HHS Public Access

Author manuscript

IEEE Trans Signal Process. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

IEEE Trans Signal Process. 2018 January ; 66(1): 236–250. doi:10.1109/TSP.2017.2750111.

Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples

Greg Ongie [Member, IEEE],
Department of EECS, University of Michigan, Ann Arbor, MI 48108 USA

Sampurna Biswas [Student Member, IEEE], and
Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, 52245 USA

Mathews Jacob [Senior Member, IEEE]
Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, 52245 USA

Abstract

We consider the recovery of a continuous domain piecewise constant image from its non-uniform Fourier samples using a convex matrix completion algorithm. We assume the discontinuities/edges of the image are localized to the zero level-set of a bandlimited function. This assumption induces linear dependencies between the Fourier coefficients of the image, which results in a two-fold block Toeplitz matrix constructed from the Fourier coefficients being low-rank. The proposed algorithm reformulates the recovery of the unknown Fourier coefficients as a structured low-rank matrix completion problem, where the nuclear norm of the matrix is minimized subject to structure and data constraints. We show that exact recovery is possible with high probability when the edge set of the image satisfies an incoherency property. We also show that the incoherency property is dependent on the geometry of the edge set curve, implying higher sampling burden for smaller curves. This paper generalizes recent work on the super-resolution recovery of isolated Diracs or signals with finite rate of innovation to the recovery of piecewise constant images.

Index Terms

Off-the-Grid Image Recovery; Structured Low-Rank Matrix Completion; Finite Rate of Innovation

I. Introduction

The direct recovery of continuous domain signals by convex optimization is emerging as a powerful alternative to traditional discrete domain compressed sensing [1]–[3]. The ability of these continuous domain “off-the-grid” schemes to minimize discretization errors makes them attractive in practical applications, where only the low-pass measurements of the signal are available. The history of such continuous domain signal recovery algorithms dates back to Prony [4], where the recovery of a linear combination of exponentials from uniform samples is considered. Prony-like algorithms recover the signal by estimating an annihilating polynomial whose zeros correspond to the frequencies of the exponentials. Work by Liang et

Correspondence to: Mathews Jacob.

al. [5], [6] and the finite rate of innovation (FRI) framework [7] extended Prony-like methods to recover more general signals that reduce to a sparse linear combination of Dirac delta functions under an appropriate transformation (e.g., differential operators, convolution). Recently, several authors have further extended FRI methods to recover such signals from their non-uniform Fourier samples [3], [8]– [11] by exploiting the low-rank structure of an enhanced matrix built from Fourier data (e.g., a Hankel matrix in 1-D). Recovery guarantees exists for certain classes of these signals when the singularities are isolated and well-separated [2], [3], [12].

The signal models discussed above have limited flexibility in exploiting the extensive additional structure present in multi-dimensional imaging problems. In particular, the edges in multidimensional images are connected and can be modeled as smooth curves or surfaces. While discrete image representations to capture this structure have been the subject extensive research [13], [14], similar continuous domain representations have attracted less attention. We recently introduced a novel framework recover piecewise polynomial images, whose edges are localized to smooth curves, from their uniform [15], [16] and non-uniform [11] Fourier samples; our framework generalizes a recent extension of FRI models to curves [17]. We assume that the partial derivatives of the signal vanish outside the zero level-set of a bandlimited function, which is only true for piecewise smooth signals. This relation translates to a linear system of convolution equations involving the uniform Fourier samples of the partial derivatives, which can be compactly represented as the multiplication of a specific structured matrix with the Fourier coefficients of the bandlimited function. We have introduced theoretical guarantees for the recovery of such images from uniform samples [15], [16]. Our earlier work has shown that the structured matrix built from the Fourier coefficients of piecewise constant images is low-rank [11], [16], which we used to recover the image from its non-uniform Fourier samples with good performance in practical applications. We have also introduced an computationally efficient algorithm termed as GIRAF, which works on the original signal samples rather than the structured high-dimensional matrix [18], [19]; the computational complexity of this algorithm is comparable to discrete total variation regularization, which makes this scheme readily applicable to large-scale imaging problems, such as undersampled dynamic magnetic resonance image reconstruction [20].

The main focus of the present paper is to introduce theoretical guarantees on the recovery of continuous domain piecewise constant images from *non-uniform* Fourier samples via a convex structured low-rank matrix completion algorithm. Our main result shows number of non-uniform samples to recovery the image is proportional to the complexity of the edge set, as measured by the bandwidth of the edge set function, and an incoherence measure related to the edge set geometry. We additionally show that the recovery is robust to noise and model-mismatch.

The proof of the main result builds off of [3], which proved similar recovery guarantees for the recovery of multi dimensional isolated Diracs from non-uniform Fourier samples by minimizing the nuclear norm of an “enhanced” multi-level Hankel matrix. This work showed that the number of samples necessary for recovery depends the number of Diracs and on an incoherence measure of the signal, that can be defined solely in terms of the

relative locations of the Diracs. However, the theory in [3] relies heavily on an explicit factorization of the enhanced matrix (e.g., Vandermonde factorization of a Hankel matrix in the 1-D case), which is only available when the number of singularities are isolated and finite. Since the singularities in the proposed class of piecewise constant images (i.e., the image edges) are not isolated nor finite, the recovery guarantees in [3] cannot be directly extended to our setting. Instead, to achieve our result, we give a new characterization of the row and column spaces of the structured matrix arising in our setting. We show this new characterization allows us to derive an incoherence measure based solely on geometric properties of the edge set. In particular, we derive an upper bound for the incoherence measure that is related to the size of edge set curve. The results show that high sampling burden is associated with the estimation of images with smaller piecewise constant regions, which is consistent with intuition.

We note that the signal models in [1]–[3] do not include the class of piecewise constant images considered in this work. In particular, all of the above models assume the discontinuities to be finite in number and well separated, unlike in our setting. Recently, [12] adapted the results in [3] to introduce recovery guarantees for Fourier interpolation of a variety of finite-rate-of-innovation signal models [7], including piecewise constant functions. However, these results are limited to the 1-D setting and share the assumption than the discontinuities/innovations of the signal are finite and isolated. Furthermore, the structured matrix lifting considered in this work is different than those considered in [3] and [12]. Specifically, the structured matrix lifting in this work consists of two vertically concatenated multi-level Toeplitz matrices (i.e., block Toeplitz with Toeplitz blocks), whose entries are built from the weighted Fourier coefficients of the images. This is substantially different from the structured matrix liftings considered in [3] (unweighted, one block, single block multilevel Hankel) and [12] (weighted, one block, single-level Hankel). Finally, we note that a preliminary version of the results presented in this has been published previously in the conference paper [21] without proofs. The present work includes considerably more details and proofs, and major improvements to the main theorem.

A. Notation

Bold lower-case letters $\boldsymbol{x}$ are used to indicate vector quantities, bold upper-case $\boldsymbol{X}$ to denote matrices, and calligraphic script $\mathscr{X}$ for general linear operators. We typically reserve lower-case greek letters μ, γ, *etc.* for trigonometric polynomials (3) and upper-case greek letters Λ,Ω, *etc.* for their coefficient index sets, i.e. finite subsets of the integer lattice $\mathbb{Z}^2$, with cardinality denoted by $|\Lambda|$. We write $\Lambda+\Omega$ for the dilation of the index set Ω by Λ, i.e. the Minkowski sum $\{\boldsymbol{k}+\boldsymbol{\ell}\ \boldsymbol{k}\in\Lambda, \boldsymbol{\ell}\in\Omega\}$, and write 2Λ to mean $\Lambda+\Lambda$, $3\Lambda=2\Lambda+\Lambda$, *etc.* We also denote the contraction of Ω by Λ by $\Omega:\Lambda=\{\boldsymbol{\ell}\in\Omega:\boldsymbol{\ell}-\boldsymbol{k}\in\Omega \text{ for all } \boldsymbol{k}\in\Lambda\}$.

II. Background

A. 2-D Piecewise Constant Images with Bandlimited Edges

In this work we consider a continuous domain *piecewise constant* model for images,

$$f(\boldsymbol{r}) = \sum_{i=1}^{N} a_i 1_{U_i}(\boldsymbol{r}), \quad \text{for all} \quad \boldsymbol{r} = (x, y) \in [0, 1]^2, \quad (1)$$

where $a_i \in \mathbb{C}$, 1_U denotes the characteristic function of the set U, and each $U_i \subset [0, 1]^2$ is a simply connected regions with piecewise smooth boundaries ∂U_i. We study the recovery of such an image from a sampling of its Fourier coefficients $\hat{f}$ specified by

$$\hat{f}[\boldsymbol{k}] = \int_{[0,1]^2} f(\boldsymbol{r}) e^{-j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}; \boldsymbol{k} \in \Omega \subset \mathbb{Z}^2. \quad (2)$$

Following [16], we further assume that the edge set of the piecewise constant image, specified by $E := \cup_i \partial U_i$, coincides with the zero set of a 2-D bandlimited function:

$$E = \{\boldsymbol{r} \in [0, 1]^2 : \mu(\boldsymbol{r}) = 0\}, \text{ with } \mu(\boldsymbol{r}) = \sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}] e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}, \quad (3)$$

where the coefficients $c[\boldsymbol{k}] \in \mathbb{C}$, and Λ is a finite subset of $\mathbb{Z}^2$. We call any function μ in the form (3) a *trigonometric polynomial*, and we say μ is bandlimited to Λ, i.e., the Fourier coefficients $\hat{\mu}$ are supported within Λ. For short, we will write $\{\mu = 0\}$ for the zero set of μ considered as a subset of $[0, 1]^2$.

Define the *degree* of a trigonometric polynomial μ, denoted by $deg(\mu) = (K,L)$ to be the linear dimensions of the smallest rectangle containing the support set $\{\boldsymbol{k} : \hat{\mu}[\boldsymbol{k}] \neq 0\}$. In [16] we proved that for every curve E given by the zero set of a trigonometric polynomial, there exists a unique minimal degree trigonometric polynomial[1] μ_0 such that $E = \{\mu_0 = 0\}$ and if μ is any other trigonometric polynomial with $\{\mu_0 = 0\} \subset \{\mu = 0\}$, then $deg(\mu_0) \leq deg(\mu)$ entrywise. By extension, we define the *degree* of a curve E to be equal to the degree of of its minimial degree polynomial μ_0. We also say the curve E is *bandlimited* to $\Lambda_0 \subset \mathbb{Z}^2$, where Λ_0 is the minimal rectangular index set containing the support of $\hat{\mu}$. Intuitively, the degree/bandwidth of a curve gives a quantitative measure of its complexity. For example, in [16] we show the number of connected components of a curve is bounded in terms of its degree.

B. Recovery from uniform Fourier samples

We have shown in [16] that when μ is any trigonometric polynomial that vanishes on the edge set of the piecewise constant image f, the gradient $\nabla f = (\partial_x f, \partial_y f)$ satisfies the property

$$\mu \nabla f = 0, \quad (4)$$

[1] More precisely, μ_0 is unique up to multiplication by a phase factor $e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}$ for some $\boldsymbol{k} \in \mathbb{Z}^2$.

where equality in (4) is understood in the sense of distributions (see, e.g., [22]). The spatial domain annihilation relation (4) translates directly to the following convolution annihilation relation in Fourier domain:

$$\sum_{\boldsymbol{k}\in\Lambda} \widehat{\partial f}[\boldsymbol{\ell}-\boldsymbol{k}]\,\hat{\mu}[\boldsymbol{k}] = \mathbf{0}, \quad \forall \boldsymbol{\ell}\in\mathbb{Z}^2. \quad (5)$$

Here $\widehat{\partial f}[\boldsymbol{k}] = j2\pi(k_x\hat{f}[\boldsymbol{k}], k_y\hat{f}[\boldsymbol{k}])$ for $\boldsymbol{k} = (k_x, k_y)$. Note the equations in (5) are linear with respect to the coefficients $\hat{\mu}$.

Suppose we have access to samples of the Fourier coefficients $\hat{f}$ on a finite rectangular grid $\Gamma \subset \mathbb{Z}^2$, and suppose μ is bandlimited to $\Lambda_1 \subset \Gamma$. Then we can build the system of equations in (5) for all $\boldsymbol{\ell}$ belonging to the index set $\Lambda_2 \subset \Gamma$, where Λ_2 is the set of all integer shifts of Λ_1 contained in Γ. In this case (5) can be compactly represented in matrix form as

$$\mathcal{T}(\hat{f})\boldsymbol{h} = \begin{bmatrix} \mathcal{T}_x(\hat{f}) \\ \mathcal{T}_y(\hat{f}) \end{bmatrix} \boldsymbol{h} = \mathbf{0}, \quad (6)$$

where $\mathcal{T}_x(\hat{f}), \mathcal{T}_y(\hat{f}) \in \mathbb{C}^{|\Lambda_2|\times|\Lambda_1|}$ are matrices corresponding to the discrete 2-D convolution with the arrays $k_x\hat{f}[k_x, k_y]$ and $k_x\hat{f}[k_x, k_y]$ for $(k_x, k_y) \in \Gamma$, respectively (after omitting the inconsequential factor $j2\pi$). Here we use $\boldsymbol{h}$ to denote the vectorized version of the filter $(\hat{\mu}[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda_1)$, where the index set Λ_1 is called the *filter support*. The matrices $\mathcal{T}_x(\hat{f})$ and $\mathcal{T}_y(\hat{f})$ have a block Toeplitz with Toeplitz blocks structure. See Figure 2 for an illustration of the construction of $\mathcal{T}(\hat{f})$.

