Beyond incoherence:
stable and robust sampling strategies for compressive imaging

Felix Krahmer and Rachel Ward

December 20, 2012

Abstract

In many signal processing applications, one wishes to acquire images that are sparse in transform domains such as spatial finite differences or wavelets using frequency domain samples. For such applications, overwhelming empirical evidence suggests that superior image reconstruction can be obtained through *variable density* sampling strategies that concentrate on lower frequencies. The wavelet and Fourier transform domains are not incoherent because low-order wavelets and low-order frequencies are correlated, so compressed sensing theory does not immediately imply sampling strategies and reconstruction guarantees. In this paper we turn to a more refined notion of coherence – the so-called *local coherence* – measuring for each sensing vector separately how correlated it is to the sparsity basis. For Fourier measurements and Haar wavelet sparsity, the local coherence can be controlled, so for matrices comprised of frequencies sampled from suitable power-law densities, we can prove the restricted isometry property with near-optimal embedding dimensions. Consequently, the variable-density sampling strategies we provide — which are independent of the ambient dimension up to logarithmic factors — allow for image reconstructions that are stable to sparsity defects and robust to measurement noise. Our results cover both reconstruction by ℓ_1-minimization and by total variation minimization.

1 Introduction

The measurement process in a wide range of imaging applications such as radar, sonar, astronomy, and computer tomography, can be modeled – after appropriate approximation and discretization – as taking samples from weighted discrete Fourier transforms [16]. Similarly, it is well known in the medical imaging literature that the measurements taken in Magnetic Resonance Imaging (MRI) are well modeled as Fourier coefficients of the desired image. Within all of these scenarios, one seeks strategies for taking frequency domain measurements so as to reduce the number of measurements without degrading the quality of image reconstruction. A central feature of natural images that can be exploited in this process is that they allow for approximately sparse representation in suitable bases or dictionaries.

The theory of compressed sensing, also termed compressive sampling, as introduced in [14, 8], fits comfortably into this set-up: its key observation is that signals which allow for a sparse or approximately sparse representation can be recovered from relatively few linear measurements via convex approximation, provided these measurements are sufficiently *incoherent* with the basis in which the signal is sparse.

1.1 Imaging with partial frequency measurements

Much work in compressed sensing has focused on the setting of imaging with frequency domain measurements [8, 7, 34] and, in particular, towards accelerating the MRI measurement process [22, 21]. For images that are localized in the pixel domain, the incoherence between pixel and Fourier bases implies that uniformly subsampled discrete Fourier transform measurements can be used to achieve near-optimal oracle reconstruction bounds: up to logarithmic factors in the discretization size, any image can be approximated from s such frequency measurements up to the error that would be incurred if the image were first acquired in full, and then compressed by setting all but the s largest-magnitude pixels to zero [34, 31, 7, 9]. Compressed sensing recovery guarantees hold more generally subject to incoherence between sampling and sparsity transform domains. Unfortunately, natural images are generally *not* directly sparse in the pixel basis, but rather with respect to transform domains more closely resembling wavelet bases. As low-scale wavelets are highly correlated (coherent) with low frequencies, sampling theorems for compressive imaging with partial Fourier transform measurements have remained elusive.

A number of empirical studies, including the very first papers on compressed sensing MRI [22, 21], suggest that better image restoration is possible by subsampling frequency measurements from variable densities preferring low frequencies to high frequencies. In fact, variable-density MRI sampling had been proposed previously in a number of works outside the context of compressed sensing, although there did not seem to be a consensus on an optimal density [23, 36, 38, 28, 18]. For a thorough empirical comparison of different densities from the compressed sensing perspective, see [39].

Guided by these observations, the authors of [30] estimated the coherence between each element of the sensing basis with elements of the sparsity basis separately as a means to derive optimal sampling strategies in the context of compressed sensing MRI. In particular, they observe that incoherence-based results in compressed sensing imply exact recovery results for more general systems if one samples a row from the measurement basis proportionally to its squared maximal correlation with the sparsity-inducing basis. For given problem dimensions, they find the optimal distribution as the solution to a convex problem. In [29], the approach of optimizing the sampling distribution is combined with premodulation by a chirp, which by itself is another measure to reduce the coherence [3, 20]. A similar variable-density analysis already appeared in [33] in the context of sampling strategies and reconstruction guarantees for functions with sparse orthogonal polynomial expansions and will also be the guiding strategy of this paper (cf. Section 5).

1.2 Contributions of this paper

In this paper we derive near-optimal reconstruction bounds for variable-density subsampled discrete Fourier transform measurements for both wavelet sparsity and gradient sparsity models. More precisely, up to logarithmic factors in the discretization size, any image can be approximated from s such measurements up to the error that would be incurred if the wavelet transform or the gradient of the image, respectively, were first computed in full, and then compressed by setting all but the s largest-magnitude coefficients to zero. Note that the reconstruction results that have been derived for uniformly-subsampled frequency measurements [8, 15] are not optimal in this sense.

A major role in determining the best sampling density will be played by the *local coherence* of the sensing basis with respect to the sparsity basis, as introduced in Section 5. Consequently, an important ingredient of our analysis is Theorem 6.2, which provides frequency-dependent bounds on inner products between rows of the orthonormal discrete Fourier transform and rows of the orthonormal discrete Haar wavelet transform. In particular, the maximal correlation between a *fixed* row in the discrete Fourier transform and *any* row of the discrete Haar wavelet transform decreases according to an inverse power law of the frequency, and decays sufficiently quickly that the sum of squared maximal correlations scales only logarithmically with the discretization size N. This implies, according to the techniques used in [33, 32, 4], that subsampling rows of the discrete Fourier matrix proportionally to the squared correlation results in a matrix that has the restricted isometry property of near-optimal order subject to appropriate rescaling of the rows.

For reconstruction, total variation minimization [35, 5, 27, 37, 11, 10] will be our algorithm of choice. In the papers [25] and [26], total variation minimization was shown to provide stable and robust image reconstruction provided that the sensing matrix is incoherent with the Haar wavelet basis. Following the approach of [25], we prove that from variable density frequency samples, total variation minimization can be used for stable image recovery guarantees. Due to the effects of variable density sampling we will no longer need the samples to be incoherent to the Haar basis.

1.3 Outline

The remainder of this paper is organized as follows. Preliminary notation is introduced in Section 2. The main results of this paper along with a numerical illustration are contained in Section 3. Section 4 reviews compressed sensing theory and Section 5 presents recent results on sampling strategies for coherent systems. The main results on the coherence between Fourier and Haar wavelet bases is provided in Section 6, and proofs of the main results are contained in Section 7. We conclude with a summary and a discussion of open problems in Section 8.

2 Preliminaries

2.1 Notation

In this paper, we consider discrete images, that is, $N \times N$ blocks of pixels, and represent them as discrete functions $f \in \mathbb{C}^{N\times N}$. We write $f(t_1, t_2)$ to denote any particular pixel value and for a positive integer N, we denote the set $\{1, 2, \ldots, N\}$ by $[N]$. By $f_1 \circ f_2$ we denote the Hadamard product, i.e., the image resulting from pointwise products of the pixel values, $f_1 \circ f_2(t_1, t_2) = f_1(t_1, t_2)f_2(t_1, t_2)$. On the space of such images, the ℓ_p vector norm is given by $\|f\|_p = \left(\sum_{t_1,t_2} |f(t_1, t_2)|^p\right)^{1/p}, 1 \le p < \infty$, and $\|f\|_\infty = \max_{(t_1,t_2)} |f(t_1, t_2)|$. The inner product inducing the ℓ_2 vector norm is $\langle f, g\rangle = \sum_{t_1,t_2} \bar{f}(t_1, t_2)g(t_1, t_2)$, where $\bar{z}$ denotes the complex conjugate of number $z \in \mathbb{C}$. By an abuse of notation, the "ℓ_0-norm" $\|f\|_0 = \#\{(t_1, t_2) : f(t_1, t_2) \neq 0\}$ counts the number of non-zero entries of f.

An image f is called s-sparse if $\|f\|_0 \le s$. The error of best s-term approximation of an image f in ℓ_p is defined as

$$\sigma_s(f)_p = \inf_{g:\|g\|_0\le s} \|f - g\|_p.$$

Clearly, $\sigma_s(f)_p = 0$ if f is s-sparse. Informally, f is called compressible if $\sigma_s(f)_1$ decays quickly as s increases.

For two nonnegative functions $f(t)$ and $g(t)$ on the real line, we write $f \gtrsim g$ (or $f \lesssim g$) if there exists a constant $C > 0$ such that $f(t) \ge Cg(t)$ (or $f(t) \le Cg(t)$, respectively) for all $t > 0$.

