Combining geometry and combinatorics: a unified approach to sparse signal recovery

R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss

***Abstract*— There are two main algorithmic approaches to sparse signal recovery: geometric and combinatorial. The geometric approach utilizes geometric properties of the measurement matrix Φ. A notable example is the Restricted Isometry Property, which states that the mapping Φ preserves the Euclidean norm of sparse signals; it is known that random dense matrices satisfy this constraint with high probability. On the other hand, the combinatorial approach utilizes sparse matrices, interpreted as adjacency matrices of sparse (possibly random) graphs, and uses combinatorial techniques to recover an approximation to the signal.**

In this paper we present a unification of these two approaches. To this end, we extend the notion of Restricted Isometry Property from the Euclidean ℓ_2 norm to the Manhattan ℓ_1 norm. Then we show that this new ℓ_1-based property is essentially equivalent to the combinatorial notion of *expansion* of the sparse graph underlying the measurement matrix. At the same time we show that the new property suffices to guarantee correctness of both geometric and combinatorial recovery algorithms.

As a result, we obtain new measurement matrix constructions and algorithms for signal recovery which, compared to previous algorithms, are superior in either the number of measurements or computational efficiency of decoders.

I. Introduction

With the rise in high-speed data transmission and the exponential increase in data storage, it is imperative that we develop effective data compression techniques, techniques which accomodate both the volume and speed of data streams. A new approach to compressing n-dimensional vectors (or signals) begins with linear observations or measurements. For a signal x, its compressed representation is equal to $\mathbf{\Phi}x$, where $\mathbf{\Phi}$ is a carefully chosen $m \times n$ matrix, $m \ll n$, often chosen at random from some distribution. We call the vector $\mathbf{\Phi}x$ the *measurement vector* or a *sketch* of x. Although the dimension of $\mathbf{\Phi}x$ is much smaller than that of x, it retains many of the essential properties of x.

There are several reasons why linear compression or sketching is of interest. First, we can easily maintain a linear sketch $\mathbf{\Phi}x$ under linear updates to the signal x. For example, after incrementing the i-th coordinate x_i, we simply update the sketch as $\mathbf{\Phi}(x+e_i) = \mathbf{\Phi}x+\mathbf{\Phi}e_i$. Similarly, we also easily obtain a sketch of a sum of two signals given the sketches for individual signals x and y, since $\mathbf{\Phi}(x+y) = \mathbf{\Phi}x + \mathbf{\Phi}y$. Both properties are useful in several computational areas, notably computing over data streams [AMS99], [Mut03], [Ind07], network measurement [EV03], query optimization and answering in databases [AMS99].

Another scenario where linear compression is of key importance is *compressed sensing* [CRT06], [Don06a], a rapidly developing area in digital signal processing. In this setting, x is a physical signal one wishes to sense (e.g., an image obtained from a digital camera) and the linearity of the observations stems from a physical observation process. Rather than first observing a signal in its entirety and then compressing it, it may be less costly to sense the compressed version directly via a physical process. A camera "senses" the vector by computing a dot product with a number of pre-specified measurement vectors. See [TLW+06], [DDT+08] for a prototype camera built using this framework. Other applications of linear sketching include database privacy [DMT07].

Although the sketch is considerably smaller than the original vector, we can still recover a large amount of information about x. See the surveys [Mut03], [Ind07] on streaming and sublinear algorithms for a broad overview of the area. In this paper, we focus on retrieving a *sparse approximation* x_* of x. A vector is called *k-sparse* if it has at most k non-zero elements in the canonical basis (or, more generally, k non-zero coefficients in some basis B). The goal of the sparse approximation is to find a vector x_* such that the ℓ_p *approximation error* $\|x - x_*\|_p$ is at most $C > 0$ times the smallest possible ℓ_q approximation error $\|x - x'\|_q$, where x' ranges over all k-sparse vectors. Note that the error $\|x - x'\|_q$ is minimized when x' consists of the k largest (in magnitude) coefficients of x.

There are many algorithms for recovering sparse approximations (or their variants) of signals from their sketches. The early work on this topic includes the *algebraic* approach of [Man92](cf. [GGI+02a]). Most of the known algorithms,

Berinde is with the Department of Electrical Engineering and Computer Science, MIT. Email: texel@mit.edu

Gilbert is with the Department of Mathematics, The University of Michigan at Ann Arbor. Email: annacg@umich.edu

Indyk is with the Computer Science and Artificial Intelligence Laboratory, MIT. Email: indyk@theory.lcs.mit.edu

Karloff is with AT&T Labs - Research. Email: howard@research.att.com

Strauss is with the Department of Mathematics and the Department of Electrical Engineering and Computer Science, The University of Michigan at Ann Arbor. Email: martinjs@umich.edu

ACG is an Alfred P. Sloan Research Fellow and has been supported in part by NSF DMS 0354600. MJS has been supported in part by NSF DMS 0354600 and NSF DMS 0510203. ACG and MJS have been partially supported by DARPA ONR N66001-06-1-2011. RB and PI are supported in part by David and Lucille Packard Fellowship. PI is partially supported by MADALGO (Center for Massive Data Algorithmics, funded by the Danish National Research Association) and NSF grant CCF-0728645.

however, can be roughly classified as either *combinatorial* or *geometric*.

Combinatorial approach. In the combinatorial approach, the measurement matrix $\mathbf{\Phi}$ is sparse and often binary. Typically, it is obtained from an adjacency matrix of a sparse bipartite random graph. The recovery algorithm proceeds by iteratively identifying and eliminating "large" coefficients[1] of the vector x. The identification uses non-adaptive binary search techniques. Examples of combinatorial sketching and recovery algorithms include [GGI+02b], [CCFC02], [CM04], [GKMS03], [DWB05], [SBB06b], [SBB06a], [CM06], [GSTV06], [GSTV07], [Ind08], [XH07] and others.

