Off-the-Grid Recovery of Piecewise Constant Images from Few Fourier Samples[*]

Greg Ongie[†] and Mathews Jacob[‡]

April 20, 2016

Abstract

We introduce a method to recover a continuous domain representation of a piecewise constant two-dimensional image from few low-pass Fourier samples. Assuming the edge set of the image is localized to the zero set of a trigonometric polynomial, we show the Fourier coefficients of the partial derivatives of the image satisfy a linear annihilation relation. We present necessary and sufficient conditions for unique recovery of the image from finite low-pass Fourier samples using the annihilation relation. We also propose a practical two-stage recovery algorithm which is robust to model-mismatch and noise. In the first stage we estimate a continuous domain representation of the edge set of the image. In the second stage we perform an extrapolation in Fourier domain by a least squares two-dimensional linear prediction, which recovers the exact Fourier coefficients of the underlying image. We demonstrate our algorithm on the super-resolution recovery of MRI phantoms and real MRI data from low-pass Fourier samples, which shows benefits over standard approaches for single-image super-resolution MRI.

1   Introduction

Many studies in image processing show that natural images can be represented efficiently as piecewise smooth functions [39, 54, 15]. For instance, the Mumford-Shah formulation poses the recovery of a piecewise smooth image as a variational optimization problem that jointly estimates a segmentation of the image and the smooth functions on the regions defined by the segmentation [42, 43]. This problem is typically solved using iterative algorithms that alternate between the evolution a level-set curve and the recovery of the full signal [8, 60, 62]. These constrained level-set methods have been demonstrated to yield considerably improved reconstructions in challenging inverse problems [67, 32, 56, 3, 4, 17, 49]. Unfortunately, these algorithms are often slow and vulnerable to local minima problems, which restrict their utility in applications.

At the same time, the recovery of piecewise smooth one-dimensional signals using harmonic retrieval/linear prediction has been actively investigated in signal processing [37, 63, 40, 18, 2]. This includes the so-called finite-rate-of-innovation (FRI) framework [2] which extends Prony's method [55, 13] to the recovery of a much wider class of signals, including non-uniform 1-D splines with unknown knot locations. The FRI framework is also related to the earlier work [25, 27, 37, 26], where Haacke and Liang proposed a super-resolution reconstruction technique for MRI using piecewise 1-D polynomial modeling. However, all of these techniques are fundamentally tied to the recovery of 1-D signals. While some extensions of 1-D FRI schemes to multidimensional polygonal signals are available [41, 52, 11], these specialized models do

[†]Department of Mathematics, University of Iowa, Iowa City, Iowa. (*gregory-ongie@uiowa.edu*).

[‡]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, Iowa. (*mathews-jacob@uiowa.edu*)

not provide good approximations of natural images. Recent work on atomic norm minimization can also be applied to the recovery of a weighted linear combination of 1-D Diracs at arbitrary locations from few Fourier samples [6, 59, 7]; these methods are off-the-grid continuous domain generalizations of compressed sensing theory [16, 1]. While the framework can account for piecewise constant 1-D signals by analyzing the derivative of the signal [6], and linear combination of Diracs in 2-D [66], its extension to piecewise constant signals in two or higher dimensions is not straightforward. For instance, the partial derivatives piecewise constant images have singularities supported on curves that bound the constant regions. We observe that this violates two of the main assumptions required for super-resolution recovery in [6]: (1) the number of singularities is finite, and (2) the singularities are well-separated. Moreover, the 2-D setting in [66, 34] fails to capture the extensive structure that is present in imaging applications. Namely, the image discontinuities are not often isolated, but are instead localized to smooth curves in 2-D.

Recently, Pan *et al.* [47] proposed an extension of FRI methods to recover piecewise complex analytic images in continuous domain. Specifically, they assume that the complex derivative of the image is supported on the zero level-set of a band-limited periodic function, called the annihilating polynomial. The authors rely on the annihilation of the Fourier transform of the derivatives by convolution with the Fourier coefficients of the annihilating polynomial to recover the edge set, and eventually the image. The work in this paper is mainly motivated by [47], and aims to overcome several of its limitations. For instance, the recovery of an arbitrary piecewise complex analytic signal from fintely many Fourier coefficients is ill-posed, since this class of functions has infinite degrees of freedom. Even in recovering the edge set, no sampling conditions that guarantee perfect recovery were available in [47]. Another problem is that the complex analytic signal model in [47] is not realistic for natural images.

The main focus of this paper is to develop a novel theoretical framework and efficient algorithms to recover piecewise constant images, whose discontinuities are localized to zero level-sets of band-limited functions, from few Fourier samples. We introduce a two step recovery scheme, which has similarities to Prony's method and traditional FRI recovery. The first step involves the determination of the discontinuities of the image as the zero set of a band-limited function, which translates to a set of linear annihilation conditions in Fourier domain. Once the edge set is recovered, we recover the full image as the solution to an optimization problem. This work connects the classical Mumford-Shah algorithm, which alternates between level-set estimation and signal estimation, with FRI theory, which provides recovery guarantees from a finite number of Fourier samples.

Unlike the piecewise analytic model in [47], the number of degrees of freedom of our proposed piecewise constant representation is finite, which enables the successful recovery of the signal from finite number of Fourier coefficients. We introduce necessary and sufficient conditions for the recovery of the edge set and the edge set aware recovery of the signal. The one-to-one correspondence between the degree of the annihilating polynomial and the number of singularities in the signal, which is used to establish guarantees in the traditional FRI settings [63], cannot be exploited in our setting since the number of singularities is no longer finite. Instead, we introduce novel proof techniques based on the algebraic geometry of zero sets of trigonometric polynomials to establish the uniqueness guarantees.

We also introduce several refinements to make the two-step algorithm robust to noise and model mismatch. Since the model order–the number of coefficients necessary to describe the edge set–is not typically known in advance, we rely on a singular value decomposition of the block Toeplitz matrix corresponding to the linear annihilation relation for model order determination. Once the null-space of the matrix is estimated, we pose the extrapolation of the Fourier coefficients of the signal from the known samples as an optimization problem. The algorithm determines the signal such that the energy of the convolution between its Fourier coefficients and the null-space filters is minimized. We show that this approach is approximately equivalent to estimating the edge set as a sum of squares polynomial built from a basis of null-space filters. This approach is similar to the projection step in the multiple signal classification algorithm (MUSIC) algorithm [48], used in direction-of-arrival estimation and other signal

processing applications.

We demonstrate the proposed two-stage recovery scheme on the the super-resolution of magnetic resonance (MR) images from low-resolution Fourier samples. The experiments show our scheme has benefits over other standard discrete spatial domain approaches to single-image super-resolution in MRI. In particular, our scheme is better able to resolve fine image features, preserve strong edges, and suppress noise artifacts. Our work is also fundamentally different from several super-resolution approaches in image processing and MRI, where edge information from other sources (co-registered high resolution datasets, anatomical information, *etc.*) is used to constrain reconstructions from few measurements; see e.g. [30, 65, 20, 33, 35, 29, 61, 38, 23]. Specifically, we seek to estimate the edge locations using only the partial Fourier data itself rather than using additional prior information. This approach can be viewed as a fully 2-D version of the 1-D super-resolution approach investigated in [25, 27, 37, 26].

Our work also has conceptual similarities to recent low-rank structured matrix completion methods in MRI [51, 28, 34]. In particular, [28] and [34] exploit the low-rank structure of a multi-fold Hankel matrix obtained built from Fourier samples. The low-rank property of the structured matrix results from assuming signal sparsity and other constraints. For example, in [28] MR images are modeled as signals with limited spatial support and smoothly varying phase. We observe that these image models are far less constrained than the piecewise constant model considered in this work, which makes them less suitable for super-resolution recovery. The model in [34] assumes that the signal sparse in a transform domain, i.e. consists of isolated singularities, which allows them to adapt the theory in [12] for recovery. In contrast, we model the partial derivatives of the continuous signal to be localized to curves. The direct extension of the theory from [12] as well as the traditional FRI setting is not possible in our context. Since our model captures the smoothness of the edge contours often found in images, it is more constrained than models like [34] that assume isolated discontinuities.

Preliminary versions of this work were presented in conference papers [45] and [44]. In particular, Theorems 1 and 2 were stated in [44] without proof. In addition to supplying the proofs of these results, we provide sampling guarantees for edge aware signal recovery that were absent in [44]. While the basic version of the two step algorithm presented here was introduced in [45], it is considerably generalized and validated using MRI data in this work.

1.1   Notation

We collect notation used throughout the paper here for easy reference. Bold lower-case letters $\boldsymbol{x}$ are used to indicate vector quantities, bold upper-case $\boldsymbol{X}$ to denote matrices, and calligraphic script $\mathcal{X}$ for general linear operators. We typically reserve lower-case greek letters μ, γ, *etc.* for trigonometric polynomials (9) and upper-case greek letters Λ, Γ, *etc.* for their coefficient index sets, i.e. finite subsets of the integer lattice $\mathbb{Z}^2$, with cardinality denoted by $|\Lambda|$. A set of complex coefficients indexed by Λ is denoted by $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, which we also treat as a vector in $\mathbb{C}^{|\Lambda|}$ when convenient. We write $\Lambda + \Gamma$ for the dilation of the index set Γ by Λ, i.e. the Minkowski sum $\{\boldsymbol{k} + \boldsymbol{\ell} : \boldsymbol{k} \in \Lambda, \ \boldsymbol{\ell} \in \Gamma\}$, and write 2Λ to mean $\Lambda + \Lambda$, $3\Lambda = 2\Lambda + \Lambda$, *etc.* We also denote the contraction of Γ by Λ by $\Gamma : \Lambda = \{\boldsymbol{\ell} \in \Gamma : \boldsymbol{\ell} - \boldsymbol{k} \in \Gamma \text{ for all } \boldsymbol{k} \in \Lambda\}$. For two coefficient vectors $\boldsymbol{c} = (c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ and $\boldsymbol{d} = (d[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, we denote their Hermitian inner product by $\langle \boldsymbol{c}, \boldsymbol{d} \rangle = \sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}]\overline{d[\boldsymbol{k}]}$. For any square-integrable functions f and g we use $\langle f, g \rangle$, to denote the usual Hermitian inner product $\langle f, g \rangle = \int f(\boldsymbol{r})\overline{g(\boldsymbol{r})}d\boldsymbol{r}$. Finally, $\langle D, \varphi \rangle$ is also used to denote the dual pairing for a tempered distribution D acting on the Schwartz space $\mathcal{S}$ of (complex) test functions φ.

2   Overview

We introduce a two step method to recover a piecewise constant image from few low-pass Fourier samples, assuming the edge set of the image is well-approximated by the zero level-set of a bandlimited function.

In the first step, we estimate the edge set by solving for the level-set function. In the second step, we use the level-set function to perform an edge-aware recovery of the full signal. To make the presentation self-contained, and to motivate our proposed 2-D recovery scheme, we start with a brief review of FRI recovery of piecewise constant signals in 1-D.

2.1 Review of 1-D FRI theory

The FRI framework introduced in [63] can be viewed as an extension of Prony's method to the recovery of a wide class of signals from sub-Nyquist sampling rates, including piecewise polynomial and non-uniform spline signals (similar extensions of Prony's method for the recovery of 1-D piecewise polynomials was considered in Haacke and Liang's earlier work [25, 27, 37, 26]). In order to draw parallels to our 2-D recovery scheme, we focus here on the FRI recovery of a continuous domain 1-D piecewise constant signal $f : [0, 1] \to \mathbb{C}$, given as

$$f(x) = \sum_{i=1}^{K-1} a_i \, 1_{[x_i, x_{i+1})}(x); \tag{1}$$

from its low-pass Fourier coefficients

$$\widehat{f}[k] = \int_0^1 f(x) e^{-j2\pi kx} dx; \quad k \in \mathbb{Z}, \, |k| \le N.$$

for some fixed N. In (1), 1_U denotes the characteristic function of the set U, $a_i \in \mathbb{C}$ are the signal amplitudes, and $0 < x_1 < x_2 < \cdots < x_K < 1$ are the edge locations. See Figure 1a for an illustration of this signal model and the two-step FRI recovery described below.

First, observe that the derivative ∂f (in the sense of distributions) is a stream of Diracs:

$$\partial f(x) = \sum_{i=1}^{K} b_i \, \delta(x - x_i); \tag{2}$$

where $b_i = a_i - a_{i-1}$, for $i = 1, \ldots, K$ with $a_K = a_0 = 0$. Therefore, the Fourier coefficients of ∂f are a linear combination of exponentials

$$\widehat{\partial f}[k] = \sum_{i=1}^{K} b_i \, e^{-j2\pi kx_i}, \tag{3}$$

which can be computed from the known samples $\widehat{f}[k]$ by the Fourier domain relation $\widehat{\partial f}[k] = j2\pi k \widehat{f}[k]$. Recovering f now reduces to estimating the $2K$ unknown parameters $\{b_i, x_i\}_{i=1}^{K}$ from the known $2N + 1$ samples. This can be achieved using Prony's method, or one of its robust variants [13]. The solution relies on determining the so-called *annihilating polynomial*, i.e. the trigonometric polynomial defined as

$$\mu(x) = \prod_{i=1}^{K} (e^{j2\pi x} - e^{j2\pi x_i}) = \sum_{k=0}^{K} c[k] \, e^{j2\pi kx}. \tag{4}$$

which is zero at each of the edge locations $\{x_i\}_{i=1}^{K}$, and these are its only zeros. An easy computation shows that the Fourier coefficients $\widehat{\partial f}[k]$ are annihilated by convolution with the annihilating polynomial coefficients $c[k]$:

$$\sum_{k=0}^{K} c[k] \, \widehat{\partial f}[\ell - k] = 0, \quad \text{for all} \ \ell \in \mathbb{Z}. \tag{5}$$

4

This can also be seen from fact that in spatial domain,

$$\mu \, \partial f = 0,$$

in the sense of distributions. Therefore, provided we have at least $2K$ contiguous samples of $\widehat{\partial f}$ (i.e. $N \geq K$) we can form a linear Toeplitz system from (5), and solve for the $K+1$ unknown coefficients $c[k]$. This allows us to recover the edge locations $\{x_i\}_{i=1}^{K}$ as the roots of $\mu(x)$. Then we may solve for the coefficients b_i, and hence the amplitudes a_i, by substituting the $\{x_i\}_{i=1}^{K}$ into the model (3) and solving an overdetermined linear system.

The main tool enabling FRI recovery is the existence of a trigonometric polynomial whose zero set coincides with the edge locations of the piecewise constant signal, or equivalently, the singularities of its derivative. Similar to [47], we propose extending this scheme to 2-D by considering the infinite one-dimensional zero sets of two-variable trigonometric polynomials.

2.2 Image Model

We focus on a continuous domain representation of images, that is, as functions $f : \mathbb{R}^2 \to \mathbb{C}$ with compact support. For simplicity we will assume the support is always contained within $[0,1]^2$. We are particularly interested in images that can be modeled as 2-D *piecewise constant* functions

$$f(\boldsymbol{r}) = \sum_{i=1}^{n} a_i \, 1_{U_i}(\boldsymbol{r}), \quad \text{for all} \quad \boldsymbol{r} = (x,y) \in [0,1]^2, \tag{6}$$

where $a_i \in \mathbb{C}$, and 1_U is the characteristic function of the region U:

$$1_U(\mathbf{r}) = \begin{cases} 1 & \text{if } \boldsymbol{r} \in U \\ 0 & \text{else.} \end{cases} \tag{7}$$

Here we assume each region $U_i \subseteq [0,1]^2$ has piecewise smooth boundary ∂U_i. We also assume the representation (6) is expressed with the fewest number of characteristic functions such that the regions U_i are pairwise disjoint. In this case the set of discontinuous points of f coincides with $E = \cup_{i=1}^{n} \partial U_i$, which we call the *edge set* of the image.

Our aim is to recover the piecewise constant image $f(\boldsymbol{r})$ from few of its Fourier coefficients

$$\widehat{f}[\boldsymbol{k}] = \int_{[0,1]^2} f(\boldsymbol{r}) e^{-j2\pi \boldsymbol{k} \cdot \boldsymbol{r}} d\boldsymbol{r}, \quad \boldsymbol{k} \in \Gamma \subseteq \mathbb{Z}^2 \tag{8}$$

where Γ is some finite sampling set. Without any further assumptions on the edge set, unique recovery of images in the form (6) from finitely many Fourier samples is clearly ill-posed since the class of piecewise smooth curves has infinite degrees of freedom. Hence, we restrict the edge set to belong to a parametric model described by finitely many parameters. Motivated by the FRI curves framework [47], we will constrain the edge set to be the zero set of a 2-D band-limited periodic function:

$$C = \{\mathbf{r} \in [0,1]^2 : \mu(\boldsymbol{r}) = 0\}, \quad \text{where} \quad \mu(\boldsymbol{r}) = \sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}] \, e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}, \quad \forall \boldsymbol{r} \in [0,1]^2, \tag{9}$$

where the coefficients $c[\boldsymbol{k}] \in \mathbb{C}$, and Λ is any finite subset of $\mathbb{Z}^2$. We call any function μ described by (9) a *trigonometric polynomial*, and the zero set $C = \{\mu = 0\}$ a *trigonometric curve*, provided C is infinite and has no isolated points. Note a piecewise constant image cannot have isolated points in its edge set, so this restriction on the zero set is fully compatible with our image model. We elaborate more on the properties of trigonometric curves in §2.4; see also Fig. 2.

