

NISSUNA UMANA INVESTIGAZIONE SI PUO DIMANDARE VERA SCIENZA
S'ESSA NON PASSA PER LE MATEMATICHE DIMOSTRAZIONI
LEONARDO DA VINCI

vol. 1

no. 1

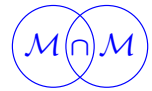
2013

MATHEMATICS AND MECHANICS
of
Complex Systems

ALBERT FANNJIANG

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





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

ALBERT FANNJIANG

This paper proposes a general framework for compressed sensing of constrained joint sparsity (CJS) which includes total variation minimization (TV-min) as an example. The gradient- and 2-norm error bounds, independent of the ambient dimension, are derived for the CJS version of basis pursuit and orthogonal matching pursuit. As an application the results extend Candès, Romberg, and Tao’s proof of exact recovery of piecewise constant objects with noiseless incomplete Fourier data to the case of noisy data.

1. Introduction

One of the most significant developments of the last decade in imaging and signal processing is compressive sensing (CS) which promises reconstruction with fewer data than the ambient dimension. CS capability [Candès and Tao 2005; Donoho 2006] hinges on favorable sensing matrices and enforcing a key piece of prior knowledge, that is, sparse objects.

Consider the linear inverse problem $Y = \Phi X + E$ 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 $\|X\|_0 \equiv \#$ of nonzero elements in X , can be effectively enforced by ℓ_1 -minimization (ℓ_1 -min) [Chen et al. 2001; Donoho and Huo 2001]:

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

with favorable sensing matrices Φ .

The ℓ_1 -min idea dates back to geophysics research in the 1970s [Claerbout and Muir 1973; Taylor et al. 1979]. The ℓ_1 -minimizer is often a much better approximation of the sparse object than the traditional minimum energy solution

Communicated by Micol Amar.

This research is partially supported by the U.S. National Science Foundation under grant DMS-0908535.

MSC2010: 15A29.

Keywords: total variation, joint sparsity, multiple measurement vectors, compressive sensing.

via ℓ_2 -minimization because the 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 a number of data points n much smaller than m if the sensing matrix Φ satisfies some favorable conditions such as the restricted isometry property (RIP) [Candès and Tao 2005]: Φ is said to satisfy the RIP of order k if

$$(1 - \delta_k)\|Z\|_{2,2}^2 \leq \|\Phi Z\|_2^2 \leq (1 + \delta_k)\|Z\|_2^2 \quad (2)$$

for any k -sparse vector Z where the minimum of such a constant δ_k is the restricted isometry constant (RIC) of order k .

The drawback of the RIP is that only a few special types of matrices are known to satisfy it, including independently and identically distributed (i.i.d.) random matrices and random partial Fourier matrices formed by random row selections of the discrete Fourier transform.

A more practical alternative CS criterion is furnished by the incoherence property as measured by one minus the mutual coherence [Donoho and Elad 2003; Tropp 2004]:

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

A parallel development in image denoising pioneered by Osher and coworkers [Rudin et al. 1992; Rudin and Osher 1994] seeks to enforce edge detection by total variation minimization (TV-min):

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

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, the TV-min approach to deblurring can be formulated as follows. Let $f \in \mathbb{C}^{p \times q}$ be a noisy complex-valued data set of $p \times q$ pixels. Let T be the transformation from the true object to the ideal sensors, modeling the imaging process. Replacing the total variation in (4) by the discrete total variation

$$\|g\|_{\text{TV}} \equiv \sum_{i,j} \sqrt{|\Delta_1 g(i, j)|^2 + |\Delta_2 g(i, j)|^2},$$

$$\Delta g = (\Delta_1 g, \Delta_2 g)(i, j) \equiv (g(i+1, j) - g(i, j), g(i, j+1) - g(i, j)),$$

we obtain

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

(see [Chambolle and Lions 1997; Chan and Shen 2005]).

In a breakthrough paper, Candès et al. [2006] 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 piecewise constant objects from (5).

A main application of this present work is to extend the result of Candès et al. 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, for example, $L^2(\mathbb{R}^d)$. To fit into CS's discrete framework, 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, the 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} , that is, $m = \mathcal{O}(\ell^{-d})$. As we will assume that the original object is well approximated by the discrete model in the limit $\ell \rightarrow 0$, the sparsity s of the edges of a piecewise constant object is proportional to ℓ^{1-d} , that is, the object is nonfractal. 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.

First we introduce the notation for multivectors $\mathbf{Y} \in \mathbb{C}^{n \times d}$:

$$\|\mathbf{Y}\|_{b,a} = \left(\sum_{j=1}^n \|\text{row}_j(\mathbf{Y})\|_a^b \right)^{1/b}, \quad a, b \geq 1, \quad (6)$$

where $\text{row}_j(\mathbf{Y})$ is the j -th row of \mathbf{Y} . The 2, 2-norm is exactly the Frobenius norm. To avoid confusion with the *subordinate* matrix norm [Golub and Van Loan 1996], it is more convenient to view \mathbf{Y} as multivectors rather than a matrix.

We aim at the following error bounds. Let V be the discretized object and \hat{V} an estimate of V . We will propose a compressive sampling scheme that leads to the error bound for the TV-minimizer \hat{V} :

$$\|\Delta V - \Delta \hat{V}\|_{2,2} = \mathcal{O}(\varepsilon), \quad \ell \rightarrow 0, \quad (7)$$

implying via the discrete Poincaré inequality that

$$\|V - \hat{V}\|_2 = \mathcal{O}(\varepsilon/\ell) \quad (8)$$

independent of the ambient dimension d .

If \hat{V} is the reconstruction by using a version of the greedy algorithm, orthogonal matching pursuit (OMP) [Pati et al. 1993; Davis et al. 1997], for multivectors then in addition to (7) we also have

$$\|V - \hat{V}\|_2 = \mathcal{O}(\varepsilon/\sqrt{\ell}) \quad (9)$$

independent of the ambient dimension d (see Section 3). We do not know if the bound (9) applies to the TV-minimizer.

A key advantage of the greedy algorithm used to prove (9) is the exact recovery of the gradient support (that is, the edge location) under proper conditions (Theorem 2 in Section 3). On the one hand, TV-min requires fewer data points for recovery: $\mathcal{O}(s)$ for TV-min under the RIP versus $\mathcal{O}(s^2)$ for the greedy algorithm under incoherence where the sparsity $s = \mathcal{O}(\ell^{1-d})$ is as already mentioned. On the other hand, the greedy algorithm is computationally more efficient and incoherent measurements are much easier to design and verify than the RIP.

At heart our theory is based on reformulation of TV-min as CS of joint sparsity with linear constraints (such as the curl-free constraint in the case of TV-min): basis pursuit denoising (BPDN) for constrained joint sparsity (CJS) is formulated as

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

where

$$\varphi(\mathbf{Z}) = [\Phi_1 Z_1, \dots, \Phi_d Z_d], \quad Z_j = \text{the } j\text{-th column of } \mathbf{Z},$$

and \mathcal{L} represents a linear constraint. Without loss of generality, we assume the matrices $\{\Phi_j\} \subset \mathbb{C}^{n \times m}$ all have *unit 2-norm* columns.

In connection to TV-min, Z_j is the j -th directional gradient of the discrete 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. As shown below, for inverse scattering we have $\Phi_j = \Phi$ for all j , and \mathcal{L} is the curl-free constraint which takes the form

$$\Delta_1 Z_2 = \Delta_2 Z_1$$

for $d = 2$ (see (53)). Our main results, Theorems 1 and 2, constitute performance guarantees for CJS based, respectively, on the RIP and incoherence of the measurement matrices Φ_j .

