Overcoming the coherence barrier 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 applications. Although completely general, one of the principal applications for our framework is the Magnetic Resonance Imaging (MRI) problem. Our theory provides an explanation for the abundance of numerical evidence demonstrating the advantage of so-called variable density sampling strategies in compressive MRI. 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 significant benefits.**

I. Introduction

In this paper we present a new mathematical framework for compressed sensing (CS). Our framework generalizes the three traditional pillars of CS—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.

Our second contribution is an analysis of an intriguing effect that occurs in asymptotically sparse and asymptotically incoherent problems. Namely, *the success of CS 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 consequently little to be gained from CS over classical linear reconstruction techniques. However, once the resolution of the problem is sufficiently large, CS can and will offer a substantial advantage.

The phenomenon has two important consequences for practitioners seeking to use CS in applications:

(i) Consider a CS 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 CS reconstruction done at high resolutions will give much higher quality when compared to full sampling than one done at a low resolution. Hence a practitioner working at low resolution may well conclude that CS 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 (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 give substantially better results. This is true because the multilevel random sampling strategy successfully exploits asymptotic sparsity and asymptotic incoherence. Thus, with the same total number of measurements, CS with multilevel sampling works as a *resolution enhancer*: it recovers fine details of an image in a way that is not possible with the lower resolution reconstruction.

Such resolution dependence suggests the following advisory. It is critical that simulations with CS 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 CS as a practical tool.

An important application of our work is the MRI problem. This served as one of the original motivations for CS, 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], [2]—demonstrated that the standard random sampling strategies of CS theory lead to substandard reconstructions. This is due to a phenomenon known as the *coherence barrier*.

On the other hand, random sampling according to some nonuniform density was shown empirically to lead to substantially improved reconstruction quality. It is now standard in MR applications to sample in this way [1]–[3]. However, whilst MRI is now viewed as a successful application area for CS, a mathematical theory addressing these sampling strategies is largely lacking. Despite some recent work [4], a substantial gap remains between the standard theorems of CS and its implementation in such problems (see [5] for a detailed discussion). Our framework bridges this gap. In particular, we provide a mathematical foundation for CS for such problems, and gives credence to the abundance of empirical studies demonstrating the success of variable density sampling in overcoming the coherence barrier.

Whilst the MRI problem will serve as our main application, we stress that our theory is general in that it holds for almost arbitrary sampling and sparsity systems. Moreover, standard CS results, in particular those of Candès & Plan [6], are specific instances of our main results.

For brevity, we shall provide only the most salient aspects of our framework. A substantially more detailed discussion can be found in [5]. We shall also only consider the finite-dimensional case. However, we remark that everything that follows can be extended to infinite-dimensional signals in separable Hilbert spaces [5]. This generalizes the theory of infinite-dimensional CS introduced in [7].

II. Background

A. Compressed sensing

A typical setup in CS 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 let

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

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

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

and 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

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

and let $\Omega \subseteq \{1, \ldots, N\}$ be of cardinality m and chosen uniformly at random. According to a result of Candès & Plan [6] and Adcock & Hansen [7], 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 $\hat{f} = (\hat{f}_1, \ldots, \hat{f}_N)^\top$, 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 estimate (2) shows that the number of measurements m 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 CS [8].

On the other hand, when $\mu(U)$ is large, one cannot expect to reconstruct an s-sparse vector f from highly subsampled measurements, regardless of the recovery algorithm employed [6]. 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 basis and $\{\psi_j\}_{j=1}^N$ corresponds to the rows of the $N \times N$ discrete Fourier transform (DFT) matrix, then the matrix $U = \mathrm{DFT} \cdot \mathrm{DWT}^{-1}$ satisfies $\mu(U) = \mathcal{O}(1)$ for any N [4], [9]. Hence, although signals and images are typically sparse in wavelet bases, they cannot be recovered from highly subsampled measurements using the standard CS algorithm.

III. New concepts

We now introduce our new framework that overcomes the aforementioned coherence barrier. We first require the following three new concepts.

A. Asymptotic incoherence

Consider the above example. It is known that, whilst the global coherence $\mu(U)$ is $\mathcal{O}(1)$, 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 $\{K+1, \ldots, N\}$.

Note that, for the wavelet example discussed above, one has $\mu(P_K^\perp U), \mu(U P_K^\perp) = \mathcal{O}\left(K^{-1}\right)$ [9] for all large N.

