5-22-2013

Near-optimal Compressed Sensing Guarantees for Total Variation Minimization

Deanna Needell
Claremont McKenna College

R. Ward

Near-optimal compressed sensing guarantees for total variation minimization

Deanna Needell[*] and Rachel Ward[†]

March 26, 2013

Abstract

Consider the problem of reconstructing a multidimensional signal from an underdetermined set of measurements, as in the setting of compressed sensing. Without any additional assumptions, this problem is ill-posed. However, for signals such as natural images or movies, the minimal *total variation* estimate consistent with the measurements often produces a good approximation to the underlying signal, even if the number of measurements is far smaller than the ambient dimensionality. This paper extends recent reconstruction guarantees for two-dimensional images $\boldsymbol{x} \in \mathbb{C}^{N^2}$ to signals $\boldsymbol{x} \in \mathbb{C}^{N^d}$ of arbitrary dimension $d \geq 2$ and to isotropic total variation problems. To be precise, we show that a multidimensional signal $\boldsymbol{x} \in \mathbb{C}^{N^d}$ can be reconstructed from $\mathcal{O}(sd \log(N^d))$ linear measurements $\boldsymbol{y} = \boldsymbol{Ax}$ using total variation minimization to within a factor of the best s-term approximation of its gradient. The reconstruction guarantees we provide are necessarily optimal up to polynomial factors in the spatial dimension d.

1 Introduction

Compressed sensing (CS) is an emerging signal processing methodology where signals are acquired in compressed form as undersampled linear measurements. The applications of CS are abundant, ranging from radar and error correction to many areas of image processing [18]. The underlying assumption that makes such acquisition and reconstruction possible is that most natural signals are *sparse* or *compressible*. We say that a signal $\boldsymbol{x} \in \mathbb{C}^p$ is s-sparse when

$$\|\boldsymbol{x}\|_0 \stackrel{\text{def}}{=} |\operatorname{supp}(\boldsymbol{x})| \leq s \ll p. \tag{1}$$

Compressible signals are those which are well-approximated by sparse signals. In the CS framework, we acquire $m \ll p$ nonadaptive linear measurements of the form

$$\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}) + \boldsymbol{\xi},$$

[*]Claremont McKenna College, 850 Columbia Ave, Claremont CA, 91711, email: dneedell@cmc.edu.

[†]University of Texas at Austin, 2515 Speedway, Austin, Texas, 77842, email: rward@math.utexas.edu. R. Ward has been supported in part by a Donald D. Harrington Faculty Fellowship, Alfred P. Sloan Research Fellowship, and DOD-Navy grant N00014-12-1-0743.

where $\mathcal{M} : \mathbb{C}^p \to \mathbb{C}^m$ is an appropriate linear operator and $\boldsymbol{\xi}$ is vector modeling additive noise. The theory of CS ([22, 9, 21]) ensures that under suitable assumptions on $\mathcal{M}$, compressible signals can be approximately reconstructed by solving an ℓ_1 minimization problem:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{w}}{\mathrm{argmin}} \|\boldsymbol{w}\|_1 \quad \text{such that} \quad \|\mathcal{M}(\boldsymbol{w}) - \boldsymbol{y}\|_2 \leq \varepsilon; \qquad (L_1)$$

above, $\|\boldsymbol{w}\|_1 = \sum_i |w_i|$ and $\|\boldsymbol{w}\|_2 = \left(\sum_i |w_i|^2\right)^{1/2}$ denote the standard ℓ_1 and Euclidean norms, ε bounds the noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$, and argmin denotes the set of minimizers (we use the notation of equality rather than set notation here and use the convention that if the minimizer is not unique, we may choose a solution lexicographically; however, in this particular case under standard compressed sensing assumptions it is known that the solution is indeed unique [51]). The program (L_1) may be cast as a second order cone program (SOCP) and can be solved efficiently using standard convex programming methods (see e.g. [3, 17]).

One property of the operator $\mathcal{M}$ that guarantees sparse signal recovery via (L_1), as introduced by Candès and Tao in [9], is the restricted isometry property (RIP).

Definition 1. *A linear operator $\mathcal{M} : \mathbb{C}^p \to \mathbb{C}^m$ is said to have the restricted isometry property (RIP) of order $s \in \mathbb{N}$ and level $\delta \in (0, 1)$ if*

$$(1 - \delta)\|\boldsymbol{x}\|_2^2 \leq \|\mathcal{M}(\boldsymbol{x})\|_2^2 \leq (1 + \delta)\|\boldsymbol{x}\|_2^2 \qquad \textit{for all } s\textit{-sparse } \boldsymbol{x} \in \mathbb{C}^p. \qquad (2)$$

Many distributions of random matrices of dimension $m \times p$, including randomly subsampled rows from the discrete Fourier transform [47] or from a bounded orthonormal system more generally [47, 45, 43, 44, 5], and randomly-generated circulant matrices [28], are known to generate RIP matrices of order s and level $\delta \leq c < 1$ if the number of measurements satisfies $m \approx \delta^{-2} s \log^4(p)$. Note that here and throughout we have used the notation $u \approx v$ (analogously $u \lesssim v$) to indicate that there exists some absolute constant $C > 0$ such that $u = Cv$ ($u \leq Cv$). Moreover, a matrix whose entries are independent and identical (i.i.d.) realizations of a properly-normalized subgaussian random variable will have the RIP with probability exceeding $1 - e^{-cm}$ once $m \approx \delta^{-2} s \log(p/s)$ [10, 36, 47, 1].

Candès, Romberg, and Tao [8] showed that when the measurement operator $\mathcal{M}$ has the RIP of order $\mathcal{O}(s)$ and sufficiently small constant δ, the program (L_1) recovers an estimation $\hat{\boldsymbol{x}}$ to $\boldsymbol{x}$ that satisfies the error bound

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq C \left(\frac{\|\boldsymbol{x} - \boldsymbol{x_s}\|_1}{\sqrt{s}} + \varepsilon \right), \qquad (3)$$

where $\boldsymbol{x_s}$ denotes the best s-sparse approximation to the signal $\boldsymbol{x}$. Using properties about Gel'fand widths of the ℓ_1 ball due to Kashin [26] and Garnaev–Gluskin [23], this is the optimal minimax reconstruction rate for ℓ_1-minimization using $m \approx s \log(p/s)$ nonadaptive linear measurements. Due to the rotational-invariance of an RIP matrix with randomized column signs [29], a completely analogous theory holds for signals that are compressible with respect to a known orthonormal basis or tight frame $\boldsymbol{D}$ by replacing $\boldsymbol{w}$ with $\boldsymbol{D}^*\boldsymbol{w}$ inside the ℓ_1-norm of the minimization problem (L_1) [6, 31].

1.1 Imaging with CS

Natural images are highly compressible with respect to their gradient representation. For an image $\boldsymbol{x} \in \mathbb{C}^{N^2}$ one defines its discrete directional derivatives by

$$\boldsymbol{x}_u : \mathbb{C}^{N^2} \to \mathbb{C}^{(N-1) \times N}, \qquad (\boldsymbol{x}_u)_{j,k} = \boldsymbol{x}_{j+1,k} - \boldsymbol{x}_{j,k} \qquad (4)$$

$$\boldsymbol{x}_v : \mathbb{C}^{N^2} \to \mathbb{C}^{N \times (N-1)}, \qquad (\boldsymbol{x}_v)_{j,k} = \boldsymbol{x}_{j,k+1} - \boldsymbol{x}_{j,k}. \qquad (5)$$

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

$$
\left((\nabla \boldsymbol{x})_{j,k,1}, (\nabla \boldsymbol{x})_{j,k,2}\right) \overset{\text{def}}{=}
\begin{cases}
\left((\boldsymbol{x}_u)_{j,k}, (\boldsymbol{x}_v)_{j,k}\right), & 1 \le j \le N-1, \quad 1 \le k \le N-1 \\
\left(0, (\boldsymbol{x}_v)_{j,k}\right), & j = N, \quad 1 \le k \le N-1 \\
\left((\boldsymbol{x}_u)_{j,k}, 0\right), & k = N, \quad 1 \le j \le N-1 \\
(0,0), & j = k = N
\end{cases}
$$

The *anisotropic* total variation seminorm is defined as

$$
\|\boldsymbol{x}\|_{TV_1} \overset{\text{def}}{=} \sum_{j,k=1}^{N} |(\nabla \boldsymbol{x})_{j,k,1}| + |(\nabla \boldsymbol{x})_{j,k,2}|, \tag{6}
$$

and the *isotropic* total variation seminorm as

$$
\|\boldsymbol{x}\|_{TV_2} \overset{\text{def}}{=} \sum_{j,k=1}^{N} \left((\nabla \boldsymbol{x})_{j,k,1}^2 + (\nabla \boldsymbol{x})_{j,k,2}^2\right)^{1/2}. \tag{7}
$$

The recovery guarantees we derive apply to both anisotropic and isotropic total variation semi norms, and we use the notation $\|\boldsymbol{x}\|_{TV}$ to refer to either choice of seminorm. For brevity of presentation, we provide details only for the isotropic total variation seminorm, but refer the reader to [37] for the analysis of the anisotropic variant.