1.1. Comparison of existing theories. The gradient-based method of [Patel et al. 2012] modifies the original Fourier measurements to obtain Fourier measurements of the corresponding vertical and horizontal edge images which then are separately reconstructed by the standard CS algorithms. This approach attempts to take advantage of typically lower *separate* sparsity and is different from TV-min. Nevertheless, a similar 2-norm error bound [Patel et al. 2012, Proposition V.2] to (8) is obtained.

Needell and Ward [2012] obtain interesting results for *anisotropic* total variation (ATV) minimization in terms of the objective function

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

While for *real*-valued objects in two dimensions, the isotropic TV seminorm is

equivalent to the anisotropic version, the two seminorms are, however, not the same in dimension greater than 3 and/or for complex-valued objects. A rather remarkable result of Needell and Ward is the bound $\|V - \hat{V}\|_2 = \mathcal{O}(\varepsilon)$, modulo a logarithmic factor, for $d = 2$. This is achieved by proving a strong Sobolev inequality for two dimensions under the additional assumption of the RIP with respect to the bivariate Haar transform. Unfortunately, this latter assumption prevents the results in [Needell and Ward 2012] 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 consider these existing approaches from the perspective of the CJS framework for arbitrary d . The approach of [Patel et al. 2012] can be reformulated as solving d standard BPDNs

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

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 becomes an important part of the approach in [Patel et al. 2012].

From the CJS perspective, the ATV-min considered in [Needell and Ward 2012] can be reformulated as follows. Let $\tilde{\mathbf{Z}} \in \mathbb{C}^{dm}$ be the image gradient vector by stacking the d directional gradients and let $\tilde{\mathbf{Y}} \in \mathbb{C}^{dn}$ be the similarly concatenated data vector. Likewise let $\tilde{\Phi} = \text{diag}(\Phi_1, \dots, \Phi_d) \in \mathbb{C}^{dn \times dm}$ be the block-diagonal matrix with blocks $\Phi_j \in \mathbb{C}^{n \times m}$. Then ATV-min is equivalent to BPDN for a *single* constrained and concatenated vector:

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

where $\tilde{\mathcal{L}}$ is the same constraint \mathcal{L} reformulated for concatenated vectors. Repeating verbatim the proofs of Theorems 1 and 2 we obtain the same error bounds as (7)–(9) for ATV-min as formulated in (11) under the same conditions for Φ_j separately.

In [Needell and Ward 2012], ATV-min is formulated in terms of the image, instead of the image gradient, to get rid of the curl-free constraint. To proceed the differently concatenated matrix $[\Phi_1, \dots, \Phi_d]$ is then assumed to satisfy the RIP of higher order demanding $2dn$ measurement data points. For $d = 2$, Needell and Ward assume the RIP of order $5s$ with $\delta_{5s} < \frac{1}{3}$ for $[\Phi_1, \Phi_2]$, which is much more stringent than the RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$ for Φ_1 and Φ_2 *separately* in (11). In particular, $\Phi_1 = \Phi_2$ is allowed for (11) but not for Needell and Ward. To get the aforementioned favorable $\mathcal{O}(\varepsilon)$ 2-norm error bound for $d = 2$, an additional measurement matrix satisfying the RIP with respect to the bivariate Haar basis is needed, which, as mentioned above, excludes partial Fourier measurements.

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

2. BPDN for CJS

Consider the linear inversion problem

$$\mathbf{Y} = \varphi(\mathbf{X}) + \mathbf{E}, \quad \mathcal{L}\mathbf{X} = 0, \quad (12)$$

where

$$\varphi(\mathbf{X}) = [\Phi_1 X_1, \Phi_2 X_2, \dots, \Phi_d X_d], \quad \Phi_j \in \mathbb{C}^{n \times m},$$

and the corresponding BPDN

$$\min \|\mathbf{Z}\|_{1,2}, \quad \text{s.t.} \quad \|\mathbf{Y} - \varphi(\mathbf{Z})\|_{2,2} \leq \varepsilon = \|\mathbf{E}\|_{2,2}, \quad \mathcal{L}\mathbf{Z} = 0. \quad (13)$$

For TV-min in d dimensions, $\Phi_j = \Phi$ for all j , the vector \mathbf{X} represents the discrete gradient of the unknown object V , and \mathcal{L} is the curl-free constraint. Without loss of generality, we assume the matrices $\{\Phi_j\}$ all have unit 2-norm columns.

We say that \mathbf{X} is s -row sparse if the number of nonzero rows in \mathbf{X} is at most s . With a slight abuse of terminology we call \mathbf{X} the object (of CJS).

In the following theorems, we let the object \mathbf{X} be general, not necessarily s -row sparse. Let $\mathbf{X}^{(s)}$ consist of the 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 1. *Suppose that the linear map φ satisfies the RIP of order $2s$*

$$(1 - \delta_{2s}) \|\mathbf{Z}\|_{2,2}^2 \leq \|\varphi(\mathbf{Z})\|_{2,2}^2 \leq (1 + \delta_{2s}) \|\mathbf{Z}\|_{2,2}^2 \quad (14)$$

for any $2s$ -row sparse \mathbf{Z} with $\delta_{2s} < \sqrt{2} - 1$. Let $\hat{\mathbf{X}}$ be the minimizer of (13). Then

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{2,2} \leq C_1 s^{-1/2} \|\mathbf{X} - \mathbf{X}^{(s)}\|_{1,2} + C_2 \varepsilon \quad (15)$$

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

Remark 1. Note that the RIP for joint sparsity (14) follows straightforwardly from the assumption of the separate RIP:

$$(1 - \delta_{2s}) \|\mathbf{Z}\|_2^2 \leq \|\Phi_j \mathbf{Z}\|_2^2 \leq (1 + \delta_{2s}) \|\mathbf{Z}\|_2^2 \quad \text{for all } j,$$

with a common RIC.

Remark 2. For the standard Lasso with a particular choice of regularization parameter, [Candès and Plan 2009, Theorem 1.3] guarantees exact support recovery under a favorable sparsity constraint. In our setting and notation, their TV-min principle suggests

$$\min_{\mathcal{L}_{\mathbf{Z}=0}} \lambda \sigma \|\mathbf{Z}\|_{1,2} + \frac{1}{2} \|\mathbf{Y} - \varphi(\mathbf{Z})\|_{2,2}^2, \quad \lambda = 2\sqrt{2 \log m}, \quad (16)$$

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 Candès and Plan can be extended to (16), it is inadequate for our purpose because they assume independently selected support and signs, which is clearly not satisfied by the gradient of a piecewise constant object.

The proof of Theorem 1 is given in Appendix A.

The error bound (15) implies (7) for s -row sparse \mathbf{X} . For the 2-norm bound (8), we apply the discrete Poincaré inequality [Cheung 1998]

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

to get

$$\|V - \hat{V}\|_2 \leq \frac{m^{1/d}}{2d^{1/2}} C_2 \varepsilon = \mathcal{O}\left(\frac{\varepsilon}{\ell}\right), \quad (17)$$

since $\ell \sim m^{-1/d}$.

3. Greedy pursuit for CJS

One idea to improve the error bound is through exact recovery of the support. This can be achieved by greedy algorithms. As before, we consider the general linear inversion with CJS (12) with $\|\mathbf{E}\|_{2,2} = \varepsilon$.

Algorithm 1 on the next page is an extension of the joint-sparsity greedy algorithms of [Cotter et al. 2005; Chen and Huo 2006; Tropp et al. 2006] to a setting with multiple sensing matrices.

Note that the linear constraint is not enforced in Algorithm 1.

A natural indicator of the performance of OMP is the mutual coherence (3) [Tropp 2004; Donoho et al. 2006]. Let

$$\mu_{\max} = \max_j \mu(\Phi_j).$$

Then, analogous to [Donoho et al. 2006, Theorem 5.1], we have the following performance guarantee.

Algorithm 1. OMP for joint sparsity

Input: $\{\Phi_j\}$, \mathbf{Y} , $\varepsilon > 0$.

