Constrained adaptive sensing

Mark A. Davenport, Andrew K. Massimino, Deanna Needell, and Tina Woolf

June 22, 2015

Abstract

Suppose that we wish to estimate a vector $\mathbf{x} \in \mathbb{C}^n$ from a small number of noisy linear measurements of the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$, where $\mathbf{z}$ represents measurement noise. When the vector $\mathbf{x}$ is sparse, meaning that it has only s nonzeros with $s \ll n$, one can obtain a significantly more accurate estimate of $\mathbf{x}$ by adaptively selecting the rows of $\mathbf{A}$ based on the previous measurements provided that the signal-to-noise ratio (SNR) is sufficiently large. In this paper we consider the case where we wish to realize the potential of adaptivity but where the rows of $\mathbf{A}$ are subject to physical constraints. In particular, we examine the case where the rows of $\mathbf{A}$ are constrained to belong to a finite set of allowable measurement vectors. We demonstrate both the limitations and advantages of adaptive sensing in this constrained setting. We prove that for certain measurement ensembles, the benefits offered by adaptive designs fall far short of the improvements that are possible in the unconstrained adaptive setting. On the other hand, we also provide both theoretical and empirical evidence that in some scenarios adaptivity does still result in substantial improvements even in the constrained setting. To illustrate these potential gains, we propose practical algorithms for constrained adaptive sensing by exploiting connections to the theory of optimal experimental design and show that these algorithms exhibit promising performance in some representative applications.

1 Introduction

Suppose that we wish to estimate a sparse vector from a small number of noisy linear measurements. In the setting where the measurements are selected in advance (independently of the signal) we now have a rich understanding of both practical algorithms and the theoretical limits on the performance of these algorithms. A typical result from this literature states that for a suitable measurement design, one can estimate a sparse vector with an accuracy that matches the minimax lower bound up to a constant factor [6]. Such results have had a tremendous impact in a variety of practical settings. In particular, they provide the mathematical foundation for "compressive sensing," a paradigm for efficient sampling that has inspired a range of new sensor designs over the last decade.

M. A. Davenport and A. K. Massimino are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA (e-mail: mdav@gatech.edu, massimino@gatech.edu) .

D. Needell is with the Department of Mathematics & Computer Science, Claremont McKenna College, Claremont, CA, 91711 USA (e-mail: dneedell@cmc.edu).

T. Woolf is with the Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA, 91711 USA (e-mail: tina.woolf@cgu.edu).

A distinguishing feature of the standard compressive sensing paradigm is that the measurements are *nonadaptive*, meaning that a fixed set of measurements are designed and acquired without allowing for any possibility of adapting as the measurements begin to reveal the structure of the signal. While this can be attractive in the sense that it enables simpler hardware design, in the context of sparse estimation this also leads to some clear drawbacks. In particular, this would mean that even once the acquired measurements show us that portions of the signal are very likely to be zero, we may still expend significant effort in "measuring" these zeros! In such a case, by *adaptively* choosing the measurements, dramatic improvements may be possible.

Inspired by this potential, recent investigations have shown that we can often acquire a sparse (or compressible) signal via far fewer measurements or far more accurately if we choose them adaptively (e.g., see [12, 17, 18, 21]). This body of work, which will be discussed in greater detail in Section 1.2, demonstrates that adaptive sensing indeed offers the potential for dramatic improvements over nonadaptive sensing in many settings. However, the existing approaches to adaptive sensing, which rely on being able to acquire *arbitrary* linear measurements, cannot be applied in most real-world applications where the measurements must respect certain physical *constraints*. In this paper, our focus is on *constrained adaptive sensing*, where our measurements are restricted to be chosen from a particular set of allowable measurements. We will see that new algorithms are required and explore the theoretical limits within this more restrictive setting. Before describing the constrained adaptive setting in more detail, we first provide a brief review of existing approaches to nonadaptive and adaptive sensing of sparse signals.

1.1   Nonadaptive sensing

In the standard nonadaptive compressive sensing framework [5, 7, 13, 14], we acquire a signal $\mathbf{x}$ via the linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$, where $\mathbf{A}$ is an $m \times n$ matrix representing the sensing system and $\mathbf{z}$ represents measurement noise. The goal is to design $\mathbf{A}$ so that m is smaller than n by exploiting the fact that $\mathbf{x}$ is *sparse* (or nearly sparse). Given a basis $\boldsymbol{\Psi}$, we say that a signal $\mathbf{x} \in \mathbb{C}^n$ is s-sparse if it can be represented by a linear combination of just s elements from $\boldsymbol{\Psi}$, i.e., we can write $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$, with $\|\boldsymbol{\alpha}\|_0 \leq s$, where $\|\boldsymbol{\alpha}\|_0 := |\text{supp}(\boldsymbol{\alpha})|$ denotes the number of nonzeros in $\boldsymbol{\alpha}$. We will typically be interested in the case where $s \ll n$.

There is now a rich literature that describes a wide range of techniques for designing an appropriate $\mathbf{A}$ and efficient algorithms for recovering $\mathbf{x}$. In much of this literature, the matrix $\mathbf{A}$ is chosen via randomized constructions that are known to satisfy certain desirable properties such as the so-called *restricted isometry property* (RIP).[1] Under the assumption that $\mathbf{A}$ satisfies the RIP (or that $\mathbf{A}\boldsymbol{\Psi}$ satisfies the RIP in the case where $\boldsymbol{\Psi} \neq \mathbf{I}$), if each entry of $\mathbf{z}$ is independent white Gaussian noise with variance σ^2 then one can show that techniques based on ℓ_1-minimization produce an approximation $\widehat{\mathbf{x}}$ satisfying

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq C \frac{n \log n}{\|\mathbf{A}\|_F^2} s\sigma^2, \tag{1}$$

where $C > 1$ is a fixed constant [13].

One can show that this result is essentially optimal in the sense that there is no alternative method to choose $\mathbf{A}$ or perform the reconstruction that can do better than this (up to the precise

---

[1]See Section 2 for a more detailed discussion of the RIP and its implications in the context of adaptive sensing. Note that the RIP is typically stated to require $\|\mathbf{A}\mathbf{x}\|_2 \approx \|\mathbf{x}\|_2$ for all s-sparse $\mathbf{x}$, which implies a fixed scaling for the matrix $\mathbf{A}$ where $\|\mathbf{A}\|_F^2 \approx n$. To ease the comparison with results that arise in contexts with alternative scalings, in the result stated in (1) we make no assumption on the scaling of $\mathbf{A}$ and merely require $\|\mathbf{A}\mathbf{x}\|_2 \approx \beta\|\mathbf{x}\|_2$ for some $\beta > 0$.

value of the constant C) [6]. In the event that the signal $\mathbf{x}$ is not exactly s-sparse, it is also possible to extend these results by introducing an additional term in the error bound that measures the error incurred by approximating $\mathbf{x}$ as s-sparse. See [13] and references therein for further details.

1.2 Adaptive sensing

A defining feature of the approach described above is that it is completely nonadaptive. When we consider the effect of noise, this nonadaptive approach might draw some severe skepticism. To see why, note that in the nonadaptive scenario, most of the "sensing energy" is used to measure the signal at locations where there is no information, i.e., where the signal vanishes. Specifically, one consequence of using the randomized constructions for $\mathbf{A}$ typically considered in the literature, or alternatively, any matrix satisfying the RIP, is that the available sensing energy (i.e., $\|\mathbf{A}\|_F^2$) is evenly distributed across all possible indices. This is natural since *a priori* we do not know where the nonzeros may lie, however, since most of the coordinates $\mathbf{x}_j$ are zero, it also means that the vast majority of the sensing energy is seemingly wasted. In other words, by design, the sensing vectors are approximately orthogonal to the signal, yielding a poor signal-to-noise ratio (SNR).

