TV-MIN AND GREEDY PURSUIT FOR CONSTRAINED JOINT SPARSITY AND APPLICATION TO INVERSE SCATTERING

ALBERT FANNJIANG

Abstract. This paper proposes an appealing framework for analyzing total variation minimization (TV-min) and extends Candès, Romberg and Tao's proof of exact recovery of piecewise constant objects with noiseless incomplete Fourier data to the case of noisy data. The approach is based on reformulation of TV-min as compressed sensing of constrained joint sparsity (CJS). TV- and 2-norm error bounds, independent of the ambient dimension, are derived for Basis Pursuit and for Orthogonal Matching Pursuit for CJS.

1. Introduction

One of the most promising developments in imaging and signal processing of the last decade is compressive sensing (CS) which promises reconstruction with fewer data than the ambient dimension. The CS capability [5, 15] hinges on favorable sensing matrices and enforcing a key prior knowledge, i.e. sparseness of the object.

Consider the linear inverse problem $Y = \mathbf{\Phi}X + E$ where $X \in \mathbb{C}^m$ is the *sparse* object vector to be recovered, $Y \in \mathbb{C}^n$ is the measurement data vector and $E \in \mathbb{C}^n$ represents the (model or external) errors. The great insight of CS is that the sparseness of X, as measured by the sparsity $\|X\|_0 \equiv \#$ nonzero elements in X, can be effectively enforced by ℓ^1-minimization [10, 17]

$$\min \|Z\|_1 \quad \text{subject to (s.t.)} \quad \|\mathbf{\Phi}Z - Y\|_2 \le \|E\|_2. \tag{1}$$

The idea of ℓ^1-minimization dates back to geophysics research in 1970's [12, 29]. The ℓ^1-minimizer is often a much better approximation to the sparse object than the traditional minimum energy solution via ℓ^2-minimization because 1-norm is closer to $\|\cdot\|_0$ than the 2-norm. Moreover, the ℓ^1-min principle is a convex optimization problem and can be efficiently computed. The ℓ^1-min principle is effective in recovering the sparse object with the number of data n much less than the ambient dimension m if the sensing matrix $\mathbf{\Phi}$ satisfies some favorable conditions such as the restricted isometry property (RIP) [5]: $\mathbf{\Phi}$ is said to satisfy RIP of order k if

$$(1-\delta_k)\|Z\|_{2,2}^2 \le \|\mathbf{\Phi}Z\|_2^2 \le (1+\delta_k)\|Z\|_2^2 \tag{2}$$

for any k-sparse vector Z where δ_k is the restricted isometry constant (RIC) of order k. RIP has been directly established for only a few special types of matrices including independently and identically distributed (i.i.d.) random matrices and random partial Fourier matrix which is the randomly selected row submatrices of the discrete Fourier transform. Alternatively, CS

techniques also enjoy performance guarantee under incoherence as measured by the mutual coherence defined by

$$\mu(\mathbf{\Phi}) = \max_{i \neq j} \frac{|\sum_k \Phi^*_{ik} \Phi_{kj}|}{\sqrt{\sum_k |\Phi_{ki}|^2} \sqrt{\sum_k |\Phi_{kj}|^2}}. \tag{3}$$

[15, 16]

A parallel development in image denoising pioneered by Osher and coworkers [27, 28] seeks to enforce edge detection by total variation minimization (TV-min)

$$\min \int |\nabla g| \quad \text{s.t.} \quad \int |g - f|^2 \leq \varepsilon^2 \tag{4}$$

where f is the noisy image and ε is the noise level. The idea is that for the class of piecewise constant functions, the gradient is sparse and can be effectively enforced by TV-minimization.

For digital images, TV-minimization approach to deblurring can be formulated as follows. Let $f \in \mathbb{C}^{p \times q}$ be a noisy complex-valued data of $p \times q$ pixels. Let T be the transformation from the true object to the noiseless measurement, modeling physical processes such as propagation and diffraction. Replacing the total variation in (4) by the discrete total variation

$$\begin{aligned} \|g\|_{\text{TV}} &\equiv \sum_{i,j} \sqrt{|\Delta_1 g(i,j)|^2 + |\Delta_2 g(i,j)|^2}, \\ \Delta g = (\Delta_1 g, \Delta_2 g)(i,j) &\equiv (g(i+1,j) - g(i,j), g(i,j+1) - g(i,j)) \end{aligned}$$

we obtain

$$\min \|g\|_{\text{TV}} \quad \text{s.t.} \quad \|Tg - f\|_2 \leq \varepsilon \tag{5}$$

cf. [7, 9].

In a breakthrough paper [3], Candès *et al.* show the equivalence of (5) to (1) for a random partial Fourier matrix with noiseless data ($\varepsilon = 0$) and obtain a performance guarantee of exact reconstruction of piece-wise constant objects from (5).

The main purpose of this note is to extend the result of [3] to inverse scattering with noisy data. In this context it is natural to work with the continuum setting in which the object is a vector in an infinite dimensional function space, e.g. $\ell^2(\mathbb{R}^d)$. To tap into CS techniques, we discretize the object function by pixelating the ambient space with a regular grid of equal spacing ℓ. The grid spacing ℓ can be thought of as the resolution length and the most fundamental parameter of the discrete model from which all other parameters are derived. For example, the total number of resolution cells is proportional to ℓ^{-d}. As we assume that the original object is well approximated by the discrete model in the limit $\ell \to 0$, the sparsity s of the edges of a piecewise constant object is proportional to ℓ^{1-d}, i.e. the object is non-fractal. It is important to keep in mind the continuum origin of the discrete model in order to avoid confusion about the small ℓ limit throughout the paper.

Specifically we aim at the following error bounds: Let V be the discretized object and $\hat{V}$ a reconstruction of V. We will propose a compressive sampling scheme that leads to the error

bound for the TV-minimizer $\hat{V}$ (Theorem 3, Section 4)

$$\|V - \hat{V}\|_{\mathrm{TV}} = \mathcal{O}(\varepsilon), \quad \ell \to 0 \tag{6}$$

implying via the discrete Poincare inequality that

$$\|V - \hat{V}\|_2 = \mathcal{O}(\frac{\varepsilon}{\ell}) \tag{7}$$

independent of the ambient dimension.

If $\hat{V}$ is the reconstruction by using a vectorial version of the greedy algorithm, Orthogonal Matching Pursuit (OMP), then in addition to (6) we also have

$$\|V - \hat{V}\|_2 = \mathcal{O}(\frac{\varepsilon}{\sqrt{\ell}}) \tag{8}$$

independent of the ambient dimension (Section 6). We do not know if the bound (8) applies to the TV-minimizer. A key advantage of the greedy algorithm, which is used to prove (8), is the exact recovery of the gradient support (i.e. the edge location) under suitable conditions (Theorem 4, Section 6). On the one hand, TV-min requires fewer data for recovery: $\mathcal{O}(s)$ for TV-min under RIP versus $\mathcal{O}(s^2)$ for the greedy algorithm under incoherence where the sparsity s of the object gradient is proportional to ℓ^{1-d} for $\ell \ll 1$. On the other hand, the greedy algorithm is computationally more efficient and moreover incoherent measurement is easier to design and verify in practice.

We emphasize that the scope of our approach is general and by no means limited to inverse scattering discussed in detail in the present work. At heart our theory is based on reformulation of TV-min as CS of joint sparsity with linear constraints (such as curl-free constraint in the case of TV-min): For $\mathbf{Y} \in \mathbb{C}^{m \times p}$ define the notation

$$\|\mathbf{Y}\|_{b,a} \;=\; \Big(\sum_{j=1}^{m} \|\mathrm{row}_j(\mathbf{Y})\|_a^b \Big)^{1/b}, \quad a, b \geq 1 \tag{9}$$

where $\mathrm{row}_j(\mathbf{Y})$ is the jth row of $\mathbf{Y}$. The 2,2-norm is exactly the Frobenius norm. To avoid confusion with the *subordinate* matrix norm [20], it is best to view $\mathbf{Y}$ as multiple vectors rather than a matrix.

BPDN for constrained joint sparsity (CJS) is

$$\min \|\mathbf{Z}\|_{1,2}, \quad \text{s.t.} \quad \|\mathbf{Y} - \varphi(\mathbf{Z})\|_{2,2} \leq \varepsilon, \quad \mathcal{L}\mathbf{Z} = 0 \tag{10}$$

where

$$\varphi(\mathbf{Z}) = [\mathbf{\Phi}_1 Z_1, \ldots, \mathbf{\Phi}_d Z_d], \quad \mathbf{Z} = [Z_j].$$

Without loss of generality, we assume the matrices $\{\mathbf{\Phi}_j\}$ all have unit-norm columns.

In connection to TV-min, Z_j is the j-th directional gradient of the object V. And from the definition of discrete gradients, it is clear that every measurement of Z_j can be deduced from two measurements of the object V, slightly shifted in the j-th direction with respect to each other. For inverse scattering $\mathbf{\Phi}_j = \mathbf{\Phi}, \forall j$ and $\mathcal{L}$ is the curl-free constraint. Our results, Theorem 2 and Theorem 4, constitute performance guarantees for CJS and hinge respectively on RIP and incoherence of the measurement matrices $\mathbf{\Phi}_j$.

1.1. **Comparison of existing theories.** The gradient-based method of [23] modifies the original Fourier measurements to obtain Fourier measurements of the corresponding vertical and horizontal edge images which are separately reconstructed by the standard CS algorithms. This approach attempts to take advantage of usually lower separate sparsity and is different from TV-min. Nevertheless, a similar 2-norm error bound (Proposition V.2, [23]) to (7) is obtained.