Initialization: $\mathbf{X}^0 = \mathbf{0}$, $\mathbf{R}^0 = \mathbf{Y}$, and $\mathcal{S}^0 = \emptyset$.

Iteration:

1) $i_{\max} = \arg \max_i \sum_{j=1}^d |\Phi_{j,i}^* \mathbf{R}_j^{k-1}|$, where $\Phi_{j,i}^*$ is the conjugate transpose of the i -th column of Φ_j .

2) $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{i_{\max}\}$.

3) $\mathbf{X}^k = \arg \min \|\Phi \mathbf{Z} - \mathbf{Y}\|_{2,2}$ s.t. $\text{supp}(\mathbf{Z}) \subseteq \mathcal{S}^k$.

4) $\mathbf{R}^k = \mathbf{Y} - \varphi(\mathbf{X}^k)$.

5) Stop if $\sum_j \|\mathbf{R}_j^k\|_2 \leq \varepsilon$.

Output: \mathbf{X}^k .

Theorem 2. Suppose the sparsity s satisfies

$$s < \frac{1}{2} \left(1 + \frac{1}{\mu_{\max}} \right) - \frac{\sqrt{d}\varepsilon}{\mu_{\max} X_{\min}}, \quad X_{\min} = \min_k \|\text{row}_k(\mathbf{X})\|_1. \quad (18)$$

Let \mathbf{Z} be the output of [Algorithm 1](#), with the stopping rule that the residual drops to the level ε or below. Then $\text{supp}(\mathbf{Z}) = \text{supp}(\mathbf{X})$.

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

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{B}} \|\mathbf{Y} - \Phi \mathbf{B}\|_{2,2}, \quad \text{s.t.} \quad \text{supp}(\mathbf{B}) \subseteq \text{supp}(\mathbf{X}), \quad \mathcal{L} \mathbf{B} = \mathbf{0}. \quad (19)$$

Then

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{2,2} \leq \frac{2\varepsilon}{\sqrt{1 - \mu_{\max}(s-1)}}. \quad (20)$$

The proof of [Theorem 2](#) is given in [Appendix B](#).

The main advantage of [Theorem 2](#) over [Theorem 1](#) is the guarantee of exact recovery of the support of \mathbf{X} . Moreover, a better 2-norm error bound follows because now the gradient error is guaranteed to vanish outside a set of cardinality $\mathcal{O}(\ell^{1-d})$: Let $\mathbb{L} \subset \mathbb{Z}^d$ be a finite lattice of $\mathcal{O}(\ell^{-d})$ cardinality and $\{\mathbb{L}_l : l = 1, \dots, L\}$ a partition of \mathbb{L} , i.e., $\mathbb{L} = \bigcup_l \mathbb{L}_l$ and $\mathbb{L}_l \cap \mathbb{L}_k = \emptyset$ for $l \neq k$. Let the scaled sets $\ell \mathbb{L}_l$, $l = 1, \dots, L$, be the level sets of the object V such that

$$V = \sum_{l=1}^L v_l \mathbf{I}_{\ell \mathbb{L}_l},$$

where $\mathbf{I}_{\ell \mathbb{L}_l}$ is the indicator function of $\ell \mathbb{L}_l$. The reconstructed object \hat{V} from $\hat{\mathbf{X}}$ given in [\(19\)](#) also takes the same form:

$$\hat{V} = \sum_{l=1}^L \hat{v}_l \mathbf{I}_{\ell \mathbb{L}_l}.$$

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

$$\|\Delta(V - \hat{V})\|_{2,2} = \mathcal{O}(\varepsilon)$$

by (20) and the gradient error occurs only on the boundaries of $\ell\mathbb{L}_l$ of cardinality $\mathcal{O}(\ell^{1-d})$, we have

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

Namely

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

and thus

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

4. Application: inverse scattering

In this section, we discuss the main application of the CJS formulation, that is, the TV-min for the inverse scattering problem.

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

$$\nabla^2 u(\mathbf{r}) + \omega^2(1 + v(\mathbf{r}))u(\mathbf{r}) = 0, \quad (21)$$

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

$$u^i(\mathbf{r}) = e^{i\omega\mathbf{r}\cdot\hat{\mathbf{d}}}, \quad (22)$$

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

$$\nabla^2 u^s + \omega^2 u^s = -\omega^2 v u \quad (23)$$

which can be written as the Lippmann–Schwinger equation:

$$u^s(\mathbf{r}) = \omega^2 \int_{\mathbb{R}^d} v(\mathbf{r}') (u^i(\mathbf{r}') + u^s(\mathbf{r}')) G(\mathbf{r}, \mathbf{r}') d\mathbf{r}', \quad (24)$$

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

The scattered field necessarily satisfies Sommerfeld's radiation condition

$$\lim_{r \rightarrow \infty} r^{(d-1)/2} \left(\frac{\partial}{\partial r} - i\omega \right) u^s = 0,$$

reflecting the fact that the energy which is radiated from the sources represented by the right-hand side of (23) must scatter to infinity.

Thus the scattered field has the far-field asymptotic

$$u^s(\mathbf{r}) = \frac{e^{i\omega|\mathbf{r}|}}{|\mathbf{r}|^{(d-1)/2}} (A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) + \mathcal{O}(|\mathbf{r}|^{-1})), \quad \hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|, \quad (25)$$

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 [Colton and Kress 1998]

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

which under the Born approximation becomes

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) = \frac{\omega^2}{4\pi} \int d\mathbf{r}' v(\mathbf{r}') e^{i\omega\mathbf{r}' \cdot (\hat{\mathbf{d}} - \hat{\mathbf{r}})}. \quad (26)$$

For 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 :

$$\ell\mathbb{L} = \{\mathbf{r}_j = \ell(p_1, p_2) : j = (p_1 - 1)\sqrt{m} + p_2, \mathbf{p} = (p_1, p_2) \in \mathbb{L}\}. \quad (27)$$

In the context of inverse scattering, it is natural to treat the size of the discrete ambient domain $\ell\mathbb{L}$ as 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 , $j = 1, \dots, m$ be the strength of the scatterers. In other words, the total object is a sum of δ -functions:

$$v(\mathbf{r}) = \sum_j v_j \delta(\mathbf{r} - \mathbf{r}_j). \quad (28)$$

Let $\mathcal{S} = \{\mathbf{r}_{i_j} : j = 1, \dots, s\}$ be the locations of the scatterers. Hence $v_j = 0$ for all $\mathbf{r}_j \notin \mathcal{S}$.

For point objects the scattering amplitude becomes a finite sum:

$$A(\hat{\mathbf{r}}, \hat{\mathbf{d}}, \omega) = \frac{\omega^2}{4\pi} \sum_{j=1}^m v_j e^{i\omega\mathbf{r}_j \cdot (\hat{\mathbf{d}} - \hat{\mathbf{r}})}. \quad (29)$$

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

5. Sampling schemes

Next we review the sampling schemes introduced in [Fannjiang 2010] for *point* objects (28).

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

$$y_l = \frac{4\pi}{\omega^2 \sqrt{n}} A(\hat{\mathbf{r}}_l, \hat{\mathbf{d}}_l, \omega_l), \quad l = 1, \dots, n. \quad (30)$$

The measurement vector is related to the *point* object vector $X = (v_j) \in \mathbb{C}^m$ by the sensing matrix Φ as

$$Y = \Phi X + E, \quad (31)$$

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

$$n^{-1/2} e^{-i\omega_l \hat{\mathbf{r}}_l \cdot \mathbf{r}_j} e^{i\omega_l \hat{\mathbf{d}}_l \cdot \mathbf{r}_j} = n^{-1/2} e^{i\omega_l \ell (p_2(\sin \theta_l - \sin \tilde{\theta}_l) + p_1(\cos \theta_l - \cos \tilde{\theta}_l))}, \quad (32)$$

with $j = (p_1 - 1) + p_2$. Note that Φ has unit 2-norm columns.

