Irish Math. Soc. Bulletin
Number 81, Summer 2018, 5–22
ISSN 0791-5578

Sparsification of Matrices and Compressed Sensing

FINTAN HEGARTY, PADRAIG Ó CATHÁIN AND YUNBIN ZHAO

Abstract. Compressed sensing is a signal processing technique whereby the limits imposed by the Shannon–Nyquist theorem can be exceeded provided certain conditions are imposed on the signal. Such conditions occur in many real-world scenarios, and compressed sensing has emerging applications in medical imaging, big data, and statistics. Finding practical matrix constructions and computationally efficient recovery algorithms for compressed sensing is an area of intense research interest. Many probabilistic matrix constructions have been proposed, and it is now well known that matrices with entries drawn from a suitable probability distribution are essentially optimal for compressed sensing.

Potential applications have motivated the search for constructions of sparse compressed sensing matrices (i.e., matrices containing few non-zero entries). Various constructions have been proposed, and simulations suggest that their performance is comparable to that of dense matrices. In this paper, extensive simulations are presented which suggest that sparsification leads to a marked improvement in compressed sensing performance for a large class of matrix constructions and for many different recovery algorithms.

1. Introduction

Compressed sensing is a new paradigm in signal processing, developed in a series of ground-breaking publications by Donoho, Candès, Romberg, Tao and their collaborators over the past ten years or so [14, 8, 9]. Many real-world signals have the special property of being *sparse* — they can be stored much more concisely than a random signal. Instead of sampling the whole signal and then applying data compression algorithms, sampling and compression of sparse signals can be achieved simultaneously. This process requires dramatically

2010 *Mathematics Subject Classification.* 94A12, 90C05.
Key words and phrases. compressed sensing, optimisation, sparsification.
Received on 0-0-0000; revised 0-0-0000.

fewer measurements than the number dictated by the Shannon–Nyquist Theorem, but requires complex measurements which are *incoherent* with respect to the signal. The compressed sensing paradigm has generated an explosion of interest over the past few years within both the mathematical and electrical engineering research communities.

A particularly significant application has been to Magnetic Resonance Imaging (MRI), for which compressed sensing can speed up scans by a factor of five [23], either allowing increased resolution from a given number of samples or allowing real-time imaging at clinically useful resolutions. A major breakthrough achieved with compressed sensing has been real-time imaging of the heart [35, 24]. The US National Institute for Biomedical Imaging and Bioengineering published a news report in September 2014 describing compressed sensing as offering a "vast improvement" in paediatric MRI imaging [19]. Emerging applications of compressed sensing in data mining and computer vision were described by Candès in a plenary lecture at the 2014 International Congress of Mathematicians [7].

The central problems in compressed sensing can be framed in terms of linear algebra. In this model, a signal is a vector v in some high-dimensional vector space, $\mathbb{R}^N$. The sampling process can be described as multiplication by a specially chosen $n \times N$ matrix Φ, called the *sensing matrix*. Typically we will have $n \ll N$, so that the problem of recovering v from Φv is massively under-determined.

A vector is *k-sparse* if it has at most k non-zero entries. The set of k-sparse vectors in $\mathbb{R}^N$ plays the role of the set of compressible signals in a communication system. The problem now is to find necessary and sufficient conditions so that the inverse problem of finding v given Φ and Φv is efficiently solvable.

If u and v are distinct k-sparse vectors for which $\Phi u = \Phi v$, then one of them is not recoverable. Clearly, therefore, we require that the images of all k-sparse vectors under Φ are distinct, which is equivalent to requiring that the null-space of Φ does not contain any $2k$-sparse vectors. There is no known polynomial time algorithm to certify this property. We refer to the problem of finding the sparsest solution $\hat{x}$ to the linear system $\Phi\hat{x} = \Phi x$ as the *sparse recovery problem*. Natarajan has shown that certain instances of this problem are NP-hard [27].

Compressed sensing (CS) can be regarded as the study of methods for solving the sparse recovery problem and its generalizations (e.g., sparse approximations of non-sparse signals, solutions in the presence of noise) in a computationally efficient way. Most results in CS can be characterized either as certifications that the sparse recovery problem is solvable for a restricted class of matrices, or as the development of efficient computational methods for sparse recovery for some given class of matrices.

One of the most important early developments in CS was a series of results of Candès, Romberg, Tao and their collaborators. They established fundamental constraints for sparse recovery: one cannot hope to recover k-sparse signals of length N in less than $O(k \log N)$ measurements under any circumstances[1]. (For $k = 1$, standard results from complexity theory show that $O(\log N)$ measurements are required.) The main tools used to prove this result are the *restricted isometry parameters* (RIP), which measure how the sensing matrix Φ distorts the ℓ_2-norm of sparse vectors. Specifically, Φ has the RIP(k, ϵ) property if, for every k-sparse vector v, the following inequalities hold:

$$(1 - \epsilon)|v|_2^2 \leq |\Phi v|_2^2 \leq (1 + \epsilon)|v|_2^2.$$

Tools from Random Matrix Theory allow precise estimations of the RIP parameters of certain random matrices. In particular, it can be shown that the *random Gaussian ensemble*, which has entries drawn from a standard normal distribution, is asymptotically optimal for compressed sensing, i.e., the number of measurements required is $O(k \log N)$. A slightly weaker result is known for the *random Fourier ensemble*, a random selection of rows from the discrete Fourier transform matrix [8, 9, 33].

