Published in final edited form as:
SIAM J Imaging Sci. 2021 ; 14(2): 580–619. doi:10.1137/20M1340654.

Recovery of surfaces and functions in high dimensions: sampling theory and links to neural networks*

Qing Zou[†], **Mathews Jacob**[‡]

[†]Applied Mathematics and Computational Sciences, University of Iowa, Iowa City, IA 52242.

[‡]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242.

Abstract

Several imaging algorithms including patch-based image denoising, image time series recovery, and convolutional neural networks can be thought of as methods that exploit the manifold structure of signals. While the empirical performance of these algorithms is impressive, the understanding of recovery of the signals and functions that live on manifold is less understood. In this paper, we focus on the recovery of signals that live on a union of surfaces. In particular, we consider signals living on a union of smooth band-limited surfaces in high dimensions. We show that an exponential mapping transforms the data to a union of low-dimensional subspaces. Using this relation, we introduce a sampling theoretical framework for the recovery of smooth surfaces from few samples and the learning of functions living on smooth surfaces. The low-rank property of the features is used to determine the number of measurements needed to recover the surface. Moreover, the low-rank property of the features also provides an efficient approach, which resembles a neural network, for the local representation of multidimensional functions on the surface. The direct representation of such a function in high dimensions often suffers from the curse of dimensionality; the large number of parameters would translate to the need for extensive training data. The low-rank property of the features can significantly reduce the number of parameters, which makes the computational structure attractive for learning and inference from limited labeled training data.

Keywords

level set; surface recovery; function representation; image denoising; neural networks

1 Introduction

Several imaging algorithms were introduced to exploit the extensive redundancy with images to recover them from noisy and possibly undersampled measurements. For instance, several patch-based image denoising methods were introduced in the recent past. Algorithms such as non-local means perform averaging of similar patches within the image to achieve

*Parts of this work were presented at SIAM conference on applied algebraic geometry (SIAM AG 2019), at the IEEE International Symposium on Biomedical Imaging (ISBI 2020) and at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020).

zou-qing@uiowa.edu.

denoising [6]. Similar patch-based regularization strategies are used for image recovery from undersampled data [27,56]. Similar approaches are also used for the recovery of images in a time series by exploiting their non-local similarity [36,38]. The success of these methods could be attributed to the manifold assumption [11,47], which states that signals in real-world datasets (e.g. patches in images) are restricted to smooth manifolds in high dimensional spaces. In particular, the regularization penalty used in non-local methods can be viewed as the energy of the signal gradients on the patch manifold rather than in the original domain, facilitating the collective recovery of the patch manifold from noisy measurements [4]. In particular, non-local methods estimate the interpatch weights, which are used for denoising; the interpatch weights are equivalent to the manifold Laplacian, which captures the structure of the manifold. Similarly, image denoising approaches such as BM3D [8] that cluster patches, followed by PCA approximations of the cluster, can also be viewed as modeling the tangent subspaces of the patch manifold in each neighborhood. Patch dictionary based schemes, which allow the coefficients to be adapted to the specific patch, could also be viewed as tangent subspace approximation methods.

Convolutional neural networks are now emerging as very powerful alternatives for image denoising [52, 58] and image recovery [2,17]. Rather than averaging similar patches, neural networks learn how to denoise the image neighborhoods from example pairs of noisy and noise-free patches. These frameworks can be viewed as learning a multidimensional function in high dimensional patch spaces. In particular, the inputs to the network are noisy patches and the corresponding outputs are the denoised patches/pixels. We note that the learning of such functions using conventional methods will suffer from the curse of dimensionality. Specifically, large amounts of training data may be needed to learn the parameters of such a high-dimensional function, if represented using conventional methods. While the empirical performance of neural networks is impressive, the mathematical understanding of why and how they can learn complex multidimensional functions in high-dimensional spaces from relatively limited training data is still emerging. We note that the manifold assumption is also used in the CNN literature to explain the good performance of neural networks.