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

$$(\xi_l, \zeta_l) = \rho_l (\cos \phi_l, \sin \phi_l), \quad \rho_l = \sqrt{\xi_l^2 + \zeta_l^2} \leq \sqrt{2}. \quad (33)$$

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

$$\theta_l + \tilde{\theta}_l = 2\phi_l + \pi, \quad (34)$$

and set the frequency ω_l to be

$$\omega_l = \frac{\Omega \rho_l}{\sqrt{2} \sin(\theta_l - \tilde{\theta}_l/2)}, \quad (35)$$

where Ω is a control parameter. Then the entries (32) of the sensing matrix Φ under the condition

$$\Omega \ell = \pi / \sqrt{2} \quad (36)$$

are those of the random partial Fourier matrix

$$e^{i\pi(p_1 \xi_l + p_2 \zeta_l)}, \quad l = 1, \dots, n, \quad p_1, p_2 = 1, \dots, \sqrt{m}. \quad (37)$$

We consider two particular sampling schemes: The first employs multiple frequencies with the sampling angle always in the back-scattering direction, resembling the imaging geometry of synthetic aperture radar; the second employs only a 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 Ω -band limited probes, that is, $\omega_l \in [-\Omega, \Omega]$. This and (35) lead to the constraint

$$\left| \sin \frac{\theta_l - \tilde{\theta}_l}{2} \right| \geq \frac{\rho_l}{\sqrt{2}}. \quad (38)$$

A simple way to satisfy (34) and (38) is to set

$$\phi_l = \tilde{\theta}_l = \theta_l - \pi, \quad (39)$$

$$\omega_l = \frac{\Omega \rho_l}{\sqrt{2}}, \quad (40)$$

where $l = 1, \dots, n$. In this case the scattering amplitude is sampled exactly in the backward direction, resembling synthetic-aperture radar imaging. In contrast, exact forward sampling with $\tilde{\theta}_l = \theta_l$ almost surely violates the constraint (38).

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

$$\omega_l = \gamma \Omega, \quad \gamma \geq 1, \quad l = 1, \dots, n. \quad (41)$$

We set

$$\theta_l = \phi_l + \arcsin \frac{\rho_l}{\gamma \sqrt{2}}, \quad (42)$$

$$\tilde{\theta}_l = \phi_l - \arcsin \frac{\rho_l}{\gamma \sqrt{2}}. \quad (43)$$

The difference between the incident angle and the sampling angle is

$$\theta_l - \tilde{\theta}_l = 2 \arcsin \frac{\rho_l}{\gamma \sqrt{2}}, \quad (44)$$

which diminishes as $\gamma \rightarrow \infty$. In other words, in the high-frequency limit, the sampling angle approaches the incident angle, resembling X-ray tomography [Natterer 1986].

6. Piecewise constant objects

Next let us consider the following class of piecewise constant objects:

$$v(\mathbf{r}) = \sum_{\mathbf{p} \in \mathbb{L}} v_{\mathbf{p}} \mathbf{I}_{\square} \left(\frac{\mathbf{r}}{\ell} - \mathbf{p} \right), \quad \square = \left[-\frac{1}{2}, \frac{1}{2} \right]^2, \quad (45)$$

where \mathbf{I}_{\square} is the indicator function of the unit square \square . As remarked in Section 1, we think of the pixelated v as a discrete approximation of some compactly supported function on \mathbb{R}^2 and having a well-defined limit as $\ell \rightarrow 0$. Set $V = (v_j) \in \mathbb{C}^m$ and $j = (p_1 - 1)\sqrt{m} + p_2$.

The discrete version of (26) is, however, not exactly the same as (29) since extended objects have different scattering properties from those of point objects.

The integral on the right-hand side of (26), modulo the discretization error, is

$$\int dr' v(r') e^{i\omega r' \cdot (\hat{d} - \hat{r})} = \sum_{p \in \mathbb{L}} v_p e^{i\omega \ell p \cdot (\hat{d} - \hat{r})} \int_{\square} e^{i\omega r' \cdot (\hat{d} - \hat{r})} \mathbf{I}_{\square} \left(\frac{r'}{\ell} \right) dr'.$$

Now letting \hat{d}_l , \hat{r}_l , and ω_l , $l = 1, \dots, n$, be selected according to Scheme I or II and substituting them into the above equation, we obtain

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

Let

$$x_j = \ell^2 v_p, \quad j = (p_1 - 1)\sqrt{m} + p_2,$$

and

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

where

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

and $E = (e_l)$ is the noise vector.

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

$$\phi_{kp} = \frac{1}{\sqrt{n}} e^{i\pi(p_1 \xi_k + p_2 \eta_k)}, \quad p = (p_1 - 1)\sqrt{m} + p_2, \quad p_1, p_2 = 1, \dots, \sqrt{m}. \quad (46)$$

Then the system above can be written in the same form as (31):

$$Y = \Phi X + E, \quad X = (x_j), \quad (47)$$

where the data and error vectors have been modified as above to account for the differences between extended and point objects.

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

$$\min \|Z\|_{\text{TV}}, \quad \text{s.t.} \quad \|Y - \Phi Z\|_2 \leq \|E\|_2. \quad (48)$$

We accomplish this by transforming (48) into BPDN for CJS (13).

Define $X = (X_1, X_2)$ with

$$(X_1, X_2) = \ell^2 (\Delta_1 V, \Delta_2 V) \in \mathbb{C}^{m \times 2}.$$

Suppose the support of $\{v_{\mathbf{p}+e_1}, v_{\mathbf{p}+e_2}\}$ is contained in \mathbb{L} . Simple calculation yields

$$y_l = \frac{\ell^2}{\sqrt{n}} e^{i\pi\xi_l} \sum_{\mathbf{p} \in \mathbb{L}} v_{\mathbf{p}+e_1} e^{i\pi(p_1\xi_l+p_2\eta_l)} = \frac{\ell^2}{\sqrt{n}} e^{i\pi\eta_l} \sum_{\mathbf{p} \in \mathbb{L}} v_{\mathbf{p}+e_2} e^{i\pi(p_1\xi_l+p_2\eta_l)}$$

and thus

$$(e^{-i\pi\xi_l} - 1)y_l = \frac{\ell^2}{\sqrt{n}} \sum_{\mathbf{p} \in \mathbb{L}} (v_{\mathbf{p}+e_1} - v_{\mathbf{p}}) e^{i\pi(p_1\xi_l+p_2\eta_l)}, \quad (49)$$

$$(e^{-i\pi\eta_l} - 1)y_l = \frac{\ell^2}{\sqrt{n}} \sum_{\mathbf{p} \in \mathbb{L}} (v_{\mathbf{p}+e_2} - v_{\mathbf{p}}) e^{i\pi(p_1\xi_l+p_2\eta_l)}. \quad (50)$$

Define $\mathbf{Y} = (Y_1, Y_2)$ with

$$Y_1 = ((e^{-i\pi\xi_l} - 1)y_l), \quad Y_2 = ((e^{-i\pi\eta_l} - 1)y_l) \in \mathbb{C}^n,$$

and $\mathbf{E} = (E_1, E_2)$ with

$$E_1 = ((e^{-i\pi\xi_l} - 1)e_l), \quad E_2 = ((e^{-i\pi\eta_l} - 1)e_l) \in \mathbb{C}^n. \quad (51)$$

We rewrite (47) in the form

$$\mathbf{Y} = \Phi \mathbf{X} + \mathbf{E}, \quad (52)$$

subject to the constraint

$$\Delta_1 X_2 = \Delta_2 X_1 \quad (53)$$

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 (that is, path-independent).

