Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing

Ben Adcock
Department of Mathematics
Purdue University

Anders C. Hansen
DAMTP
University of Cambridge

Clarice Poon
DAMTP
University of Cambridge

Bogdan Roman
Computer Laboratory
University of Cambridge

***Abstract*—We introduce a mathematical framework that bridges a substantial gap between compressed sensing theory and its current use in real-world applications. Although completely general, one of the principal applications for our framework is the Magnetic Resonance Imaging (MRI) problem. Our theory provides a comprehensive explanation for the abundance of numerical evidence demonstrating the advantage of so-called variable density sampling strategies in compressive MRI. Besides this, another important conclusion of our theory is that the success of compressed sensing is resolution dependent. At low resolutions, there is little advantage over classical linear reconstruction. However, the situation changes dramatically once the resolution is increased, in which case compressed sensing can and will offer substantial benefits.**

I. Introduction

In this paper we present a new mathematical framework for overcoming the so-called coherence barrier in compressed sensing[1]. Our framework generalizes the three traditional pillars of compressed sensing—namely, *sparsity*, *incoherence* and *uniform random subsampling*—to three new concepts: *asymptotic sparsity*, *asymptotic incoherence* and *multilevel random subsampling*. As we explain, asymptotic sparsity and asymptotic incoherence are more representative of real-world problems—e.g. imaging—than the usual assumptions of sparsity and incoherence. For instance, problems in MRI are both asymptotically sparse and asymptotically incoherent, and hence amenable to our framework.

The second important contribution of the paper is an analysis of a novel and intriguing effect that occurs in asymptotically sparse and asymptotically incoherent problems. Namely, *the success of compressed sensing is resolution dependent*.

As suggested by their names, asymptotic incoherence and asymptotic sparsity are only truly witnessed for reasonably large problem sizes. When the problem size is small, there is little to be gained from compressed sensing over classical linear reconstruction techniques. However, once the resolution of the problem is sufficiently large, compressed sensing can and will offer a substantial advantage.

The phenomenon, which we call resolution dependence, has two important consequences for practitioners seeking to use compressed sensing in applications:

(i) Suppose one considers a compressed sensing experiment where the sampling device, the object to be recovered, the sampling strategy and subsampling percentage are all fixed, but the resolution is allowed to vary. Resolution dependence means that a compressed sensing reconstruction done at high resolutions (e.g. 2048×2048) will yield much higher quality when compared to full sampling than one done at a low resolution (e.g. 256×256). Hence a practitioner carrying out an experiment at low resolution may well conclude that compressed sensing imparts limited benefits. However, a markedly different conclusion would be reached if the same experiment were to be performed at higher resolution.

(ii) Suppose we conduct a similar experiment, but we now use the same total number of samples N (instead of the same percentage) at low resolution as we take at high resolution. Intriguingly, the above result still holds: namely, the higher resolution reconstruction will yield substantially better results. This is true because the multilevel random sampling strategy successfully exploits asymptotic sparsity and asymptotic incoherence. Thus, with the same amount of total effort, i.e. the number of measurements, compressed sensing with multilevel sampling works as a *resolution enhancer*: it allows one to recover the fine details of an image in a way that is not possible with the lower resolution reconstruction.

On a broader note, resolution dependence and its consequences suggest the following advisory: it is critical that simulations with compressed sensing be carried out with a careful understanding of the influence of the problem resolution. Naïve simulations with standard, low-resolution test images may very well lead to incorrect conclusions about the efficacy of compressed sensing as a tool for image reconstruction.

An important application of our work is the problem of MRI. This served as one of the original motivations for compressed sensing, and continues to be a topic of substantial research. Some of the earliest work on this problem—in particular, the research of Lustig et al. [1]–[3]—demonstrated that, due to high coherence, the standard random sampling strategies of compressed sensing theory lead to substandard reconstructions. Conversely, random sampling according to some nonuniform density was shown empirically to lead to substantially improved reconstruction quality. It is now

[1]This paper is part of a larger project on subsampling in applications. Further details, as well as codes and numerical examples, can be found on the project website http://subsample.org.

standard in MR applications to use a variable density strategy to overcome the coherence barrier. See [2]–[5].

This work has culminated in the extremely successful application of compressed sensing to MRI. However, a mathematical theory addressing these sampling strategies is largely lacking. Despite some recent work of Krahmer & Ward [6], a substantial gap remains between the standard theorems of compressed sensing and its implementation in such problems. Our framework aims to bridge precisely this gap. In particular, our theorems provide a mathematical foundation for compressed sensing for coherent problems, and gives credence to the above empirical studies demonstrating the success of nonuniform density sampling.