Needell and Ward [22] obtain interesting results for the *anisotropic* total variation (ATV) minimization in terms of the objective function

$$\|g\|_{\mathrm{ATV}} \equiv \sum_{i,j} |\Delta_1 g(i,j)| + |\Delta_2 g(i,j)|.$$

While for *real*-valued objects in two dimensions, the isotropic TV semi-norm is equivalent to the anisotropic version. The two semi-norms are, however, not the same in dimension ≥ 3 and/or for complex-valued objects. A rather remarkable result of [22] is the bound $\|V - \hat{V}\|_2 = \mathcal{O}(\varepsilon)$, modulo a logarithmic factor, for dimension $d = 2$. This is achieved by proving a strong Sobolev inequality for two dimensions under the additional assumption of RIP with respect to the bivariate Haar transform. Unfortunately, this latter assumption prevents the results in [22] from being directly applicable to structured measurement matrices such as Fourier-like matrices which typically have high mutual coherence with any compactly supported wavelet basis when adjacent subbands are present. Their approach also does not guarantee exact recovery of the gradient support.

It is worthwhile to further compare our formulation with the existing ones but unlike [22, 23] we will consider the case of arbitrary dimensions. From our perspective, the approach of [23] is equivalent to solving d standard BPDN's

$$\min \|Z_\tau\|_1, \quad \text{s.t.} \quad \|Y_\tau - \mathbf{\Phi} Z_\tau\|_2 \leq \varepsilon, \quad \tau = 1, \ldots, d.$$

separately without the curl-free constraint $\mathcal{L}$ where Z_τ and Y_τ are, respectively, the τ-th columns of $\mathbf{Z}$ and $\mathbf{Y}$. To recover the original image from the directional gradients, an additional step of consistent integration is necessary and becomes an important part of the approach in [23].

From our perspective, the ATV-min considered in [22] can be reformulated as follows. Let $\tilde{Z} \in \mathbb{C}^{dn}$ be the image gradient vector by stacking the d directional gradients where n is the number of image pixels and let $\tilde{Y} \in \mathbb{C}^{dm}$ be the similarly concatenated data vector. Likewise let $\tilde{\mathbf{\Phi}} = \mathrm{diag}(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_d) \in \mathbb{C}^{dm \times dn}$ be the block-diagonal matrix with blocks $\mathbf{\Phi}_j \in \mathbb{C}^{m \times n}$. Then ATV-min is equivalent to BPDN for a *single* constrained sparse vector

$$\min \|\tilde{Z}\|_1, \quad \text{s.t.} \quad \|\tilde{Y} - \tilde{\mathbf{\Phi}} \tilde{Z}\|_2 \leq \varepsilon, \quad \tilde{\mathcal{L}} \tilde{Z} = 0. \tag{11}$$

where $\tilde{\mathcal{L}}$ is the same constrain $\mathcal{L}$ reformulated for concatenated vectors. Repeating *verbatim* the proofs of Theorems 2 and 4 we obtain the same error bounds as (6)-(8) for ATV-min as formulated in (11) under the same conditions for $\mathbf{\Phi}_j$ separately.

ATV-min is formulated differently in [22]. Instead of image gradient, it is formulated in terms of the image to do without the curl-free constraint. But the concatenated matrix $[\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_d] \in \mathbb{C}^{m \times nd}$ is assumed to satisfy RIP of higher order and $2dm$ measurement data

are used. For $d = 2$, RIP of order $5s$ with $\delta_{5s} < 1/3$ is assumed for $[\mathbf{\Phi}_1, \mathbf{\Phi}_2]$ in [22] which is much more stringent than RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$ for $\mathbf{\Phi}_1, \mathbf{\Phi}_2$ *separately* in (11) (cf. Theorem 2). In particular, $\mathbf{\Phi}_1 = \mathbf{\Phi}_2$ is allowed for (11) (and our constrained joint sparsity framework) but not for [22]. To get the favorable $\mathcal{O}(\varepsilon)$ 2-norm error bound for $d = 2$, additional measurement matrix satisfying RIP with respect to the bivariate Haar basis is needed, which, as commented above, excludes Fourier measurements.

1.2. **Organization.** The rest of the paper is organized as follows. In Section 2, we review the scattering problem starting from the continuum setting and introduce the discrete model. In Section 3, we discuss various sampling schemes including the forward and backward sampling schemes for inverse scattering for point objects. In Section 4 we formulate TV-min for piecewise constant objects as BPDN for CJS. In Section 5, we present a performance guarantee for BPDN for CJS and obtain error bounds. In Section 6, we analyze the greedy approach to sparse recovery of CJS and derive error bounds, including an improved 2-norm error bound. We present numerical examples and conclude in Section 7. We present the proofs in the Appendices.

2. Scattering theory

A monochromatic wave u propagating in a heterogeneous medium characterized by a variable refractive index $n^2(\mathbf{r}) = 1 + V(\mathbf{r})$ is governed by the Helmholtz equation

$$\nabla^2 u(\mathbf{r}) + \omega^2(1 + V(\mathbf{r}))u(\mathbf{r}) = 0 \tag{12}$$

where V describes the medium inhomogeneities. For simplicity, the wave velocity is assumed to be unity and hence the wavenumber ω equals the frequency.

Consider the scattering of the incident plane wave

$$u^{\mathrm{i}}(\mathbf{r}) = e^{i\omega \mathbf{r}\cdot\hat{\mathbf{d}}} \tag{13}$$

where $\hat{\mathbf{d}}$ is the incident direction. The scattered field $u^{\mathrm{s}} = u - u^{\mathrm{i}}$ then satisfies

$$\nabla^2 u^{\mathrm{s}} + \omega^2 u^{\mathrm{s}} = -\omega^2 \nu u \tag{14}$$

which can be written as the Lippmann-Schwinger equation:

$$u^{\mathrm{s}}(\mathbf{r}) \;=\; \omega^2 \int_{\mathbb{R}^d} \nu(\mathbf{r}') \left(u^{\mathrm{i}}(\mathbf{r}') + u^{\mathrm{s}}(\mathbf{r}')\right) G(\mathbf{r}, \mathbf{r}') d\mathbf{r}' \tag{15}$$

where G is the Green function for the operator $-(\nabla^2 + \omega^2)$.

The scattered field necessarily satisfies Sommerfeld's radiation condition

$$\lim_{r\to\infty} r^{(d-1)/2}\Big(\frac{\partial}{\partial r} - i\omega\Big)u^{\mathrm{s}} = 0$$

reflecting the fact that the energy which is radiated from the sources represented by the right hand side of (14) must scatter to infinity. Thus the scattered field has the far-field asymptotic

$$u^{\mathrm{s}}(\mathbf{r}) = \frac{e^{i\omega|\mathbf{r}|}}{|\mathbf{r}|^{(d-1)/2}} \left(A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) + \mathcal{O}(|\mathbf{r}|^{-1})\right), \quad \hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|, \tag{16}$$

where A is the scattering amplitude and d the spatial dimension. In inverse scattering theory, the scattering amplitude is the measurement data determined by the formula [13]

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) = \frac{\omega^2}{4\pi} \int d\mathbf{r}' V(\mathbf{r}') u(\mathbf{r}') e^{-i\omega \mathbf{r}' \cdot \hat{\mathbf{r}}}$$

which under the Born approximation becomes

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) = \frac{\omega^2}{4\pi} \int d\mathbf{r}' V(\mathbf{r}') e^{i\omega \mathbf{r}' \cdot (\hat{\mathbf{d}} - \hat{\mathbf{r}})} \tag{17}$$

For the simplicity of notation we consider the two dimensional case in detail. Let $\mathbb{L} \subset \mathbb{Z}^2$ be a square sublattice of m integral points. Suppose that s point scatterers are located in a square lattice of spacing ℓ

$$\ell\mathbb{L} = \left\{ (x_j, z_j) = \ell(p_1, p_2) : j = (p_1 - 1)\sqrt{m} + p_2, \mathbf{p} = (p_1, p_2) \in \mathbb{L} \right\}. \tag{18}$$

In the context of inverse scattering, it is natural to treat the size of the discrete ambient domain $\ell\mathbb{L}$ being fixed independent of the resolution length ℓ. In particular, $m \sim \ell^{-2}$ in two dimensions.

First let us motivate the inverse scattering sampling scheme in the case of point scatterers and let $V_j, j = 1, ..., m$ be the strength of the scatterers. Let $\mathcal{S} = \left\{ \mathbf{r}_{i_j} : j = 1, ..., s \right\}$ be the locations of the scatterers. Hence $V_j = 0, \forall \mathbf{r}_j \notin \mathcal{S}$.

For point objects the scattering amplitude becomes a finite sum

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) = \frac{\omega^2}{4\pi} \sum_{j=1}^{m} V_j e^{i\omega \mathbf{r}_j \cdot (\hat{\mathbf{d}} - \hat{\mathbf{r}})}. \tag{19}$$

In the Born approximation the exciting field $u(\mathbf{r}_j)$ is replaced by the incident field $u^{\mathrm{i}}(\mathbf{r}_j)$.

3. Sampling schemes

Let $\hat{\mathbf{d}}_l, \hat{\mathbf{r}}_l, l = 1, ..., n$ be various incident and sampling directions for the frequencies $\omega_l, l = 1, ..., n$ to be determined later. Define the measurement vector $Y = (Y_l) \in \mathbb{C}^n$ with

$$Y_l = \frac{4\pi}{\omega^2 \sqrt{n}} A(\hat{\mathbf{r}}_l, \hat{\mathbf{d}}_l, \omega_l), \quad l = 1, ..., n. \tag{20}$$

The measurement vector is related to the object vector $X = (V_j) \in \mathbb{C}^m$ by the sensing matrix $\mathbf{\Phi}$ as

$$Y = \mathbf{\Phi} X + E \tag{21}$$

where E is the measurement error. Let $\theta_l, \tilde{\theta}_l$ be the polar angles of $\hat{\mathbf{d}}_l, \hat{\mathbf{r}}_l$, respectively. The (l, j)-entry of $\mathbf{\Phi} \in \mathbb{C}^{n \times m}$ is

$$n^{-1/2} e^{-i\omega_l \hat{\mathbf{r}}_l \cdot \mathbf{r}_j} e^{i\omega_l \hat{\mathbf{d}}_l \cdot \mathbf{r}_j} = n^{-1/2} e^{i\omega_l \ell (p_2(\sin\theta_l - \sin\tilde{\theta}_l) + p_1(\cos\theta_l - \cos\tilde{\theta}_l))}, \quad j = (p_1 - 1) + p_2. \tag{22}$$

Note that $\mathbf{\Phi}$ has unit-norm columns.