To see that (53) 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 and Δ_2 we have

$$v_{1,0} = v_{0,0} + (\Delta_1 V)_{0,0}, \quad v_{0,1} = v_{0,0} + (\Delta_2 V)_{0,0}.$$

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

$$v_{\mathbf{p}+e_1} = v_{\mathbf{p}} + (\Delta_1 V)_{\mathbf{p}}, \quad v_{\mathbf{p}+e_2} = v_{\mathbf{p}} + (\Delta_2 V)_{\mathbf{p}},$$

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

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

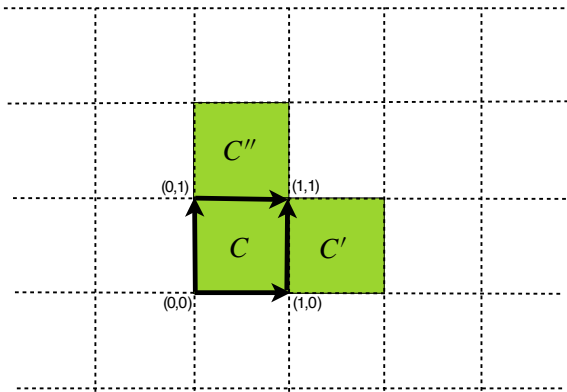


Figure 1. Consistency among cells C , C' , and C'' .

which is equivalent to (53).

Now (47) is equivalent to (52) with the constraint (53) provided that the value of V at (any) one grid point is known. The equivalence between the original TV-min (48) and the CJS formulation (13) with $\Phi_j = \Phi$ for all j , then hinges on the equivalence of their respective feasible sets which can be established under the assumption of Gaussian noise. When E in (47) is Gaussian noise, then so is \mathbf{E} , and vice versa, with variances precisely related to each other.

The random partial Fourier measurement matrix satisfies the RIP with $n = \mathcal{O}(s)$, up to a logarithmic factor [Candès et al. 2006], while its mutual coherence μ behaves like $\mathcal{O}(n^{-1/2})$ [Fannjiang et al. 2010]. Therefore (18) implies the sparsity constraint $s = \mathcal{O}(\sqrt{n})$ for the greedy approach which is more stringent than $s = \mathcal{O}(n)$ for the BPDN approach.

7. Conclusion

We have developed a general compressive sensing (CS) theory (Theorems 1 and 2) for constrained joint sparsity with multiple sensing matrices and obtained performance guarantees parallel to those for CS theory for a single measurement vector and matrix.

From the general theory we have derived 2-norm error bounds for the object and the gradient, independent of the ambient dimension, for total variation minimization (TV-min) and greedy estimates of piecewise constant objects.

In addition, the constrained joint sparsity (CJS) greedy algorithm can recover exactly the gradient support (that is, the edges of the object) leading to an improved 2-norm error bound. Although the CJS greedy algorithm needs a higher number of measurement data points than TV-min for Fourier measurements the incoherence property required is much easier to check, and is often the only practical way to

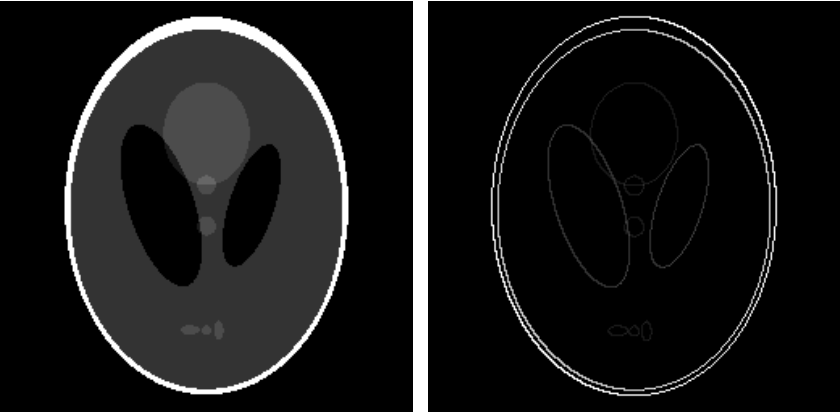


Figure 2. The original 256×256 Shepp–Logan phantom (left) and the Shepp–Logan phantom and the magnitudes of its gradient (right) with sparsity $s = 2,184$.

verify the restricted isometry property when the measurement matrix is not independently and identically distributed or Fourier.

We end by presenting a numerical example demonstrating the noise stability of TV-min. Efficient algorithms for TV-min denoising/deblurring exist [Beck and Teboulle 2009; Weiss et al. 2009]. We use the open source code ℓ_1 -MAGIC for our simulation.

Figure 2 shows the 256×256 image of the Shepp–Logan phantom and the modulus of its gradient. Clearly the sparsity ($s = 2,184$) of the gradient is much smaller than that of the original image. We take 10,000 Fourier measurement data points 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, shown in Figure 3. The relative error is 66.8% in

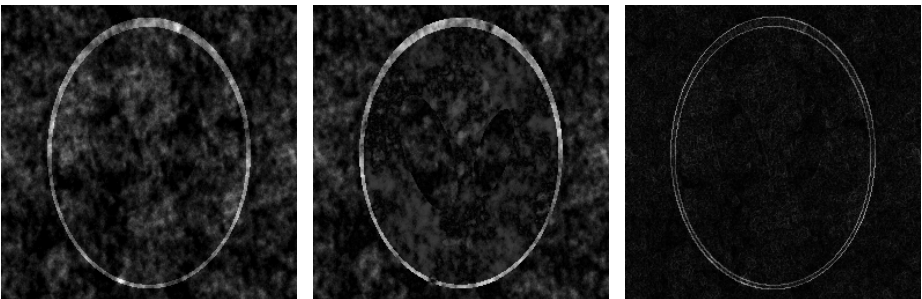


Figure 3. Noiseless ℓ_1 -min reconstructed image (left) and the differences (middle) from the original image. The plot on the right is the gradient of the reconstructed image.

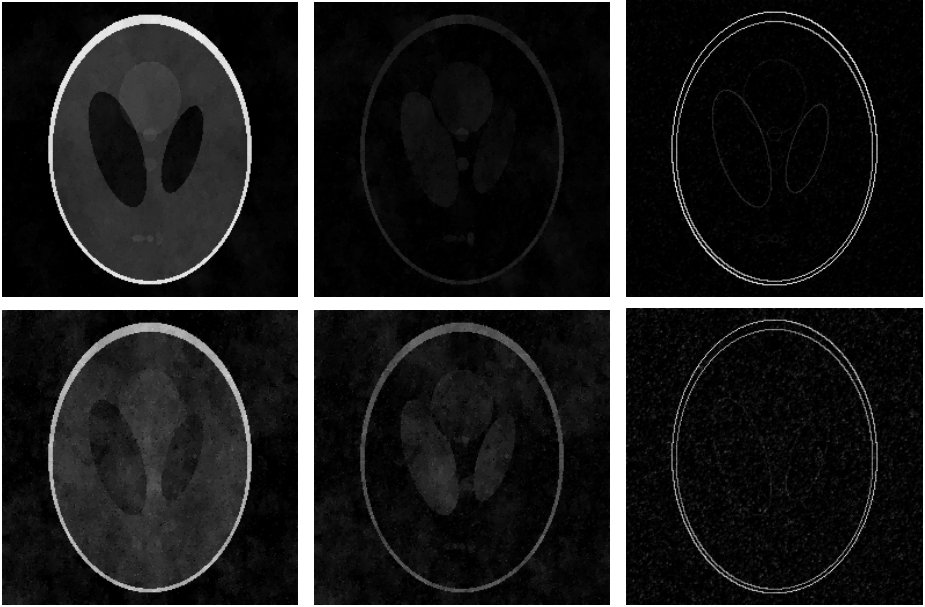


Figure 4. TV-reconstructed image with 5% (top left) and 10% (bottom left) and the respective differences (middle) from the original image. The plots on the right column are the magnitudes of the reconstructed image gradients.

the ℓ_2 norm and 72.8% in the TV norm. Only the outer boundary, which has the largest pixel values, 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.