The total variation seminorm is a regularizer of choice in many image processing applications. That is, given a collection of noisy linear measurements $\mathcal{M}(\boldsymbol{x}) + \boldsymbol{\xi}$ with $\|\boldsymbol{\xi}\|_2 \le \varepsilon$ of an underlying image $\mathbf{x} \in \mathbb{C}^{N^2}$, total variation minimization is used to pick from among the possibly infinitely-many images consistent with these measurements:

$$
\hat{\boldsymbol{x}} = \underset{\boldsymbol{z}}{\operatorname{argmin}} \|\boldsymbol{z}\|_{TV} \quad \text{such that} \quad \|\mathcal{M}(\boldsymbol{z}) - \boldsymbol{y}\|_2 \le \varepsilon \tag{TV}
$$

Properties of TV minimizers in inverse problems have been studied in the discrete and continuous settings [2, 40, 20, 48, 46, 41, 12, 13], and convergence rates of stability measures for TV have also been established [4, 24]. In the setting of compressed sensing and more broadly in other imaging applications, total variation regularization has been used for denoising, deblurring, and inpainting (see e.g. [8, 11, 7, 42, 14, 33, 34, 32, 39, 35, 25, 27, 50, 38] and the references therein). **In this article, we focus on recovery guarantees for (TV) in the compressed sensing setting. Along the way, we derive strengthened Sobolev inequalities for discrete signals lying near the null space of operators incoherent with the Haar wavelet basis, and we believe that such bounds should be useful in a broader context for understanding the connection between total variation minimization and ℓ_1-wavelet coefficient minimization.**

While (TV) is similar to the ℓ_1-minimization program (L_1), the RIP-based theoretical guarantees for (L_1) do not directly translate to recovery guarantees for (TV) because the gradient map $\boldsymbol{z} \to \nabla \boldsymbol{z}$ is not well-conditioned on the orthogonal complement of $\ker(\nabla)$. In fact, viewed as an invertible operator over mean-zero images, the condition number of the gradient map is proportional to the image side length N.[1] For anisotropic (TV), recovery guarantees in the compressed sensing setting were obtained in [38] for two-dimensional images.

---

[1] One sees that the norm of ∇ is a constant whereas the norm of its inverse is proportional to N (one can observe this scaling, for example, by noting it is obtained by the image whose entries are constant).

Theorem A (from [38]). *For a number of measurements $m \gtrsim s \log(N^2/s)$, there are choices of linear operators $\mathcal{M} : \mathbb{C}^{N^2} \to \mathbb{C}^m$ for which the following holds for any image $\boldsymbol{x} \in \mathbb{C}^{N^2}$: Given noisy measurements $\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}) + \boldsymbol{\xi}$ with noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$, the reconstructed image*

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{z}}{\operatorname{argmin}} \|\boldsymbol{z}\|_{TV_1} \quad \text{such that} \quad \|\mathcal{M}(\boldsymbol{z}) - \boldsymbol{y}\|_2 \leq \varepsilon \tag{8}$$

satisfies the error bound

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \lesssim \log(N^2/s) \Big(\frac{\|\nabla \boldsymbol{x} - (\nabla \boldsymbol{x})_s\|_1}{\sqrt{s}} + \varepsilon \Big). \tag{9}$$

Here and throughout, $\boldsymbol{z}_s$ denotes the best s-term approximation to the array $\boldsymbol{z}$.

In words, the total variation minimizer estimates $\boldsymbol{x}$ to within a factor of the noise level and best s-term approximation error of its gradient. The bound in (9) is optimal up to the logarithmic factor $\log(N^2/s)$.

The contribution of this paper is twofold: We extend the recovery guarantees of Theorem A to the multidimensional setting $x \in \mathbb{C}^{N^d}$ and to the setting of isotropic total variation minimization, for arbitrary dimension $d \geq 2$. The precise statement of results is given in Theorem 3. The proofs involve extending the Sobolev inequalities for random subspaces from [38] to higher-dimensional signal structures, using bounds of Cohen, Dahmen, Daubechies, and DeVore in [15] on the compressibility of wavelet representations in terms of the bounded variation of a function, which hold for functions in dimension $d \geq 2$. Hence, our results for total variation, do not hold in dimension $d = 1$. See [49] for results on one-dimensional total variation under assumptions other than the RIP.

1.2 Organization

The article is organized as follows. In Section 2 we recall relevant background material on the multidimensional total variation seminorm and multidimensional orthonormal wavelet transform. Section 3 states our main result: total variation minimization provides stable signal recovery for signals of arbitrary dimension $d \geq 2$. The proof of this result will occupy the remainder of the paper; in Section 4 we prove that the signal *gradient* is recovered stably, while in Section 5 we pass from stable gradient recovery to stable signal recovery using strengthened Sobolev inequalities which we derive for random subspaces. The proofs of propositions and theorems used along the way are contained in the appendix.

2 Preliminaries for multidimensional signal analysis

The setting for this article is the space $\mathbb{C}^{N^d}$ of multidimensional arrays of complex numbers,

$$\boldsymbol{x} = (x_\alpha), \quad \alpha \equiv (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \{1, 2, \ldots, N\}^d.$$

From here on out, we will use the shorthand $[N]^d = \{1, 2, \ldots, N\}^d$. We will also use the convention that vectors such as $\boldsymbol{x}$ (apart from index vectors α) are boldface and their scalar components such as x_α are normal typeface. We also treat $\mathbb{C}^{N^d}$ as the Hilbert space equipped with inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{\alpha \in [N]^d} x_\alpha \cdot \bar{y}_\alpha, \tag{10}$$

4

where $\bar{y}$ denotes the conjugate of y. This Hilbert space is isometric[2] to the subspace $\Sigma_N^d \subset L_2([0,1)^d)$ of functions which are constant over cubes $[\frac{\alpha_i-1}{N}, \frac{\alpha_i}{N})_{i=1}^d$ of side length N^{-1}, and the isometry is provided by identifying $x_\alpha \in \mathbb{C}^{N^d}$ with the function $f \in \Sigma_N^d$ satisfying $f(u) = N^{d/2}x_\alpha$ for $u \in [\frac{\alpha_i-1}{N}, \frac{\alpha_i}{N})_{i=1}^d$. More generally, we denote by $\|x\|_p = \left(\sum_{\alpha \in [N]^d} |x_\alpha|^p\right)^{1/p}$ the entrywise ℓ_p-norm of the signal x.