The typical advantages of the combinatorial approach include fast recovery (often sub-linear in the signal length n if $k \ll n$), as well as fast and incremental (under coordinate updates) computation of the sketch vector $\mathbf{\Phi}x$. In addition, it is possible to construct efficient (albeit suboptimal) measurement matrices *explicitly*, at least for simple type of signals. For example, it is known [Ind08], [XH07] how to explicitly construct matrices with $k2^{(\log\log n)^{O(1)}}$ measurements, for signals x that are exactly k-sparse. The main disadvantage of the approach is the suboptimal sketch length.

Geometric approach. This approach was first proposed in the papers [CRT06], [Don06a] and has been extensively investigated since then (see [Gro06] for a bibliography). In this setting, the matrix $\mathbf{\Phi}$ is dense, with at least a constant fraction of non-zero entries. Typically, each row of the matrix is independently selected from a sub-exponential n-dimensional distribution, such as Gaussian or Bernoulli. The key property of the matrix $\mathbf{\Phi}$ which yields efficient recovery algorithms is the *Restricted Isometry Property* [CRT06], which requires that for any k-sparse vector x we have $\|\mathbf{\Phi}x\|_2 = (1 \pm \delta)\|x\|_2$. If a matrix $\mathbf{\Phi}$ satisfies this property, then the recovery process can be accomplished by finding a vector x_* using the following linear program:

$$\min \ \|x_*\|_1 \text{ subject to } \mathbf{\Phi}x_* = \mathbf{\Phi}x. \qquad \text{(P1)}$$

The advantages of the geometric approach include a small number of measurements ($O(k\log(n/2k))$ for Gaussian matrices and $O(k\log^{O(1)} n)$ for Fourier matrices) and resiliency to measurement errors[2]. The main disadvantage is the running time of the recovery procedure, which involves solving a linear program with n variables and $n+m$ constraints. The computation of the sketch $\mathbf{\Phi}x$ can be done efficiently for some matrices (e.g., Fourier); however, an efficient sketch update is not possible. In addition, the problem of finding an explicit construction of efficient matrices satisfying the RIP property is open [Tao07]; the best known explicit construction [DeV07] yields $\Omega(k^2)$ measurements.

Connections. There has been some recent progress in obtaining the advantages of both approaches by decoupling the algorithmic and combinatorial aspects of the problem. Specifically, the papers [NV07], [DM08], [NT08] show that one can use *greedy* methods for data compressed using *dense* matrices satisfying the RIP property. Similarly [GLR08], using the results of [KT07], show that sketches from (somewhat) sparse matrices can be recovered using linear programming.

The best results (up to $O(\cdot)$ constants) obtained prior to this work are shown in Figure 1[3]. We ignore some aspects of the algorithms, such as explicitness or universality of the measurement matrices. Furthermore, we present only the algorithms that work for arbitrary vectors x, while many other results are known for the case where the vector x itself is exactly k-sparse; e.g., see [TG05], [DWB05], [SBB06b], [Don06a], [XH07]. The columns describe:

- citation,
- whether the recovery algorithms hold with high probability for All signals or for Each signal,
- sketch length,
- time to compute $\mathbf{\Phi}x$ given x,
- time to update $\mathbf{\Phi}x$ after incrementing one of the coordinates of x,
- time[4] to recover an approximation of x given $\mathbf{\Phi}x$,
- approximation guarantee, and
- whether the algorithm is robust to noisy measurements.

In the approximation error column, $\ell_p \leq C\ell_q$ means that the algorithm returns a vector x_* such that $\|x - x_*\|_p \leq C \min_{x'} \|x - x'\|_q$, where x' ranges over all k-sparse vectors. In [CDD06], the authors show that an approximation guarantee of the form "$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$" implies a "$\ell_1 \leq (1+O(C))\ell_1$" guarantee, and that it is impossible to achieve "$\ell_2 \leq C\ell_2$" deterministically (or for all signals simultaneously) unless the number of measurements is $\Omega(n)$. The parameters $C > 1$, $c \geq 2$ and $a > 0$ denote absolute constants, possibly different in each row. We assume that $k < n/2$.

In addition, in Figure 2 we present very recent results discovered during the course of our research. Some of the running times of the algorithms depend on the "precision parameter" D, which is always bounded from the above by

[1] In the non-sketching world, such methods algorithms are often called "weak greedy algorithms", and have been studied thoroughly by Temlyakov [Tem02]

[2] Historically, the geometric approach resulted also in the first *deterministic* or *uniform* recovery algorithms, where a fixed matrix $\mathbf{\Phi}$ was guaranteed to work for *all* signals x. In contrast, the early combinatorial sketching algorithms only guaranteed $1 - 1/n$ probability of correctness for *each* signal x. However, the papers [GSTV06], [GSTV07] showed that combinatorial algorithms can achieve deterministic or uniform guarantees as well.

[3] Some of the papers, notably [CM04], are focused on a somewhat different formulation of the problem. However, it is known that the guarantees presented in the table hold for those algorithms as well. See Lecture 4 in [Ind07] for a more detailed discussion.

[4] In the decoding time column LP=LP(n, m, T) denotes the time needed to solve a linear program defined by an $m \times n$ matrix $\mathbf{\Phi}$ which supports matrix-vector multiplication in time T. Heuristic arguments indicate that LP$(n, m, T) \approx \sqrt{n}T$ if the interior-point method is employed. In addition, the paper [NV07] does not discuss the running time of the algorithm. Our bound is obtained by multiplying the number of algorithm iterations (i.e., k) by the number of entries in the matrix $\mathbf{\Phi}$ (i.e., $nk\log^c n$). See [NT08] for an in-depth discussion of the running times of OMP-based procedures.