The discrete directional derivatives of $f \in \mathbb{C}^{N\times N}$ are defined pixel-wise as

$$\begin{aligned} f_x \in \mathbb{C}^{N-1\times N}, \qquad f_x(t_1, t_2) &= f(t_1 + 1, t_2) - f(t_1, t_2) \\ f_y \in \mathbb{C}^{N\times N-1}, \qquad f_y(t_1, t_2) &= f(t_1, t_2 + 1) - f(t_1, t_2) \end{aligned}$$

The discrete gradient transform $\nabla : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N\times 2}$ is defined in terms of the directional derivatives via

$$\nabla f(t_1, t_2) := \Big(f_x(t_1, t_2),\ f_y(t_1, t_2)\Big),$$

where the directional derivatives are extended to $N \times N$ by adding zero entries. The *total variation* semi-norm is the ℓ_1 norm of the image gradient,

$$\|f\|_{TV} := \|\nabla f\|_1 = \sum_{t_1,t_2} \big(|f_x(t_1, t_2)| + |f_y(t_1, t_2)|\big).$$

Here we note that our definition is the *anisotropic* version of the total variation semi-norm. The *isotropic* total variation semi-norm becomes the sum of terms

$$\left|f_x(t_1,t_2)+if_y(t_1,t_2)\right| = \left(f_x(t_1,t_2)^2 + f_y(t_1,t_2)^2\right)^{1/2}.$$

The isotropic and anisotropic total variation semi-norms are thus equivalent up to a factor of $\sqrt{2}$.

2.2 Bases for sparse representation and measurements

The Haar wavelet basis is a simple basis which allows for good sparse approximations of natural images. We will work primarily in two dimensions, but first introduce the univariate Haar wavelet basis as it will nevertheless serve as a building block for higher dimensional bases.

Definition 2.1 (Univariate Haar wavelet basis)**.** The univariate discrete Haar wavelet system is an orthonormal basis of $\mathbb{C}^{2^p}$ consisting of the constant function $h^0(t) \equiv 2^{-p/2}$, the step function $h^1_{0,0} = h^1$ given by

$$h^1(t) = \begin{cases} 2^{-p/2}, & 1 \le t \le 2^{p-1}, \\ -2^{-p/2}, & 2^{p-1} < t \le 2^p, \end{cases}$$

along with the dyadic step functions

$$\begin{aligned} h^1_{n,\ell}(t) &= 2^{\frac{n}{2}} h^1(2^n t - \ell) \\ &= \begin{cases} 2^{\frac{n-p}{2}} & \text{for} \quad \ell 2^{p-n} \le t < \ell 2^{p-n} + 2^{p-n-1} \\ -2^{\frac{n-p}{2}} & \text{for} \quad \ell 2^{p-n} + 2^{p-n-1} \le t < \ell 2^{p-n} + 2^{p-n} \\ 0 & \text{else,} \end{cases} \end{aligned}$$

for $(n,\ell) \in \mathbb{Z}^2$ satisfying $0 < n < p$ and $0 \le \ell < 2^n$.

To define the bivariate Haar wavelet basis of $\mathbb{C}^{2^p \times 2^p}$, we extend the univariate system by the window functions

$$h^0_{n,\ell}(t) = 2^{\frac{n}{2}} h^0(2^n t - \ell) = \begin{cases} 2^{\frac{n-p}{2}} & \text{for} \quad \ell 2^{p-n} \le t < (\ell+1) 2^{p-n} \\ 0 & \text{else.} \end{cases}$$

The bivariate Haar wavelet system can now be defined via tensor products of functions in the extended univariate system. In order for the system to form an orthonormal basis of $\mathbb{C}^{2^p \times 2^p}$, only tensor products of univariate functions with the same scaling parameter n are included.

Definition 2.2 (Bivariate Haar wavelet basis)**.** The bivariate Haar system of $\mathbb{C}^{2^p \times 2^p}$ consists of the constant function $h^{(0,0)}$ given by

$$h^{(0,0)}(t_1, t_2) = h^0(t_1)h^0(t_2) \equiv 2^{-p}$$

and the functions $h^e_{n,\ell}$ with indices in the range $0 \le n < p$, $\ell = (\ell_1, \ell_2) \in \mathbb{Z}^2 \cap [0, 2^n)^2$, and $e = (e_1, e_2) \in \big\{\{0,1\}, \{1,0\}, \{1,1\}\big\}$ given by

$$h^e_{n,\ell}(t_1, t_2) = h^{e_1}_{n,\ell_1}(t_1)h^{e_2}_{n,\ell_2}(t_2).$$

We denote by $\mathcal{H}$ the bivariate Haar transform $f \to \big(\langle f, h_{n,\ell}\rangle\big)_{n,\ell}$ and, by a slight abuse of notation, also the unitary matrix representing this linear map.

We will also work with discrete Fourier measurements.

Definition 2.3 (Discrete Fourier basis)**.** Let $N = 2^p$. The one-dimensional discrete Fourier system is an orthonormal basis of $\mathbb{C}^N$ consisting of the vectors

$$\varphi_k(t) = \frac{1}{\sqrt{N}} e^{i2\pi tk/N}, \quad -N/2 + 1 \le t \le N,$$

indexed by discrete frequencies in the range $-N/2+1 \le k \le N/2$. The two-dimensional discrete Fourier basis of $\mathbb{C}^{N\times N}$ is just a tensor product of one-dimensional bases, namely

$$\varphi_{k_1,k_2}(t_1, t_2) = \frac{1}{N} e^{i2\pi(t_1k_1+t_2k_2)/N}, \quad -N/2 + 1 \le t_1, t_2 \le N/2, \tag{2.3}$$

indexed by discrete frequencies in the range $-N/2 + 1 \le k_1, k_2 \le N/2$.

We denote by $\mathcal{F}$ the two-dimensional discrete Fourier transform $f \to \big(\langle f, \varphi_{k_1,k_2}\rangle\big)_{k_1,k_2}$ and, again, also the associated unitary matrix. Finally, we denote by $\mathcal{F}_\Omega$ its restriction to a set of frequencies $\Omega \subset [N]^2$.

3 Main results

Our main results say that appropriate variable density subsampling of the discrete Fourier transform will with high probability result in a set of measurements admitting stable image reconstruction via total variation minimization or ℓ_1-minimization.

While our recovery guarantees are robust to measurement noise, the model we consider corresponds to a *weighted* ℓ_2-norm such that high-frequency measurements have higher sensitivity to noise. Loosely speaking, this compensates for the smaller number of samples in regions of low sampling probability. Our first result concerns stable recovery guarantees for total variation minimization.

Theorem 3.1. *Fix integers $N = 2^p, m,$ and s such that $s \gtrsim \log(N)$ and*

$$m \gtrsim s \log^3(s) \log^5(N).$$

Select m frequencies $\{(\omega_1^j, \omega_2^j)\}_{j=1}^m \subset \{-N/2+1, \dots, N/2\}^2$ i.i.d. according to

$$\operatorname{Prob}\big[(\omega_1^j, \omega_2^j) = (k_1, k_2)\big] = C_N \min\left(C, \frac{1}{k_1^2 + k_2^2}\right) =: \eta(k_1, k_2), \quad -N/2+1 \le k_1, k_2 \le N/2, \tag{3.1}$$

where C is an absolute constant and C_N is chosen such that η is a probability distribution.
Consider the weight vector $\rho = (\rho_j)_{j=1}^m$ with $\rho_j = (1/\eta(\omega_1^j, \omega_2^j))^{1/2}$, and assume that the noise vector $\xi = (\xi_j)_{j=1}^m$ satisfies $\|\rho \circ \xi\|_2 \le \varepsilon\sqrt{m}$, for some $\epsilon > 0$. Then with probability exceeding $1 - N^{-C\log^3(s)}$, the following holds for all images $f \in \mathbb{C}^{N \times N}$:
Given noisy partial Fourier measurements $y = \mathcal{F}_\Omega f + \xi$, the estimation

$$f^\# = \operatorname*{argmin}_{g \in \mathbb{C}^{N\times N}} \|g\|_{TV} \quad \textit{such that} \quad \|\rho \circ (\mathcal{F}_\Omega g - y)\|_2 \le \varepsilon\sqrt{m}, \tag{3.2}$$

approximates f up to the noise level and best s-term approximation error of its gradient:

$$\|f - f^\#\|_2 \lesssim \frac{\|\nabla f - (\nabla f)_s\|_1}{\sqrt{s}} + \varepsilon.$$

Disregarding measurement noise, the error rate provided in Theorem 3.1 (and also the one in Theorem 3.2 below) is optimal up to logarithmic factors in the ambient image dimension. This follows from classical results about the Gel'fand width of the ℓ_1-ball due to Kashin [19] and Garnaev–Gluskin [17].

Numerical results such as those detailed in [39] and illustrated below in Figure 1 confirm that variable-density sampling strategies significantly outperform uniform sampling strategies as well as deterministic sampling strategies, and Theorem 3.1 provides theoretical justification for such observations.

Our second result focuses on stable image reconstruction by ℓ_1-minimization in the Haar wavelet transform domain. It is a direct consequence of applying the Fourier-wavelet incoherence estimates derived in Theorem 5.2 to Theorem 6.2.