For $\ell = 1, 2, \dots d$, the discrete derivative of x in the direction of r_ℓ is the array $x_{r_\ell} \in \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}}$ defined component-wise by

$$(x_{r_\ell})_\alpha \stackrel{\text{def}}{=} x_{(\alpha_1,\alpha_2,\dots,\alpha_\ell+1,\dots,\alpha_d)} - x_{(\alpha_1,\alpha_2,\dots,\alpha_\ell,\dots\alpha_d)}, \tag{11}$$

and we define the d-dimensional discrete gradient transform $\nabla : \mathbb{C}^{N^d} \to \mathbb{C}^{d \times N^d}$ through its components

$$(\nabla x)_\alpha = (\nabla x)_{\alpha,\ell} \stackrel{\text{def}}{=} \begin{cases} (x_{r_\ell})_\alpha, & \alpha_\ell \leq N-1, \\ 0, & \text{else} \end{cases} \tag{12}$$

The d-dimensional *anisotropic* total variation seminorm is defined as $\|x\|_{TV_1} \stackrel{\text{def}}{=} \|\nabla x\|_1$, while the *isotropic* total variation seminorm is a mixed ℓ_1-ℓ_2 norm of the d-dimensional discrete gradient,

$$\|x\|_{TV_2} \stackrel{\text{def}}{=} \sum_{\alpha \in [N]^d} \left(\sum_{\ell=1}^d (\nabla x)_{\alpha,\ell}^2\right)^{1/2} \stackrel{\text{def}}{=} \|\nabla x\|_{1,2}. \tag{13}$$

A linear operator $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^r$ can be regarded as a sequence of multidimensional arrays componentwise,

$$y_k = [\mathcal{A}(x)]_k = \langle a_k, x \rangle, \tag{14}$$

and a linear operator $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^{N^d}$ can be expressed similarly through its components $y_\alpha = [\mathcal{A}(x)]_\alpha = \langle a_\alpha, x \rangle$. If $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^{r_1}$ and $\mathcal{B} : \mathbb{C}^{N^d} \to \mathbb{C}^{r_2}$ then the *row direct sum operator* $\mathcal{M} = \mathcal{A} \oplus_r \mathcal{B}$ is the linear operator from $\mathbb{C}^{N^d}$ to $\mathbb{C}^{r_1+r_2}$ with component arrays $\mathcal{M} = (m_k)_{k=1}^{r_1+r_2}$ given by

$$m_k = \begin{cases} a_k, & 1 \leq k \leq r_1, \\ b_{k-r_1}, & 1 + r_1 \leq k \leq r_1 + r_2. \end{cases}$$

Alternatively, for linear operators $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^r$ and $\mathcal{B} : \mathbb{C}^{N^d} \to \mathbb{C}^r$, the column direct sum operator $\mathcal{N} = \mathcal{A} \oplus_c \mathcal{B} : \mathbb{C}^{N^d \times 2} \to \mathbb{C}^r$ has component arrays $\mathcal{N} = (n_k)_{k=1}^r$ given by

$$(n_k)_{\alpha,\ell} = \begin{cases} (a_k)_\alpha, & \ell = 1, \\ (b_k)_\alpha, & \ell = 2. \end{cases}$$

2.1 The multidimensional Haar wavelet transform

The Haar wavelet transform provides a sparsifying basis for natural signals such as images and movies, and is closely related to the discrete gradient. For a comprehensive introduction to wavelets, we refer the reader to [19].

---

[2]Recall that $f \in L_2(Q)$ if $\int_Q |f(u)|^2 du < \infty$, and $L_2(Q)$ is a Hilbert space equipped with the inner product $\langle f, g \rangle = \int_Q f(u) \cdot \bar{g}(u) du$.

The (continuous) multidimensional Haar wavelet basis is derived from a tensor-product representation of the univariate Haar basis, which forms an orthonormal system for square-integrable functions on the unit interval and consists of the constant function

$$h^0(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0, & \text{otherwise,} \end{cases}$$

the step function

$$h^1(t) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \end{cases}$$

and dyadic dilations and translations of the step function,

$$h_{j,k}(t) = 2^{j/2} h^1(2^j t - k); \quad j \in \mathbb{N}, \quad 0 \leq k < 2^j. \tag{15}$$

The Haar basis for the higher dimensional space $L_2(Q)$ of square-integrable functions on the unit cube $Q = [0,1)^d$ consists of tensor-products of the univariate Haar wavelets. Concretely, for $V = \{0,1\}^d - \{0\}^d$ and $\boldsymbol{e} = (e_1, e_2, \ldots, e_d) \in V$, we define the multivariate functions $h^{\boldsymbol{e}} : L_2(Q) \rightarrow L_2(Q)$ by

$$h^{\boldsymbol{e}}(u) = \prod_{e_i} h^{e_i}(u_i).$$

The orthonormal Haar system on $L_2(Q)$ is then comprised of the constant function along with all functions of the form

$$h_{j,k}^{\boldsymbol{e}}(u) = 2^{jd/2} h^{\boldsymbol{e}}(2^j u - k), \quad \boldsymbol{e} \in V, \quad j \geq 1, \quad k \in \mathbb{Z}^d \cap 2^j Q. \tag{16}$$

The *discrete* multidimensional Haar transform is derived from the continuous construction via the isometric identification between $\mathbb{C}^{N^d}$ and $\Sigma_N \subset L_2(Q)$: defining

$$\boldsymbol{h}_0(\alpha) = N^{-d/2}, \qquad \boldsymbol{h}_{j,k,\boldsymbol{e}}(\alpha) = N^{-d/2} h_{j,k}^{\boldsymbol{e}}(\alpha/N), \quad \alpha \in [N]^d, \tag{17}$$

the matrix product computing the discrete Haar transform can be expressed as $\mathcal{H}(\boldsymbol{x}) = \left(\langle \boldsymbol{h}_0, \boldsymbol{x} \rangle, (\langle \boldsymbol{h}_{j,k,\boldsymbol{e}}, \boldsymbol{x} \rangle) \right)$, where the indices (j, k, e) are in the range (16) but with $j \leq d - 1$, which is a set of size $[N]^d - 1$. Note that with this normalization, the transform is orthonormal.

2.2 Gradient versus wavelet sparsity

The following is a corollary of a remarkable result from Cohen, Dahmen, Daubechies, and De-Vore [15] which bounds the compressibility of a function's wavelet representation by its bounded variation, and will be very useful in our analysis.

Proposition 2 (Corollary of Theorem 1.1 from [15])**.** *There is a universal constant $C > 0$ such that the following holds for any mean-zero $\boldsymbol{x} \in \mathbb{C}^{N^d}$ in dimension $d \geq 2$: if the Haar transform coefficients $\boldsymbol{c} = \mathcal{H}(\boldsymbol{x})$ are partitioned by their support into blocks $\boldsymbol{c}_{j,k} = (\langle \boldsymbol{h}_{j,k,e}, \boldsymbol{x} \rangle)_{e \in V}$ of cardinality $|\boldsymbol{c}_{j,k}| = 2^d - 1$, then the coefficient block of kth largest ℓ_2-norm, denoted by $\boldsymbol{c}_{(k)}$, has ℓ_2-norm bounded by*

$$\|\boldsymbol{c}_{(k)}\|_2 \leq C \frac{\|\boldsymbol{x}\|_{TV_1}}{k \cdot 2^{d/2-1}}$$

Thus by equivalence of the anisotropic and isotropic total variation seminorms up to a factor of $\sqrt{d}$,

$$\|\boldsymbol{c}_{(k)}\|_2 \leq C \frac{\sqrt{d}\|\boldsymbol{x}\|_{TV_2}}{k \cdot 2^{d/2-1}}.$$

6

Proposition 2, whose derivation from Theorem 1.1 of [15] is outlined in the appendix, will be crucial in our proofs of robust recovery via total variation.

3 The main result

Our main result concerns near-optimal recovery guarantees for multidimensional total variation minimization from compressed measurements. Recall that a linear operator $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^r$ is said to have the restricted isometry property (RIP) of order s and level $\delta \in (0, 1)$ when

$$(1 - \delta)\|\boldsymbol{x}\|_2^2 \leq \|\mathcal{A}(\boldsymbol{x})\|_2^2 \leq (1 + \delta)\|\boldsymbol{x}\|_2^2 \qquad \text{for all } s\text{-sparse } \boldsymbol{x} \in \mathbb{C}^{N^d}. \tag{18}$$

A linear operator $\mathcal{A} = (\boldsymbol{a}_k) : \mathbb{C}^{N^d} \to \mathbb{C}^r$ satisfies the RIP if and only if the $r \times N^d$ matrix $\boldsymbol{A}$ whose kth row consists of the unraveled entries of the kth multidimensional array $\boldsymbol{a}_k$ satisfies the classical RIP, (1), and so without loss of generality we treat both definitions of the RIP as equivalent.