Equation (6) shows that $\mathcal{T}(\hat{f})$ is rank deficient, since it has the non-trivial vector $\boldsymbol{h}$ in its nullspace. In addition, when the filter support Λ_1 defining $\mathcal{T}(\hat{f})$ is sufficiently big, we can also show $\mathcal{T}(\hat{f})$ is low-rank. This is because if μ_0 is the minimal degree polynomial for the edge set, then *any* multiple of $\mu = \gamma \cdot \mu_0$ bandlimited to Λ_1 will satisfy the annihilation equation (4). In Fourier domain, this means the vector

$$\boldsymbol{h} = ((\hat{\mu}_0 * \hat{\gamma})[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda_1) \quad (7)$$

is in the nullspace of $\mathcal{T}(\hat{f})$. Hence if the filter support Λ_1 is larger than support Λ_0 of μ_0, $\mathcal{T}(\hat{f})$ has a large nullspace and is low-rank. The following result from [16] gives an exact characterization of the rank of $\mathcal{T}(\hat{f})$, which will be important for this work:

Theorem 1—[16] *Suppose f is a piecewise constant image* (1) *whose edge set* $E = \{\mu_0 = 0\}$ *is the zero set of a trigonometric polynomial* μ_0 *bandlimited to* Λ_0. *Let* $\mathcal{T}(\hat{f})$ *be built with filter size* $\Lambda_1 \supseteq \Lambda_0$, *then*

$$rank\, \mathscr{T}(\hat{f}) \leq |\Lambda_1| - |\Lambda_1 : \Lambda_0| \quad (8)$$

where $|\Lambda_1|$ *is the number of indices in* Λ_1 *and* $|\Lambda_1 : \Lambda_0|$ *is the number of integer shifts of* Λ_0 *contained in* Λ_1. *Moreover, equality holds in* (8) *if* $\Gamma \supseteq 2\Lambda_1 + \Lambda_0$ *and if the edge set does not contain any singular points. In this case, the nullspace of* $\mathscr{T}(\hat{f})$ *consists of all vectors in the form* (7).

Note that $R := |\Lambda_1| - |\Lambda_1 : \Lambda_0|$ is a measure of the bandwidth of μ_0 and hence is indicative of the complexity of the edge set curve $E = \{\mu_0 = 0\}$. In the remainder of this work we assume the conditions in Theorem 1 that guarantee the equality rank $\mathscr{T}(\hat{f}) = R$ holds, in particular $\Gamma \supseteq 2\Lambda_1 + \Lambda_0$.

If we take $\Lambda_1 = \Lambda_0$, the above result shows Fourier samples of $\hat{f}$ in $\Gamma \supseteq 3\Lambda_0$ is sufficient for the recovery of the minimal degree polynomial μ_0, since in this case $\hat{\mu_0}$ can be identified as the unique non-trivial nullspace vector of $\mathscr{T}(\hat{f})$. The following theorem states that once μ_0 is available, f is the unique solution to the annihilation equations (4) and (5):

Theorem 2—[16]. *Suppose* f *is a piecewise constant image* (1) *whose edge set* $E = \{\mu_0 = 0\}$ *is the zero set of a trigonometric polynomial* μ_0 *bandlimited to* Λ_0. *Suppose the Fourier sampling set* $\Gamma \supseteq \Lambda_0$. *If* $g \in L^1([0,1]^2)$ *satisfies*

$$\mu_0 \nabla g = 0 \text{ subject to } \hat{g}[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}] \text{ for all } \boldsymbol{k} \in \Gamma, \quad (9)$$

then $g = f$ *almost everywhere.*

In principle, this result allows us to solve for the amplitudes of regions of the piecewise constant function f by plugging in the known μ_0 into the equation (9) and solving a linear system, similar to Prony's method. However, for complicated piecewise constant images with many regions, it may be more practical to use the approximations introduced in [16].

III. Recovery from non-uniform Fourier samples

The theory presented in Section I shows that the exact recovery of a continuous domain piecewise constant image with a bandlimited edge set is possible when we collect Fourier samples of the image on a sufficiently large uniform grid in Fourier domain. However, the recovery procedure breaks down when we have non-uniform or missing samples, which is often the case in practical settings, e.g., compressed sensing MRI [23]. Therefore, we propose and analyze a method to interpolate the missing samples to a uniform grid in Fourier domain, which guarantees full recovery of the image in spatial domain.

Recall that Theorem 1 says that the structured matrix $\mathscr{T}(\hat{f})$ built from the Fourier coefficients $\hat{f}[\boldsymbol{k}]$, $\boldsymbol{k} \in \Gamma$, where $\Gamma \subset \mathbb{Z}^2$ is a uniform rectangular grid, is known to be low-rank precisely when f is a piecewise constant image with a bandlimited edge set. Hence we

propose to recover $\hat{f}[\boldsymbol{k}]$, $\boldsymbol{k} \in \Gamma$ from its samples at non-uniform locations $\Omega \subset \Gamma$ as the solution to the convex matrix completion problem: min

$$\min_{\hat{g}[\boldsymbol{k}],\, \boldsymbol{k} \in \Gamma} \|\mathscr{T}(\hat{g})\|_* \text{ subject to } \hat{g}[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}] \text{ for all } k \in \Omega \quad (10)$$

where $\|\cdot\|^*$ denotes the nuclear norm, i.e., the sum of the singular values of a matrix, which is the convex relation of the rank functional. Note that (10) is different than the standard low-rank matrix completion setting studied in [24], [25] in that the low-rank matrix $\mathscr{T}(\hat{f})$ is structured and parameterized by the coefficient vector $\hat{f}$. Similar structured low-rank matrix completion schemes have been proposed for the recovery of signals from non-uniform Fourier samples [3], [12] and used with empirical success in MRI applications [10], [11], [26]. The main focus of this paper is to determine the sufficient number of samples that will ensure exact recovery of the Fourier coefficients of f on the reconstruction grid Γ with high probability.

A. Role of incoherence

Several authors have shown that the sufficient number of samples for low-rank matrix recovery by nuclear norm minimization to succeed is dependent on the *incoherence* of the sampling basis with respect to the matrix to be to be recovered [3], [25]. Similarly, our results depend on an incoherence measure derived from the structure of the matrix $\mathscr{T}(\hat{f})$ and properties of the piecewise constant image f. In particular, define $\mathscr{P}_U$ and $\mathscr{P}_V$ to be the orthogonal projections onto the column space and row space of $\mathscr{T}(\hat{f})$, respectively, i.e., if $\mathscr{T}(\hat{f}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ is the rank-R singular value decomposition then $\mathscr{P}_U\boldsymbol{X} = \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{X}$, $\mathscr{P}_V\boldsymbol{X} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^*$. In Appendix B, we show that the structured matrix $\mathscr{T}(\hat{f})$ can be expanded using orthonormal basis of matrices $\mathbf{A}_k$ such that

$$\mathscr{T}(\hat{f}) = \sum_{\boldsymbol{k} \in \Gamma \setminus \{0\}} \hat{f}[\boldsymbol{k}] w[\boldsymbol{k}] \boldsymbol{A}_{\boldsymbol{k}} \quad (11)$$

where $w[\boldsymbol{k}]$, $\boldsymbol{k} \in \Gamma \setminus \{0\}$ are a set of positive weights that do not depend on $\hat{f}$. Similar to results in [3], [12], [25], we prove that nuclear norm minimization (10) recovers the exact low-rank solution with high probability provided we can uniformly bound the norms of the projections of the sampling basis matrices $\boldsymbol{A}_{\boldsymbol{k}}$ onto the row and column spaces of $\mathscr{T}(\hat{f})$:

Proposition 3—*Consider $\mathscr{T}(\hat{f})$ of rank R corresponding to a piecewise constant function f whose edge set coincides with the zero set of μ_0, let ρ be the incoherency measure associated with μ_0 to be defined in the sequel, and set $c_s = |\Gamma|/|\Lambda_1|$. Then we have*

$$\max_{\boldsymbol{k} \in \Gamma} \|\mathscr{P}_U \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 \leq \frac{\rho R c_s}{|\Gamma|}, \quad (12)$$

$$\max_{k \in \Gamma} \|\mathscr{P}_V \boldsymbol{A}_k\|_F^2 \leq \frac{\rho R c_s}{|\Gamma|} \quad (13)$$

The proof in Section VIII-F relies on the row and column spaces of $\mathscr{T}(\hat{f})$ derived in Lemma 8 and Lemma 6 in the next section. These results will be used in the derivation of the main theorem in Section IX.

B. Main Results

We now present our main results, which determine the sufficient number of random Fourier samples for the convex structured low-rank matrix completion program (10) to succeed with high probability. Our first theorem addresses the case of recovery from noiseless Fourier samples:

Theorem 4—*Let f be a continuous domain piecewise constant image* (1)*, whose edge-set is described by the zero-set of the trigonometric polynomial μ_0 bandlimited to Λ_0 (see* (3)*). Let $\Omega \subset \Gamma$ be an index set drawn uniformly at random within Γ. Then there exists a universal constant $c > 0$ such that the solution to* (10) *is $\hat{f}$ with probability exceeding $1 - |\Gamma|^{-2}$, provided*

$$|\Omega| > c \rho c_s R \log^4 |\Gamma|, \quad (14)$$

where $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0| = \text{rank}\, \mathscr{T}(\hat{f})$, $c_s = |\Gamma|/|\Lambda_1|$, c is a universal constant, and $\rho \geq 1$ is an incoherence measure depending on the geometry of the edge-set, to be defined in the sequel.

To better understand the dependence of the bound in (14) on the filter size Λ_1 and the edge set bandwidth Λ_0, assume for simplicity that Λ_1 is some dilation of Λ_0, that is, $\Lambda_1 = \alpha\Lambda_0$, where $\alpha > 1$ is an integer. In this case, the factor $c_s R$ in (14) simplifies to

$$\left(\frac{|\Lambda_1| - |\Lambda_1 : \Lambda_0|}{|\Lambda_1|}\right)|\Gamma| \leq \left(\frac{\alpha^2 - (\alpha-1)^2}{\alpha^2}\right)|\Gamma| \leq \frac{2|\Gamma|}{\alpha}. \quad (15)$$

Therefore, assuming the other constants in (14) are fixed, the number of measurements sufficient for exact recovery is proportional to the reciprocal of the dilation factor α. This suggests taking the filter size Λ_1 to be as large as allowed by Theorem 4. Namely, Λ_1 should satisfy $2\Lambda_1 + \Lambda_0 = \Gamma$, i.e., the side-lengths of filter support Λ_1 should be roughly half those of the reconstruction grid Γ. Fixing the filter support Λ_1 to obey this bound, then $\Gamma = (2\alpha + 1)\Lambda_0$, and so $|\Gamma| \leq (2\alpha + 1)^2|\Lambda_0|$. Inserting this bound into (15) gives

$$c_s R = O(\alpha \mid \Lambda_0 \mid). \quad (16)$$

Combined with (14), this shows that the number of measurements sufficient for exact recovery is on the order of $|\Lambda_0|$, up to incoherence and log factors.

The proof of Theorem 4, detailed in Appendix B, is in line with the approach of [3]. In particular, we prove the result by constructing an approximate dual certificate using the well-known "golfing scheme" of [25]. The main differences between in the proof of the above result and that in [3] results from the differences in the matrix structure and hence the characterization of the incoherency between the row and column subspaces of $\mathcal{T}(\hat{f})$ with the sampling basis. In particular, the matrix $\mathcal{T}(\hat{f})$ we consider is obtained by stacking two block Toeplitz with Toeplitz blocks (BTTB) matrices whose entries are the weighted Fourier coefficients of f, as opposed to a single unweighted BTTB matrix in [3]. The approach in [3] relies on an explicit low-rank factorization of a BTTB matrix in terms of Vandermonde-like matrices[2]. Since this factorization is not available in our setting, we use algebraic properties of trigonometric polynomials to give a new characterization of the row and column spaces of the matrix. In particular, we show in Section IV that similar Vandermonde-like basis matrices exist for the row and column space of the lifted matrix, and use these to derive a related incoherence measure that satisfies the bounds in Prop. 3.

C. Recovery in the presence of noise and model-mismatch

We now generalize (66) to the setting where we have noisy or corrupted Fourier samples

$$\hat{f}_n[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}] + \eta[\boldsymbol{k}], \boldsymbol{k} \in \Omega, \quad (17)$$

where $\eta[\boldsymbol{k}] \in \mathbb{C}$ is a vector of noise. In this case, we pose recovery as

$$\min_{\hat{g}} \|\mathcal{T}(\hat{g})\|_* \text{ subject to} \|\mathcal{P}_\Omega(\hat{f}_n - \hat{g})\|_2 \leq \delta. \quad (18)$$

where $\delta > 0$ is an estimate of the ℓ^2-norm of the error $\|\eta\|$, and $\mathcal{P}_\Omega$ denotes projection onto Ω. We make no assumptions on the statistics of the noise η. In particular, η can represent errors due to model-mismatch, such as when the image is not perfectly piecewise constant, or when the edge set of the image does not coincide perfectly with the zero level-set of a bandlimited function.

[2]The structured matrices considered in [3] are block Hankel with Hankel block matrices (BHHB), but this difference is purely cosmetic: every BTTB matrix can be re-expressed as BHHB after a permutation of its rows and columns. In particular, the Vandermonde-like factorization of BHHB matrices in [3] carries over to BTTB matrices.