Whilst the MR problem will serve as our main application, we stress that our theory is extremely gen2‘eral in that it holds for almost arbitrary sampling and sparsity systems. In particular, standard compressed sensing results, such as those of Candès, Romberg & Tao [7], Candès & Plan [8], are specific instances of our main results.

For brevity, we shall provide only the most salient aspects of our framework. A more detailed discussion and analysis, containing proofs, further discussion and numerical experiments can be found in the paper [9] and the website [10].

II. Background

A. Compressed sensing

A typical setup in compressed sensing is as follows. Let $\{\psi_j\}_{j=1}^N$ and $\{\varphi_j\}_{j=1}^N$ be two orthonormal bases of $\mathbb{C}^N$, the *sampling* and *sparsity* bases respectively, and write

$$U = (u_{ij})_{i,j=1}^N \in \mathbb{C}^{N\times N}, \qquad u_{ij} = \langle \varphi_j, \psi_i \rangle.$$

Note that U is an isometry. The *coherence* of U is given by

$$\mu(U) = \max_{i,j=1,\dots,N} |u_{ij}|^2 \in [N^{-1}, 1]. \tag{1}$$

We say that U is *perfectly* incoherent if $\mu(U) = N^{-1}$.

Let $f \in \mathbb{C}^N$ be s-sparse in the basis $\{\varphi_j\}_{j=1}^N$. In other words, $f = \sum_{j=1}^N x_j \varphi_j$, and the vector $x = (x_j)_{j=1}^N \in \mathbb{C}^N$ satisfies $|\mathrm{supp}(x)| \leq s$, where

$$\mathrm{supp}(x) = \{j : x_j \neq 0\}.$$

Suppose now we have access to the samples

$$\hat{f}_j = \langle f, \psi_j \rangle, \quad j = 1, \dots, N,$$

and let $\Omega \subseteq \{1, \dots, N\}$ be of cardinality m and chosen uniformly at random. According to a result of Candès & Plan [8] and Adcock & Hansen [11], f can be recovered exactly with probability exceeding $1 - \epsilon$ from the subset of measurements $\{\hat{f}_j : j \in \Omega\}$, provided

$$m \gtrsim \mu(U) \cdot N \cdot s \cdot \left(1 + \log(\epsilon^{-1})\right) \cdot \log N. \tag{2}$$

In practice, recovery is achieved by solving the convex optimization problem:

$$\min_{\eta \in \mathbb{C}^N} \|\eta\|_{l^1} \text{ subject to } P_\Omega U \eta = P_\Omega \hat{f}, \tag{3}$$

where $\hat{f} = (\hat{f}_1, \dots, \hat{f}_N)^\top$, and $P_\Omega \in \mathbb{C}^{N\times N}$ is the diagonal projection matrix with j^{th} entry 1 if $j \in \Omega$ and zero otherwise.

B. The coherence barrier

The key estimate (2) shows that the number of measurements m required is, up to a log factor, on the order of the sparsity s, provided the coherence $\mu(U) = \mathcal{O}\left(N^{-1}\right)$. This is the case, for example, when U is the DFT matrix; a problem which was studied in some of the first papers on compressed sensing [7] (this example is actually perfectly incoherent).

On the other hand, when $\mu(U)$ large, one cannot expect to reconstruct an s-sparse vector f from highly subsampled measurements, regardless of the recovery algorithm employed [8]. We refer to this as the *coherence barrier*.

The MRI problem gives an important instance of this barrier. If $\{\varphi_j\}_{j=1}^N$ is a discrete wavelet and $\{\psi_j\}_{j=1}^N$ corresponds to the tows of the $N \times N$ discrete Fourier transform (DFT) matrix, then the matrix $U = \mathrm{DFT} \cdot \mathrm{DWT}^{-1}$ satisfies $\mu(U) = \mathcal{O}\left(1\right)$ for any N [6], [12]. Hence, although signals and images are typically sparse in wavelet bases, they cannot be recovered from highly subsampled measurements using the standard compressed sensing algorithm.

III. New concepts

In order to overcome the coherence barrier, we require three new concepts.