Theorem 3.2. *Fix integers $N = 2^p, m,$ and s such that $s \gtrsim \log(N)$ and*

$$m \gtrsim s \log^2(N) \log^3(s).$$

Select m frequencies $\Omega = \{(\omega_1^j, \omega_2^j)\}_{j=1}^m \subset \{-N/2+1, \dots, N/2\}^2$ i.i.d. according to the density η as in (3.1) *and assume again that the noise vector $\xi = (\xi_j)_{j=1}^m$ satisfies the weighted ℓ_2-constraint with weight ρ and noise level ε as in Theorem 3.1. Then with probability exceeding $1 - N^{-C\log^3(s)}$, the following holds* *for all images $f \in \mathbb{C}^{N\times N}$: Given noisy measurements $y = \mathcal{F}_\Omega f + \xi$, the estimation*

$$f^\# = \operatorname*{argmin}_{g\in\mathbb{C}^{N\times N}} \|\mathcal{H}g\|_1 \quad \textit{such that} \quad \|\rho \circ (\mathcal{F}_\Omega g - y)\|_2 \le \varepsilon\sqrt{m}$$

approximates f up to the noise level and best s-term approximation error in the bivariate Haar basis:

$$\|f - f^\#\|_2 \lesssim \frac{\|\mathcal{H}f - (\mathcal{H}f)_s\|_1}{\sqrt{s}} + \varepsilon.$$

Even though the required number of samples m in Theorem 3.2 is smaller than the number of samples required for the total variation minimization guarantees in Theorem 3.1, we find that total variation minimization requires fewer measurements empirically. This may be due to the fact that the gradient of a natural image has stronger sparsity than its Haar wavelet representation. For this reason we focus on total variation minimization. Independent of this observation, we strongly suspect that the additional logarithmic factors in the number of measurements stated in Theorem 3.1 are an artifact of the proof, and that it should be possible to strengthen the result to obtain a similar recovery guarantee with the number of measurements as in Theorem 3.2.

4 Compressed sensing background

4.1 The restricted isometry property

Under very mild assumptions on the matrix $\Phi : \mathbb{C}^N \to \mathbb{C}^m$, any k-sparse $x \in \mathbb{C}^N$ can be recovered from $y = \Phi x$ as the solution to the optimization problem:

$$x = \operatorname{argmin} \|z\|_0 \quad \text{such that} \quad \Phi z = y$$

One of the fundamental results in compressed sensing is that this optimization problem, which is NP-hard in general, can be relaxed to an ℓ_1-minimization problem if one asks that the matrix Φ restricted to any subset of $2k$ columns be well-conditioned. This property is quantified via the so-called *restricted isometry property* as introduced in [9]:

Definition 4.1 (Restricted isometry property)**.** Let $\Phi \in \mathbb{C}^{m\times N}$. For $s \le N$, the restricted isometry constant δ_s associated to Φ is the smallest number δ for which

$$(1-\delta)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+\delta)\|x\|_2^2$$

for all s-sparse vectors $x \in \mathbb{C}^N$. If $\delta_s \le \delta$, one says that Φ has the restricted isometry property (RIP) of order s and level δ.

The restricted isometry property ensures stability: not only sparse vectors, but also compressible vectors can be recovered from the measurements via ℓ_1-minimization. It also ensures robustness to measurement errors.

Proposition 4.2 (Sparse recovery for RIP matrices)**.** *Assume that the restricted isometry constant* δ_{5s} *of* $\Phi \in \mathbb{C}^{m\times N}$ *satisfies* $\delta_{5s} < \frac{1}{3}$*. Let* $x \in \mathbb{C}^N$ *and assume noisy measurements* $y = \Phi x + \xi$ *with* $\|\xi\|_2 \leq \varepsilon$*. Then*

$$x^{\#} = \arg\min_{z\in\mathbb{C}^N} \quad \|z\|_1 \text{ subject to } \quad \|\Phi z - y\|_2 \leq \varepsilon$$

satisfies

$$\|x - x^{\#}\|_2 \leq \frac{2\sigma_s(x)_1}{\sqrt{s}} + \varepsilon.$$

In particular, reconstruction is exact, $x^{\#} = x$*, if* x *is* s*-sparse and* $\varepsilon = 0$*.*

There are stronger versions of this result which allow for weaker constraints on the restricted isometry constant [24]. However, our version is a corollary of the following proposition, which generalizes the results from [8] and seem to have appeared first in [25]. This proposition will also play an important role in the proof of our main results.

Proposition 4.3 (Stable recovery for RIP matrices, [25])**.** *Suppose that* $\gamma \geq 1$ *and* $\Phi \in \mathbb{C}^{m\times N}$ *satisfies the restricted isometry property of order* $5k\gamma^2$ *and level* $\delta < 1/3$*, and suppose that* $u \in \mathbb{C}^N$ *satisfies a tube constraint*

$$\|\Phi u\|_2 \lesssim \varepsilon.$$

Suppose further that for a subset S *of cardinality* $|S| = k$*, the signal* u *satisfies a cone constraint*

$$\|u_{S^c}\|_1 \leq \gamma \|u_S\|_1 + \xi.$$

Then

$$\|u\|_2 \lesssim \frac{\xi}{\gamma\sqrt{k}} + \varepsilon. \tag{4.1}$$

Proposition 4.2 follows from Proposition 4.3 by noting that the minimality of $x^{\#}$ implies a cone constraint for the residual $x - x^{\#}$ over the support of the s largest-magnitude entries of x. For completeness, the proof of Proposition 4.3 is recalled in the appendix.

4.2 Bounded orthonormal systems

While the strongest known results on the restricted isometry property concern random matrices with independent entries such as Gaussian or Bernoulli, a scenario that has proven particularly useful for applications is that of structured random matrices with rows chosen from a basis incoherent to the basis inducing sparsity (see below for a detailed discussion on the concept of incoherence). The resulting sampling schemes correspond to *bounded orthonormal systems*, and such systems have been extensively studied in the compressed sensing literature (see [31] for an expository article including many references).

Definition 4.4 (Bounded orthonormal system)**.** Consider a set T equipped with probability measure ν.

- A set of functions $\{\psi_j : T \to \mathbb{C}, j \in [N]\}$ is called an *orthonormal system with respect to* ν if $\int_T \bar{\psi}_j(x)\psi_k(x)d\nu(x) = \delta_{jk}$, where δ_{jk} denotes the Kronecker delta.
- An orthonormal system is said to be *bounded* with bound K if $\sup_{j\in[N]} \|\psi_j(x)\|_\infty \leq K$.

For example, the basis of complex exponentials $\psi_j(x) = \exp(i2\pi jx)$ forms a bounded orthonormal system with optimally small constant $K = 1$ with respect to the uniform measure on $T = \{0, \frac{1}{N}, \dots, \frac{N-1}{N}\}$, and d-dimensional tensor products of complex exponentials form bounded orthonormal systems with respect to the uniform measure on the set T^d. A *random sample* of an orthonormal system is the vector $(\psi_1(x), \dots, \psi_N(x))$, where x is a random variable drawn according to the associated distribution ν. Any matrix whose rows are independent random samples of a bounded orthonormal system, such as the uniformly subsampled discrete Fourier matrix, will have the restricted isometry property:

Proposition 4.5 (RIP for bounded orthonormal systems, [31])**.** *Consider the matrix $\Psi \in \mathbb{C}^{m\times N}$ whose rows are independent random samples of a orthonormal system $\{\psi_j,\ j \in [N]\}$ with bound $K \geq 1$ and orthogonalization measure ν. If*

$$m \gtrsim \delta^{-2} K^2 s \log^3(s) \log(N),$$

for some $s \gtrsim \log(N)$[1]*, then with probability at least $1 - N^{-C\log^3(s)}$, the restricted isometry constant δ_s of $\frac{1}{\sqrt{m}}\Psi$ satisfies $\delta_s \leq \delta$.*

An important special case of a bounded orthonormal system arises by sampling a function with sparse representation in one basis using measurements from a different, incoherent basis. The *mutual coherence* between a unitary matrix $A \in \mathbb{C}^{N\times N}$ with rows $(a_j)_{j=1}^N$ and a unitary matrix $B \in \mathbb{C}^{N\times N}$ with rows $(b_j)_{j=1}^N$ is given by

$$\mu(A, B) = \sup_{j,k} |\langle a_j, b_k \rangle|$$

[1] For matrices consisting of uniformly subsampled rows of the discrete Fourier matrix, it has been shown in [12] that this constraint is not necessary.

The smallest possible mutual coherence is $\mu = N^{-1/2}$, as realized by the discrete Fourier matrix and the identity matrix. We call two orthonormal bases A and B *mutually incoherent* if $\mu = O(N^{-1/2})$ or $\mu = O(\log^{\alpha}(N)N^{-1/2})$. In this case, the rows $(\tilde{b}_j)_{j=1}^N$ of the basis $\widetilde{B} = \sqrt{N}BA^*$ constitute a bounded orthonormal system with respect to the uniform measure. Propositions 4.5 and 4.2 then imply that, with high probability, signals $f \in \mathbb{C}^N$ of the form $f = Ax$ for x sparse can be stably reconstructed from uniformly subsampled measurements $y = Bf = \widetilde{B}x$, as $\widetilde{B}$ has the restricted isometry property.