Paper	A/E	Sketch length	Encode time	Column sparsity/ Update time	Decode time	Approx. error	Noise
[CCFC02], [CM06]	E	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^c n$	$\ell_2 \leq C\ell_2$	
	E	$k \log n$	$n \log n$	$\log n$	$n \log n$	$\ell_2 \leq C\ell_2$	
[CM04]	E	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^c n$	$\ell_1 \leq C\ell_1$	
	E	$k \log n$	$n \log n$	$\log n$	$n \log n$	$\ell_1 \leq C\ell_1$	
[CRT06]	A	$k \log(n/k)$	$nk \log(n/k)$	$k \log(n/k)$	LP	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	Y
	A	$k \log^c n$	$n \log n$	$k \log^c n$	LP	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	Y
[GSTV06]	A	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^c n$	$\ell_1 \leq C \log n \ell_1$	Y
[GSTV07]	A	$k \log^c n$	$n \log^c n$	$\log^c n$	$k^2 \log^c n$	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	
[NV07]	A	$k \log(n/k)$	$nk \log(n/k)$	$k \log(n/k)$	$nk^2 \log^c n$	$\ell_2 \leq \frac{C(\log n)^{1/2}}{k^{1/2}}\ell_1$	Y
	A	$k \log^c n$	$n \log n$	$k \log^c n$	$nk^2 \log^c n$	$\ell_2 \leq \frac{C(\log n)^{1/2}}{k^{1/2}}\ell_1$	Y
[GLR08] (k "large")	A	$k(\log n)^{c \log\log\log n}$	kn^{1-a}	n^{1-a}	LP	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	
This paper	A	$k \log(n/k)$	$n \log(n/k)$	$\log(n/k)$	LP	$\ell_1 \leq C\ell_1$	Y

Fig. 1. Summary of the best prior results.

Paper	A/E	Sketch length	Encode time	Update time	Decode time	Approx. error	Noise
[DM08]	A	$k \log(n/k)$	$nk \log(n/k)$	$k \log(n/k)$	$nk \log(n/k) \log D$	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	Y
[NT08]	A	$k \log(n/k)$	$nk \log(n/k)$	$k \log(n/k)$	$nk \log(n/k) \log D$	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	Y
	A	$k \log^c n$	$n \log n$	$k \log^c n$	$n \log n \log D$	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	Y
[IR08]	A	$k \log(n/k)$	$n \log(n/k)$	$\log(n/k)$	$n \log(n/k)$	$\ell_1 \leq C\ell_1$	Y

Fig. 2. Recent work.

$\|x\|_2$ if the coordinates of x are integers.

A. Our results

In this paper we give a sequence of results which indicate that the combinatorial and geometric approaches are, in a rigorous sense, different manifestations of a common underlying phenomenon. This enables us to achieve a unifying perspective on both approaches, as well as obtaining several new concrete algorithmic results.

We consider matrices which are *binary* and *sparse*; i.e., they have only a small number d of ones in each column, and all the other entries are equal to zero. It has been shown recently [Cha08] that such matrices cannot satisfy the RIP property with parameters k and δ, unless the number of rows is $\Omega(k^2)$. Our first result is that, nevertheless, such matrices satisfy a different form of the RIP property, that we call the *RIP-p property*, where the ℓ_2 norm is replaced by the ℓ_p norm. Formally, the matrix $\mathbf{\Phi}$ satisfies $\mathrm{RIP}_{p,k,\delta}$ property if for any k-sparse vector x we have $\|\mathbf{\Phi}x\|_p = (1 \pm \delta)\|x\|_p$. In particular, we show that this property holds for $1 \leq p \leq 1 + O(1)/\log n$ if the matrix $\mathbf{\Phi}$ is an adjacency matrix of a high-quality *unbalanced expander graph*. Thus we have a large class of natural such measurement matrices. We also exhibit an RIP-2 matrix that is not an RIP-1 matrix, so that, with [Cha08], we conclude that these two conditions are incomparable—neither one is stronger than the other.

Theorem 1: Consider any $m \times n$ matrix $\mathbf{\Phi}$ that is the adjacency matrix of an (k, ϵ)-unbalanced expander $G = (A, B, E)$, $|A| = n$, $|B| = m$, with left degree d, such that $1/\epsilon, d$ are smaller than n. Then the scaled matrix $\mathbf{\Phi}/d^{1/p}$ satisfies the $\mathrm{RIP}_{p,k,\delta}$ property, for $1 \leq p \leq 1 + 1/\log n$ and $\delta = C\epsilon$ for some absolute constant $C > 1$.

The fact that the unbalanced expanders yield matrices with RIP-p property is not an accident. In particular, we show in Section II that any binary matrix $\mathbf{\Phi}$ in which each column has d ones[5] and which satisfies RIP-1 property with proper parameters, must be an adjacency matrix of a good unbalanced expander. That is, an RIP-p matrix and the adjacency matrix of an unbalanced expander are essentially equivalent. Therefore, RIP-1 provides an interesting "analytic" formulation of expansion for unbalanced graphs. Also, without significantly improved explicit constructions

[5]In fact, the latter assumption can be removed without loss of generality. The reason is that, from the RIP-1 property alone, it follows that each column must have roughly the same number of ones. The slight unbalance in the number of ones does not affect our results much; however, it does complicate the notation somewhat. As a result, we decided to keep this assumption throughout the paper.

of unbalanced expanders with parameters that match the probabilistic bounds (a longstanding open problem), we do not expect significant improvements in the explicit constructions of RIP-1 matrices.

Theorem 2: Consider any $m \times n$ binary matrix $\mathbf{\Phi}$ such that each column has exactly d ones. If for some scaling factor $S > 0$ the matrix $S\mathbf{\Phi}$ satisfies the $\mathrm{RIP}_{1,k,\delta}$ property, then the matrix $\mathbf{\Phi}$ is an adjacency matrix of an (k, ϵ)-unbalanced expander, for

$$\epsilon = \left(1 - \frac{1}{1+\delta}\right)/(2 - \sqrt{2}).$$

In the next step in Section III, we show that the RIP-1 property of a binary matrix (or, equivalently, the expansion property) alone suffices to guarantee that the linear program P_1 recovers a good sparse approximation. In particular, we show the following

Theorem 3: Let $\mathbf{\Phi}$ be an $m \times n$ matrix of an unbalanced $(2k, \epsilon)$-expander. Let $\alpha(\epsilon) = (2\epsilon)/(1 - 2\epsilon)$. Consider any two vectors x, x_*, such that $\mathbf{\Phi}x = \mathbf{\Phi}x_*$, and $\|x_*\|_1 \leq \|x\|_1$. Then

$$\|x - x_*\|_1 \leq 2/(1 - 2\alpha(\epsilon)) \cdot \|x - x_k\|_1.$$

where x_k is the optimal k-term representation for x.
We also provide a noise-resilient version of the theorem; see Section III for details.