The idea behind adaptive sensing is that we should focus our sensing energy on locations where the signal is nonzero in order to increase the SNR, or equivalently, not waste sensing energy. In other words, one should try to learn as much as possible about the signal while acquiring it in order to design more effective subsequent measurements. Roughly speaking, one would like to (*i*) detect those entries which are nonzero or significant, (*ii*) progressively concentrate the sensing vectors on those entries, and (*iii*) estimate the signal from such localized linear functionals. Such a strategy is employed by the *compressive binary search* and *compressive adaptive sense and search* strategies of [12] and [21]. These algorithms operate by examining successively smaller pieces of the signal to accurately determine the location of signal energy. These techniques can yield dramatic improvements in recovery accuracy.

To quantify the potential benefits of an adaptive scheme, suppose that we observe

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle + z_i \tag{2}$$

where the z_i are independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2)$ entries and the $\mathbf{a}_i$ are allowed to depend on previous observations $(y_1, \cdots, y_{i-1})$, with the only constraint being that $\sum_i \|\mathbf{a}_i\|_2^2 = \|\mathbf{A}\|_F^2$ is fixed. Consider a simple procedure that uses half of the sensing energy in a nonadaptive way to identify the support of an s-sparse vector $\mathbf{x}$ and then adapts to use the remaining half of the sensing energy to estimate the values of the nonzeros. If such a scheme identifies the correct support, then it is easy to show that this procedure will yield an estimate satisfying

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 = \frac{2s}{\|\mathbf{A}\|_F^2} \, s\sigma^2. \tag{3}$$

If we contrast this result to that in (1), which represents the best possible performance in the nonadaptive setting, we see that this simple adaptive scheme can potentially improve upon the nonadaptive scheme by a factor of roughly $(n/s) \log n$, which represents a *dramatic* improvement in the typical scenario where $s \ll n$. Of course, this is predicated on the assumption that the first stage of support identification succeeds, which is not always the case.

A fundamental question is thus: *in practice, how much lower can the mean squared error (MSE) be when we are allowed to sense the signal adaptively?* The answer is a subtle one. In [1], it is shown that for the worst-case s-sparse vector $\mathbf{x}$, even the best possible adaptive scheme will satisfy

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \geq C \frac{n}{\|\mathbf{A}\|_F^2} \, s\sigma^2 \tag{4}$$

3

for a fixed constant $C > 0$. This lower bound improves upon the nonadaptive performance (1) by only a factor of $\log n$, coming far short of the improvement that (3) indicates might be possible. Similar results are also obtained in [9]. What these results find is that the same vectors that determine the worst-case performance in the nonadaptive setting are so difficult to estimate that it is impossible to obtain a reliable estimate of their support, and so adaptive algorithms offer limited room for improvement.

The result (4) does not say that adaptive sensing *never* helps. In fact, in practice it *almost always* does help. For example, when some or most of the nonzero entries in $\mathbf{x}$ are only slightly larger than the worst-case amplitude identified in [1], we *can* detect them sufficiently reliably to enable the dramatic improvements predicted in (3). More concretely, provided that σ^2 is not too large relative to the nonzero entries of $\mathbf{x}$, a well-designed adaptive scheme, where the $\mathbf{a}_i$ are chosen sequentially as in [12, 21], can achieve

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq C' \frac{s}{\|\mathbf{A}\|_F^2} s\sigma^2 \tag{5}$$

for a fixed constant C', which represents an enormous improvement when $s \ll n$, and demonstrates that the potential benefits suggested in (3) can be realized in certain regimes.

1.3 Constrained sensing

Up to this point, we have discussed results in which we essentially have complete freedom to design both the adaptive and nonadaptive measurements in an optimal fashion (that is, up to a constraint on $\|\mathbf{A}\|_F$). However, there are many applications where such freedom does not exist, and there are significant constraints on the kind of measurements that we can actually acquire. Such constraints arise in various hardware devices inspired by compressive sensing. For example, the single-pixel camera [15] acquires samples of an image by computing inner products with binary patterns. In this application we could still utilize adaptive measurements, but they must be binary. In other applications, we may be restricted to obtaining point samples of the signal of interest. For example, in standard sampling systems we are restricted to individually measuring each signal coefficient over time or space. Finally, in tomography and magnetic resonance imaging (MRI), as well as other medical imaging settings, we cannot acquire inner products with arbitrary linear functionals—we are limited to Fourier measurements.

In all of these settings, the measurements are *constrained*; we still have the flexibility to design measurements adaptively, but we can only select measurements from a fixed ensemble of predetermined measurements. Thus, the constrained setting will typically preclude the use of any of the adaptive sensing algorithms referenced above, and a new approach is required. Specifically, if we let $\mathcal{M} \subset \mathbb{C}^n$ denote the set of candidate measurement vectors, then the constrained adaptive sensing problem becomes one of sequentially selecting the rows $\mathbf{a}_i$ of our sensing matrix from the set $\mathcal{M}$. In this work, we assume the multiplicity of a particular measurement from $\mathcal{M}$ is allowed to be greater than one; that is, repeated measurements are permitted. For the methods discussed in this paper, we will restrict our attention to the case where $\mathcal{M}$ is a finite set. For a majority of our discussion and examples, we will focus on the setting where $|\mathcal{M}| = n$ and $\mathcal{M} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ is the ensemble of discrete Fourier measurements.

With this restriction on the measurements, Figure 1 illustrates the large potential difference between a completely nonadaptive sensing scheme, where the measurements are selected uniformly at random, and an "oracle" adaptive sensing scheme which uses a priori knowledge of the true locations of the nonzeros in a signal to carefully adapt the choice of measurement vectors to minimize the expected recovery error using the strategy outlined in Section 3. In both cases, the Compressive
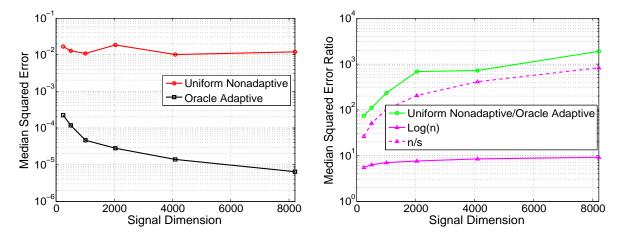
Figure 1: (Left) The median squared error versus the signal dimension n for nonadaptive recovery with uniformly random selected measurements (red) and oracle adaptive recovery (black). (Right) The ratio (green) of the nonadaptive median squared recovery error to the oracle adaptive median squared recovery error versus the signal dimension n, with $\log n$ (solid magenta) and n/s (dashed magenta) included for reference. In these simulations, the values on the support of $\boldsymbol{\alpha}$ are distributed i.i.d. according to $\mathcal{N}(\sqrt{n}, 1)$, and the number of measurements taken is always $m = 0.6n$ (rounding when necessary). The measurement noise $\mathbf{z}$ is distributed i.i.d. as $\mathcal{N}(0, \sigma^2 = 10^{-4})$.

Sampling Matching Pursuit (CoSaMP) [22] algorithm is used for the signal recovery. The median[2] squared error over 200 trials is displayed against the signal dimension n. Here, the signal is 10-sparse with a Haar wavelet decomposition that is supported on a tree (that is, $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$ where $\boldsymbol{\Psi}$ is the Haar wavelet basis, with $\boldsymbol{\alpha}$ such that $\|\boldsymbol{\alpha}\|_0 = 10$ and the support is on a tree).