The following theorem shows that when the deviation of $\hat{f}_n$ from $\hat{f}$ is small, the modified recovery program (18) recovers a solution that is close in norm to $\hat{f}$ under the same sampling conditions as Theorem 4.

Theorem 5—*Let f be specified by* (1)*, whose edge-set is described by the zero-set of the trigonometric polynomial μ_0 bandlimited to Λ_0 with associated incoherence measure ρ. Let $\Omega \subset \Gamma$ be an index set drawn uniformly at random within Γ such that $|\Omega|$ satisfies the bound* (14) *in Theorem 4. If the measurements $\hat{f}_n$ satisfy $\|\mathcal{P}_\Omega(\hat{f}_n - \hat{f})\|_2 \leq \delta$, then the solution $\hat{g}$ to* (18) *satisfies*

$$\|\mathcal{T}(\hat{f}) - \mathcal{T}(\hat{g})\|_F \leq 5\,|\,\Gamma\,|^2\delta\,. \quad (19)$$

with probability exceeding $1 - |\Gamma|^{-2}$.

See Section IV in the Supplementary Materials for proof. The bound (19) allows us to quantify the effect of model-mismatch on recovery. In particular, suppose the image f_n represents a perturbation from an ideal piecewise constant image f such that their difference in L^2-norm is δ-small:

$$\|f_n - f\|_{L^2}^2 = \left(\int_{[0,1]^2} |\,f_n(r) - f(\boldsymbol{r})\,|^2 d\boldsymbol{r}\right)^{\frac{1}{2}} \leq \delta\,. \quad (20)$$

Then by Parseval's theorem, the measurements of $\hat{f}_n$ satisfy $\|\mathcal{P}_\Omega(\hat{f}_n - \hat{f})\|^2 \leq \delta$, hence Theorem 5 applies. From (19) we obtain the bound $\|\mathcal{T}(\hat{f}) - \mathcal{T}(\hat{g})\|_F \leq 5|\Gamma|^2\|f_n - f\|_{L^2}$. This shows that if the image f_n is close to the ideal piecewise constant image f in spatial domain L^2-norm, then the matrix $\mathcal{T}(\hat{g})$ we recover using (18) will be close in norm to $\mathcal{T}(\hat{f})$ with high probability.

IV. Row and column spaces of $\mathcal{T}(\hat{f})$ and incoherence

In this section we define an incoherence measure ρ that satisfies the desired bounds in Prop. 3. We show that the incoherence measure depends only on the geometry of the edge set of the image. The incoherence measure is derived from a new characterization of the row and column spaces of the matrix $\mathcal{T}(\hat{f})$ in terms of Vandermonde-like basis matrices.

A. Row and column spaces of $\mathcal{T}(\hat{f})$

Our first lemma gives a basis for the row space of $\mathcal{T}(\hat{f})$:

Lemma 6—*A basis of the row space of $\mathcal{T}(\hat{f})$ is given by the columns of the $|\Lambda_1| \times R$ Vandermonde-like matrix*

$$\boldsymbol{E}_{row}(P) := \frac{1}{\sqrt{|\Lambda_1|}} \begin{pmatrix} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_1} & \cdots & e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_R} \\ \vdots & & \vdots \\ e^{j2\pi \boldsymbol{k}_{|\Lambda_1|} \cdot \boldsymbol{r}_1} & \cdots & e^{j2\pi \boldsymbol{k}_{|\Lambda_1|} \cdot \boldsymbol{r}_R} \end{pmatrix} \quad (21)$$

where $\{\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{|\Lambda_1|}\}$ *is a linear indexing of elements in* Λ_1*, and* $P = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_R\}$ *is a set of* $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$ *distinct points on the edge set curve* $\{\mu_0 = 0\}$ *chosen such that the columns of* $\boldsymbol{E}_{row}$ *are linearly independent.*

The careful reader will have noticed that Lemma 6 takes for granted the existence of a set of points $P = \{\boldsymbol{r}_1, ..., \boldsymbol{r}_R\} \subset \{\mu_0 = 0\}$ such that the columns of $\boldsymbol{E}_{\text{row}}(P)$ is linearly independent. Call such a set P a set of *admissible nodes* for the curve $\{\mu_0 = 0\}$. The following result shows that sets of admissible nodes always exist and are easy to construct:

Lemma 7—*Let* μ_0 *be bandlimited to* Λ_0*. Any set of* $M \geq R + |\Lambda_0|$ *distinct points on the curve* $\{\mu_0 = 0\}$ *contains a subset of R points that are a set of admissible nodes.*

The next lemma shows that we can characterize the column space of $\mathscr{T}(\hat{f})$ in a similar way as the row space:

Lemma 8—*A basis of the column space of* $\mathscr{T}(\hat{f})$ *is given by the columns of the* $2|\Lambda_2| \times R$ *weighted Vandermonde-like matrix:*

$$\boldsymbol{E}_{col}(P) = \frac{1}{\sqrt{|\Lambda_2|}} \begin{pmatrix} \frac{w_{1,x}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_1} & \cdots & \frac{w_{R,x}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_R} \\ \vdots & & \vdots \\ \frac{w_{1,x}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_1} & \cdots & \frac{w_{R,x}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_R} \\ \hline \frac{w_{1,y}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_1} & \cdots & \frac{w_{R,y}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_R} \\ \vdots & & \vdots \\ \frac{w_{1,y}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_1} & \cdots & \frac{w_{R,y}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_R} \end{pmatrix}, \quad (22)$$

where where $\{\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{|\Lambda_2|}\}$ *is a linear indexing of elements in* Λ_2 *and* $P = \{\boldsymbol{r}_1, ..., \boldsymbol{r}_R\}$ *is a set of admissible nodes for the curve* $\{\mu_0 = 0\}$*. The weight vectors* $\boldsymbol{w}_i = (w_{i,x}, w_{i,y})$*, are described by the formula* (52) *in* Appendix VIII*, and depend only on the edge set* $\{\mu_0 = 0\}$*, the nodes P, and the filter support* Λ_1*.*

See Section VIII-C for the proofs of Lemmas 6 and 7, and Section VIII-E for the proof of Lemma 8.

B. Incoherence measure

We now show how to define an incoherence measure ρ that satisfies the desired bounds in Prop. 3. Consider the Gram matrix $\boldsymbol{G}(P) = [\boldsymbol{E}_{\text{row}}(P)]^* \boldsymbol{E}_{\text{row}}(P)$, where P is any set of R points $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_R$ on the edge set curve $\{\mu = 0\}$. It is easy to see from the definition (21) that the entries of $\boldsymbol{G}(P)$ are specified by

$$(\boldsymbol{G}(P))_{i,j} = \frac{1}{|\Lambda_1|} D_{\Lambda_1}(r_i - r_j), \quad 1 \le i, j \le R, \quad (23)$$

where $D_{\Lambda_1}(\boldsymbol{r}) := \Sigma_{\boldsymbol{k} \in \Lambda_1} e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}$ is the *Dirichlet kernel* supported on Λ_1. Note that $\boldsymbol{G}(P)$ has ones along the diagonal, and the magnitude of the off-diagonal entries is dictated by the distances $|\boldsymbol{r}_i - \boldsymbol{r}_j|$ and the filter support Λ_1. We now define the *incoherence measure* ρ associated with the edge set $E = \{\mu_0 = 0\}$ in terms of $\boldsymbol{G}(P)$.

Definition 9—*Suppose the edge set curve $E = \{\mu_0 = 0\}$ has bandwidth Λ_0 (see* (3)*), and set $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$. Define the incoherence measure ρ by*

$$\rho = \min_{\substack{P \subset \{\mu_0 = 0\} \\ |P| = R}} \frac{1}{\lambda_{min}[\boldsymbol{G}(P)]}, \quad (24)$$

where $\lambda_{min}[\boldsymbol{G}(P)]$ is the minimum eigenvalue of $\boldsymbol{G}(P)$.

Put in words, among all possible arrangements of R points along the edge-set $\{\mu_0 = 0\}$, we seek the arrangement such that the minimum eigenvalue $\boldsymbol{G}(P)$ is as large as possible. Intuitively, the optimal arrangement will maximize the minimum separation distance among the R points, and ρ can be thought of as a measure of this geometric property. In particular, edge set curves that enclose a small area, and hence require the points P to be closely spaced along the curve, will result in a large value of ρ. According to Theorem 4, the measurement burden will be high for such curves.

Note that curves corresponding to a particular bandwidth can come in different sizes. Specifically, for a fixed μ_0 with bandwidth Λ_0 consider the family of curves $\{\mu_0 = \alpha\}$, where α is a scalar. One can change α to obtain multiple curves with exactly the same bandwidth, each of which correspond to a different levelset of μ_0. These level-sets will have different incoherence measures, depending on how large or small the level-set curves are. This shows the incoherence of an edge set captures something besides its bandwidth. See Figure 3 for an illustration.

We can give incoherency measure of an edge set a more precise geometric interpretation based on the minimum separation distance of a set of admissible nodes. We generalize a bound on the condition number of Vandermonde matrices derived in [27] to the case of the

Vandermonde-like matrix (21), and use this to derive a bound for the incoherence parameter ρ.

Theorem 10—*Assume that the points* $P = \{(x_i, y_i)\}_{i=1}^{R}$ *belonging to the curve* $\{\mu_0 = 0\}$ *satisfy* $|x_i - x_j| > \Delta$ *and* $|y_i - y_j| > \Delta$ *for all* $i \neq j$. *Assume the filter support* $\Lambda_1 \subset \mathbb{Z}^2$ *is a square region symmetric around the origin of size* $\sqrt{|\Lambda_1|} \times \sqrt{|\Lambda_1|}$. *Then*

$$\rho \leq \left(1 - \frac{1}{\sqrt{|\Lambda_1|}\Delta}\right)^{-2}, \quad (25)$$

where ρ *is the incoherence parameter* (24) *associated with the curve* $\{\mu_0 = 0\}$.

See Section I of the Supplementary Materials for the proof. The bound in (25) shows that the incoherence is close to one (i.e., is as small as possible) when $\Delta \gg 1/\sqrt{|\Lambda_1|}$. Since Δ is the spacing between each pair of points on the curve, to achieve a larger Δ spacing, and hence a smaller ρ, requires a larger curve. This suggests that fewer measurements are required to recover a larger curve, which is consistent with the findings in the isolated Dirac setting [27], [28].

V. Numerical Experiments

A. Algorithms

For small to moderate problem sizes the nuclear norm minimization problem (10) can be solved efficiently with the alternating directions method of multipliers (ADMM) algorithm, which results in a modification of the singular value thresholding (SVT) algorithm [29]. This approach has been proposed for related structured low-rank matrix completion problems in several works, e.g., [3], [11], [12], [30]. We adopt this approach here as well for our small-scale numerical experiments. A detailed implementation of this algorithm can be found in, e.g., [28]. However, we note that for large scale problems, such as those encountered in realistic imaging applications, more efficient approaches need to be adopted, because often in these cases the lifted matrix is too large to be held in memory. A fast algorithm for solving an approximation to (10) for large-scale problems is given in [19].

B. Phase transitions

In Fig. 4, we study the probability of exact recovery under different assumptions on the filter size and edge set of the image. For these experiments the reconstruction grid Γ was of size 65×65. We generated synthetic random piecewise constant functions with known edge set bandwidth (see Fig. 3(c)), and attempted to recover their Fourier coefficients in Γ from random samples in Ω at the specified undersampling factor. For each set of parameters we ran 10 random trials. We count the recovery as “exact” if the recovered coefficients $\hat{f}$ satisfied $\|\hat{f} - f_0\|/\|\hat{f}_0\| < 10^{-3}$, where $\hat{f}_0$ is the ground truth. The exact recovery rate was then obtained by averaging over the 10 trials.

First, in Fig. 4(a), we studied the effect of changing the filter size Λ_1 on the recovery while keeping other parameters constant. We fixed the edge-set bandwidth to $|\Lambda_0| = 9 \times 9$ and varied the filter size as $|\Lambda_1| = (2K+1) \times (2K+1)$ for $K = 1, \ldots, 30$. We call K the filter bandwidth. Note that Theorem 4 has restrictions on how large Λ_1 can be. The maximum filter bandwidth for which Theorem 4 holds in this case was $K = 15$ (red line in Figure 4(a)), however we extended the filter size to observe the behavior of the algorithm outside of this regime. As predicted by Theorem 4, we find that the optimal performance is obtained when Λ_1 is the largest as allowed by Theorem 4 (roughly half the size of Γ in each dimension).

Next, in Fig. 4(b), we study the recovery as a function of the bandwidth of the edge-set of the image. The filter bandwidth was fixed at $K = 15$, and we varied the edge-set bandwidth as $|\Lambda_0| = (2K_0+1) \times (2K_0+1)$. The phase transition shows dependence $|\Omega| = O(|\Lambda_0|)$ as predicted by Theorem 4.