With the goal of understanding the above algorithms from a geometrical perspective, we consider the following conceptual problems **(a)** when can we learn and recover a manifold or surface in high dimensional space from few samples or training data, **(b)** when can we exactly learn and recover a function that lives on a surface, from few input-output examples, **(c)** can these results explain the good performance of imaging algorithms that use manifold structure. We note that many different surface models including parametric shape models [18,22], local and multi-resolution representations [34,44], and implicit level-set [21,32,43] shape representations have been used in low-dimensional settings (e.g. 2D/3D). Our main focus in this paper is on surface recovery in high-dimensional spaces with application to machine learning and learning surfaces of patches and images. The utility of the above algorithms in such high dimensional applications have not been well-studied, to the best of our knowledge; the direct extension of the low-dimensional algorithms is expected to be associated with high computational complexity. We consider implicit level set representation of the surface to deal with shapes of arbitrary topology. Specifically, we model the surface or union of surfaces as the zero level set of a multidimensional function ψ. To restrict the degrees of freedom of the surface, we consider the level set function to be bandlimited. The bandwidth of ψ can be viewed as a measure of the complexity of the surface; a more band-limited function will translate to a smoother surface. We refer the readers to our earlier works [30,37,38,59,60] for examples of 2D/3D recovery of shapes, where the above representation is used to represent and recover shapes with sharp corners and edges.

We show that under the above assumptions on the surface, a non-linear mapping of the points on the surface will live on a low-dimensional feature subspace, whose dimension depends on the complexity of the surface. Specifically, one can transform each data point to a feature vector, whose size is equal to the number of basis functions used for the surface representation. Since we use a linear combination of complex exponentials to represent the surface, the lifting in our setting is an exponential mapping. We use the low-rank property of the feature matrix to estimate the surface from few of its samples. Our sampling results show that an irreducible surface can be perfectly recovered from very few samples, whose number is dependent on the bandwidth. Our experiments in these settings show the good recovery of the surfaces from few noisy points. Our results also show that the union of irreducible surfaces can also be recovered from few samples, provided each of the irreducible components are adequately sampled. We also use a kernel low-rank algorithm to recover a surface from its noisy samples, which bears close similarity with non-local means algorithms widely used in image processing.

We also show that the low-rank property can be used to efficiently represent multidimensional functions of points living on the surface. In particular, we are only interested in the good representation of the function when the input is on or in the vicinity of the surface. We assume the functions are linear combination of the same basis functions (exponentials in our case). Since such representations are linear in the feature space, the low-rank nature of the exponential features provides an elegant approach to represent the function using considerably fewer parameters. In particular, we show that the feature vectors of a few anchor points on the surface span the space, which allows us to efficiently represent the function as the interpolation of the function values at the anchor points using a Dirichlet kernel. The significant reduction in the number of free parameters offered by this local representation makes the learning of the function from finite samples tractable. We note that the computational structure of the representation is essentially a one-layer kernel network. Note that the approximation is highly local; the true function and the local representation match only on the surface, while they may deviate significantly on points which are not on the surface. We demonstrate the preliminary utility of this network in denoising, which shows improved performance compared to some state-of-the-art methods. Here, we model the denoiser as a function $f:\mathbb{R}^{p^2} \to \mathbb{R}$ that provides a *noise-free* center pixel of a p $\times p$ noisy patch. The noisy patch is assumed to a point in p^2 dimensional space, close to the low-dimensional patch surface or union of surfaces. We also show that this framework can be used to learn a manifold, which can be viewed as the signal subspace version of the null-space based kernel low-rank algorithm considered above. In this case, the network structure is an auto-encoder.

This work is related to kernel methods, which are widely used for the approximation of functions [4,7,24]. It is well-known that an arbitrary function can be approximated using kernel methods, and the computational structure resembles a single hidden layer neural network. Our work has two key distinctions with the above approaches: (a) unlike most kernel methods that choose infinite bandwidth kernels (e.g. Gaussians), we restrict our attention to a band-limited kernel. (b) We focus on a restrictive data model, where the data samples are localized or close to the zero set of a band-limited function. These two restrictions allow us to come up with theoretical results on when such a surface can be perfectly recovered from few samples. The results also provide clues on how many training data pairs are needed to learn functions on such surfaces. We note that such sampling theoretical results are not available in kernel literature, to the best of our knowledge. This work is inspired by the recent work on algebraic varieties [31], which also considers surfaces with finite degrees of freedom. The main distinction of this paper is the novel theoretical guarantees on recovery of the surface and functions living on the surface, which go beyond the empirical results in [31]. We focus on bandlimited surfaces in this work to borrow the theoretical tools from [37,38,60]. This work extends the results in [37,38,60] in three important ways **(i).** The planar results are generalized to the high dimensional setting. **(ii).** The worst-case sampling conditions are replaced by high-probability results, which are far less conservative, and are in good agreement with experimental results. **(iii).** The sampling results are extended to the local representation of functions. While we focus on bandlimited functions to come up with theoretical bounds, the results could be generalized to arbitrary surface representations including most basis functions, such as polynomial basis functions considered in [31,53] and shift invariant representations [5,54]. We note that this work uses parametric level-set representations unlike non-parametric level-set models (e.g. [21,43] for image segmentation). The narrow-band evolution used by these approaches to manage computational complexity makes these algorithms highly vulnerable to initial guess, unlike our algorithms as illustrated in [60]. While we illustrate our algorithms in 2D/3D applications for visualization purposes, we stress that our main focus is on high-dimensional ($\gg 3$) extensions of the level set approach and generalization to shape recovery. Non-parametric and even parametric level-set methods [5, 54] will be associated with very high computational complexity in this setting without the proposed computational approaches, and has not been reported to the best of our knowledge.