A. Asymptotic incoherence

Consider the above example. It is known that, whilst the global coherence $\mu(U)$ is $\mathcal{O}\left(1\right)$, the coherence decreases as either the Fourier frequency or wavelet scale increases. We refer to this property as *asymptotic incoherence*:

Definition 1. *Let $U \in \mathbb{C}^{N\times N}$ be an isometry. Then U is asymptotically incoherent if*

$$\lim_{\substack{K,N\to\infty \\ K<N}} \mu(P_K^\perp U) = \lim_{\substack{K,N\to\infty \\ K<N}} \mu(U P_K^\perp) = 0, \tag{4}$$

where $P_K^\perp : \mathbb{C}^{N\times N}$ is the projection matrix corresponding to the index set $\{K+1, \dots, N\}$.

Note that, for the wavelet example discussed above, one has $\mu(P_K^\perp U), \mu(U P_K^\perp) = \mathcal{O}\left(K^{-1}\right)$ [12].

B. Multilevel sampling

Asymptotic incoherence suggests a different subsampling strategy should be used instead of standard random sampling. High coherence in the first few rows of U means that important information about the signal to be recovered may well be contained in its corresponding low frequency measurements. Hence to ensure good recovery we should fully sample these rows. Once outside of this region, when the coherence starts to decrease, we can begin to subsample. Thus, instead of sampling uniform at random, we consider the following *multilevel* random sampling scheme:

Definition 2. *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r$, $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \dots, r$, and suppose that*

$$\Omega_k \subseteq \{N_{k-1}+1, \dots, N_k\}, \quad |\Omega_k| = m_k, \quad k = 1, \dots, r,$$

are chosen uniformly at random, where $N_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{N},\mathbf{m}} := \Omega_1 \cup \ldots \cup \Omega_r$ as an $(\mathbf{N},\mathbf{m})$-multilevel sampling scheme.

Note that similar sampling strategies are found in most empirical studies on compressive MRI [2]–[5]. Closely related strategies were also considered in [12], as well as in [13].

C. Asymptotic sparsity in levels

In the case of perfect incoherence, the standard random sampling strategies of compressed sensing are ideally suited for sparse signals. However, in asymptotically incoherent setting, the notion of sparsity can be substantially relaxed.

To explain this, let $x = (x_j)_{j=1}^N$ be vector of coefficients of a signal f in the basis $\{\varphi_j\}_{j\in\mathbb{N}}$. Suppose that x was very sparse in its entries $j = 1, \ldots, M_1$. Since the matrix U is highly coherent in its corresponding rows, there is no way we can exploit this sparsity to achieve subsampling. High coherence forces us to sample fully the first M_1 rows, otherwise we run the risking of missing critical information about x.

This means that there is nothing to be gained from high sparsity of x in its first few entries. However, we can expect to achieve significant subsampling if x is *asymptotically sparse*:

Definition 3. *For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \le M_1 < \ldots < M_r$ and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, with $s_k \le M_k - M_{k-1}$, $k = 1, \ldots, r$, where $M_0 = 0$. We say that $x \in l^2(\mathbb{N})$ is $(\mathbf{s},\mathbf{M})$-sparse if, for each $k = 1, \ldots, r$,*

$$\Delta_k := \mathrm{supp}(x) \cap \{M_{k-1}+1, \ldots, M_k\},$$

satisfies $|\Delta_k| \le s_k$.

As we next explain, signals possessing this sparsity pattern—which we henceforth refer to as being *asymptotically sparse in levels*—are ideally suited to multilevel sampling schemes. Roughly speaking, the number of measurements m_k required in each band Ω_k is determined by the sparsity of f in the corresponding band Δ_k and the asymptotic coherence.

This naturally leads to the question: which types of images are asymptotically sparse? The answer is that most images possess exactly this sparsity structure. Natural, real-life images are asymptotically sparsity in wavelet bases. At coarse scales, most images are not sparse, but sparsity rapidly increases with refinement.

IV. Main results

For brevity, we consider only the case of a two-level sampling strategy. The multilevel case is described in [9].