C. Comparison with TV minimization on real MRI data

We also compare the proposed Fourier domain interpolation scheme against standard discrete TV minimization in spatial domain:

$$\min_{\boldsymbol{u} \in \mathbb{C}^N} TV(\boldsymbol{u}) \text{ subject to } P_\Omega(\boldsymbol{F}\boldsymbol{u}) = P_\Omega(\boldsymbol{F}\boldsymbol{u}_0). \quad (26)$$

Here $\boldsymbol{u} \in \mathbb{C}^N$ with $N = N_x N_y$ is a 2-D array representing a discrete $N_x \times N_y$ image, $\boldsymbol{u}_0 \in \mathbb{C}^N$ is the image to be recovered, $\boldsymbol{F} \in \mathbb{C}^{N \times N}$ denotes the unitary 2-D discrete Fourier transform (DFT) matrix acting on $N_1 \times N_2$ arrays, $\boldsymbol{P}_\Omega$ is projection onto the index of sampling locations $\Omega \subset [N_x] \times [N_y]$, and $TV(\cdot)$ denotes the (isotropic) total variation semi-norm:

$$TV(\boldsymbol{u}) = \sum_{i=1}^{N} \left(|(\partial_1 \boldsymbol{u})_i|^2 + |(\partial_2 \boldsymbol{u})_i|^2 \right)^{\frac{1}{2}} \quad (27)$$

where ∂_1 and ∂_2 are finite difference operators in the horizontal and vertical directions, respectively. The problem (26) has been studied extensively [31]–[36] as a model for undersampled MRI reconstruction and other inverse problems in imaging.

In Fig. 5 we perform an experiment comparing against TV minimization and the proposed approach on real MRI data. For this experiment we used a fully-sampled four-coil single-slice acquisition consisting of 256×256 Cartesian k-space samples, which was compressed to a single virtual coil using an SVD-based technique [37]. The data in the single virtual coil was observed to have smoothly varying complex phase in image domain. To compensate for this source of model-mismatch, we further pre-preprocessed the data by removing the complex phase in image domain. We note that this preprocessing step is unrealistic for a true MRI experiment. However, the optimization problem (10) could be modified to incorporate an estimate of the smoothly varying phase in the measurement model; we omit this step for simplicity. Finally, we retrospectively undersampled the pre-processed virtual single coil data, taking 50% uniform random samples. We find that the proposed structured low-rank

recovery shows significant improvement recovery error over standard total variation as measured by SNR = $20\log_{10}(\|\hat{f}\|/\|\hat{f}^* - \hat{f}\|)$, where $\hat{f}^*$ is the recovered data and $\hat{f}$ is the ground truth. The error images indicate the proposed method more faithfully recovers the true edges of the image.

VI. Discussion

Discrete domain total-variation minimization has played a central role in compressed sensing from its inception [31], [32], which models the image to be recovered as (approximately) piecewise constant. Since the present work can be thought of as an extension of compressed sensing type guarantees to the continuous domain setting, it is fruitful to explore the connections between our continuous domain model and discrete domain total variation.

At first glance, the structured low-rank matrix completion problem (10) may seem far removed from the TV-minimization problem (26). But, in fact, one can show TV-minimization (26) is equivalent to nuclear norm minimization of a related structured matrix lifting in Fourier domain. Specifically, (26) is equivalent to

$$\min_{\boldsymbol{v}} \|\mathscr{C}(\boldsymbol{F}\boldsymbol{u})\|_* \text{ subject to } P_\Omega(\boldsymbol{F}\boldsymbol{u}) = P_\Omega(\boldsymbol{F}\boldsymbol{u}_0). \quad (28)$$

Here

$$\mathscr{C}(\boldsymbol{F}\boldsymbol{u}) = \begin{bmatrix} \mathscr{C}_x(\boldsymbol{F}\boldsymbol{u}) \\ \mathscr{C}_y(\boldsymbol{F}\boldsymbol{u}) \end{bmatrix} \in \mathbb{C}^{2N \times N} \quad (29)$$

and $\mathscr{S}_x(\boldsymbol{F}\boldsymbol{u})$, $\mathscr{S}_y(\boldsymbol{F}\boldsymbol{u})$ are block circulant with circulant blocks matrices whose first column is specified by the arrays $\boldsymbol{v}_x = \boldsymbol{F}\partial_x\boldsymbol{u}$ and $\boldsymbol{v}_y = \boldsymbol{F}\partial_x\boldsymbol{u}$. Assuming circular boundary conditions, we can write $(\boldsymbol{v}_x)[k_x, k_y] = (1 - e^{j2\pi k_x/N_x})(\boldsymbol{F}\boldsymbol{u})[k_x, k_y]$ and $(\boldsymbol{v}_y)[k_x, k_y] = (1 - e^{j2\pi k_y/N_y})(\boldsymbol{F}\boldsymbol{u})[k_x, k_y]$.

We find it interesting to use this re-formulation of TV-minimization to better understand the proposed approach. In Table I we summarize the similarities and differences. One essential difference is the dimensions of the matrix liftings. In particular, the matrix lifting we propose has dimensions $2|\Lambda_2| \times |\Lambda_1|$, with $|\Lambda_1| \ll |\Lambda_2|$ whereas the matrix lifting associated with TV in (28) has dimensions $2N \times N$. If the reconstruction grid size is the same in both cases, i.e., $|\Gamma| = N$, then the proposed matrix lifting has substantially fewer columns than the one associated with TV. This is due to our assumption that edge set of the image has low bandwidth. In other words, we restrict the degrees of freedom of the model by constructing a lifting with fewer columns. We believe this difference may explain the success of the proposed method over TV-minimization observed empirically in Section V.

VII. Conclusion

We derived performance guarantees for the recovery of piecewise constant images from random non-uniform Fourier samples via a convex structured low-rank matrix completion problem. This was achieved by adapting results in [3] to the case of a low-rank block two-fold Toeplitz matrix with an additional weighting scheme that arises naturally when considering piecewise constant images. We also define incoherence measures that rely only on geometric properties of the edge set, which indicate that the sampling burden is higher for images with smaller piecewise constant regions.

The recovery guarantees in this work studied the case of uniform random samples. However, in practice we observe that recovery works well with when considering other types of variable density random sampling, where the low spatial frequencies are more heavily sampled. It would be interesting to adapt our results to a wider variety of sampling distributions, and to identify the optimal sampling strategy for signals belonging to our image model.

Acknowledgments

This work is supported by grants NIH 1R01EB019961-01A1 and ONR N00014-13-1-0202.

References

1. Bhaskar BN, Recht B. Atomic norm denoising with applications to line spectral estimation. Communication, Control, and Computing, 2011 49th Annual Allerton Conference on; IEEE; 2011. 261–268.
2. Candès EJ, Fernandez-Granda C. Towards a mathematical theory of super-resolution. Communications on Pure and Applied Mathematics. 2014; 67(6):906–956.
3. Chen Y, Chi Y. Robust spectral compressed sensing via structured matrix completion. Information Theory, IEEE Trans on. 2014; 60(10):6576–6601.
4. Stoica P, Moses RL. Introduction to spectral analysis. Vol. 1. Prentice hall; Upper Saddle River, NJ: 1997.
5. Haacke E, Liang ZP, Izen S. Superresolution reconstruction through object modeling and parameter estimation. Acoustics, Speech and Signal Processing, IEEE Trans on. Apr; 1989 37(4):592–595.
6. Haacke EM, Liang ZP, Izen SH. Constrained reconstruction: A superresolution, optimal signal-to-noise alternative to the fourier transform in magnetic resonance imaging. Medical Physics. 1989; 16(3):388–397. [PubMed: 2739620]
7. Vetterli M, Marziliano P, Blu T. Sampling signals with finite rate of innovation. Signal Processing, IEEE Trans on. 2002; 50(6):1417–1428.
8. Jin KH, Lee D, Ye JC. A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank hankel matrix. 2015 preprint arXiv:1504.00532.
9. Jin KH, Lee D, Ye JC. A novel k-space annihilating filter method for unification between compressed sensing and parallel mri. IEEE ISBI. 2015
10. Haldar JP. Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. Medical Imaging, IEEE Trans on. 2014; 33(3):668–681.
11. Ongie G, Jacob M. Recovery of piecewise smooth images from few fourier samples. SampTA. 2015:543–547.
12. Ye JC, Kim JM, Jin KH, Lee K. Compressive sampling using annihilating filter-based low-rank interpolation. IEEE Trans on Information Theory. 2016
13. Starck JL, Candès EJ, Donoho DL. The curvelet transform for image denoising. Image Processing, IEEE Trans on. 2002; 11(6):670–684.

14. Do MN, Vetterli M. The contourlet transform: an efficient directional multiresolution image representation. Image Processing, IEEE Trans on. 2005; 14(12):2091–2106.
15. Ongie G, Jacob M. Super-resolution MRI using finite rate of innovation curves. IEEE ISBI. 2015
16. Ongie G, Jacob M. Off-the-grid recovery of piecewise constant images from few fourier samples. May.2015 :543–547. arXiv:1510.00384.
17. Pan H, Blu T, Dragotti PL. Sampling curves with finite rate of innovation. Signal Processing, IEEE Trans. on; 2014.
18. Ongie G, Jacob M. A fast algorithm for structured low-rank matrix recovery with applications to undersampled MRI reconstruction. IEEE International Symposium on Biomedical Imaging; 2016;
19. Ongie G, Jacob M. Giraf: A fast algorithm for structured low-rank matrix recovery giraf: A fast algorithm for structured low-rank matrix recovery. 2016 arXiv:1609.07429.
20. Balachandrasekaran A, Ongie G, Jacob M. Accelerated dynamic MRI using structured low rank matrix completion. 2016 IEEE International Conference on Image Processing (ICIP); Institute of Electrical and Electronics Engineers (IEEE); 2016.
21. Ongie G, Biswas S, Jacob M. Structured matrix recovery of piecewise constant signals with performance guarantees. International Conference on Image Processing; 2016;
22. Strichartz RS. A Guide to Distribution Theory and Fourier Transforms. World Scientific Pub Co Pte Lt; 2003.
23. Lustig M, Donoho D, Pauly JM. Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic resonance in medicine. 2007; 58(6):1182–1195. [PubMed: 17969013]
24. Candès E, Recht B. Exact matrix completion via convex optimization. Commun ACM. Jun; 2012 55(6):111–119.
25. Gross D. Recovering low-rank matrices from few coefficients in any basis. Information Theory, IEEE Trans on. 2011; 57(3):1548–1566.
26. Shin PJ, Larson PE, Ohliger MA, Elad M, Pauly JM, Vigneron DB, Lustig M. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. Magnetic Resonance in Medicine. 2013
27. Moitra A. Super-resolution, extremal functions and the condition number of vandermonde matrices. Proceedings of the 47th Annual ACM on Symposium on Theory of Computing; 2015; 821–830.
28. Ye JC, Kim JM, Jin KH. Compressive sampling using structured low-rank interpolation. 2015 preprint arXiv:1511.08975.
29. Cai JF, Candès EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization. 2010; 20(4):1956–1982.
30. Fazel M, Pong TK, Sun D, Tseng P. Hankel matrix rank minimization with applications to system identification and realization. SIAM Journal on Matrix Analysis and Applications. 2013; 34(3): 946–977.
31. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans on information theory. 2006; 52(2):489–509.
32. Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics. 2006; 59(8):1207–1223.
33. Needell D, Ward R. Near-optimal compressed sensing guarantees for total variation minimization. IEEE Trans on Image Processing. 2013; 22(10):3941–3949.
34. Needell D, Ward R. Stable image reconstruction using total variation minimization. SIAM Journal on Imaging Sciences. 2013; 6(2):1035–1058.
35. Krahmer F, Ward R. Stable and robust sampling strategies for compressive imaging. IEEE Trans on image processing. 2014; 23(2):612–622.
36. Poon C. On the role of total variation in compressed sensing. SIAM Journal on Imaging Sciences. 2015; 8(1):682–720.
37. Zhang T, Pauly JM, Vasanawala SS, Lustig M. Coil compression for accelerated imaging with cartesian sampling. Magnetic resonance in medicine. 2013; 69(2):571–582. [PubMed: 22488589]

38. Li T, Wang X. The BKK root count in c^n. Mathematics of Computation of the American Mathematical Society. 1996; 65(216):1477–1484.
39. Tropp J. User-friendly tail bounds for sums of random matrices. Foundations of computational Math. 2012; 12(4):389–434.