For our main result it will be convenient to define for a multidimensional array $\boldsymbol{a} \in \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}}$ the associated arrays $\boldsymbol{a}_{0_\ell} \in \mathbb{C}^{N^d}$ and $\boldsymbol{a}^{0_\ell} \in \mathbb{C}^{N^d}$ obtained by concatenating a block of zeros to the beginning and end of $\boldsymbol{a}$ oriented in the ℓth direction:

$$(\boldsymbol{a}^{0_\ell})_\alpha = \begin{cases} 0, & \alpha_\ell = 1 \\ a_{\alpha_1, \ldots, \alpha_\ell - 1, \ldots, \alpha_d}, & 2 \leq \alpha_\ell \leq N \end{cases} \tag{19}$$

and

$$(\boldsymbol{a}_{0_\ell})_\alpha = \begin{cases} 0, & \alpha_\ell = N \\ a_{\alpha_1, \ldots, \alpha_\ell, \ldots, \alpha_d}, & 1 \leq \alpha_\ell \leq N - 1. \end{cases} \tag{20}$$

The following lemma relating gradient measurements with $\boldsymbol{a}$ to signal measurements with $\boldsymbol{a}^{0_\ell}$ and $\boldsymbol{a}_{0_\ell}$ can be verified by direct algebraic manipulation and thus the proof is omitted.

Lemma 3. *Given* $\boldsymbol{x} \in \mathbb{C}^{N^d}$ *and* $\boldsymbol{a} \in \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}}$,

$$\langle \boldsymbol{a}, \boldsymbol{x}_{r_\ell} \rangle = \langle \boldsymbol{a}^{0_\ell}, \boldsymbol{x} \rangle - \langle \boldsymbol{a}_{0_\ell}, \boldsymbol{x} \rangle,$$

where the directional derivative $\boldsymbol{x}_{r_\ell}$ *is defined in* (11).

For a linear operator $\mathcal{A} = (\boldsymbol{a}_k) : \mathbb{C}^{N^{\ell-1} \times (N-1) \times N^{d-\ell}} \to \mathbb{C}^m$ we define the operators $\mathcal{A}^{0_\ell} : \mathbb{C}^{N^d} \to \mathbb{C}^m$ and $\mathcal{A}_{0_\ell} : \mathbb{C}^{N^d} \to \mathbb{C}^m$ as the sequences of arrays $(\boldsymbol{a}_k^{0_k})_{k=1}^m$ and $(\boldsymbol{a}_{0_k k})_{k=1}^m$, respectively. We conclude from Lemma 3 that $\mathcal{A}(\boldsymbol{x}_{r_\ell}) = \mathcal{A}^{0_\ell}(\boldsymbol{x}) - \mathcal{A}_{0_\ell}(\boldsymbol{x})$.

We are now prepared to state our main result which shows that total variation minimization yields stable recovery of N^d-dimensional signals from RIP measurements.

Main Theorem. *Let* $N = 2^n$. *Fix integers* p *and* q. *Let* $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^p$ *be such that, composed with the orthonormal Haar wavelet transform,* $\mathcal{A}\mathcal{H}^* : \mathbb{C}^{N^d} \to \mathbb{C}^p$ *has the restricted isometry property of order* $2ds$ *and level* $\delta < 1$. *Let* $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_d$ *with* $\mathcal{B}_j : \mathbb{C}^{N^{d-1}(N-1)} \to \mathbb{C}^q$ *be such that* $\mathcal{B} = \mathcal{B}_1 \oplus_c \mathcal{B}_2 \oplus_c \cdots \oplus_c \mathcal{B}_d : \mathbb{C}^{N^{d-1}(N-1)} \to \mathbb{C}^{dq}$ *has the restricted isometry property of order* $5ds$ *and level* $\delta < 1/3$. *Set* $m = 2dq + p$, *and consider the linear operator* $\mathcal{M} : \mathbb{C}^{N^d} \to \mathbb{C}^m$ *given by*

$$\mathcal{M} = \mathcal{A} \oplus_r \left[\mathcal{B}_1 \right]^{0_1} \oplus_r \left[\mathcal{B}_1 \right]_{0_1} \oplus_r \cdots \oplus_r \left[\mathcal{B}_\ell \right]^{0_\ell} \oplus_r \left[\mathcal{B}_\ell \right]_{0_\ell} \oplus_r \cdots \oplus_r \left[\mathcal{B}_d \right]^{0_d} \oplus_r \left[\mathcal{B}_d \right]_{0_d}. \tag{21}$$

7

The following holds for any $\boldsymbol{x} \in \mathbb{C}^{N^d}$. From noisy measurements $\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}) + \boldsymbol{\xi}$ with noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$, the solution to

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{z}} \|\boldsymbol{x}\|_{TV_2} \quad such\ that \quad \|\mathcal{M}(\boldsymbol{z}) - \boldsymbol{y}\|_2 \leq \varepsilon \tag{22}$$

satisfies:

i) $\|\nabla(\boldsymbol{x} - \hat{\boldsymbol{x}})\|_2 \lesssim \frac{\|\nabla\boldsymbol{x} - (\nabla\boldsymbol{x})_S\|_{1,2}}{\sqrt{s}} + \sqrt{d}\varepsilon,$

ii) $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_{TV_2} \lesssim \|\nabla\boldsymbol{x} - (\nabla\boldsymbol{x})_S\|_{1,2} + \sqrt{sd}\varepsilon,$

iii) $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \lesssim \log(N^d)\Big(\frac{\|\nabla\boldsymbol{x} - (\nabla\boldsymbol{x})_S\|_{1,2}}{\sqrt{s}} + \sqrt{d}\varepsilon\Big),$

where $\|\boldsymbol{z}\|_{TV_2} = \|\nabla\boldsymbol{z}\|_{1,2}$ is the isotropic total variation seminorm, $\|\boldsymbol{x}\|_{1,2} = \sum_{\alpha \in [N]^d} \big(\sum_{\ell=1}^d x_{\alpha,\ell}^2\big)^{1/2}$ is the associated mixed $\ell_1 - \ell_2$ norm, and $(\nabla\boldsymbol{x})_S$ is the signal gradient $\nabla\boldsymbol{x}$ restricted to the subset $S = \big\{\{\alpha_{(1)}, \ldots, \alpha_{(s)}\} \times [d]\big\}$ of s largest-magnitude block norms $\|(\nabla\boldsymbol{x})_{\alpha_{(k)}}\|_2$.

Remarks.

  1. Following the same lines of reasoning as the proof of this Main Theorem, one may derive error bounds for *anisotropic* TV minimization that are a bit tighter, namely, one only requires RIP of order $5s$. We believe that the suboptimal bounds for isotropic TV are merely an artifact of our proof technique, as isotropic TV minimization is preferred in practice.

  2. The third bound shows that the reconstruction error is proportional (up to a logarithmic factor) to the noise level ε and the *tail of the gradient* of the signal $\boldsymbol{x}$. A number of $m \approx sd \log(N^d)$ i.i.d. and properly normalized Gaussian measurements can be used to construct the measurement operator $\mathcal{M}$ which, with high probability, satisfies the required RIP conditions of the theorem [1, 47]. From this number m of measurements, the error guarantees are optimal up to the factor of d required RIP measurements, factor of $\sqrt{d}$ on the noise dependence, and logarithmic factor in the signal dimension N^d. We emphasize here that the specific construction of the measurement ensemble is likely only an artifact of the proof, and that more general RIP measurements are likely possible. See also [30] for results using Fourier measurements (for $d = 2$).

  3. The main theorem recovers the total variation guarantees of [38] when $d = 2$ up to a $\log(1/s)$ term. This term is lost only because in the higher-dimensional analysis, our proofs require blocking of the wavelet coefficients. This term can be recovered by applying a more efficient blocking strategy and by writing in terms of p in (32) of the proof. We write the bound as-is for simplicity.

  4. The requirement of sidelength $N = 2^n$ is not an actual restriction, as signals with arbitrary side-length N can be extended via reflections across each dimension to a signal of side-length $N = 2^n$ without increasing the total variation by more than a factor of 2^d. This requirement again seems to be only an artifact of the proof and one need not perform such changes in practice.

  We now turn to the proof of the main theorem. We will first prove the gradient-level recovery bounds (*i*) and (*ii*), and then use these to prove the signal-level recovery bound (*iii*) via Sobolev inequalities for incoherent subspaces. Along the way, we must take care to balance estimates involving the gradient vectors $(\nabla\boldsymbol{x}_{\alpha,\ell})_{\ell=1}^d \in \mathbb{C}^d$ and their block norms $\|\nabla\boldsymbol{x}_\alpha\|_2$, as well as the blocks of wavelet coefficients $\mathbf{c}_{(k)} \in \mathbb{C}^{2^d-1}$ associated to a dyadic cube and their block norms $\|\mathbf{c}_{(k)}\|_2$.