As well as providing examples of asymptotically optimal compressed sensing matrices, Candès et al. provided an efficient recovery algorithm: they showed that, under modest additional assumptions on the RIP parameters of a matrix, ℓ_1-minimization can be used

[1]Throughout this paper we use some standard notation for asymptotics: for functions $f, g : \mathbb{N} \to \mathbb{N}$, we write that $f = O(g)$ if there exists a constant C, not depending on n, such that $f(n) < Cg(n)$ for all sufficiently large n. Intuitively, the function g eventually dominates f, up to a constant factor. We say that $f = \Theta(g)$ if there exist two constants c, C such that $cg(n) < f(n) < Cg(n)$ for all sufficiently large n. Hence, f and g grow at the same rate.

for signal recovery. Thus efficient signal recovery is possible in large systems, making applications to real-world problems feasible.

Generation and storage of random matrices are potential obstacles to implementations of CS. It is also difficult to design efficient signal recovery algorithms capable of exploiting the structure of a random matrix. For implementation in real-world systems, it is desirable that CS constructions produce matrices that are *sparse* (possess relatively few non-zero entries), *structured*, and *deterministically constructed*. Systems with these properties can be stored implicitly, and efficient recovery algorithms can be designed to take advantage of their known structure. If Φ is $n \times N$ with d non-zero entries per column, then computing Φv takes $\Theta(dN)$ operations, which is a significant saving when $d \ll n$. In some applications, signals are frequently subject to rank-one updates (i.e., v is replaced by $v+\alpha e_i$), in which case the image vector can be updated in time $O(d)$, see [38].

Motivated by real-world applications, a number of papers have explored CS constructions where the Gaussian ensemble is replaced by a sparse random matrix (e.g., coming from an expander graph or an LDPC code) [3, 2, 18], or by a matrix obtained from a deterministic construction [11, 15, 16]. But to date, constructions meeting all three criteria have either been asymptotic in nature (i.e., the results only produce matrices that are too large for practical implementations), or are known only to exist for a very restricted range of parameters. This investigation was inspired by work of the second author on constructions of sparse CS matrices from pairwise balanced designs and complex Hadamard matrices [6, 5]. Some related work on constructing CS matrices from finite geometry is contained in [20, 37].

In this paper we take a new approach. Rather than constructing a sparse matrix and examining its CS properties, we begin with a matrix which is known to possess good CS properties (with high probability) and explore the effect of sparsification on this matrix. That is, we set many of the entries in the original matrix to zero, and compare the performance of the sparse matrix with the original. Results of Guo, Baron and Shamai suggest that sparse matrices should behave similarly to dense matrices in our regime [17]. Surprisingly, we actually observe an **improvement** in signal recovery as the sparsity increases.

First we survey some previous work on sparse compressed sensing matrices. Then in Section 3, we give a formal definition of sparsification, and describe algorithms used to generate random matrices and random vectors, as well as the recovery algorithms. In Section 4 we describe the results of extensive simulations. These provide substantial computational evidence which suggests that sparsification is a robust phenomenon, providing benefits in both recovery time and proportion of successful recoveries for a wide range of random and structured matrices occurring in the CS literature. In particular, Table 2 shows the benefits of sparsification for a range of matrix constructions, while Figure 2 illustrates how sparsification improves performance for a range of CS recovery algorithms. Finally, in Section 5 we conclude with some observations and open questions motivated by our numerical experiments.

2. Tradeoffs between sparsity and compressed sensing

A number of authors have investigated ways of replacing random ensembles with more computationally tractable sensing matrices. As previously mentioned, foundational results of Candès et al. establish asymptotically sharp results: to recover signals of length N with k non-zero entries, $n = \Theta(k \log N)$ measurements are necessary. Work of Chandar established that when $n = \Theta(k \log N)$, then the columns of Φ must contain at least $\Theta(\min\{k, N/n\})$ non-zero entries [10]. In [29], Nelson and Nguyen establish an essentially optimal result when $n = \Theta(k \log N)$ and $k < N/\log^3 N$. They show that each column of Φ necessarily contains $\Theta(k \log N)$ non-zero entries; i.e., the proportion of non-zero entries in Φ cannot tend to zero as N tends to ∞.

Observe that some restriction on k as a function of N is necessary: in the limiting case $k = N$, the identity matrix clearly suffices as a sparse sensing matrix. Furthermore, combinatorial constructions of sparse matrices are known which have near optimal recovery guarantees with a mutual incoherence property[2] [6]. In such matrices $n = O(k^2)$, and for certain infinite families of matrices (e.g., those

[2] Informally, the k-RIP property requires that k-sets of columns of Φ approximate an orthonormal basis. The incoherence of a matrix is the maximal inner product of a pair of columns, which is essentially the 2-RIP of Φ. In contrast to k-RIP, efficient constructions of matrices with near-optimal incoherence are known [6], but they have sub-optimal compressed sensing performance. Using 2-RIP alone, asymptotically one requires at least k^2 measurements to recover k-sparse

coming from projective planes) the number of non-zero entries in each column is $\Theta(k)$. Results bounding errors in the ℓ_1 norm (so-called RIP-1 guarantees) have been obtained using expander graphs. In particular, Bah and Tanner have shown that essentially optimal RIP-1 recovery can be achieved when $\lim_{n\to\infty} N/n = \alpha$ for some fixed α, with a constant number of non-zero entries per column [1]. (See also the discussion of dense versus sparse matrices in Section 3 of this paper.) These bounds are strictly weaker than RIP-2 bounds, though fast specialised algorithms have been developed for signal recovery with such matrices [32].