B. Multilevel sampling

When U is asymptotically incoherent a different subsampling strategy should be used instead of standard random sampling. High coherence in the first few rows of U means that we cannot subsample in this region without risking losing important information about the signal to be recovered. Hence we fully sample these rows. However, once outside of this region, where the coherence is less, we are free to subsample. Therefore, instead of sampling uniformly at random, we now consider the following *multilevel* random sampling scheme:

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

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

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 [1]–[3].

C. Asymptotic sparsity in levels

Having introduced the new sampling strategy for asymptotically incoherent problems, we now consider the following question: what is an appropriate signal model for such a sampling strategy? In the case of incoherence and uniform random subsampling, sparsity is an appropriate model. However, in this new setting we require a somewhat different notion.

To explain this, let $x = (x_j)_{j=1}^N$ be vector of coefficients of a signal f in the basis $\{\varphi_j\}_{j=1}^N$. Suppose that x was very sparse

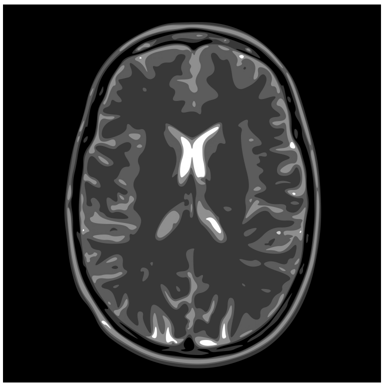

Fig. 1. The GPLU phantom.

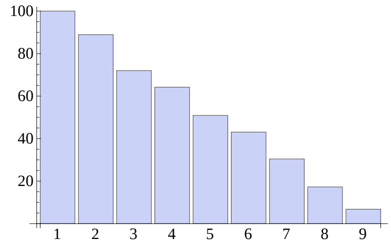

Fig. 2. The percentage of Haar wavelet coefficients at each scale for the image in Fig. 1 which are greater than 10^{-3} in magnitude.

in its entries $j = 1, \dots, M_1$. 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 of U, otherwise we risk 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 subsampling if the sparsity patten of x matches the incoherence pattern of the matrix U. We therefore consider:

Definition 3. *For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \dots < M_r$ and $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \dots, r$, where $M_0 = 0$. We say that $x \in \mathbb{C}^N$, where $N = M_r$, is $(\mathbf{s}, \mathbf{M})$-sparse if, for each $k = 1, \dots, r$, the quantity $\Delta_k := \mathrm{supp}(x) \cap \{M_{k-1}+1, \dots, M_k\}$ satisfies $|\Delta_k| \leq s_k$.*

In other words, we allow x to be split up into r levels, each with a different amount of sparsity. If the sparsity ratios $s_k/(M_k - M_{k-1})$ decrease with k, then we refer to x as being *asymptotically sparse in levels*.

As we shall see, signals possessing this sparsity pattern are ideally suited to multilevel sampling schemes. Roughly speaking, the concomitance of asymptotic sparsity and asymptotic incoherence means that the number of measurements m_k required in each band Ω_k is determined primarily by the sparsity of f in the corresponding band Δ_k times by a small asymptotic coherence factor.

This leads to the question: is asymptotic sparsity in levels a realistic signal model? The answer is emphatically yes. Most images possess exactly this type of sparsity structure. To illustrate, in Fig. 2 we plot the percentage of significant wavelet coefficients at each scale for the image given in Fig. 1. Note that this image is the analytic phantom introduced by Guerquin–Kern, Lejeune, Pruessmann and Unser in [10]. As is evident, there is little sparsity at coarse scales, but sparsity rapidly increases with refinement.

IV. Main result

For brevity, we shall only address the two-level case (the multilevel case is described in [5]). Thus, we consider signals with a two-level sparsity structure, with the first part being nonsparse, and the second part sparse, and a two-level sampling strategy that corresponds to full sampling in the first rows, and uniform random subsampling in the remaining rows.

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 $r = 2$, $\mathbf{s} = (s_1, s_2)$ and $\mathbf{M} = (M_1, M_2)$ with $s_1 = M_1$ and $M_2 = N$. Suppose that*

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

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

$$m \gtrsim (N - N_1) \cdot (\log((s_1 + s_2)\epsilon^{-1}) + 1) \cdot \mu_{N_1} \cdot s_2 \cdot \log(N) .$$

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 $(\mathbf{s}, \mathbf{M})$-sparse, and if the measurements $\hat{f} = Ux + z$ are corrupted by noise z satisfying $\|z\| \leq \delta$, then one can also prove that under essentially the same conditions the minimization

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

recovers f exactly, up to an error depending only on δ and the error $\sigma_{\mathbf{s},\mathbf{M}}(f)$ of the best approximation of x by an $(\mathbf{s}, \mathbf{M})$-sparse vector. We refer to [5] for details.