It is well-known that the Fourier and Haar wavelet transforms are not incoherent, which implies that $\mathbf{A}\boldsymbol{\Psi}$ does not satisfy the RIP; hence, we would not expect blind nonadaptive sensing to do well in this setting. However, Figure 1 does illustrate the large potential for improvement over nonadaptive sensing. In this case, the adaptive algorithm can potentially improve the recovery error over nonadaptive sensing by roughly a factor of n/s, which represents a substantial gain when $s \ll n$. While we will see below that there are also nonadaptive strategies to address the coherence of Fourier and Haar which somewhat reduce the gap between adaptive and nonadaptive sensing in this case, we believe that this clearly illustrates the potential for adaptive sensing, even in the constrained setting.

1.4   Organization

The remainder of the paper is organized as follows. In Section 2, we show a simple lower bound on the adaptive performance of systems limited to Discrete Fourier Transform (DFT) measurements. We then generalize this result to the larger class of measurements satisfying the RIP. In both cases, the signal is assumed to be sparse in the canonical basis. In Section 3, we give a method for measurement selection based on optimal experimental design. In Section 4, we provide simulations in a more realistic setting and display numerical results when Fourier measurements are used and the signal is assumed to be sparse in the Haar wavelet basis. We also present some analytical justification using 1-sparse signals in this constrained adaptive setting. Finally, we conclude in

---

[2]Note that the *median* and *mean* curves exhibit the same overall behavior; however, we display the median error across all trials rather than the mean error throughout because the median, being a more robust measure, resulted in smoother curves with clearer trends between the methods.

Section 5 with a brief discussion.

2 Lower bounds on adaptive performance

The main result of this section shows that adaptive sensing cannot offer substantial improvements over the nonadaptive scheme when the measurements are restricted to certain specific classes of ensembles and the signal is sparse in the canonical basis (i.e., $\boldsymbol{\Psi} = \mathbf{I}$). We first consider the Fourier ensemble, where the sensing vectors are chosen from the rows of the DFT matrix $\mathbf{F} \in \mathbb{C}^{n \times n}$, where $\mathbf{F}$ has entries

$$f_{jk} = \frac{1}{\sqrt{n}} \exp(-2\pi\sqrt{-1}jk/n) \tag{6}$$

for $j, k = 0, 1, \ldots, n-1$. In this constrained setting we have the following lower bound.

Theorem 1. *Let* $\mathbf{x} \in \mathbb{R}^n$ *be s-sparse and suppose that we observe m measurements of the form* $y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle + z_i$ *where* $z_i \sim \mathcal{N}(0, \sigma^2)$ *and the* $\mathbf{a}_i$ *are chosen (potentially adaptively) by selecting rows from the DFT matrix* (6), *where at least s unique measurements are chosen and duplicates are otherwise permitted. Then for any estimate* $\widehat{\mathbf{x}}$ *by any adaptive scheme, we have*

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \geq \frac{n}{m} s\sigma^2. \tag{7}$$

This shows that even using an optimal choice of sensing vectors, the recovery error is still proportional to $\frac{n}{m}s\sigma^2$, which falls far short of the potential gains possible in the unconstrained setting shown in (3). This result is intuitive given the incoherence of the DFT and canonical bases. This is somewhat reminiscent of the main results of [1] and [9], which also show (in an unconstrained setting) that for a certain range of worst-case SNRs, adaptive schemes do not result in a substantial improvement. However, these arguments only apply for a narrow range of SNRs, and if the SNR improves (by only a small constant factor), there is a dramatic transition and adaptivity yields significant improvements [12, 21]. In contrast, Theorem 1 applies *no matter how large the SNR*, and so in a sense is far more pessimistic.

Proof of Theorem 1. Let $\mathbf{F}'$ be the $m \times n$ sensing matrix consisting of the m adaptively chosen vectors from the rows of $\mathbf{F}$, and let $\mathbf{F}'_\Lambda$ denote the $m \times s$ submatrix of $\mathbf{F}'$ whose column indices correspond to the indices of the support Λ of $\mathbf{x}$. We first observe that for any $s \times m$ matrix $\mathbf{B}$,

$$\mathbb{E}\|\mathbf{Bz}\|_2^2 = \mathbb{E}\,\mathrm{tr}(\mathbf{Bzz}^*\mathbf{B}^*) = \mathrm{tr}(\mathbf{B}\mathbb{E}[\mathbf{zz}^*]\mathbf{B}^*)$$
$$= \mathrm{tr}(\mathbf{B}\sigma^2\mathbf{I}\mathbf{B}^*) = \sigma^2\,\mathrm{tr}(\mathbf{BB}^*)$$
$$= \sigma^2\|\mathbf{B}\|_F^2.$$

Using the rows of $\mathbf{F}'$ to acquire the measurements as in (2), we obtain $\mathbf{y} = \mathbf{F}'\mathbf{x} + \mathbf{z} = \mathbf{F}'_\Lambda\mathbf{x}_\Lambda + \mathbf{z}$. It is not difficult to show (e.g., see [6]) that the optimal approximation to $\widehat{\mathbf{x}}$ has nonzero entries

$$\widehat{\mathbf{x}}_\Lambda = (\mathbf{F}'_\Lambda)^\dagger(\mathbf{F}'\mathbf{x} + \mathbf{z}) = (\mathbf{F}'_\Lambda)^\dagger(\mathbf{F}'_\Lambda\mathbf{x}_\Lambda + \mathbf{z}) = \mathbf{x}_\Lambda + (\mathbf{F}'_\Lambda)^\dagger\mathbf{z},$$

where $(\mathbf{F}'_\Lambda)^\dagger$ denotes the Moore–Penrose pseudoinverse of $\mathbf{F}'_\Lambda$ and we have assumed $\mathbf{F}'_\Lambda$ to be full rank (this requires at least s *unique* adaptively chosen measurements to comprise $\mathbf{F}'$, but duplicates are otherwise permitted). Thus, the optimal estimator yields an estimate with error $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 = \|(\mathbf{F}'_\Lambda)^\dagger\mathbf{z}\|_2$. Using this fact and denoting by $\sigma_i(\mathbf{B})$ the ith singular value of $\mathbf{B}$, we have

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \geq \mathbb{E}\|(\mathbf{F}'_\Lambda)^\dagger\mathbf{z}\|_2^2 = \sigma^2\|(\mathbf{F}'_\Lambda)^\dagger\|_F^2$$

6

$$= \sigma^2 \sum_{i=1}^{s} \sigma_i^2((\mathbf{F}'_\Lambda)^\dagger) = \sigma^2 \sum_{i=1}^{s} \frac{1}{\sigma_i^2(\mathbf{F}'_\Lambda)}$$

$$\geq \sigma^2 \frac{s^2}{\sum_{i=1}^{s} \sigma_i^2(\mathbf{F}'_\Lambda)} = \sigma^2 \frac{s^2}{\|\mathbf{F}'_\Lambda\|_F^2} \tag{8}$$

where the second inequality follows from Jensen's inequality. Since $\|\mathbf{F}'_\Lambda\|_F^2 = \frac{sm}{n}$, this completes the proof. $\qquad\square$

Our next result generalizes this type of lower bound to any ensemble whose submatrices satisfy the RIP with high probability. This statement is significant because it suggests that in some constrained situations, specifically many commonly studied in compressive sensing, there is little benefit from adaptivity. Formally, let $\mathbf{A}$ be any $n \times n$ matrix with unit-norm rows such that a (uniformly) random $m \times n$ submatrix $\widetilde{\mathbf{A}}$ satisfies the RIP with probability $1 - \exp(-cn)$ with RIP constant $\delta = 0.5$, i.e.,

$$0.5\frac{m}{n}\|\mathbf{u}\|_2^2 \leq \|\widetilde{\mathbf{A}}\mathbf{u}\|_2^2 \leq 1.5\frac{m}{n}\|\mathbf{u}\|_2^2, \tag{9}$$

where $\mathbf{u}$ is $2s$-sparse and we have normalized to account for the unit-norm row assumption. Theorem 2 makes rigorous the claim that selecting rows intelligently from such a matrix yields no substantial improvement over a nonadaptive scheme.