The crux of the method is to transform the scattering matrix into the random Fourier matrix by suitable sampling schemes so that X can be effectively and efficiently reconstructed as the solution of the ℓ^1-minimization (1).

Let (ξ_l, ζ_l) be i.i.d. uniform random variables on $[-1,1]^2$ and let ρ_l, ϕ_l be the polar coordinates as in

$$(\xi_l, \zeta_l) = \rho_l(\cos\phi_l, \sin\phi_l), \quad \rho_l = \sqrt{\xi_l^2 + \zeta_l^2} \leq \sqrt{2} \tag{23}$$

Let the sampling angle $\tilde{\theta}_l$ be related to the incident angle θ_l via

$$\theta_l + \tilde{\theta}_l = 2\phi_l + \pi, \tag{24}$$

and set the frequency ω_l to be

$$\omega_l = \frac{\Omega\rho_l}{\sqrt{2}\sin\frac{\theta_l - \tilde{\theta}_l}{2}} \tag{25}$$

where Ω is a control parameter. Then the entries (22) of the sensing matrix $\mathbf{\Phi}$ have the form

$$e^{i\sqrt{2}\Omega\ell(p_1\xi_l + p_2\zeta_l)}, \quad l = 1, ..., n, \quad p_1, p_2 = 1, ..., \sqrt{m}. \tag{26}$$

We consider two particular sampling schemes: The first one employs multiple frequencies with the sampling angle always in the back-scattering direction resembling the imaging geometry of synthetic aperture radar (SAR); the second employs only single high frequency with the sampling angle in the forward direction, resembling the imaging geometry of X-ray tomography.

I. Backward Sampling This scheme employs $\Omega-$band limited probes, i.e. $\omega_l \in [-\Omega, \Omega]$. This and (25) lead to the constraint:

$$\left|\sin\frac{\theta_l - \tilde{\theta}_l}{2}\right| \geq \frac{\rho_l}{\sqrt{2}}. \tag{27}$$

A simple way to satisfy (24) and (27) is to set

$$\phi_l = \tilde{\theta}_l = \theta_l - \pi, \tag{28}$$

$$\omega_l = \frac{\Omega\rho_l}{\sqrt{2}} \tag{29}$$

$l = 1, ..., n$. In this case the scattering amplitude is sampled exactly in the backward direction, resembling SAR imaging. In contrast, the exact forward sampling with $\tilde{\theta}_l = \theta_l$ almost surely violates the constraint (27).

II. Forward Sampling This scheme employs single frequency probes no less than Ω:

$$\omega_l = \gamma\Omega, \quad \gamma \geq 1, \quad l = 1, ..., n. \tag{30}$$

We set

$$\theta_l = \phi_l + \arcsin\frac{\rho_l}{\gamma\sqrt{2}} \tag{31}$$

$$\tilde{\theta}_l = \phi_l - \arcsin\frac{\rho_l}{\gamma\sqrt{2}}. \tag{32}$$

The difference between the incident angle and the sampling angle is

$$\theta_l - \tilde{\theta}_l = 2 \arcsin \frac{\rho_l}{\gamma\sqrt{2}} \tag{33}$$

which diminishes as $\gamma \to \infty$. In other words, in the high frequency limit, the sampling angle approaches the incident angle, resembling X-ray tomography [21].

A following performance guarantee then follows from a general result in compressive sensing [18].

Theorem 1. *Let the probe frequencies ω_l, the incident angles θ_l and the sampling angles $\tilde{\theta}_l$ satisfy (24) and (25), for example, by the Backward or Forward Sampling scheme.*

Suppose

$$\Omega\ell = \pi/\sqrt{2} \tag{34}$$

and suppose

$$\frac{n}{\ln n} \geq C_0 \delta^{-2} s \ln^2 s \ln m \ln \frac{1}{\alpha}, \quad \alpha \in (0,1) \tag{35}$$

holds for some constant C_0 and any $\delta < \sqrt{2} - 1$. Then the Basis Pursuit minimizer $\hat{X}$ satisfies

$$\|\hat{X} - X\|_2 \;\; \leq \;\; C_1 s^{-1/2} \|X - X^{(s)}\|_1 + C_2 \varepsilon \tag{36}$$

for some constants C_1 and C_2. with probability at least $1 - \alpha$.

When $V = (V_j)$ is not sparse, but $\Delta V = (\Delta_1 V, \Delta_2 V)$ is, as in the case of piecewise constant objects, then it is natural to consider the TV-min formulation (5).

Our goal is to extend the error bound (36) to the case of piecewise constant objects.

4. Inverse scattering as TV-min

Let us consider the following class of piecewise constant object:

$$V(\mathbf{r}) \;\; = \;\; \sum_{\mathbf{p} \in \mathbb{L}} V_{\mathbf{p}} \, \mathrm{I}_{\square}(\frac{\mathbf{r}}{\ell} - \mathbf{p}), \quad \square = \left[-\frac{1}{2}, \frac{1}{2}\right]^2 \tag{37}$$

where $\mathrm{I}_{\square}$ is the indicator function of the unit square $\square$. As remarked in the Introduction, we think of the pixelated V as discrete approximation of some compactly support function on $\mathbb{R}^2$ and having a well-defined limit as $\ell \to 0$.

Let the discrete total variation $\|V\|_{\mathrm{TV}}$ be defined by

$$\|V\|_{\mathrm{TV}} = \sum_{\mathbf{p} \in \mathbb{L}} \sqrt{|V_{\mathbf{p}+\mathbf{e}_1} - V_{\mathbf{p}}|^2 + |V_{\mathbf{p}+\mathbf{e}_2} - V_{\mathbf{p}}|^2}, \quad \mathbf{e}_1 = (1,0), \quad \mathbf{e}_2 = (0,1). \tag{38}$$

The discrete version of (17) is, however, not exactly (19) since extended objects have different scattering properties from those of point objects.

The integral on the right hand side of (17), modulo the discretization error, is

$$\int d\mathbf{r}' V(\mathbf{r}') e^{i\omega \mathbf{r}'\cdot(\hat{\mathbf{d}}-\hat{\mathbf{r}})} = \sum_{\mathbf{p}\in\mathbb{L}} V_{\mathbf{p}} e^{\mathrm{i}\omega\ell\mathbf{p}\cdot(\hat{\mathbf{d}}-\hat{\mathbf{r}})} \int e^{\mathrm{i}\omega\mathbf{r}'\cdot(\hat{\mathbf{d}}-\hat{\mathbf{r}})}\, \mathrm{I}_{\square}(\frac{\mathbf{r}'}{\ell}) d\mathbf{r}'.$$

Now letting $\hat{\mathbf{d}}_l, \hat{\mathbf{r}}_l, \omega_l, l = 1, \cdots, n$ be selected according to Scheme I or II under the condition (34) and substituting them in the above equation, we obtain

$$\begin{aligned}\int d\mathbf{r}' V(\mathbf{r}') e^{i\omega_l \mathbf{r}'\cdot(\hat{\mathbf{d}}_l-\hat{\mathbf{r}}_l)} &= \ell^2 \sum_{\mathbf{p}\in\mathbb{L}} V_{\mathbf{p}} e^{i\pi(p_1\xi_l+p_2\eta_l)} \int_{\square} e^{i\pi(x\xi_l+y\eta_l)} dxdy \\ &= \ell^2 \sum_{\mathbf{p}\in\mathbb{L}} V_{\mathbf{p}} e^{i\pi(p_1\xi_l+p_2\eta_l)} \frac{2\sin(\pi\xi_l/2)}{\pi\xi_l} \frac{2\sin(\pi\eta_l/2)}{\pi\eta_l}.\end{aligned}$$

Let

$$X_j = \ell^2 V_{\mathbf{p}}, \quad j = (p_1-1)\sqrt{m} + p_2$$

and

$$Y_l = \frac{4\pi}{\omega_l^2 \tilde{g}_l \sqrt{n}} A(\hat{\mathbf{r}}_l, \hat{\mathbf{d}}_l, \omega_l) + E_l, \quad l = 1, \cdots, n$$

where

$$\tilde{g}_l = \frac{2\sin(\pi\xi_l/2)}{\pi\xi_l} \frac{2\sin(\pi\eta_l/2)}{\pi\eta_l}$$

where $E = (E_l)$ is the measurement error.

Define the sensing matrix $\mathbf{\Phi} = [\phi_{kp}]$ as

$$\phi_{kp} = \frac{1}{\sqrt{n}} e^{\mathrm{i}\pi(p_1\xi_k+p_2\eta_k)}, \quad p = (p_1-1)\sqrt{m}+p_2, \quad p_1, p_2 = 1, ..., \sqrt{m}. \tag{39}$$

Then (17) can be written as

$$Y = \mathbf{\Phi} X + E. \tag{40}$$

Our goal is to establish the performance guarantee for TV-min

$$\min \|Z\|_{\mathrm{TV}}, \quad \text{subject to} \quad \|Y - \mathbf{\Phi} Z\|_2 \le \|E\|_2. \tag{41}$$

To this end, we need to transform (40) into the form suitable for compressive sensing.