5
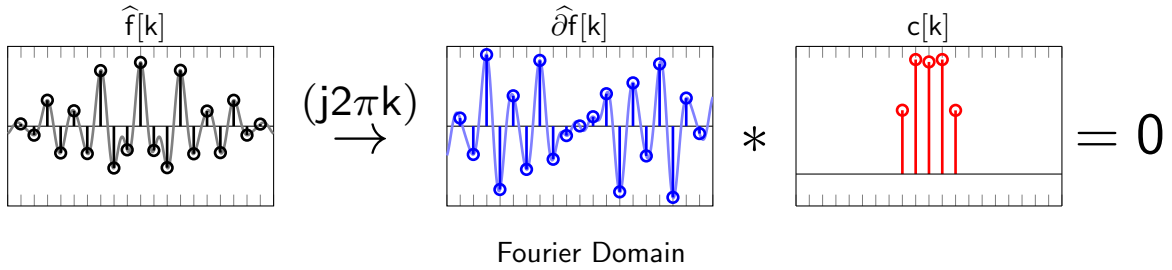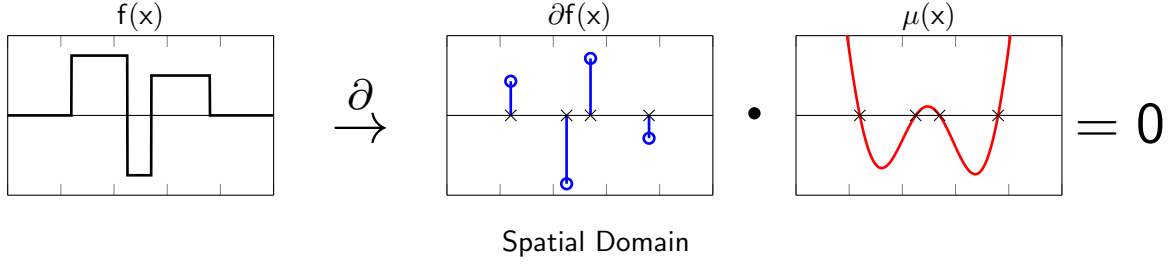
(a) FRI recovery of 1-D piecewise constant signals
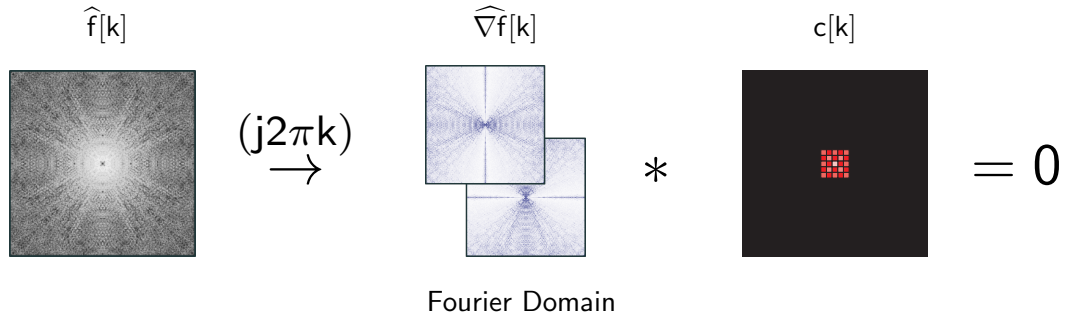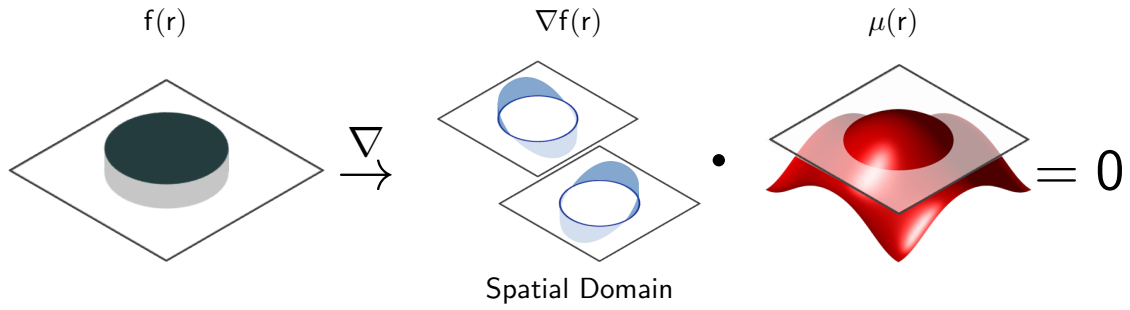
(b) Proposed extension to 2-D piecewise constant images

Figure 1: *Overview of proposed extension of FRI to 2-D piecewise constant images.* In (a) we illustrate the FRI recovery of a 1-D continuous domain piecewise constant signal $f(x)$, as described in §2.1. The distributional derivative $\partial f(x)$ is a stream of Diracs supported on the edge set of the signal. Hence, multiplication by a trigonometric polynomial $\mu(x)$ which is zero on the edge set will annihilate $\partial f(x)$ in spatial domain. This descends to a linear convolution relation in Fourier domain. Therefore, given sufficiently many uniform low-pass Fourier samples $\widehat{f}[k]$ we may compute $\partial \widehat{f}[k]$, and form a linear system to solve for the finite Fourier coefficients $c[k]$ of $\mu(x)$. In (b) we illustrate the proposed extension of this scheme to 2-D presented in §2.3. Here $f(\boldsymbol{r})$, $\boldsymbol{r} = (x, y)$, is now the characteristic function for a region U in the plane. We show the distributional gradient of f is supported on the one-dimensional edge set ∂U. Multiplication by a trigonometric polynomial $\mu(\boldsymbol{r})$ which vanishes on ∂U will annihilate ∇f in spatial domain, and this likewise translates to a linear convolution annihilation relation in Fourier domain, which we may use to solve for coefficients $c[\boldsymbol{k}]$ of $\mu(\boldsymbol{r})$.

6

2.3 Annihilation property for piecewise constant images

The annihilation relation (5) for 1-D piecewise constant signals uses the fact that the derivative of the signal can be represented as a sum of isolated Diracs. Similar annihilation relations extend to multi-dimensional signals when they can expressed as sum of isolated Diracs in a transform domain [52, 11, 34]. However, the derivative of a piecewise constant image cannot be represented as a sum of isolated Diracs, which prohibits the direct extension of these methods to our setting. However, [47] showed it is still possible to derive an annihilation relationship for 2-D complex analytic signals having singularities supported on trigonometric curves (9). Similarly, we now show that the Fourier coefficients of the partial derivatives of piecewise constant images (6) having edge set described by a trigonometric curve satisfy a 2-D annihilation relation in Fourier domain.

First consider the case of a single characteristic function 1_U where U is a simply connected region with piecewise smooth boundary ∂U. We will show that the product of the gradient of 1_U by any smooth periodic function ψ vanishing on ∂U is identically zero:

$$\psi \nabla 1_U = 0. \tag{10}$$

However, since 1_U is non-smooth, the above equality can only be understood in a weak sense. In particular, we will find it useful to interpret the partial derivatives of 1_U as tempered distributions, i.e. as linear functionals acting on test functions of Schwartz class $\mathcal{S}$ (see, e.g., [57] or [19]). Since the domain $[0,1]^2$ is compact, in this case $\mathcal{S}$ coincides with all C^∞ smooth complex-valued periodic functions on $[0,1]^2$. Letting $\varphi \in \mathcal{S}$ denote any such test function, we have:

$$\langle \partial_x 1_U, \varphi \rangle = -\langle 1_U, \partial_x \varphi \rangle = - \int_U \partial_x \varphi \, d\boldsymbol{r} = - \oint_{\partial U} \varphi \, dy \tag{11}$$

where the last step follows by Green's theorem. Likewise, one can show

$$\langle \partial_y 1_U, \varphi \rangle = \oint_{\partial U} \varphi \, dx. \tag{12}$$

We can generalize the above relation to the gradient $\nabla 1_U = (\partial_x 1_U, \partial_y 1_U)$ by considering it to be a distribution acting on test fields $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ where $\varphi_1, \varphi_2 \in \mathcal{S}$. For all such $\boldsymbol{\varphi}$ we have

$$\langle \nabla 1_U, \boldsymbol{\varphi} \rangle = -\langle 1_U, \nabla \cdot \boldsymbol{\varphi} \rangle = - \int_U \nabla \cdot \boldsymbol{\varphi} \, d\mathbf{r} = - \oint_{\partial U} \boldsymbol{\varphi} \cdot \boldsymbol{n} \, ds \tag{13}$$

where $\boldsymbol{n}$ is the unit outward normal to the curve ∂U.

Now suppose ψ is any smooth function that vanishes on ∂U, i.e. $\psi(\boldsymbol{r}) = 0$ for all $\boldsymbol{r} \in \partial U$. Since ψ is smooth the product $\psi \nabla 1_U$ is well-defined as a tempered distribution, and for all test fields $\boldsymbol{\varphi}$ we have

$$\langle \psi \nabla 1_U, \boldsymbol{\varphi} \rangle = \langle \nabla 1_U, \psi \, \boldsymbol{\varphi} \rangle = - \oint_{\partial U} \psi \, (\boldsymbol{\varphi} \cdot \boldsymbol{n}) \, ds = 0. \tag{14}$$

Therefore we have shown (10) holds in the distributional sense. The preceeding analysis easily generalizes to piecewise constant functions by linearity: if f is a piecewise constant function (6), and ϕ is any smooth function that vanishes on the edge set $E = \cup_{i=1}^n \partial U_i$, then

$$\phi \nabla f = \sum_{i=1}^n a_i \underbrace{(\phi \nabla 1_{U_i})}_{0} = 0. \tag{15}$$

Now assume that $\phi = \mu$ is a trigonometric polynomial, and hence is described by a finite number of Fourier coefficients, which we denote by $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, where $\Lambda \subseteq \mathbb{Z}^2$ is its finite Fourier support set. Taking Fourier transforms of (15), and applying the convolution theorem for tempered distributions yields the following annihilation relation:

Proposition 1. *Let $f = \sum_{i=1}^{n} a_i \, 1_{U_i}$ be piecewise constant where the edge set $E = \cup_i^n \partial U_i$ is a subset of the trigonometric curve $\{\mu = 0\}$. Then the Fourier coefficients of the gradient of f are annihilated by convolution with the Fourier coefficients $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ of μ, i.e.*

$$\sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}] \, \widehat{\nabla f}[\boldsymbol{\ell} - \boldsymbol{k}] = \boldsymbol{0}, \quad \forall \, \boldsymbol{\ell} \in \mathbb{Z}^2, \tag{16}$$

where $\widehat{\nabla f}[\boldsymbol{k}] = (\widehat{\partial_x f}[\boldsymbol{k}], \widehat{\partial_y f}[\boldsymbol{k}])$.

Due to the above property, we call μ an *annihilating polynomial* for f, and the coefficients $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ an *annihilating filter*. We show in §3 that the above Fourier domain annihilation relation can be used to recover an annihilating polynomial, and hence the edge set, from finitely many low-pass Fourier coefficients of the image.

We note that a similar annihilation relationship to (16) was derived for piecewise complex analytic signals in [47]. Specifically, in [47] it is shown that functions $f(z) = f(x + jy)$ complex analytic on a region U with boundary $\partial\Omega$ given by a trigonometric curve $\{\mu = 0\}$ satisfy the annihilation relation

$$\mu \, \partial_{\bar{z}} f = 0, \tag{17}$$

where $\partial_{\bar{z}} := \partial_x + j\partial_y$ is the complex Wirtinger derivative. While the proofs in [47] rely on a representation of f as a complex curve integral, (17) can also be understood in terms of distributions. Also note that if f is piecewise constant, then it satisfies (17). The main difference with our approach is that we assume that both partial derivatives of f are jointly annihilated by μ. This rules out functions f that are non-constant inside the regions defined by the curve $\{\mu = 0\}$.

2.4 Properties of the edge set model: Trigonometric curves

Our image model, which assumes the edge set coincides with a trigonometric curve (9), may seem overly restrictive at first. However, trigonometric curves can represent a wide variety of curve topologies, even with relatively few coefficients. For instance, trigonometric curves can have multiple connected components, non-smooth cusps, and can self-intersect in complicated ways. Moreover, any compactly supported plane curve can be approximated to an arbitrary degree of accuracy by a trigonometric curve, provided the number of coefficients is large enough; see Fig. 2.

Trigonometric curves can be characterized in terms similar to algebraic plane curves, i.e. the zero sets of polynomials in two complex variables. In particular, one can associate with every algebraic plane curve a unique minimal degree polynomial whose zero set generates the curve. Similarly, we prove there is a unique minimal degree trigonometric polynomial associated with any trigonometric curve C, which we call the *minimal polynomial* for C. Here we define the *degree* of a trigonometric polynomial μ to be the dimensions of the smallest rectangle that contains the frequency support set Λ, which we denote as $deg(\mu) = (K, L)$. In Appendix A we prove the following:

Proposition 2. *For every trigonometric curve C there is a unique (up to scaling) real-valued trigonometric polynomial μ_0 with $C = \{\mu_0 = 0\}$ such that for any other trigonometric polynomial μ with $C = \{\mu = 0\}$ we have $deg(\mu_0) \leq deg(\mu)$ componentwise.*

This proposition shows trigonometric curves can always be described as the zero set of a *real-valued* trigonometric polynomial, a property we will use implicitly from now on. Moreover, it allows us to uniquely define the *degree* of a trigonometric curve as the degree of its associated minimal polynomial. Intuitively, the degree also gives some measure of inherent complexity of a trigonometric curve. For example, the degree puts bounds on how many non-smooth points the curve may have and the number of connected curve components. One particularly important bound for this work is the following:

Proposition 3. *If C is a trigonometric curve of degree (K, L), then the number of connected components of the complement set $[0, 1]^2/C$ is at most $2KL$.*

Consequently, a piecewise constant image with edge set given by a trigonometric curve of degree (K, L) cannot consist of more than $2KL$ distinct regions, i.e. the number of terms in the sum (6) is bounded by $2KL$. This means the total degrees of freedom of such an image is at most roughly $3KL$, since we need at most $(K + 1)(L + 1)$ parameters to describe the edge set, and at most $2KL$ parameters for the region amplitudes.

Some further topological and algebraic properties of trigonometric curves and minimal polynomials can be found in Appendix B, which are used in proofs of the sampling theorems in §3 and §4.

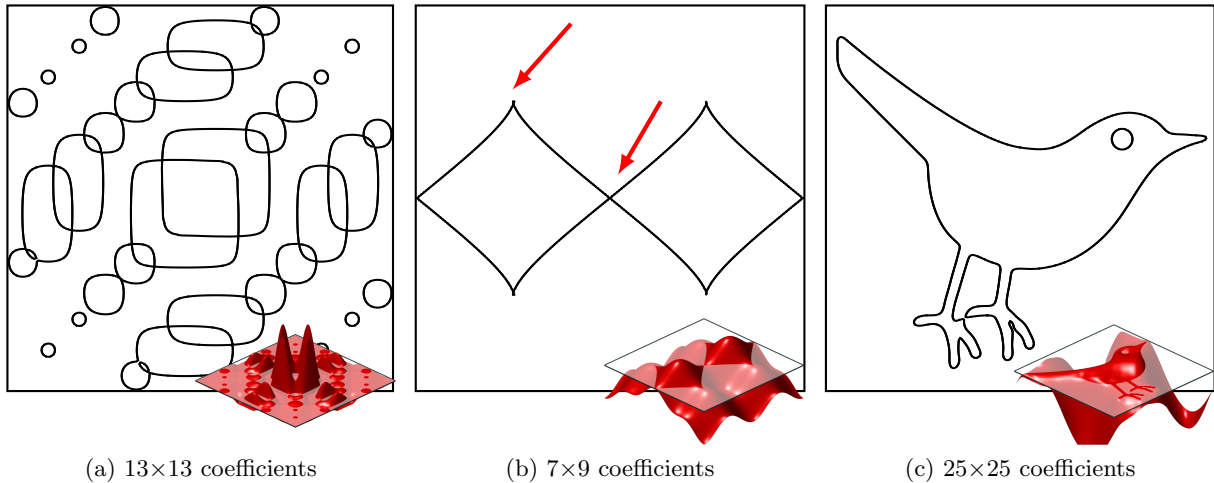(a) 13×13 coefficients   (b) 7×9 coefficients   (c) 25×25 coefficients

Figure 2: *Examples of trigonometric curves.* The one-dimensional zero sets of two-variable trigonometric polynomials are a topologically diverse class of plane curves, which can (a) have many distinct curve components, (b) non-smooth cusps and self-intersections (indicated by arrows), and (c) approximate arbitrary plane curves using few coefficients. Shown inset is a surface plot of the trigonometric polynomial whose zero set gives the corresponding trigonometric curve.

3 Recovery of edge set from finite Fourier samples

In this section, we focus on sampling guarantees for the first step of our recovery scheme. We give necessary and sufficient conditions on the number of samples required for unique recovery of the edge set of the signal from the annihilation relation. Similar sampling guarantees for edge-aware recovery of the full signal are provided in the next section.

Let f be any piecewise constant image (6) whose edge set is described by a trigonometric curve $\{\mu = 0\}$ where μ is a trigonometric polynomial with coefficients $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$. If we have access to all Fourier samples $\widehat{f}[\boldsymbol{k}]$ for all $\boldsymbol{k} \in \Gamma$, then from the annihilation relation (16) we can form the *finite* linear system of equations:

$$\sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}] \, \widehat{\nabla f}[\boldsymbol{\ell} - \boldsymbol{k}] = \boldsymbol{0}, \quad \forall \, \boldsymbol{\ell} \in \Gamma : \Lambda \tag{18}$$

where the index set $\Gamma : \Lambda \subseteq \mathbb{Z}^2$ is the set $\boldsymbol{\ell} \in \mathbb{Z}^2$ such that $\boldsymbol{\ell} - \boldsymbol{k} \in \Gamma$ for all $\boldsymbol{k} \in \Lambda$. Assuming Λ is symmetric, i.e. for all $\boldsymbol{k} \in \Lambda$ we have $-\boldsymbol{k} \in \Lambda$, then $\Gamma : \Lambda$ is the simply the set of all integer shifts $\boldsymbol{\ell}$ such that the translated support $\Lambda + \boldsymbol{\ell}$ is contained in Γ. Here the expression $\widehat{\nabla f}[\boldsymbol{k}] = (\widehat{\partial_x f}[\boldsymbol{k}], \widehat{\partial_y f}[\boldsymbol{k}])$ is computed from the known samples using the Fourier domain relations $\widehat{\partial_x f}[\boldsymbol{k}] = j2\pi k_x \widehat{f}[\boldsymbol{k}]$, and $\widehat{\partial_y f}[\boldsymbol{k}] = j2\pi k_y \widehat{f}[\boldsymbol{k}]$,

for any $\boldsymbol{k} = (k_x, k_y) \in \mathbb{Z}^2$. Provided we have enough samples of $\widehat{f}$ to construct a determined linear system of equations (18), we can potentially solve for the unknown coefficients $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$. We now derive conditions for the necessary and sufficient number of samples needed for unique recovery via the linear system (18).