By combining Theorem 3 with the best known probabilistic constructions of expanders (Section II) we obtain a scheme for sparse approximation recovery with parameters as in Figure 1. The scheme achieves the best known bounds for several parameters: the scheme is deterministic (one matrix works for all vectors x), the number of measurements is $O(k \log(n/k))$, the update time is $O(\log(n/k))$ and the encoding time is $O(n \log(n/k))$. In particular, this provides the first known scheme which achieves the best known measurement and encoding time bounds *simultaneously*. In contrast, the Gaussian and Fourier matrices are known to achieve only one optimal bound at a time. The fast encoding time also speeds up the LP decoding, given that the linear program is typically solved using the interior-point method, which repeatedly performs matrix-vector multiplications. In addition to theoretical guarantees, random sparse matrices offer an attractive empirical performance. We show in Section IV that the empirical behavior of binary sparse matrices with LP decoding is consistent with the analytic performance of Gaussian random matrices.

In the full version of this paper [BGI+08], we show that adjacency matrices of unbalanced expanders can be augmented to facilitate sub-linear time combinatorial recovery. This fact has been implicit in the earlier work [GSTV07], [Ind08]; we verify that indeed the expansion property is the sufficient condition guaranteeing correctness of those algorithms. As a result, we obtain an explicit construction of matrices with $O(k2^{(\log \log n)^{O(1)}})$ rows that are amenable to a sublinear decoding algorithm for all vectors (similar to that in [GSTV07]). Previous explicit constructions for sublinear time algorithms either had $\Omega(k^2)$ rows [CM06] or had $O(k2^{(\log \log n)^{O(1)}})$ rows [Ind08], [XH07] but were restricted to k-sparse signals or their slight generalizations. An additional (and somewhat unexpected) consequence is that the algorithm of [Ind08] is simple, effectively mimicking the well-known "parallel bit-flip" algorithm for decoding low-density parity-check codes.

Theorem 4: Let $\epsilon > 0$ be a fixed constant, and $p = 1 + 1/\log n$. Consider $x \in \mathbb{R}^n$ and a sparsity parameter k. There is a measurement matrix $\mathbf{\Psi}$, which we can construct explicitly or randomly, and an algorithm $\mathrm{HHS}(p)$ that, given measurements $v = \mathbf{\Psi}x$ of x, returns an approximation $\widehat{x}$ of x with $O(k/\epsilon)$ nonzero entries. The approximation satisfies

$$\|x - \widehat{x}\|_p \leq \epsilon k^{1/p-1} \|x - x_k\|_1.$$

where x_k is the optimal k-term representation for x. Let R denote the size of the measurements for either an explicit or random construction. Then the $\mathrm{HHS}(p)$ algorithm runs in time $\mathrm{poly}(R)$.

Figure 3 summarizes the connections among all of our results. We show the relationship between the combinatorial and geometric approaches to sparse signal recovery

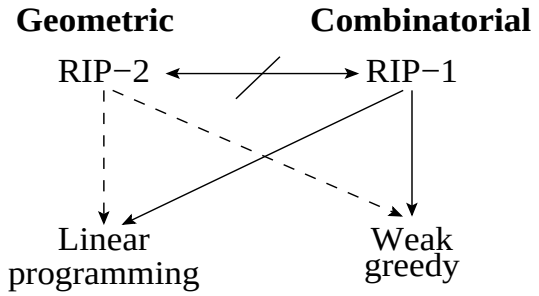

Fig. 3. The above diagram captures the relations between the combinatorial and geometric approaches and the two main classes of decoding algorithms. Connections established in prior work are shown with dashed lines. Our work connects both approaches, with the ultimate goal of obtaining the "best of both worlds."

II. Unbalanced expanders and RIP matrices

A. Unbalanced expanders

In this section we show that RIP-p matrices for $p \approx 1$ can be constructed using high-quality expanders. The formal definition of the latter is as follows.

Definition 5: A (k, ϵ)*-unbalanced expander* is a bipartite simple graph $G = (A, B, E)$ with left degree d such that for any $X \subset A$ with $|X| \leq k$, the set of neighbors $N(X)$ of X has size $|N(X)| \geq (1 - \epsilon)d|X|$.

In constructing such graphs, our goal is to make $|B|$, d, and ϵ as small as possible, while making k as close to $|B|$ as possible.

The following well-known proposition can be shown using the probabilistic method.

Proposition 6: For any $n/2 \geq k \geq 1$, $\epsilon > 0$, there exists a (k, ϵ)-unbalanced expander with left degree $d = O(\log(n/k)/\epsilon$ and right set size $O(kd/\epsilon) = O(k \log(n/k)/\epsilon^2)$.

Proposition 7: For any $n \geq k \geq 1$ and $\epsilon > 0$, one can explicitly construct a (k, ϵ)-unbalanced expander with left degree $d = 2^{O(\log(\log(n)/\epsilon)))^3}$, left set size n and right set size $m = kd/\epsilon^{O(1)}$.

Proof: The construction is given in [CRVW02], Theorem 7.3. Note that the theorem refers to notion of *lossless conductors*, which is equivalent to unbalanced expanders, modulo representing all relevant parameters (set sizes, degree, etc.) in the log-scale. After an additional $O(nd)$-time postprocessing, we can ensure that the graph is simple; i.e., it contains no duplicate edges. ■

B. Construction of RIP matrices

Definition 8: An $m \times n$ matrix $\mathbf{\Phi}$ is said to satisfy $\mathrm{RIP}_{p,k,\delta}$ if, for any k-sparse vector x, we have

$$\|x\|_p \leq \|\mathbf{\Phi}x\|_p \leq (1+\delta)\|x\|_p$$

Observe that the definitions of $\mathrm{RIP}_{1,k,\delta}$ and $\mathrm{RIP}_{2,k,\delta}$ matrices are incomparable. In what follows below, we present sparse binary matrices with $O(k \log(n/k))$ rows that are $\mathrm{RIP}_{1,k,\delta}$; it has been shown recently [Cha08] that sparse binary matrices cannot be $\mathrm{RIP}_{2,k,\delta}$ unless the number of rows is $\Omega(k^2)$. In the other direction, consider an appropriately scaled random Gaussian matrix G of $R \approx k \log(n)$ rows. Such a matrix is known to be $\mathrm{RIP}_{2,k,\delta}$. To see that this matrix is *not* $\mathrm{RIP}_{1,k,\delta}$, consider the signal x consisting of all zeros except a single 1 and the signal y consisting of all zeros except k terms with coefficient $1/k$. Then $\|x\|_1 = \|y\|_1$ but $\|Gx\|_1 \approx \sqrt{k}\|Gy\|_1$.