Since the k-RIP property is difficult to establish in practice, some authors have relaxed this in various directions. Berinde, Gilbert, Indyk, Karloff and Strauss [3] considered random binary matrices with constant column sum and related these to the incidence matrices of expander graphs. We reinterpret these matrices as sparsifications of the all-ones matrix below. Sarvotham, Baron and Baraniuk [2] and Dimakis, Smarandache and Vontobel [12] have considered the use of LDPC matrices. In particular, they have provided a strong correspondence between error-correcting performance of LDPC codes (when considered over $\mathbb{F}_2$) and CS performance of the same binary matrices (when considered over $\mathbb{R}$). While both groups obtained essentially optimal CS performance guarantees, their constructions are limited by the lack of known explicit constructions for expander graphs and LDPC codes respectively. Moghadam and Radha have previously considered a two step construction of sparse random matrices, involving construction of a random $(0,1)$-matrix followed by replacing each entry 1 with a sample from some probability distribution, [25, 26].

If one is content with recovery of each sparse vector with high probability, then much sparser matrices become useful. A strong result in this direction is due to Gilbert, Li, Porat and Strauss, who show that there exist matrices with $n = O(k \log N)$ rows and $\Theta(\log^2 k \log N)$ non-zero entries per column which recover sparse vectors with probability 0.75 [16] (see also [34]). Their matrices also come with efficient encoding, updating and recovery algorithms. While essentially optimal results are known for sparsity bounds on

signals. No practical deterministic construction is known which uses asymptotically fewer measurements; this obstruction is known as the *square-root bottleneck*, and finding more efficient deterministic constructions is a major open problem in compressed sensing.

CS matrices with an optimal number of rows, much less is known when either some redundant rows are allowed in the construction, or when RIP is replaced with a slightly weaker condition.

Several authors have compared the performance of sparse and dense CS matrices [17, 36, 13, 21]. Guo, Baron and Shamai have essentially shown that in certain limiting cases of the sparse recovery problem, dense and sparse sensing matrices behave in a surprisingly similar manner. In particular, they consider a variant of the recovery problem: given Φ and Φx, what can one say about any single component of x? They show that, as the size of the system becomes large (in a suitably controlled way), the problem of estimating x_i becomes independent of estimating x_j. In fact, the problem is equivalent to recovering a single measurement of x_i contaminated by additive Gaussian noise. They also apply their philosophy to sparse matrices, where as the size of the matrix becomes large, estimation of **all** signal components becomes independent, and each can be recovered independently. We refer the reader to the original paper for technical details [17]. As a result, under their assumptions, there should be no essential difference between CS performance in the sparse and dense cases.

For any compressed sensing matrix Φ, denote by $\delta(\Phi)$ the proportion of non-zero entries in Φ. Suppose that the number of columns of Φ is a linear function of the number of rows, i.e., that $N = \alpha n$ for some fixed $\alpha \in \mathbb{R}$. Suppose further that $\lim_{n\to\infty} \delta(\Phi) n^{1-\epsilon} = 0$ for any $\epsilon > 0$, but that $\lim_{n\to\infty} \delta(\Phi) n$ diverges. (Consider $\delta(\Phi) = \log n/n$ for example.) Under these hypotheses, Guo, Baron and Shamai claim that sparse and dense matrices exhibit identical CS performance. In particular, there exist matrices with $k \log k$ rows and $\Theta(\log k)$ non-zero entries per row which recover vectors of sparsity $O(k)$.

This appears to be in conflict with Nelson and Nguyen's result, which requires at least $\Theta(n \log\log n/ \log n)$ non-zero entries in such a matrix. The difference is that Nelson and Nguyen's result holds only when $n = O(k \log^3 N)$, whereas Guo, Baron and Shamai consider the case $n = \Theta(N)$.

Wang, Wainwright and Ramchandran's analysis of the number of measurements required for signal recovery depends on the quantity $(1 - \delta(\Phi))k$, which can be considered a measure of how much information about x is captured in each co-ordinate of Φx. They show that if $(1-\delta(\Phi))k \to \infty$ as $N \to \infty$ (which corresponds to relatively dense matrices, where, in the limit, each component of the signal is sampled infinitely often), then sparsification has no effect on recovery, while if $(1-\delta(\Phi))k$ remains bounded (so each component of the signal is sampled only finitely many times in expectation), then what the authors term "dramatically more measurements" are required. We refer the reader to the original paper for more details [36].

Our simulations are close in spirit to those considered by Guo, Baron and Shamai. Our computations are rather surprising as they suggest a modest **improvement** in signal recovery as we apply a sparsifying process to certain families of CS matrices. This improvement seems to persist across different recovery algorithms and different matrix constructions, and does not appear to have been noted in any of the work discussed in this section. (Though Lu, Li, Kpalma and Ronsin have observed some improvement in CS performance for sparse binary matrices [22].) We also observed a substantial improvement in the running times of the recovery algorithms, which may be of interest in practical applications.

3. Sparsification

We begin with a formal definition of sparsification.

Definition 3.1. The matrix Φ' is a *sparsification* of Φ if $\Phi'_{i,j} = \Phi_{i,j}$ for every non-zero entry of Φ'. The *density* of Φ, denoted $\delta(\Phi)$, is the proportion of non-zero entries that it contains, and the *relative density* of Φ' is the ratio $\delta(\Phi')/\delta(\Phi)$. We write $\mathrm{Sp}(\Phi, s)$ for the set of all sparsifications of Φ of relative density s.

In general, we have that $\mathrm{Sp}(\Phi, 1) = \Phi$, and that $\mathrm{Sp}(\Phi, 0)$ is the zero matrix. We also have a transitive property: if $\Phi' \in \mathrm{Sp}(\Phi, s_1)$ and $\Phi'' \in \mathrm{Sp}(\Phi', s_2)$ then $\Phi'' \in \mathrm{Sp}(\Phi, s_1 s_2)$. Two independent sparsifications will not in general be comparable: there is a partial ordering on the set of sparsifications of a matrix, but not a total order.