Theorem 2. *Let $\mathbf{x} \in \mathbb{R}^n$ be s-sparse and suppose that we observe m measurements of the form $y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle + z_i$ where $z_i \sim \mathcal{N}(0, \sigma^2)$ and the $\mathbf{a}_i$ are chosen (potentially adaptively) by selecting rows from a matrix $\mathbf{A}$ satisfying (9), where at least s unique measurements are chosen and duplicates are otherwise permitted. Then for any estimate $\widehat{\mathbf{x}}$ obtained from any adaptive scheme, we have*

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \geq \frac{n}{2m^2}s\sigma^2. \tag{10}$$

In fact, one usually anticipates m to be on the order of $s \log n$, in which case this bound becomes

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \geq \frac{sn}{2ms\log n}\sigma^2 = \frac{n}{2m\log n}\sigma^2,$$

which is roughly a factor of $s \log^2 n$ lower than the upper bound in (1). This result shows that the recovery error with any adaptive measurements selected from some standard RIP ensemble again falls short of the possible gains shown in (3).

Proof of Theorem 2. Let $\widetilde{\mathbf{A}}$ be a randomly selected $m \times n$ submatrix of $\mathbf{A}$. By assumption, with probability at least $1 - \exp(-cn)$ we have that $\widetilde{\mathbf{A}}$ satisfies (9). By choosing $\mathbf{x}$ to be a standard basis vector in (9), we observe that any column $\mathbf{a}^{(j)}$ of $\widetilde{\mathbf{A}}$ then satisfies

$$0.5\frac{m}{n} \leq \|\mathbf{a}^{(j)}\|_2^2 \leq 1.5\frac{m}{n}.$$

Thus, any $m \times s$ submatrix of $\widetilde{\mathbf{A}}$ can have squared Frobenius norm at most $\frac{1.5sm}{n}$.

Let $\mathbf{A}'$ be the $m \times n$ sensing matrix consisting of m adaptively chosen vectors from the rows of $\mathbf{A}$, with duplicates permitted. For the sake of a contradiction, suppose that there is some $m \times s$ submatrix $\mathbf{A}'_\Lambda$ of $\mathbf{A}'$, where Λ is some index set of cardinality s such that $\|\mathbf{A}'_\Lambda\|_F^2 > \frac{2sm^2}{n}$. Then at least one of the rows $\mathbf{a}_i$ of $\mathbf{A}'_\Lambda$ must be such that $\|\mathbf{a}_i\|_2^2 > \frac{2sm}{n}$. But this implies that any $m \times n$ submatrix of $\mathbf{A}$ that contains row i will have an $m \times s$ submatrix whose squared Frobenius norm exceeds $\frac{2sm}{n}$. Since row i is selected with probability $1/n$ in a randomly selected $m \times n$

submatrix $\widetilde{\mathbf{A}}$ of $\mathbf{A}$, and $1/n > \exp(-cn)$ (where c is chosen appropriately in (9)), we have arrived at a contradiction.

Therefore, it must be that all $m \times s$ matrices $\mathbf{A}'_\Lambda$ have squared Frobenius norm at most $\frac{2sm^2}{n}$. By (8), and assuming at least s unique rows of $\mathbf{A}$ are chosen in the construction of $\mathbf{A}'_\Lambda$, we thus have that

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \geq \frac{s^2}{\|\mathbf{A}'_\Lambda\|_F^2}\sigma^2 \geq \frac{s^2 n}{2sm^2}\sigma^2 = \frac{n}{2m^2}s\sigma^2.$$

This completes the proof. □

3 Adaptivity through optimal experimental design

Although there are some settings where constrained adaptive sensing does not offer substantial improvement over the nonadaptive scheme, one can of course ask if there are other settings where notable gains are still possible. In order to address this question, we consider the simplified constrained adaptive sensing problem where we assume the support Λ of the signal $\mathbf{x}$, with respect to the sparsity basis $\mathbf{\Psi}$, is known, or some estimate of the support is provided. How would we choose the measurements to best make use of this information, while still respecting that the measurements are constrained to be from the measurement ensemble $\mathcal{M}$?

Let $\{\mathbf{a}_i\}_{i=1}^m$ denote a sequence of length m with elements $\mathbf{a}_i \in \mathcal{M}$ corresponding to the measurements of $\mathcal{M}$ that are chosen.[3] Then, denote by $\mathbf{A}'$ the $m \times n$ matrix (recall $\mathcal{M} \subset \mathbb{C}^n$) whose i^{th} row is $\mathbf{a}_i$. If $\Lambda = \text{supp}(x)$, then it can be shown by following the arguments in the proof of Theorem 1 that the optimal MSE satisfies

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 = \|(\mathbf{A}'\mathbf{\Psi}_\Lambda)^\dagger\|_F^2\sigma^2 = \text{tr}\left(((\mathbf{A}'\mathbf{\Psi}_\Lambda)^*\mathbf{A}'\mathbf{\Psi}_\Lambda)^{-1}\right)\sigma^2, \tag{11}$$

where σ^2 is the variance of the noise term as in (2), $\mathbf{\Psi}_\Lambda$ is the submatrix of $\mathbf{\Psi}$ restricted to the columns indexed by Λ, and $\mathbf{A}'\mathbf{\Psi}_\Lambda$ is assumed to have full (column) rank. Our goal is to find a length-m measurement sequence $\{\mathbf{a}_i\}_{i=1}^m$ that minimizes (11), which is equivalent to solving

$$\{\widehat{\mathbf{a}}_i\}_{i=1}^m = \underset{\left\{\{\mathbf{a}_i\}_{i=1}^m \mid \mathbf{a}_i \in \mathcal{M}\right\}}{\text{argmin}} \ \text{tr}\left(((\mathbf{A}'\mathbf{\Psi}_\Lambda)^*\mathbf{A}'\mathbf{\Psi}_\Lambda)^{-1}\right), \tag{12}$$

where $\mathbf{A}' = \mathbf{A}'(\{\mathbf{a}_i\}_{i=1}^m)$ is constructed as described above. However, the optimization problem (12) that we would ideally like to solve is a computationally demanding discrete optimization problem; hence, we instead consider the relaxation

$$\widehat{\mathbf{S}} = \underset{\text{diagonal matrices } \mathbf{S} \succeq 0}{\text{argmin}} \ \text{tr}\left(((\mathbf{A}\mathbf{\Psi}_\Lambda)^*\mathbf{S}\mathbf{A}\mathbf{\Psi}_\Lambda)^{-1}\right) \quad \text{subject to } \text{tr}(\mathbf{S}) \leq \mathcal{E}, \tag{13}$$

where $\mathbf{A}$ is the $|\mathcal{M}| \times n$ matrix containing all possible measurement vectors from $\mathcal{M}$ and the constraint $\text{tr}(\mathbf{S}) \leq \mathcal{E}$ ensures that the resulting "weighted" sensing matrix $\sqrt{\mathbf{S}}\mathbf{A}$ satisfies the "sensing energy" constraint $\|\sqrt{\mathbf{S}}\mathbf{A}\|_F^2 \leq \mathcal{E}$ when the rows of $\mathbf{A}$ are normalized. Note that this is equivalent to the continuous design for the A-optimality criterion studied in the optimal experimental design literature [24].

Fortunately, (13) is a convex problem [4] and can be efficiently solved by a number of methods. Whereas the problem in (12) would tell us which measurements and how many of each to use from

---

[3]We use a *sequence* of elements from $\{1, \ldots, |\mathcal{M}|\}$ rather than: 1) a *subset* to emphasize that the m measurements from $\mathcal{M}$ need not be distinct, and 2) a *multiset* because there is a significance to the order in which the measurements are selected in adaptive sensing.