Appendix A: Proof of Theorem 1

The argument is patterned after [Candès 2008] with adaptation to the CJS setting.

Proposition 1. *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 $T, T' \subset \{1, \dots, m\}$ with $|S| \leq s$ and $|S'| \leq s'$.

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 (14)

$$2(1 - \delta_{s+s'}) \leq \|\varphi(\mathbf{Z} \pm \mathbf{Z}')\|_{2,2}^2 \leq 2(1 + \delta_{s+s'}). \quad (54)$$

By the parallelogram identity and (54)

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

which proves the proposition. \square

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

$$\|\varphi(\hat{\mathbf{X}} - \mathbf{X})\|_{2,2} \leq \|\varphi(\hat{\mathbf{X}}) - \mathbf{Y}\|_{2,2} + \|\mathbf{Y} - \varphi(\mathbf{X})\|_{2,2} \leq 2\varepsilon. \quad (55)$$

Set $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{D}$ and decompose \mathbf{D} into a sum of $\mathbf{D}_{S_0}, \mathbf{D}_{S_1}, \mathbf{D}_{S_2}, \dots$, 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

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

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

$$\sum_{j \geq 2} \|\mathbf{D}_{S_j}\|_{2,2} \leq s^{-1/2} \sum_{j \geq 1} \|\mathbf{D}_{S_j}\|_{1,2} \leq s^{-1/2} \|\mathbf{D}_{S_0^c}\|_{1,2}. \quad (56)$$

This yields, by the Cauchy–Schwarz inequality,

$$\|\mathbf{D}_{(S_0 \cup S_1)^c}\|_{2,2} = \left\| \sum_{j \geq 2} \mathbf{D}_{S_j} \right\|_{2,2} \leq \sum_{j \geq 2} \|\mathbf{D}_{S_j}\|_{2,2} \leq s^{-1/2} \|\mathbf{D}_{S_0^c}\|_{1,2}. \quad (57)$$

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

$$\|\mathbf{D}_{S_0^c}\|_{1,2} \leq 2\|\mathbf{X}_{S_0^c}\|_{1,2} + \|\mathbf{D}_{S_0}\|_{1,2}. \quad (58)$$

Note that $\|\mathbf{X}_{S_0^c}\|_{1,2} = \|\mathbf{X} - \mathbf{X}^{(s)}\|_{1,2}$ by definition. Applying (57), (58), and the Cauchy–Schwarz inequality to $\|\mathbf{D}_{S_0}\|_{1,2}$ gives

$$\|\mathbf{D}_{(S_0 \cup S_1)^c}\|_{2,2} \leq \|\mathbf{D}_{S_0}\|_{2,2} + 2e_0, \quad (59)$$

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 that

$$\begin{aligned} & \|\varphi(\mathbf{D}_{S_0 \cup S_1})\|_{2,2}^2 \\ &= \langle \varphi(\mathbf{D}_{S_0 \cup S_1}), \varphi(\mathbf{D}) \rangle - \left\langle \varphi(\mathbf{D}_{S_0 \cup S_1}), \sum_{j \geq 2} \varphi(\mathbf{D}_{S_j}) \right\rangle \\ &= \Re \langle \varphi(\mathbf{D}_{S_0 \cup S_1}), \varphi(\mathbf{D}) \rangle - \sum_{j \geq 2} \Re \langle \varphi(\mathbf{D}_{S_0 \cup S_1}), \varphi(\mathbf{D}_{S_j}) \rangle \\ &= \Re \langle \varphi(\mathbf{D}_{S_0 \cup S_1}), \varphi(\mathbf{D}) \rangle - \sum_{j \geq 2} [\Re \langle \varphi(\mathbf{D}_{S_0}), \varphi(\mathbf{D}_{S_j}) \rangle + \Re \langle \varphi(\mathbf{D}_{S_1}), \varphi(\mathbf{D}_{S_j}) \rangle]. \end{aligned} \quad (60)$$

From (55) and the RIP (14) 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 1 that

$$|\Re \langle \varphi(\mathbf{D}_{S_0}), \varphi(\mathbf{D}_{S_j}) \rangle| \leq \delta_{2s} \|\mathbf{D}_{S_0}\|_{2,2} \|\mathbf{D}_{S_j}\|_{2,2}, \quad (61)$$

$$|\Re \langle \varphi(\mathbf{D}_{S_1}), \varphi(\mathbf{D}_{S_j}) \rangle| \leq \delta_{2s} \|\mathbf{D}_{S_1}\|_{2,2} \|\mathbf{D}_{S_j}\|_{2,2}, \quad (62)$$

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 (60)–(62) and the RIP,

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

Therefore from (56) 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 (58) 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 (14) 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 2

We prove the theorem by induction.

Let $\text{supp}(\mathbf{X}) = \mathcal{S} = \{J_1, \dots, J_s\}$ and

$$X_{\max} = \|\text{row}_{J_1}(\mathbf{X})\|_1 \geq \|\text{row}_{J_2}(\mathbf{X})\|_1 \geq \dots \geq \|\text{row}_{J_s}(\mathbf{X})\|_1 = X_{\min}.$$

In the first step,

$$\begin{aligned} \sum_{j=1}^d |\Phi_{j,J_1}^* Y_j| &= \sum_{j=1}^d |X_{J_1 j} + X_{J_2 j} \Phi_{j,J_1}^* \Phi_{j,J_2} + \dots + X_{J_s j} \Phi_{j,J_1}^* \Phi_{j,J_s} + \Phi_{j,J_1}^* E_j| \\ &\geq X_{\max} - X_{\max}(s-1)\mu_{\max} - \sum_j \|E_j\|_2. \end{aligned} \quad (\text{B.1})$$

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

$$\begin{aligned} \sum_{j=1}^d |\Phi_{j,l}^* Y_j| &= \sum_{j=1}^d |X_{J_1 j} \Phi_{j,l}^* \Phi_{j,J_1} + X_{J_2 j} \Phi_{j,l}^* \Phi_{j,J_2} + \dots + X_{J_s j} \Phi_{j,l}^* \Phi_{j,J_s} + \Phi_{j,l}^* E_j| \\ &\leq X_{\max} s \mu_{\max} + \sum_j \|E_j\|_2. \end{aligned} \quad (\text{B.2})$$

Hence, if

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

then the right-hand side of (B.1) is greater than the right-hand side of (B.2) 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 [Donoho et al. 2006, Lemma 5.3].

Proposition 2. *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 $\mathbf{A} \in \mathbb{C}^{n \times d}$ with $\text{supp}(\mathbf{A}) = \mathcal{S}^k$. Define*

$$\mathbf{Y}' = \mathbf{Y} - \varphi(\mathbf{A}) \quad (\text{B.3})$$

and

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

Clearly, $\mathbf{Y}' = \mathbf{Z}' + \mathbf{E}$. If $\mathcal{G}^k \subsetneq \text{supp}(\mathbf{X})$ 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 .

Proposition 2 says that the selection of a column, followed by the formation of the residual signal, leads to a situation like before, where the ideal noiseless signal has no more representing columns than before, and the noise level is the same.