VIII. Appendix A: Incoherence Bounds

A. Notation and Preliminaries

To simplify our arguments, we will convert the linear operators $\mathscr{T}(\hat{f})$ and $\mathscr{T}(\hat{f})^*$ defined in Fourier domain to linear operators acting on spaces of trigonometric polynomials (3) in spatial domain. Specifically, for any index set $\Omega \subset \mathbb{Z}^2$, let B_Ω denote the vector space of all trigonometric polynomials that have coefficients supported within Ω. Similarly, we denote the space of vector fields $\boldsymbol{\rho} = (\rho_1, \rho_2)$ with components $\rho_1, \rho_2 \in B_\Omega$ as B_Ω^2. We set $\mathscr{S}(f) = \mathcal{F}\,\mathscr{T}(\hat{f})\,\mathcal{F}^{-1}$, where $\mathcal{F}$ is the Fourier transform of a periodic function on $[0, 1]^2$. For any index set Λ, define the Dirichlet kernel $D_{\Lambda_1}(\boldsymbol{r}) := \Sigma_{\boldsymbol{k}\in\Lambda_1} e^{j2\pi\boldsymbol{k}\cdot\boldsymbol{r}}$. For all $\varphi \in B_{\Lambda_1}$, the action of the linear operator $\mathscr{S}(f): B_{\Lambda_1} \to B_{\Lambda_2}^2$ can be expressed compactly as

$$\mathscr{S}(f)\varphi = D_{\Lambda_2} * (\varphi\, \nabla f) \in B_{\Lambda_1}^2, \quad (30)$$

where $\varphi\nabla f$ is understood as a tempered distribution, and the convolution is applied separately to each vector field component. Here convolution with D_{Λ_2} is a bandlimiting operation. Simliarly, for $\boldsymbol{\rho} = (\rho, \rho_2) \in B_{\Lambda_2}^2$, the adjoint $\mathscr{S}(f)^*$ acts as

$$\mathscr{S}(f)^*\boldsymbol{\rho} = D_{\Lambda_1} * (\boldsymbol{\rho} \cdot \nabla f) \in B_{\Lambda_1} \quad (31)$$

which is the spatial domain equivalent of the adjoint matrix $\mathscr{T}(\hat{f})^*$. More expliclty, if $f = 1_U$ where U is a simply connected region with smooth boundary ∂U, a straightforward argument using the divergence theorem shows that the function $\mathscr{S}(f)\boldsymbol{\varphi}$ is given pointwise as the weighted curve integral

$$(\mathscr{S}(f)\varphi)(\boldsymbol{r}) = \oint_{\partial U} D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}')\, \boldsymbol{n}(\boldsymbol{r}')\, ds(r'), \quad (32)$$

for all $\boldsymbol{r} \in [0, 1]^2$, where $\boldsymbol{n}(\boldsymbol{r}')$ is the outward unit normal to the curve ∂U at r', and ds is the arc-length element. Likewise, $\mathscr{S}(f)^*\boldsymbol{\rho}$ is the function given pointwise by

$$(\mathcal{S}(f)^{*}\boldsymbol{\rho})(\boldsymbol{r}) = \oint_{\partial U} D_{\Lambda_1}(\boldsymbol{r}-\boldsymbol{r}')\,[\boldsymbol{\rho}(\boldsymbol{r}')\cdot\boldsymbol{n}(\boldsymbol{r}')]\,ds(r'), \quad (33)$$

for all $\boldsymbol{r} \in [0, 1]^2$. These formulas can be generalized to an arbitrary piecewise constant function $f = \Sigma_i a_i 1_{U_i}$ by linearity. However, in the remainder we focus on the case where $f = 1_U$ to simplify our arguments.

B. Fundamental subspaces of $\mathcal{S}(f)$ and dimensions

Under the conditions of Theorem 1, the nullspace of $\mathcal{T}(\hat{f})$ is spanned by shifts of the minimal annihilating filter, $\widehat{\mu_0}$. In spatial domain, this space consists of all multiples of the minimal degree polynomial $\gamma = \eta\ \mu_0$ such that γ is bandlimited to Λ_1. We denote this space by

$$(\mu_0)_{\Lambda_1} := \{\eta\,\mu_0 : \eta \in B_{\Lambda_1:\Lambda_0}\}. \quad (34)$$

Note that $(\mu_0)_{\Lambda_1}$ is a subspace of B_{Λ_1} with dimension $|\Lambda_1:\Lambda_0|$. Therefore, the dimension of the kernel of $\mathcal{S}(f)$, denoted by $\ker \mathcal{S}(f)$, is given by

$$\dim\ \ker\ \mathcal{S}(f) = |\,\Lambda_1:\Lambda_0\,|\,. \quad (35)$$

By the rank-nullity theorem, the dimension of the image of $\mathcal{S}(f)$, denoted by $\operatorname{im} \mathcal{S}(f)$, is

$$\dim \operatorname{im} \mathcal{S}(f) = |\,\Lambda_1\,| - |\,\Lambda_1:\Lambda_0\,| = R. \quad (36)$$

Likewise, the dimension of the coimage $\operatorname{im} \mathcal{S}(f)^*$ is also R. Furthermore, since $\operatorname{im} \mathcal{S}(f)^* = [\ker \mathcal{S}(f)]^{\perp}$, we have

$$\operatorname{im} \mathcal{S}(f)^* = (\mu_0)^{\perp}_{\Lambda_1} \quad (37)$$

This means that any $\gamma \in B_{\Lambda_1}$ is in the row space if and only if γ is orthogonal to every trigonometric polynomial of the form $\eta\ \mu_0 \in B_{\Lambda_1}$, or equivalently,

$$\langle \gamma, \eta, \mu_0 \rangle = \int_{[0,1]^2} \gamma(\boldsymbol{r})\overline{\eta(\boldsymbol{r})\mu_0(\boldsymbol{r})}\,d\boldsymbol{r} = 0 \quad (38)$$

for all $\eta \in B_{\Lambda_1:\,\Lambda_0}$.

C. Basis for the coimage of $\mathscr{S}(\mathbf{f})$ (corresponding to the row space of $\mathscr{T}(\hat{\mathbf{f}})$)

Let $\boldsymbol{s} \in [0, 1]^2$, and set $\boldsymbol{\varphi_s} \in B_{\Lambda_1}$ to be the translated Dirichlet kernel:

$$\varphi_{\boldsymbol{s}}(\boldsymbol{r}) = D_{\Lambda_1}(\boldsymbol{r} - \boldsymbol{s}) \quad \text{for all} \quad \boldsymbol{r} \in [0, 1]^2. \qquad (39)$$

Equivalently, $\boldsymbol{\varphi_s} \in B_{\Lambda_1}$ is the trigonometric polynomial specified in Fourier domain as

$$\widehat{\varphi_{\boldsymbol{s}}}[\boldsymbol{k}] = \begin{cases} e^{-j2\pi \boldsymbol{s} \cdot \boldsymbol{k}} & \text{if } \boldsymbol{k} \in \Lambda_1 \\ 0 & \text{if } \boldsymbol{k} \notin \Lambda_1 \end{cases}. \qquad (40)$$

Observe that the inner product of $\boldsymbol{\varphi_s}$ with any other trigonometric polynomial $\eta \in B_{\Lambda_1}$ is given by the point-evaluation of η at $\boldsymbol{s}$:

$$\langle \eta, \varphi_{\boldsymbol{s}} \rangle = \sum_{\boldsymbol{k} \in \Lambda_1} \hat{\eta}[\boldsymbol{k}] e^{j2\pi k \cdot \boldsymbol{s}} = \eta(\boldsymbol{s}). \qquad (41)$$

Suppose now that the point $\boldsymbol{s}$ satisfies $\mu_0(\boldsymbol{s}) = 0$. In this case, we see that $\boldsymbol{\varphi_s}$ is necessarily in the coimage $\operatorname{im} \mathscr{S}(f)^* = (\mu_0)^{\perp}_{\Lambda_1}$ since we have

$$\langle \gamma \mu_0, \varphi_{\boldsymbol{s}} \rangle = \gamma(\boldsymbol{s}) \mu_0(\boldsymbol{s}) = 0. \qquad (42)$$

for any multiple of the minimal polynomial $\gamma\mu_0 \in B_{\Lambda_1}$, i.e., any element in $\ker \mathscr{S}(f) = (\mu_0)_{\Lambda_1}$.

We will now show how to construct a basis for the coimage of $\mathscr{S}(f)$ out of elements having the form $\boldsymbol{\varphi_{r_i}}$ for some $\boldsymbol{r}_i$, $i = 1, \ldots, R$ belonging to the zero set of μ_0. For an arbtirary collection of R points $\{\boldsymbol{r}_i\}_{i=1}^{R} \subset \{\mu_0 = 0\}$, we are not guaranteed that the set of functions $\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{R}$ is linearly independent. However, we will show that there exists a constant $M = M(\Lambda_0, \Lambda_1)$ such that *for any* M distinct points $\{r_i\}_{i=1}^{M} \subset \{\mu_0 = 0\}$ we can always find a subset of R linearly independent basis functions from the collection $\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{M}$. The constant M is related the maximum number of isolated zeros that a system of two trigonometric polynomials can have. The following lemma, which is a consequence of the BKK bound in

enumerative algebraic geometry (see, e.g., [38]), puts a bound on M. See section II of the supplementary material for proof.

Lemma 11

Let Λ_1 and Λ_0 be rectangular index sets such that $\Lambda_0 \subset \Lambda_1$, and set $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$. For any μ_0, μ_1 trigonometric polynomials bandlimited to Λ_0 and Λ_1, respectively, the maximum number M of isolated solutions of $\mu_0(\mathbf{r}) = \mu_1(\mathbf{r}) = 0$ is bounded as

$$M < R + |\Lambda_0| . \quad (43)$$

We now prove equivalents of Lemma 6 and Lemma 7 in terms of the spatial domain operator $\mathcal{S}(f)$:

Lemma 12

Let $\{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ be any collection of N distinct points on the curve $\{\mu_0 = 0\}$, where $N \geq R + |\Lambda_0|$. Then the coimage space $\mathrm{im}\, \mathcal{S}(f)^* = (\mu_0)^{\perp}_{\Lambda_1}$ is spanned by the set of shifted Dirichlet kernels $\varphi_i(\mathbf{r}) = D_{\Lambda_1}(\mathbf{r} - \mathbf{r}_i)$ for all $i = 1, \ldots, N$, i.e.,

$$span\{\varphi_{\mathbf{r}_i}\}_{i=1}^{N} = (\mu_0)^{\perp}_{\Lambda_1} . \quad (44)$$

In particular, there exists a subset of $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$ elements of $\{\varphi_{\mathbf{r}_i}\}_{i=1}^{N}$ that is a basis for the coimage space $(\mu_0)^{\perp}_{\Lambda_1}$.

Proof—All the functions $\varphi_{\mathbf{r}_i}$ are in $(\mu_0)^{\perp}_{\Lambda_1}$ since we have $\ll \varphi_i, \gamma\mu_0 \gg = \gamma(\mathbf{r}_i)\mu_0(\mathbf{r}_i) = 0$ because each $\mathbf{r}_i$ belong to the zero set of μ_0. This implies that

$$\mathrm{span}\{\varphi_{\mathbf{r}_i}\}_{i=1}^{M} \subseteq (\mu_0)^{\perp}_{\Lambda_1} . \quad (45)$$

Our focus is on proving (44) with equality. For this, it is sufficient to show that any vector orthogonal to $\mathrm{span}\{\varphi_i\}_{i=1}^{N}$ is in $(\mu_0)_{\Lambda}$. Assume that there is a vector $\eta(\mathbf{r}) \in B_{\Lambda_1}$ that is in the orthogonal complement space of $\mathrm{span}\{\varphi_i\}_{i=1}^{N}$. This is only possible if

$$\langle \eta, \varphi_i \rangle = \eta(\boldsymbol{r}_i) = 0, \quad \text{for all} \quad i = 1, \ldots, N. \quad (46)$$

Therefore, both η and μ_0 have N zeros in common. By Lemma 11 this is only possible if η contains μ_0 as a factor. This implies that all vectors in the orthogonal complement space of $\text{span}\{\varphi_i\}_{i=1}^{M}$ are in $(\mu_0)_{\Lambda_1}$, or equivalently

$$\text{span}\,\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{M} \supseteq (\mu_0)_{\Lambda_1}^{\perp}, \quad (47)$$

which together with (45) proves (44).

Finally, we also know that the dimension of $(\mu_0)_{\Lambda_1}^{\perp}$ is equal to $R < M$. Thus, one can select a subset of R basis functions φ_i that are linearly independent and hence a basis for $(\mu_0)_{\Lambda_1}^{\perp}$.

Translating this result to Fourier domain, we see that the row space of $\mathscr{T}(\hat{f})$ is spanned by the vectors of Fourier coefficients $(\widehat{\varphi_i}[\boldsymbol{k}]:\boldsymbol{k} \in \Lambda_1) \in \mathbb{C}^{|\Lambda_1|}$, for $i = 1, \ldots, R$. Equivalently, this can be expressed as the columns of the Vandermonde-like matrix $\boldsymbol{E}_{\text{row}}$ specified by (21), which proves Lemma 6 and 7.

D. Discretization of curve integrals: quadrature formula

Using the results from the previous subsection, we now introduce a quadrature formula for curve integrals, which we will use to determine the range space im $\mathscr{S}(f)$ in the next subsection.

Let γ be any function in B_Λ for any $\Lambda \supseteq \Lambda_0$. Then from the orthogonal decomposition $B_\Lambda = (\mu_0)_\Lambda \oplus (\mu_0)_\Lambda^{\perp}$ we can decompose γ as

$$\gamma(\boldsymbol{r}) = \sum_{i=1}^{S} a_i D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i) + \varphi(\boldsymbol{r})\mu_0(\boldsymbol{r}), \quad (48)$$

where $S = |\Lambda|-|\Lambda: \Lambda_0|$, and where $\{D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i)\}_{i=1}^{S}$ defines a basis of $(\mu_0)_\Lambda^{\perp}$. Here, the coefficients a_i in (48) are obtained uniquely as

$$\begin{bmatrix} a_1 \\ \vdots \\ a_S \end{bmatrix} = \boldsymbol{D}^{-1} \begin{bmatrix} \gamma(\boldsymbol{r}_1) \\ \vdots \\ \gamma(\boldsymbol{r}_S) \end{bmatrix}, \quad (49)$$

where $\boldsymbol{D} \in \mathbb{R}^{S \times S}$ is the symmetric matrix with entries $[\boldsymbol{D}]_{i,j} = D_\Lambda(\boldsymbol{r}_i - \boldsymbol{r}_j)$ for $1 \leq i, j \leq S$. The above expression can be compactly expressed as $\boldsymbol{a} = \boldsymbol{D}^{-1}\boldsymbol{g}$, where $\boldsymbol{g} = (\gamma(\boldsymbol{r}_1), \ldots, \gamma(\boldsymbol{r}_S))^T$.