Theorem 1 *Consider any $m \times n$ matrix $\mathbf{\Phi}$ that is the adjacency matrix of an (k, ϵ)-unbalanced expander $G = (A, B, E)$ with left degree d, such that $1/\epsilon, d$ are smaller than n. Then the scaled matrix $\mathbf{\Phi}/d^{1/p}$ satisfies the $\mathrm{RIP}_{p,k,\delta}$ property, for $1 \leq p \leq 1 + 1/\log n$ and $\delta = C\epsilon$ for some absolute constant $C > 1$.*

Proof: Let $x \in \mathbb{R}^n$ be a k-sparse vector. Without loss of generality, we assume that the coordinates of x are ordered such that $|x_1| \geq \ldots \geq |x_n|$.

The proof proceeds in two stages. In the first part, we show that the theorem holds for the case of $p = 1$. In the second part, we extend the theorem to the case where p is slightly larger than 1.

The case of $p = 1$. We order the edges $e_t = (i_t, j_t)$, $t = 1 \ldots dn$ of G in a lexicographic manner. It is helpful to imagine that the edges $e_1, e_2 \ldots$ of E are being added to the (initially empty) graph. An edge $e_t = (i_t, j_t)$ causes a *collision* if there exists an earlier edge $e_s = (i_s, j_s), s < t$, such that $j_t = j_s$. We define E' to be the set of edges which do *not* cause collisions, and $E'' = E - E'$.

Lemma 9: We have

$$\sum_{(i,j) \in E''} |x_i| \leq \epsilon d \|x\|_1$$

Proof: For each $t = 1 \ldots dn$, we use an indicator variable $r_t \in \{0, 1\}$, such that $r_t = 1$ iff $e_t \in E''$. Define a vector $z \in \mathbb{R}^{dn}$ such that $z_t = |x_{i_t}|$. Observe that

$$\sum_{(i,j) \in E''} |x_i| = \sum_{e_t = (i_t, j_t) \in E} r_t |x_{i_t}| = r \cdot z$$

To upper bound the latter quantity, observe that the vectors satisfy the following constraints:

- The vector z is non-negative.
- The coordinates of z are monotonically non-increasing.
- For each *prefix set* $P_i = \{1 \ldots di\}$, $i \leq k$, we have $\|r_{|P_i}\|_1 \leq \epsilon di$ - this follows from the expansion properties of the graph G.
- $r_{|P_1} = 0$, since the graph is simple.

It is now immediate that for any r, z satisfying the above constraints, we have $r \cdot z \leq \|z\|_1 \epsilon$. Since $\|z\|_1 = d\|x\|_1$, the lemma follows. ■

Lemma 9 immediately implies that $\|\mathbf{\Phi}x\|_1 \geq d\|x\|_1 (1 - 2\epsilon)$. Since for any x we have $\|\mathbf{\Phi}x\|_1 \leq d\|x\|_1$, it follows that $\mathbf{\Phi}/d$ satisfies the $\mathrm{RIP}_{1,k,2\epsilon}$ property.

The case of $p \leq 1 + 1/\log n$. See the full version of this paper [BGI+08].

■

The above theorem shows that one can construct RIP-p matrices for $p \approx 1$ from good unbalanced expanders. In following, we show that this is not an accident: any binary matrix Φ in which satisfies RIP-1 property with proper parameters, and with each column having exactly d ones, must be an adjacency matrix of a good unbalanced expander. This reason behind this is that if some set of vertices does *not* expand too well, then there are many collisions between the edges going out of that set. If the signs of the coordinates "following" those edges are different, the coordinates will cancel each other out, and thus the ℓ_1 norm of a vector will not be preserved.

The assumption that each column has exactly d ones is not crucial, since the RIP-1 property itself implies that the number of ones in each column can differ by at most factor of $1 + \delta$. All proofs in this paper are resilient to this slight unbalance. However, we decided to keep this assumption for the ease of notation.

Theorem 2 *Consider any $m \times n$ binary matrix $\mathbf{\Phi}$ such that each column has exactly d ones. If for some scaling factor $S > 0$ the matrix $S\mathbf{\Phi}$ satisfies the $\mathrm{RIP}_{1,s,\delta}$ property, then the matrix $\mathbf{\Phi}$ is an adjacency matrix of an (s, ϵ)-unbalanced expander, for*

$$\epsilon = \left(1 - \frac{1}{1+\delta}\right) / \left(2 - \sqrt{2}\right).$$

Note that for small values of $\delta > 0$, we have

$$\left(1-\frac{1}{1+\delta}\right)/(2-\sqrt{2}) \approx \delta/(2-\sqrt{2})$$

Proof: Let $G=(A,B,E)$ be the graph with adjacency matrix $\mathbf{\Phi}$. Assume that there exists $X \subset A$, $|X|=k' \leq k$ such that $|N(X)| < dk'(1-\epsilon)$. We will construct two n-dimensional vectors y,z such that $\|y\|_1 = \|z\|_1 = k'$, but $\|\mathbf{\Phi}z\|_1 / \|\mathbf{\Phi}y\|_1 \leq 1-\epsilon(2-\sqrt{2})$, which is a contradiction.

The vector y is simply the characteristic vector of the set X. Clearly, we have $\|y\|_1 = k'$ and $\|\mathbf{\Phi}y\|_1 = dk$.