A. Discussion

Theorem 4 shows that asymptotic incoherence and two-level sampling overcomes the coherence barrier for two-level sparse signals. To see this, we note:

(i) The condition $\|P_{N_1}^\perp U P_{M_1}\| \leq 2/(5\sqrt{M_1})$ (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.

Let us explain how this relates to the MRI problem. With Fourier samples and wavelets as the sparsity system, (i) gives

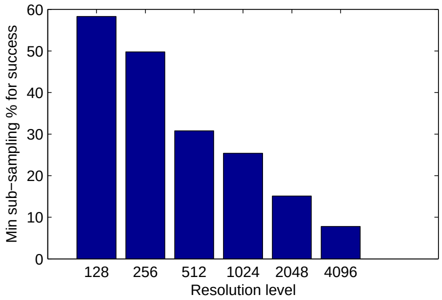

Fig. 3. The minimum subsampling percentage p.

that we recover the nonsparse part of the signal with $N_1 \approx M_1$ measurements. The fact that $N_1 \approx M_1$ in this case was shown in [11]. Since $\mu_{N_1} = \mathcal{O}\left(N_1^{-1}\right)$, (ii) gives that an additional $m_2 \gtrsim s_2$ measurements are required to recover the sparse part of the signal. Hence this result is nearly optimal for signals with two-level asymptotic sparsity. Namely, the full and the sparse parts of the signal are recovered using (up to constants and log factors) optimal numbers of measurements.

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 [5] governing $(\mathbf{s},\mathbf{M})$-compressible signals shows 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 strictly necessary for implementing the method.

V. Resolution dependence and numerical results

As explained, natural 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 when the problem resolution is low. On the 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 function $f(t) = \mathrm{e}^{-t}\chi_{[0.2,0.8]}(t)$, $t \in [0,1]$, from its Fourier samples using 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 smaller than that obtained by taking all possible samples of f.

In Fig. 3 we plot p against the resolution N. The difference between low resolution ($N = 128$) and high resolution ($N = 4096$) is clear and dramatic. We conclude that the success of the reconstruction is highly *resolution dependent*.

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. In Fig. 4 we display a segment of the reconstruction. For the purposes of comparison, artificial fine details were added to the image to be recovered. As is clear, CS with multilevel sampling acts a *resolution enhancer*. By sampling higher in the Fourier spectrum, one recovers fine details of the image whilst taking the same number of measurements.

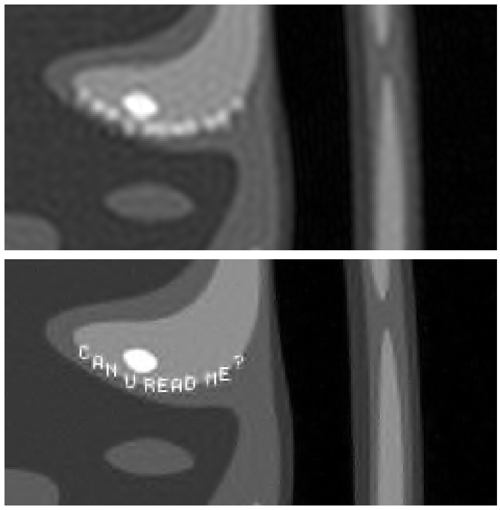

Fig. 4. The reconstruction of the 2048×2048 GPLU phantom (Fig. 1) from 512^2 Fourier samples. Top: linear reconstruction using the first 512^2 Fourier samples and zero padding elsewhere. Bottom: multilevel random CS reconstruction. Note that standard uniform random sampling CS would give an extremely poor reconstruction in this case, due to the $\mathcal{O}(1)$ global coherence.

For further numerical examples and discussion, see [5].

References

[1] 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.

[2] 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.

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

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

[5] 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.

[6] 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.

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

[8] 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.

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

[10] 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.

[11] B. Adcock, A. C. Hansen, and C. Poon, "On optimal wavelet reconstructions from Fourier samples: linearity and universality of the stable sampling rate," *Technical report NA2012/07, DAMTP, University of Cambridge*, 2012.