$\mathcal{M}$, (13) instead tells us, through the diagonal matrix $\widehat{\mathbf{S}}$ of weights, "how much" of each measurement to use. If we use the sensing energy model, the way to use $\widehat{\mathbf{S}}$ from (13) is straightforward: the sensing matrix is simply $\sqrt{\widehat{\mathbf{S}}}\mathbf{A}$ From the relation

$$\text{tr}\left(((\mathbf{A\Psi}_\Lambda)^*\widehat{\mathbf{S}}\mathbf{A\Psi}_\Lambda)^{-1}\right) = \frac{m}{\mathcal{E}}\,\text{tr}\left[\sum_{i=1}^{n}\frac{m\widehat{s}_{ii}}{\mathcal{E}}(\mathbf{A\Psi}_\Lambda)_i^*(\mathbf{A\Psi}_\Lambda)_i\right]^{-1}, \tag{14}$$

where $(\mathbf{A\Psi}_\Lambda)_i$ denotes the ith row of $\mathbf{A\Psi}_\Lambda$ and $\widehat{s}_{ii}$ denotes the ith element on the diagonal of $\widehat{\mathbf{S}}$, it is clear that if q is an integer, setting $m\widehat{s}_{ii}/\mathcal{E} = q$ is equivalent to repeating the corresponding measurement q times.

If, on the other hand, we use the measurement model where we must choose m unweighted measurements from $\mathcal{M}$, the practical use of $\widehat{\mathbf{S}}$ from (13) is less obvious. We experimented with several different approaches to using the weights in $\widehat{\mathbf{S}}$, and the following method empirically seemed to produce the best results. In this work, we use a simple (though likely sub-optimal) sampling scheme to obtain a discrete design. Specifically, we set $\mathcal{E} = m$ and draw exactly m measurements, with replacement, according to the probability mass function

$$p_i = \frac{\widehat{s}_{ii}}{m}. \tag{15}$$

We guarantee that the resulting matrix $\mathbf{A}'$ is at least rank s by rejecting any construction for which this constraint is not satisfied. These m measurements then form the rows of the sensing matrix $\mathbf{A}'$.

4   Case study: Fourier measurements of Wavelet sparse signals

The results of Section 2 demonstrate that adaptive sensing cannot offer substantial improvements over nonadaptive sensing for certain classes of measurement ensembles when the signal is sparse in the canonical basis. We next explore the case when $\mathbf{\Psi}$ is instead a wavelet basis and we acquire Fourier measurements (this is indeed the setting of Figure 1, which suggests dramatic potential improvements from constrained adaptive sensing). This setting serves as a somewhat idealized model for a number of applications in tomography and other medical imaging since physical limitations would entail that we can only acquire Fourier measurements, and natural images are generally sparse with respect to wavelet bases [11]. In this setting we might receive one Fourier measurement at a time, and from those, we can (potentially in real time) request the next Fourier coefficient to be measured.

For the remainder of this discussion we will assume the sparsity basis $\mathbf{\Psi}$ is the Haar wavelet basis. We will denote the $n \times n$ discrete Haar wavelet transform by $\mathbf{H}$, with entries h_{jk} for $j, k = 0, 1, \ldots, n-1$ and n is assumed to be some power of 2. When $j = 0$, we have

$$h_{0k} = \frac{1}{\sqrt{n}}. \tag{16}$$

For indices $j > 0$, we write $j = 2^p + q - 1$, where p and q are nonnegative integers, and define

$$h_{jk} = \frac{1}{\sqrt{n}}\begin{cases} 2^{p/2} & \frac{(q-1)n}{2^p} \le k < \frac{(q-0.5)n}{2^p} \\ -2^{p/2} & \frac{(q-0.5)n}{2^p} \le k < \frac{qn}{2^p} \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Since, however, the Haar wavelet basis $\mathbf{H}$ is a sparsifying transformation, for a signal (or image) $\mathbf{x}$ we have that $\mathbf{Hx} = \boldsymbol{\alpha}$, with $\|\boldsymbol{\alpha}\|_0 \leq s$. This means $\mathbf{x} = \mathbf{H}^*\boldsymbol{\alpha}$, where $\mathbf{H}^*$ denotes the adjoint of $\mathbf{H}$, for which $\mathbf{H}^* = \mathbf{H}^{-1}$ since $\mathbf{H}$ is unitary.

With this notation in hand and recalling that $\mathbf{F}$ is the $n \times n$ DFT, (11) becomes

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 = \|(\mathbf{F}'\mathbf{H}_\Lambda^*)^\dagger\|_F^2 \sigma^2, \tag{18}$$

where $\mathbf{F}'$ is the $m \times n$ sensing matrix consisting of the m adaptively chosen vectors from $\mathbf{F}$ and $\mathbf{H}_\Lambda^*$ is the $n \times s$ submatrix of $\mathbf{H}^*$ restricted to the columns indexed by $\Lambda = \mathrm{supp}(\boldsymbol{\alpha}) = \mathrm{supp}(\mathbf{Hx})$. Thus, we see that the optimal MSE depends on the correlations of the Fourier and Haar basis elements. We now present a suite of numerical simulations in this setting that employ the relaxation (13) followed by the sampling scheme described in Section 3 to select a sequence of m measurement vectors from $\mathbf{F}$. We then follow with a short analysis for the simple case of 1-sparse signals.

4.1 Simulations

Here we present a practical implementation of adaptive sensing obtained via the relaxation (13) which we then compare with the results of traditional nonadaptive sensing. To implement (13), we use the Templates for First-Order Conic Solvers (TFOCS) software package [2, 3]. We set $\mathcal{M}$ to be the ensemble of n measurements from the $n \times n$ DFT matrix $\mathbf{F}$. For all of the numerical experiments, $\mathbf{x}$ is a 10-sparse signal in the Haar wavelet basis (i.e., $\boldsymbol{\Psi} = \mathbf{H}^\star$) with the values on the support of $\boldsymbol{\alpha}$ distributed i.i.d as $\mathcal{N}(\sqrt{n}, 1)$ and the measurement noise $\mathbf{z}$ is distributed as i.i.d. $\mathcal{N}(0, 10^{-4})$. Unless otherwise stated, the signal is of dimension $n = 1024$. We consider signals whose support is chosen uniformly at random, and also those whose support obeys a tree structure. Briefly, in the latter case the support is organized on a binary tree, plus an extra node at the top. The first scaling (or lowest frequency) coefficient has just one child; the second and further wavelet coefficients have two children each (see [10, 16] for similar wavelet-tree constructions). An s-sparse support is filled by choosing the first scaling location, and then in each of the $s-1$ remaining rounds, choosing one node randomly among the unfilled nodes which currently have a chosen parent.

Nonadaptive sensing. Due to the lack of incoherence between the DFT and Haar bases, it has been observed (and recently theoretically shown [19, 20]) that so-called *Variable-Density Sampling* (VDS) is often preferable to standard uniform random selection of Fourier measurements. In VDS[4], sampling can be concentrated on the lower frequencies, producing superior recovery results. We test recovery using either ℓ_1-minimization [8, 25, 26] or the greedy pursuit CoSaMP [22].

Adaptive sensing. In the more realistic setting, we employ a simple strategy which uses $m/2$ nonadaptive measurements (using either VDS or uniform sampling) to construct an estimate of Λ. This is done by executing either ℓ_1-minimization (followed by thresholding) or CoSaMP. We then solve the relaxation (13) using this estimated support, and the remaining $m/2$ measurements are selected adaptively using the distribution given by (15). To recover the signal, either ℓ_1-minimization or CoSaMP is again used to obtain an updated estimate $\hat{\Lambda}$ using all m measurements. The final signal coefficient estimate is calculated as $\widehat{\boldsymbol{\alpha}}_{\hat{\Lambda}} = (\mathbf{F}'\mathbf{H}_{\hat{\Lambda}}^*)^\dagger \mathbf{y}$, where $\mathbf{y}$ is the m-dimensional vector of measurements, $\mathbf{F}'$ is the $m \times n$ vector of Fourier measurements selected, and $\mathbf{H}_{\hat{\Lambda}}^*$ is the $n \times 10$ submatrix of $\mathbf{H}^*$ restricted to the columns indexed by $\hat{\Lambda}$. One could alternatively use the signal estimate returned directly from the recovery algorithm, which we have observed to perform similarly to (or only slightly worse than) our implemented method.