Define

$$\mathbf{X} = \ell^2((\Delta_1 V)_{\mathbf{p}}, (\Delta_2 V)_{\mathbf{p}}) \in \mathbb{C}^{m\times 2}, \quad (\Delta_1 V)_{\mathbf{p}} = V_{\mathbf{p}+\mathbf{e}_1} - V_{\mathbf{p}}, \quad (\Delta_2 V)_{\mathbf{p}} = V_{\mathbf{p}+\mathbf{e}_2} - V_{\mathbf{p}}.$$

Suppose the support of $\{V_{\mathbf{p}+\mathbf{e}_1}, V_{\mathbf{p}+\mathbf{e}_2}\}$ is contained in $\mathbb{L}$. Simple calculation yields that

$$\begin{aligned} Y_l &= \frac{\ell^2}{\sqrt{n}} e^{i\pi\xi_l} \sum_{\mathbf{p}\in\mathbb{L}} V_{\mathbf{p}+\mathbf{e}_1} e^{\mathrm{i}\pi(p_1\xi_l+p_2\eta_l)} \\ &= \frac{\ell^2}{\sqrt{n}} e^{i\pi\eta_l} \sum_{\mathbf{p}\in\mathbb{L}} V_{\mathbf{p}+\mathbf{e}_2} e^{\mathrm{i}\pi(p_1\xi_l+p_2\eta_l)}\end{aligned}$$

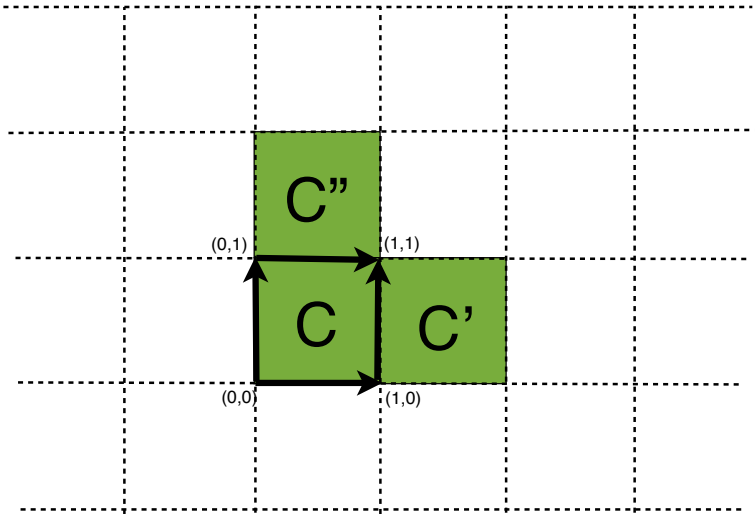

FIGURE 1. Consistency among cells C, C' and C''.

and thus

$$(e^{-\mathrm{i}\pi\xi_l} - 1)Y_l = \frac{\ell^2}{\sqrt{n}} \sum_{\mathbf{p}\in\mathbb{L}} (V_{\mathbf{p}+\mathbf{e}_1} - V_{\mathbf{p}}) e^{\mathrm{i}\pi(p_1\xi_l + p_2\eta_l)} \tag{42}$$

$$(e^{-\mathrm{i}\pi\eta_l} - 1)Y_l = \frac{\ell^2}{\sqrt{n}} \sum_{\mathbf{p}\in\mathbb{L}} (V_{\mathbf{p}+\mathbf{e}_2} - V_{\mathbf{p}}) e^{\mathrm{i}\pi(p_1\xi_l + p_2\eta_l)}. \tag{43}$$

Define the $n \times 2$ data matrix

$$\mathbf{Y} = \big((e^{-\mathrm{i}\pi\xi_l} - 1)Y_l, (e^{-\mathrm{i}\pi\eta_l} - 1)Y_l\big)_{l=1}^n \in \mathbb{C}^{n\times 2}$$

and the $n \times 2$ error matrix

$$\mathbf{E} = \big((e^{-\mathrm{i}\pi\xi_l} - 1)E_l, (e^{-\mathrm{i}\pi\eta_l} - 1)E_l\big)_{l=1}^n \in \mathbb{C}^{n\times 2}. \tag{44}$$

We rewrite (40) in the form

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{E}. \tag{45}$$

subject to the constraint

$$(\Delta_1\Delta_2 V)_{\mathbf{p}} = (\Delta_2\Delta_1 V)_{\mathbf{p}}, \quad \forall \mathbf{p} \in \mathbb{L} \tag{46}$$

which is the discrete version of curl-free condition. This ensures that the reconstruction by line integration of $(V_{\mathbf{p}})$ from $\mathbf{X}$ is consistent.

To see that (46) is necessary and sufficient for the recovery of $(V_{\mathbf{p}})$, consider, for example, the notations in Figure 1 and suppose $V_{0,0}$ is known. By definition of the difference operators Δ_1, Δ_2 we have

$$\begin{aligned} V_{1,0} &= V_{0,0} + (\Delta_1 V)_{0,0} \\ V_{0,1} &= V_{0,0} + (\Delta_2 V)_{0,0} \end{aligned}$$

In general, we can determine $V(\mathbf{p}), \mathbf{p} \in \mathbb{L}$ iteratively from the relationship

$$\begin{aligned} V_{\mathbf{p}+\mathbf{e}_1} &= V_{\mathbf{p}} + (\Delta_1 V)_{\mathbf{p}} \\ V_{\mathbf{p}+\mathbf{e}_2} &= V_{\mathbf{p}} + (\Delta_2 V)_{\mathbf{p}} \end{aligned}$$

and the knowledge of V at any grid point. The path-independence in evaluating V_{p_1+1,p_2+1}

$$\begin{aligned} V_{p_1+1,p_2+1} &= V_{p_1,p_2} + (\Delta_1 V)_{p_1,p_2} + (\Delta_2 V)_{p_1+1,p_2} \\ &= V_{p_1,p_2} + (\Delta_2 V)_{p_1,p_2} + (\Delta_1 V)_{p_1,p_2+1} \end{aligned}$$

implies that

$$(\Delta_2 V)_{p_1+1,p_2} - (\Delta_2 V)_{p_1,p_2} = (\Delta_1 V)_{p_1,p_2+1} - (\Delta_1 V)_{p_1,p_2}$$

which is equivalent to (46).

5. BPDN for constrained joint sparsity

Now eq. (40) is equivalent to (45) with the constraint (46) provided that the value of V at (any) one grid point is known. Our analysis is based on the constrained joint sparsity formulation (45)-(46) and the extension of BPDN to CJS.

Consider the linear inversion problem

$$\mathbf{Y} = \varphi(\mathbf{X}) + \mathbf{E}, \quad \mathcal{L}\mathbf{X} = 0 \tag{47}$$

where

$$\varphi(\mathbf{X}) = [\mathbf{\Phi}_1 X_1, \mathbf{\Phi}_2 X_2, \ldots, \mathbf{\Phi}_d X_d], \quad \mathbf{\Phi}_j \in \mathbb{C}^{m \times n}$$

and the corresponding BPDN

$$\min \|\mathbf{Z}\|_{1,2}, \quad \text{s.t.} \quad \|\mathbf{Y} - \varphi(\mathbf{Z})\|_{2,2} \le \varepsilon = \|\mathbf{E}\|_{2,2}, \quad \mathcal{L}\mathbf{Z} = 0. \tag{48}$$

For TV-min in d dimensions, $p = d$, $\mathbf{\Phi}_j = \mathbf{\Phi}, \forall j$, $\mathbf{Z}$ represents the discrete gradient of the underlying object and $\mathcal{L}$ is the curl-free constraint. Without loss of generality, we assume the matrices $\{\mathbf{\Phi}_j\}$ all have unit-norm columns.

The equivalence between the original TV-min (41) and the gradient-based formulation (48) hinges on the equivalence of their respective feasible sets. When E in (40) is Gaussian noise, then so is $\mathbf{E}$ and their variances are precisely related to each other. If if ε can not be precisely derived, then by the simple observation

$$\|\mathbf{E}\|_{2,2} \le 2\sqrt{2}\|E\|_2$$

in the case of Fourier measurement, we can use the larger feasible set

$$\|\mathbf{Y} - \mathbf{\Phi}\mathbf{Z}\|_{2,2} \leq 2\sqrt{2}\|E\|_2$$

instead to capture the TV minimizer in (41).

CS method relies on the sparseness of the object which is represented by a matrix in the gradient-based formulation (45). We say that $\mathbf{X}$ is s-row sparse if the number of nonzero rows in $\mathbf{X}$ is exactly s.

In the following theorems, we let the object gradient $\mathbf{X}$ be general, not necessarily s-row-sparse. Let $\mathbf{X}^{(s)}$ consist of s largest rows in the 2-norm of $\mathbf{X}$. Then $\mathbf{X}^{(s)}$ is the best s-row-sparse approximation of $\mathbf{X}$.

Theorem 2. *Suppose that the linear map φ satisfies the RIP of order $2s$*

$$(1 - \delta_{2s})\|\mathbf{Z}\|_{2,2}^2 \leq \|\varphi(\mathbf{Z})\|_{2,2}^2 \leq (1 + \delta_{2s})\|\mathbf{Z}\|_{2,2}^2 \tag{49}$$

for any $2s$-row-sparse $\mathbf{Z}$ with

$$\delta_{2s} < \sqrt{2} - 1.$$

Let $\hat{\mathbf{X}}$ be the minimizer of (48). Then

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{2,2} \;\leq\; C_1 s^{-1/2}\|\mathbf{X} - \mathbf{X}^{(s)}\|_{1,2} + C_2\varepsilon \tag{50}$$

for absolute constants C_1, C_2 depending only on δ_{2s}.

Remark 1. *Note that the RIP for joint sparsity (49) follows straightforwardly from the assumption of separate RIP*

$$(1 - \delta_{2s})\|Z\|_2^2 \leq \|\mathbf{\Phi}_j Z\|_2^2 \leq (1 + \delta_{2s})\|Z\|_2^2, \quad \forall j$$

for any $2s$-sparse vector Z.

The proof of Theorem 2 is given in Appendix A.

To conclude our result about TV-min, let us quote the following useful estimate for RIC [25].