Corollary 4.6 (RIP for incoherent systems, [34])**.** *Consider orthonormal bases $A, B \in \mathbb{C}^{N\times N}$ with mutual coherence bounded by $\mu(A, B) \leq KN^{-1/2}$. Fix $\delta > 0$ and integers N, m, and s such that $s \gtrsim \log(N)$ and*

$$m \gtrsim \delta^{-2}K^2 s \log^3(s)\log(N). \tag{4.2}$$

Consider the matrix $\Phi \in \mathbb{C}^{m\times N}$ formed by uniformly subsampling m rows i.i.d. from the the matrix $\widetilde{B} = \sqrt{N}BA^*$. Then with probability at least $1 - N^{-c\log^3(s)}$, the restricted isometry constant δ_s of $\frac{1}{\sqrt{m}}\Phi$ satisfies $\delta_s \leq \delta$.

5 Local coherence

The coherence-based sparse recovery results implied by Corollary 4.6 do not take advantage of the full range of applicability of bounded orthonormal systems. As argued in [33], Proposition 4.5 implies comparable sparse recovery guarantees for a much wider class of sampling/sparsity bases through preconditioning resampled systems. In the following, we formalize this approach through the notion of local coherence.

Definition 5.1 (Local coherence)**.** The *local coherence* of an orthonormal basis $\{\varphi_j\}_{j=1}^N$ of $\mathbb{C}^N$ with respect to the orthonormal basis $\{\psi_k\}_{k=1}^N$ of $\mathbb{C}^N$ is the function $\mu^{loc}(\Phi, \Psi) \in \mathbb{R}^N$ defined coordinate-wise by

$$\mu_j^{loc}(\Phi, \Psi) = \sup_{1\leq k\leq N} |\langle \varphi_j, \psi_k\rangle|.$$

The following result shows that we can reduce the number of measurements m in (4.6) by replacing the bound K on the coherence in (4.2) by a bound on the ℓ_2-norm of the local coherence, provided we sample rows from Φ appropriately using the local coherence function. It can be seen as a direct finite-dimensional analog to Theorem 2.1 in [33], but for completeness, we include a short self-contained proof.

Theorem 5.2. *Let $\Phi = \{\varphi_j\}_{j=1}^N$ and $\Psi = \{\psi_k\}_{k=1}^N$ be orthonormal bases of $\mathbb{C}^N$. Assume the local coherence of Φ with respect to Ψ is pointwise bounded by the function κ, that is $\sup_{1\leq k\leq N} |\langle \varphi_j, \psi_k\rangle| \leq \kappa_j$. Let $s \gtrsim \log(N)$, suppose*

$$m \gtrsim \delta^{-2}\|\kappa\|_2^2 s \log^3(s)\log(N),$$

and choose m (possibly not distinct) indices $j \in \Omega \subset [N]$ i.i.d. from the probability measure ν on $[N]$ given by

$$\nu(j) = \frac{\kappa_j^2}{\|\kappa\|_2^2}.$$

Consider the matrix $A \in \mathbb{C}^{m \times N}$ with entries

$$A_{j,k} = \langle \varphi_j, \psi_k \rangle, \quad j \in \Omega, k \in [N],$$

and consider the diagonal matrix $D = \operatorname{diag}(d) \in \mathbb{C}^N$ with $d_j = \|\kappa\|_2/\kappa_j$. Then with probability at least $1 - N^{-c\log^3(s)}$, the restricted isometry constant δ_s of the preconditioned matrix $\frac{1}{\sqrt{m}}DA$ satisfies $\delta_s \leq \delta$.

Proof. We show that the system $\{\widetilde{\varphi}_j\} = \{d_j\varphi_j\}$ is an orthonormal system with respect to ν in the sense of Definition 4.4. Indeed,

$$\begin{aligned}\sum_{j=1}^{N} \widetilde{\varphi}_j(k_1)\widetilde{\varphi}_j(k_2)\nu(j) &= \sum_{j=1}^{N} \Big(\frac{\|\kappa\|_2}{\kappa_j}\varphi_j(k_1)\Big)\Big(\frac{\|\kappa\|_2}{\kappa_j}\varphi_j(k_2)\Big)\frac{\kappa_j^2}{\|\kappa\|_2^2} \\ &= \sum_{j=1}^{N} \varphi_j(k_1)\varphi_j(k_2) = \delta_{k_1,k_2};\end{aligned}$$

hence the $\widetilde{\varphi}_j$ form an orthonormal system with respect to ν. Noting that this system is bounded with bound $\|\kappa\|_2$, the result follows from Proposition 4.5. □

Remark 5.3. Note that the local coherence not only appears in the embedding dimension m, but also in the sampling measure. Hence a priori, one cannot guarantee the optimal embedding dimension if one only has suboptimal bounds for the local coherence. That is why the sampling measure in Theorem 5.2 is defined via the (known) upper bounds κ and $\|\kappa\|_2$ rather than the (usually unknown) exact values μ_{loc} and $\|\mu_{loc}\|_2$, showing that suboptimal bounds still lead to meaningful bounds on the embedding dimension.

Remark 5.4. For $\mu \leq KN^{-1/2}$ (as in Corollary 4.6), one has $\|\mu^{loc}\|_2 \leq K$, so Theorem 5.2 is a direct generalization of Corollary 4.6. As one has equality if and only if μ^{loc} is constant, however, Theorem 5.2 will be stronger in most cases.

6 Local coherence estimates for frequencies and wavelets

Due to the tensor product structure of both of these bases, the two-dimensional local coherence of the two-dimensional Fourier basis with respect to bivariate Haar wavelets will follow by first bounding the local coherence of the one-dimensional Fourier basis with respect to the set of univariate building block functions of the bivariate Haar basis.

Lemma 6.1. *Fix $N = 2^p$ with $p \in \mathbb{N}$. For the space $\mathbb{C}^N$, the one-dimensional Fourier basis vectors φ_k, $k \neq 0$, and the one-dimensional Haar wavelet basis building blocks $h^e_{n,k}$, $e = 0, 1$, satisfy the incoherence relation*

$$|\langle \varphi_k, h^e_{n,\ell} \rangle| \leq \min\Big(\frac{6 \cdot 2^{\frac{n}{2}}}{|k|}, 3\pi 2^{-\frac{n}{2}}\Big).$$

Proof. We estimate

$$\begin{aligned}
\langle \varphi_k, h^e_{n,\ell} \rangle &= \sum_{j=2^{p-n}\ell}^{2^{p-n}\ell+2^{p-n-1}-1} 2^{\frac{n-p}{2}} 2^{-\frac{p}{2}} e^{2\pi i 2^{-p} kj} + (-1)^e \sum_{j=2^{p-n}\ell+2^{p-n-1}}^{2^{p-n}\ell+2^{p-n}-1} 2^{\frac{n-p}{2}} 2^{-\frac{p}{2}} e^{2\pi i 2^{-p} kj} \\
&= e^{2\pi i 2^{-n}\ell k} \Big(1 + (-1)^e e^{2\pi i 2^{-n-1}k}\Big) 2^{\frac{n}{2}-p} \sum_{j=0}^{2^{p-n-1}-1} e^{2\pi i 2^{-p} kj} \\
&= e^{2\pi i 2^{-n}\ell k} \Big(1 + (-1)^e e^{2\pi i 2^{-n-1}k}\Big) 2^{\frac{n}{2}-p} \frac{1 - e^{2\pi i 2^{-n-1}k}}{1 - e^{2\pi i 2^{-p}k}}.
\end{aligned}$$

To estimate this expression, we note that

$$|1 - e^{2\pi i 2^{-n-1}k}| \leq \min(2, \pi 2^{-n}|k|) \tag{6.1}$$

and distinguish two cases:

If $0 \neq |k| \leq 2^{p-2}$, we bound $|1 - e^{2\pi i 2^{-p}k}| \geq 2^{-p}|k|$ and apply (6.1) to obtain

$$\begin{aligned}
|\langle \varphi_k, h^e_{n,\ell} \rangle| &\leq 2 \cdot 2^{\frac{n}{2}-p} \frac{\min(2, \pi 2^{-n}|k|)}{2^{-p}|k|} \\
&\leq \min(\frac{4 \cdot 2^{\frac{n}{2}}}{|k|}, 2\pi 2^{-\frac{n}{2}}).
\end{aligned}$$

For $2^{p-2} < |k| \leq 2^{p-1}$, and hence $2^{-p} \leq \frac{1}{2}|k|^{-1}$, we note that $|1 - e^{2\pi i 2^{-p}k}| \geq \frac{\sqrt{2}}{2}$ and bound, again using (6.1),

$$\begin{aligned}
|\langle \varphi_k, h^e_{n,\ell} \rangle| &\leq 2 \cdot 2^{\frac{n}{2}} |k|^{-1} \frac{\min(2, \pi 2^{-n}|k|)}{\frac{\sqrt{2}}{2}} \\
&\leq \min\Big(\frac{6 \cdot 2^{\frac{n}{2}}}{|k|}, 3\pi 2^{-\frac{n}{2}}\Big).
\end{aligned}$$

□

This lemma enables us to derive the following incoherence estimates for the bivariate case.