We illustrate our notation. Consider a Bernoulli random variable which takes value 1 with probability p and value 0 with probability $1-p$ and let Φ be an $n \times N$ matrix with entries drawn from this distribution; in short, a *Bernoulli ensemble* with expected value p. Then the expected density of Φ is p. Writing J for the all-ones matrix, a randomly chosen $\Phi' \in \mathrm{Sp}(J, p)$ will also have density p, and can be considered a good approximation of a Bernoulli matrix. If Φ''

is an independent random sparsification (i.e., all non-zero entries of the matrix have an equal probability of being set to zero) of Φ' with relative density p', then Φ'' approximates the Bernoulli ensemble with expected value pp'. So we have both $\Phi' \in \mathrm{Sp}(\Phi', p')$ and $\Phi' \in \mathrm{Sp}(J, pp')$. Later, we will consider successive sparsifications where we begin with a dense matrix whose entries are drawn from, e.g., a normal distribution.

Bernoulli ensembles have previously been considered in the compressed sensing literature, see [31] for example, though note that the matrices here take values in $\{0, 1\}$, not $\{\pm 1\}$. Such $\{\pm 1\}$-matrices are an affine transformation of ours: $M' = 2M - J$; as a result, the compressed sensing performance of either matrix is essentially the same.

In this paper, we will mostly be interested in *pseudo-random* sparsifications of an $n \times N$ compressed sensing matrix Φ. Specifically, for $s = t/n$, we obtain a matrix $\Phi' \in \mathrm{Sp}(\Phi, s)$ by generating a pseudo-random $\{0, 1\}$-matrix S with sn randomly located ones per column, and returning the entry-wise product $\Phi' = \Phi * S$. We will generally re-normalize Φ' so that every column has unit ℓ_2-norm.

Given a matrix Φ, we test its CS performance by running simulations. Since many different methodologies occur in the literature, we specify ours here.

Our k-sparse vectors always contain exactly k non-zero entries, in positions chosen uniformly at random from the $\binom{N}{k}$ possible supports of this size. The entries, unless otherwise specified, are drawn from a uniform distribution on the open interval $(0, 1)$. The vector is then scaled to have unit ℓ_2-norm. Simulations where the non-zero entries were drawn from the absolute value of a Gaussian distribution produced similar results. Note that many authors use $(0, 1)$- or $(0, \pm 1)$-vectors for their simulations. Appropriate combinations of matrices and algorithms often exhibit dramatic improvements of performance on this restricted set of signals.

We recover signals using ℓ_1-minimization. In this paper we will use the `matlab` LP-solver and the implementations of *Orthogonal Matching Pursuit* (OMP) and *Compressive Sampling Matching Pursuit* (CoSaMP) algorithms which were developed by Needell and Tropp [28]. Specifically, given a matrix Φ and signal vector x, we compute $y = \Phi x$, and solve the ℓ_1-minimization problem $\Phi\hat{x} = y$ for $\hat{x}$. The objective function is the ℓ_1-norm of $\hat{x}$ and it is assumed

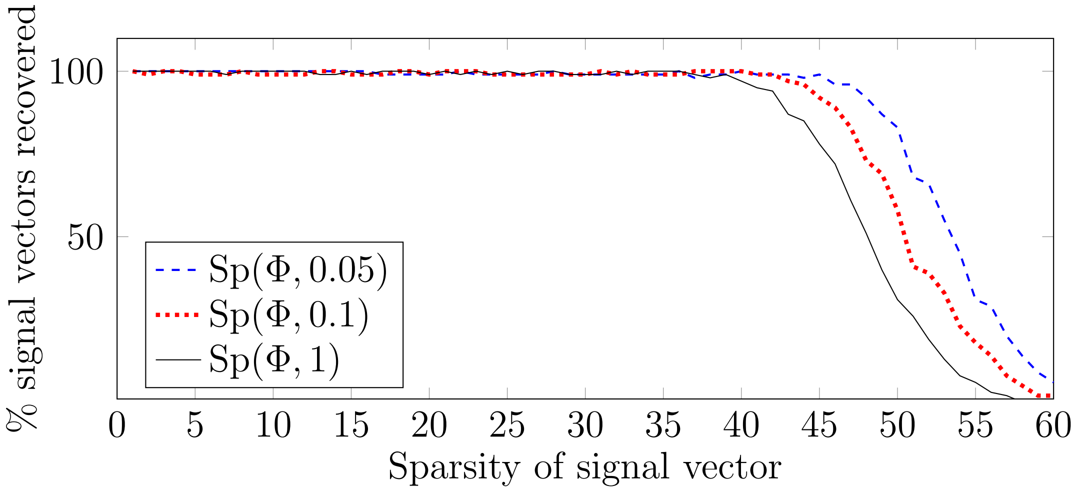

FIGURE 1. Effect of sparsification on signal recovery.

that all variables are non-negative. We consider the recovery successful if $|x - \hat{x}|_1 \leq c$ for some constant c. We take $c = 10^{-6}$ in all the simulations presented in this paper.

We conclude this section with an example illustrating the potential benefits of sparsification. In Figure 1, we explore the effect of sparsification on a 200×2000 matrix Φ with entries uniformly distributed on $(0, 1)$. The results for this case were compared with matrices drawn from $\mathrm{Sp}(\Phi, 0.1)$ and $\mathrm{Sp}(\Phi, 0.05)$. For each signal sparsity between 1 and 60, we generated 500 random vectors as described above and recorded the number of successful recoveries using the `matlab` LP-solver. To avoid bias we generated a new random matrix for each trial.