The vector z is defined via a random process. For $i \in X$, define r_i to be i.i.d. random variables uniformly distributed over $\{-1,1\}$. We define $z_i = r_i$ if $i \in X$, and $z_i = 0$ otherwise. Note that $\|z\|_1 = \|y\|_1 = k'$.

Let $C \subset N(X)$ be the "collision set", i.e., the set of all $j \in N(X)$ such that the number u_j of the edges from j to X is at least 2. Let $|C|=l$. By the definition of the set C we have $\sum_j u_j \geq 2l$. Moreover, from the assumption it follows that $\sum_j u_j \geq 2\epsilon dk'$.

Let $v = \mathbf{\Phi}z$. We split v into v_C and v_{C^c}. Clearly, $\|v_{C^c}\|_1 = k'd - \sum_j u_j$. It suffices to show that $\|v_C\|_1$ is significantly smaller than $\sum_j u_j$ for *some* z.

Claim 10: The expected value of $\|v_C\|_2^2$ is equal to $\sum_j u_j$.

Proof: For each $j \in C$, the coordinate v_j is a sum of u_j independent random variables uniformly distributed over $\{-1,1\}$. The claim follows by elementary analysis. ■

By Claim 10 we know that there *exists* z such that $\|v_C\|_2 \leq \sqrt{\sum_j u_j} \leq \frac{\sum_j u_j}{\sqrt{2l}}$. This implies that $\|v_C\|_1 \leq \sqrt{l}\,\|v_C\|_2 \leq \frac{\sum_j u_j}{\sqrt{2}}$. Therefore

$$\begin{aligned}
\|v\|_1 &\leq \|v_C\|_1 + \|v_{C^c}\|_1 \\
&\leq \frac{\sum_j u_j}{\sqrt{2}} + dk' - \sum_j u_j \\
&= dk' - (1-1/\sqrt{2})\sum_j u_j \\
&\leq dk' - (1-1/\sqrt{2})\cdot 2\epsilon dk' \\
&= dk'[1-\epsilon(2-\sqrt{2})]
\end{aligned}$$

■

III. LP DECODING

In this section we show that if A is an adjacency matrix of an expander graph, then the LP decoding procedure can be used for recovering sparse approximations.

Let Φ be an $m \times n$ adjacency matrix of an unbalanced $(2k,\epsilon)$-expander G with left degree d. Let $\alpha(\epsilon) = (2\epsilon)/(1-2\epsilon)$. We also define $E(X:Y) = E \cap (X \times Y)$ to be the set of edges between the sets X and Y.

A. L1 Uncertainty Principle

In this section we show that any vector from the kernel of a an adjacency matrix $\mathbf{\Phi}$ of an expander graph is "smooth"; i.e., the ℓ_1 norm of the vector cannot be concentrated on a small subset of its coordinates. An analogous result for RIP-2 matrices and with respect to the ℓ_2 norm has been used before (e.g., in [KT07]) to show guarantees for LP-based recovery procedures.

Lemma 11: Consider any $y \in \mathbb{R}^n$ such that $\mathbf{\Phi}y = 0$, and let S be any set of k coordinates of y. Then we have

$$\|y_S\|_1 \leq \alpha(\epsilon)\|y\|_1.$$

Proof: Without loss of generality, we can assume that S consists of the largest (in magnitude) coefficients of y. We partition coordinates into sets $S_0, S_1, S_2, \ldots S_t$, such that (i) the coordinates in the set S_l are not larger (in magnitude) than the coordinates in the set S_{l-1}, $l \geq 1$, and (ii) all sets but S_t have size k. Therefore, $S_0 = S$. Let $\mathbf{\Phi}'$ be a submatrix of $\mathbf{\Phi}$ containing rows from $N(S)$.

From the equivalence of expansion and RIP-1 property we know that $\|\mathbf{\Phi}'y_S\|_1 = \|\mathbf{\Phi}y_S\|_1 \geq d(1-2\epsilon)\|y_S\|_1$. At the same time, we know that $\|\mathbf{\Phi}'y\|_1 = 0$. Therefore

$$\begin{aligned}
0 &= \|\mathbf{\Phi}'y\|_1 \\
&\geq \|\mathbf{\Phi}'y_S\|_1 - \sum_{l\geq 1} \sum_{(i,j)\in E, i\in S_l, j\in N(S)} |y_i| \\
&\geq d(1-2\epsilon)\|y_S\|_1 - \sum_{l\geq 1} |E(S_l : N(S))| \min_{i\in S_{l-1}} |y_i| \\
&\geq d(1-2\epsilon)\|y_S\|_1 - \frac{1}{k}\sum_{l\geq 1} |E(S_l : N(S))| \cdot \|y_{S_{l-1}}\|_1
\end{aligned}$$

From the expansion properties of G it follows that, for $l \geq 1$, we have $|N(S \cup S_l)| \geq d(1-\epsilon)|S \cup S_l|$. It follows that at most $d\epsilon 2k$ edges can cross from S_l to $N(S)$, and therefore

$$\begin{aligned}
0 &\geq d(1-2\epsilon)\|y_S\|_1 - \frac{1}{k}\sum_{l\geq 1} |E(S_l : N(S))| \cdot \|y_{S_{l-1}}\|_1 \\
&\geq d(1-2\epsilon)\|y_S\|_1 - d\epsilon 2 \sum_{l\geq 1} \|y_{S_{l-1}}\|_1/k \\
&\geq d(1-2\epsilon)\|y_S\|_1 - 2d\epsilon\|y\|_1
\end{aligned}$$

It follows that $d(1-2\epsilon)\|y_S\|_1 \leq 2d\epsilon\|y\|_1$, and thus $\|y_S\|_1 \leq (2\epsilon)/(1-2\epsilon)\|y\|_1$. ■

B. LP recovery

The following theorem provides recovery guarantees for the program P_1, by setting $u=x$ and $v=x_*$.