Proposition 1. *Suppose*

$$\frac{n}{\ln n} \geq C\delta^{-2} k \ln^2 k \ln m \ln \frac{1}{\alpha}, \quad \alpha \in (0, 1) \tag{51}$$

for given sparsity k where C is an absolute constant. Then the restricted isometry constant of the matrix (26) with (34) satisfies

$$\delta_k \leq \delta$$

with probability at least $1 - \alpha$.

Using Theorem 2 and Proposition 1 with $\mathbf{\Phi}_j = \mathbf{\Phi}$, $k = 2s$ and $\delta < \sqrt{2} - 1$, we obtain the error bound for inverse scattering.

Theorem 3. *Let the probe frequencies ω_l, the incident angles θ_l and the sampling angles $\tilde{\theta}_l$ are determined by by the Backward or Forward Sampling scheme.*

Suppose

$$\Omega\ell = \pi/\sqrt{2} \tag{52}$$

and suppose

$$\frac{n}{\ln n} \geq C_0 \delta^{-2} s \ln^2 s \ln m \ln \frac{1}{\alpha}, \quad \alpha \in (0,1) \tag{53}$$

holds for some constant C_0 *and any* $\delta < \sqrt{2} - 1$. *Then the estimate (50) holds true with probability at least* $1 - \alpha$.

Now suppose the object gradient is s-row-sparse. Theorem 3 says that the number of data needed for an accurate TV-min solution is $\mathcal{O}(s)$, modulo logarithmic factors, which is $\mathcal{O}(\ell^{-1})$ for piecewise constant objects in two dimensions. See [23] for a discussion of practical methods of reconstructing the original object from the recovered gradient $\hat{\mathbf{X}}$.

The error bound (50) implies (6) for a s-row-sparse gradient since $\mathbf{X}, \hat{\mathbf{X}}$ are exactly the gradients of V and $\hat{V}$, respectively. For the 2-norm bound (7), we apply the discrete Poincare inequality [11]

$$\|f\|_2^2 \leq \frac{m^{2/d}}{4d} \|\Delta f\|_2^2$$

to get

$$\|V - \hat{V}\|_2 \leq \frac{m^{1/d}}{2d^{1/2}} C_2 \varepsilon = \mathcal{O}(\frac{\varepsilon}{\ell}). \tag{54}$$

For an improved 2-norm error bound, we turn to the greedy algorithm in the next section.

6. OMP for constrained joint sparsity

One idea to improve the error bound is through exact recovery of the support of the object gradient. This can be achieved by greedy algorithms such as Orthogonal Matching Pursuit (OMP) [14, 16, 24, 30]. As before, we first consider the general linear inversion with joint sparsity (47), except that we assume $\sum_j \|E_j\|_2 = \varepsilon$.

Algorithm 1. OMP for joint sparsity

Input: $\{\mathbf{\Phi}_j\}, \mathbf{Y}, \eta > 0$
Initialization: $\mathbf{X}^0 = 0, \mathbf{R}^0 = \mathbf{Y}$ and $\mathcal{S}^0 = \emptyset$
Iteration:
1) $i_{\max} = \arg\max_i \sum_j |\Phi_{j,i}^* R_j^{k-1}|$
2) $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{i_{\max}\}$
3) $\mathbf{X}^k = \arg\min_{\mathbf{Z}} \sum_j \|\mathbf{\Phi}_j Z_j - Y_j\|_2$ s.t. $\text{supp}(\mathbf{Z}) \in S^k$
4) $\mathbf{R}^k = \mathbf{Y} - \varphi(\mathbf{X}^k)$
5) Stop if $\sum_j \|R_j^k\|_2 \leq \varepsilon$.
Output: $\mathbf{X}^k$.

A natural indicator of the performance of OMP is the mutual coherence defined in (3). Let

$$\mu_{\max} = \max_j \mu(\mathbf{\Phi}_j).$$

Then analogous to Theorem 5.1 of [16], we have the following performance guarantee.

Theorem 4. *Let $X_k = row_k(\mathbf{X})$ be the k-th row of $\mathbf{X}$. Suppose $\sum_j \|E_j\|_2 = \varepsilon$. Suppose the gradient sparsity s (i.e. the number of nonzero rows) satisfies*

$$s < \frac{1}{2}(1 + \frac{1}{\mu_{\max}}) - \frac{\varepsilon}{\mu_{\max} X_{\min}}, \quad X_{\min} = \min_k \|X_k\|_1. \tag{55}$$

Let $\mathbf{Z}$ be the output of OMP as applied to the linear inverse problem (47) without *the constraint, with the stopping rule that the residual drops to the level ε or below. Then $supp(\mathbf{Z}) = supp(\mathbf{X})$.*

Let $\hat{\mathbf{X}}$ solve the least squares problem

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{B}} \sum_j \|Y_j - \mathbf{\Phi}_j B_j\|_2, \quad s.t. \quad supp(\mathbf{B}) \subseteq supp(\mathbf{X}), \quad \mathcal{L}\mathbf{B} = 0. \tag{56}$$

Then

$$\sum_j \|X_j - \hat{X}_j\|_2 \leq \frac{2\varepsilon}{\sqrt{1 - \mu_{\max}(s-1)}}. \tag{57}$$

Remark 2. *The least squares solution $\hat{\mathbf{X}}$ is not necessarily the discrete gradient of the following least squares solution on the level of object itself*

$$\hat{X} = \arg\min_Z \|Y - \mathbf{\Phi} Z\|_2 \tag{58}$$

where the gradient of Z is supported on $supp(\mathbf{X})$, since the objective functions are different.

Remark 3. *For the random partial Fourier measurement matrix, the mutual coherence μ behaves like $\mathcal{O}(n^{-1/2})$ [19]. Therefore (55) impies the sparsity constraint $s = \mathcal{O}(\sqrt{n})$ which is more stringent than (53).*

Remark 4. *For the standard Lasso with a particular choice of regularization parameter, Theorem 1.3 of* [4] *guarantees exact support recovery under a favorable sparsity constraint similar to (53). In our setting and notation, their TV-min principle corresponds to*

$$\min_{\mathcal{L}\mathbf{Z}=0} \lambda\sigma\|\mathbf{Z}\|_{1,2} + \frac{1}{2}\|\mathbf{Y} - \varphi(\mathbf{Z})\|_{2,2}^2, \quad \lambda = 2\sqrt{2\log m} \tag{59}$$

where $\sigma^2 = \varepsilon^2/(2n)$ is the variance of the assumed Gaussian noise in each entry of $\mathbf{Y}$. Unfortunately, even if the result of [4] *can be extended to (59), it is inadequate for our purpose because* [4] *assumes independently selected support and signs, which is clearly not satisfied by the gradient of a piecewise constant object.*

The proof of Theorem 4 is given in Appendix B. The main advantage of Theorem 4 over Theorem 3 is the guarantee of exact recovery of the gradient support. Not only the edges of the object can be exactly recovered, a better 2-norm error bound follows because now the gradient error is guaranteed to vanish outside a set of $\mathcal{O}(\ell^{1-d})$ cardinality: First we have

$$|V_{\mathbf{p}} - \hat{V}_{\mathbf{p}}| \leq \sum_{\mathbf{q}} \|\Delta V_{\mathbf{q}} - \Delta\hat{V}_{\mathbf{q}}\|_2 \tag{60}$$

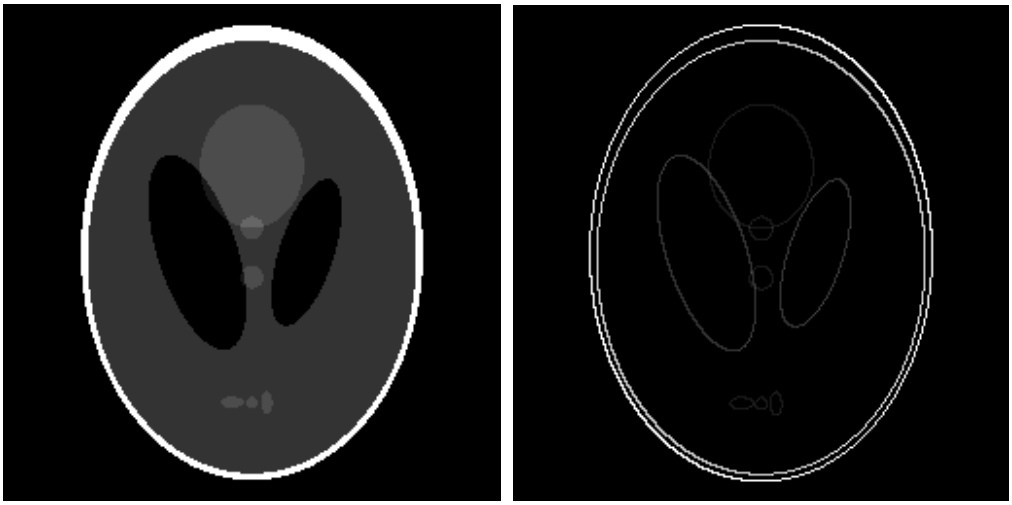

FIGURE 2. The original 256×256 Shepp-Logan phantom (left), the Shepp-Logan phantom and the magnitudes of its gradient with sparsity $s = 2184$.

where the summation runs along any path connecting the given point $\mathbf{p}_0$ to $\mathbf{p}$.