We observe that for signal vector sparsities between 45 and 55, matrices in $\mathrm{Sp}(\Phi, 0.05)$ achieve substantially better recovery than those from $\mathrm{Sp}(\Phi, 1)$. The code used to generate this simulation as well as others in this paper is available in full, along with data from multiple simulations at a webpage dedicated to this project: `http://fintanhegarty.com/compressed_sensing.html` .

4. RESULTS

Our simulations produce large volumes of data. To highlight the interesting features of these data-sets, we propose the following measure for acceptable signal recovery in practice.

Definition 4.1. For a matrix Φ and for $0 \leq t \leq 1$, we define the *t-recovery threshold*, denoted R_t, to be the largest value of k for which Φ recovers k-sparse signal vectors with probability exceeding t.

We construct an estimate $\hat{R}_t$ for R_t by running simulations. As the number of simulations that we run increases, $\hat{R}_t$ converges to R_t. In practice this convergence is rapid. The definition of R_t generalizes naturally to a space of matrices (say $n \times N$ Gaussian ensembles): it is simply the expected value of R_t for a matrix chosen uniformly at random from the space. To estimate R_t with reasonably high confidence, we proceed as follows: beginning with signals of sparsity $k = 1$, we attempt 50 recoveries. We increment the value k by 1 and repeat until we reach the first sparsity k_0 where less than $50t$ signals are recovered. Beginning at $k_0 - 3$, we attempt 200 recoveries at each signal sparsity. When we reach a signal sparsity k_1 where less than $200t$ signals are recovered, we attempt 1000 signal recoveries at each signal sparsity starting at $k_1 - 3$. When we reach a signal sparsity k_2 where less than $1000t$ signals are recovered, we set $\hat{R}_t = k_2 - 1$.

We typically find that $k_1 = k_2$, which gives us confidence that $\hat{R}_t = R_t$. Unless otherwise specified, we use the assumptions outlined in Section 3.

4.1. **Recovery algorithms with sparsification.** As suggested already in Figure 1, taking $\Phi' \in \mathrm{Sp}(\Phi, s)$ for some value of $s \sim 0.05$ seems to offer considerable improvements when using linear programming for signal recovery. Similar results hold for OMP and CoSaMP, though note that in each case we supply these algorithms with the sparsity of the signal vector. (While there is an option to withhold this data, the recovery performance of CoSaMP seems to suffer substantially without it — and we wish to be able to perform comparisons with linear programming.) In Figure 2, we graph $R_{0.98}$ of $\mathrm{Sp}(\Phi, s)$ as a function of s, where Φ is a 200×2000 matrix with entries drawn from the absolute values of samples from a standard normal distribution.

For each algorithm, $R_{0.98}$ appears to obtain a maximum for matrices of density between 0.15 and 0.05. It is perhaps interesting to note that the percentage improvement obtained by CoSaMP is far greater than that for either of the other algorithms.

Table 1 shows the average time taken for one hundred vector recovery attempts using 200×2000 measurement matrices with entries drawn from the absolute values of samples from a normal distribution, over a range of vector sparsities. We note an improvement in running time of an order of magnitude for linear programming when using sparsified matrices, and an improvement when using CoSaMP.

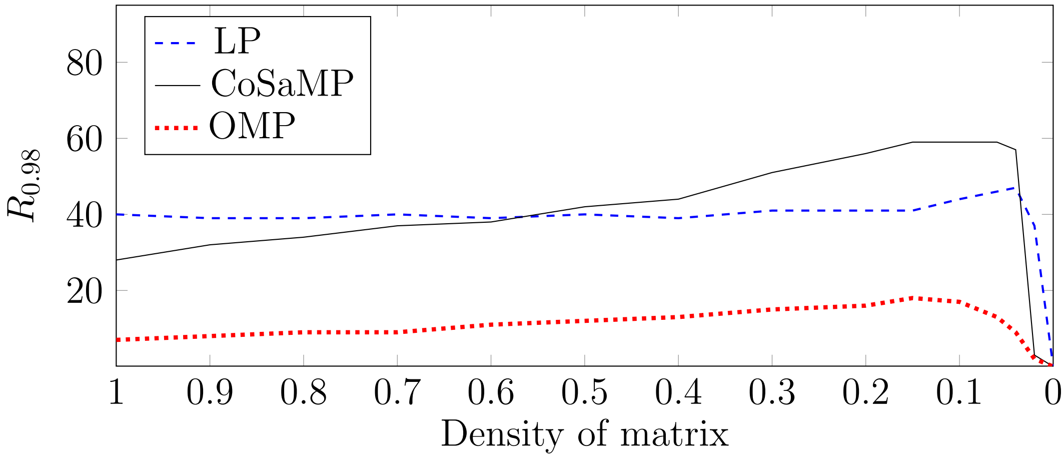

FIGURE 2. Signal recovery as a function of matrix density for LP, OMP and CoSaMP.