Oracle adaptive sensing. For sake of comparison, we also consider the case where the true support Λ of the signal is known a priori, and the measurements are selected as in the adaptive

---

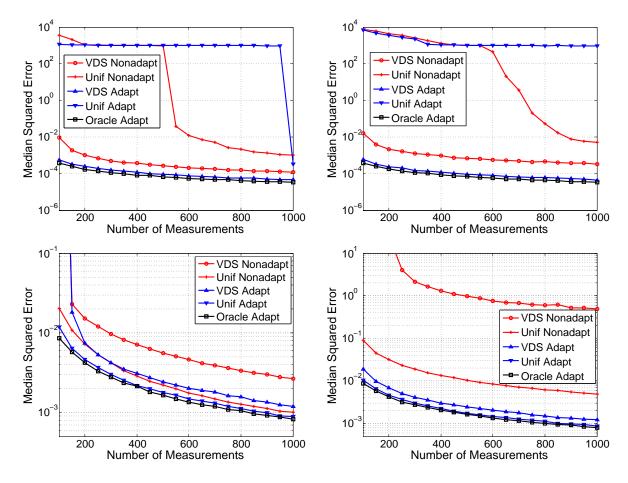[4]Following the experiments in [20], we also do not apply any preconditioning to the sensing matrix.

Figure 2: The median squared error versus the number of measurements m when the nonzero locations of $\boldsymbol{\alpha}$ are selected on a sparse tree (top) or uniformly at random (bottom). The nonadaptive (red) and adaptive (blue) recovery is shown when either VDS or uniform sampling is used for the nonadaptive measurements, and CoSaMP (left) or ℓ_1-minimization (right) is used; the oracle adaptive (black) recovery is also included for comparison.

sensing case using this Λ. Recovery is then performed simply by applying the pseudoinverse: $\widehat{\boldsymbol{\alpha}}_\Lambda = (\mathbf{F}'\mathbf{H}_\Lambda^*)^\dagger \mathbf{y}$.

Figure 2 compares recovery results over 1000 trials for nonadaptive, adaptive, and oracle adaptive sensing versus the number of measurements m. We see that when the signal is supported on a tree, uniform sampling performs poorly for both nonadaptive and adaptive sensing, as might be expected. The performance of the uniform sampling methods can be understood via the empirical observation that in this case we require roughly 500 measurements before we can reliably estimate the support. When using the CoSaMP algorithm, the sudden improvement at $m \approx 500$ for uniform nonadaptive and at $m \approx 1000$ for uniform adaptive (which uses $m \approx 500$ measurements for support identification) corresponds to the threshold where more than half of the trials resulted in a correct support recovery. In contrast, sampling with VDS offers dramatic improvements for both nonadaptive and adaptive sensing with either reconstruction algorithm, with adaptive sensing performing almost as well as the oracle. In this case, VDS is already capturing much of the potential improvement offered by adaptivity because the energy of the signal is heavily biased towards the lower frequencies, although adaptivity still results in somewhat improved performance. In contrast to the tree-sparse case, when the signal support is selected randomly, uniform nonadaptive sampling
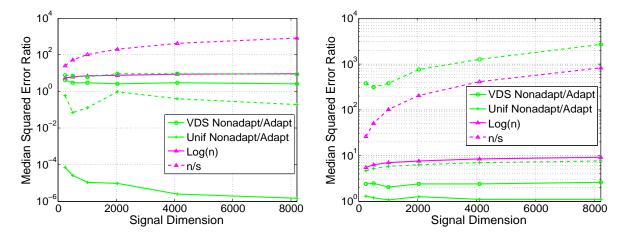
Figure 3: The ratio (green) of the nonadaptive median squared recovery error to the adaptive median squared recovery error versus the signal dimension n when the support locations of $\boldsymbol{\alpha}$ are selected on a sparse tree (left) or uniformly at random (right). The ratio is shown when either VDS or uniform sampling is used for the nonadaptive measurements, and CoSaMP (solid line) or ℓ_1-minimization (dashed line) is used. The curves for $\log n$ (solid magenta) and n/s (dashed magenta) are included for comparison.

actually performs better than VDS, whereas adaptive sensing performs similarly regardless of the type of nonadaptive measurements taken. Thus if one is not sure of the signal structure in general, adaptive sensing can offer improvements in either case. This flexibility represents one of the main advantages of adaptive sensing.

The results of our second simulation are shown in Figure 3, where we compare the ratio of nonadaptive to adaptive sensing recovery over 200 trials against the dimension n of the signal $\mathbf{x}$; the number of measurements used is always $m = 0.6n$. We note that since the norms of $\boldsymbol{\alpha}$ and $\mathbf{z}$ both scale with n, the SNR remains roughly the same for all signal dimensions n. We observe similar results as Figure 2, demonstrating the behavior holds as a function of dimension.

4.2   Analysis of the 1-sparse case

We now provide some analytical justification explaining why adaptive sensing can achieve a lower MSE than nonadaptive sensing for the Haar wavelet basis with Fourier measurements, but show that the largest gains are realized for a small fraction of the possible signal support sets. We consider the simple case when $s = 1$ and the support is eventually known (either by oracle or by utilizing some method for estimation, as in the above experiments), and use this toy problem as motivational justification for the general setting. If we denote the s singular values of $\mathbf{F}'\mathbf{H}_\Lambda^*$ by $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s > 0$, then in general we want to minimize

$$\|(\mathbf{F}'\mathbf{H}_\Lambda^*)^\dagger\|_F^2 = \sum_{i=1}^s \frac{1}{\sigma_i^2}. \tag{19}$$

When $s = 1$, (19) becomes $\|(\mathbf{F}'\mathbf{H}_\Lambda^*)^\dagger\|_F^2 = \frac{1}{\sigma_1^2}$. However, minimizing this quantity is the same as maximizing $\|\mathbf{F}'\mathbf{H}_\Lambda^*\|_F^2 = \sigma_1^2$. It is easy to see that $\|\mathbf{F}'\mathbf{H}_\Lambda^*\|_F^2$ is maximized when a measurement of $\mathbf{F}$ that is most correlated with $\mathbf{H}_\Lambda^*$ is chosen for every measurement in $\mathbf{F}'$. That is, if such a row can be identified, the best way to sense 1-sparse signals adaptively once the support is known is to simply repeat that measurement until the number of allotted measurements has been reached. Note that $\mathbf{F}'\mathbf{H}_\Lambda^*$ is $m \times 1$ in this case, and is still full rank when selecting the measurements in this

way; thus, the theory leading to (18) still holds. In this setting, we can determine explicitly what the MSE looks like, and provide bounds on the MSE that depend on the support location of the 1-sparse signal. The result is provided in Theorem 3.