Let $\ell\mathcal{L}_l \subseteq \ell\mathbb{L}, l = 1, ..., L$ be the level sets of the object V such that

$$V(\mathbf{r}) = \sum_{l=1}^{L} v_l \, \mathrm{I}_{\ell\mathcal{L}_l}(\mathbf{r})$$

where $\mathcal{L}_l \cap \mathcal{L}_k = \emptyset, l \neq k, \mathbb{L} = \cup_l \mathcal{L}_l$. The reconstructed object $\hat{V}$ corresponding to $\hat{\mathbf{X}}$ in (57) also takes the same form

$$\hat{V}(\mathbf{r}) = \sum_{l=1}^{L} \hat{v}_l \, \mathrm{I}_{\ell\mathcal{L}_l}(\mathbf{r}).$$

To fix the undetermined constant, we may assume that $v_1 = \hat{v}_1$. Since

$$\sum_j \|\Delta_j(V - \hat{V})\|_2 = \mathcal{O}(\varepsilon)$$

and the gradient error occurs on the perimeter of $\ell\mathcal{L}_l$ whose length is $\mathcal{O}(\ell^{1-d})$, we have

$$|v_l - \hat{v}_l| = \mathcal{O}(\varepsilon \ell^{(d-1)/2}), \quad \forall l.$$

The path in (60) goes through $\mathcal{O}(1)$ number of level sets and thus

$$\|V - \hat{V}\|_\infty = \mathcal{O}(\varepsilon \ell^{(d-1)/2})$$

and

$$\|V - \hat{V}\|_2 = \mathcal{O}(\frac{\varepsilon}{\sqrt{\ell}}).$$

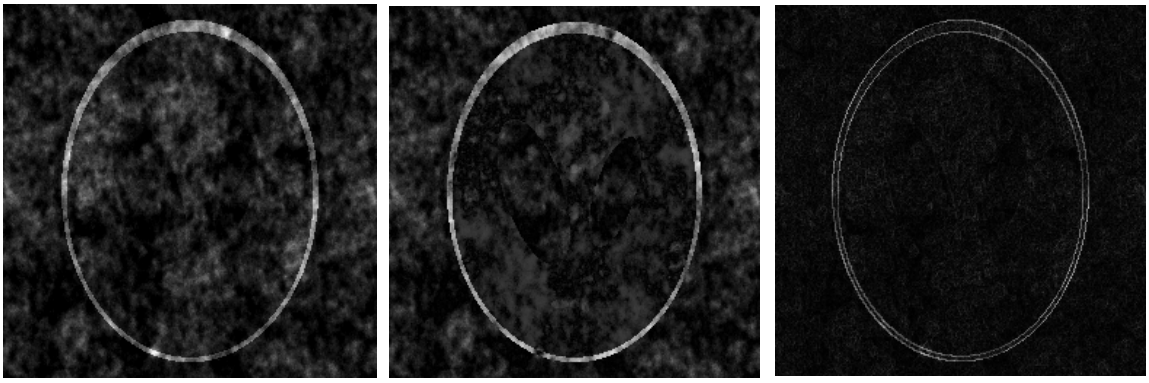

FIGURE 3. Noiseless ℓ^1-min reconstructed image (left) and the differences (middle) from the original image. The plot on the right is the gradient of the reconstructed image.

7. Conclusion

We have developed a general CS theory (Theorems 2 and 4) for constrained joint sparsity and obtained performance guarantees parallel to those for the CS theory for single measurement vector. From the general theory we have derived TV- and 2-norm error bounds, independent of the ambient dimension, for TV-min (Theorem 3) and OMP (Theorem 4) approaches to inverse scattering of piecewise constant objects.

In addition, OMP can recover exactly the gradient support (i.e. the edges of the object) leading to an improved 2-norm error bound. Although OMP needs a higher number of measurement data than TV-min for Fourier measurements the incoherence property required for OMP is much easier to check and often the only practical way to verify RIP when the measurement matrix is not i.i.d. or Fourier.

We end by presenting a numerical example demonstrating the noise stability of the TV-min. Efficient algorithms for TV-min denoising/deblurring exist [1, 31]. We use the open source code *L1-MAGIC* (`http://users.ece.gatech.edu/~justin/l1magic/`) for our simulation.

Figure 2 shows the 256×256 image of the Shepp-Logan Phantom (left) and the modulus of its gradient (right). Clearly the sparsity ($s = 2184$) of the gradient is much smaller than that of the original image. We take 10000 Fourier measurement data for the ℓ^1-min (1) and TV-min (5) reconstructions.

Because the image is not sparse, ℓ^1-min reconstruction produces a poor result even in the absence of noise, Figure 3. The relative error is 66.8% in the ℓ^2 norm and 72.8% in the TV norm. Only the outer perimeter is reasonably recovered.

Figure 4 shows the results of TV-min reconstruction in the presence of 5% (top) or 10% (bottom) noise. Evidently, the performance is greatly improved.

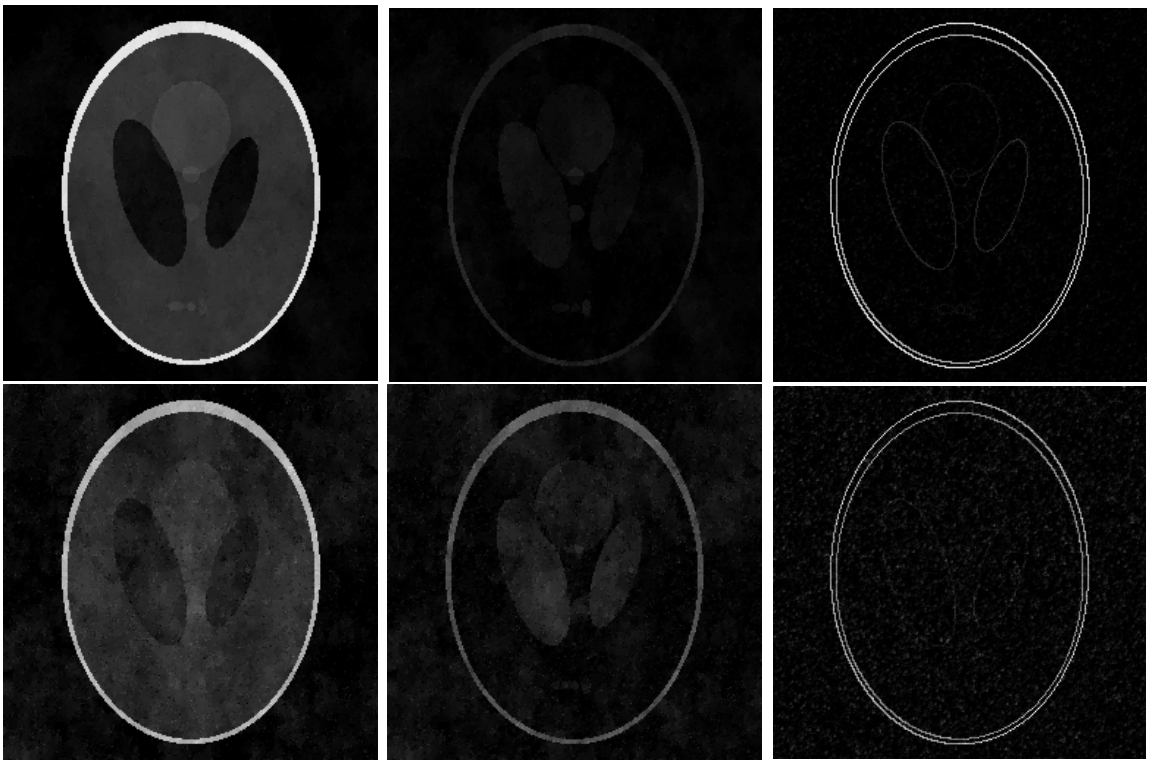

FIGURE 4. TV-reconstructed image with 5% (top left) and 10% (bottom left) and the respective differences (middle) from the original image. The plots on the right column are the magnitudes of the reconstructed image gradients.

APPENDIX A. PROOF OF THEOREM 2

The argument is patterned after [2] with adaptation to the setting of constrained joint sparsity.

Proposition 2. *We have*

$$|\Re\langle\varphi(\mathbf{Z}),\varphi(\mathbf{Z}')\rangle| \leq \delta_{s+s'}\|\mathbf{Z}\|_{2,2}\|\mathbf{Z}'\|_{2,2}$$

for all $\mathbf{Z},\mathbf{Z}'$ *supported on disjoint subsets* $T,T'\subset\{1,...,m\}$ *with* $|S|\leq s, |S'|\leq s'$.

Proof. Without loss of generality, suppose that $\|\mathbf{Z}\|_{2,2}=\|\mathbf{Z}'\|_{2,2}=1$. Since $\mathbf{Z}\perp\mathbf{Z}'$, $\|\mathbf{Z}\pm\mathbf{Z}'\|_{2,2}^2=2$. Hence we have from the RIP (49)

$$2(1-\delta_{s+s'}) \leq \|\varphi(\mathbf{Z}\pm\mathbf{Z}')\|_{2,2}^2 \leq 2(1+\delta_{s+s'}) \tag{61}$$

By the parallelogram identity and (61)

$$|\Re\langle\varphi(\mathbf{Z}),\varphi(\mathbf{Z}')\rangle| = \frac{1}{4}\left|\|\varphi(\mathbf{Z})+\varphi(\mathbf{Z}')\|_{2,2}^2 - \|\varphi(\mathbf{Z})-\varphi(\mathbf{Z}')\|_{2,2}^2\right| \leq \delta_{s+s'}$$

which proves the proposition. □

By the triangle inequality and the fact that $\mathbf{X}$ is in the feasible set we have

$$\|\varphi(\hat{\mathbf{X}} - \mathbf{X})\|_{2,2} \leq \|\varphi(\hat{\mathbf{X}}) - \mathbf{Y}\|_{2,2} + \|\mathbf{Y} - \varphi(\mathbf{X})\|_{2,2} \leq 2\varepsilon. \tag{62}$$

Set $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{D}$ and decompose $\mathbf{D}$ into a sum of $\mathbf{D}_{S_0}, \mathbf{D}_{S_1}, \mathbf{D}_{S_2}, ...$, each of row-sparsity at most s. Here S_0 corresponds to the locations of the s largest rows of $\mathbf{X}$; S_1 the locations of the s largest rows of $\mathbf{D}_{S_0^c}$; S_2 the locations of the next s largest rows of $\mathbf{D}_{S_0^c}$, and so on.