1.1 Terminology and Notation

We introduce some commonly used terminologies and notations throughout the paper. We term the zero level set of a trigonometric polynomial as a surface. Usually, the lower-case Greek letters ψ, η, etc. are used to represent the trigonometric polynomials. The calligraphic letter $\mathcal{S}$ or $\mathcal{S}[\psi]$ is used to represent the zero level sets of the trigonometric polynomials and hence the surfaces. The bold lower-case letters $\mathbf{x}$ denotes the real variable in $[0,1)^n$ and sometimes the points on the surface. The indexed bold lower-case letters $\mathbf{x}_i$ represent the samples on the surface. The upper-case Greek letters $\Lambda, \Gamma \subset \mathbb{Z}^n$ are used to denote the bandwidth of the trigonometric polynomials. In other words, the upper-case Greek letters are the coefficients index set. The coefficients set is shown as $\{\mathbf{c}_\mathrm{k} : \mathbf{k} \in \Lambda\}$. The cardinality of bandwidth Λ is given by $|\Lambda|$, which will serve as a measure of the complexity of the surface.

The notation $\Gamma \ominus \Lambda$ indicates the set of all the possible uniform shifts of the set Λ within the set Γ. The specific Greek letter Φ (sometimes subscripts are used to identify the corresponding bandwidth) is used to represent the lifting map (feature map) of the point on the surface. The notation $\Phi(\mathbf{X})$ denotes the feature matrix of the sampling set $\mathbf{X}$.

1.2 Background on non-local means; reinterpretation as manifold regularization

Non-local means (NLM) methods average patches in an image based on their similarity to obtain a denoised image. In particular, they compute a weight matrix, whose entries are $\mathbf{W}_{i,j} = \exp\left(-\frac{\|\mathbf{P}_{\mathbf{r}_i}(f) - \mathbf{P}_{\mathbf{r}_j}(f)\|^2}{\sigma^2}\right)$, where $\mathbf{P}_{\mathbf{r}}$ (f) denotes a patch in the image f, centered at $\mathbf{r}$.

The smoothing approach in NLM can be viewed as the minimization problem

$$\{\mathbf{f}^*\} = \underset{f}{\operatorname{argmin}}\|\mathbf{f} - \mathbf{g}\|^2 + \eta \sum_{i=1}^{N} \sum_{i=1}^{N} \mathbf{W}_{i,j}\|\mathbf{P}_{\mathbf{r}_i}(f) - \mathbf{P}_{\mathbf{r}_j}(f)\|^2 . \tag{1}$$

This optimization problem can be viewed as the discretization of the manifold smoothness regularization strategy used in machine learning [4], which considers the recovery of a multidimensional function $\mathbf{f}(\mathbf{s})$ on a manifold $\mathscr{M}$ from its noisy samples $\mathbf{f}(\mathbf{s}_k) = \mathbf{y}_k$:

$$\{\mathbf{F}^*\} = \underset{f}{\operatorname{argmin}}\|\mathbf{f}(s_k) - \mathbf{y}_k\|^2 + \eta \int_{\mathscr{M}} \| \nabla_{\mathscr{M}}\mathbf{f}\|^2 dx . \tag{2}$$