Lemma 13

Let $f = 1_U$ where U is a simply connected region with smooth boundary ∂U, which is the zero levelset of $\mu_0 \in B_{\Lambda_0}$ and let $\gamma \in B_\Lambda$. Consider the curve integral of the form

$$\boldsymbol{q} = \oint_{\partial U} \gamma(\boldsymbol{r})\, \boldsymbol{n}(\boldsymbol{r})\, ds(\boldsymbol{r}), \quad (50)$$

where $\mathbf{n}(\mathbf{r}) = \nabla f(\mathbf{r})/|\nabla f(\mathbf{r})|$ is the unit normal on the curve ∂U. The curve integral can be evaluated using the quadrature formula

$$\boldsymbol{q} = \sum_{i=1}^{S} \gamma(\boldsymbol{r}_i)\, \boldsymbol{w}_i, \quad (51)$$

where the $S = |\Lambda|/|\Lambda : \Lambda_0|$ points $\{r_i\}_{i=1}^S$ belong to the curve $\{\mu_0 = 0\}$, and the cooresponding weight vectors $\mathbf{w}_i \in \mathbb{R}^2$, $i = 1, .., S$, are specified by

$$\begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_S \end{bmatrix} = \boldsymbol{D}^{-1} \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_S \end{bmatrix}. \quad (52)$$

where $\mathbf{v}_i = \oint_{\partial U} D_\Lambda(\mathbf{r} - \mathbf{r}_i)\mathbf{n}(\mathbf{r})ds(\mathbf{r}) \in \mathbb{R}^2$.

Proof—Decomposing $\gamma(r)$ using (48), we obtain

$$\oint_{\partial U} \gamma(\boldsymbol{r})\, \boldsymbol{n}(\boldsymbol{r})\, ds(\boldsymbol{r}) = \sum_{i=1}^{S} a_i \underbrace{\oint_{\partial U} D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{n}(\boldsymbol{r})\, ds(\boldsymbol{r})}_{:= \boldsymbol{v}_i} \quad (53)$$

The above sum can be expressed in the vector form as

$$\sum_{i=1}^{S} a_i \boldsymbol{v}_i = \boldsymbol{a}^* \boldsymbol{V} = \boldsymbol{g}^* \boldsymbol{D}^{-1} \boldsymbol{V} \quad (54)$$

where $\boldsymbol{V} = [\boldsymbol{v}_1^T, \ldots, \boldsymbol{v}_S^T]^T \in \mathbb{C}^{R \times 2}$. Setting $\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{V} = [\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_S^T]^T \in \mathbb{C}^{R \times 2}$ we obtain (51).

E. Basis for the range of $\mathcal{S}(\mathbf{f})$ (corresponding to the column space of $\mathcal{T}(\hat{\mathbf{f}})$)

We now introduce a basis set for im $\mathcal{S}(f)$, which will be used to prove Lemma 8.

Lemma 14

The range of $\mathcal{S}(f)$*, denoted by* im $\mathcal{S}(f)$ *is specified by*

$$\text{im } \mathcal{S}(f) = span\{\boldsymbol{w}_i D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\}_{i=1}^{R} \quad (55)$$

for an appropriate choice of points $\{\boldsymbol{r}_i\}_{i=1}^{R} \subset \{\mu_0 = 0\}$ *with* $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$*, and where the weight vectors* $\boldsymbol{w}_i$ *are specified by* (52).

Proof—Consider an arbitrary element $\boldsymbol{\rho} = (\rho, \rho_2) \in$ im $\mathcal{S}(f)$. We can express $\boldsymbol{\rho}$ as $\boldsymbol{\rho} = \mathcal{S}(f)\psi = \mathcal{B}_{\Lambda_2}(\psi \nabla f) = D_{\Lambda_2} * (\psi \nabla f)$ for some $\psi \in B_{\Lambda_1}$. By the definition in (33), we have

$$\begin{aligned} \boldsymbol{\rho}(\boldsymbol{r}) &= \oint_{\partial U} \psi(\boldsymbol{s})\, D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{s})\, \boldsymbol{n}(\boldsymbol{s})\, ds \quad (56) \\ &= \sum_{i=1}^{S} \psi(\boldsymbol{r}_i)\, D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{w}_i, \end{aligned}$$

where we Lemma 13 in the last step with $S = |\Gamma| - |\Gamma : \Lambda_0|$ since the integrand $\psi(\boldsymbol{s})^* D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{s})$ belongs to B_Γ. The above relation shows that any $\boldsymbol{\rho}(\boldsymbol{s}) \in$ im $\mathcal{S}(f)$ can be expressed as the linear combination of the functions $D_{\Lambda_2}(\boldsymbol{s} - \boldsymbol{r}_i)\boldsymbol{w}_i$, for $i = 1, .., S$. Thus, we have im $\mathcal{S}(f) \subset \text{span}\{D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{w}_i\}_{i=1}^{S}$. We also know that dim (im $\mathcal{S}(f)$) $= R < S$. This implies that we can select a subset of R vectors from the set $\{D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\boldsymbol{w}_i\}_{i=1}^{S}$ that are linearly independent, which will span im $\mathcal{S}(f)$, and hence define a basis.

Correspondingly, the column space of $\mathcal{T}(\hat{f})$ is spanned by the Fourier coefficients of the basis vectors $\boldsymbol{w}_i D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)$, or the columns of the $2|\Lambda_2| \times R$ weighted Vandermonde-like matrix $\boldsymbol{E}_{\text{col}}$ specified by (22).

F. Incoherence Bounds

1) Projection onto row subspace

Let $\boldsymbol{E}_{\text{row}} = \boldsymbol{E}_{\text{row}}(P)$ be any basis for the row space V of $\mathscr{T}(\hat{f})$ specified by (21), whose columns are vectorized Fourier coefficients of the translated and normalized Dirichlet kernels $\varphi_i(\boldsymbol{r}) = \frac{1}{\sqrt{|\Lambda_1|}} D_{\Lambda_1}(\boldsymbol{r} - \boldsymbol{r}_i)$, $i = 1, \ldots, R$, for some set of admissible nodes $P = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_R\} \subset \{\mu_0 = 0\}$. Projecting the measurement basis matrix $\boldsymbol{A}_{\boldsymbol{k}}$ onto V, we have

$$\begin{aligned}\|\mathscr{P}_V \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 &= \|\boldsymbol{A}_{\boldsymbol{k}} \boldsymbol{E}_{\text{row}} (\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}})^{-1} \boldsymbol{E}_{\text{row}}^*\|_F^2 \\ &\leq [\lambda_{min}(\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}})]^{-1} \|\boldsymbol{A}_{\boldsymbol{k}} \boldsymbol{E}_{\text{row}}\|_F^2\end{aligned}$$

Since $\boldsymbol{A}_{\boldsymbol{k}}$ selects $|\omega(\boldsymbol{k})|$ rows of $\boldsymbol{E}_{\text{row}}$, each of which has R entries of magnitude $1/\sqrt{|\Lambda_1|}$, we have

$$\|\boldsymbol{A}_{\boldsymbol{k}} \boldsymbol{E}_{\text{row}}\|_F^2 = \frac{1}{|\omega(\boldsymbol{k})|} \cdot R \cdot |\omega(\boldsymbol{k})| \cdot \frac{1}{|\Lambda_1|} = \frac{R}{|\Lambda_1|} = \frac{R c_s}{|\Gamma|} \quad (57)$$

where $c_s = |\Gamma| / |\Lambda_1|$. Hence,

$$\|\mathscr{P}_V \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 \leq [\lambda_{min}(\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}})]^{-1} \frac{R c_s}{|\Gamma|}. \quad (58)$$

Minimizing over all sets of admissible nodes P in the construction of $\boldsymbol{E}_{\text{row}}$ gives the final bound

$$\|\mathscr{P}_V \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 \leq \frac{\rho R c_s}{|\Gamma|}. \quad (59)$$

2) Projection onto column space

Let $\boldsymbol{E}_{\text{col}} = \boldsymbol{E}_{\text{col}}(P)$ be a basis for the column space of $\mathscr{T}(\hat{f})$ specified by (22), whose columns are vectorized Fourier coefficients of the translated and weighted Dirichlet kernels $\frac{1}{\sqrt{|\Lambda_2|}} \frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|} D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)$, for some set of admissible nodes $P = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_R\} \subset \{\mu_0 = 0\}$. Observe the columns of $\boldsymbol{E}_{\text{col}}$ are defined to have unit ℓ^2-norm. Following the same steps as in the row space bound, we have

$$\|\mathscr{P}_U \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 = \|\boldsymbol{E}_{\text{col}}(\boldsymbol{E}_{\text{col}}^* \boldsymbol{E}_{\text{col}})^{-1} \boldsymbol{E}_{\text{col}}^* \boldsymbol{A}_{\boldsymbol{k}}\|_F^2$$
$$\leq [\lambda_{min}(\boldsymbol{E}_{\text{col}}^* \boldsymbol{E}_{\text{col}})]^{-1} \|\boldsymbol{E}_{\text{col}}^* \boldsymbol{A}_{\boldsymbol{k}}\|_F^2$$

Expanding the norm $\|\boldsymbol{E}_{\text{col}}^* \boldsymbol{A}_{\boldsymbol{k}}\|_F^2$ gives

$$\|\boldsymbol{E}_{\text{col}}^* \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 = \frac{1}{|\Lambda_2|} \sum_{i=1}^{R} \frac{1}{|\omega(\boldsymbol{k})|} \sum_{\boldsymbol{\ell} \in \omega(\boldsymbol{k})} \left| \left\langle \frac{\boldsymbol{\ell}}{\|\boldsymbol{\ell}\|}, \frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|} \right\rangle \right|^2$$
$$\leq \frac{R}{|\Lambda_2|} \leq \frac{R c_s}{|\Gamma|}.$$

Hence, we have

$$\|\mathscr{P}_U \boldsymbol{A}_{\boldsymbol{k}}\|_F^2 \leq \frac{\rho' R c_s}{|\Gamma|}. \quad (60)$$

where ρ' is defined similarly to ρ as:

$$\rho' = \min_{\substack{P \subset \{\mu_0 = 0\} \\ |P| = R}} \frac{1}{\lambda_{min}[\boldsymbol{E}_{\text{col}}(P) * \boldsymbol{E}_{\text{col}}(P)]}, \quad (61)$$

Finally, we show how to bound ρ' by ρ in (60). Observe that we can re-define ρ and ρ' in terms of the minimum singular value of the basis matrices $\boldsymbol{E}_{\text{row}}(P)$ and $\boldsymbol{E}_{\text{col}}(P)$, according to the correspondences:

$$\lambda_{min}(\boldsymbol{E}_{\text{col}}(P)^* \boldsymbol{E}_{\text{col}}(P)) = \sigma_{min}^2(\boldsymbol{E}_{\text{col}}(P)),$$
$$\lambda_{min}(\boldsymbol{E}_{\text{row}}(P)^* \boldsymbol{E}_{\text{row}}(P)) = \sigma_{min}^2(\boldsymbol{E}_{\text{row}}(P)).$$

We will show $\sigma_{min}^2(\boldsymbol{E}_{\text{row}}(P)) \leq \sigma_{min}^2(\boldsymbol{E}_{\text{col}}(P))$, or equivalently, $[\lambda_{min}(\boldsymbol{E}_{\text{col}}(P)^* \boldsymbol{E}_{\text{col}}(P))]^{-1} \leq [\lambda_{min}(\boldsymbol{E}_{\text{row}}(P)^* \boldsymbol{E}_{\text{row}}(P)]^{-1}$, for any set P consisting of R points on the edge set. The claim then follows immediately by taking the minimum over all such sets P.

To ease notation, we drop the dependence on the set P in the following. Observe that we can express $\boldsymbol{E}_{\text{col}}$ as

$$\boldsymbol{E}_{\text{col}} = \begin{bmatrix} \widetilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_x \\ \widetilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_y \end{bmatrix} \quad (62)$$

where $\boldsymbol{W}_x = \text{diag}(\frac{w_{1,x}}{\|\boldsymbol{w}_1\|}, \ldots, \frac{w_{R,x}}{\|\boldsymbol{w}_R\|})$, $\boldsymbol{W}_y = \text{diag}(\frac{w_{1,y}}{\|\boldsymbol{w}_1\|}, \ldots, \frac{w_{R,y}}{\|\boldsymbol{w}_R\|})$, and $\widetilde{\boldsymbol{E}}_{\text{col}} \in \mathbb{C}^{|\Lambda_2| \times R}$ is the Vandermonde-like matrix given entrywise by $[\widetilde{\boldsymbol{E}}_{\text{col}}]_{i,j} = e^{\imath 2\pi \boldsymbol{k}_i^T \boldsymbol{r}_j}$, for all $\boldsymbol{k}_i \in \Lambda_2$, $1 \le j \le R$. In other words, $\widetilde{\boldsymbol{E}}_{\text{col}}$ has the same structure as $\boldsymbol{E}_{\text{row}}$, but is built with respect to Λ_2 instead of Λ_1. In particular, since we always assume $\Lambda_1 \subset \Lambda_2$, the matrix $\boldsymbol{E}_{\text{row}}$ can be embedded as a submatrix of $\widetilde{\boldsymbol{E}}_{\text{col}}$ by restricting the rows of $\widetilde{\boldsymbol{E}}_{\text{col}}$ to those indexed by Λ_1. By the variational characterization of the minimum singular value of a matrix, we have

$$\begin{aligned} \sigma_{min}^2(\boldsymbol{E}_{\text{col}}) &= \min_{\|\boldsymbol{u}\|=1} \|\boldsymbol{E}_{\text{col}} \boldsymbol{u}\|^2 \quad (63) \\ &= \min_{\|\boldsymbol{u}\|=1} \|\widetilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_x \boldsymbol{u}\|^2 + \|\widetilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_y \boldsymbol{u}\|^2 \\ &\ge \sigma_{min}^2(\widetilde{\boldsymbol{E}}_{\text{col}}) \underbrace{(\|\boldsymbol{W}_x \boldsymbol{u}\|^2 + \|\boldsymbol{W}_y \boldsymbol{u}\|^2)}_{=1} \end{aligned}$$

Finally, since $\boldsymbol{E}_{\text{row}}$ is a submatrix of $\widetilde{\boldsymbol{E}}_{\text{col}}$, we also have $\sigma_{min}^2(\boldsymbol{E}_{\text{row}}) \le \sigma_{min}^2(\widetilde{\boldsymbol{E}}_{\text{col}})$, which together with (63) gives the desired inequality.