3.1 Necessary condition for recovery of edge set

Suppose we know the exact Fourier support $\Lambda \subseteq \mathbb{Z}^2$ of the minimal polynomial μ describing the edge set of the image, and let $\Gamma \subseteq \mathbb{Z}^2$ be an arbitrary sampling set. To allow for the unique recovery of the coefficients $(c[k] : \boldsymbol{k} \in \Lambda)$ of μ by solving (18), we need enough equations so that the nullspace of the system is at most one-dimensional. This means we need the number of equations to match $|\Lambda| - 1$. For every shift of the filter support Λ contained in the sampling set Γ, i.e. every $\boldsymbol{\ell} \in \Gamma : \Lambda$, we obtain two equations corresponding to each partial derivative. Therefore, a necessary condition for the recovery of the $(c[k] : \boldsymbol{k} \in \Lambda)$ is

$$2|\Gamma : \Lambda| \geq |\Lambda| - 1. \tag{19}$$

However, in general we will not know in advance the exact Fourier support of the minimal polynomial. To gain some intuition on the number of samples required to recover the signal, we focus on the case of rectangular Fourier supports. In particular, we assume the degree of the minimal polynomial is known, i.e. the dimensions of the smallest rectangle containing the true Fourier support. Recall that the degree also gives some measure of the complexity of the edge-set (see Prop. 2 & Prop. 3). To recover the minimal polynomial coefficients, we can instead solve (18) for all coefficients within the minimal rectangular support Λ_R. When the sampling set Γ is also rectangular, by counting shifts of Λ_R in Γ, the expression in (19) reduces to the following necessary condition:

Proposition 4. *Let f be piecewise constant with edge set described by a trigonometric curve of degree (K, L). A necessary condition to recover the edge set from the annihilation equations* (18) *is to collect samples of $\widehat{f}$ on a $(K' + 1) \times (L' + 1)$ rectangular grid in $\mathbb{Z}^2$ with $K' \geq K$ and $L' \geq L$ such that*

$$2\left(K' - K + 1\right)(L' - L + 1) \geq (K + 1)(L + 1) - 1.$$

To illustrate this bound, suppose μ has degree (K, L), and we take Fourier samples from a rectangular grid of size $(1 + \alpha)K \times (1 + \alpha)L$, $\alpha > 0$. Then it is necessary that $\alpha \geq \frac{1}{\sqrt{2}}$, or that we collect roughly $3KL$ Fourier samples to recover the edge set.

Our numerical experiments on simulated data (see Fig. 3) indicate the above necessary conditions might also be *sufficient* for unique recovery; that is, we hypothesize the coefficients of the minimal polynomial are the only non-trivial solution to the system of equations (18) whenever it is fully determined.

3.2 Sufficient conditions for recovery of edge set

We now focus on sufficient conditions for the recovery of the edge set of a piecewise constant image, or equivalently, the coefficients of its minimal polynomial, from (18). For simplicity, we first state our result in the case of a single characteristic function. We give the proof of this result, and all the remaining proofs in §3–5, in Appendix B.

Theorem 1. *Let $f = 1_U$ be the characteristic function of a simply connected region U with boundary $\partial U = \{\mu = 0\}$, where μ is the minimal polynomial. Then the coefficients $\boldsymbol{c} = (c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ of μ can be uniquely recovered (up to scaling) from samples of $\widehat{f}$ in 3Λ as the only non-trivial solution to the equations*

$$\sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}] \, \widehat{\nabla f}[\boldsymbol{\ell} - \boldsymbol{k}] = \mathbf{0}, \quad \textit{for all} \ \ \boldsymbol{\ell} \in 2\Lambda. \tag{20}$$

(a) Original image   (b) Recovery from necessary number of samples   (c) Recovery from fewer than necessary samples

(d) Original image   (e) Recovery from necessary number of samples   (f) Recovery from fewer than necessary samples
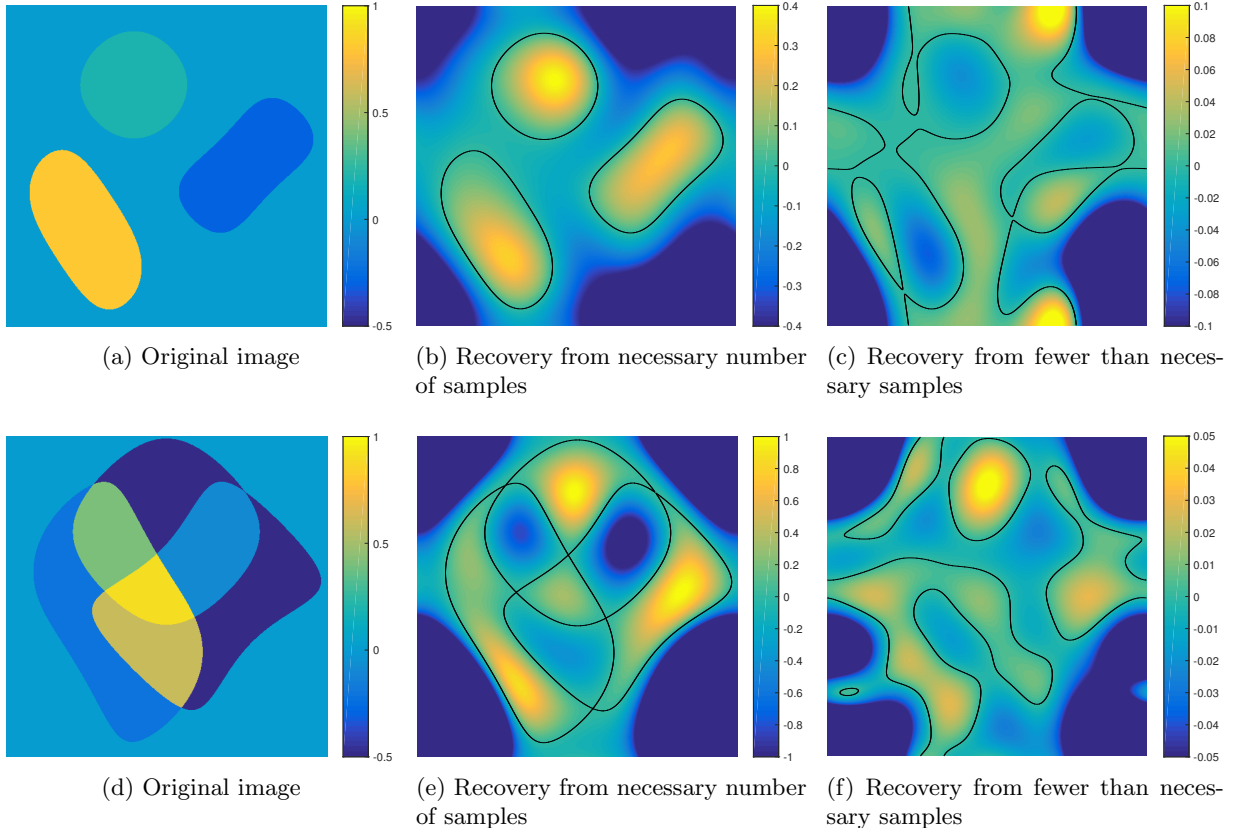
Figure 3: *Exact recovery of edge set from the minimum necessary number of Fourier samples.* The piecewise constant images shown in (a)&(d) were generated to have edge set given by the trigonometric curve $\{\mu = 0\}$ of degree $(6,6)$, i.e. the known minimal polynomial μ is described by $7{\times}7$ coefficients. We recover μ by solving the linear system (18) built from 11×11 Fourier samples of the original image, which were approximated numerically to high precision. This is the necessary minimum number of samples specified by Prop. 4, and is less than the sufficient number of samples specified by Thm. 2 ($19{\times}19$ samples). The recovered minimal polynomial μ is shown in (b)&(e) with $\{\mu = 0\}$ plotted in black; we observe the recovered coefficients match the known minimal polynomial coefficients to within machine precision. In (c)&(f) we show a solution obtained from fewer than the necessary minimum number of samples. Note that in this case the true edge set does not coincide with the zero set of the recovered trigonometric polynomial.

The proof of Theorem 1 entails showing any other trigonometric polynomial η having coefficients $(d[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ satisfying (20) must vanish on ∂U. From this it follows that η is a scalar multiple of μ by properties of minimal polynomials. We now extend the above result to piecewise constant signals, provided the boundaries of the regions do not intersect:

Theorem 2. *Let $f = \sum_{i=1}^{n} a_i \, 1_{U_i}$ be piecewise constant such that the boundary curves $\partial U_i, i = 1, \dots, n$, are connected, pairwise disjoint, and given by trigonometric curves: $\partial U_i = \{\mu_i = 0\}$, where μ_i is the minimal polynomial. Then, the coefficients $\boldsymbol{c} = (c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ of $\mu = \mu_1 \cdots \mu_n$, and equivalently the edge set $E = \cup_{i=1}^{n} \partial U_i$, can be uniquely recovered (up to scaling) from samples of $\widehat{f}$ in 3Λ as the only non-trivial solution of*

$$\sum_{\boldsymbol{k}\in\Lambda} c[\boldsymbol{k}] \, \widehat{\nabla f}[\boldsymbol{\ell} - \boldsymbol{k}] = \boldsymbol{0}, \quad \text{for all} \ \ \boldsymbol{\ell} \in 2\Lambda. \tag{21}$$

We remark that the proof of Theorem 2 can be adapted to many instances where the boundary curves

11

are allowed to intersect, provided we take slightly more samples than 3Λ. The number of additional samples required depends on the precise intersection geometry, and is difficult to characterize for arbitrary curves, and so we do not state a general result here.

Note the sampling requirement of 3Λ in Theorems 1 and 2 is greater than the necessary number of samples given in Prop. 4. For example, if the edge set has degree (K, L), the necessary number of samples is roughly $3KL$ whereas the sufficient number is roughly $9KL$. In other words, there is a gap in our results between the necessary and sufficient number of samples needed for unique recovery of the edge set. We conjecture that Theorems 1 and 2 can in fact be sharpened to the necessary number of samples, and extended to edge sets with arbitrary intersection.

Finally, we provide a generalization of Theorem 2 that will be useful in practical applications when the degree of the edge set is unknown. Specifically, assume that the system (18) is built with a larger coefficient support set Λ' than the minimal support Λ. In this case the solution to (18) is not necessarily unique: any multiple of the minimal polynomial $\eta = \mu\gamma$ such that γ has coefficients in $\Lambda' : \Lambda$ is also a solution. This is because η satisfies $\{\mu = 0\} \subseteq \{\eta = 0\}$, hence is also an annihilating polynomial. We now show that *every* solution to (18) must be the coefficients of a multiple of the minimal polynomial, under the same restriction as Theorem 2:

Theorem 3. *Let f be as in Theorem 2. Suppose $\mu = \mu_1 \cdots \mu_n$ has coefficients in Λ, and let Λ' be another index set with $\Lambda \subseteq \Lambda'$. If the coefficients $(d[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda')$ of a trigonometric polynomial η are a non-trivial solution of*

$$\sum_{\boldsymbol{k} \in \Lambda'} d[\boldsymbol{k}] \, \widehat{\nabla f}[\boldsymbol{\ell} - \boldsymbol{k}] = \boldsymbol{0}, \quad \text{for all} \quad \boldsymbol{\ell} \in \Lambda' + \Lambda, \tag{22}$$

then μ divides η, that is $\eta = \mu\gamma$ where γ is another trigonometric polynomial with coefficients supported in $\Lambda' : \Lambda$.

4 Edge set aware recovery of the image

In this section we establish conditions for the unique recovery of a piecewise constant image from few of its low-pass Fourier samples with knowledge of the minimal polynomial describing its edge set. We will study the problem of identifiability of the signal in continuous domain: given ideal Fourier samples $\widehat{f}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma$, and the minimal polynomial μ, when is f is the unique function satisfying the annihilation relation, $\mu\nabla f = 0$? Here we will restrict solutions f to belong to $L^1([0,1]^2)$, and interpret ∇f in a a distributional sense. First, we show any function $g \in L^1([0,1]^2)$ satisfying the annihilation relation can be compactly parametrized as a piecewise constant function:

Proposition 5. *Let μ be any trigonometric polynomial. Suppose $g \in L^1([0,1]^2)$ satisfies the annihilation relation*

$$\mu \nabla g = 0 \tag{23}$$

where equality holds in the distributional sense. Then, g can be parameterized as the piecewise constant function

$$g = \sum_{i=1}^{n} b_i \, 1_{U_i} \tag{24}$$

almost everywhere, where $b_i \in \mathbb{C}$ and U_i are the connected components of $\{\mu = 0\}^C$.

The above proposition implies that the full recovery of the piecewise constant image now simplifies to the estimation of the unknown amplitudes b_i, provided the regions U_i are known. One approach of recovering the signal then would be to substitute the parametric model (24) back into the annihilation

system (18), and solve for the unknown amplitudes, analogous to the amplitude recovery step in Prony's method. However, this direct approach is not feasible in practice, since to identify the regions U_i would require us to factor a high degree multivariate polynomial having arbitrary complex coefficients. For this reason, obtaining an exact off-the-grid parametric representation of the full signal in spatial domain is especially challenging in our setting.

We avoid this direct approach and instead pose recovery as the solution to the following optimization problem in spatial domain:

$$\min_{g \in L^1([0,1]^2)} \|\mu \nabla g\| \quad \text{such that} \quad \hat{g}[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}], \text{ for all } \boldsymbol{k} \in \Gamma, \tag{25}$$

where $\|\cdot\|$ is any suitable norm. Alternatively, we can solve for the solution entirely in Fourier domain:

$$\min_{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \mathbb{Z}^2} \left\| \boldsymbol{c} * \widehat{\nabla g} \right\| \quad \text{such that} \quad \hat{g}[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}], \text{ for all } \boldsymbol{k} \in \Gamma. \tag{26}$$

where $\boldsymbol{c} = (c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ are the finitely supported coefficients of the minimal polynomial μ, $*$ denotes discrete 2-D convolution, and again $\|\cdot\|$ is some suitable norm. In the formulation (26), the solution is an extrapolation in Fourier domain from the known samples in Γ to the unknown coefficients, which can be viewed as a linear prediction in 2-D. Note that (25) and (26) are still infinite-dimensional optimization problems, and not solvable in practice. However, (25) and (26) will motivate the practical algorithms we introduce in §5. We now derive necessary and sufficient conditions for the unique recovery of the signal from (25) or equivalently (26).

4.1 Necessary conditions

A piecewise constant function g in the form (24) with known regions $\{U_i\}_{i=1}^n$ has n remaining degrees of freedoms corresponding to the unknown region amplitudes $\{b_i\}_{i=1}^n$. Hence the necessary minimum number of samples for unique recovery of the full signal—using *any* algorithm—is n. However, typically this quantity will be much smaller than the number of available samples required to estimate the edge set. For instance, if the edge set of the image is a trigonometric curve of degree (K, L), then by Proposition 3, then the number of distinct regions of the image is at most $2KL$. However, by Proposition 4 the necessary minimum number of samples required to uniquely reconstruct the edge set from the annihilation equations is roughly $3KL$, which is more than the maximum bound on the number of regions.

Here we note the parallels with Prony's method and 1-D FRI theory: To recover a signal with K singularities it is necessary (and sufficient) to collect $2K + 1$ uniform Fourier samples [63]. In the first stage, the full $2K + 1$ samples are used to identify the locations of the singularities. However, in the second stage only K of the $2K + 1$ samples are necessary (and sufficient) to recover the signal amplitudes. Likewise, in our setting, once the necessary number of samples is obtained to reconstruct the edge set, no further samples should be needed to recover the amplitudes.

4.2 Sufficient conditions

Under the same non-intersecting assumption on the edge set of the piecewise constant function f as in the Theorems of §3, we can show the low-pass samples of $\widehat{f}$ are sufficient to uniquely identify f from the annihilation relation with knowledge of minimal polynomial. Specifically, we show that f is identifiable from samples of $\widehat{f}$ in Λ, the coefficient support set of the minimal polynomial:

Theorem 4. *Let f be as in Theorem 2, with known minimal polynomial μ having coefficients in Λ, and let the Fourier sampling set $\Gamma \supseteq \Lambda$. If $g \in L^1([0,1]^2)$ satisfies*

$$\mu \nabla g = 0 \quad and \quad \widehat{g}[\boldsymbol{k}] = \widehat{f}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma \tag{27}$$

then g = f almost everywhere.

Again, we conjecture that this result also holds for piecewise constant functions having an edge set described by an arbitrary trigonometric curve C, not just those consisting of non-intersecting curves.

This result shows one benefit of our piecewise constant image model versus the piecewise complex analytic signal model considered in [47]: in the piecewise constant case, recovery from the annihilation relation given finitely many Fourier samples is well-posed. However, this is not the case for piecewise analytic signals using the annihilation relation (17), as considered in [47]. This is because, in analogy with Proposition 5, one can show

$$\mu \, \partial_{\bar{z}} g = 0 \tag{28}$$

if and only if g is complex analytic on each connected component U_i of $\{\mu = 0\}^C$. Or, equivalently:

$$g(z) = \sum_{i=1}^{n} h_i(z) 1_{U_i}(z), \tag{29}$$

where h_i is any function complex analytic on U_i. Signals in this class have infinite degrees of freedom since each $h_i(z)$ is described by infinitely many complex parameters, namely its power series coefficients. Therefore, recovery of the signal g from the annihilation relation (28) given finitely many Fourier samples is ill-posed. In particular, for any finite Fourier sampling set Γ, there are infinitely many signals g in the form (29) satisfying $\widehat{f}[\boldsymbol{k}] = \widehat{g}[\boldsymbol{k}]$, for all $\boldsymbol{k} \in \Gamma$.