Suppose that the set $\mathcal{G}^k \subseteq \text{supp}(\mathbf{X})$ of k distinct indices has been selected and that \mathbf{A} in **Proposition 2** solves the following least-squares problem:

$$\mathbf{A} = \arg \min \|\mathbf{Y} - \Phi \mathbf{B}\|_{2,2}, \quad \text{s.t.} \quad \text{supp}(\mathbf{B}) \subseteq \mathcal{G}^k, \quad (\text{B.4})$$

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

$$A_j = \arg \min_{B_j} \|Y_j - \Phi_j B_j\|_2, \quad \text{s.t.} \quad \text{supp}(B_j) \subseteq \mathcal{G}^k. \quad (\text{B.5})$$

Let Φ_{j, \mathcal{G}^k} be the column submatrix of Φ_j indexed by the set \mathcal{G}^k . By **(B.3)** and **(B.5)**, we have $\Phi_{j, \mathcal{G}^k}^* Y_j' = 0$ for all j , which implies that no element of \mathcal{G}^k gets selected at the $(k+1)$ -st step.

In order to ensure that some element in $\text{supp}(\mathbf{X}) \setminus \mathcal{G}^k$ gets selected at the $(k+1)$ -st step we only need to repeat the calculation **(B.1)**–**(B.2)** to obtain the condition

$$(2s-1)\mu_{\max} + \frac{2 \sum_j \|E_j\|_2}{\|X_{J_{k+1}}\|_1} < 1. \quad (\text{B.6})$$

Since $\sum_j \|E_j\|_2 \leq \sqrt{d} \|\mathbf{E}\|_{2,2} = \sqrt{d} \varepsilon$, **(B.6)** follows from

$$(2s-1)\mu_{\max} + \frac{2\sqrt{d}\varepsilon}{X_{\min}} < 1, \quad (\text{B.7})$$

which is the same as **(18)** and allows us to apply **Proposition 2** repeatedly.

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

$$\begin{aligned} \sum_j \|Y_j'\|_2 &\geq \sum_{j=1}^d |\Phi_{j, J_{k+1}}^* Y_j'| = \sum_j |X_{J_{k+1}j} + \sum_{i=k+2}^s X_{J_{k+1}i} \Phi_{j, J_{k+1}}^* \Phi_{i, J_i} + \Phi_{j, J_{k+1}}^* E_j| \\ &\geq \|\text{row}_{J_{k+1}}(\mathbf{X})\|_1 - \mu_{\max}(s-k-1) \|\text{row}_{J_{k+2}}(\mathbf{X})\|_1 - \sum_j \|E_j\|_2 \\ &\geq (1 - \mu_{\max}(s-k-1)) \|\text{row}_{J_{k+1}}(\mathbf{X})\|_1 - \sum_j \|E_j\|_2, \end{aligned}$$

and (B.7) (equivalently, $X_{\min}(1 - \mu_{\max}(2s - 1)) > 2\sqrt{d}\varepsilon$) that $\|\mathbf{Y}\|_{1,2} > \sqrt{d}\varepsilon$ for $k = 0, 1, \dots, s - 1$. Thus the iteration does not stop until $k = s$.

Since $\hat{\mathbf{X}}$ is the solution of the least-squares problem (19), we have

$$\|\mathbf{Y} - \Phi \hat{\mathbf{X}}\|_{2,2} \leq \|\mathbf{Y} - \Phi \mathbf{X}\|_{2,2} \leq \varepsilon$$

and

$$\|\Phi(\mathbf{X} - \hat{\mathbf{X}})\|_{2,2}^2 \leq 2\|\mathbf{Y} - \Phi \mathbf{X}\|_{2,2}^2 + 2\|\mathbf{Y} - \Phi \hat{\mathbf{X}}\|_{2,2}^2 \leq 2\varepsilon^2,$$

which implies

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{2,2} \leq \sqrt{2}\varepsilon/\lambda_{\min},$$

where

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

The desired error bound (20) can now be obtained from the following result [Donoho et al. 2006, Lemma 2.2].

Proposition 3. *Suppose $s < 1 + \mu(\Phi_j)^{-1}$. Every $m \times s$ column submatrix of Φ_j has the s -th singular value bounded below by $\sqrt{1 - \mu(\Phi_j)(s - 1)}$.*

By Proposition 3, $\lambda_{\min} \geq \sqrt{1 - \mu_{\max}(s - 1)}$ and thus

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{2,2} \leq \frac{\sqrt{2}\varepsilon}{\sqrt{1 - \mu_{\max}(s - 1)}}.$$

Acknowledgements

I thank Stan Osher and Justin Romberg for the suggestion of publishing this note at the Institute for Pure and Applied Mathematics workshop ‘‘Challenges in Synthetic Aperture Radar,’’ February 6–10, 2012. I thank the anonymous referees and Deanna Needell for pointing out the reference [Needell and Ward 2012] which helps me appreciate more deeply the strengths and weaknesses of my approach. I am grateful to Wenjing Liao for preparing Figures 2–4.

References

- [Beck and Teboulle 2009] A. Beck and M. Teboulle, ‘‘Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems’’, *IEEE Trans. Image Process.* **18**:11 (2009), 2419–2434.
- [Candès 2008] E. J. Candès, ‘‘The restricted isometry property and its implications for compressed sensing’’, *C. R. Acad. Sci. Paris Sér. I Math.* **346**:9–10 (2008), 589–592.
- [Candès and Plan 2009] E. J. Candès and Y. Plan, ‘‘Near-ideal model selection by ℓ_1 minimization’’, *Ann. Stat.* **37**:5A (2009), 2145–2177.
- [Candès and Tao 2005] E. J. Candès and T. Tao, ‘‘Decoding by linear programming’’, *IEEE Trans. Inf. Theory* **51**:12 (2005), 4203–4215.

- [Candès et al. 2006] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information”, *IEEE Trans. Inf. Theory* **52:2** (2006), 489–509.
- [Chambolle and Lions 1997] A. Chambolle and P.-L. Lions, “Image recovery via total variation minimization and related problems”, *Numer. Math.* **76:2** (1997), 167–188.
- [Chan and Shen 2005] T. F. Chan and J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*, Society for Industrial and Applied Mathematics, Philadelphia, 2005.
- [Chen and Huo 2006] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors”, *IEEE Trans. Signal Process.* **54:12** (2006), 4634–4643.
- [Chen et al. 2001] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit”, *SIAM Rev.* **43:1** (2001), 129–159.
- [Cheung 1998] W.-S. Cheung, “Discrete Poincaré-type inequalities”, *Tamkang J. Math.* **29:2** (1998), 145–153.
- [Claerbout and Muir 1973] J. F. Claerbout and F. Muir, “Robust modeling with erratic data”, *Geophys.* **38:5** (1973), 826–844.
- [Colton and Kress 1998] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, 2nd ed., Applied Mathematical Sciences **93**, Springer, Berlin, 1998. 3rd ed. in 2013.
- [Cotter et al. 2005] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors”, *IEEE Trans. Signal Process.* **53:7** (2005), 2477–2488.
- [Davis et al. 1997] G. Davis, S. Mallat, and M. Avellaneda, “Adaptive greedy approximations”, *Constr. Approx.* **13:1** (1997), 57–98.
- [Donoho 2006] D. L. Donoho, “Compressed sensing”, *IEEE Trans. Inf. Theory* **52:4** (2006), 1289–1306.
- [Donoho and Elad 2003] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization”, *Proc. Nat. Acad. Sci. USA* **100:5** (2003), 2197–2202.
- [Donoho and Huo 2001] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition”, *IEEE Trans. Inf. Theory* **47:7** (2001), 2845–2862.
- [Donoho et al. 2006] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise”, *IEEE Trans. Inf. Theory* **52:1** (2006), 6–18.
- [Fannjiang 2010] A. C. Fannjiang, “Compressive inverse scattering, II: Multi-shot SISO measurements with born scatterers”, *Inverse Probl.* **26:3** (2010), Article ID #035009.
- [Fannjiang et al. 2010] A. C. Fannjiang, T. Strohmer, and P. Yan, “Compressed remote sensing of sparse objects”, *SIAM J. Imaging Sci.* **3:3** (2010), 595–618.
- [Golub and Van Loan 1996] G. H. Golub and C. F. Van Loan, *Matrix computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
- [Natterer 1986] F. Natterer, *The mathematics of computerized tomography*, B. G. Teubner, Stuttgart, 1986.
- [Needell and Ward 2012] D. Needell and R. Ward, “Stable image reconstruction using total variation minimization”, preprint, 2012. [arXiv 1202.6429v6](https://arxiv.org/abs/1202.6429v6)
- [Patel et al. 2012] V. M. Patel, R. Maleh, A. C. Gilbert, and R. Chellappa, “Gradient-based image recovery methods from incomplete Fourier measurements”, *IEEE Trans. Image Process.* **21:1** (2012), 94–105.