Theorem 3 *Consider any two vectors u,v, such that for $y = v-u$ we have $\mathbf{\Phi}y = 0$, and $\|v\|_1 \leq \|u\|_1$. Let S be the set of k largest (in magnitude) coefficients of u, then*

$$\|v-u\|_1 \leq 2/(1-2\alpha(\epsilon)) \cdot \|u-u_S\|_1$$

Proof: We have

$$\begin{aligned}
\|u\|_1 \geq \|v\|_1 &= \|(u+y)_S\|_1 + \|(u+y)_{S^c}\|_1 \\
&\geq \|u_S\|_1 - \|y_S\|_1 + \|y_{S^c}\|_1 - \|u_{S^c}\|_1 \\
&= \|u\|_1 - 2\|u_{S^c}\|_1 + \|y\|_1 - 2\|y_S\|_1 \\
&\geq \|u\|_1 - 2\|u_{S^c}\|_1 + (1-2\alpha(\epsilon))\|y\|_1
\end{aligned}$$

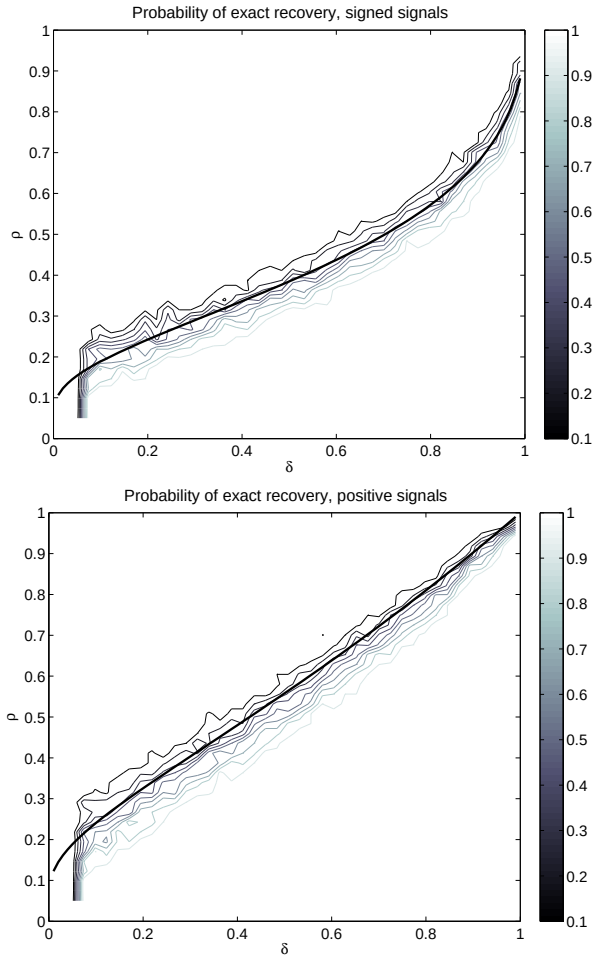

Fig. 4. Probability of correct signal recovery of a random k-sparse signal $x \in \{-1, 0, 1\}^n$ (left) and $x \in \{0, 1\}^n$ (right) as a function of $k = \rho m$ and $m = \delta n$, for $n = 200$. The probabilities were estimated by partitioning the domain into 40×40 data points and performing 50 independent trials for each data point, using random sparse matrices with $d = 8$. The thick curve demarcates a phase transition in the ability of LP decoding to find the sparsest solution to $Gx_* = Gx$ for G a Gaussian random matrix (see [DT06]). The empirical behavior for binary sparse matrices is consistent with the analytic behavior for Gaussian random matrices.

where we used Lemma 11 in the last line. It follows that

$$2\|u_{S^c}\|_1 \geq (1 - 2\alpha(\epsilon))\|y\|_1$$

■

Theorem 12: Consider any two vectors u, v, such that for $y = v - u$ we have $\|\Phi y\|_1 = \beta \geq 0$, and $\|v\|_1 \leq \|u\|_1$. Let S be the set of k largest (in magnitude) coefficients of u. Then

$$\|v - u_S\|_1 \leq 2/(1 - 2\alpha(\epsilon)) \cdot \|u_{S^c}\|_1 + \frac{2\beta}{d(1 - 2\epsilon)(1 - 2\alpha)}$$

Proof: Analogous to the proof of Theorem 3. ■

IV. Experimental Results

Our theoretical analysis shows that, up to constant factors, our scheme achieves the best known bounds for sparse approximate recovery. In order to determine the exact values of those constant factors, we show in Figure 4 the empirical probability of correct recovery of a random k-sparse signal of dimension $n = 200$ as a function of $k = \rho m$ and $m = \delta n$. As one can verify, the empirical $O(\cdot)$ constants involved are quite low. The thick curve shows the analytic computation of the phase transition between the survival of typical l-faces of the cross-polytope (left) and the polytope (right) under projection by G a Gaussian random matrix. This line is equivalent to a phase transition in the ability of LP decoding to find the sparsest solution to $Gx_* = Gx$, and, in effect, is representative of the performance of Gaussian matrices in this framework (see [Don06b] and [DT06] for more details). Gaussian measurement matrices with $m = \delta n$ rows and n columns can recover signals with sparsity $k = \rho m$ below the thick curve and cannot recover signals with sparsity k above the curve. This figure thus shows that the empirical behavior of binary sparse matrices with LP decoding is consistent with the analytic performance of Gaussian random matrices. Furthermore, the empirical values of the asymptotic constants seem to agree. See [BI08] for further experimental data.

V. Conclusion

We show in this paper that the geometric and the combinatorial approaches to sparse signal recovery are different manifestations of a common underyling phenomenon. Thus, we are able to show a unified perspective on both approaches—the key unifiying elements are the adjacency matrices of unbalanced expanders.

In most of the recent applications of *compressed sensing*, a physical device instantiates the measurement of x and, as such, these applications need measurement matrices which are conducive to physical measurement processes. This paper shows that there is another, quite different, large, natural class of measurement matrices, combined with the same (or similar) recovery algorithms for sparse signal approximation. These measurement matrices may or may not be conducive to physical measurement processes but they are quite amenable to computational or digital signal measurement. Our work suggests a number of applications in digital or computational "sensing" such as efficient numerical linear algebra and network coding.