4 Stable gradient recovery

In this section we prove statements (i) and (ii) of the main theorem concerning stable gradient recovery, using standard results in compressed sensing combined with a summation by parts trick provided by the specific form of the measurements in the Main Theorem and Lemma 3.

Recall that when a signal obeys a tube and cone constraint we can bound the norm of the entire signal, as in [11]. We refer the reader to Section A.1 of [38] for a complete proof.

Proposition 4. *Suppose that $\mathcal{B}$ is a linear operator satisfying the restricted isometry property of order $5ds$ and level $\delta < 1/3$, and suppose that the signal $\boldsymbol{h}$ satisfies a tube constraint*

$$\|\mathcal{B}(\boldsymbol{h})\|_2 \leq \sqrt{2d}\varepsilon.$$

Suppose further that using the notation of the Main Theorem, for a subset $R = R' \times [d]$ of cardinality $|R| \leq sd$ (meaning $|R'| \leq s$), $\boldsymbol{h}$ satisfies a cone-constraint

$$\|\boldsymbol{h}_{R^c}\|_{1,2} \leq \|\boldsymbol{h}_R\|_{1,2} + \sigma. \tag{23}$$

Then

$$\|\boldsymbol{h}\|_2 \lesssim \frac{\sigma}{\sqrt{s}} + \sqrt{d}\varepsilon \tag{24}$$

and

$$\|\boldsymbol{h}\|_{1,2} \lesssim \sigma + \sqrt{sd}\varepsilon. \tag{25}$$

Proposition 4 generalizes results in [8] and its proof is included in the appendix. Using Proposition 4 and RIP assumptions on the operator $\mathcal{B}$, the gradient-level recovery guarantees (i) and (ii) reduce to proving that the discrete gradient of the residual signal error satisfies the tube and cone constraints.

Proof. (Main Theorem, statements (i) and (ii).)
Let $\boldsymbol{v} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ be the residual error, and set $\boldsymbol{h} = \nabla \boldsymbol{v} = \nabla \boldsymbol{x} - \nabla \hat{\boldsymbol{x}} \in \mathbb{C}^{N^d \times d}$.
Then we have

Cone Constraint. Consider the block norm $\|(\nabla\mathbf{x})_\alpha\|_2 = \left(\sum_{\ell=1}^d x_{\alpha,\ell}^2\right)^{1/2}$ associated to the index $\alpha \in [N]^d$, and denote by $\alpha_{(j)}$ the index of the jth largest block norm $\|(\nabla\mathbf{x})_{\alpha_{(j)}}\|_2$, and let $S = \{\alpha_{(1)}, \ldots, \alpha_{(s)}\} \times [d]$. Since $\hat{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{v}$ is a minimizer of (TV) and $\boldsymbol{x}$ satisfies the feasibility constraint in (TV), we have that $\|\nabla\hat{\boldsymbol{x}}\|_{1,2} \leq \|\nabla\boldsymbol{x}\|_{1,2}$. By the reverse triangle inequality,

$$\begin{aligned}
\|(\nabla\boldsymbol{x})_S\|_{1,2} - \|\boldsymbol{h}_S\|_{1,2} - \|(\nabla\boldsymbol{x})_{S^c}\|_{1,2} + \|\boldsymbol{h}_{S^c}\|_{1,2} & \\
&\leq \|(\nabla\boldsymbol{x})_S - \boldsymbol{h}_S\|_{1,2} + \|(\nabla\boldsymbol{x})_{S^c} - \boldsymbol{h}_{S^c}\|_{1,2} \\
&= \|\nabla\hat{\boldsymbol{x}}\|_{1,2} \\
&\leq \|\nabla\boldsymbol{x}\|_{1,2} \\
&= \|(\nabla\boldsymbol{x})_S\|_{1,2} + \|(\nabla\boldsymbol{x})_{S^c}\|_{1,2}.
\end{aligned}$$

This yields the cone constraint

$$\|\boldsymbol{h}_{S^c}\|_{1,2} \leq \|\boldsymbol{h}_S\|_{1,2} + 2\|(\nabla\boldsymbol{x})_{S^c}\|_{1,2}$$

9

Tube constraint. Recall that $\boldsymbol{v} = \boldsymbol{x} - \hat{\boldsymbol{x}}$. Since both $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ are feasible solutions to (TV), Jensen's inequality gives

$$\|\mathcal{M}(\boldsymbol{v})\|_2^2 \leq 2\|\mathcal{M}(\boldsymbol{x}) - \boldsymbol{y}\|_2^2 + 2\|\mathcal{M}(\hat{\boldsymbol{x}}) - \boldsymbol{y}\|_2^2 \leq 4\varepsilon^2$$

By Lemma 3, we have for each component operator $\mathcal{B}_j$,

$$\mathcal{B}_j(\boldsymbol{v}_{r_j}) = [\mathcal{B}_j]^{0_j}(\boldsymbol{v}) - [\mathcal{B}_j]_{0_j}(\boldsymbol{v}) \tag{26}$$

Then $\mathcal{B}(\nabla\boldsymbol{v}) = \sum_{j=1}^{d} \mathcal{B}_j(\boldsymbol{v}_{r_j})$, (where we assume that $\nabla\boldsymbol{v}$ is ordered appropriately) and

$$
\begin{aligned}
\|\mathcal{B}(\nabla\boldsymbol{v})\|_2^2 &= \|\sum_{j=1}^{d} \mathcal{B}_j(\boldsymbol{v}_{r_j})\|_2^2 \\
&\leq d\sum_{j=1}^{d} \|\mathcal{B}_j(\boldsymbol{v}_{r_j})\|_2^2 \\
&\leq 2d\sum_{j=1}^{d} \left(\|[\mathcal{B}_j]^{0_j}(\boldsymbol{v})\|_2^2 + \|[\mathcal{B}_j]_{0_j}(\boldsymbol{v})\|_2^2 \right) \\
&\leq 2d\|\mathcal{M}(\boldsymbol{v})\|_2^2 \\
&\leq 8d\varepsilon^2.
\end{aligned}
\tag{27}
$$

In light of Proposition 4 the proof is complete.

$\square$

Remark 5. *The component operator $\mathcal{A}$ from the main theorem was not used at all in deriving properties (i) and (ii); on the other hand, only the measurements in $\mathcal{A}$ will be used to derive property (iii) from (i) and (ii). We conjecture that all measurements in the main result apart from those in the component operator $\mathcal{A}$ are artifacts of the proof techniques herein.*

5 A Sobolev inequality for incoherent subspaces

We now derive a strengthened Sobolev inequality for signals lying near the null space of a matrix which is incoherent to the Haar wavelet basis.

Theorem 6 (Sobolev inequality for incoherent subspaces)**.** *Let $d \geq 2$ and let $N = 2^n$. Let $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^m$ be a linear map such that, composed with the multivariate Haar wavelet transform $\mathcal{H} : \mathbb{C}^{N^d} \to \mathbb{C}^{N^d}$, the resulting operator $\mathcal{A}\mathcal{H}^* : \mathbb{C}^{N^d} \to \mathbb{C}^m$ satisfies the restricted isometry property of order $2s$ and level $\delta < 1$. Then there is a universal constant $C > 0$ such that the following holds: if $\boldsymbol{v} \in \mathbb{C}^{N^d}$ satisfies the tube constraint $\|\mathcal{A}(\boldsymbol{v})\|_2 \leq \varepsilon$, then*

$$\|\boldsymbol{v}\|_2 \leq C\frac{\|\boldsymbol{v}\|_{TV_1}}{\sqrt{s}} \log(N^d) + \varepsilon. \tag{28}$$

and so, by equivalence of the isotropic and anisotropic total variation seminorms up to a factor of $\sqrt{d}$,

$$\|\boldsymbol{v}\|_2 \leq C\frac{\|\boldsymbol{v}\|_{TV_2}}{\sqrt{s/d}} \log(N^d) + \varepsilon. \tag{29}$$

10

Remark 7. *The RIP assumptions on $\mathcal{A}\mathcal{H}^* = \mathcal{A}\mathcal{H}^{-1}$ imply that for $\boldsymbol{v}$ with s-sparse wavelet representation,*

$$\|\mathcal{A}(\boldsymbol{v})\|_2 = \|\mathcal{A}\mathcal{H}^*\mathcal{H}(\boldsymbol{v})\|_2 \approx (1 \pm \delta)\|\mathcal{H}(\boldsymbol{v})\|_2 \approx (1 \pm \delta)\|\boldsymbol{v}\|_2,$$

with the final equality holding because $\mathcal{H}$ is unitary. This implies that the null space of $\mathcal{A}$ cannot contain any signals admitting an s-sparse wavelet expansion, apart from the zero vector. In particular, the null space of $\mathcal{A}$ contains no nontrivial constant signals and thus intersects the null space of ∇ trivially.