Theorem 6.2. *Let $N = 2^p$ for $\mathbb{N} \ni p \geq 8$. Then the local coherence μ^{loc} of the orthonormal two-dimensional Fourier basis $\{\varphi_{k_1,k_2}\}$ with respect to the orthonormal bivariate Haar wavelet basis $\{h^e_{n,\ell}\}$ in*

$\mathbb{C}^{N\times N}$, *as defined in* (2.3) *and* (2.2), *respectively, is bounded by*

$$\mu^{loc}_{k_1,k_2} \leq \kappa(k_1,k_2) := \min\left(1, \frac{18\pi}{\max(|k_1|,|k_2|)}\right)$$
$$\leq \kappa'(k_1,k_2) := \min\left(1, \frac{18\pi\sqrt{2}}{(|k_1|^2+|k_2|^2)^{1/2}}\right),$$

and one has $\|\kappa\|_2 \leq \|\kappa'\|_2 \leq 52\sqrt{p} = 52\sqrt{\log_2(N)}$.

Proof. First note that the bivariate Fourier coefficients decompose into the product of univariate Fourier coefficients:

$$\langle \varphi_{k_1,k_2}, h^e_{n,\ell}\rangle = \langle \varphi_{k_1}, h^{e_1}_{n,\ell_1}\rangle\langle \varphi_{k_2}, h^{e_2}_{n,\ell_2}\rangle.$$

For $k_i \neq 0$, the factors can be bounded using Lemma 6.1, which, for $k_1 \neq 0 \neq k_2$, yields the bound

$$|\langle \varphi_{k_1,k_2}, h^e_{n,\ell}\rangle| \leq \min\left(\frac{6\cdot 2^{\frac{n}{2}}}{|k_1|}, 3\pi 2^{-\frac{n}{2}}\right)\min\left(\frac{6\cdot 2^{\frac{n}{2}}}{|k_2|}, 3\pi 2^{-\frac{n}{2}}\right) \leq \frac{18\pi}{\max(|k_1|,|k_2|)}.$$

Next we consider the case where either $k_1 = 0$ or $k_2 = 0$; w.l.o.g., assume $k_1 = 0$. We use that in one dimension, we have $\langle \varphi_0, h^1_{n,\ell}\rangle = 0$ as well as $\langle \varphi_0, h^0_{n,\ell}\rangle = 2^{-\frac{n}{2}}$. So we only need to consider the case that $e_1 = 0$ and hence $e_2 = 1$. Thus we obtain

$$|\langle \varphi_{0,k_2}, h^e_{n,\ell}\rangle| \leq 2^{-\frac{n}{2}}\frac{6\cdot 2^{\frac{n}{2}}}{|k_2|} = \frac{6}{\max(|k_1|,|k_2|)}.$$

In both cases, we obtain $\mu^{loc}_{k_1,k_2} \leq \frac{18\pi}{\max(|k_1|,|k_2|)}$. The bound $\mu^{loc}_{k_1,k_2} \leq 1$ follows directly from the Cauchy-Schwartz inequality. We conclude $\mu^{loc}_{k_1,k_2} \leq \kappa(k_1,k_2) \leq \kappa'(k_1,k_2)$.

For the ℓ_2-bound, we use an integral estimate,

$$\|\kappa'\|_2^2 \leq \#\{(k_1,k_2) : k_1^2 + k_2^2 \leq 648\pi^2\} + \sum_{\substack{k_1,k_2=-2^{p-1}+1 \\ |k_1|^2+|k_2|^2>648\pi^2}}^{2^{p-1}} \frac{648\pi^2}{|k_1|^2+|k_2|^2}$$
$$\leq 20600 + \iint_{r=18\pi\sqrt{2}-1}^{2^{p-\frac{1}{2}}} 18\pi\sqrt{2}r^{-1}dr d\phi$$
$$\leq 17200 + 502\log_2(N) \leq 2700\log_2(N) = 2700p,$$

where we used that $p \geq 8$. Taking square root implies the result. □

As the infimum of a strictly decreasing function and a strictly increasing function is bounded uniformly by its value at the intersection point of the two functions, Lemma 6.1 also gives frequency-dependent bounds for the local coherence between frequencies and wavelets in the univariate setting.

Corollary 6.3. *Fix $N = 2^p$ with $p \in \mathbb{N}$. For the space $\mathbb{C}^N$, the one-dimensional Fourier basis vectors φ_k, $k \neq 0$, and the one-dimensional Haar wavelets satisfy the incoherence relation*

$$|\langle \varphi_k, h_{n,\ell} \rangle| \leq 3\sqrt{2\pi}/\sqrt{k}.$$

7 Recovery guarantees

7.1 Proof of Theorem 3.2

The proof of Theorem 3.2 concerning recovery from ℓ_1-minimization in the bivariate Haar transform domain follows by combining the local incoherence estimate of Theorem 6.2 with the local coherence based reconstruction guarantees of Theorem 5.2. Under the conditions of Theorem 5.2, the stated recovery results follow directly from Theorem 4.2. The weighted ℓ_2-norm in the noise model is a consequence of the preconditioning.

7.2 Preliminary lemmas for the proof of Theorem 3.1

The proof of Theorem 3.1 proceeds along similar lines to that of Theorem 3.2, but we need a few more preliminary results relating the bivariate Haar transform to the gradient transform. Each of the following results are derived from a more general statement involving the continuous bivariate Haar system and the bounded variation seminorm.

Proposition 7.1. *Suppose $f \in \mathbb{C}^{N^2}$, and suppose its bivariate Haar transform $w = \mathcal{H}f \in \mathbb{C}^{N^2}$ is arranged such that $w_{(k)}$ is the k-th largest-magnitude coefficient. Then there is a universal constant $C > 0$ such that for all $k \geq 1$,*

$$|w_{(k)}| \leq C \frac{\|f\|_{TV}}{k}$$

In words, this proposition says that the bivariate Haar coefficient sequence of a function f is in weak ℓ_1 and its weak ℓ_1 semi-norm is bounded by the total variation semi-norm of f. See [25] for a derivation of Proposition 7.1 from Theorem 8.1 of [13].

We also have the following result about the bivariate Haar system.

Lemma 7.2. *Let $N = 2^p$. For any indices (t_1, t_2) and $(t_1, t_2 + 1)$, there are at most $6p$ bivariate Haar wavelets $h^e_{n,\ell}$ satisfying $|h^e_{n,\ell}(t_1, t_2 + 1) - h^e_{n,\ell}(t_1, t_2)| > 0$.*

Proof. The lemma follows by showing that for fixed dyadic scale n in $0 < n \leq p$, there are at most 6 Haar wavelets with edge length 2^{p-n} satisfying $|h^e_{n,\ell}(t_1, t_2 + 1) - h^e_{n,\ell}(t_1, t_2)| > 0$. If the edge between (t_1, t_2) and $(t_1, t_2 + 1)$ coincides with a dyadic edge at scale n, then the 3 wavelets supported on each of the bordering dyadic squares transition from being zero to nonzero along this edge. On the other hand, if (t_1, t_2) coincides with a dyadic edge at dyadic scale $n+1$ but does not coincide with a dyadic edge at scale n, then 2 of the 3 wavelets supported on the dyadic square having $(t_1, t_2+1), (t_1, t_2)$ as a center edge can satisfy the stated bound. □

Lemma 7.3.

$$\|\nabla h^e_{n,\ell}\|_1 \leq 8 \qquad \forall n, \ell, e.$$

Proof. $h^e_{n,\ell}$ is supported on a dyadic square of side-length 2^{p-n}, and on its support, its absolute value is constant, $|h^e_{n,\ell}| = 2^{n-p}$. Thus at the four boundary edges of the square, there is a jump of 2^{n-p}, at the (at most two) dyadic edges in the middle of the square where the sign changes there is a jump of $2 \cdot 2^{n-p}$. Hence $\|\nabla h^e_{n,\ell}\|_1 \leq \|\nabla h^{\{1,1\}}_{n,\ell}\|_1 \leq 8 \cdot 2^{p-n} \cdot 2^{n-p} = 8$. □

We are now ready to prove Theorem 3.1.

7.3 Proof of Theorem 3.1

Recall that $\mathcal{H} : \mathbb{C}^{N^2} \to \mathbb{C}^{N^2}$ denotes the bivariate Haar transformation $f \mapsto \left(\left\langle f, h^e_{n,\ell}\right\rangle\right)_{n,\ell,e}$, let $w^f_{(j)}$ denote the j-th largest-magnitude Haar coefficient, and let $h_{(j)}$ denote the associated Haar wavelet.