Step (i). Define the norm

$$\|\mathbf{Z}\|_{\infty,2} = \max_j \|\mathrm{row}_j(\mathbf{Z})\|_2.$$

For $j \geq 2$,

$$\|\mathbf{D}_{S_j}\|_{2,2} \leq s^{1/2}\|\mathbf{D}_{S_j}\|_{\infty,2} \leq s^{-1/2}\|\mathbf{D}_{S_{j-1}}\|_{2,2}$$

and hence

$$\sum_{j\geq 2} \|\mathbf{D}_{S_j}\|_{2,2} \leq s^{-1/2} \sum_{j\geq 1} \|\mathbf{D}_{S_j}\|_{1,2} \leq s^{-1/2}\|\mathbf{D}_{S_0^c}\|_{1,2}. \tag{63}$$

This yields by the Cauchy-Schwarz inequality

$$\|\mathbf{D}_{(S_0\cup S_1)^c}\|_{2,2} = \|\sum_{j\geq 2} \mathbf{D}_{S_j}\|_{2,2} \leq \sum_{j\geq 2} \|\mathbf{D}_{S_j}\|_{2,2} \leq s^{-1/2}\|\mathbf{D}_{S_0^c}\|_{1,2}. \tag{64}$$

Also we have

$$\begin{aligned} \|\mathbf{X}\|_{1,2} &\geq \|\hat{\mathbf{X}}\|_{1,2} \\ &= \|\mathbf{X}_{S_0} + \mathbf{D}_{S_0}\|_{1,2} + \|\mathbf{X}_{S_0^c} + \mathbf{D}_{S_0^c}\|_{1,2} \\ &\geq \|\mathbf{X}_{S_0}\|_{1,2} - \|\mathbf{D}_{S_0}\|_{1,2} - \|\mathbf{X}_{S_0^c}\|_{1,2} + \|\mathbf{D}_{S_0^c}\|_{1,2} \end{aligned}$$

which implies

$$\|\mathbf{D}_{S_0^c}\|_{1,2} \leq 2\|\mathbf{X}_{S_0^c}\|_{1,2} + \|\mathbf{D}_{S_0}\|_{1,2}. \tag{65}$$

Note that $\|\mathbf{X}_{S_0^c}\|_{1,2} = \|\mathbf{X} - \mathbf{X}^{(s)}\|_{1,2}$ by definition. Applying (64), (65) and the Cauchy-Schwartz inequality to $\|\mathbf{D}_{S_0}\|_{1,2}$ gives

$$\|\mathbf{D}_{(S_0\cup S_1)^c}\|_{2,2} \leq \|\mathbf{D}_{S_0}\|_{2,2} + 2e_0 \tag{66}$$

where $e_0 \equiv s^{-1/2}\|\mathbf{X} - \mathbf{X}^{(s)}\|_{1,2}$.

Step (ii). Define the inner product

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} A_{ij}^* B_{ij}$$

Observe

$$
\begin{aligned}
(67) \qquad & \|\varphi(\mathbf{D}_{S_0\cup S_1})\|_{2,2}^2 \\
&= \langle \varphi(\mathbf{D}_{S_0\cup S_1}), \varphi(\mathbf{D})\rangle - \left\langle \varphi(\mathbf{D}_{S_0\cup S_1}), \sum_{j\geq 2} \varphi(\mathbf{D}_{S_j}) \right\rangle \\
&= \Re\langle \varphi(\mathbf{D}_{S_0\cup S_1}), \varphi(\mathbf{D})\rangle - \sum_{j\geq 2} \Re\left\langle \varphi(\mathbf{D}_{S_0\cup S_1}), \varphi(\mathbf{D}_{S_j})\right\rangle \\
&= \Re\langle \varphi(\mathbf{D}_{S_0\cup S_1}), \varphi(\mathbf{D})\rangle - \sum_{j\geq 2} \left[\Re\left\langle \varphi(\mathbf{D}_{S_0}), \varphi(\mathbf{D}_{S_j})\right\rangle + \Re\left\langle \varphi(\mathbf{D}_{S_1}), \varphi(\mathbf{D}_{S_j})\right\rangle\right].
\end{aligned}
$$

From (62) and the RIP (49) it follows that

$$
|\langle \varphi(\mathbf{D}_{S_0\cup S_1}), \varphi(\mathbf{D})\rangle| \leq \|\varphi(\mathbf{D}_{S_0\cup S_1})\|_{2,2}\|\varphi(\mathbf{D})\|_{2,2} \leq 2\varepsilon\sqrt{1+\delta_{2s}}\|\mathbf{D}_{S_0\cup S_1}\|_{2,2}.
$$

Moreover, it follows from Proposition 2 that

$$
\begin{aligned}
(68) \qquad & \left|\Re\left\langle \varphi(\mathbf{D}_{S_0}), \varphi(\mathbf{D}_{S_j})\right\rangle\right| \leq \delta_{2s}\|\mathbf{D}_{S_0}\|_{2,2}\|\mathbf{D}_{S_j}\|_{2,2} \\
(69) \qquad & \left|\Re\left\langle \varphi(\mathbf{D}_{S_1}), \varphi(\mathbf{D}_{S_j})\right\rangle\right| \leq \delta_{2s}\|\mathbf{D}_{S_0}\|_{2,2}\|\mathbf{D}_{S_j}\|_{2,2}
\end{aligned}
$$

for $j \geq 2$. Since S_0 and S_1 are disjoint

$$
\|\mathbf{D}_{S_0}\|_{2,2} + \|\mathbf{D}_{S_1}\|_{2,2} \leq \sqrt{2}\sqrt{\|\mathbf{D}_{S_0}\|_{2,2}^2 + \|\mathbf{D}_{S_1}\|_{2,2}^2} = \sqrt{2}\|\mathbf{D}_{S_0\cup S_1}\|_{2,2}.
$$

Also by (67)-(69) and RIP

$$
(1-\delta_{2s})\|\mathbf{D}_{S_0\cup S_1}\|_{2,2}^2 \leq \|\varphi(\mathbf{D}_{S_0\cup S_1})\|_{2,2}^2 \leq \|\mathbf{D}_{S_0\cup S_1}\|_{2,2}\left(2\varepsilon\sqrt{1+\delta_{2s}} + \delta_{2s}\sum_{j\geq 2}\|\mathbf{D}_{S_j}\|_{2,2}\right).
$$

Therefore from (63) we obtain

$$
\|\mathbf{D}_{S_0\cup S_1}\|_{2,2} \leq \alpha\varepsilon + \rho s^{-1/2}\|\mathbf{D}_{S_0^c}\|_{1,2}, \quad \alpha = \frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}, \quad \rho = \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}
$$

and moreover by (65) and the definition of e_0

$$
\|\mathbf{D}_{S_0\cup S_1}\|_{2,2} \leq \alpha\varepsilon + \rho\|\mathbf{D}_{S_0}\|_{2,2} + 2\rho e_0
$$

after applying the Cauchy-Schwartz inequality to bound $\|\mathbf{D}_{S_0}\|_{1,2}$ by $s^{1/2}\|\mathbf{D}_{S_0}\|_{2,2}$. Thus

$$
\|\mathbf{D}_{S_0\cup S_1}\|_{2,2} \leq (1-\rho)^{-1}(\alpha\varepsilon + 2\rho e_0)
$$

if (49) holds.

Finally,

$$
\begin{aligned}
\|\mathbf{D}\|_{2,2} &\leq \|\mathbf{D}_{S_0\cup S_1}\|_{2,2} + \|\mathbf{D}_{(S_0\cup S_1)^c}\|_{2,2} \\
&\leq 2\|\mathbf{D}_{S_0\cup S_1}\|_{2,2} + 2e_0 \\
&\leq 2(1-\rho)^{-1}(\alpha\varepsilon + (1+\rho)e_0)
\end{aligned}
$$

which is the desired result.

Appendix B. Proof of Theorem 4

We prove the theorem by induction.

Suppose $\text{supp}(\mathbf{X}) = \mathcal{S} = \{J_1, \ldots, J_s\}$ and $X_{\max} = \|X_{J_1}\|_1 \geq \|X_{J_2}\|_1 \geq \cdots \geq \|X_{J_s}\|_1 = X_{\min}$.

In the first step,

$$\begin{aligned} \sum_{j=1}^{d} |Y_j^* \phi_{J_1 j}| &= \sum_{l=1}^{d} |x^*_{J_1 j} + x^*_{J_2 j}\phi^*_{J_2 j}\phi_{J_1 j} + \ldots + x^*_{J_s j}\phi^*_{J_s j}\phi_{J_1 j} + E_j^*\phi_{J_1 j}| \\ &\geq X_{\max} - X_{\max}(s-1)\mu_{\max} - \sum_j \|E_j\|_2. \end{aligned} \tag{70}$$

On the other hand, $\forall l \notin \text{supp}(\mathbf{X})$,

$$\begin{aligned} \sum_{j=1}^{d} |Y_j^* \phi_{l j}| &= \sum_{j=1}^{d} |x^*_{J_1 j}\phi^*_{J_1 j}\phi_{l j} + x^*_{J_2 j}\phi^*_{J_2 j}\phi_{l j} + \ldots + x^*_{J_s j}\phi^*_{J_s j}\phi_{l j} + E_j^*\phi_{l j}| \\ &\leq X_{\max} s\mu_{\max} + \sum_j \|E_j\|_2. \end{aligned} \tag{71}$$

Hence, if

$$(2s-1)\mu_{\max} + \frac{2\sum_j \|E_j\|_2}{X_{\max}} < 1,$$

then the right hand side of (70) is greater than the right hand side of (71) which implies that the first index selected by OMP must belong to $\text{supp}(\mathbf{X})$.

To continue the induction process, we state the straightforward generalization of a standard uniqueness result for sparse recovery to the joint sparsity setting (Lemma 5.3, [16]).