Theorem 6 admits corollaries for various families of random matrices with restricted isometries. For Gaussian random matrices, the theorem implies the following.

Corollary 8. *Let $\mathcal{A} : \mathbb{C}^{N^d} \to \mathbb{C}^m$ be a linear map realizable as an $N^d \times m$ matrix whose entries are mean-zero i.i.d. Gaussian random variables. Then there is a universal constant $c > 0$ such that with probability exceeding $1 - e^{-cm}$, the following bound holds for any $\boldsymbol{x} \in \mathbb{C}^{N^d}$ lying in the null-space of $\mathcal{A}$:*

$$\|\boldsymbol{x}\|_2 \lesssim \frac{\|\boldsymbol{x}\|_{TV_1}}{\sqrt{m}}[\log(N^d)]^2. \tag{30}$$

and

$$\|\boldsymbol{x}\|_2 \lesssim \frac{\sqrt{d}\|\boldsymbol{x}\|_{TV_2}}{\sqrt{m}}[\log(N^d)]^2. \tag{31}$$

Proof of Corollary 8. From results on Gaussian matrices and the restricted isometry property (see e.g. [9, 36, 1, 47]), $\mathcal{A}$ satisfies the RIP of order $2s$ and level $\delta < 1$ with probability exceeding $1 - e^{-cm}$ when s is proportional to $m/\log(N^d)$. Substituting this value for s into (28) and (29) yields the claim. $\square$

Proof of Theorem 6. Consider the signal error $\boldsymbol{v} = \boldsymbol{x} - \hat{\boldsymbol{x}}$, and consider its orthogonal decomposition $\boldsymbol{v} = \boldsymbol{v^0} + \boldsymbol{v^1}$ where $\boldsymbol{v^0} = \langle \boldsymbol{h}_0, \boldsymbol{v} \rangle \boldsymbol{h}_0$ is constant and $\boldsymbol{v^1} = \boldsymbol{v} - \boldsymbol{v^0}$ is mean-zero (recall that $\boldsymbol{h}_0$ has constant entries equal to $N^{-d/2}$). Let $\boldsymbol{c} = \mathcal{H}(\boldsymbol{v}) \in \mathbb{C}^{N^d}$ represent the orthonormal Haar transform of $\boldsymbol{v}$. Suppose without loss of generality that the desired sparsity level s is either smaller than $2^d - 1$ or a positive multiple of $2^d - 1$, and write $s = p(2^d - 1)$ where either $p \in \mathbb{N}$ or $p \in (0, 1)$ (for arbitrary $s \in \mathbb{N}$, we could consider $s' = \lceil s/(2^d - 1) \rceil$ which satisfies $s' \leq 2s$).

Let $S = S_0 \subset [N]^d$ be the index set of cardinality s which includes the constant Haar coefficient $c_0 = \langle \boldsymbol{h}_0, \boldsymbol{v} \rangle$ along with the $s - 1$ largest-magnitude entries of $\boldsymbol{c}$ (not including c_0). Let S_1 be the set of s largest-magnitude entries of $\boldsymbol{c}$ in $[N]^d \setminus S_0$, and so on. Note that $\boldsymbol{c}_S$ and similar expressions above and below can have both the meaning of restricting $\boldsymbol{c}$ to the indices in S as well as being the array whose entries are set to zero outside S.

Now, since $\boldsymbol{v^1}$ is mean-zero and $\boldsymbol{c}_{S_0^c} = \mathcal{H}(\boldsymbol{v^1})_{S_0^c}$, we may apply Proposition 2. To that end, consider the decomposition of $\boldsymbol{c}$ into blocks $\boldsymbol{c}_{(k)}$ of cardinality $2^d - 1$ as in Proposition 2, according to the support of their wavelets. By definition, $\|\boldsymbol{c}_{S_0}\|_1$ is at least as large as $\|\boldsymbol{c}_\Omega\|_1$ for any other $\Omega \subset [N]^d$ of cardinality $s - 1$. Consequently, $\|\boldsymbol{c}_{S_0^c}\|_1$ is *smaller* than $\|\boldsymbol{c}_{\Omega^c}\|_1$ for any other $\Omega \subset [N]^d$

11

of cardinality $s - 1$. Thus, because $s = p(2^d - 1)$,

$$
\begin{aligned}
\|\boldsymbol{c}_{S_0^c}\|_1 &\leq \sum_{j \geq p+1} \|\boldsymbol{c}_{(j)}\|_1 \\
&\leq (2^d - 1)^{1/2} \sum_{j \geq p+1} \|\boldsymbol{c}_{(j)}\|_2 \\
&\lesssim \|\boldsymbol{v}\|_{TV_1} \sum_{\ell=p+1}^{N^d} \frac{1}{\ell} \\
&\lesssim \|\boldsymbol{v}\|_{TV_1} \log(N^d),
\end{aligned}
\tag{32}
$$

where the second to last inequality follows from Proposition 2 and the last inequality from properties of the geometric summation.

We use a similar procedure to bound the ℓ_2-norm of the residual,

$$
\begin{aligned}
\|\boldsymbol{c}_{S_0^c}\|_2^2 &\lesssim \sum_{j \geq p+1} \|\boldsymbol{c}_{(j)}\|_2^2 \\
&\lesssim \frac{\|\boldsymbol{v}\|_{TV_1}^2}{2^d} \sum_{\ell=p+1}^{N^d} \frac{1}{\ell^2} \\
&\lesssim \frac{(\|\boldsymbol{v}\|_{TV_1})^2}{2^d \max(1, p)} \\
&\lesssim \frac{(\|\boldsymbol{v}\|_{TV_1})^2}{s}.
\end{aligned}
\tag{33}
$$

Then, $\|\boldsymbol{c}_{S_0^c}\|_2 \lesssim \|\boldsymbol{v}\|_{TV_1}/\sqrt{s}$.

By assumption, $\boldsymbol{v}$ satisfies the tube constraint $\|\mathcal{A}(\boldsymbol{v})\|_2 \leq \varepsilon$ and $\mathcal{A}\mathcal{H}^* = \mathcal{A}\mathcal{H}^{-1}$ satisfies the restricted isometry property of order $5ds$. We conclude that

$$
\begin{aligned}
\varepsilon &\geq \|\mathcal{A}(\boldsymbol{v})\|_2 = \|\mathcal{A}\mathcal{H}^*(\boldsymbol{c})\|_2 \\
&\geq \|\mathcal{A}\mathcal{H}^*(\boldsymbol{c}_{S_0} + \boldsymbol{c}_{S_1})\|_2 - \sum_{k=2}^{r} \|\mathcal{A}\mathcal{H}^*(\boldsymbol{c}_{S_k})\|_2 \\
&\geq (1 - \delta)\|\boldsymbol{c}_{S_0} + \boldsymbol{c}_{S_1}\|_2 - (1 + \delta) \sum_{k=2}^{r} \|\boldsymbol{c}_{S_k}\|_2 \\
&\geq (1 - \delta)\|\boldsymbol{c}_{S_0}\|_2 - (1 + \delta) \frac{1}{\sqrt{s}} \|\boldsymbol{c}_{S_0^c}\|_1,
\end{aligned}
\tag{34}
$$

the last inequality holding because the magnitude of each entry in the array $\boldsymbol{c}_{S_k}$ is smaller than the average magnitude of the entries in the array $\boldsymbol{c}_{S_{k-1}}$. Along with the tail bound (32), we can then conclude that, up to a constant in the restricted isometry level δ,

$$
\|\boldsymbol{c}_{S_0}\|_2 \lesssim \varepsilon + \log(N^d)\Big(\frac{\|\boldsymbol{v}\|_{TV_1}}{\sqrt{s}} \Big).
\tag{35}
$$

Combining this bound with the ℓ_2-tail bound (33) and recalling that the Haar transform is unitary, we find that

$$
\|\boldsymbol{v}\|_2 = \|\mathcal{H}^* \boldsymbol{c}\|_2 = \|\boldsymbol{c}\|_2 \leq \|\boldsymbol{c}_{S_0}\|_2 + \|\boldsymbol{c}_{S_0^c}\|_2 \lesssim \varepsilon + \log(N^d)\Big(\frac{\|\boldsymbol{v}\|_{TV_1}}{\sqrt{s}} \Big),
\tag{36}
$$

which completes the proof.