Let $D \in \mathbb{C}^{N^2 \times N^2}$ be the diagonal matrix encoding the weights in the noise model, i.e., $D = \mathrm{diag}(\rho)$, where, for κ' as in Theorem 6.2, $\rho(k_1, k_2) = \|\kappa'\|_2 / \kappa'(k_1, k_2) = C\sqrt{\log_2(N)} \max\Big(1, (|k_1|^2 + |k_2|^2)^{1/2}/18\pi\Big)$. Note that $Dg \equiv \rho \circ g$.

By Theorem 5.2 combined with the bivariate incoherence estimates from Theorem 6.2, we know that with high probability $\mathcal{A} := \frac{1}{\sqrt{m}} D \mathcal{F}_\Omega \mathcal{H}^*$ has the restricted isometry property of order s and level δ once

$$m \gtrsim s\delta^{-2} \log^3(s) \log^5 N.$$

Thus, for the stated number of measurements m with an appropriate hidden constant, we can assume that $\mathcal{A}$ has the restricted isometry property of order

$$\bar{s} = 24\widetilde{C}^2 s \log^3(N),$$

where the exact value of the constant $\widetilde{C}$ will be determined below. In the remainder of the proof we show that this event implies the result.

Let $u = f - f^{\#}$ denote the residual error of (3.2). Then we have

- **Cone Constraint on ∇u.** Let S denote the support of the best s-sparse approximation to ∇f.

Since $f^{\#} = f - u$ is the minimizer of (TV) and f is also a feasible solution,

$$\begin{aligned}
\|(\nabla f)_S\|_1 - \|(\nabla u)_S\|_1 - \|(\nabla f)_{S^c}\|_1 + \|(\nabla u)_{S^c}\|_1 \\
&\leq \|(\nabla f)_S - (\nabla u)_S\|_1 + \|(\nabla f)_{S^c} - (\nabla u)_{S^c}\|_1 \\
&= \|\nabla f^{\#}\|_1 \\
&\leq \|\nabla f\|_1 \\
&= \|(\nabla f)_S\|_1 + \|(\nabla f)_{S^c}\|_1
\end{aligned}$$

Rearranging yields the cone constraint

$$\|(\nabla u)_{S^c}\|_1 \leq \|(\nabla u)_S\|_1 + 2\|\nabla f - (\nabla f)_S\|_1. \tag{7.1}$$

- **Cone Constraint on $w^u = \mathcal{H}u$.** Proposition 7.1 allows us to pass from a cone constraint on the gradient to a cone constraint on the Haar transform. More specifically, we obtain

$$|w^u_{(j)}| \leq C\frac{\|\nabla u\|_1}{j}.$$

Now consider the set $\widetilde{S}$ consisting of the s edges indexed by S. By Lemma 7.2, the set Λ indexing those wavelets which change sign across edges in $\widetilde{S}$ has cardinality at most $|\Lambda| = 6s\log(N)$. Decompose u as

$$u = \sum_j w^u_{(j)} h_{(j)} = \sum_{j\in\Lambda} w^u_{(j)} h_{(j)} + \sum_{j\in\Lambda^c} w^u_{(j)} h_{(j)} =: u_\Lambda + u_{\Lambda^c}$$

and note that by linearity of the gradient,

$$\nabla u = \nabla u_\Lambda + \nabla u_{\Lambda^c}.$$

Now, by construction of the set Λ, we have that $(\nabla u_{\Lambda^c})_S = 0$ and so $(\nabla u)_S = (\nabla u_\Lambda)_S$. By Lemma 7.3 and the triangle inequality,

$$\begin{aligned}
\|(\nabla u)_S\|_1 = \|(\nabla u_\Lambda)_S\|_1 \leq \|\nabla u_\Lambda\|_1 \\
&\leq \sum_{j\in\Lambda} |w_{(j)}|\|\nabla h_{(j)}\|_1 \\
&\leq 8\sum_{j\in\Lambda} |w_{(j)}|.
\end{aligned}$$

Combined with Lemma 7.1 concerning the decay of the wavelet coefficients and the cone constraint

(7.1), and letting

$$\widetilde{s} = 6s\log(N) = |\Lambda|,$$

this gives rise to a cone constraint on the wavelet coefficients:

$$\begin{aligned}
\sum_{j=\widetilde{s}+1}^{N^2} |w^u_{(j)}| &\leq \sum_{j=s+1}^{N^2} |w^u_{(j)}| \\
&\leq C\log(N^2/s)\|\nabla u\|_1 \\
&= C\log(N^2/s)\Big(\|(\nabla u)_S\|_1 + \|(\nabla u)_{S^c}\|_1\Big) \\
&\leq C\log(N^2/s)\Big(\|2(\nabla u)_S\|_1 + 2\|\nabla f - (\nabla f)_S\|_1\Big) \\
&\leq C\log(N^2/s)\Big(16\sum_{j\in\Lambda}|w_{(j)}| + 2\|\nabla f - (\nabla f)_S\|_1\Big) \\
&\leq \widetilde{C}\log(N^2/s)\Big(\sum_{j=1}^{\widetilde{s}}|w_{(j)}| + \|\nabla f - (\nabla f)_S\|_1\Big)
\end{aligned}$$

- **Tube constraint,** $\|\mathcal{A}\mathcal{H}u\|_2 \leq \sqrt{2}\varepsilon$.

  By assumption, $\mathcal{A} = \frac{1}{\sqrt{m}} D\mathcal{F}_\Omega\mathcal{H}^* : \mathbb{C}^{N^2} \to \mathbb{C}^m$ has the RIP of order $\overline{s} > s$. Also by assumption, $\|D\mathcal{F}_\Omega f - Dy\|_2 = \|\rho \circ (\mathcal{F}_\Omega f - y)\|_2 \leq \sqrt{m}\varepsilon$, so f is a feasible solution to (3.2).

  Since both f and $f^\#$ are in the feasible region of (3.2), we have for $u = f - f^\#$,

$$\begin{aligned}
m\|\mathcal{A}\mathcal{H}u\|_2^2 = \|D\mathcal{F}_\Omega\mathcal{H}^*\mathcal{H}u\|_2^2 = \|D\mathcal{F}_\Omega u\|_2^2 \\
&\leq \|D\mathcal{F}_\Omega f - Dy\|_2^2 + \|D\mathcal{F}_\Omega f^\# - Dy\|_2^2 \\
&\leq 2m\varepsilon^2.
\end{aligned}$$

- Using the derived cone and tube constraints on $\mathcal{H}u$ along with the assumed RIP bound on $\mathcal{A}$, the proof is complete by applying Proposition 4.3 using $\gamma = \widetilde{C}\log(N^2/s) \leq 2\widetilde{C}\log(N)$, $k = 6s\log N$, and $\xi = \widetilde{C}\log(N^2/s)\|\nabla f - (\nabla f)_S\|_1$. In fact, this is where we need that the RIP order is $\overline{s}$, to accomodate for the factors γ and k. □

8 Summary and outlook

We established reconstruction guarantees for variable-density discrete Fourier measurements in both the wavelet sparsity and gradient sparsity setup. Our results build on local coherence estimates between

Fourier and wavelet bases. The resulting sampling strategies are specific to two-dimensional discrete images, that is, $N \times N$ blocks of pixels. A priori, our results do not directly generalize to higher dimensional or univariate signal models. In particular, optimal sampling strategies as well as stable image recovery guarantees remain open for higher-dimensional signals.

Variable density sampling in compressive imaging has often been justified as taking into account the tree-like sparsity structure of natural images in wavelet bases (e.g., in [39]). We note that our theory does not directly exploit this additional structure, and depends only on the local incoherence between Fourier and wavelet bases. We expect, however, that this additional structure can be used to derive sampling strategies with stronger reconstruction guarantees, as indicated by the suboptimality of the sampling densities predicted by our results in numerical simulations (Figure 1).

All the recovery guarantees in this paper are uniform, that is, we seek measurement ensembles which allow for approximate reconstruction of all images. For non-uniform recovery guarantees, we expect that the number of measurements required in our main results can be reduced by several logarithmic factors by following a probabilistic and "RIP-less" approach [6].

It should also be noted that this paper does not address the important issue of errors arising from discretization of the image and Fourier measurements. In particular, as observed for example in [1], the use of discrete rather than continuous Fourier representations can be a significant source of error in compressed sensing. The authors of [1] propose to resolve this issue using uneven sections, that is, the number of discretization points in frequency is chosen to be larger than the number of discretization points in time. Nevertheless, the results in [1] are again just formulated for incoherent samples. Recently, it has been proposed to overcome this issue by sampling all of the low frequencies in addition to uniformly sampling the higher frequencies [2], but to date, no provable reconstruction guarantees have been provided. We expect that our approach can be applied to this setup – due to the variable density, it may even be possible to sample from the infinite set rather than restricting to a finite subset based on an intricate criterion. Such a generalization is out of reach for the optimization-based approaches such as in [30], which will always be specific to the given problem dimension. In this sense, we expect that the additional understanding provided by this paper can eventually lead to optimized sampling schemes. All these questions, however, are left for future work.

A Proof of Proposition 4.3.