Write $\mu_K = \mu(P_K^{\perp}U)$. We now have:

Theorem 4. *Let $U \in \mathbb{C}^{N\times N}$ be an isometry and $x \in \mathbb{C}^N$ be $(\mathbf{s},\mathbf{M})$-sparse, where $r = 2$, $\mathbf{s} = (M_1, s_2)$, $s = M_1 + s_2$ and $\mathbf{M} = (M_1, M_2)$ with $M_2 = N$. Suppose that*

$$\|P_{N_1}^{\perp} U P_{M_1}\| \le \frac{\gamma}{\sqrt{M_1}}, \tag{5}$$

for some $1 \le N_1 \le N$ and $\gamma \in (0, 2/5]$, and that $\gamma \le s_2\sqrt{\mu_{N_1}}$. For $\epsilon > 0$, let $m \in \mathbb{N}$ satisfy

$$m \gtrsim (N - N_1) \cdot (\log(s\epsilon^{-1}) + 1) \cdot \mu_{N_1} \cdot s_2 \cdot \log(N).$$

Let $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ be a two-level sampling scheme, where $\mathbf{N} = (N_1, N_2)$ and $\mathbf{m} = (m_1, m_2)$ with $N_2 = N$, $m_1 = N_1$ and $m_2 = m$, and suppose that $\xi \in \mathbb{C}^N$ is a minimizer of (3), where $\hat{f} = Ux$. Then, with probability exceeding $1 - \epsilon$, ξ is unique and $\xi = x$.

Note that if f is not exactly sparse, and if the measurements $\hat{f} = Ux + z$ are corrupted by noise z satisfying $\|z\| \le \delta$, then one can also prove that under essentially the same conditions the minimization

$$\inf_{\eta\in\mathcal{H}} \|\eta\|_{l^1} \text{ subject to } \|P_\Omega U\eta - y\| \le \delta. \tag{6}$$

recovers f exactly, up to an error depending only on δ and the error of the best approximation $\sigma_{\mathbf{s},\mathbf{M}}(f)$ of x by an $(\mathbf{s},\mathbf{M})$-sparse vector. We refer to [9] for details.

A. Discussion

Theorem 4 demonstrates that asymptotic incoherence and two-level sampling overcomes the coherence barrier. To see this, note the following:

(i) The condition $\|P_{N_1}^{\perp} U P_{M_1}\| \le \frac{2}{5\sqrt{M_1}}$ (which is always satisfied for some N_1, since U is an isometry) implies that fully sampling the first N_1 measurements allows one to recover the first M_1 coefficients of f.

(ii) To recover the remaining s_2 coefficients we require, up to log factors, an additional

$$m_2 \gtrsim (N - N_1) \cdot \mu_{N_1} \cdot s_2,$$

measurements, taken randomly from the range $M_1 + 1, \ldots, M_2$. In particular, if N_1 is a fixed fraction of N, and if $\mu_{N_1} = \mathcal{O}\left(N_1^{-1}\right)$, such as for wavelets with Fourier measurements, then one requires only $m_2 \gtrsim s_2$ additional measurements to recover the sparse part of the signal.

(iii) When f is asymptotically sparse, such is the case for natural images, then the relative size of s_2 will become smaller as M and N grow. In particular, the percentage $\left(\frac{N_1+m_2}{N}\right)\times 100$ of measurements required will decrease. Hence the subsampling rate possible will improve as the problem resolution becomes larger (see Section V).

We remark that it is not necessary to know the sparsity structure, i.e. the values $\mathbf{s}$ and $\mathbf{M}$, of the image f in order to implement the multilevel sampling technique. Given a multilevel scheme $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$, the result of [9] governing asymptotically compressible signals demonstrates that f will be recovered exactly up to an error on the order of $\sigma_{\mathbf{s},\mathbf{M}}(f)$, where $\mathbf{s}$ and $\mathbf{M}$ are determined implicitly by $\mathbf{N}$, $\mathbf{m}$ and the conditions of the theorem. Of course, some *a priori* knowledge of $\mathbf{s}$ and $\mathbf{M}$ will greatly assist in selecting the parameters $\mathbf{N}$ and $\mathbf{m}$ so as to get the best recovery results. However, this is not necessary for implementing the method.

V. Resolution dependence and numerical results

As explained, natural, real life images are not sparse at coarse wavelet scales, nor is there substantial asymptotic incoherence. Hence, regardless of how we choose to recover f, there is little possibility for substantial subsampling. On the

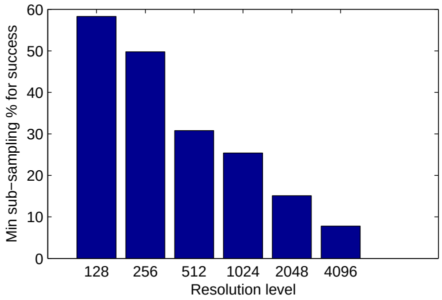

Fig. 1. The minimum subsampling percentage p.

Fig. 2. The GPLU phantom.