Proposition 3. *Let $\mathbf{Z} = \varphi(\mathbf{X})$ and $\mathbf{Y} = \mathbf{Z} + \mathbf{E}$. Let $\mathcal{S}^k$ be a set of k indices and let $\mathbf{A} \in \mathbb{C}^{m\times p}$ with supp$(\mathbf{A}) = \mathcal{S}^k$. Define*

$$\mathbf{Y}' = \mathbf{Y} - \varphi(\mathbf{A})$$

and

$$\mathbf{Z}' = \mathbf{Z} - \varphi(\mathbf{A}).$$

Clearly, $\mathbf{Y}' = \mathbf{Z}' + \mathbf{E}$. If $\mathcal{S}^k \subsetneq$ supp$(\mathbf{X})$ and the sparsity s of $\mathbf{X}$ satisfies $2s < 1 + \mu_{\max}^{-1}$, then $\mathbf{Z}'$ has a unique sparsest representation $\mathbf{Z}' = \varphi(\mathbf{X}')$ with the sparsity of $\mathbf{X}'$ at most s.

Let us suppose that the set $\mathcal{S}^k \subseteq \text{supp}(\mathbf{X})$ of k distinct indices have been selected and that $\mathbf{A}$ in Proposition 3 solves the following least squares problem

$$\mathbf{A} = \arg\min_{\mathbf{B}} \sum_j \|Y_j - \mathbf{\Phi}_j B_j\|_2, \quad \text{s.t.} \quad \text{supp}(\mathbf{B}) \subseteq \mathcal{S}^k \tag{72}$$

without imposing the constraint $\mathcal{L}$. This is equivalent to the concatenation $\mathbf{A} = [A_j]$ of p separate least squares solutions

$$A_j = \arg\min_{B_j} \|Y_j - \mathbf{\Phi}_j B_j\|_2, \quad \text{s.t.} \quad \text{supp}(B_j) \subseteq \mathcal{S}^k \tag{73}$$

Let $\mathbf{\Phi}_{j,\mathcal{S}^k}$ be the column submatrix of $\mathbf{\Phi}_j$ indexed by the set $\mathcal{S}^k$. By (73), $Y_j'^*\mathbf{\Phi}_{j,\mathcal{S}^k} = 0, \forall j$, which implies that no element of $\mathcal{S}^k$ gets selected at the $(k+1)$-st step. In order to ensure that some element in $\text{supp}(\mathbf{X}) \setminus \mathcal{S}^k$ gets selected at the $(k+1)$-st step we only need to repeat the calculation (70)-(71) to obtain the condition

$$(2s-1)\mu_{\max} + \frac{2\sum_j \|E_j\|_2}{\|X_{J_{k+1}}\|_1} < 1$$

which follows from the assumption (55) or equivalently

$$(2s-1)\mu_{\max} + \frac{2\sum_j \|E_j\|_2}{X_{\min}} < 1. \tag{74}$$

By the s-th step, all elements of the support set are selected and by the nature of the least squares solution the 2-norm of the residual is at most ε. Thus the stopping criterion is met and the iteration stops after s steps.

On the other hand, it follows from the calculation

$$\begin{aligned}
\sum_j \|Y_j'\|_2 &\geq \sum_j \left|\phi^*_{J_{k+1}j} Y_j'\right| \\
&= \sum_j \left|x_{J_{k+1}j} + \sum_{i=k+2}^{s} x_{J_i i}\phi^*_{J_{k+1}j}\phi_{J_i i} + \phi^*_{J_{k+1}j}E_j\right| \\
&\geq \|X_{J_{k+1}}\|_1 - \mu_{\max}(s-k-1)\|X_{J_{k+2}}\|_1 - \sum_j \|E_j\|_2 \\
&\geq (1-\mu_{\max}(s-k-1))\|X_{J_{k+1}}\|_1 - \sum_j \|E_j\|_2
\end{aligned}$$

and (74) or, equivalently,

$$X_{\min}(1-\mu_{\max}(2s-1)) > 2\varepsilon \tag{75}$$

that $\sum_j \|Y_j'\|_2 > \varepsilon$ for $k = 0, 1, \cdots, s-1$. Thus the iteration does not stop until $k = s$.

Since $\hat{\mathbf{X}}$ be the solution of the least squares problem (56), we have

$$\sum_j \|Y_j - \mathbf{\Phi}_j \hat{X}_j\|_2 \leq \sum_j \|Y_j - \mathbf{\Phi}_j X_j\|_2 \leq \varepsilon$$

and

$$\sum_j \|\mathbf{\Phi}_j(X_j - \hat{X}_j)\|_2 \leq \sum_j \|Y_j - \mathbf{\Phi}_j X_j\|_2 + \|Y_j - \mathbf{\Phi}_j \hat{X}_j\|_2 \leq 2\varepsilon$$

which implies

$$\sum_j \|\hat{X}_j - X_j\|_2 \leq 2\varepsilon/\lambda_{\min}$$

where

$$\lambda_{\min} = \min_j\{\text{the } s\text{-th singular value of the column submatrix of } \mathbf{\Phi}_j \text{ indexed by } \mathcal{S}\}$$

.

The desired error bound (57) can now be obtained from the following result (Lemma 2.2, [16]).

Proposition 4. *Suppose $s < 1 + \mu(\mathbf{\Phi}_j)^{-1}$. Every $m \times s$ column submatrix of $\mathbf{\Phi}_j$ has the s-th singular value bounded below by $\sqrt{1 - \mu(\mathbf{\Phi}_j)(s-1)}$.*

Proposition 4 implies that $\lambda_{\min} \geq \sqrt{1 - \mu_{\max}(s-1)}$ and thus the desired result

$$\sum_j \|\hat{X}_j - X_j\|_2 \leq \frac{2\varepsilon}{\sqrt{1 - \mu_{\max}(s-1)}}.$$

Acknowledgement. I thank Stan Osher and Justin Romberg for suggestion of publishing this note at the IPAM workshop "Challenges in Synthetic Aperture Radar" February 6-10, 2012. I thank the anonymous referees and Deanna Needell for pointing out the reference [22] which helps me appreciate more deeply the strength and weakness of my approach. I am grateful to Wenjing Liao for preparing Fig. 2-4. The research is partially supported by the NSF grant DMS - 0908535.

References

[1] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring Problems", *IEEE Trans. Image Proc.* **18** (11), 2419-2434, 2009.

[2] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I.* **346** (2008) 589-592.

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

[4] E.J. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Stat.* **37** (2009), 2145-2177.

[5] E. J. Candès and T. Tao, " Decoding by linear programming," *IEEE Trans. Inform. Theory* **51** (2005), 4203 – 4215.

[6] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vision* **20** (2004), 89-97.

[7] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems, " *Numer. Math.* **76** (1997), 167-188.

[8] T. F. Chan, G. H. Golub, and P. Mulet, A nonlinear primal-dual method for total variation-based image restoration., *SIAM J. Sci. Comput.* **20** (6), pp. 19641977, 1999,

[9] T. Chan and J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet and Stochastic Methods*, Society for Industrial and Applied Mathematics, 2005.

[10] S.S. Chen, D.L. Donoho and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.* **43** (2001), 129-159.

[11] W.-S. Cheung, "Discrete Poincaré-type inequalities", *Tamkang J. Math.* **29 (2)** (1998), 145-153.

[12] Claerbout, J. F., and Muir, F., "Robust modeling with erratic data," *Geophysics* **38** (1973), no. 5, 826-844.

[13] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory.* 2nd edition, Springer, 1998.

[14] G.M. Davis, S. Mallat and M. Avellaneda, "Adaptive greedy approximations", *J. Constructive Approx.* **13** (1973), 57-98.
[15] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory* **52** (2006) 1289 – 1306.
[16] D.L. Donoho, M. Elad and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory* **52** (2006) 6-18.
[17] D.L. Donoho and X. Huo, "Uncertainty principle and ideal atomic decomposition, " *IEEE Trans. Inform. Theory* **47** (2001), 2845-2862.
[18] A. Fannjiang, " Compressive inverse scattering II. Multi-shot SISO measurements with Born scatterers," *Inverse Problems* **26** (2010), 035009
[19] A. Fannjiang, T. Strohmer and P. Yan, "Compressed Remote Sensing of Sparse Object," *SIAM J. Imag. Sci.* **3** (2010) 596-618.
[20] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd edition. The Johns Hopkins University Press, 1996.
[21] F. Natterer, *The Mathematics of Computerized Tomography,* John Wiley & Sons, 1986.
[22] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," `arXiv`:1202.6429v6, May 31, 2012.
[23] V. Patel, R Maleh, A. Gilbert, and R. Chellappa, " Gradient-based image recovery methods from incomplete Fourier measurements," *IEEE Trans. Image Process.* **21** (2012), 94-104.
[24] Y.C. Pati, R. Rezaiifar and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Proceedings of the 27th Asilomar Conference in Signals, Systems and Computers*, 1993.
[25] H. Rauhut, "Stability results for random sampling of sparse trigonometric polynomials," *IEEE Trans. Inform. Th.* **54** (2008), 5661-5670.
[26] J. Romberg, "Imaging via compressive sampling," *IEEE Sign. Proc. Mag.* **20**, 14-20, 2008.
[27] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," *Proc. IEEE ICIP* **1** (1994), pp. 3135.
[28] L.I. Rudin, S. Osher and E. Fatemi, " Nonlinear total variation based noise removal algorithms," *Physica D* **60** (1992) 259-268.
[29] Taylor, H. L., S. C. Banks, J. F. McCoy, "Deconvolution with the ℓ-1 norm," *Geophysics* **44** (1979), no. 1, 39-52.
[30] J.A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit," *IEEE Trans. Inf. Theory* **53** (2007) 4655-4666.
[31] P. Weiss, L. Blanc-Féraud and G. Aubert, "Efficient schemes for total variation minimization under constraints in image processing," *SIAM J. Sci. Comput* **31** (3), pp. 20472080, 2009.

E-mail address: `fannjiang@math.ucdavis.edu`

Department of Mathematics, University of California, Davis, CA 95616-8633