Theorem 3. *Denote by* $\mathbf{x} = \mathbf{H}^* \boldsymbol{\alpha}$ *the signal of interest, and suppose* $\mathbf{x}$ *becomes 1-sparse after applying the Haar wavelet transformation* $\mathbf{H}$ *(that is,* $\mathbf{H}\mathbf{x} = \boldsymbol{\alpha}$ *and* $\|\boldsymbol{\alpha}\|_0 = 1$*), and let* $\operatorname{supp}(\boldsymbol{\alpha}) = \Lambda$*. Suppose after* $\frac{m}{2}$ *nonadaptive Fourier measurements, the support* Λ *is correctly identified. For the remaining* $\frac{m}{2}$ *Fourier measurements, we measure repeatedly with a particular measurement from the* $n \times n$ *DFT* $\mathbf{F}$ *defined in (6); denote this measurement by* $\mathbf{f}_j$*, where* $j \in \{0, 1, \ldots, n-1\}$ *is some row index. Then, our observations are of the form*

$$y_i = \langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \boldsymbol{\alpha}_\Lambda \rangle + z_i, \tag{20}$$

for $i = \frac{m}{2} + 1, \ldots, m$*, where the noise* z_i *are i.i.d.* $N(0, \sigma^2)$*. Then the MSE is given by*

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 = \frac{2/m}{|\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|^2} \sigma^2, \tag{21}$$

and is bounded by

$$\frac{2\sigma^2}{m} \le \mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \le \frac{n\sigma^2}{m}, \tag{22}$$

where the expectation is taken with respect to $\mathbf{z}$*.*

The upper bound on $\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2$ in Theorem 3 is precisely the lower bound from Theorem 1 when $s = 1$. This means that there is indeed some room for improvement with adaptive sensing when the sparsity basis is the Haar wavelet transform rather than the canonical basis.

Note that in standard compressive sensing when we rely on the RIP, the DFT matrix $\mathbf{F}$ is normalized by $\frac{1}{\sqrt{m}}$ rather than $\frac{1}{\sqrt{n}}$. If we make this normalization, the bound in (22) becomes

$$\frac{2\sigma^2}{n} \le \mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \le \sigma^2.$$

Including pre-conditioning and other scalings of course yields an analogous bound.

The proof of Theorem 3 relies on three lemmas that provide bounds for the term $|\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|$ appearing in the MSE in (21). Specifically, one term of interest is

$$\min_{\Lambda \in \{0, \ldots, n-1\}} \max_{j \in \{0, \ldots, n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|. \tag{23}$$

The maximization over j corresponds to selecting the best DFT measurement $\mathbf{f}_j$, and the minimization accounts for the worst case signal (i.e., the worst case support Λ). On the other hand, we also want to obtain a value that represents the best case signal so we are also interested in

$$\max_{\Lambda \in \{0, \ldots, n-1\}} \max_{j \in \{0, \ldots, n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|. \tag{24}$$

Before proving Theorem 3, let us set some notation. The Haar wavelet transform matrix $\mathbf{H}$, defined in (16) and (17) consists of *blocks* of consecutive rows with the same nonzero entry magnitudes. Let $1 \le a \le \log_2 n$ denote the block of $\mathbf{H}$, where $a = 1$ corresponds to the $\frac{n}{2}$ rows indexed by $j = \frac{n}{2}, \ldots, n-1$ (i.e., the "bottom" half of $\mathbf{H}$), $a = 2$ corresponds to the $\frac{n}{4}$ rows indexed by $j = \frac{n}{2} - \frac{n}{4}, \ldots, \frac{n}{2} - 1$, and so on. Similarly, for $\mathbf{H}^*$, instead of blocks of *rows*, we have blocks of *columns*; the block corresponding to $a = \log_2 n$ represents the lowest frequency wavelets, and the block corresponding to $a = 1$ represents the highest frequency wavelets.

Proof of Theorem 3. This proof requires the following three lemmas. Lemmas 1 and 2 are used to prove Lemma 3, and Lemma 3 is used to complete the proof of Theorem 3. The lemmas can be derived using elementary trigonometric bounds, and we omit the proofs here.

Lemma 1. *Fix $j \in \mathbb{Z}$ where $1 \le j \le n-1$ and let $a = 1, \ldots, \log_2 n$. Choose $k \in \mathbb{Z}$, $0 \le k \le n - \frac{2^a}{2}$. Then*

$$\left| \sum_{q=k}^{k+\frac{2^a}{2}-1} e^{-2\pi i j q/n} \right| = \sqrt{\frac{1 - \cos(\frac{2^a \pi j}{n})}{1 - \cos(\frac{2\pi j}{n})}}. \tag{25}$$

Lemma 2. *Let $\mathbf{f}_j$, $j \in \{0, \ldots, n-1\}$, be row j from the $n \times n$ DFT and let $\mathbf{H}_\Lambda^*$ be the inverse discrete Haar wavelet transform restricted to the column indexed by Λ. Let $a = 1, \ldots, \log_2 n$ denote the block of $\mathbf{H}^*$ and let $\Lambda \in \{1, 2, \ldots, n-1\}$, $|\Lambda| = 1$, be a column in the set corresponding to block a. Then,*

$$|\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle| = \frac{1}{\sqrt{n 2^{a-1}}} \frac{1 - \cos(\frac{2^a \pi j}{n})}{\sqrt{1 - \cos(\frac{2\pi j}{n})}}, \tag{26}$$

where $j = 1, \ldots, n-1$. When $j = 0$,

$$|\langle \mathbf{f}_0, \mathbf{H}_\Lambda^* \rangle| = \begin{cases} 1 & \Lambda = 0 \\ 0 & \Lambda \in \{1, \ldots, n-1\}. \end{cases}$$

When $\Lambda = \{0\}$,

$$|\langle \mathbf{f}_j, \mathbf{H}_0^* \rangle| = \begin{cases} 1 & j = 0 \\ 0 & j = 1, 2, \ldots, n-1. \end{cases}$$

Lemma 3. *Let $\mathbf{f}_j$, $j \in \{0, \ldots, n-1\}$, be a row from the $n \times n$ DFT matrix $\mathbf{F}$ and let $\mathbf{H}_\Lambda^*$ be the inverse discrete Haar wavelet transform restricted to the column indexed by Λ. Let $\Lambda \in \{0, 1, \ldots, n-1\}$ so that $|\Lambda| = 1$. Then*

$$\min_{\Lambda \in \{0, \ldots, n-1\}} \max_{j \in \{0, \ldots, n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle| = \sqrt{\frac{2}{n}} \tag{27}$$

and

$$\max_{\Lambda \in \{0, \ldots, n-1\}} \max_{j \in \{0, \ldots, n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle| = 1. \tag{28}$$

Since $|\Lambda| = 1$, $\boldsymbol{\alpha}_\Lambda$ is just a scalar, so that the measurements (20) can be written as

$$y_i = \langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle \boldsymbol{\alpha}_\Lambda + z_i. \tag{29}$$

This can be concisely written as

$$\mathbf{y} = \mathbf{A} \boldsymbol{\alpha}_\Lambda + \mathbf{z}, \tag{30}$$

where $\mathbf{A}$ is the $m/2$-dimensional column vector with each entry equal to $\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle$. To estimate $\boldsymbol{\alpha}_\Lambda$, we apply $\mathbf{A}^\dagger$ to $\mathbf{y}$. In this case, $\mathbf{A}^\dagger$ is an $m/2$-dimensional row vector, with each entry equal to $\frac{2}{m \langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle}$. Therefore,

$$\hat{\boldsymbol{\alpha}}_\Lambda = \mathbf{A}^\dagger \mathbf{y} = \mathbf{A}^\dagger (\mathbf{A}\boldsymbol{\alpha}_\Lambda + \mathbf{z})$$

$$= \left[\frac{2}{m\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle} \quad \cdots \quad \frac{2}{m\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle} \right] \left(\begin{bmatrix} \langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle \\ \vdots \\ \langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle \end{bmatrix} \boldsymbol{\alpha}_\Lambda + \mathbf{z} \right)$$

$$= \frac{2}{m} \sum_{i=1}^{m/2} \left(\boldsymbol{\alpha}_\Lambda + \frac{z_i}{\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle} \right) = \boldsymbol{\alpha}_\Lambda + \frac{2}{m} \sum_{i=1}^{m/2} \frac{z_i}{\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle}.$$