4.2. **Matrix constructions under sparsification.** In this section we explore the effect of sparsification on a number of different constructions proposed for CS matrices. We have already encountered the Gaussian, Uniform and Bernoulli ensembles. We will also consider some *structured random matrices*, which still have entries drawn from a probability distribution, but the matrix entries are no longer independent. The *partial circulant ensemble* [30] consists of rows sampled randomly from a circulant matrix, the first row of which contains entries drawn uniformly at random from some suitable probability distribution. Table 2 compares $R_{0.98}$ for $\mathrm{Sp}(\Phi, 1)$ and $\mathrm{Sp}(\Phi, 0.05)$ for 200×2000 matrices from each of the classes listed. Note that in the case of the Bernoulli ensemble, we actually compare $\mathrm{Sp}(J_{200,2000}, 0.5)$ with $\mathrm{Sp}(J_{200,2000}, 0.05)$, where $J_{200,2000}$ is an all-ones matrix. The entries of the partial circulant matrix were drawn from a normal distribution.

k	Time for 100 recovery attempts				% vectors successfully recovered			
	CoSaMP		LP		CoSaMP		LP	
	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$	$\delta = 1$
1	0.94	0.44	18.1	106.61	100	100	100	100
10	0.78	1.25	39.78	157.53	100	100	100	100
20	0.56	1.98	27.50	177.17	100	100	100	100
30	1.56	4.48	27.39	171.84	100	99	100	100
40	3.68	33.98	27.02	207.22	100	55	99	94
50	11.09	66.77	33.59	375.17	99	7	78	38
60	48.38	81.54	43.82	364.25	75	0	1	1
70	89.87	93.62	41.03	329.22	0	0	0	0

TABLE 1. Effect of sparsification on recovery time.

We denote by $\hat{k}$ the signal sparsity k for which the greatest difference in recovery between Φ and $\Phi' \in \mathrm{Sp}(\Phi, 0.05)$ occurs.

Construction	$R_{0.98}$		Maximal performance difference		
	$\delta = 1$	$\delta = 0.05$	$\hat{k}$	$\delta = 1$	$\delta = 0.05$
Normal	39	46	51	25	81
Uniform	39	45	51	24	73
Bernoulli	39	42	49	38	67
Partial Circulant	39	46	52	22	76

Table 2. Benefit of sparsification for different matrix constructions.

4.3. **Varying the matrix parameters.** Finally, we investigate the effect of sparsification on matrices of varying parameters. In particular, we explore the effect of sparsification on a family of matrices with entries drawn from the absolute value of the Gaussian distribution. First we explore the effect of sparsification as the ratio of columns to rows in the sensing matrix increases. For Figure 3, we use signal vectors whose entries were drawn from the absolute value of the normal distribution. We observe a modest improvement in performance which appears to persist.

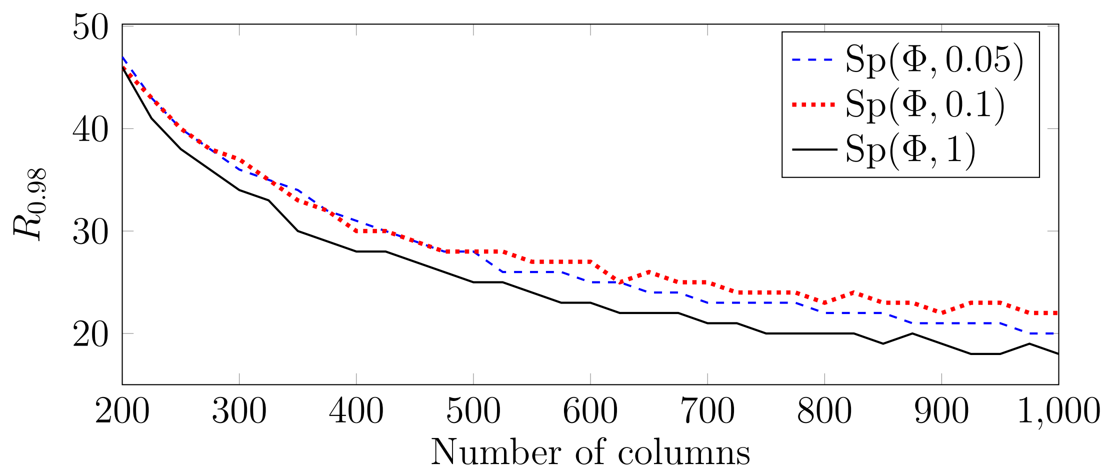

Figure 3. Recovery capability of matrices with 100 rows and varying number of columns under sparsification.

Now for Figure 4, we fix the ratio of columns to rows of Φ to be 10, and vary the number of rows. We know from the results of Candès et al. that $R_{0.98} = \Theta(n/\log n)$ in all cases. Nevertheless, the clear difference in slopes for recovery at different sparsities offers compelling evidence that the benefits of sparsification persist for large matrices.