### 4.3 Recovery from finitely many annihilation equations

The conditions derived in §4.1 and §4.2 settle the problem of unique recovery from using the infinite dimensional annihilation relation (27). A more subtle question, relevant to practical applications, is whether the Fourier extrapolation solution is uniquely determined from finitely many annihilation equations, i.e. for some large extrapolation set $\Delta \subseteq \mathbb{Z}^2$, and some small sampling set $\Gamma \subseteq \mathbb{Z}^2$ do the conditions

$$\sum_{\boldsymbol{k} \in \Lambda} c[\boldsymbol{k}] \, \nabla g[\boldsymbol{\ell} - \boldsymbol{k}] = 0 \ \text{ for all } \ \boldsymbol{\ell} \in \Delta : \Lambda \quad and \quad \widehat{g}[\boldsymbol{k}] = \widehat{f}[\boldsymbol{k}] \ \text{ for all } \ \boldsymbol{k} \in \Gamma. \tag{30}$$

imply $\widehat{g}[\boldsymbol{k}] = \widehat{f}[\boldsymbol{k}]$ for all $\boldsymbol{k} \in \Delta$? In other words, given few samples of $\widehat{f}$ and knowledge of the minimal polynomial coefficients $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, can we uniquely extrapolate $\widehat{f}$ to a larger grid $\Delta \subseteq \mathbb{Z}^2$ using the Fourier domain annihilation relation (30)? Note that (27) may be thought of as the limiting case of (30), where the extrapolation set Δ expands to all of $\mathbb{Z}^2$. Our numerical experiments (see Fig. 4) indicate that, given the minimal polynomial, unique recovery is in fact possible from (30) with the necessary minimum number of samples given in §4.1. We hypothesize that a sufficient condition analogous to Theorem 4 also holds in this case. However, establishing this condition is less straightforward than in the infinite dimensional setting (27) since in this context we lack the parameterization result (24). We plan to pursue this question in a future work.

5 Algorithms

We rely on a two step algorithm similar to Prony's method, which involves (1) the estimation of the edge set of an image (2) the edge set aware recovery of the image. Theorems 2 and 3 together guarantee the unique recovery of the edge set from the linear annihilation relation (18), while Theorem 4 guarantees the unique recovery of the full piecewise constant image given its edge set. However, these guarantees only hold under ideal conditions, making direct estimation of the signal from (18) infeasible in practice. Here we introduce robust refinements of this procedure to enable recovery of the image in the presence of noise or model-mismatch.
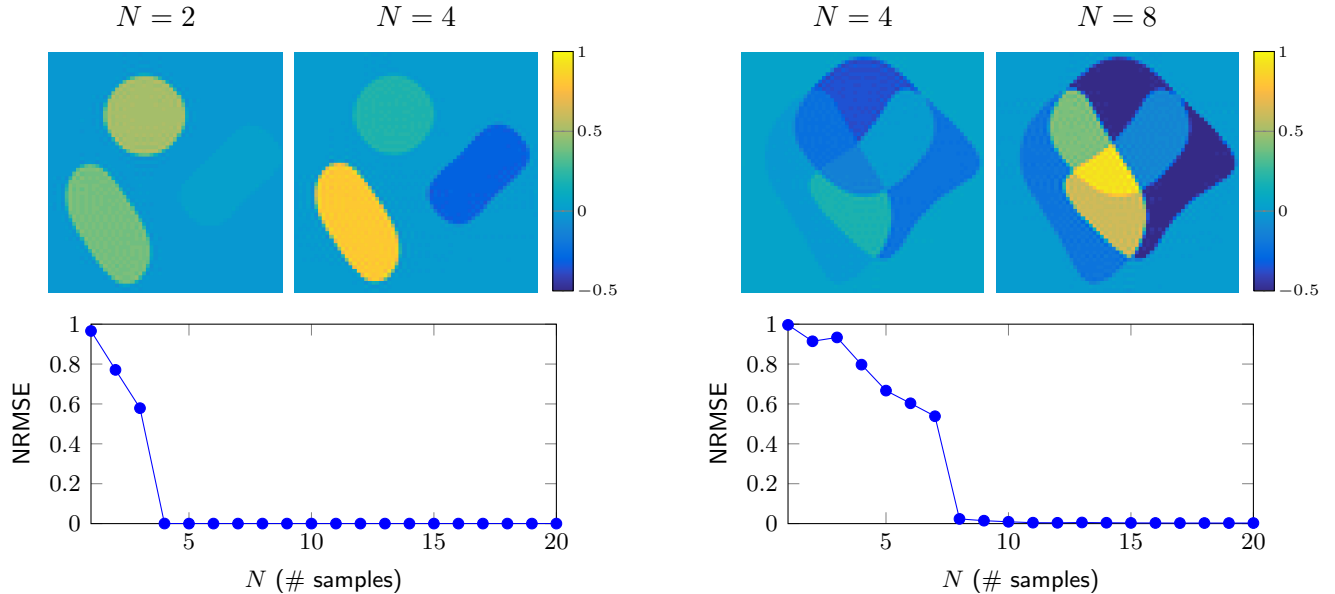
(a) Recovery of $n = 4$ regions (including background) (b) Recovery of $n = 8$ regions (including background)

Figure 4: *Edge set aware recovery of synthetic data.* Piecewise constant images with $n = 4$ and $n = 8$ distinct connected regions (including the background region) were generated with edge set given by a known minimal annihilating polynomial having 7×7 coefficients (as shown in Fig. 3). Choosing N samples uniformly at random from the 7×7 center Fourier coefficients of the image, we extrapolate the image in Fourier domain to 65×65 coefficients by solving (30) in a least-squares fashion. We plot the normalized root mean square error (NRMSE) of the recovery averaged over ten trials. Pictured above the plots is the inverse DFT of a reconstruction obtained from the indicated number of samples N. Note that in both cases we are able to recover the image nearly exactly from the necessary minimum number of samples, $N = 4$ and $N = 8$ respectively, corresponding to the number of regions. Below this number, we still recover the Fourier coefficients of a piecewise constant image, however one with the incorrect region amplitudes.

5.1 Step 1: Estimation of an annihilating subspace

Directly estimating the edge set via the linear system (18) in practical applications is challenging due to the following problems:

1. *Model-order selection*: Theorems 2 and 3 assume the knowledge of the degree of the minimal polynomial describing the edge set of the image. However, this is often unknown in practical applications.

2. *Noise and model-mismatch*: The available Fourier samples are often corrupted by noise in practical settings. In addition, the assumption that the underlying image is piecewise constant with edge set described by a trigonometric curve will only be approximately true in practice.

We now introduce efficient approaches to overcome the above problems. The key idea is that rather than estimating a single annihilating filter, we estimate an entire *annihilating subspace*—a collection of linearly independent annihilating filters which jointly encode the edge set. This can be thought of as an extension of *subspace methods* in spectral estimation to the 2-D setting. Specifically, we draw parallels between our algorithm and the well-known MUSIC algorithm [48]. This gives us an efficient means to address the model-order selection problem. We also introduce an efficient algorithm for the robust estimation of the annihilating subspace in §5.1.2, which address the problem of model-mismatch and noisy measurements. This scheme is related to Cadzow denoising methods [5, 14] used for robust spectral estimation, and

15

recent low-rank structured matrix completion approaches in MRI [28, 34]. Finally, once the annihilating subspace is estimated, in §5.2 we propose recovering the piecewise constant signal in Fourier domain by constraining its extrapolated Fourier coefficients to be orthogonal to the annihilating subspace.

For convenience, we rewrite the annihilation system (18) in matrix notation. First, we fix rectangular index sets $\Gamma, \Lambda \subseteq \mathbb{Z}^2$, which represent the sampling set, and annihilating filter support, respectively, and set $N = |\Gamma : \Lambda|$, $M = |\Lambda|$. Define $\mathcal{T}_x$ and $\mathcal{T}_y$ to be the linear operators mapping samples $(\widehat{f}[\boldsymbol{k}] : \boldsymbol{k} \in \Gamma)$ to matrices $\mathcal{T}_x(\widehat{f}) \in \mathbb{C}^{N \times M}$ and $\mathcal{T}_y(\widehat{f}) \in \mathbb{C}^{N \times M}$ which represent the discrete convolution of $\widehat{\partial_x f}$ and $\widehat{\partial_y f}$ with filters $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, whose output is restricted to the index set $\Gamma : \Lambda$. In particular, $\mathcal{T}_x(\widehat{f})$ and $\mathcal{T}_y(\widehat{f})$ are block Toeplitz matricies with Toeplitz blocks, where the rows of $\mathcal{T}_x(\widehat{f})$ indexed by $\boldsymbol{\ell} \in \Gamma : \Lambda$ are given by $(\widehat{\partial_x f}[\boldsymbol{\ell} - \boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ in vectorized form, and similarly for $\mathcal{T}_y(\widehat{f})$; see Fig. 5. Finally, we define the linear operator $\mathcal{T}$ as the vertical concatenation of $\mathcal{T}_x$ and $\mathcal{T}_y$, i.e.

$$\mathcal{T}(\widehat{f}) = \begin{bmatrix} \mathcal{T}_x(\widehat{f}) \\ \mathcal{T}_y(\widehat{f}) \end{bmatrix} \in \mathbb{C}^{2N \times M}. \tag{31}$$

Treating $\boldsymbol{c} = (c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ as vector in $\mathbb{C}^M$ we may reproduce the annihilation system (18) by writing

$$\mathcal{T}(\widehat{f})\,\boldsymbol{c} = \boldsymbol{0}$$

Therefore, recovering an annihilating polynomial for f is equivalent to finding a vector in the nullspace of the structured matrix $\mathcal{T}(\widehat{f})$, which we call the *annihilation matrix*.

We note that our annihilation matrix is structurally similar to the to the multi-fold Hankel matrix construction in the matrix enhancement pencil matrix (MEMP) method, used for 2-D harmonic retrieval [31, 12]. The main difference in our construction is that we apply weights $j2\pi\boldsymbol{k}$ to the Fourier data before creating the structured matrix. Also, similar multi-fold Toeplitz/Hankel liftings of Fourier data have been considered in MRI reconstruction problems to model spatially limited and phase constrained images [51, 28], and to model transform sparse images [34]. Similar to us, [34] also applies weights to Fourier data before constructing the structured matrix. However, our construction (31) is unique in that we consider joint annihilation of both partial derivatives with respect to a common filter.

5.1.1 Annihilating subspace estimation

When the degree of the minimal polynomial associated with the edge set is unknown in advance—which will typically be the case—we cannot build the annihilation matrix $\mathcal{T}(\widehat{f})$ with the correct filter support to ensure its null space is one-dimensional. For instance, if the assumed filter support Λ' is strictly larger than support Λ of the minimal filter $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, the annihilation matrix will have a larger null space. This is because any filter of the form $\boldsymbol{c} * \boldsymbol{d}$ having Fourier support contained in Λ' is also an annihilating filter by the associativity of convolution. In fact, Theorem 3 says that under certain conditions on f, *every* null space vector of $\mathcal{T}(\widehat{f})$ is of this form. Hence, under these conditions, $\mathcal{T}(\widehat{f})$ is rank deficient with the following bound:

Proposition 6. *Let f be a piecewise constant image with edge set $C = \{\mu = 0\}$, where μ is the minimal polynomial with coefficients in Λ. If the assumed annihilating filter support Λ' strictly contains Λ, then*

$$rank\ \mathcal{T}(\widehat{f}) \leq |\Lambda'| - |\Lambda' : \Lambda|. \tag{32}$$

Furthermore, equality holds if f satisfies the conditions of Thm. 2 and if the sampling set $\Gamma \subseteq 2\Lambda + \Lambda'$.

We call the null space of $\mathcal{T}(\widehat{f})$ the *annihilating subspace*. Since identifying the minimal annihilating filter within the annihilating subspace when it has large dimension is challenging, we propose to leverage the entire annihilating subspace for both edge set estimation and image reconstruction.
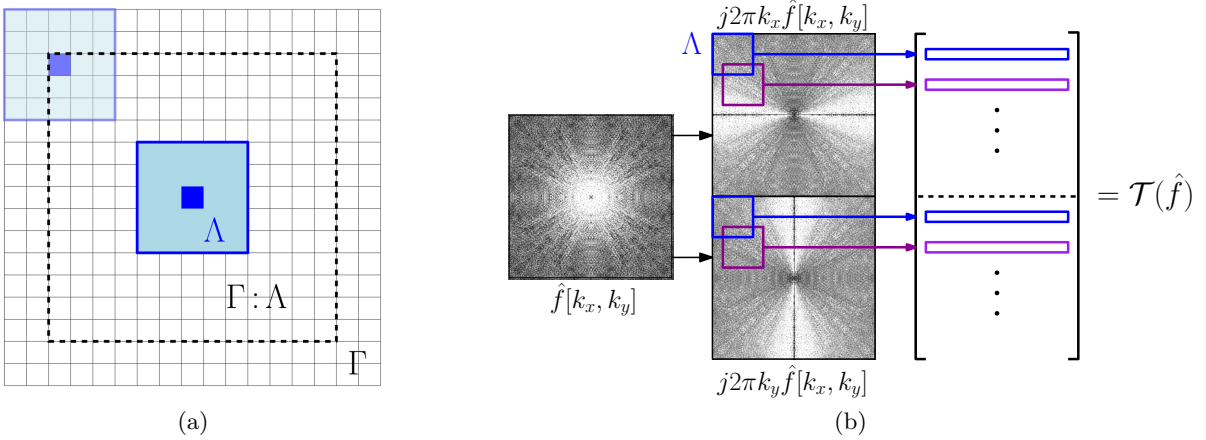
Figure 5: *Construction of the annihilation matrix.* Shown in (a) are the index sets in $\mathbb{Z}^2$ used to construct the annihilation matrix $\mathcal{T}(\widehat{f})$ defined in (31). Here Γ in is the set of Fourier sampling locations, Λ (in black) is the assumed annihilating filter support, and $\Gamma:\Lambda$ (interior of dashed line) is the set of all integer shifts $\boldsymbol{\ell}$ such that Λ shifted by $\boldsymbol{\ell}$ is contained in Γ. The construction of $\mathcal{T}(\widehat{f})$ is shown schematically in (b). From known samples $\widehat{f}[k_x, k_y]$ for $(k_x, k_y) \in \Gamma$ we compute $\widehat{\partial_x f}[k_x, k_y] = j2\pi k_x \hat{f}[k_x, k_y]$ and $\widehat{\partial_y f}[k_x, k_y] = j2\pi k_y \hat{f}[k_x, k_y]$ by a Fourier domain multiplication. Then each row of $\mathcal{T}(\widehat{f})$ is obtained by vectorizing a rectangular Fourier domain patch of size Λ from the computed samples of $\widehat{\partial_x f}$ and $\widehat{\partial_y f}$.

Observe that the edge set is still uniquely recoverable from the annihilating subspace under ideal conditions. To see this, let $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_R]$ be any orthonormal basis for the annihilating subspace. Define μ_i to be the trigonometric polynomial having coefficients given by $\boldsymbol{d}_i$. By the preceding discussion μ_i must always have the form $\mu_i = \gamma_i \mu$ for μ the minimal polynomial and γ_i another trigonometric polynomial. We define a function $\overline{\mu}$ as the following sum-of-squares average

$$\overline{\mu}(\boldsymbol{r}) = \sqrt{\sum_{i=1}^{R} |\mu_i(\boldsymbol{r})|^2} = \sqrt{\sum_{i=1}^{R} |\gamma_i(\boldsymbol{r})\mu(\boldsymbol{r})|^2}. \tag{33}$$

Note we always have $\overline{\mu}(\boldsymbol{r}) = 0$ when $\mu(\boldsymbol{r}) = 0$. We can also prove the following:

Proposition 7. *The function $\overline{\mu}$ has at most finitely many zeros on $[0,1]^2/\{\mu = 0\}$.*

Therefore, we can identify the edge set by finding the zeros of $\overline{\mu}(\boldsymbol{r})$, and removing any isolated points. This method of estimating the edge set can also be interpreted as a higher dimensional analogue of the technique used in spectral MUSIC to estimate the locations of isolated spectral peaks [48]. Recall that with spectral MUSIC one performs this estimate by computing the distance of pure exponential signals from a particular eigenspace of the signal covariance matrix. The following result shows we may also interpret $\overline{\mu}(\boldsymbol{r})$ in an analogous way:

Proposition 8. *Let $\boldsymbol{D} \in \mathbb{C}^{M \times R}$ be an orthonormal basis of the nullspace of $\mathcal{T}(\widehat{f})$, and for all $\boldsymbol{r} \in [0,1]^2$, define the filter $\boldsymbol{e_r} \in \mathbb{C}^M$ by $\boldsymbol{e_r}[\boldsymbol{k}] = e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}$ for all $\boldsymbol{k} \in \Lambda$. Then we have*

$$\overline{\mu}(\boldsymbol{r}) = \|\boldsymbol{D}\boldsymbol{D}^H \boldsymbol{e_r}\|_2$$

that is, $\overline{\mu}(\boldsymbol{r})$ is the ℓ^2-norm of the projection of $\boldsymbol{e_r}$ onto the annihilating subspace.

While we do not directly estimate the edge set in this fashion, the function $\overline{\mu}$ will play an important role in the second step of our proposed algorithm in §5.2.