Here, $\mathscr{M}$ is a smooth surface/manifold and $\nabla_{\mathscr{M}}$ denotes the gradient of the function on the manifold. The weight matrix $\mathbf{W}$ captures the geometry of the patch manifold in (1). Specifically, closer point pairs on $\mathscr{M}$ will have higher weights, while distant point pairs will have smaller weights. The equivalence with NLM can be seen by viewing the noisy patches as noisy samples $\mathbf{y}_k$ on the patch manifold. We note that the weighted sum is often expressed in a compact form as

$$\sum_{i=1}^{N} \sum_{i=1}^{N} \mathbf{W}_{i,j}\|\mathbf{f}(x_i) - \mathbf{f}\left(x_j\right)\|^2 = \operatorname{trace}\left(\mathbf{F}\mathbf{L}\mathbf{F}^T\right).$$

Here, $\mathbf{F} = \left[\mathrm{f}_1 \ \cdots \ \mathrm{f}_N\right]$ and $\mathbf{L}$ is the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$, which captures the structure of the manifold and $\mathbf{D}$ is a diagonal matrix $D = \operatorname{diag}(\sum_j \mathbf{W}_{i,j})$. $\mathbf{L}$ can be viewed as the discrete approximation of the Laplace-Beltrami operator on the continuous surface/manifold [4].

2 Parametric surface representation

In this work, we use the level set representation to describe a (hyper-)surface. We model a (hyper-)surface $\mathscr{S}$ in $[0,1)^n$; $n \geq 2$ as the zero level set of a function ψ:

$$\mathcal{S}[\psi] = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \psi(\mathbf{x}) = 0 \right\}. \tag{3}$$

For example, when $n = 2$, $\mathcal{S}$ is a (hyper-)surface of dimension 1, which is typically a curve. We note that the level set representation is widely used in image segmentation [21]. The normal practice in image segmentation is the non-parametric level set representation of a time-dependent evolution function ψ, which results in the PDE-driven models. Note that the initialization of these models affects the stability and the rate of convergence of the methods. So good initialization of level set functions is usually a requirement for good segmentation.

Several authors have recently proposed to represent the level set function ψ as a linear combination of basis functions $\varphi_{\mathbf{k}}(\mathbf{x})$ [5,54]. These schemes argue that the reduced number of parameters translate to fast and efficient algorithms. Besides, we do not require the good initialization in this setting. Motivated by these schemes, we represent $\psi(\mathbf{x})$ as

$$\psi(\mathbf{x}) = \sum_{\mathbf{k} \in \Lambda} \mathbf{c}_{\mathbf{k}} \varphi_{\mathbf{k}}(\mathbf{x}). \tag{4}$$

Since the level set function is the linear combination of some basis functions, we term the corresponding zero level set as parametric zero level set. We note that the surface properties would depend on the specific basis functions and will indeed decide the type of the kernel used in the algorithms in Section 4.3. We now provide some examples of parametric representations, depending on the choices of the basis functions.

2.1 Shift invariant surface representation

A popular choice for the basis functions is the shift invariant representation, where compactly supported basis functions such as B-splines are used. Specifically, the basis functions are shifted copies of a template φ, denoted

$$\varphi_{\mathbf{k}}(\mathbf{x}) = \varphi\left(\frac{\mathbf{x}}{T} - \mathbf{k}\right). \tag{5}$$

Here T is the grid spacing, which controls the degrees of freedom of the representation. The number of B-splines in the above representation is $1/(T - 1)^n$. One may also choose a multi-resolution or sparse wavelet surface representation, when the basis functions are shifted and dilated copies of a template. This approach allows the surface to have different smoothness properties at different spatial regions.

2.2 Polynomial surface representation

The surface can also be represented as a linear combination of polynomials [31]. The polynomial degree will control the degrees of freedom in this setting. This work is inspired by [31]. However, we note that the recovery of the surface and functions from few points are not considered in the polynomial setting. In addition, the stability of polynomial representations is not fully clear, which may be needed to represent complex surfaces; we note that low degree polynomials were considered in the examples in [31].