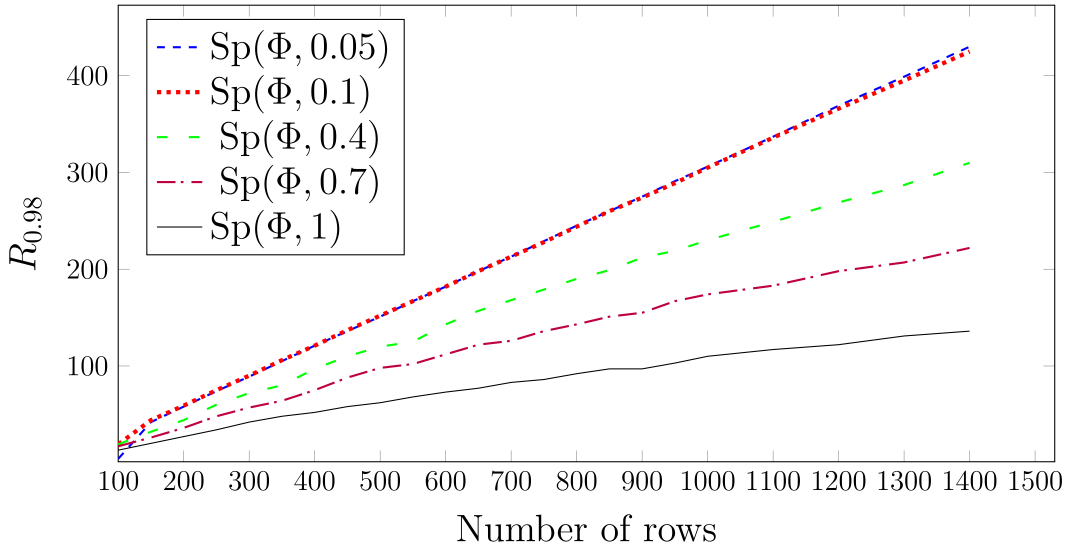

FIGURE 4. Recovery with CoSaMP for matrices with fixed row to column ratio under sparsification.

5. CONCLUSION

Some of the most important open problems in compressed sensing relate to the development of efficient matrix constructions and effective algorithms for sparse recovery. Deterministic constructions are essentially limited by the Welch bound: using known methods it is not possible to guarantee recovery of vectors of sparsity exceeding $\Theta(\sqrt{n})$, where n is the number of rows in the recovery matrix (see, e.g., [6]). Probabilistic constructions are much better: Candès and Tao's theory of restricted isometry parameters allows the provable recovery of vectors of sparsity k in dimension N with $\Theta(k \log(N))$ measurements. Such guarantees hold with overwhelming probability for Gaussian ensembles and many other classes of random matrices. But the random nature of these matrices can make the design of efficient recovery algorithms difficult. In this paper we have demonstrated that sparsification offers potential improvements for computational compressed sensing. In particular, Figures 1 and 4 show that sparsification results in the recovery of vectors of higher sparsity. Table 1 shows a substantial improvement in runtimes for linear programming arising from sparsification. These appear to be robust phenomena, which persist under a variety of recovery algorithms and matrix constructions. At the problem sizes that we explored, matrices with densities between 0.05 and 0.1 seemed to provide optimal performance.

We conclude with a small number of observations and conjectures which we believe to be suitable for further investigation. Since a Bernoulli ensemble in our terminology can be regarded as a sparsification of the all-ones matrix, it is clear that sparsification can improve CS performance. The necessary decay in CS performance as the density approaches zero shows that the effect of sparsification cannot be monotone. Extensive simulations suggest that when recovery is achieved with a general purpose linear programming solver, matrices with approximately 10% non-zero entries have substantially better CS properties than dense matrices. A catastrophic decay of compressed sensing performance occurs in many of the examples we investigated between densities of 0.05 and 0.01. We pose two questions which we think suitable for further research.

Question 5.1. *As the number of rows of* Φ *increases, the optimal matrix density appears to decrease. This effect does not appear to depend strongly on the matrix construction chosen. Does there exist a function* $\Gamma(n, N, k)$ *which describes the optimal level of sparsification for an* $n \times N$ *matrix recovering* k*-sparse vectors? Our simulations suggest that when* $N < n^{\alpha}$*, the optimal density of a CS matrix will be approximately* $\alpha n^{-1/2}$ *when* $k = o(n^{1-\epsilon})$*.*

Question 5.2. *We have considered pseudo-random sparsifications in this paper. In general, this should not be necessary. Are there deterministic constructions for* $(0, 1)$*-matrices with the property that their entry-wise product with a CS matrix improves CS performance? A natural class of candidates would be the incidence matrices of* t*-*(v, k, λ) *designs (see* [4] *for example). Some related work is contained in* [5, 6]*.*

Acknowledgements

The authors acknowledge the comments of anonymous referees.

The first and third authors have been supported by the Engineering and Physical Sciences Research Council grant EP/K00946X/1. The second author acknowledges the support of the Australian Research Council via grant DP120103067, and Monash University where much of this work was completed. This research was partially supported by the Academy of Finland (grants #276031, #282938, and #283262).

References

[1] B. Bah and J. Tanner. Vanishingly sparse matrices and expander graphs, with application to compressed sensing. *IEEE Trans. Inform. Theory*, 59(11):7491–7508, Nov 2013.

[2] D. Baron, S. Sarvotham, and R. Baraniuk. Bayesian compressive sensing via belief propagation. *IEEE Trans. Signal Process.*, 58(1):269–280, Jan 2010.

[3] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Proc. 46th Ann. Allerton Conference on Communication, Control, and Computing*, pages 798–805, Sept 2008.

[4] T. Beth, D. Jungnickel, and H. Lenz. *Design theory. Vol. I*, volume 69 of *Encyclopedia Math. Appl.*. Cambridge University Press, second edition, 1999.

[5] D. Bryant, C. Colbourn, D. Horsley, and P. Ó Catháin. Compressed sensing and designs: theory and simulations. *IEEE Trans. Inform. Theory*, 63(8):4850-4859, 2017.

[6] D. Bryant and P. Ó Catháin. An asymptotic existence result on compressed sensing matrices. *Linear Algebra Appl.*, 475:134–150, 2015.

[7] E. Candès. Mathematics of sparsity (and a few other things). *International Congress of Mathematicians, Plenary Lecture 3*, Seoul, Aug 2014.

[8] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, Feb 2006.

[9] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[10] V. B. Chandar. Sparse graph codes for compression, sensing and secrecy. *PhD thesis*, MIT, 2010.

[11] R. A. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complexity*, 23(4-6):918–925, 2007.