17

5.1.2 Annihilation subspace estimation in the presence of noise

When the measurements are corrupted by noise or there is model mismatch, we propose performing a denoising step to better identify the annihilating subspace. This denoising step is similar to Cadzow methods used in FRI algorithms [2, 47], as well as recently proposed algorithms for structured low-rank matrix completion problems [28, 12, 46, 34]. Given noisy Fourier samples $b[\boldsymbol{k}] = \widehat{f}[\boldsymbol{k}] + n[\boldsymbol{k}]$, $\boldsymbol{k} \in \Gamma$, where we assume $n[\boldsymbol{k}]$ is a vector of i.i.d. mean-zero Gaussian noise, we attempt find the $\widehat{f}^*$ closest in Euclidean norm that satisfies the rank bound in Proposition 6:

$$\widehat{f}^* = \underset{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma}{\arg\min} \sum_{\boldsymbol{k} \in \Gamma} |\widehat{g}[\boldsymbol{k}] - b[\boldsymbol{k}]|^2 \quad \text{such that} \quad \operatorname{rank}[\mathcal{T}(\widehat{g})] \leq r. \tag{34}$$

Here the parameter r is estimated according to (32). This can be viewed as a maximum likelihood estimation, where the rank constraint on $\mathcal{T}(\widehat{g})$ enforces our modeling assumption; see [14] for a similar approach in 1-D. Note the rank constraint in (34) is non-convex, which means in general we are only able to compute an local minimizer of (34). One common heuristic for solving problems of the type (34) is to perform a Cadzow denoising on a lifted matrix. However, as observed in [47], this procedure is slow to converge in this context. Instead, we attempt to solve (34) more directly with a variable splitting approach adopted from [28]:

$$\min_{\substack{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma \\ \boldsymbol{T}:\operatorname{rank}\boldsymbol{T} \leq r}} \sum_{\boldsymbol{k} \in \Gamma} |\widehat{g}[\boldsymbol{k}] - b[\boldsymbol{k}]|^2 + \lambda \|\mathcal{T}(\widehat{g}) - \boldsymbol{T}\|_F^2 \tag{35}$$

where λ is a regularization parameter balancing data fidelity and the low-rank penalty term. This suggests the following alternating minimization scheme: set $\widehat{f}^{(0)} = b$ and for $n \geq 1$, solve

$$\boldsymbol{T}^{(n+1)} = \underset{\boldsymbol{T}:\operatorname{rank}\boldsymbol{T} \leq r}{\arg\min} \|\mathcal{T}(\widehat{f}^{(n)}) - \boldsymbol{T}\|_F^2 \tag{36}$$

$$\widehat{f}^{(n+1)} = \underset{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma}{\arg\min} \sum_{\boldsymbol{k} \in \Lambda} |\widehat{g}[\boldsymbol{k}] - b[\boldsymbol{k}]|^2 + \lambda \|\mathcal{T}(\widehat{g}) - \boldsymbol{T}^{(n+1)}\|_F^2 \tag{37}$$

The exact solution to subproblem (36) is easily computed by taking the singular value decomposition of $\mathcal{T}(\widehat{f}^{(n)}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$, and setting $\boldsymbol{T}^{n+1} = \boldsymbol{U}\boldsymbol{\Sigma}_r\boldsymbol{V}^*$, where $\boldsymbol{\Sigma}_r$ is the diagonal matrix of singular values σ_i with $\sigma_i = 0$ for all $i > r$. Also, subproblem (37) is quadratic in $\widehat{g}$, and has an exact solution given by $\widehat{g} = (b + \mathcal{T}^*(\boldsymbol{T})) / (1 + \lambda \operatorname{diag}(\mathcal{T}^*\mathcal{T}))$, where the operations above are understood element-wise. Here $\mathcal{T}^*$ denotes the adjoint of $\mathcal{T}$, which computes a weighted sum of entries of $\boldsymbol{T} \in \mathbb{C}^{2N \times M}$. In particular, one can show $\mathcal{T}^*\mathcal{T}$ is a diagonal matrix. Due to the alternating scheme, the cost in (35) is guaranteed to monotonically decrease with updates (37). A detailed convergence analysis of the algorithm is beyond the scope of this work; further analysis can be found in [28]. However, we observe the algorithm converges well in practice. In all our experiments conducted in §6 a suitable solution was reached in 5–10 iterations.

5.2 Step 2: Edge set aware image recovery

We now focus on the edge-aware recovery of the image using (26), which amounts to the Fourier domain extrapolation of the measured samples. This recovery can still be considered as "off-the-grid", in the sense that we reconstruct the exact Fourier series coefficients of the underlying continuous domain image. We will also briefly consider the case where we solve for the image "on-the-grid" in spatial domain similar to [47].

5.2.1 Fourier domain recovery: Extrapolation by linear prediction

Suppose that we have an estimate of a basis for the annihilating subspace $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_R]$. Motivated by the Fourier domain recovery formulations (26) and (30), we pose the recovery of the Fourier domain of the signal within an arbitrary rectangular domain $\Delta \subseteq \mathbb{Z}^2$ as

$$\min_{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Delta} \sum_{i=1}^{R} \|\widehat{\nabla g} * \boldsymbol{d}_i\|^2 \quad \text{subj. to} \quad \widehat{g}[\boldsymbol{k}] = \widehat{f}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma, \tag{38}$$

i.e. we extrapolate the signal by enforcing its Fourier derivatives to be annihilated by filters in the annihilating subspace. Here we restrict the convolution to the valid region $\Delta : \Lambda \subset \mathbb{Z}^2$. When the measurements are corrupted by noise (i.e. $b[\boldsymbol{k}] = \widehat{f}[\boldsymbol{k}] + n[\boldsymbol{k}]$, $\boldsymbol{k} \in \Gamma$, where n is a vector i.i.d. complex white Gaussian noise), we relax the equality constraint in (38) and instead solve

$$\min_{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Delta} \sum_{i=1}^{R} \|\widehat{\nabla g} * \boldsymbol{d}_i\|^2 + \lambda \|P_\Gamma \widehat{g} - \boldsymbol{b}\|^2. \tag{39}$$

Here, λ is a tunable parameter balancing annihilation and data fidelity, and P_Γ is projection onto the sampling set Γ. Note that (39) is a linear least-squares problem, and can be solved efficiently with a conjugate gradient iterative solver, provided we have an fast method of computing matrix-vector products with the gradient of the objective in (39). Towards this end, we rewrite (39) in terms of a linear operator $\mathcal{Q}$ acting on $\widehat{g}$:

$$\min_{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Delta} \|\mathcal{Q}\widehat{g}\|_F^2 + \lambda \|P_\Gamma \widehat{g} - \boldsymbol{b}\|^2. \tag{40}$$

where $\| \cdot \|_F$ denotes the Frobenius norm. We can express $\mathcal{Q}$ as

$$\mathcal{Q} = [\mathcal{Q}_{x,1}, \mathcal{Q}_{y,1}, ..., \mathcal{Q}_{x,R}, \mathcal{Q}_{y,R}]; \quad \mathcal{Q}_{x,i} = \boldsymbol{P}\boldsymbol{C}_i\boldsymbol{M}_x, \quad \mathcal{Q}_{y,i} = \boldsymbol{P}\boldsymbol{C}_i\boldsymbol{M}_y \text{ for all } i = 1, ..., R, \tag{41}$$

with $\boldsymbol{M}_x$ and $\boldsymbol{M}_y$ representing elementwise multiplication by $j2\pi k_x$ and $j2\pi k_y$, respectively, $\boldsymbol{C}_i$ discrete convolution with the annihilating filter $\boldsymbol{d}_i$, and $\boldsymbol{P}$ projection onto the index set $\Delta : \Lambda$. Note that since $\widehat{g}$ is restricted to Δ, we may take $\boldsymbol{C}_i$ to be a 2-D circular convolution, provided we zero-pad the filter $\boldsymbol{d}_i$ and the input $\widehat{g}$ sufficiently. Thus, we can write $\boldsymbol{C}_i = \boldsymbol{F}\boldsymbol{S}_i\boldsymbol{F}^H$ where $\boldsymbol{F}$ is the 2-D DFT on any rectangular grid containing the set $\Delta + \Lambda$, and $\boldsymbol{S}_i$ is elementwise multiplication by the inverse DFT of $\boldsymbol{d}_i$, which is given by a spatial gridding of the annihilating polynomial $\mu_i(\boldsymbol{r})$ having coefficients $\boldsymbol{d}_i$.

Taking the gradient of the objective in (40) and setting it to zero yields

$$(\mathcal{Q}^*\mathcal{Q} + \lambda P_\Gamma^* P_\Gamma)\widehat{g} = P_\Gamma^* \boldsymbol{b}.$$

From (41), the operator $\mathcal{Q}^*\mathcal{Q}$ can be expanded as

$$\mathcal{Q}^*\mathcal{Q} = \sum_{i=1}^{R} (\boldsymbol{M}_x^* \boldsymbol{C}_i^* \boldsymbol{P}^* \boldsymbol{P}\boldsymbol{C}_i\boldsymbol{M}_x + \boldsymbol{M}_y^* \boldsymbol{C}_i^* \boldsymbol{P}^* \boldsymbol{P}\boldsymbol{C}_i\boldsymbol{M}_y) \tag{42}$$

Unfortunately, when the number of annihilating filters is large (i.e. R is large), applying this operator directly will be computationally costly, requiring $8R$ two-dimensional FFT operations per application. However, if the resolution of the reconstruction grid Δ is large enough, the projection operator $\boldsymbol{P}^*\boldsymbol{P}$ will be close to the identity matrix $\boldsymbol{I}$. Making the approximation $\boldsymbol{P}^*\boldsymbol{P} \approx \boldsymbol{I}$ in (42), we obtain

$$\mathcal{Q}^*\mathcal{Q} \approx \boldsymbol{M}_x^* \boldsymbol{F}^* \boldsymbol{R}\boldsymbol{F}\boldsymbol{M}_x + \boldsymbol{M}_y^* \boldsymbol{F}^* \boldsymbol{R}\boldsymbol{F}\boldsymbol{M}_y. \tag{43}$$

where $\boldsymbol{R} = \sum_{i=1}^{R} \boldsymbol{S}_i^* \boldsymbol{S}_i$ is diagonal and represents pointwise multiplication by spatially gridded samples of the sum-of-squares polynomial $\overline{\mu}^2 = \sum_{i=1}^{R} |\mu_i|^2$, as defined in (33). Note that after computing the entries of $\boldsymbol{R}$, computing matrix-vector products with (43) requires only four 2-D FFTs per iteration to evaluate, as opposed to $8R$. Working backwards, we can show approximating $\mathcal{Q}^* \mathcal{Q}$ in this manner is equivalent to minimizing

$$\min_{\widehat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Delta} \|\overline{\mu} \, \nabla \mathcal{B}(g)\|_{L^2}^2 + \lambda \|P_\Gamma \, \widehat{g} - \boldsymbol{b}\|^2 \tag{44}$$

where here $\mathcal{B}(g)$ is the function obtained by bandlimiting g to the Fourier extrapolation set Δ, and $\| \cdot \|_{L^2}^2$ is the spatial domain L^2-norm of vector-valued functions on $[0,1]^2$. This shows that linear prediction using the entire annihilating subspace is roughly equivalent to (25). The main difference is that the minimal polynomial μ is replaced by the sum-of-squares average polynomial $\overline{\mu}$ defined in (33). We propose solving (44) as a surrogate for (39) when the dimension of the annihilating subspace is large, and call (44) the least-squares linear prediction (LSLP) approach.

5.2.2 Spatial domain recovery: Weighted total variation

We also consider solving for the discrete samples of a function $g : [0,1]^2 \to \mathbb{C}^2$ on a rectangular grid $\{\boldsymbol{r}_{i,j}\} \subset [0,1]^2$ in spatial domain:

$$\min_g \sum_{i,j} \overline{\mu}(\boldsymbol{r}_{i,j}) \, |\nabla g(\boldsymbol{r}_{i,j})| + \lambda \|P_\Gamma \, \widehat{g} - \boldsymbol{b}\|^2, \tag{45}$$

This approach of solving for the image on discrete spatial grid is similar to the approach in [47], and can be thought of as a discrete approximation of (25). To obtain a fast algorithm, we use finite differences to approximate the gradient ∇g in (45), and replace the Fourier projection $P_\Gamma \, \widehat{g}$ with its DFT counterpart. In this case, (45) can be cast as a weighted total variation minimization which can be solved efficiently with a primal-dual algorithm, such as [69]. We call (45) the weighted total variation approach (WTV).

We note that several edge-aware recovery formulations similar to our WTV (45) and LSLP approach (44) have been proposed previously in the literature; see e.g. [65, 33, 29, 61, 38, 23]. The difference of the proposed approach lies in the estimation of the edge weights, which are obtained from the low-resolution data itself, rather than from *a priori* known high resolution edge information, as is assumed in the above cited works.

6 Experiments

In this section we demonstrate the utility of our proposed algorithms for super-resolution in magnetic resonance (MR) imaging. A single-coil MR acquisition restricted to a single plane with sampling locations on a uniform Cartesian grid can be accurately modeled as the Fourier coefficients of the underlying 2-D image [21]. This suggests we can apply our proposed two-stage recovery scheme to recover MR data from low-pass Fourier samples at high resolution, provided the underlying image is well-approximated as piecewise constant.

First, to test the validity of the proposed scheme, we experiment on simulated data generated from piecewise constant phantoms. To ensure resolution-independence, we compute ideal Fourier samples from analytical Fourier domain expressions of the MR phantoms, as done in [24]. Taking low-pass samples located within a centered rectangular window, we solve for an annihilating filter of the maximum possible size allowable using the procedure outlined in Section 5. We then recover the image using the proposed algorithms in §5, and compare against a standard total variation (TV) regularized recovery, and a non-convex version of TV (ncvx-TV). For ncvx-TV, we adopt the approach in [10, 9] using a non-convex
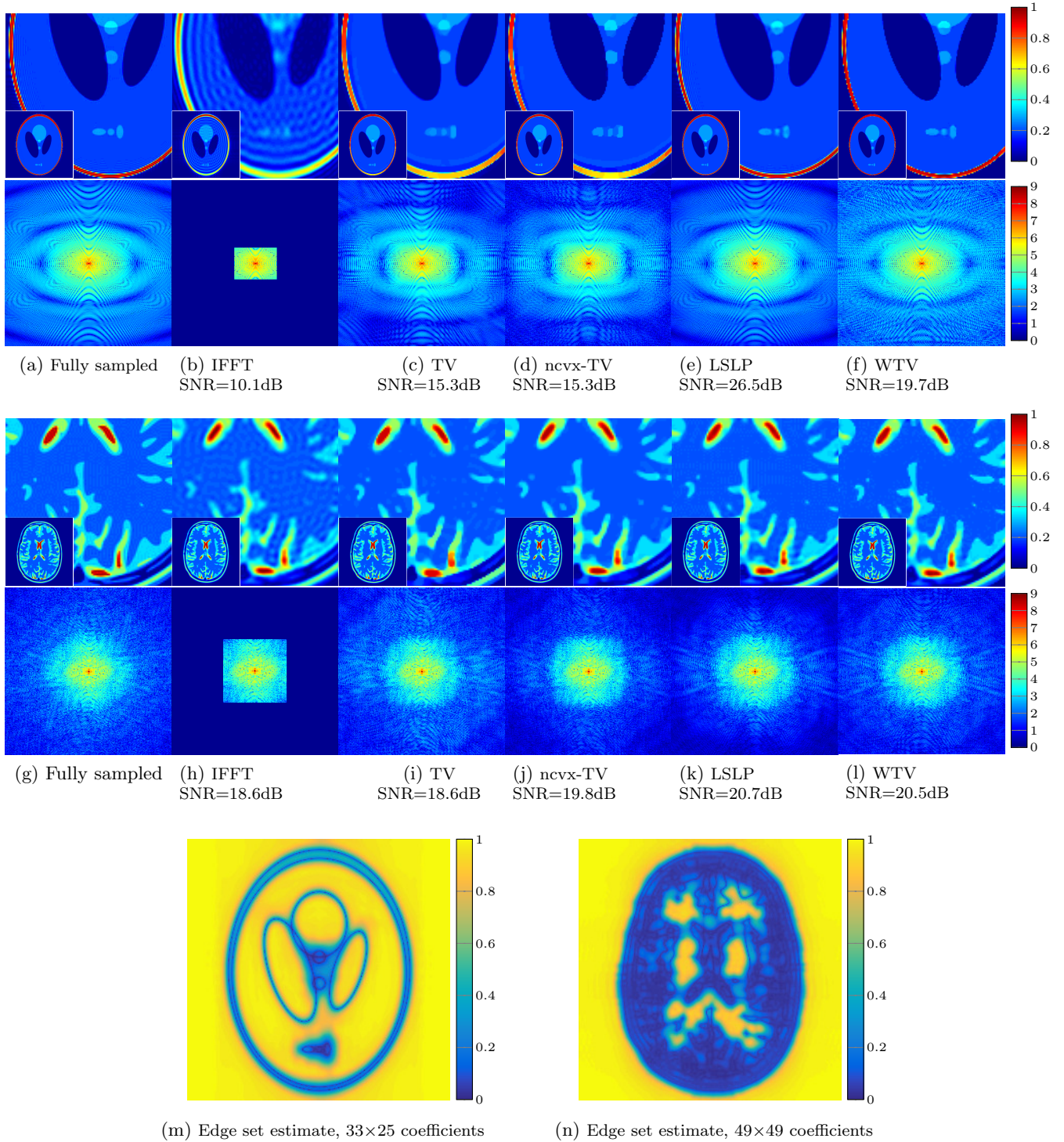
(a) Fully sampled   (b) IFFT   (c) TV   (d) ncvx-TV   (e) LSLP   (f) WTV

SNR=10.1dB   SNR=15.3dB   SNR=15.3dB   SNR=26.5dB   SNR=19.7dB

(g) Fully sampled   (h) IFFT   (i) TV   (j) ncvx-TV   (k) LSLP   (l) WTV

SNR=18.6dB   SNR=18.6dB   SNR=19.8dB   SNR=20.7dB   SNR=20.5dB

(m) Edge set estimate, 33×25 coefficients   (n) Edge set estimate, 49×49 coefficients

Figure 6: *Super-resolution of piecewise constant phantoms from ideal Fourier samples.* Recovery of Shepp-Logan phantom from 65×49 ideal low-pass Fourier samples, and a brain phantom reconstructed onto 256×256 spatial grid reconstructed from 97×97 samples. In (a)-(l), spatial domain is shown above, and Fourier domain below (in log scale). Shown in (a)&(g) are the fully sampled images, (b)&(h) is the IFFT of the zero-padded samples, (c)&(i) is a standard TV regularized recovery, (d)&(j) is the non-convex TV used in [10], (e)&(k) is the recovery using the proposed least squares linear prediction (LSLP) method, and (f)&(l) the recovery using the proposed weighted-TV (WTV) formulation. Note the the LSLP more accurately extrapolates the true Fourier coefficients of the image. The sum-of-squares polynomial used as the edge set estimate for the proposed schemes in shown (m) and (n).

ℓ^p, $p < 1$, penalty in place of the ℓ^1 penalty, and fix $p = 1/2$ in our experiments. For all recovery schemes, we tune regularization parameters to optimize the signal-to-noise ratio (SNR), defined as SNR = $20 \log_{10}(\|\boldsymbol{x}_0\|_2/\|\boldsymbol{x} - \boldsymbol{x}_0\|_2)$ where $\boldsymbol{x}$ is the reconstructed image, and $\boldsymbol{x}_0$ is the ground truth image.