IX. Appendix B: Proof of Main Theorem

A. Reformulation in lifted domain

We now reformulate the recovery of $\hat{f}$ as a matrix recovery problem in the lifted domain. The matrices $\mathscr{T}_x(\hat{f})$ and $\mathscr{T}_y(\hat{f})$ contain several copies of the weighted entries $k_x \hat{f}[\boldsymbol{k}]$ and $k_y \hat{f}[\boldsymbol{k}]$, respectively. We use $\omega(\boldsymbol{k})$ to denote the set of locations (α_1, α_2) in the matrix $\mathscr{T}_x(\hat{f})$ or $\mathscr{T}_y(\hat{f})$ that contain the entry $k_x \hat{f}[\boldsymbol{k}]$ or $k_y \hat{f}[\boldsymbol{k}]$ (this set is the same in either case).

We define the sampling matrices $\boldsymbol{A}_{\boldsymbol{k}} = \begin{bmatrix} \boldsymbol{A}_{1,\boldsymbol{k}} \\ \boldsymbol{A}_{2,\boldsymbol{k}} \end{bmatrix} \in \mathbb{C}^{2|\Lambda_2| \times |\Lambda_1|}$, for each $\boldsymbol{k} = (k_1, k_2) \in \Gamma$, where

$$(\boldsymbol{A}_{i,\boldsymbol{k}})_{\boldsymbol{\alpha}} = \begin{cases} \frac{k_i}{\|\boldsymbol{k}\| \sqrt{|\omega_i(\boldsymbol{k})|}}, & \text{if} \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \omega(\boldsymbol{k}) \\ 0 & \text{else} \end{cases} \quad (64)$$

for $i = 1, 2$. The matrices $\{\boldsymbol{A}_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\Gamma}$ form an orthonormal basis for the space of matrices defined by the range of the matrix lifting $\mathscr{T}$; we will call any matrix in the range of $\mathscr{T}$ a *structured matrix*. For any set of coefficients $\{\hat{g}[\boldsymbol{k}]\}_{\boldsymbol{k}\in\Gamma}$ we can expand the structured matrix $\mathscr{T}(\hat{g})$ as

$$\mathscr{T}(\hat{g}) = \sum_{\boldsymbol{k}\in\Gamma} \hat{g}[\boldsymbol{k}]\, \|k\| \sqrt{|\, w_i(k)\,|}\, \boldsymbol{A}_{\boldsymbol{k}}. \quad (65)$$

We denote the projection operator corresponding to a single sampling location $\boldsymbol{k}$ by $\mathscr{A}_{\boldsymbol{k}}(\boldsymbol{X}) = \langle \boldsymbol{A}_{\boldsymbol{k}}, \boldsymbol{X}\rangle \boldsymbol{A}_{\boldsymbol{k}}$. Since $\{\boldsymbol{A}_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\Gamma}$ is an orthonormal basis, for any structured matrix $\boldsymbol{X}$, we have $\Sigma_{k\in\Gamma}\, \mathscr{A}_{\boldsymbol{k}}(\boldsymbol{X}) = \mathscr{A}(\boldsymbol{X}) = \boldsymbol{X}$. Since $\boldsymbol{A}_{\boldsymbol{k}}$ is not the basis for a general $\boldsymbol{X} \in \mathbb{C}^{2/\Lambda_2/\times/\Lambda_1/}$, we also define the projection operator to the space orthogonal to the space of structured matrices by $\mathscr{A}_\perp(\boldsymbol{X}) = (\mathscr{I} - \mathscr{A})(\boldsymbol{X})$, where $\mathscr{I}$ is the identity operator. In particular, the constraint $\mathscr{A}_\perp(\boldsymbol{X}) = \boldsymbol{0}$ implies that $\boldsymbol{X}$ is a structured matrix.

The recovery of f from its partial Fourier samples $\hat{f}[\boldsymbol{k}]$, $\boldsymbol{k} \in \Omega$, can thus be reformulated as the completion of a structured matrix $\boldsymbol{X}$ from its measurements $\mathscr{A}_{\boldsymbol{k}}$, $\boldsymbol{k} \in \Omega$. Since the matrix is structured, we have $\mathscr{A}_\perp(\boldsymbol{X}) = \boldsymbol{0}$. We thus reformulate (10) as the structured low-rank recovery problem:

$$\text{minimize}_{\boldsymbol{X}} \|\boldsymbol{X}\|_* \text{ subject to } \mathscr{Q}_\Omega(\boldsymbol{X}) = \mathscr{Q}_\Omega(\mathscr{T}(\hat{f})), \quad (66)$$

where $\mathscr{Q}_\Omega$ that satisifies $\mathbb{E}[\mathscr{Q}_\Omega] = \mathscr{I}$ is defined as:

$$\mathscr{Q}_\Omega = \frac{|\,\Gamma\,|}{|\,\Omega\,|}\mathscr{A}_\Omega + \mathscr{A}^\perp \quad (67)$$

B. Conditions for perfect recovery

The tangent space T of the matrix $\mathbf{X}$ is defined as $T := \{\mathbf{U}\mathbf{X}_1^H + \mathbf{X}_2\mathbf{V}^H : \mathbf{X}_1 \in \mathbb{C}^{|\,\Lambda_2\,|\times R}, \mathbf{X}_2 \in \mathbb{C}^{|\,\Lambda_1\,|\times R}\}$ where $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ is the singular value decomposition of $\boldsymbol{X}$. The orthogonal complement of T is denoted by $T^\perp$. We first show that if $\mathscr{P}_T \approx \mathscr{P}_T\mathscr{Q}_\Omega\mathscr{P}_T$, and if an approximate dual certificate that satisfies certain conditions exist, we obtain perfect recovery.

Lemma 15

Consider a multiset Ω *that contains* m *random indices. Suppose the sampling operator* $\mathscr{Q}_\Omega$ *obeys*

$$\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T\| \leq \frac{1}{2} \quad (68)$$

and there exists a dual certificate matrix $\boldsymbol{W}$ *satisfying*

$$\mathcal{Q}_\Omega^\perp(\boldsymbol{W}) = 0 \quad (69)$$

$$\|\mathcal{P}_T(\boldsymbol{W} - \boldsymbol{U}\boldsymbol{V}^*)\|_F \leq \frac{1}{6n} \quad (70)$$

$$\|\mathcal{P}_T^\perp(\boldsymbol{W})\| \leq \frac{1}{2}. \quad (71)$$

Then, $\mathcal{T}(\hat{f})$ *is the unique solution to* (66)*, where* $n = |\Gamma|$ *and* $m = |\Omega|$.

See Section III-A of supplementary material for proof. Equation (68) suggests that $\mathcal{Q}_\Omega \approx \mathcal{I}$ on the tangent space. The conditions (69), (70), and (71) indicates the existence of a $\boldsymbol{W}$, which approximates the exact dual certificate $\boldsymbol{U}\boldsymbol{V}^*$. The above lemma is in line with [3, lemma 1], with the exception of the third condition, indicated by (70). To satisfy (68), we bound the deviation of $\mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T$ from $\mathcal{P}_T$ in the following lemma.

Lemma 16

Suppose (12) *holds. Then we have*

$$\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T\| \leq \varepsilon \leq \frac{1}{2} \quad (72)$$

with probability exceeding $1 - n^{-4}$*, provided that* $m > c_1\rho R\, c_s \log(n)$.

We prove this using [39, Theorem 1.6]. (See Section III-B of supplementary material)

C. Construction of the approximate dual certificate W

We will now use the golfing scheme of [3], [25] to construct an approximate dual certificate $\boldsymbol{W}$, which satisfies (69), (70), and (71). In particular, we generate j_0 independent random sampling sets Ω_i; $1 \leq i \leq j_0$, each containing $\tilde{m} = m/j_0$ samples corresponding to sampling with replacement. We start with $\boldsymbol{F}_0 = \boldsymbol{U}\boldsymbol{V}^*$, and follow the following steps:

1. $\boldsymbol{F}_0 = \boldsymbol{U}\boldsymbol{V}^*$ and set $j_0 - 3\log_{\frac{1}{\varepsilon}} n$.
2. $\forall i (1 \le i \le j_0)$, $\boldsymbol{F}_i = \mathcal{P}_T(\mathcal{I} - \mathcal{Q}_{\Omega_i})\mathcal{P}_T(\boldsymbol{F}_{i-1})$
3. $\boldsymbol{W} = \sum_{j=1}^{j_0} \mathcal{Q}_{\Omega_i} \boldsymbol{F}_{j-1}$

Step 3 ensures that $\boldsymbol{W}$ satisfies (69) since each term $\boldsymbol{W}_i = \mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}$ satisfies $\mathcal{Q}_{\Omega}^{\perp}(\boldsymbol{W}_i) = 0$. The recursive construction also satisfies (70). In particular,

$$\begin{aligned}\|\mathcal{P}_T(\boldsymbol{W} - \boldsymbol{U}\boldsymbol{V}^*)\|_F &= \|\mathcal{P}_T \boldsymbol{F}_{j_0}\|_F \\ &\le \varepsilon^{j_0}\|\boldsymbol{F}_0\|_F = \varepsilon^{j_0}\sqrt{R} \le \varepsilon^{j_0} n\end{aligned}$$

Now we focus on showing that $\boldsymbol{W}$ satisfies (71). Note that if j_0 is chosen as $3\log_{\frac{1}{\varepsilon}} n$, assuming $n > 6$, we have $(\varepsilon)^{j_0} n < \frac{1}{6n}$.

Lemma 17

For any matrix $\boldsymbol{M}$*, there exists some numerical constant* c_2 *such that*

$$\|(\mathcal{I} - \mathcal{Q}_{\Omega})(\boldsymbol{M})\| \le c_2\sqrt{\frac{n\log n}{m}}\|\boldsymbol{M}\|_{\mathcal{A},2} + \frac{c_2 n\log n}{m}\|\boldsymbol{M}\|_{\mathcal{A},\infty}, \quad (73)$$

with probability at least $1 - n^{-10}$*. Here,*

$$\|\boldsymbol{M}\|_{\mathcal{A},\infty} = \max_{k\in\Gamma}\left|\frac{\langle \mathbf{A}_k, \mathbf{M}\rangle}{|\omega_k|}\right| \quad (74)$$

$$\|\boldsymbol{M}\|_{\mathcal{A},2} = \sqrt{\sum_{k\in\Gamma}\frac{|\langle \mathbf{A}_k, \mathbf{M}\rangle|^2}{|\omega_k|}} \quad (75)$$

See Section III-C of supplementary material for proof.

Lemma 18

Assume that there exists a constant μ_5 *such that* $\omega_{\boldsymbol{k}}\|\mathscr{P}_T(\boldsymbol{A}_{\boldsymbol{k}})\|_{\mathscr{A},2} \leq \frac{\mu_5 R}{n}$. *For any matrix* $\boldsymbol{M}$, *we have*

$$\|\mathscr{P}_T[(\mathscr{I}-\mathscr{Q}_\Omega)(\boldsymbol{M})]\|_{\mathscr{A},2} \leq c_3\sqrt{\frac{\mu_5 R\log n}{m}}\left(\|\boldsymbol{M}\|_{\mathscr{A},2}+\sqrt{\frac{n\log n}{m}}\|\boldsymbol{M}\|_{\mathscr{A},\infty}\right),$$

with probability at least $1-n^{-10}$.

See Section III-D of the Supplementary Materials for proof.

Lemma 19

For any matrix $\boldsymbol{M}\in T$, *there exists some numerical constant* c_4, *such that*

$$\|\mathscr{P}_T[(\mathscr{I}-\mathscr{Q}_\Omega)(\boldsymbol{M})]\|_{\mathscr{A},\infty} \leq c_4\sqrt{\frac{\rho c_s R\log n}{m}}\sqrt{\frac{\rho c_s R}{n}}\|\boldsymbol{M}\|_{\mathscr{A},2}+\frac{c_4\rho c_s R\log n}{m}\|\boldsymbol{M}\|_{\mathscr{A},\infty}, \quad (76)$$

with probability at least $1-n^{-10}$.

