 2009.

[54] H. Jung, K. Sung, K. Nayak, E. Kim, and J. Ye, "k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, 2009.

[55] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proc. Picture Coding Symposium (PCS)*, (Chicago, Illinois), May 2009.

[56] R. Marcia and R. Willet, "Compressive coded aperture video reconstruction," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2008.

[57] C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic MR image reconstruction using Kalman filtered compressed sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, (Taipei, Taiwan), April 2009.

[58] D. Angelosante and G. B. Giannakis, "RLS-weighted LASSO for Adaptive Estimation of Sparse Signals," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, (Taipei, Taiwan), April 2009.

[59] S. Ji, D. Dunson, and L. Carin, "Multi-task compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.

[60] Y. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," 2008. preprint.

[61] M. Fornasier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM J. Numer. Anal*, vol. 46, no. 2, pp. 577–613, 2008.

[62] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 655–687, 2008.

[63] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimenstionality," in *Proc. Symp. Theory of Computing*, pp. 604–613, 1998.

[64] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, pp. 51–77, February 2009.

[65] P. Indyk and A. Naor, "Nearest neighbor preserving embeddings," *ACM Trans. Algorithms*, 2006. To appear.

[66] P. Agarwal, S. Har-Peled, and H. Yu, "Embeddings of surfaces, curves, and moving points in euclidean space," in *Proceedings of the twenty-third annual symposium on Computational geometry*, pp. 381–389, ACM New York, NY, USA, 2007.

[67] S. Dasgupta and Y. Freund, "Random projection trees and low dimensional manifolds," in *Proceedings of the 40th annual ACM symposium on Theory of computing*, pp. 537–546, ACM New York, NY, USA, 2008.

[68] M. B. Wakin, "Manifold-based signal recovery and parameter estimation from compressive measurements," 2008. Preprint.

[69] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. Wavelets XI at SPIE Optics and Photonics*, (San Diego, California), August 2005.

[70] C. Hegde, M. Wakin, and R. Baraniuk, "Random projections

for manifold learning," in *Proc. Neural Information Processing Systems (NIPS)*, December 2007.

[71] M. B. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," in *Proc. Picture Coding Symposium (PCS 2009)*, (Chicago, Illinois), May 2009.

[72] M. Davenport, C. Hegde, M. Duarte, and R. Baraniuk, "A Theoretical Analysis of Joint Manifolds," *Arxiv preprint arXiv:0901.0760*, 2009.

[73] H. David and H. Nagaraja, *Order statistics*. Wiley-Interscience, 2004.

[74] H. Alzer, "On some inequalities for the incomplete gamma function," *Mathematics of Computation*, vol. 66, no. 218, pp. 771–778, 1997.