- [Pati et al. 1993] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition”, pp. 40–44 in *Conference record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA, 1993), vol. 1, edited by A. Singh, IEEE Computer Society Press, Los Alamitos, CA, 1993.
- [Rudin and Osher 1994] L. I. Rudin and S. Osher, “Total variation based image restoration with free local constraints”, *Proc. IEEE ICIP* **1** (1994), 31–35.
- [Rudin et al. 1992] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms”, *Physica D* **60**:1–4 (1992), 259–268.
- [Taylor et al. 1979] H. L. Taylor, S. C. Banks, and J. F. McCoy, “Deconvolution with the $\ell - 1$ norm”, *Geophys.* **44**:1 (1979), 39–52.
- [Tropp 2004] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation”, *IEEE Trans. Inf. Theory* **50**:10 (2004), 2231–2242.
- [Tropp et al. 2006] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation, I: Greedy pursuit”, *Signal Process.* **86**:3 (2006), 572–588.
- [Weiss et al. 2009] 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 (2009), 2047–2080.

Received 11 May 2012. Revised 17 Sep 2012. Accepted 12 Nov 2012.

ALBERT FANNJIANG: fannjiang@math.ucdavis.edu
 Department of Mathematics, University of California, Davis, One Shields Ave.,
 Davis, CA 95616-8633, United States



MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS

msp.org/memocs

EDITORIAL BOARD

ANTONIO CARCATERRA	Università di Roma “La Sapienza”, Italia
ERIC A. CARLEN	Rutgers University, USA
FRANCESCO DELL’ISOLA	(CO-CHAIR) Università di Roma “La Sapienza”, Italia
RAFFAELE ESPOSITO	(TREASURER) Università dell’Aquila, Italia
ALBERT FANNJIANG	University of California at Davis, USA
GILLES A. FRANCFORT	(CO-CHAIR) Université Paris-Nord, France
PIERANGELO MARCATI	Università dell’Aquila, Italy
JEAN-JACQUES MARIGO	École Polytechnique, France
PETER A. MARKOWICH	DAMTP Cambridge, UK, and University of Vienna, Austria
MARTIN OSTOJA-STARZEWSKI	(CHAIR MANAGING EDITOR) Univ. of Illinois at Urbana-Champaign, USA
PIERRE SEPPECHER	Université du Sud Toulon-Var, France
DAVID J. STEIGMANN	University of California at Berkeley, USA
PAUL STEINMANN	Universität Erlangen-Nürnberg, Germany
PIERRE M. SUQUET	LMA CNRS Marseille, France

MANAGING EDITORS

MICOL AMAR	Università di Roma “La Sapienza”, Italia
CORRADO LATTANZIO	Università dell’Aquila, Italy
ANGELA MADEO	Université de Lyon–INSA (Institut National des Sciences Appliquées), France
MARTIN OSTOJA-STARZEWSKI	(CHAIR MANAGING EDITOR) Univ. of Illinois at Urbana-Champaign, USA

ADVISORY BOARD

ADNAN AKAY	Carnegie Mellon University, USA, and Bilkent University, Turkey
HOLM ALTENBACH	Otto-von-Guericke-Universität Magdeburg, Germany
MICOL AMAR	Università di Roma “La Sapienza”, Italia
HARM ASKES	University of Sheffield, UK
TEODOR ATANACKOVIĆ	University of Novi Sad, Serbia
VICTOR BERDICHEVSKY	Wayne State University, USA
GUY BOUCHITTÉ	Université du Sud Toulon-Var, France
ANDREA BRAIDES	Università di Roma Tor Vergata, Italia
ROBERTO CAMASSA	University of North Carolina at Chapel Hill, USA
ERIC DARVE	Stanford University, USA
FELIX DARVE	Institut Polytechnique de Grenoble, France
ANNA DE MASI	Università dell’Aquila, Italia
GIANPIETRO DEL PIERO	Università di Ferrara and International Research Center MEMOCS, Italia
EMMANUELE DI BENEDETTO	Vanderbilt University, USA
BERNOLD FIEDLER	Freie Universität Berlin, Germany
IRENE M. GAMBA	University of Texas at Austin, USA
SERGEY GAVRILYUK	Université Aix-Marseille, France
TIMOTHY J. HEALEY	Cornell University, USA
DOMINIQUE JEULIN	École des Mines, France
ROGER E. KHAYAT	University of Western Ontario, Canada
CORRADO LATTANZIO	Università dell’Aquila, Italy
ROBERT P. LIPTON	Louisiana State University, USA
ANGELO LUONGO	Università dell’Aquila, Italia
ANGELA MADEO	Université de Lyon–INSA (Institut National des Sciences Appliquées), France
JUAN J. MANFREDI	University of Pittsburgh, USA
CARLO MARCHIORO	Università di Roma “La Sapienza”, Italia
GÉRARD A. MAUGIN	Université Paris VI, France
ROBERTO NATALINI	Istituto per le Applicazioni del Calcolo “M. Picone”, Italy
PATRIZIO NEFF	Universität Duisburg-Essen, Germany
ANDREY PIATNITSKI	Narvik University College, Norway, Russia
ERRICO PRESUTTI	Università di Roma Tor Vergata, Italy
MARIO PULVIRENTI	Università di Roma “La Sapienza”, Italia
LUCIO RUSSO	Università di Roma “Tor Vergata”, Italia
MIGUEL A. F. SANJUAN	Universidad Rey Juan Carlos, Madrid, Spain
PATRICK SELVADURAI	McGill University, Canada
ALEXANDER P. SEYRANIAN	Moscow State Lomonosov University, Russia
MIROSLAV ŠILHAVÝ	Academy of Sciences of the Czech Republic
GUIDO SWEERS	Universität zu Köln, Germany
ANTOINETTE TORDESILLAS	University of Melbourne, Australia
LEV TRUSKINOVSKY	École Polytechnique, France
JUAN J. L. VELÁZQUEZ	Bonn University, Germany
VINCENZO VESPRI	Università di Firenze, Italia
ANGELO VULPIANI	Università di Roma La Sapienza, Italia

MEMOCS (ISSN 2325-3444 electronic, 2326-7186 printed) is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell’Aquila, Italy.

Cover image: “Tangle” by ©John Horigan; produced using the *Context Free* program (contextfreetart.org).

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing
<http://msp.org/>

© 2013 Mathematical Sciences Publishers

Dislocations, imperfect interfaces and interface cracks in anisotropic elasticity for quasicrystals Xu Wang and Peter Schiavone	1
Localization of point vortices under curvature perturbations Roberto Garra	19
Contraction of the proximal map and generalized convexity of the Moreau–Yosida regularization in the 2-Wasserstein metric Eric A. Carlen and Katy Craig	33
Ptolemy’s longitudes and Eratosthenes’ measurement of the earth’s circumference Lucio Russo	67
TV-min and greedy pursuit for constrained joint sparsity and application to inverse scattering Albert Fannjiang	81
On the theory of diffusion and swelling in finitely deforming elastomers Gary J. Templet and David J. Steigmann	105

MEMOCS is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell’Aquila, Italy.