See Section III-F of the Supplementary Materials for proof. From the golfing scheme, we have $\|\mathscr{P}_{T^\perp}(\boldsymbol{W})\| \leq \sum_{j=1}^{j_0}\|\mathscr{P}_{T^\perp}\mathscr{Q}_{\Omega_i}\mathscr{P}_T\boldsymbol{F}_{j-1}\|$. Using lemma 17 and substituting from lemma 18 and lemma 19, we have

$$\|\mathscr{P}_{T^\perp}\mathscr{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| \leq \left(\frac{1}{2}\right)^{j_0-1} c_2\left\{\sqrt{\frac{n\log n}{\tilde{m}}}\|\boldsymbol{F}_0\|_{\mathscr{A},2}+\frac{n\log n}{\tilde{m}}\|\boldsymbol{F}_0\|_{\mathscr{A},\infty}\right\}$$

The last inequality holds if $\tilde{m}=m/j_0 \gg \max(\mu_5,\rho c_s)R\log n$. Substituting for $j_0=3\log_{\frac{1}{\varepsilon}}(n)$ assumed in the golfing scheme, we require $m \gg c_6\max(\mu_5,\rho c_s)R\log^2 n$ to satisfy the above inquality. See Section III-G of the Supplementary Materials for details. We will now present the lemmas bounding $\|\boldsymbol{F}_0\|_{\mathscr{A},2}$ and $\|\boldsymbol{F}_0\|_{\mathscr{A},\infty}$, where $\boldsymbol{F}_0=\boldsymbol{U}\boldsymbol{V}^*$.

Lemma 20

With the incoherence measure ρ, one can bound

$$\|\boldsymbol{U}\boldsymbol{V}^*\|_{\mathscr{A},\infty} \leq \frac{\rho c_s R}{n} \quad (77)$$

$$\|\boldsymbol{U}\boldsymbol{V}^*\|_{\mathscr{A},2}^2 \leq \frac{c_7\mu_3 c_s \log^2(n)R}{n} \quad (78)$$

$$\|\mathscr{P}_T(\sqrt{\omega_\alpha}\mathbf{A}_\alpha)\|_{\mathscr{A},2}^2 \leq \frac{c_7\mu_3 c_s \log^2(n)R}{n}, \forall \boldsymbol{\alpha} \in \Gamma \quad (79)$$

for $\mu_3 = 3\rho$ *and* c_7 *is some constant.*

See Section III-H of the Supplementary Materials for proof. From (79), we see that the constant μ_5 in lemma 19 can be chosen as $\mu_5 = c_7\,\mu_3\,c_s \log^2(n)$ such that $\omega_{\boldsymbol{k}}\|\mathscr{P}_T(\boldsymbol{A}_{\boldsymbol{k}})\|_{\mathscr{A},2} \leq \frac{\mu_5 R}{n}$. Substituting for μ_5, we observe that the dominant term has its dependence on $\log^4(n)$. Thus, $\|\mathscr{P}_{T^\perp}\mathscr{D}_\Omega \boldsymbol{F}_{j-1}\| < 1/2$ if $m > c_6 c_7\, c_s\,(3\rho)\, R \log_4(n)$.

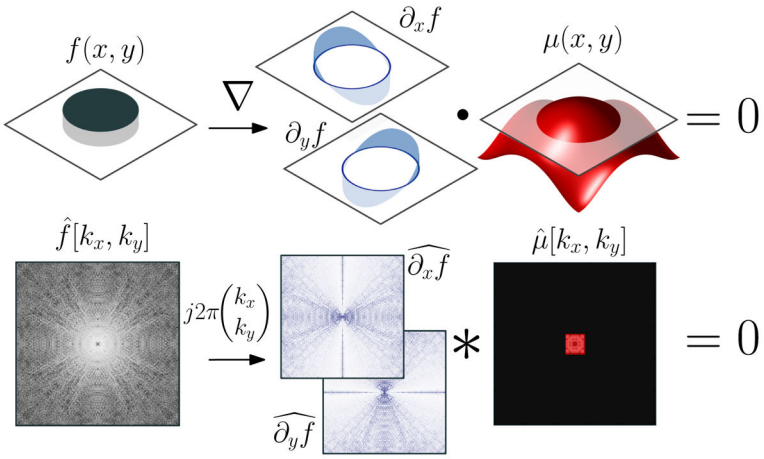


Fig. 1.

Annihilation of a piecewise constant function as a multiplication in spatial domain (top) and as a convolution in Fourier domain (bottom). The partial derivatives of a piecewise constant function are supported on the edge set. If there is a bandlimited function μ that is zero along the edge set, then the spatial domain product of μ with the gradient $\nabla f = (\partial_x f, \partial_y f)$ is identically zero. In Fourier domain, this is equivalent to the annihilation of the arrays $j2\pi k_x \hat{f}[k_x, k_y]$ and $j2\pi k_y \hat{f}[k_x, k_y]$ by 2-D convolution with a finite filter determined by the Fourier coefficients $\hat{\mu}$.

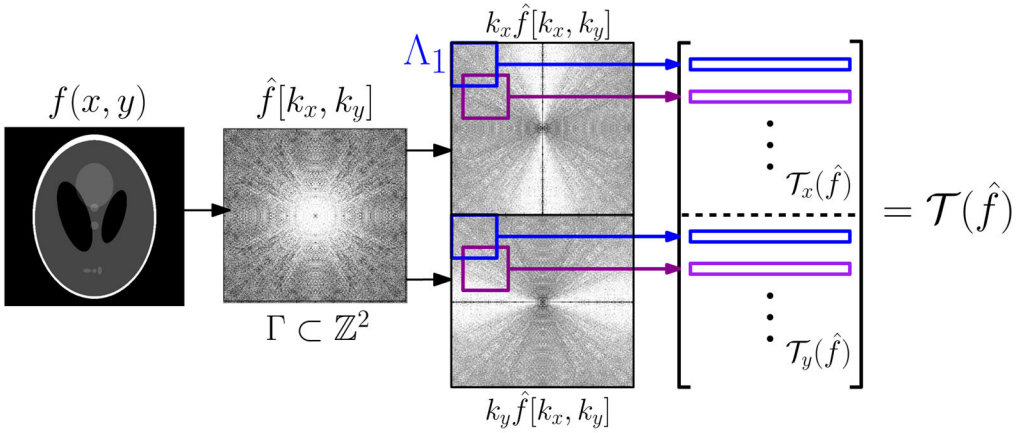


Fig. 2.

Construction of the structured matrix lifting $\mathcal{T}(\hat{f})$ considered in this work. From a rectangular array of the Fourier coefficients $\hat{f}[k_x, k_y]$ of a continuous domain image $f(x, y)$, the weighted arrays $k_x \hat{f}[k_x, k_y]$ and $k_y \hat{f}[k_x, k_y]$ are constructed. The matrices $\mathcal{T}_x(\hat{f})$ and $\mathcal{T}_y(\hat{f})$ are then obtained by extracting all vectorized patches from the weighted arrays, and loading these into the rows of $\mathcal{T}_x(\hat{f})$ and $\mathcal{T}_y(\hat{f})$. The resulting matrices $\mathcal{T}_x(\hat{f})$ and $\mathcal{T}_y(\hat{f})$ have a block Toeplitz with Toeplitz blocks structure. Finally $\mathcal{T}(\hat{f})$ is formed by vertically concatenating the blocks $\mathcal{T}_x(\hat{f})$ and $\mathcal{T}_y(\hat{f})$.

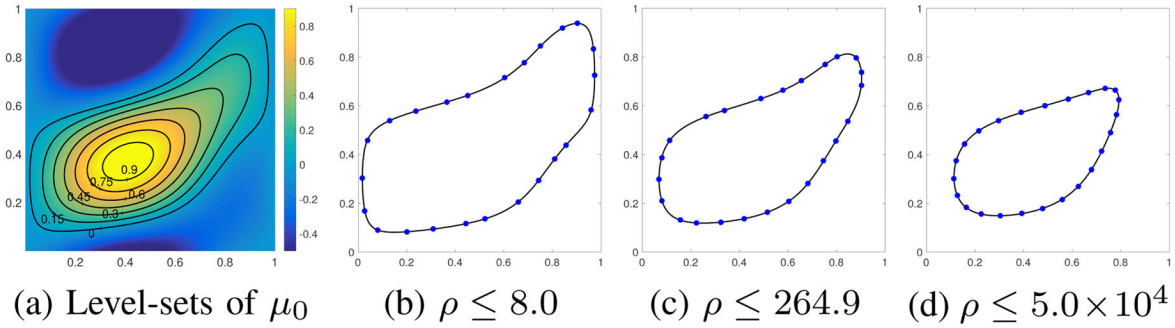
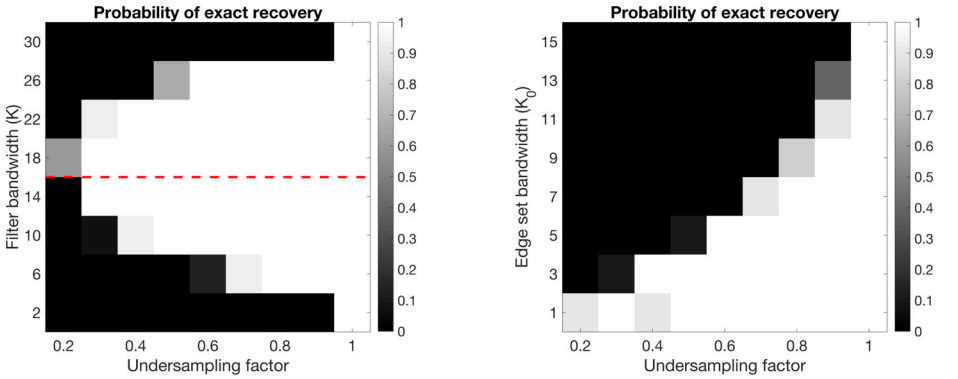
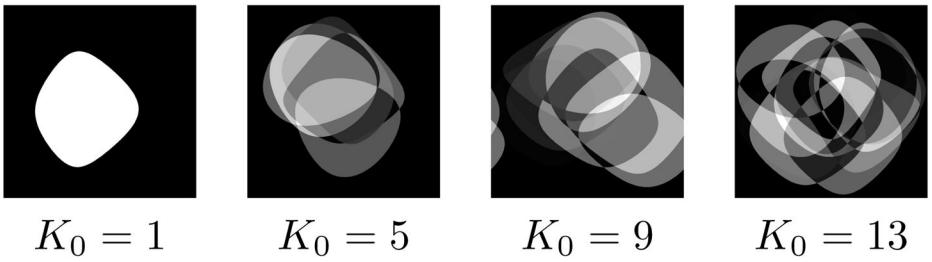


Fig. 3.

Illustration of edge set incoherence measure ρ. In (a) are the level-sets of trigonometric polynomial μ_0 bandlimited to Λ_0 of size 3×3. These curves all have the same bandwidth, Λ_0, but come in different sizes. In (b)–(d) we show $R = 24$ nodes on the curve giving the indicated bound on incoherence parameter ρ defined in (24), assuming a filter Λ_1 of size 7×7. Observe that the incoherence measure increases as the curve gets smaller. This indicates the smaller curves have a significant sampling burden.



(a) Varying filter bandwidth (b) Varying edge set bandwidth



(c) Examples of randomly generated piecewise constant images

Fig. 4.

Phase transition experiments. We generated random piecewise constant images with known edge-set bandwidth and study the success rate proposed structured low-rank matrix completion scheme under two conditions: in (a) we vary the filter size Λ_1 while keeping the edge-set bandwidth K_0 fixed, in (b) we vary the edge-set bandwidth K while keeping the filter size fixed. Examples of the randomly generated data are shown in (c).

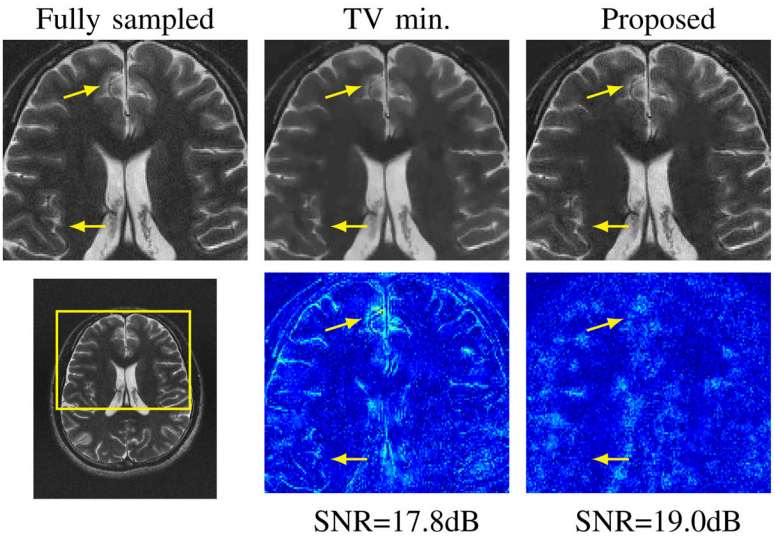


Fig. 5.

Recovery of MRI data from 2-fold random uniform undersampling. Error images shown in bottom right.

TABLE I

Comparison of proposed scheme with discrete total variation minimization

	TV-minimization	**Proposed**
Spatial domain	discrete	continuous
Derivative operator	finite differences	exact derivative
Singularity set	discrete points	connected curves
Frequency domain	discrete	discrete
Frequency weighting $w_i[\boldsymbol{k}]$	$1 - e^{j2\pi k_i/N_i}$	$j2\pi k_i$
Lifted matrix structure	two-level circulant	two-level Toeplitz
Rank of lifted matrix	sparsity of discrete gradient	bandwidth of edge set