Using this, and since $\sum_{i=1}^{m/2} z_i \sim \mathcal{N}(0, \frac{m}{2}\sigma^2)$, we find

$$\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 = \mathbb{E}\|\mathbf{H}^*(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\|_2^2$$

$$= \mathbb{E}\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2^2 = \mathbb{E}|\hat{\boldsymbol{\alpha}}_\Lambda - \boldsymbol{\alpha}_\Lambda|^2$$

$$= \mathbb{E} \left| \boldsymbol{\alpha}_\Lambda - \left(\boldsymbol{\alpha}_\Lambda + \frac{2}{m} \sum_{i=1}^{m/2} \frac{z_i}{\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle} \right) \right|^2$$

$$= \mathbb{E} \left| \frac{2}{m} \sum_{i=1}^{m/2} \frac{z_i}{\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle} \right|^2$$

$$= \left(\frac{2/m}{|\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|} \right)^2 \mathbb{E} \left| \sum_{i=1}^{m/2} z_i \right|^2$$

$$= \left(\frac{2/m}{|\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|} \right)^2 \frac{m}{2}\sigma^2 = \frac{2/m}{|\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|^2}\sigma^2. \tag{31}$$

Applying the bounds from Lemma 3 to (31), we arrive at

$$\frac{2\sigma^2}{m} \le \mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \le \frac{2n\sigma^2}{m},$$

as desired. $\square$

Theorem 3 shows that the performance of adaptive sensing, in terms of the MSE, depends on the support location of 1-sparse signals. The best adaptive recovery is possible when the support is located in the first entry of the signal while the worst recovery occurs when the support is located anywhere on the bottom half of the signal. This of course matches the intuition based on the correlations in these two bases.

In light of the discrepancy between (27) and (28), one wishes to know in some sense, what fraction of signals allow for recovery more like one versus the other. Figure 4 shows how $\max_{j\in\{0,\ldots,n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|$ varies by maximizing $\max_{j\in\{0,\ldots,n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|$ over Λ while successively removing blocks from $\mathbf{H}^*$. Using our notation for blocks, the blocks of $\mathbf{H}^*$ are removed in the following (top-down) order: $\log_2(n), \log_2(n) - 1, \ldots, 1$. Then, we plot the value of $\max_{j\in\{0,\ldots,n-1\}} |\langle \mathbf{f}_j, \mathbf{H}_\Lambda^* \rangle|$ for the remaining submatrix of $\mathbf{H}^*$. Hence, we see that the MSE is higher for signals supported on higher wavelet frequencies, and the upper bound of (22) is achieved by exactly half of the possible signal support sets, whereas the lower bound of (22) is achieved by exactly one of the possible signal supports sets (i.e., $\Lambda = \{0\}$). Fortunately, the support of natural images tends to be concentrated on lower-frequency wavelet coefficients [23].
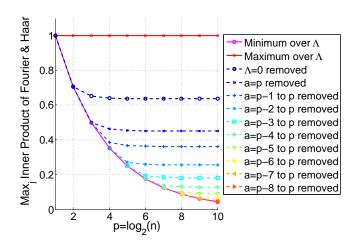
15

Figure 4: The value of $\max_{j \in \{0,\dots,n-1\}} |\langle \mathbf{f}_j, \mathbf{H}^*_\Lambda \rangle|$ is displayed against the (log of the) signal dimension $n = 2^p$. The solid magenta curve shows the optimization when minimizing over all possible supports $\Lambda \in \{0,\dots,n-1\}$, given by (27). The solid red curve shows the opposite optimization when maximizing over all possible supports $\Lambda \in \{0,\dots,n-1\}$, given by (28). The remaining dashed curves show the optimization when maximizing over all supports Λ except those in the blocks indicated.

5 Discussion

Adaptive sensing has tremendous potential to improve the accuracy of sparse recovery in a variety of settings. However, in many practical applications one does not have the freedom to choose arbitrary measurement vectors, but instead must choose from a specified pool of measurements. One example of particular interest is the setting where measurements must be taken from the Fourier ensemble, as is the case in many medical imaging applications. In this paper we established fundamental limitations on the improvements offered by adaptivity in this setting for certain sparsity bases. On the other hand, we argued that for other sparsity bases (such as the Haar wavelet basis) the role of adaptivity in the constrained setting is much less straightforward. We developed a sampling scheme which uses a simple optimization procedure to select measurements adapted to the signal support. This scheme results in significant improvements once an accurate estimate of the support is obtained, which in practice can be achieved by first dedicating a portion of the measurements to support estimation. Though this approach is not necessarily provably optimal, it nonetheless demonstrates the potential of adaptive sensing in the constrained setting. We believe future work in this area can further the understanding of both the limitations of this approach as well as the potential benefits.

Acknowledgment

The authors would like to thank Ran Zhao for his help in our theoretical analysis.

References

[1] E. Arias-Castro, E. Candès, and M. Davenport. On the fundamental limits of adaptive sensing. *to appear in IEEE Trans. Inform. Theory*, 2012.

[2] S. Becker, E. Candès, and M. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Math. Prog. Comput.*, 3(3):165–218, 2011.

[3] S. Becker, E. Candès, and M. Grant. Templates for first-order conic solvers user guide, version 1.3 release 2. Technical report, Stanford Univ., Technical report, 2014. Preprint available at http://tfocs.stanford.edu.

[4] S. Boyd and L. Vanderberghe. *Convex Optimization*. Cambridge Univ. Press, Cambridge, England, 2004.

[5] E. Candès. Compressive sampling. In *Proc. Int. Congress of Math.*, Madrid, Spain, August 2006.

[6] E. Candès and M. Davenport. How well can we estimate a sparse vector? *Appl. Comput. Harmon. Anal.*, 34(2):317–323, 2013.

[7] 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, 2006.

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

[9] R. Castro. Adaptive sensing performance lower bounds for sparse signal detection and support estimation. *Bernoulli*, 20(4):2217–2246, 2014.

[10] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using Hidden Markov Models. *IEEE Trans. Signal Processing*, 46(4):886–902, 1998.

[11] I. Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, PA, 1992.

[12] M. Davenport and E. Arias-Castro. Compressive binary search. In *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Cambridge, MA, July 2012.

[13] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok. Introduction to compressed sensing. In Yonina Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Cambridge University Press, Cambridge, UK, 2012.

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

[15] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Mag.*, 25(2):83–91, 2008.

[16] M. Duarte, M. Wakin, and R. Baraniuk. Fast reconstruction of piecewise smooth signals from random projections. In *Proc. Work. Struc. Parc. Rep. Adap. Signaux (SPARS)*, Rennes, France, November 2005.

[17] J. Haupt, R. Castro, and R. Nowak. Distilled sensing: Adaptive sampling for sparse detection and estimation. *IEEE Trans. Inform. Theory*, 57(9):6222–6235, 2011.

[18] P. Indyk, E. Price, and D. Woodruff. On the power of adaptivity in sparse recovery. In *Proc. IEEE Symp. Found. Comp. Science (FOCS)*, Palm Springs, CA, October 2011.

[19] F. Krahmer, D. Needell, and R. Ward. Compressive sensing with redundant dictionaries and structured measurements. 2015. Submitted.

[20] F. Krahmer and R. Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Processing*, 23(2):612–622, 2014.

[21] M. Malloy and R. Nowak. Near-optimal adaptive compressed sensing. *IEEE Trans. Inform. Theory*, 60(7):4001–4012, 2014.

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

[23] A. Petrosian and F. Meyer. *Wavelets in signal and image analysis: From theory to practice.* Springer Science & Business Media, 2001.

[24] F. Pukelsheim. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, 2006.

[25] E. van den Berg and M. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, June 2007. http://www.cs.ubc.ca/labs/scl/spgl1.

[26] E. van den Berg and M. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.