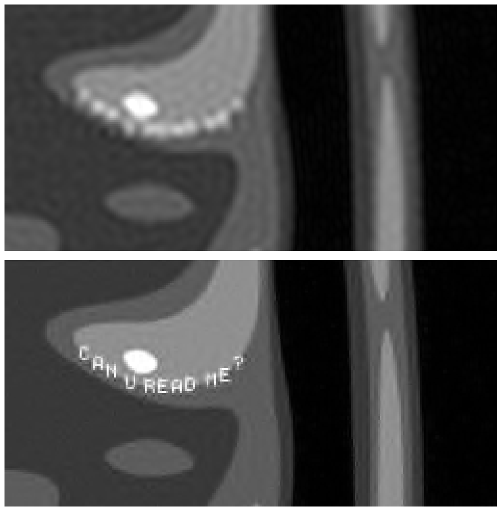

Fig. 3. The reconstruction of the 2048×2048 GPLU phantom from 512^2 Fourier samples. Top: linear reconstruction using the first 512^2 Fourier samples and zero padding elsewhere. Bottom: multilevel compressed sensing reconstruction.

other hand, asymptotic incoherence and asymptotic sparsity both kick in when the resolution increases. Multilevel sampling allows us to exploit these properties, and by doing so we achieve far greater subsampling.

To illustrate this, consider the reconstruction of the 1D image $f(t) = \mathrm{e}^{-t}\chi_{[0.2,0.8]}(t)$, $t \in [0,1]$, from its Fourier samples using orthonormal Haar wavelets. We use a two-level scheme with $p/2\%$ fixed samples and $p/2\%$ random samples, where p is the total subsampling percentage, and search for the smallest value of p such that the two-level sampling scheme succeeds: namely, it gives an error than that obtained by taking all possible samples of f.

In Figure 1 we plot p against the resolution level N. The difference between low resolution ($N = 128$) and high resolution ($N = 4096$) is clear and dramatic: the success of the reconstruction is highly *resolution dependent*. For the former we require nearly 60% of the samples, whereas with the latter this figure is reduced to less than 10%.

Now consider a different experiment, where the total number of measurements is fixed and equal to $512^2 = 262144$, but the sampling pattern is allowed to vary. Figure 2 displays the image to be recovered, which is the analytic phantom introduced by Guerquin–Kern, Lejeune, Pruessmann and Unser in [14]. For the purposes of comparison, artificial fine details were added to the image. In Figure 3 we display a segment of the reconstruction. As is clear, compressed sensing with multilevel sampling acts a *resolution enhancer*. By sampling higher in the Fourier spectrum, one recovers a far better image whilst taking the same amount of measurements.

For further numerical examples, as well as a comprehensive discussion of the multilevel sampling strategy used in Figure 2, we refer to [9], [10].

REFERENCES

[1] M. Lustig, “Sparse MRI,” Ph.D. dissertation, Stanford University, 2008.

[2] M. Lustig, D. L. Donoho, and J. M. Pauly, “Sparse MRI: the application of compressed sensing for rapid MRI imaging,” *Magn. Reson. Imaging*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed Sensing MRI,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, March 2008.

[4] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux, “Spread spectrum Magnetic Resonance Imaging,” *IEEE Trans. Med. Imaging*, vol. 31, no. 3, pp. 586–598, 2012.

[5] G. Puy, P. Vandergheynst, and Y. Wiaux, “On variable density compressive sampling,” *IEEE Signal Process. Letters*, vol. 18, pp. 595–598, 2011.

[6] F. Krahmer and R. Ward, “Compressive imaging: stable and robust recovery from variable density frequency samples,” *Preprint*, 2012.

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

[8] E. J. Candès and Y. Plan, “A probabilistic and RIPless theory of compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[9] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, “Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing,” *Arxiv 1302.0561*, 2013.

[10] (2013) http://subsample.org.

[11] B. Adcock and A. C. Hansen, “Generalized sampling and infinite-dimensional compressed sensing,” *Technical report NA2011/02, DAMTP, University of Cambridge*, 2011.

[12] E. J. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[13] V. Studer, J. Bobin, M. Chahid, H. Moussavi, E. Candès, and M. Dahan, “Compressive fluorescence microscopy for biological and hyperspectral imaging,” *Submitted*, 2011.

[14] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser, “Realistic analytical phantoms for parallel Magnetic Resonance Imaging,” *IEEE Trans. Med. Imaging*, vol. 31, no. 3, pp. 626–636, 2012.