(a) Fully-sampled

(b) Edge set estimate
(50×50 coefficients)

(c) Fully-sampled (top)
Sampling mask (mid)
Fourier domain (bottom)

(d) ncvx-TV
SNR=13.9dB
SSIM=0.744

(e) Proposed, LSLP
SNR=14.3dB
SSIM=0.788
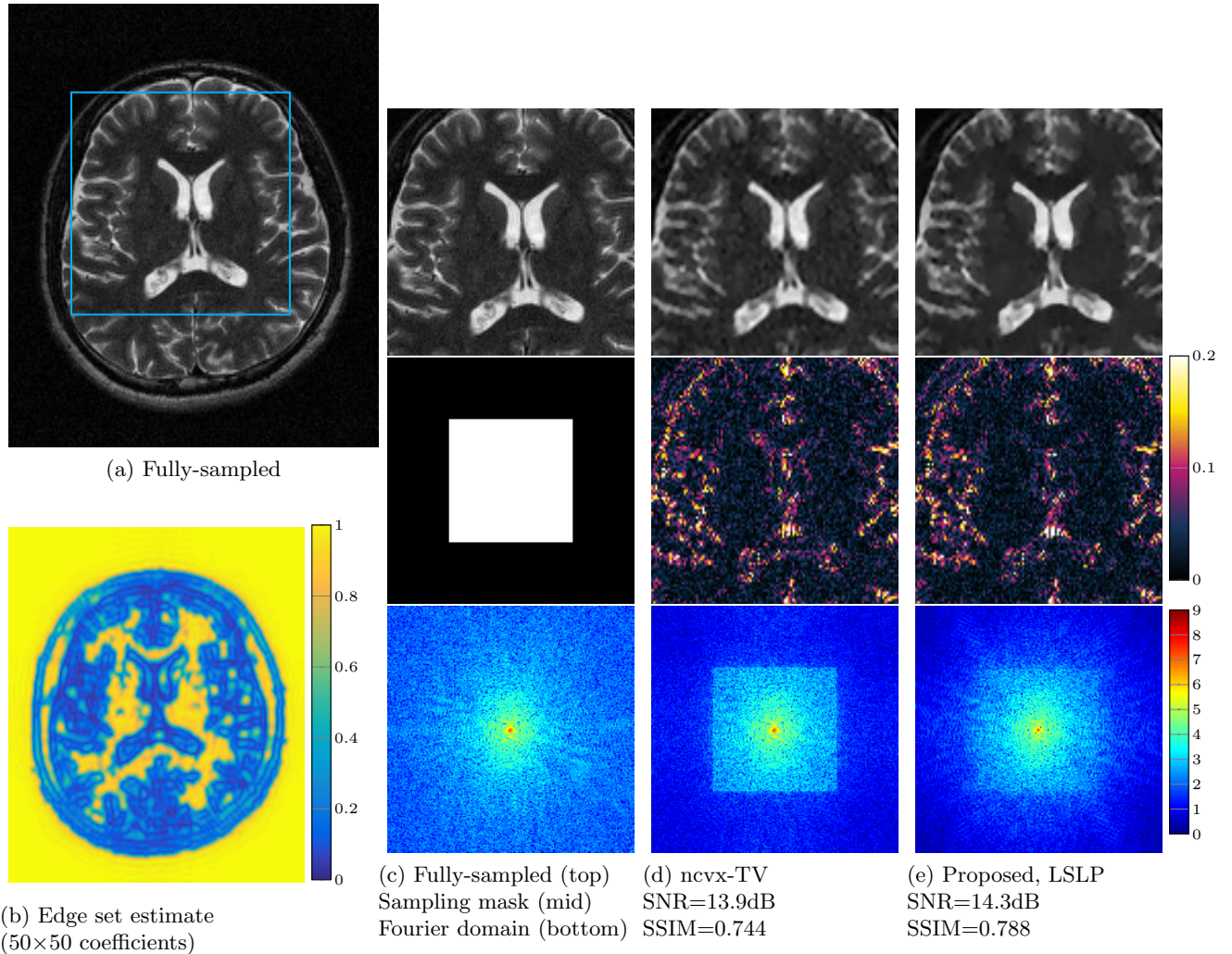
Figure 7: Super-resolution of a single-coil MR brain image from retrospectively undersampled data. Reconstruction from 100×100 center Fourier samples taken from a 200×200 acquisition reconstructed onto a 200×200 spatial grid. The original fully sampled data is shown in (a), and zoomed in (c). A nonconvex TV regularized recovery is shown in (d), and the proposed LSLP recovery is shown in (e) which uses the sum-of-squares polynomial shown in (b) as an edge set estimate. Error images are shown in middle row of (d) and (e), and the Fourier domain recovery in log scale is shown in the bottom row. Note the proposed scheme more faithfully extrapolates the true Fourier data.

The results of this first experiment are shown in Fig. 6. Our proposed scheme is able to recover the Shepp-Logan phantom with high accuracy from as few as 65×49 low-pass Fourier samples, corresponding to a 20-fold undersampling. A TV-regularized recovery from the same samples shows significant edge blurring, broadening of image features, and ringing-like artifacts in piecewise constant regions. We observe the non-convex TV approach is better able to reduce artifacts in the constant regions of the image, but still distorts the boundaries of the piecewise constant regions. We see similar benefits in the case of a brain phantom having a more complicated edge set geometry. Visually, the WTV reconstruction appears to show some advantages over LSLP in this case, in that it is better able to recovery some of the finer image features and suppress ringing artifacts. However, the LSLP recovery more faithfully recovers the

true Fourier coefficients, as measured by the SNR.

We also performed another experiment validating the method on real MR data; see Figure 7. The dataset was obtained from a fully sampled 4-coil acquisition, which we compressed into a single virtual coil using an SVD-based technique [68]. We then retrospectively collected low-pass Fourier samples from the single virtual coil. To simulate a noisy acquisition, we also added i.i.d. complex white Gaussian noise to the low-pass Fourier samples, such that the resulting SNR of the samples was approximately 30 dB.

We observed that the data from the single virtual coil was complex-valued in image domain with smoothly varying phase over its support, as is typical with MR data. Applying our edge-set estimation procedure directly on the unprocessed low-pass samples gave poor results, since our approach assumes the underlying image is piecewise constant. Therefore, prior to estimating the edge-set, we performed a phase correction step. This was done by taking the inverse DFT of the zero-padded low-pass Fourier samples, canceling out the phase in image domain, and passing back to Fourier domain. We note this pre-processing step only requires access to the low-pass samples. To ensure fair comparisons, when computing image metrics we compared the magnitude images of the reconstructions against the magnitude image obtained from the fully sampled virtual coil.

Additional noise and model-mismatch in real MR data makes estimating the true model order of the edge-set challenging. As a rule of thumb, we chose the annihilating filter dimensions to be half the size of the low-resolution input. We then performed the annihilating subspace denoising step (35) with the rank threshold r set to half the number of the coefficients in the filter. Empirically, we find this procedure is still able to accurately recover the significant edges in the image, as shown in Fig. 7(b). We then pass the "denoised" Fourier samples that result from (35) as the input to the second stage of our recovery scheme.

For the real MR data we find our proposed LSLP approach to be superior to our WTV approach, hence we restrict our comparison to this case. In Fig. 7 we compare the proposed approach with ncvx-TV, which was found to outperform standard TV (not shown). In addition to SNR, we also use the structural similarity index (SSIM) as a measure of the perceptual-quality [64]. We observe that on this dataset our proposed LSLP scheme shows some advantages over ncvx-TV in terms of SNR and SSIM. Visually, the LSLP recovery shows fewer noise artifacts in the constant regions of the image, and better preservation of edge features. Additionally, the Fourier domain images in Fig. 7 indicates the LSLP approach better extrapolates the low-pass Fourier data beyond the sampling region.

Finally, we remark that the computation time of the proposed scheme is primarily dominated by the cost in computing the SVD of the annihilation matrix in the edge-recovery stage. The dimensions of the annihilation matrix grow with the size of the annihilation filter and the number of low-resolution samples. For example, given an 128×128 array of low-resolution samples and a filter size of 64×64, the annihilation matrix has dimensions 8450×4096. The second stage of the recovery using the LSLP and WTV approaches were similar in run time to a standard TV regularized recovery, and a MATLAB implementation running on CPU took under a minute for the datasets considered here.

7   Discussion

In this work, we developed sampling guarantees for the unique recovery of a continuous domain piecewise constant image from few of its low-pass Fourier samples, assuming the edge set of the image is described by the zero set of a trigonometric polynomial. This is achieved by showing the distributional derivatives of such an image satisfy a linear annihilation relation in Fourier domain. One of our main contributions is to prove that, under ideal conditions, unique recovery of the image is possible from the annihilation relation.

We also proposed a two step algorithm for the recovery of images, which extrapolates the known Fourier samples of the image in Fourier domain. We argue such a recovery scheme deserves to be called

"off-the-grid", since we recover the exact Fourier coefficients of the underlying continuous domain image, and at no point do we need to discretize in spatial domain. In the first step of the algorithm, we estimate a collection of annihilating filters from the low-pass samples, which jointly encode the edge set of the image. In the second step, we perform a least squares linear prediction using these filters to extrapolate the signal Fourier domain. We show the proposed scheme has advantages over traditional single-image super-resolution schemes in MRI in its ability to suppress noise artifacts and preserve strong image edges.

While our experiments focused on the problem of super-resolution MRI, we note that this work has potentially wide-ranging applications beyond those presented here. For instance, the first stage our algorithm estimates a continuous domain representation of the edge set of an image, which can be useful for image segmentation. Additionally, the super-resolution problem we considered is not limited to MRI, and will work in any setting where one has a good estimate of the low-pass Fourier samples of image, such as in the image zoom problem studied in [47].

In future work, we plan to resolve the conjectures put forth in §2.4, which would extend our sampling theorems to piecewise constant images with edge set given by an arbitrary trigonometric curve, not just those that are non-intersecting. We also hope to give further insight into the robustness of this scheme with respect to noise and model mismatch, especially with regard to the rank properties of the annihilation matrix.

Finally, we observe that most of the theory presented here extends directly to higher dimensions. For example, in three dimensions we could consider piecewise constant signals whose gradient is supported on a two-dimensional surface given as the zero set of a trigonometric polynomial in three variables. However, the computational burden in this case will be much more significant. For instance, the edge set estimation step would require the singular value decomposition of an extremely large matrix, which we may not even be able to hold in memory. Therefore a goal of future research is to come up with efficient algorithms to further enable the recovery of higher dimensional signals.

A   Algebraic Properties of Trigonometric Polynomials and Curves

Here we develop the algebraic properties of trigonometric polynomials and trigonometric curves in some detail. We will slightly expand our definition of a *trigonometric polynomial* (9) to include those with frequencies $\boldsymbol{k}$ on a shifted lattice $\mathbb{Z}^2 + \boldsymbol{\sigma}$ for some fixed $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ with $\sigma_i \in \{0, \frac{1}{2}\}$. This is to ensure the results below extend to trigonometric polynomials having an even number of coefficients in one or both dimensions.

We define the *degree* of μ, denoted as $deg(\mu) = (K, L)$, to be the linear dimensions of the smallest rectangle R whose closure contains the frequency support set, and we say μ is *centered* if the frequency support rectangle R is symmetric about origin $(k, l) = (0, 0)$. For example,

$$\mu_1(x, y) = 4\sin(2\pi x)\cos(4\pi y)$$
$$= je^{j2\pi(-x-2y)} + je^{j2\pi(x-2y)} - je^{j2\pi(-x+2y)} - je^{j2\pi(x+2y)}$$

has frequency support contained in the rectangle $[-1, 1] \times [-2, 2]$, hence is centered with degree $(2, 4)$. As another example,

$$\mu_2(x, y) = 2e^{-j2\pi x} + 1 + e^{j2\pi x} - e^{j2\pi(x+y)}$$

has frequency support contained in $[-1, 1] \times [0, 1]$, hence degree $(2, 1)$, but is *not* centered. However, we can make μ_2 centered by multiplying by a phase factor $\widetilde{\mu}_2(x, y) = e^{-j\pi y}\mu_2(x, y)$, which shifts the frequency support to $[-1, 1] \times [-\frac{1}{2}, \frac{1}{2}]$. Note μ_2 and $\widetilde{\mu}_2$ have the same degree and the same zero set, since these properties are invariant under multiplication by a phase factor.

For any trigonometric polynomial μ with $deg(\mu) = (K, L)$ we can associate a unique complex polynomial $\mathcal{P}[\mu]$ in $\mathbb{C}[z, w]$ by making the substitutions $e^{j2\pi x} \mapsto z, e^{j2\pi y} \mapsto w$, and multiplying by powers

of z and w such that $\mathcal{P}[\mu]$ has degree K as a polynomial in z, and degree L as a polynomial in w. Similarly, we may define an inverse mapping acting on polynomials $q(z, w) \in \mathbb{C}[z, w]$ as $\mathcal{P}^{-1}[q](x, y) := e^{-j2\pi Sx} e^{-j2\pi Ty} q(e^{j2\pi x}, e^{j2\pi y})$, where S, T are chosen so that the resulting trigonometric polynomial is centered. For example, using the same μ_1 and μ_2 as above, we have

$$\mathcal{P}[\mu_1](z, w) = zw^2(jz^{-1}w^{-2} + jzw^{-2} - jz^{-1}w^2 - zw^2) = j(1 + z^2 - w^4 - z^2 w^4),$$
$$\mathcal{P}[\mu_2](z, w) = z(2z^{-1} + 1 + z - zw) = 2 + z + z^2 - z^2 w$$

and $\mathcal{P}^{-1}[\mathcal{P}[\mu_1]](x, y) = \mu_1(x, y)$ whereas $\mathcal{P}^{-1}[\mathcal{P}[\mu_2]](x, y) = e^{-j\pi x}\mu_2(x, y)$.

We also define $E(x, y) = (e^{j2\pi x}, e^{j2\pi y})$ for all $(x, y) \in [0, 1]^2$, which is a continuous bijective mapping of $[0, 1)^2$ onto the complex unit torus $\mathbb{CT}^2 = \{(z, w) \in \mathbb{C}^2 : |z| = |w| = 1\}$. Observe that $\mu(x, y) = 0$ if and only if $\mathcal{P}[\mu](z, w) = 0$ for $(z, w) = E(x, y)$, so that $\mu = 0$ on $C \subseteq [0, 1)^2$ if and only if $\mathcal{P}[\mu] = 0$ on $E(C)$. This shows we may study the algebraic properties of trigonometric polynomials and their zero sets by studying their corresponding complex polynomials under the mapping $\mathcal{P}$. Accordingly, for trigonometric polynomials μ and ν we will say μ *divides* ν, denoted $\mu \mid \nu$, when $\mathcal{P}[\mu]$ divides $\mathcal{P}[\nu]$ as polynomials in $\mathbb{C}[z, w]$, or equivalently, if $\nu = \mu\gamma$ for some trigonometric polynomial γ; μ and ν are said to have a *common factor* if there is a γ with positive degree such that $\gamma \mid \mu$ and $\gamma \mid \nu$; and we say μ is *irreducible* when $\mathcal{P}[\mu]$ is irreducible as a polynomial in $\mathbb{C}[z, w]$.

We will need the following well-known result regarding the maximum number of intersections of algebraic curves (see e.g. [50]):

Theorem 5 (Bézout's Theorem). *Let p_1 and p_2 be non-constant polynomials in $\mathbb{C}[z, w]$ with $deg(p_1) = d_1$ and $deg(p_2) = d_2$. If p_1 and p_2 have no common factor then the system of equations $p_1(z, w) = p_2(z, w) = 0$ has at most $d_1 d_2$ solutions.*

One important consequence of Bézout's theorem is that if $p(z, w)$ is irreducible and $q(z, w)$ is any other polynomial such that $p(z, w) = q(z, w) = 0$ has infinitely many solutions, then p must divide q.

We can readily translate these results to trigonometric polynomials μ and ν by mapping to the polynomials $p = \mathcal{P}[\mu]$ and $q = \mathcal{P}[\nu]$, and mapping back again. If $deg(\mu) = (K, L)$, then $p = \mathcal{P}[\mu]$ has total degree at most $K + L$ as a polynomial in $\mathbb{C}[z, w]$, and so we have the following corollary that we will also call "Bézout's theorem":

Corollary 1. *Let μ_1 and μ_2 be trigonometric polynomials with $deg(\mu_1) = (K_1, L_1)$ and $deg(\mu_2) = (K_2, L_2)$. If μ_1 and μ_2 have no common factor then the system of equations $\mu_1(x, y) = \mu_2(x, y) = 0$ has at most $(K_1 + L_1)(K_2 + L_2)$ solutions in $[0, 1)^2$.*

Consequently, if the zero sets of trigonometric polynomials μ and ν have infinite intersection, and μ is irreducible, this implies $\mu \mid \nu$.

A.1   Existence of a minimal polynomial (Proof of Proposition 2)

Here we prove a slightly stronger version of Proposition 2, which shows the minimal polynomial satisfies an additional unique divisibility property. Recall that we say the zero set of a trigonometric polynomial $C = \{\mu = 0\}$ is a *trigonometric curve* if C infinite and has no isolated points, i.e. if every connected component of C is one-dimensional.

Proposition 9. *For every trigonometric curve C there is unique (up to scaling) real-valued trigonometric polynomial μ_0 such that $C = \{\mu_0 = 0\}$ and for any other trigonometric polynomial μ with $\mu = 0$ on C we have $deg(\mu_0) \le deg(\mu)$ componentwise, and $\mu_0 \mid \mu$.*

Proof. Let μ be any trigonometric polynomial with $C = \{\mu = 0\}$. Set $p = \mathcal{P}[\mu] \in \mathbb{C}[z, w]$. Then p has a unique factorization (up to scaling) into irreducible factors: $p = p_1^{n_1} \cdots p_r^{n_r}$. Create a polynomial $p_0 = p_{i_1} \cdots p_{i_k}$, $k \leq r$, omitting any factors p_i whose zero sets do not intersect $\mathbb{CT}^2$, or intersect at only finitely many points. Define $\mu_0 = \mathcal{P}^{-1}[p_0]$, which by construction satisfies $\mu_0 = 0$ on C/A where A is possibly some finite set. In fact, it must be the case that $\mu_0 = 0$ on all of C, since μ_0 is continuous and C contains no isolated points by assumption. We will show μ_0 is the desired minimal polynomial.