□

5.1 Proof of the Main Theorem

Because we proved the bounds (i) and (ii) from the main theorem concerning gradient-level recovery bounds in Section 4, it remains only to prove the signal recovery error bound (iii).

By feasibility of both $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ for the constraint in the total variation minimization program, the signal error $\boldsymbol{v} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ obeys the tube-constraint $\|\mathcal{A}(\boldsymbol{v})\|_2 \leq 2\varepsilon$. Applying Theorem 6 with RIP of order $2sd$ (in place of $2s$) and the total variation bound (ii) yields

$$
\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 = \|\boldsymbol{v}\|_2
$$

$$
\lesssim \varepsilon + \log(N^d)\Big(\frac{\|\boldsymbol{v}\|_{TV_1}}{\sqrt{ds}}\Big)
$$

$$
\lesssim \varepsilon + \log(N^d)\Big(\frac{\|\boldsymbol{v}\|_{TV_2}}{\sqrt{s}}\Big)
$$

$$
\lesssim \varepsilon + \frac{\log(N^d)}{\sqrt{s}}\Big(\|\nabla\boldsymbol{x} - (\nabla\boldsymbol{x})_S\|_{1,2} + \sqrt{sd}\varepsilon\Big)
$$

$$
\lesssim \log(N^d)\Big(\sqrt{d}\varepsilon + \frac{\|\nabla\boldsymbol{x} - (\nabla\boldsymbol{x})_S\|_{1,2}}{\sqrt{s}}\Big).
$$

The proof completes.

.1 Derivation of Proposition 2

Recall that the space $L_p(\Omega)$ $(1 \leq p < \infty)$ for $\Omega \subset \mathbb{R}^d$ consists of all functions f satisfying

$$
\|f\|_{L_p(\Omega)} = \Big(\int_\Omega |f(u)|^p du\Big)^{1/p} < \infty.
$$

The space $\mathrm{BV}(\Omega)$ of functions of bounded variation over the unit cube $Q = [0,1)^d$ is often used as a continuous model for natural images. Recall that a function $f \in L_1(Q)$ has finite bounded variation if and only if its distributional gradient is a bounded Radon measure, and this measure generates the BV seminorm $|f|_{BV(\Omega)}$. More precisely,

Definition 9. *For a vector $\boldsymbol{v} \in \mathbb{R}^d$, we define the difference operator $\Delta_{\boldsymbol{v}}$ in the direction of $\boldsymbol{v}$ by*

$$
\Delta_{\boldsymbol{v}}(f, \boldsymbol{x}) := f(\boldsymbol{x} + \boldsymbol{v}) - f(\boldsymbol{x}).
$$

We say that a function $f \in L_1(Q)$ is in $BV(Q)$ if and only if

$$
V_Q(f) \overset{\text{def}}{=} \sup_{h>0} h^{-1} \sum_{j=1}^d \|\Delta_{h\boldsymbol{e}_j}(f, \cdot)\|_{L_1(Q(h\boldsymbol{e}_j))} = \lim_{h\to 0} h^{-1} \sum_{j=1}^d \|\Delta_{h\boldsymbol{e}_j}(f, \cdot)\|_{L_1(Q(h\boldsymbol{e}_j))} < \infty
$$

where $\boldsymbol{e}_j$ denotes the jth coordinate vector. The function $V_Q(f)$ provides a seminorm for $BV(Q)$:

$$
|f|_{BV(Q)} \overset{\text{def}}{=} V_Q(f).
$$

13

In particular, piecewise constant functions are in the space $BV(Q)$, and we have

Lemma 10. *Let $N = 2^n$. Let $\boldsymbol{x} \in \mathbb{C}^{N^d}$ and let $f \in \Sigma_N^d$ be its isometric embedding as a piecewise constant function. Then $|f|_{BV} \leq N^{-d/2+1}\|\boldsymbol{x}\|_{TV_1}$.*

Proof. For $h < \frac{1}{N}$,

$$\Delta_{h\boldsymbol{e_k}}(f, \boldsymbol{u}) = \begin{cases} N^{d/2}(\boldsymbol{x}_{\boldsymbol{\ell}^{(k)}} - \boldsymbol{x}_{\boldsymbol{\ell}}) & \frac{\ell_i}{N} - h \leq u_i \leq \frac{\ell_i}{N}, \\ 0, & \text{else,} \end{cases}$$

where

$$\ell_i^{(k)} = \begin{cases} \ell_i & i \neq k, \\ \ell_i + 1, & i = k. \end{cases}$$

Thus

$$
\begin{aligned}
|f|_{BV} &= \lim_{h \to 0} \frac{1}{h} \sum_{k=1}^{d} \left[\int_0^1 \int_0^1 \cdots \int_0^1 |f(\boldsymbol{u} + h\boldsymbol{e_k}) - f(\boldsymbol{u})| \, d\boldsymbol{u} \right] \\
&= \sum_{k=1}^{d} N^{d/2} \left[\sum_{\boldsymbol{\ell}} \frac{1}{N^{d-1}} |\boldsymbol{x}_{\boldsymbol{\ell}^{(k)}} - \boldsymbol{x}_{\boldsymbol{\ell}}| \right] \\
&\leq N^{-d/2+1}\|\nabla\boldsymbol{x}\|_1 = N^{-d/2+1}\|\boldsymbol{x}\|_{TV_1}.
\end{aligned}
$$

$\square$

Cohen, Dahmen, Daubechies, and DeVore showed in [15] that the properly normalized sequence of rearranged wavelet coefficients associated to a function $f \in L_2(\Omega)$ of bounded variation is in *weak-ℓ_1*, and its weak-ℓ_1 seminorm is bounded by the function BV seminorm. Using different normalizations to those used in [15] — we use the L_2-normalization for the Haar wavelets as opposed to the L_1-normalization — we consider the Haar wavelet coefficients $f_I^e = \langle f, h_I^e \rangle$ and consider the wavelet coefficient block $f_I = (f_I^e)_{e \in E} \in \mathbb{C}^{2^d-1}$ associated to those Haar wavelets supported on the dyadic cube I. With this notation, Theorem 1.1 of [15] applied to the Haar wavelet system over $L_2(Q)$ reads:

Proposition 11. *Let $d \geq 2$. Then there exists a constant $C > 0$ such that the following holds for all mean-zero $f \in BV(Q)$. Let the wavelet coefficient block with kth largest ℓ_2-norm be denoted by $f_{(k)}$, and suppose that this block is associated to the dyadic cube $I_{j,k}$ with side-length 2^{-j}. Then*

$$\|f_{(k)}\|_2 \leq C \frac{2^{j(d-2)/2}|f|_{BV}}{k}.$$

Proposition 2 results by translating Proposition 11 to the discrete setting of $\mathbb{C}^{N^d}$ and appealing to Lemma 10. We note that a stronger version of this result was provided for the 2-dimensional Haar wavelet basis in [16] and used in the proofs in [38].