The preliminary experimental analysis exhibits interesting high-dimensional geometric phenomena as well. Our results suggest that the projection of polytopes under Gaussian random matrices is similar to that of projection by sparse random matrices, despite the fact that Gaussian random matrices are quite different from sparse ones.

Acknowledgments: The authors would like to thank: Venkat Guruswami and Salil Vadhan, for their help on the expanders front; David Donoho and Jared Tanner for providing the data for the analytic Gaussian treshold curve in Figure 4; Justin Romberg for his help and clarifications regarding the ℓ_1-MAGIC package; and Tasos Sidiropoulos

for many helpful comments.

REFERENCES

[AMS99] N. Alon, Y. Matias, and M. Szegedy. The Space Complexity of Approximating the Frequency Moments. *J. Comput. System Sci.*, 58(1):137–147, 1999.

[BGI+08] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss. Combining geometry and combinatorics: a unified approach to sparse signal recovery. *arXiv:0804.4666*, 2008.

[BI08] R. Berinde and P. Indyk. Sparse recovery using sparse random matrices. *MIT-CSAIL Technical Report*, 2008.

[CCFC02] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. *ICALP*, 2002.

[CDD06] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Preprint*, 2006.

[Cha08] V. Chandar. A negative result concerning explicit matrices with the restricted isometry property. *Preprint*, 2008.

[CM04] G. Cormode and S. Muthukrishnan. Improved data stream summaries: The count-min sketch and its applications. *FSTTCS*, 2004.

[CM06] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for Compressed Sensing. In *Proc. 40th Ann. Conf. Information Sciences and Systems*, Princeton, Mar. 2006.

[CRT06] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1208–1223, 2006.

[CRVW02] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. *STOC-CCC*, 2002.

[DDT+08] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 2008.

[DeV07] R. DeVore. Deterministic constructions of compressed sensing matrices. *preprint*, 2007.

[DM08] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. *Arxiv:0803.0811*, 2008.

[DMT07] C. Dwork, F. McSherry, and K. Talwar. The price of privacy and the limits of lp decoding. *STOC*, 2007.

[Don06a] D. L. Donoho. Compressed Sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, Apr. 2006.

[Don06b] David L. Donoho. High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension. *Disc. Comput. Geometry*, 35(4):617–652, 2006.

[DT06] D. L. Donoho and J. Tanner. Thresholds for the recovery of sparse solutions via l1 minimization. *Proc. of the 40th Annual Conference on Information Sciences and Systems (CISS)*, 2006. To appear.

[DWB05] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk. Fast reconstruction of piecewise smooth signals from random projections. In *Proc. SPARS05*, Rennes, France, Nov. 2005.

[EV03] C. Estan and G. Varghese. New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice. *ACM Transactions on Computer Systems*, 2003.

[GGI+02a] A. Gilbert, S. Guha, P. Indyk, M. Muthukrishnan, and M. Strauss. Near-optimal sparse fourier representations via sampling. *STOC*, 2002.

[GGI+02b] A. C. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. J. Strauss. Fast, small-space algorithms for approximate histogram maintenance. In *ACM Symposium on Theoretical Computer Science*, 2002.

[GKMS03] A. C. Gilbert, Y. Kotidis, S. Muthukrishnan, and M. Strauss. One-Pass Wavelet Decompositions of Data Streams. *IEEE Trans. Knowl. Data Eng.*, 15(3):541–554, 2003.

[GLR08] V. Guruswami, J. Lee, and A. Razborov. Almost euclidean subspaces of l1 via expander codes. *SODA*, 2008.

[Gro06] Rice DSP Group. Compressed sensing resources. *Available at* `http://www.dsp.ece.rice.edu/cs/`, 2006.

[GSTV06] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin. Algorithmic linear dimension reduction in the ℓ_1 norm for sparse vectors. Submitted for publication, 2006.

[GSTV07] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin. One sketch for all: fast algorithms for compressed sensing. In *ACM STOC 2007*, pages 237–246, 2007.

[Ind07] P. Indyk. Sketching, streaming and sublinear-space algorithms. *Graduate course notes, available at* `http://stellar.mit.edu/S/course/6/fa07/6.895/`, 2007.

[Ind08] P. Indyk. Explicit constructions for compressed sensing of sparse signals. *SODA*, 2008.

[IR08] P. Indyk and M. Ruzic. Fast and effective sparse recovery using sparse random matrices. *Preprint*, 2008.

[KT07] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. *Preprint*, 2007.

[Man92] Y. Mansour. Randomized interpolation and approximation of sparse polynomials. *ICALP*, 1992.

[Mut03] S. Muthukrishnan. Data streams: Algorithms and applications (invited talk at soda'03). *Available at http://athos.rutgers.edu/∼muthu/stream-1-1.ps*, 2003.

[NT08] D. Needell and J. A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Arxiv math.NA 0803.2392*, 2008.

[NV07] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. 2007.

[SBB06a] S. Sarvotham, D. Baron, and R. G. Baraniuk. Compressed sensing reconstruction via belief propagation. *Technical Report ECE-0601, Electrical and Computer Engineering Department, Rice University*, 2006.

[SBB06b] S. Sarvotham, D. Baron, and R. G. Baraniuk. Sudocodes - fast measurement and reconstruction of sparse signals. *IEEE International Symposium on Information Theory*, 2006.

[Tao07] T. Tao. Open question: deterministic uup matrices. *Weblog at* `http://terrytao.wordpress.com`, 2007.

[Tem02] V. Temlyakov. Nonlinear methods of approximation. *Foundations of Comput. Math.*, July 2002.

[TG05] J. A. Tropp and A. C. Gilbert. Signal recovery from partial information via Orthogonal Matching Pursuit. Submitted to *IEEE Trans. Inform. Theory*, April 2005.

[TLW+06] Dharmpal Takhar, Jason Laska, Michael B. Wakin, Marco F. Duarte, Dror Baron, Shriram Sarvotham, Kevin Kelly, and Richard G. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Proc. IS&T/SPIE Symposium on Electronic Imaging*, 2006.

[XH07] W. Xu and B. Hassibi. Efficient compressive sensing with determinstic guarantees using expander graphs. *IEEE Information Theory Workshop*, 2007.