Note the algebraic curve $X = \{p_0 = 0\}$ is the union of irreducible curves $X_j = \{p_{i_j} = 0\} \subseteq \mathbb{C}^2$. Define $C_j = E^{-1}(X_j \cap \mathbb{CT}^2)$, then likewise we have a decomposition of C as the union of the curves C_j. If ν is any other trigonometric polynomial with $C = \{\nu = 0\}$, then $\nu = 0$ on each C_j, and for $q = P[\nu]$ we have $q = 0$ on each infinite set $E(C_j) = X_j \cap \mathbb{CT}^2$. Therefore by Bézout's theorem, $q = \mathcal{P}[\nu]$ must factor as $q = p_{i_1}^{m_1} \cdots p_{i_k}^{m_k} r$ for some $r \in \mathbb{C}[z, w]$, hence $deg(p_0) \leq deg(q)$ and $p_0 \mid q$, which implies $deg(\mu_0) \leq deg(\nu)$ and $\mu_0 \mid \nu$.

We now show μ_0 is real-valued under an appropriate choice of scaling. Without loss of generality assume $Re[\mu_0] \neq 0$. Since $Re[\mu_0] = \frac{1}{2}(\mu_0 + \mu_0^*)$ and μ_0 is centered, we have $deg(Re[\mu_0]) \leq deg(\mu_0)$. Also, since $\{\mu_0 = 0\} \subseteq \{Re[\mu_0] = 0\}$ we have $\mu_0 | Re[\mu_0]$ and so $Re[\mu_0] = c \cdot \mu_0$ for some $c \in \mathbb{C}$ by the uniqueness of μ_0. Therefore $\frac{1}{c}\mu_0$ is real-valued.

Finally, to show uniqueness suppose ν_0 is another centered trigonometric polynomial satisfying $C = \{\nu_0 = 0\}$ and $deg(\nu_0) = deg(\mu_0)$. Then $\mu_0 \mid \nu_0$, and so by degree constraints, $\nu_0 = c \cdot \mu_0$ for some $c \in \mathbb{C}$. $\square$

A.2 Additional properties of minimal polynomials

Our uniqueness theorems also make use of the following result, which shows the zero set of the minimal polynomial and the zero set of its gradient have at most a finite intersection:

Proposition 10. *Let C be a trigonometric curve with minimal polynomial μ. Then $\nabla\mu = 0$ for at most finitely many points on C.*

Proof. Note the partial derivatives $\partial_x\mu$ and $\partial_y\mu$ are trigonometric polynomials with $deg(\partial_x\mu) \leq deg(\mu)$ and $deg(\partial_y\mu) \leq deg(\mu)$. First consider the case where μ is irreducible. If $\{\nabla\mu = 0\}$ and C have an infinite intersection, then by Bézout's theorem, μ must divide $\partial_x\mu$ and $\partial_y\mu$, and so by degree considerations, $\partial_x\mu = A\,\mu$ and $\partial_y\mu = B\,\mu$ for some non-zero constants A, B. By comparing coefficients, we see this is only possible if μ is the zero polynomial.

The general case follows by induction on the number of irreducible factors in μ: Suppose $\mu = \mu_1 \cdots \mu_{n-1}\mu_n$ where each μ_i is a distinct irreducible factor (that μ must have this form follows from the proof of Prop. 2). Let $\nu = \mu_1 \cdots \mu_{n-1}$, then we may write $C = C_1 \cup C_2$ where $C_1 = \{\nu = 0\}$ and $C_2 = \{\mu_n = 0\}$. By the product rule $\nabla\mu = \mu_n\nabla\nu + \nu\nabla\mu_n$. Therefore, $\nabla\mu = 0$ on C_1 if only if $\mu_n = 0$ or $\nabla\nu = 0$ on C_1. However, $\{\mu_n = 0\} \cap C_1$ is finite by Bézout's theorem, and $\{\nabla\nu = 0\} \cap C_1$ is finite by inductive hypothesis, hence $\{\nabla\mu = 0\} \cap C_1$ is finite. A similar argument shows $\{\nabla\mu = 0\} \cap C_2$ is finite, which proves the induction. $\square$

The following proposition shows that minimal polynomials must always change sign across an infinite segment of its zero set.

Proposition 11. *Let C be a trigonometric curve with minimal polynomial μ. If U_1 and U_2 are distinct connected components of $[0, 1]^2/C$ whose boundaries share an infinite segment of C, then the sign of μ must be different on U_1 and U_2.*

Proof. Let U_1 and U_2 be any two connected components of $[0, 1]^2/C$ whose boundaries share an infinite segment $S \subseteq C$, and suppose, without loss of generality, μ is positive on both U_1 and U_2. Then for all points in S, μ has a local minimum, and since μ is smooth this implies $\nabla\mu = 0$ on all of S, which violates Proposition 10. Therefore μ must have different signs on U_1 and U_2. $\square$

Note that this result puts certain topological constraints on the geometry of trigonometric curves. For instance, if C is any trigonometric curve, and $\boldsymbol{r}_0 \in C$, then in any sufficiently small neighborhood of r_0, the set $C/\{\boldsymbol{r}_0\}$ must consist of an *even* number of branches, since otherwise the minimal polynomial would violate the parity rule in Prop. 11. This implies, for example, that within the region $[0,1]^2$ trigonometric curves must always be closed. It also rules out such features such as triple junctions, where locally three curve segments terminate at a common intersection point.

A.3 Bound on Connected Components (Proof of Proposition 3)

To bound the number of connected components in the complement of a trigonometric curve, we make use of Bernstein's theorem, also known as the BKK bound, which is a well-known refinement of Bézout's theorem (see, e.g., [58, 22]). Bernstein's theorem relates the maximum number of isolated zeros of a Laurent polynomial system to the geometry of the coefficient supports. Here we use the notation $conv(\Lambda)$ to denote the convex hull of an index set $\Lambda \subset \mathbb{Z}^2$ treated as a subset of $\mathbb{R}^2$. For any two sets $P_1, P_2 \subset \mathbb{R}^2$ we also use $\mathcal{M}(P_1, P_2)$ to denote their *mixed area*:

$$\mathcal{M}(P_1, P_2) := area(P_1 + P_2) - area(P_1) - area(P_2)$$

where $area(\cdot)$ is the usual Euclidean area and $P_1 + P_2$ is the Minkowski sum, $\{p_1 + p_2 : p_1 \in P_1, p_2 \in P_2\}$.

Theorem 6 (Bernstein's Theorem). *Let $p_1(z,w)$ and $p_2(z,w)$ be Laurent polynomials with coefficient supports Λ_1 and Λ_2, and let $P_1 = conv(\Lambda_1)$ and $P_2 = conv(\Lambda_1)$. The number of isolated solutions of system $p_1(z,w) = p_2(z,w) = 0$ in $(\mathbb{C}/\{0\})^2$ is at most $\mathcal{M}(P_1, P_2)$.*

Since a trigonometric polynomial is simply the restriction of a Laurent polynomial to the unit complex torus, we can use this result to bound the number of isolated zeros of a system of trigonometric polynomials. In particular, using that fact that the mixed area satisfies $\mathcal{M}(P_1, P_2) \leq \mathcal{M}(R_1, R_2)$ if $P_i \subseteq R_i$, $i = 1, 2$ ([22], pg. 117), we have the following direct corollary:

Corollary 2. *Let μ_1 and μ_2 be trigonometric polynomials with $deg(\mu_1) \leq (K_1, L_1)$ and $deg(\mu_2) \leq (K_2, L_2)$. The number of isolated solutions of $\mu_1(x,y) = \mu_2(x,y) = 0$ in $[0,1)^2$ is at most $K_1 L_2 + K_2 L_1$.*

Now we prove Proposition 3:

Let C be a trigonometric curve of degree (K, L), i.e. the minimal polynomial μ for C has degree (K, L), and let $U_1, ..., U_n$ be the connected components of $[0,1]^2/C$. For each $i = 1, ..., n$, μ is continuous with $\mu \neq 0$ on U_i, $\mu = 0$ on ∂U, and $\overline{U} = U \cup \partial U$ is compact, therefore μ must attain its maximum or minimum at some point $\boldsymbol{r}_i$ in the interior U, where $\nabla \mu(\boldsymbol{r}_i) = 0$. Hence, the number of components n is bounded by the cardinality of $\{\nabla \mu = 0\} = \{\partial_x \mu = 0\} \cap \{\partial_y \mu = 0\}$. When $\{\nabla \mu = 0\}$ is finite, the result $n \leq 2KL$ follows from Corollary 2 since $deg(\partial_x \mu), deg(\partial_y \mu) \leq (K, L)$.

If $\{\nabla \mu = 0\}$ is not finite, i.e. $\{\partial_x \mu = 0\}$ and $\{\partial_y \mu = 0\}$ share a common component, we may obtain the same bound by a standard perturbation argument, which we summarize here (see, e.g., the appendix [53] for further details). Let $\boldsymbol{v} \in \mathbb{R}^2$ be a sufficiently small vector. By continuity, the function $\widetilde{\mu}(\boldsymbol{r}) = \mu(\boldsymbol{r}) - \langle \boldsymbol{r}, \boldsymbol{v} \rangle$ also attains a maximum or minimum in each U_i, and $\nabla \widetilde{\mu} = \nabla \mu - \boldsymbol{v}$ vanishes at these points, which shows n is also bounded by the cardinality of $\{\nabla \mu = \boldsymbol{v}\}$. Finally, note that since $\nabla \mu$ is a smooth map whose domain and range have the same dimension, for a generic vector $\boldsymbol{v}$ the set $\{\nabla \mu = \boldsymbol{v}\}$ is finite by the regular level set theorem (Cor. 8.10, [36]), and the result again follows by Corollary 2. $\square$

We note that the bound in Proposition 3 appears to be an overestimate. Empirically, we find that the maximum number of connected components of the complement to a trigonometric curve of degree (K, L) never exceeds KL (in the topology of the torus $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$). Moreover, this bound is only obtained in artificial cases, such as when the minimal polynomial is $\mu(x,y) = \sin(\pi K x) \sin(\pi L y)$, whose zero set is the union of K vertical lines and L horizontal lines.

B   Proofs of Theorems

B.1   Proof of Theorem 1

Suppose $f = 1_U$, with U simply connected and $\partial U = \{\mu = 0\}$. where μ is the minimal polynomial having coefficients $(c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$. Let $(d[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ be any non-trivial set of coefficients satisfying

$$\sum_{\boldsymbol{k} \in \Lambda} d[\boldsymbol{k}] \widehat{\nabla f}[\boldsymbol{\ell} - \boldsymbol{k}] = 0, \quad \text{for all } \boldsymbol{\ell} \in 2\Lambda. \tag{46}$$

We will show the coefficients $d[\boldsymbol{k}]$ must be a scalar multiple of $c[\boldsymbol{k}]$.

Let η denote the trigonometric polynomial having coefficients $(d[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$. Observe that the sums in (46) represent the Fourier coefficients of $\eta \nabla f$ evaluated at $\boldsymbol{\ell} \in 2\Lambda$. Therefore, (46) can be expressed in spatial domain as

$$\mathcal{P}_{2\Lambda}(\eta \nabla f) = 0$$

where $\mathcal{P}_{2\Lambda}$ is Fourier projection onto the index set 2Λ. If we let $\boldsymbol{\varphi}$ be any smooth test field, then by duality

$$\langle \mathcal{P}_{2\Lambda}(\eta \nabla f), \boldsymbol{\varphi} \rangle = \langle \eta \nabla f, \mathcal{P}_{2\Lambda} \boldsymbol{\varphi} \rangle = 0,$$

or, equivalently,

$$\langle \eta \nabla f, \boldsymbol{\psi} \rangle = 0, \text{ for all } \boldsymbol{\psi} \in B_{2\Lambda}^2,$$

where $B_{2\Lambda}^2$ denotes the space all smooth vector fields $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ whose components φ_1, φ_2 are bandlimited to 2Λ. Since $f = 1_U$, using the distributional characterization of ∇f in (14) we have

$$\oint_{\partial U} \eta \, (\boldsymbol{\psi} \cdot \boldsymbol{n}) \, ds = 0, \text{ for all } \boldsymbol{\psi} \in B_{2\Lambda}^2. \tag{47}$$

Because ∂U is a level set of μ, the outward normal $\boldsymbol{n}$ is given pointwise by $\boldsymbol{n}(\boldsymbol{r}) = \pm \nabla \mu(\boldsymbol{r})/|\nabla \mu(\boldsymbol{r})|$ for all $\boldsymbol{r} \in \partial U$, provided $\nabla \mu(\boldsymbol{r}) \neq 0$. However, by Prop. 10, the set of points for which $\nabla \mu = 0$ on ∂U is at most finite, hence of measure zero, so the formula $\boldsymbol{n}(\boldsymbol{r}) = \pm \nabla \mu(\boldsymbol{r})/|\nabla \mu(\boldsymbol{r})|$ holds almost everywhere on ∂U. Now, without loss of generality, assume $\mu < 0$ on U. Since μ is a minimal polynomial, by Prop. 11 it must change sign across the any infinite segment of ∂U, hence $\mu > 0$ on U^C, so that $\boldsymbol{n}(\boldsymbol{r}) = \nabla \mu(\boldsymbol{r})/|\nabla \mu(\boldsymbol{r})|$ for all $\boldsymbol{r} \in \partial U$. Setting $\boldsymbol{\psi} = \eta^* \nabla \mu \in B_{2\Lambda}^2$, from (47) we have

$$\oint_{\partial U} |\eta|^2 \, |\nabla \mu| \, ds = 0,$$

which is possible if and only if $\eta = 0$ on ∂U since $|\nabla \mu(\boldsymbol{r})| = 0$ for at most finitely many $\boldsymbol{r} \in \partial U$. Therefore η must be a scalar multiple of μ by the uniqueness of minimal polynomials.

B.2   Proof of Theorems 2 and 3

Assume $(d[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$ is another solution of the system, and let η denote the associated trigonometric polynomial. Following the same steps as in the proof of Theorem 1, we may show

$$\langle \eta \, \nabla f, \boldsymbol{\psi} \rangle = 0, \text{ for all } \boldsymbol{\psi} \in B_{2\Lambda}^2,$$

Since $f = \sum_{i=1}^n a_i 1_{U_i}$, from (14) this gives

$$\sum_{i=1}^n a_i \oint_{\partial U_i} \eta \, (\boldsymbol{\psi} \cdot \boldsymbol{n}) \, ds = 0, \text{ for all } \boldsymbol{\psi} \in B_{2\Lambda}^2. \tag{48}$$

28

Recall we assume $\partial U_i = \{\mu_i = 0\}$ for all $i = 1, ..., n$, with μ_i minimal, and $\mu = \mu_1 \cdots \mu_n$. Note that $\cup_i \partial U_i = \{\mu = 0\}$, and since we assumed the boundaries to be pairwise disjoint, μ is also the minimal polynomial. Define $\widetilde{\mu}_i = \mu/\mu_i$. Then by the product rule

$$\nabla\mu = \sum_{i=1}^{n} \widetilde{\mu}_i \nabla\mu_i. \tag{49}$$

Define the vector field

$$\boldsymbol{\psi}_j = \eta^* \, \widetilde{\mu}_j \nabla\mu_j \ \text{ for all } \ j = 1, ..., n. \tag{50}$$

Note the components of $\boldsymbol{\psi}_j$ are trigonometric polynomials in $B_{2\Lambda}$, hence $\boldsymbol{\psi}_j \in B_{2\Lambda}^2$ for all $j = 1, \dots, n$. Since $\widetilde{\mu}_j = 0$ on ∂U_i for all $i \neq j$, substituting $\boldsymbol{\psi}_j$ into (48) gives

$$0 = \sum_{i=1}^{n} a_i \oint_{\partial U_i} |\eta|^2 \, \widetilde{\mu}_j (\nabla\mu_j \cdot \boldsymbol{n}) \, ds = a_j \oint_{\partial U_j} |\eta|^2 \, \widetilde{\mu}_j (\nabla\mu_j \cdot \boldsymbol{n}) \, ds = a_j \oint_{\partial U_j} |\eta|^2 \, (\nabla\mu \cdot \boldsymbol{n}) \, ds.$$

where in the last step we used that $\widetilde{\mu}_j \nabla\mu_j = \nabla\mu$ on ∂U_j by (49). Since we also assume each U_i is connected, and the curves ∂U_i do not intersect, we know μ has fixed sign on U_i. Therefore the outward normal on each ∂U_j is given by $\boldsymbol{n} = \pm\nabla\mu/|\nabla\mu|$, where the choice of sign is fixed for each i. Hence we have

$$a_j \oint_{\partial U_j} |\eta|^2 \, |\nabla\mu| = 0, \ \text{ for all } j = 1, ..., n,$$

which implies η vanishes on each ∂U_j. Since μ is the minimal polynomial, this is only possible if η is a scalar multiple of μ, which proves the claim in Theorem 2.

The proof of Theorem 3 is identical to the above, except we exchange the index set 2Λ with $\Lambda + \Lambda'$. This allows us to make the same choice of $\boldsymbol{\psi}_j$ as in (50), which again implies η vanishes on each ∂U_j. Hence by Prop. 2 we have $\eta = \mu\,\gamma$, where $\gamma \in B_{\Lambda':\Lambda}$.

B.3   Proof of Proposition 5

Suppose $g \in L^1([0,1]^2)$ satisfies $\mu\nabla g = 0$ and let U be any connected component of $\{\mu = 0\}^C$. Let $\boldsymbol{\varphi}$ be any arbitrary smooth test field with support K strictly contained within U. Since $\mu = 0$ only on ∂U, and μ is smooth, the field $\boldsymbol{\psi} = \boldsymbol{\varphi}/\mu$ is smooth, and we have

$$\langle \mu\nabla g, \boldsymbol{\psi} \rangle = \langle \nabla g, \boldsymbol{\varphi} \rangle = 0.$$

which shows the distribution ∇g restricted to U is identically zero. We now show this implies g restricted to U is a constant function almost everywhere.