Write $u_{S^c} = u^{(1)} + u^{(2)} + \cdots + u^{(r)}$ where $r = \lfloor \frac{N}{4k\gamma^2} \rfloor$. Here $u^{(1)}$ is the $4k\gamma^2$-sparse image consisting of the $4k\gamma^2$ largest-magnitude components of u_{S^c}, $u^{(2)}$ consists of the $4k\gamma^2$ largest-magnitude components among the remaining entries of u_{S^c}, and so on. Since the magnitude of each component of $u^{(j-1)}$ is at least as large as the average magnitude of the components of $u^{(j)}$, we obtain

$$\|u^{(j)}\|_2 \leq \frac{\|u^{(j-1)}\|_1}{2\gamma\sqrt{k}}, \quad j = 2, 3, \ldots$$

Combining this with the cone constraint gives

$$\sum_{j=2}^{r} \|u^{(j)}\|_2 \leq \frac{1}{2\gamma\sqrt{k}}\|u_{S^c}\|_1 \leq \frac{\gamma}{2\gamma\sqrt{k}}\|u_S\|_1 + \frac{1}{2\gamma\sqrt{k}}\xi \leq \frac{1}{2}\|u_S\|_2 + \frac{1}{2\gamma\sqrt{k}}\xi.$$

Together with the tube constraint and the RIP, we obtain

$$\begin{aligned}
\varepsilon &\gtrsim \|Au\|_2 \\
&\geq \|A(u_S + u^{(1)})\|_2 - \sum_{j=2}^{r} \|A(u^{(j)})\|_2 \\
&\geq \sqrt{1-\delta}\|u_S + u^{(1)}\|_2 - \sqrt{1+\delta}\sum_{j=2}^{r} \|u^{(j)}\|_2 \\
&\geq \sqrt{1-\delta}\|u_S + u^{(1)}\|_2 - \sqrt{1+\delta}\Big(\frac{1}{2}\|u_S\|_2 + \frac{1}{2\gamma\sqrt{k}}\xi\Big) \\
&\geq \left(\sqrt{1-\delta} - \frac{\sqrt{1+\delta}}{2}\right)\|u_S + u^{(1)}\|_2 - \sqrt{1+\delta}\frac{1}{2\gamma\sqrt{k}}\xi.
\end{aligned}$$

Then, since $\delta < 1/3$,

$$\|u_S + u^{(1)}\|_2 \leq 5\varepsilon + \frac{3\xi}{\gamma\sqrt{k}}.$$

Finally, because $\|\sum_{j=2}^{r} u^{(j)}\|_2 \leq \sum_{j=2}^{r} \|u^{(j)}\|_2 \leq \frac{1}{2}\|u_S + u^{(1)}\|_2 + \frac{1}{2\gamma\sqrt{k}}\xi$ we have

$$\|u\|_2 \leq 8\varepsilon + \frac{5\xi}{\gamma\sqrt{k}},$$

confirming (4.1). □

Acknowledgments

The authors would like to thank Anders Hansen, Deanna Needell, Holger Rauhut, Justin Romberg, Amit Singer, Mark Tygert, and Robert Vanderbei for helpful comments and suggestions. They are grateful for the stimulating research environment of the Mathematisches Forschungsinstitut Oberwolfach, where part of this work was completed. Rachel Ward was supported in part by an Alfred P Sloan Research Fellowship, a Donald D. Harrington Faculty Fellowship, and DOD-Navy grant N00014-12-1-0743.

References

[1] B. Adcock and A. Hansen. Generalized sampling and infinite dimensional compressed sensing. Preprint, 2011.

[2] B. Adcock, A. Hansen, E. Herrholz, and G. Teschke. Generalized sampling, infinite-dimensional compressed sensing, and semi-random sampling for asymptotically incoherent dictionaries. Preprint, 2011.

[3] N. Ailon and E. Liberty. Almost optimal unrestricted fast Johnson-Lindenstrauss transform. *Symposium on Discrete Algorithms (SODA '11).*

[4] N. Burq, S. Dyatlov, R. Ward, and M. Zworski. Weighted eigenfunction estimates with applications to compressed sensing. *SIAM J. Math. Anal.*, 44(5):3481–3501, 2012.

[5] E. Candès and F. Guo. New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Process.*, 82(11):1519–1543, 2002.

[6] E. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57:7235–7254, 2011.

[7] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

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

[9] E J. Candès and T Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[10] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[11] T.F. Chan, J. Shen, and H.M. Zhou. total variation wavelet inpainting. *J. Math. Imaging Vis.*, 25(1):107–125, 2006.

[12] M. Cheraghchia, V. Guruswami, and A. Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. Preprint, 2012.

[13] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space $BV(\mathbb{R}^2)$. *Am. J. of Math*, 121:587–628, 1999.

[14] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289 –1306, 2006.

[15] A. Fannjiang. TV-min and greedy pursuit for constrained joint sparsity and application to inverse scattering. Preprint, 2012.

[16] A. Fannjiang, T. Strohmer, and P. Yan. Compressed remote sensing of sparse objects. *SIAM J. Imag. Sci.*, 3:596–618, 2010.

[17] A. Garnaev and E. Gluskin. On widths of the Euclidean ball. *Sov. Math. Dokl.*, 30:200–204, 1984.

[18] A. Greiser and M. Kienlin. Efficient K-space sampling by density-weighted phase-encoding. *Magn. Reson. Med.*, 50:1266–1275, 2003.

[19] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia*, 41:334–351, 1977.

[20] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43, 2010.

[21] M. Lustig, D. Donoho, and J.M. Pauly. Sparse MRI: The application of compressed sensing for rapid MRI imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[22] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly. Compressed sensing MRI. *IEEE Sig. Proc. Mag.*, 25(2):72–82, 2008.

[23] G. Marseille, R. de Beer, M. Fuderer, A. Mehlkopf, and D. van Ormondt. Nonuniform phase-encode distributions for MRI scan time reduction. *J Magn Reson*, pages 70–75, 1996.

[24] Qun Mo and Song Li. New bounds on the restricted isometry constant δ_{2k}. *Appl. Comput. Harmon. Anal.*, 31(3):460–468, 2011.

[25] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. Preprint, 2012.

[26] D. Needell and R. Ward. Total variation minimization for stable multidimensional signal recovery. Preprint, 2012.

[27] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Model. Sim.*, 1:349–370, 2003.

[28] D Peters, F Korosec, T Ghrist, W Block, J Holden, K Vigen, and C Mistretta. Undersampled projection reconstruction applied to MR angiography. *Magn. Reson. Med.*, 43:91–101, 2000.

[29] G. Puy, J.P.Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread spectrum magnetic resonance imaging. *IEEE T. Medical Imaging*, 31(3):586–598, 2012.

[30] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *Signal Processing Letters*, 18:595–598, 2011.

[31] H. Rauhut. Compressive Sensing and Structured Random Matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.

[32] H. Rauhut and R. Ward. Sparse recovery for spherical harmonic expansions. In *Proc. SampTA*, Singapore, 2011.

[33] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1-minimization. *Journal of Approximation Theory*, 164:517–533, 2012.

[34] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[35] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[36] K. Scheffer and J. Hennig. Reduced circular field-of-view imaging. *Magn. Reson. Med.*, 40:474–480, 1998.

[37] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Probl.*, 19, 2003.

[38] C Tsai and D. Nishimura. Reduced aliasing artifacts using variable-density K-space sampling trajectories. *Magn. Reson. Med*, 43:452–458, 2000.

[39] Z. Wang and G.R. Arce. Variable density compressed image sampling. *IEEE T. Image Process.*, 19(1):264 –270, 2010.