.2 Proof of Proposition 4

Assume that the cone and tube constraints are in force. Let $R_1 = R'_1 \times [d] \subset R^c$ contain the $4s$ largest blocks of $\boldsymbol{h}$ on R^c, let $R_2 = R'_2 \times [d]$ contain the next $4s$ largest, and so on. We write $\boldsymbol{h}_{R_j}$ to mean the array $\boldsymbol{h}$ restricted to its elements indexed by R_j, and write $\boldsymbol{h}_\alpha$ to denote the array $(h_{(\alpha,\ell)})_{\ell=1}^d$ at pixel α. Thus for any $\alpha \in R'_{j+1}$,

$$\|\boldsymbol{h}_\alpha\|_2 \le \frac{1}{4s} \sum_{\beta \in R'_j} \|\boldsymbol{h}_\beta\|_2 = \frac{1}{4s} \|\boldsymbol{h}_{R_j}\|_{1,2}.$$

Therefore we have

$$\|\boldsymbol{h}_{R_{j+1}}\|_2^2 = \sum_{\alpha \in R'_{j+1}} \|\boldsymbol{h}_\alpha\|_2^2 \le \sum_{\alpha \in R'_{j+1}} \frac{1}{(4s)^2} \|\boldsymbol{h}_{R_j}\|_{1,2}^2 = \frac{1}{4s} \|\boldsymbol{h}_{R_j}\|_{1,2}^2.$$

Combining this with the cone constraint yields

$$\sum_{j \ge 2} \|\boldsymbol{h}_{R_j}\|_2 \le \frac{1}{\sqrt{4s}} \sum_{j \ge 1} \|\boldsymbol{h}_{R_j}\|_{1,2} = \frac{1}{\sqrt{4s}} \sum_{\alpha \in (R')^c} \|\boldsymbol{h}_\alpha\|_2$$

$$= \frac{1}{\sqrt{4s}} \|\boldsymbol{h}_{R^c}\|_{1,2} \le \frac{1}{\sqrt{4s}} \|\boldsymbol{h}_R\|_{1,2} + \frac{1}{\sqrt{4s}} \sigma \le \frac{1}{2} \|\boldsymbol{h}_R\|_2 + \frac{1}{\sqrt{4s}} \sigma,$$

where in the last line we have utilized the fact that $\|\boldsymbol{h}_R\|_{1,2} \le \sqrt{s} \|\boldsymbol{h}_R\|_2$. Next, the tube constraint gives

$$\sqrt{2d}\varepsilon \ge \|\mathcal{B}(\boldsymbol{h})\|_2 \ge \sqrt{1-\delta} \|\boldsymbol{h}_R + \boldsymbol{h}_{R_1}\|_2 - \sqrt{1+\delta} \sum_{j \ge 2} \|\boldsymbol{h}_{R_j}\|_2$$

$$\ge \sqrt{1-\delta} \|\boldsymbol{h}_R + \boldsymbol{h}_{R_1}\|_2 - \sqrt{1+\delta} \left(\frac{1}{2} \|\boldsymbol{h}_R\|_2 + \frac{1}{\sqrt{4s}} \sigma \right)$$

$$\ge \left(\sqrt{1-\delta} - \frac{1}{2}\sqrt{1+\delta} \right) \|\boldsymbol{h}_R + \boldsymbol{h}_{R_1}\|_2 - \sqrt{1+\delta} \left(\frac{1}{\sqrt{4s}} \sigma \right).$$

Using the fact that $\delta < 1/3$, this implies that

$$\|\boldsymbol{h}_R + \boldsymbol{h}_{R_1}\|_2 \le 10\sqrt{d}\varepsilon + \frac{3}{\sqrt{s}}\sigma.$$

The bound (24) then follows since

$$\|\boldsymbol{h}\|_2 \le \|\boldsymbol{h}_R + \boldsymbol{h}_{R_1}\|_2 + \sum_{j \ge 2} \|\boldsymbol{h}_{R_j}\|_2$$

$$\le \|\boldsymbol{h}_R + \boldsymbol{h}_{R_1}\|_2 + \frac{1}{2} \|\boldsymbol{h}_R\|_2 + \frac{1}{\sqrt{4s}} \sigma$$

$$\lesssim \sqrt{d}\varepsilon + \frac{1}{\sqrt{s}} \sigma.$$

15

Similarly, the bound (25) follows from the cone constraint,

$$\|\boldsymbol{h}\|_{1,2} \leq 2\|\boldsymbol{h}_R\|_{1,2} + \sigma$$
$$\leq 2\sqrt{s}\|\boldsymbol{h}_R\|_2 + \sigma$$
$$\lesssim \sqrt{s}(\sqrt{d}\varepsilon + \frac{1}{\sqrt{s}}\sigma) + \sigma$$
$$= \sqrt{sd}\varepsilon + \sigma,$$

which completes the proof.

Acknowledgment

References

[1] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the Restricted Isometry Property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[2] F. Baus, M. Nikolova, and G. Steidl. Fully smoothed ℓ_1-TV models: Bounds for the minimizers and parameter choice. *J. Math. Imaging Vis.*, pages 1–13, 2013.

[3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[4] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse probl.*, 20(5):1411, 2004.

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

[6] E. Candès, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Hamon. A.*, 31(1):59–73, 2011.

[7] E. Candès and J. Romberg. Signal recovery from random projections. In *Proc. SPIE Conference on Computational Imaging III*, volume 5674, pages 76–86. SPIE, 2005.

[8] 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.

[9] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.

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

[11] 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.

[12] V. Caselles, A. Chambolle, and M. Novaga. The discontinuity set of solutions of the tv denoising problem and some extensions. *Multiscale Model. Sim.*, 6(3):879–894, 2007.

[13] V. Caselles, A. Chambolle, and M. Novaga. Regularity for solutions of the total variation denoising problem. *Rev. Mat. Iberoam.*, 27(1):233–252, 2011.

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

[15] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Harmonic analysis of the space BV. *Revistaua Matematica Iberoamericana*, 19(1):235–263, 2003.

[16] 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.

[17] P. L. Combettes and J. Pesquet. Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212, 2011.

[18] Compressed sensing webpage. `http://www.dsp.ece.rice.edu/cs/`.

[19] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

[20] F. Dibos and G. Koepfler. Global total variation minimization. *SIAM J.Numer. Anal.*, 37(2):646–664, 1999.

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

[22] D. L. Donoho and P. Stark. Uncertainty principles and signal recovery. *SIAM J. on Appl. Math.*, pages 906–931, 1989.

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

[24] M. Grasmair. Locally adaptive total variation regularization. *Lect. Notes Comput. SC*, pages 331–342, 2009.

[25] B. Kai Tobias, U. Martin, and F. Jens. Suppression of MRI truncation artifacts using total variation constrained data extrapolation. *Int. J. Biomedical Imaging*, 2008.

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

[27] S. L. Keeling. Total variation based convex filters for medical imaging. *Appl. Math. Comput.*, 139(1):101–119, 2003.

[28] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *arxiv:1207.0235*, 2012.

[29] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[30] F. Krahmer and R. Ward. Beyond incoherence: stable and robust sampling strategies for compressive imaging. 2012. Submitted.

[31] Y. Liu, T. Mi, and S. Li. Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis. *IEEE T. Inform. Theory*, 58(7):4201–4214, 2012.

[32] Y. Liu and Q. Wan. Total variation minimization based compressive wideband spectrum sensing for cognitive radios. Submitted, 2011.

[33] 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.

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

[35] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty. An efficient algorithm for compressed MR imaging using total variation and wavelets. In *IEEE Conf. Comp. Vision Pattern Recog.*, 2008.

[36] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.

[37] D. Needell and R. Ward. Near-optimal compressed sensing guarantees for anisotropic and isotropic total variation minimization. Technical report, 2012. In preparation.

[38] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. *SIAM J. Imaging Sciences*, 2012. To appear.

[39] B. Nett, J. Tang, S. Leng, and G.H. Chen. Tomosynthesis via total variation minimization reconstruction and prior image constrained compressed sensing (PICCS) on a C-arm system. In *Proc. Soc. Photo-Optical Instr. Eng.*, volume 6913. NIH Public Access, 2008.

[40] M. Nikolova, Y. Wen, and R. Chan. Exact histogram specification for digital images using a variational approach. Technical report, Tech. Report, HAL-00723448, 2012.

[41] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Sim.*, 4(2):460–489, 2005.

[42] 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.

[43] 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.

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

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

[46] W. Ring. Structural properties of solutions to total variation regularization problems. *ESAIM-Math. Model. Num.*, 34(04):799–810, 2000.

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

[48] 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.

[49] S. Vaiter, G. Peyr'e, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE T. Inform. Theory*, 2012. To appear.

[50] J. Yuan, C. Schnörr, and G. Steidl. Total-variation based piecewise affine regularization. In X.-C. Tai, K. Mórken, M. Lysaker, and K.-A. Lie, editors, *Lect. Notes Comput. SC*, volume 5567 of *LNCS*, pages 552–564. Springer, 2009.

[51] H. Zhang, W. Yin, and L. Cheng. Necessary and sufficient conditions of solution uniqueness in ℓ_1 minimization. Preprint, 2012.