Let ϕ_ϵ be any smooth approximation of the identity, i.e. for all $\epsilon > 0$ define $\phi_\epsilon(\boldsymbol{r}) := \epsilon^2\phi(\epsilon\boldsymbol{r})$ where ϕ is any smooth function satisfying $\phi \geq 0$, $\int \phi(\boldsymbol{r})d\boldsymbol{r} = 1$, $supp(\phi) \subseteq \{|\boldsymbol{r}| \leq 1\}$. Let U' be any closed connected set inside U. Then there exists an $\epsilon' > 0$ such that $dist(U', \overline{U}) < \epsilon'$. Fix $\epsilon < \epsilon'$. Since $\nabla g * \phi_\epsilon$ is a smooth function, $\nabla g|_U = 0$, and $supp(\phi_\epsilon) \subseteq U$ we have

$$\nabla(g * \phi_\epsilon)(\boldsymbol{r}) = (\nabla g * \phi_\epsilon)(\boldsymbol{r}) = 0 \quad \text{for all } \boldsymbol{r} \in U'$$

which implies $g * \phi_\epsilon$ is constant on U'. By properties of approximations of identity, $g * \phi_\epsilon \to g$ pointwise almost everywhere as $\epsilon \to 0$, which shows g is the limit of constant functions on U', and therefore itself constant on U'. Since $U' \subseteq U$ was arbitrary, we may conclude g restricted to U is a constant function almost everywhere, proving the claim.

B.4 Proof of Theorem 4

Suppose g is another solution, and set $h = g - f$. We have

$$\mu \nabla h = 0, \quad \text{and} \quad \widehat{h}[\boldsymbol{k}] = 0 \text{ for all } \boldsymbol{k} \in \Gamma \supseteq \Lambda. \tag{51}$$

The proof amounts to showing that the only feasible solution to (51) is $h = 0$. By the first constraint and Prop. 5, we see that h is almost everywhere a piecewise constant function:

$$h = \sum_{i=1}^{N} b_i 1_{U_i}, \tag{52}$$

where the constants b_i are unknown, and U_i are the connected components of $\{\mu = 0\}^C$. From $\widehat{h}[\boldsymbol{k}] = 0$, $\boldsymbol{k} \in \Gamma$, we also have $\widehat{\nabla h}[\boldsymbol{k}] = 0$, $\boldsymbol{k} \in \Gamma$, and so

$$\widehat{\nabla h}[\boldsymbol{k}] = \sum_i b_i \widehat{\nabla 1_{U_i}}[\boldsymbol{k}] = \sum_i b_i \oint_{\partial U_i} e^{-j2\pi \boldsymbol{k} \cdot \boldsymbol{r}} \, \boldsymbol{n} \, ds = 0, \text{ for all } \boldsymbol{k} \in \Gamma.$$

By linearity, this implies

$$\sum_{i=1}^{N} b_i \oint_{\partial U_i} \boldsymbol{\varphi} \cdot \boldsymbol{n} \, ds = 0,$$

for all vector fields $\boldsymbol{\varphi} \in B_\Gamma^2$. The proof now follows from the same argument in the proof to Theorem 2 (with $\eta = 1$): Using the same notation as in (49), we may choose $\boldsymbol{\varphi} = \widetilde{\mu}_j \nabla \mu_j \in B_\Lambda^2 \subseteq B_\Gamma^2$, which implies

$$b_j \oint_{\partial U_j} |\nabla \mu| = 0, \text{ for all } j = 1, \dots, n,$$

and so $b_i = 0$ for all $i = 1, \dots, N$, and hence $h = 0$ proving the claim.

B.5 Proof of Proposition 6

Let μ be the minimal polynomial with coefficients $\boldsymbol{c} = (c[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda)$, and for every $\boldsymbol{\ell} \in \Lambda' : \Lambda$ define the filter $\boldsymbol{c}_{\boldsymbol{\ell}}$ supported in Λ' by

$$c_{\boldsymbol{\ell}}[\boldsymbol{k}] = \begin{cases} c[\boldsymbol{k} - \boldsymbol{\ell}] & \text{if } \boldsymbol{k} - \boldsymbol{\ell} \in \Lambda \\ 0 & \text{elsewhere,} \end{cases}$$

that is, $\boldsymbol{c}_{\boldsymbol{\ell}}$ is a shift of the minimal polynomial coefficients within the larger index set Λ. By the Fourier shift theorem, $\boldsymbol{c}_{\boldsymbol{\ell}}$ are the Fourier coefficients of $e^{j2\pi \boldsymbol{\ell} \cdot \boldsymbol{r}} \mu(\boldsymbol{r})$. Therefore, $\boldsymbol{c}_{\boldsymbol{\ell}}$ is also an annihilating filter since the zero sets of $\mu(\boldsymbol{r})$ and $e^{j2\pi \boldsymbol{\ell} \cdot \boldsymbol{r}} \mu(\boldsymbol{r})$ coincide. This implies $\mathcal{T}(\widehat{f}) \boldsymbol{c}_{\boldsymbol{\ell}} = 0$ for all $\boldsymbol{\ell} \in \Lambda' : \Lambda$. Also, note that the vectors $\boldsymbol{c}_{\boldsymbol{\ell}}$ are linearly independent because the location of their non-zero entries are distinct. Hence the nullspace of $\mathcal{T}(\widehat{f})$ has dimension at least $|\Lambda' : \Lambda|$, which gives the desired rank bound.

Finally, supposing f satisfies the conditions of Theorem 2, and $\Gamma \supseteq 2\Lambda' + \Lambda$ so that $\Gamma : \Lambda' \supseteq \Lambda' + \Lambda$. Theorem 3 shows that if $\boldsymbol{d}$ is any null space vector, then $\boldsymbol{d}$ are the Fourier coefficients of $\mu \gamma$, for some trigonometric polynomial γ having coefficients in $\Lambda' : \Lambda$, which shows that $\boldsymbol{d}$ belongs to the span of the $\boldsymbol{c}_{\boldsymbol{\ell}}$, proving the claim.

B.6 Proof of Proposition 7

Let $\boldsymbol{r} \in \{\mu = 0\}^C$. Then $\overline{\mu}(\boldsymbol{r}) = 0$ if and only if $\gamma_i(\boldsymbol{r}) = 0$ for all $i = 1, \ldots, R$, where $R = |\Lambda' : \Lambda|$. We will show the set $P = \cap_{i=1}^R \{\gamma_i = 0\}$ is at most finite. Note that if P is infinite, by Bézout's theorem, each γ_i must have a common non-constant factor ϕ, and so $\gamma_i = \rho_i \, \phi$, where each ρ_i is a trigonometric polynomial with degree strictly smaller than γ_i. This means each $\rho_i \in B_{\Lambda'}$ has coefficients supported within some index set $\Lambda_1 \subseteq \Lambda' : \Lambda$ with $\Lambda_1 \neq \Lambda' : \Lambda$. Since the coefficient vectors $\boldsymbol{d}_i$ are orthonormal, so are the μ_i as functions in $L^2([0,1]^2)$:

$$\langle \mu_i, \mu_j \rangle = \int_{[0,1]^2} \rho_i(\boldsymbol{r}) \overline{\rho_j(\boldsymbol{r})} \, |\phi(\boldsymbol{r})\mu(\boldsymbol{r})|^2 \, d\boldsymbol{r} = \delta_{i,j}.$$

where $\delta_{i,j} = 0$ if $i \neq j$ and $\delta_{i,j} = 1$ if $i = j$. However, the above weighted integral also defines an inner product in which the $\{\rho_i\}_{i=1}^R$ are mutually orthogonal, and hence linearly independent. But this is impossible, since the ρ_i are contained in a space of dimension less than R. Therefore the γ_i have no common factors, which by Bèzout's theorem implies the intersection of their zero sets is at most finite.

B.7 Proof of Proposition 8

Proof. Note that $\boldsymbol{e_r} = P_\Lambda \mathcal{F}\{\delta_{\boldsymbol{r}}\}$, where δ_r is a Dirac delta centered at $\boldsymbol{r} \in [0,1]^2$, and $\mathcal{F}$ is the (distributional) Fourier transform, and P_Λ is projection onto the index set Λ. Since each column $\boldsymbol{d}_i$ in $\boldsymbol{D}$, is a filter supported in Λ, we trivially have $P_\Lambda(\boldsymbol{d}_i) = \boldsymbol{d}_i$. Therefore, by Parseval's theorem,

$$\|\boldsymbol{D}\boldsymbol{D}^H \boldsymbol{e}(\boldsymbol{r})\|^2 = \sum_{i=1}^n |\langle \boldsymbol{e_r}, \boldsymbol{d}_i \rangle|^2 = \sum_{i=1}^n |\langle P_\Lambda \mathcal{F}\{\delta_{\boldsymbol{r}}\}, \boldsymbol{d}_i \rangle|^2$$
$$= \sum_{i=1}^n |\langle \mathcal{F}\{\delta_{\boldsymbol{r}}\}, \boldsymbol{d}_i \rangle|^2 = \sum_{i=1}^n |\langle \delta_{\boldsymbol{r}}, \mu_i \rangle|^2 = \sum_{i=1}^n |\mu_i(\boldsymbol{r})|^2.$$

$\square$

Acknowledgments

References

[1] R. G. Baraniuk. Compressive sensing. *Signal Processing Magazine, IEEE*, 24(4):118–121, July 2007.

[2] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. *Signal Processing Magazine, IEEE*, 25(2):31–40, 2008.

[3] M. Burger. A level set method for inverse problems. *Inverse problems*, 17(5):1327, 2001.

[4] M. Burger and S. J. Osher. A survey on level set methods for inverse problems and optimal design. *European Journal of Applied Mathematics*, 16(02):263–301, 2005.

[5] J. A. Cadzow. Signal enhancement-a composite property mapping algorithm. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(1):49–62, 1988.

[6] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[7] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[8] T. Chan and L. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. In *Variational and Level Set Methods in Computer Vision, 2001. Proceedings. IEEE Workshop on*, pages 161–168, 2001.

[9] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *International Symposium on Biomedical Imaging: ISBI 2009*, pages 262–265. IEEE, 2009.

[10] R. Chartrand, E. Y. Sidky, and X. Pan. Frequency extrapolation by nonconvex compressive sensing. In *International Symposium on Biomedical Imaging: ISBI 2011*, pages 1056–1060. IEEE, 2011.

[11] C. Chen, P. Marziliano, and A. C. Kot. 2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise. *Signal Processing, IEEE Transactions on*, 60(6):2851–2859, 2012.

[12] Y. Chen and Y. Chi. Robust spectral compressed sensing via structured matrix completion. *Information Theory, IEEE Transactions on*, 60(10):6576–6601, 2014.

[13] Q. Cheng and Y. Hua. A review of parametric high-resolution methods. In Y. Hua, A. Gershman, and Q. Cheng, editors, *High-resolution and robust signal processing*. CRC Press, 2003.

[14] L. Condat and A. Hirabayashi. Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements. *Sampling Theory in Signal and Image Processing*, 14(1):p–17, 2015.

[15] M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *Image Processing, IEEE Transactions on*, 14(12):2091–2106, 2005.

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

[17] O. Dorn and D. Lesselier. Level set methods for inverse scattering. *Inverse Problems*, 22(4):R67, 2006.

[18] P. L. Dragotti, M. Vetterli, and T. Blu. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix. *Signal Processing, IEEE Transactions on*, 55(5):1741–1757, 2007.

[19] R. E. Edwards. *Fourier series: a modern introduction*, volume 2. Springer Science & Business Media, 2012.

[20] R. Eslami and M. Jacob. Robust reconstruction of MRSI data using a sparse spectral model and high resolution MRI priors. *Medical Imaging, IEEE Transactions on*, 29(6):1297–1309, 2010.

[21] J. A. Fessler. Model-based image reconstruction for MRI. *Signal Processing Magazine, IEEE*, 27(4):81–89, 2010.

[22] W. Fulton. *Introduction to toric varieties*. Number 131. Princeton University Press, 1993.

[23] E. Gong, F. Huang, K. Ying, W. Wu, S. Wang, and C. Yuan. PROMISE: Parallel-imaging and compressed-sensing reconstruction of multicontrast imaging using sharable information. *Magnetic Resonance in Medicine*, 73(2):523–535, 2015.

[24] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):626–636, 2012.

[25] E. Haacke, Z.-P. Liang, and S. Izen. Superresolution reconstruction through object modeling and parameter estimation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(4):592–595, April 1989.

[26] E. M. Haacke, Z.-P. Liang, and F. E. Boada. Image reconstruction using POCS, model constraints, and linear prediction theory for the removal of phase, motion, and gibbs artifacts in magnetic resonance and ultrasound imaging. *Optical Engineering*, 29(5):555–566, 1990.

[27] E. M. Haacke, Z.-P. Liang, and S. H. Izen. Constrained reconstruction: A superresolution, optimal signal-to-noise alternative to the fourier transform in magnetic resonance imaging. *Medical Physics*, 16(3):388–397, 1989.

[28] J. P. Haldar. Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. *Transactions on Medical Imaging*, 33(3):668–681, 2014.

[29] J. P. Haldar, D. Hernando, S.-K. Song, and Z.-P. Liang. Anatomically constrained reconstruction from noisy data. *Magnetic Resonance in Medicine*, 59(4):810–818, 2008.

[30] X. Hu, D. N. Levin, P. C. Lauterbur, and T. Spraggins. SLIM: Spectral localization by imaging. *Magnetic Resonance in Medicine*, 8(3):314–322, 1988.

[31] Y. Hua. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. *Signal Processing, IEEE Transactions on*, 40(9):2267–2280, 1992.

[32] M. Jacob, Y. Bresler, V. Toronov, X. Zhang, and A. Webb. Level set algorithm for the reconstruction of functional activation in near-infrared spectroscopic imaging. *Journal of Biomedical Optics*, Jan 2006.

[33] M. Jacob, X. Zhu, A. Ebel, N. Schuff, and Z.-P. Liang. Improved model-based magnetic resonance spectroscopic imaging. *Medical Imaging, IEEE Transactions on*, 26(10):1305–1318, 2007.

[34] K. H. Jin, D. Lee, and J. C. Ye. A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix. *arXiv preprint arXiv:1504.00532*, 2015.

[35] I. Khalidov, D. Van De Ville, M. Jacob, F. Lazeyras, and M. Unser. BSLIM: Spectral localization by imaging with explicit field inhomogeneity compensation. *Medical Imaging, IEEE Transactions on*, 26(7):990–1000, 2007.

[36] J. M. Lee. *Smooth manifolds.* Springer, 2003.

[37] Z.-P. Liang, E. Haacke, and C. Thomas. High-resolution inversion of finite Fourier transform data through a localised polynomial approximation. *Inverse Problems*, 5(5):831, 1989.

[38] J. Luo, Y. Zhu, W. Li, P. Croisille, and I. E. Magnin. MRI reconstruction from 2D truncated k-space. *Journal of Magnetic Resonance Imaging*, 35(5):1196–1206, 2012.

[39] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989.

[40] I. Maravic and M. Vetterli. Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. *Signal Processing, IEEE Transactions on*, 53(8):2788–2805, 2005.

[41] P. Milanfar, G. C. Verghese, W. C. Karl, and A. S. Willsky. Reconstructing polygons from moments with connections to array processing. *Signal Processing, IEEE Transactions on*, 43(2):432–443, 1995.

[42] D. Mumford and J. Shah. Boundary detection by minimizing functionals. *Image understanding*, pages 19–43, 1988.

[43] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.

[44] G. Ongie and M. Jacob. Recovery of piecewise smooth images from few fourier samples. *Sampling Theory and Applications (SampTA)*, 2015.

[45] G. Ongie and M. Jacob. Super-resolution MRI using finite rate of innovation. *IEEE International Symposium on Biomedical Imaging: ISBI 2015.*, 2015.

[46] P. Pal and P. Vaidyanathan. A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising. *Signal Processing Letters, IEEE*, 21(6):737–741, 2014.

[47] H. Pan, T. Blu, and P. L. Dragotti. Sampling curves with finite rate of innovation. *Signal Processing, IEEE Transactions on*, 62(2), 2014.

[48] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276–280, 1986.

[49] M. Schweiger, O. Dorn, A. Zacharopoulos, I. Nissila, and S. Arridge. 3D level set reconstruction of model and experimental data in diffuse optical tomography. *Optics express*, 18(1):150–164, 2010.

[50] I. R. Shafarevich. *Basic algebraic geometry. 1. Varieties in projective space.* Springer-Verlag, Berlin, 1994.

[51] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magnetic Resonance in Medicine*, 2013.

[52] P. Shukla and P. L. Dragotti. Sampling schemes for multidimensional signals with finite rate of innovation. *Signal Processing, IEEE Transactions on*, 55(7):3670–3686, 2007.

[53] J. Solymosi and T. Tao. An incidence theorem in higher dimensions. *Discrete & Computational Geometry*, 48(2):255–280, 2012.

[54] J.-L. Starck, E. J. Candès, and D. L. Donoho. The curvelet transform for image denoising. *Image Processing, IEEE Transactions on*, 11(6):670–684, 2002.

[55] P. Stoica and R. L. Moses. *Introduction to spectral analysis*, volume 1. Prentice hall Upper Saddle River, NJ, 1997.

[56] M. Storath, A. Weinmann, and L. Demaret. Jump-sparse and sparse recovery using potts functionals. 2014.

[57] R. S. Strichartz. *A guide to distribution theory and Fourier transforms*. World Scientific, 2003.

[58] B. Sturmfels. Polynomial equations and convex polytopes. *The American Mathematical Monthly*, 105(10):907–922, 1998.

[59] G. Tang, B. N. Bhaskar, and B. Recht. Near minimax line spectral estimation. In *Information Sciences and Systems (CISS), 2013 47th Annual Conference on*, pages 1–6. IEEE, 2013.

[60] A. Tsai, A. Yezzi Jr, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *Image Processing, IEEE Transactions on*, 10(8):1169–1186, 2001.

[61] N. Vaswani and W. Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *Signal Processing, IEEE Transactions on*, 58(9):4595–4607, 2010.

[62] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International journal of computer vision*, 50(3):271–293, 2002.

[63] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *Signal Processing, IEEE Transactions on*, 50(6):1417–1428, 2002.

[64] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[65] Q.-S. Xiang. Accelerating MRI by skipped phase encoding and edge deghosting (SPEED). *Magnetic Resonance in Medicine*, 53(5):1112–1117, 2005.

[66] W. Xu, J.-F. Cai, K. V. Mishra, M. Cho, and A. Kruger. Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies. In *Information Theory and Applications Workshop (ITA), 2014*. IEEE, 2014.

[67] J. C. Ye, Y. Bresler, and P. Moulin. A self-referencing level-set method for image reconstruction from sparse Fourier samples. *International Journal of Computer Vision*, 50(3):253–270, 2002.

[68] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig. Coil compression for accelerated imaging with Cartesian sampling. *Magnetic Resonance in Medicine*, 69(2):571–582, 2013.

[69] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report*, pages 08–34, 2008.