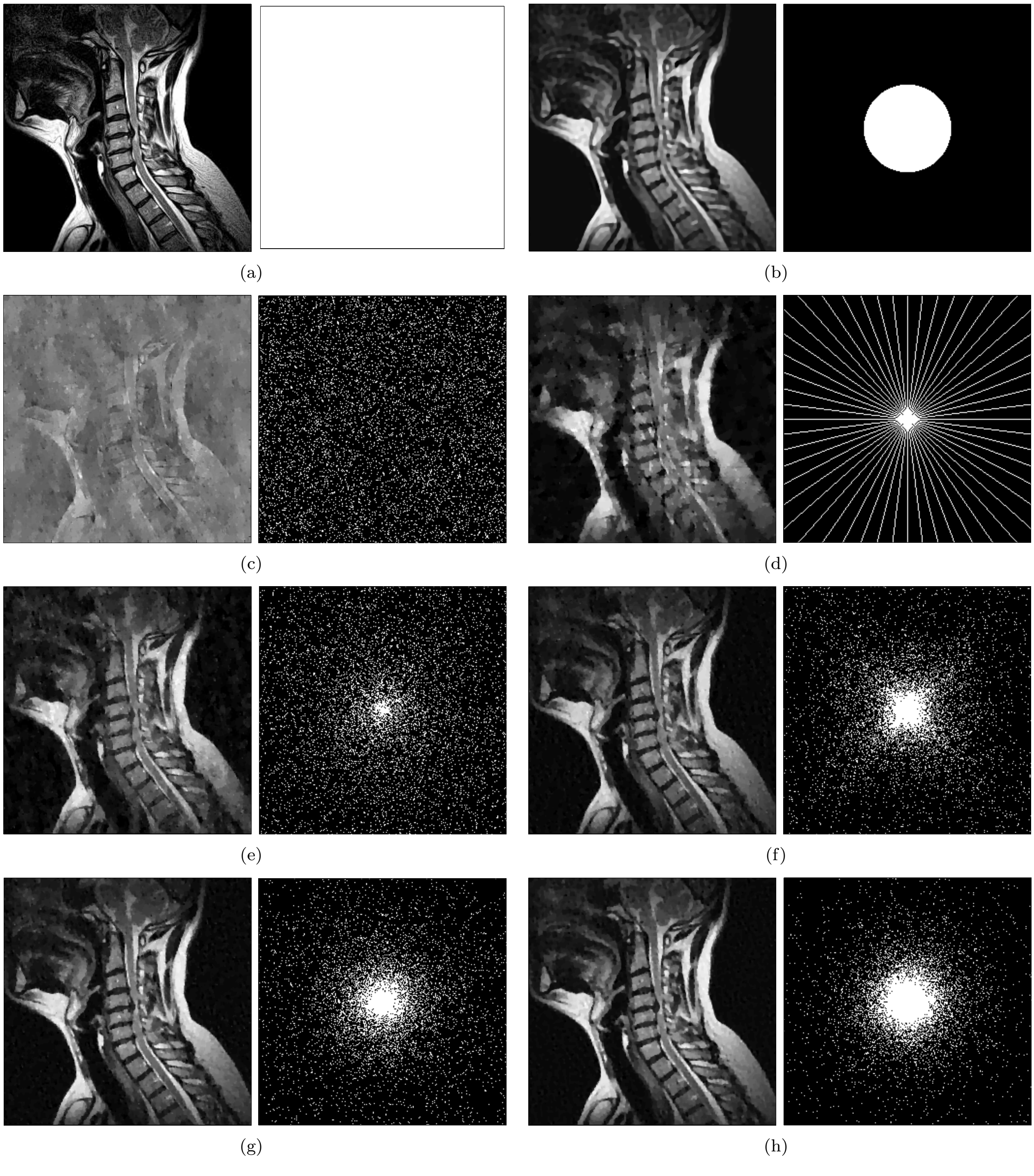

Figure 1: Various reconstructions of an MRI image $x \in \mathbb{R}^{256\times 256}$ with total variation minimization as in Theorem 3.1 with $\varepsilon = 0$ and using $m = 6400$ noiseless partial DFT measurements $y = \mathcal{F}_\Omega x$ with frequencies $\Omega = (k_1, k_2)$ sampled from various distributions. Beside each reconstruction is a plot of K-space $\{(k_1, k_2) : -N/2 + 1 \le k_1, k_2 \le N/2\}$ and the frequencies used in its reconstruction. (a) Original image (and all of K-space). (b) Reconstruction using only lowest frequencies: $\Omega = \{(k_1, k_2) : k_1^2 + k_2^2 \le 80\}$. (c) $\mathrm{Prob}\big((k_1, k_2) \in \Omega\big) \sim 1$ (Uniform subsampling) (d) Ω comprised of frequencies in equispaced radial lines. (e) $\mathrm{Prob}\big((k_1, k_2) \in \Omega\big) \propto (k_1^2 + k_2^2 + 1)^{-1/2}$(f) $\mathrm{Prob}\big((k_1, k_2) \in \Omega\big) \propto \big(\max(|k_1|, |k_2|) + 1\big)^{-1}$ (g) $\mathrm{Prob}\big((k_1, k_2) \in \Omega\big) \propto (k_1^2 + k_2^2 + 1)^{-1}$. (h) $\mathrm{Prob}\big((k_1, k_2) \in \Omega\big) \propto (k_1^2 + k_2^2 + 1)^{-3/2}$. Theorem 3.1 guarantees stable and robust recovery for the inverse square-distance distribution in (g); a slightly stronger guarantee can be obtained for the inverse-max sampling distribution given in (f) from the stronger local coherence bound in Theorem 6.2. The relative reconstruction error $\|f - f^{\#}_{TV}\|_2 / \|f\|_2$ corresponding to each reconstruction is (b) .2932, (c) .8229, (d) .4074, (e) .3192, (f) .2603, (g) .2537, and (h) .2463.

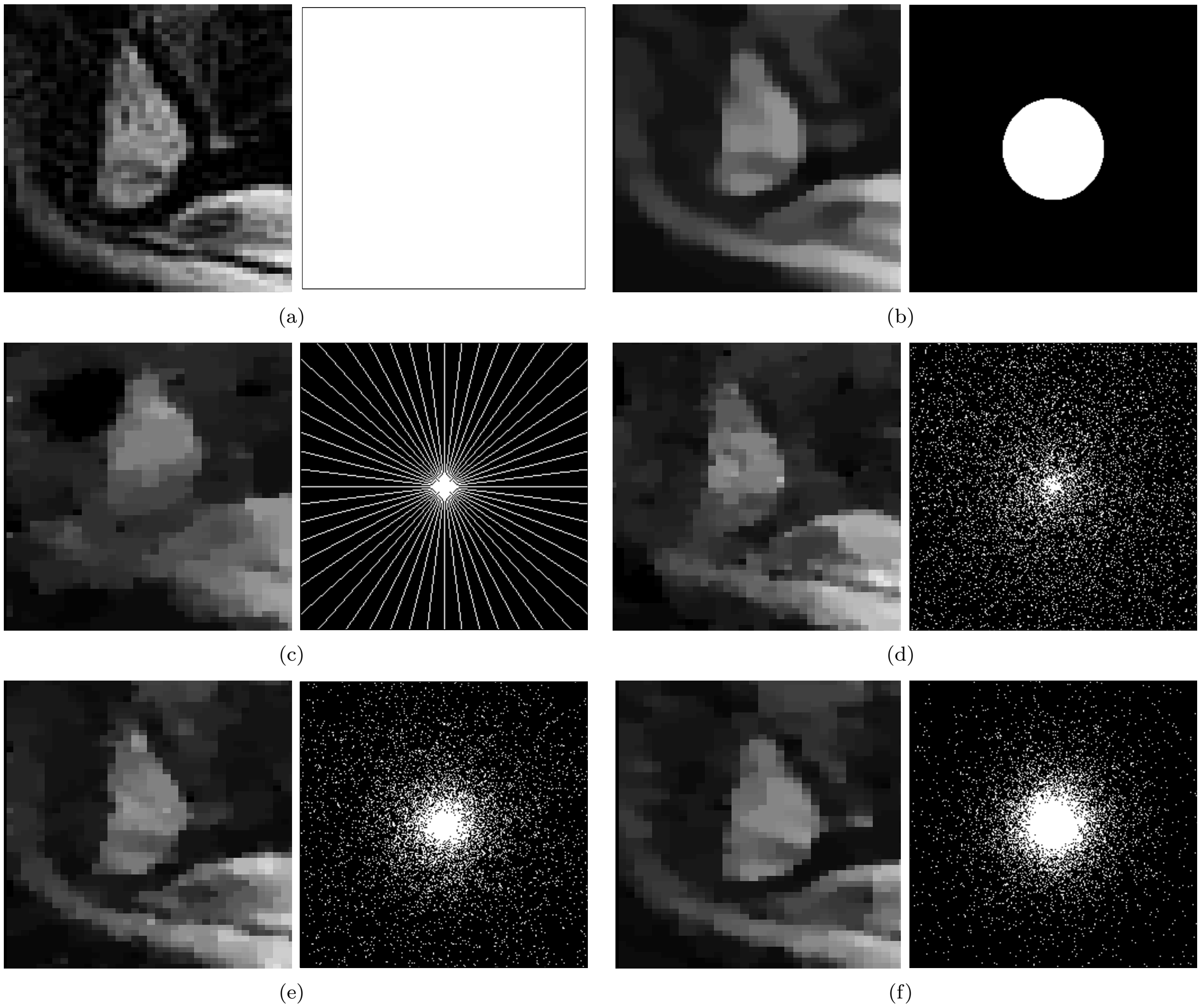

Figure 2: The same reconstructions as in Figure 1, magnified to the chin region: (a) Original image (and all of K-space). (b) Reconstruction using only lowest frequencies: $\Omega = \{(k_1, k_2) : k_1^2 + k_2^2 \leq 80\}$. (c) Ω comprised of frequencies in equispaced radial lines. (d) $\text{Prob}\big((k_1, k_2) \in \Omega\big) \propto (k_1^2 + k_2^2 + 1)^{-1/2}$ (e) $\text{Prob}\big((k_1, k_2) \in \Omega\big) \propto (k_1^2 + k_2^2 + 1)^{-1}$. (f) $\text{Prob}\big((k_1, k_2) \in \Omega\big) \propto (k_1^2 + k_2^2 + 1)^{-3/2}$. Theorem 3.1 concerns stable and robust recovery for method (e).