[12] A. Dimakis, R. Smarandache, and P. Vontobel. LDPC codes for compressed sensing. *IEEE Trans. Inform. Theory*, 58(5):3093–3114, May 2012.

[13] T. Do, L. Gan, N. Nguyen, and T. Tran. Fast and efficient compressive sensing using structurally random matrices. *IEEE Trans. Signal Process.*, 60(1):139–154, Jan 2012.

[14] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, April 2006.

[15] M. Fickus, D. G. Mixon, and J. C. Tremain. Steiner equiangular tight frames. *Linear Algebra Appl.*, 436(5):1014–1027, 2012.

[16] A. C. Gilbert, Y. Li, E. Porat, and M. J. Strauss. Approximate sparse recovery: Optimizing time and measurements. *SIAM J. Comput.*, 41(2):436–453, 2012.

[17] D. Guo, D. Baron, and S. Shamai. A single-letter characterization of optimal noisy compressed sensing. In *Proc. 47th Ann. Allerton Conference on Communication, Control, and Computing*, pages 52–59, 2009.

[18] S. Jafarpour, R. Willett, M. Raginsky, and R. Calderbank. Performance bounds for expander-based compressed sensing in the presence of Poisson

noise. In *Proc. 43rd Asilomar Conference on Signals, Systems and Computers*, pages 513–517, Nov 2009.

[19] M. Kern. Advanced technologies vastly improve MRI for children. *Science Highlights, National Institute for Biomedical Imaging and Bioengineering*, September 2014.

[20] S. Li and G. Ge. Deterministic construction of sparse sensing matrices via finite geometry. *IEEE Trans. Signal Process.*, 62(11):2850–2859, June 2014.

[21] W. Lu, K. Kpalma, and J. Ronsin. Semi-deterministic ternary matrix for compressed sensing. In *Proc. 22nd European Signal Processing Conference (EUSIPCO)*, pages 2230–2234, 2014.

[22] W. Lu, W. Li, K. Kpalma, and J. Ronsin. Near-optimal binary compressed sensing matrix. arXiv preprint abs/1304.4071, 2013.

[23] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–95, Dec 2007.

[24] M. Lustig, D. Donoho, J. Santos, and J. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, March 2008.

[25] A. Moghadam and H. Radha. Complex sparse projections for compressed sensing. In *Proc. 44th Ann. Conference on Information Sciences and Systems*, pages 1–6, March 2010.

[26] A. A. Moghadam and H. Radha. Sparse expander-like real-valued projection (SERP) matrices for compressed sensing. In *Proc. 2013 IEEE Global Conference on Signal and Information Processing*, pages 618–618, Dec 2013.

[27] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.

[28] D. Needell and J. A. Tropp. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2009.

[29] J. Nelson and H. L. Nguyen. Sparsity lower bounds for dimensionality reducing maps. In *Proc. 45th Ann. ACM Symposium on Theory of Computing*, pages 101–110, 2013.

[30] H. Rauhut, J. Romberg, and J. A. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. Anal.*, 32(2):242–254, 2012.

[31] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory*, 54(5):2210–2219, 2008.

[32] V. Ravanmehr, L. Danjean, B. Vasic, and D. Declercq. Interval-passing algorithm for non-negative measurement matrices: Performance and reconstruction analysis. *IEEE J. Emerging and Selected Topics in Circuits and Systems*, 2(3):424–432, Sept 2012.

[33] T. Tao. *Topics in random matrix theory*, volume 132 of *Grad. Stud. Math.*, American Mathematical Society, Providence, RI, 2012.

[34] A. Tehrani, A. Dimakis, and G. Caire. Optimal deterministic compressed sensing matrices. In *Proc.2013 IEEE Int. Conference on Acoustics, Speech and Signal Processing*, pages 5895–5899, May 2013.

[35] M. Uecker, S. Zhang, and J. Frahm. Nonlinear inverse reconstruction for real-time MRI of the human heart using undersampled radial FLASH. *Magnetic Resonance in Medicine*, 63(6):1456–62, 2010.

[36] W. Wang, M. Wainwright, and K. Ramchandran. Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices. *IEEE Trans. Inform. Theory*, 56(6):2967–2979, June 2010.
[37] S.-T. Xia, X.-J. Liu, Y. Jiang, and H.-T. Zheng. Deterministic constructions of binary measurement matrices from finite geometry. *IEEE Trans. Signal Process.*, 63(4):1017–1029, Feb 2015.
[38] Z. A. Zhu, R. Gelashvili, and I. Razenshteyn. Restricted isometry property for the general p-norms. *IEEE Trans. Inform. Theory*, 62:10, 5839-5854, Oct 2016.

Fintan Hegarty holds a PhD from the National University of Ireland, Galway, and spent 2013-15 as a postdoctoral researcher at the University of Birmingham. He currently works as a production editor for the Pacific Journal of Mathematics.
Padraig Ó Catháin received a PhD from the National University of Ireland, Galway in 2012. He was awarded the Kirkman medal of the Institute of Combinatorics and its Applications in 2015. He is currently an Assistant Professor at Worcester Polytechnic Institute. His research interests include combinatorial designs, Hadamard matrices and their applications.
Yunbin Zhao is a Senior Lecturer of Mathematical Optimization in the University of Birmingham. His main research interests include operations research, computational optimization, and numerical analysis.

(FH) School of Mathematics, Statistics and Applied Mathematics,
National University of Ireland, Galway
fintan.hegarty@gmail.com

(PÓC) Department of Mathematical Sciences,
Worcester Polytechnic Institute
p.ocathain@wpi.edu

(YZ) School of Mathematics,
University of Birmingham
y.zhao.2@bham.ac.uk