2.3 Band-limited surface representation

We assume that the surface is within $[0,1)^n$. A well-studied representation for support limited functions is the Fourier exponential basis, which is widely used in digital image processing [33,50,60], biomedical image processing [30,40,51], and geophysics [41]. The level set function can be assumed to be band-limited [30], when ψ is expressed as a Fourier series:

$$\psi(\mathbf{x}) = \sum_{\mathbf{k} \in \Lambda} \mathbf{c}_{\mathbf{k}} \exp\left(j2\pi \mathbf{k}^T \mathbf{x}\right), \quad \mathbf{x} \in [0,1)^n . \tag{6}$$

In the above representation, the set Λ denotes the bandwidth of the Fourier coefficients $\mathbf{c} = \{\mathbf{c}_{\mathbf{k}} : \mathbf{k} \in \Lambda\}$; its cardinality $|\Lambda|$ is the number of free parameters in the surface representation. We refer to Λ as the Fourier support of ψ and we note that we always choose the support to be symmetric with respect to the origin. This choice is governed by the relation of this representation with polynomials, described in the next subsection. The extension of Λ governs the degree of the polynomial.

In this work, we focus on the Fourier series representation due to its key benefits including well-developed theoretical tools, fast algorithms such as fast Fourier transform, orthogonality, and the property that $|\exp(j2\boldsymbol{\pi}\mathbf{k}^T\mathbf{x})\ | = 1$, which results in stable representations and also facilitate the theory. In this work, we mainly focus on this representation because it facilitates us to borrow the theoretical tools from our past work [37,38,60]. We note that these results may be extended to other basis sets but is beyond the scope of this work. We will now review some of the properties of this representation, which we will use in the following sections.

2.3.1 Relation of bandlimited representation with polynomials—

We also note that bandlimited representations (6) have an intimate relation with polynomials [30]. In particular, we note that one can transform the polynomial basis to an exponential one by the one-to-one mapping $v_i : [0,1) \rightarrow \{z \in \mathbb{C} : |z| = 1\}$:

$$v_i(x_i) = \exp(j2\pi x_i) = : z_i . \tag{7}$$

We will make use of this correspondence to study the properties of the zero sets of (6). With this transformation, the representation (6) simplifies to the complex polynomial denoted as $\mathscr{P}[\varphi]$, which is of the form

$$\mathscr{P}[\varphi](\mathbf{z}) = \sum_{\mathbf{k} \in \Lambda} c_{\mathbf{k}} \prod_{i=1}^{n} z_i^{k_i} . \tag{8}$$

Since the mapping involves powers of z_i, where z_i are specified by the trigonometric mapping (7), we term the expansion in (6) as a trigonometric polynomial.

We note that the mapping $v = (v_1, \cdots, v_n)$ defined by (7) is a bijection from $[0,1)^n$ onto the complex unit torus $\mathbb{T}^n = \{(z_1, \cdots, z_n): |z_i| = 1, i = 1, \cdots, n\}$. Hence,

$$\psi(\mathbf{x}) = 0 \Leftrightarrow \mathscr{P}[\psi][\mathbf{z}] = 0 \text{ on } \mathbb{T}^n, \quad \text{where } \; z_i = v_i(x_i), \quad i = 1, \cdots, n, \tag{9}$$

which implies that there is a one-to-one correspondence between the zero sets of ψ and the zeros of $\mathscr{P}[\psi]$ on the unit torus. Accordingly, we can study the algebraic properties of trigonometric polynomials and their zero sets by studying their corresponding complex polynomials under the mapping v.

2.3.2 Non-uniqueness of level-set representation—

We first show that the level set representation of a surface in (6) may not be unique, when the bandwidth of the representation is larger than the minimal one required to represent the surface. We first note that the function $\psi(\mathbf{x})$ with bandwidth Λ in (6) can be expressed with a larger bandwidth Γ ⊃ Λ by zero filling the additional Fourier coefficients:

$$\psi(\mathbf{x}) = \sum_{\mathbf{k} \in \Gamma} \tilde{\mathbf{c}}_{\mathbf{k}} \exp\left(j2\pi\mathbf{k}^T\mathbf{x}\right), \quad \mathbf{x} \in [0, 1)^n, \tag{10}$$

where the coefficients set $\tilde{\mathbf{c}}$ is the zero-filled version of the vector **c**, denoted by $\tilde{\mathbf{c}} \in \mathbb{C}^{|\Gamma|}$:

$$\tilde{\mathbf{c}}_{\mathbf{k}} = \begin{cases} c_{\mathbf{k}} & \text{if } \mathbf{k} \in \Lambda \\ 0 & \text{else} \end{cases}. \tag{11}$$

We note that the representation of the surface by functions with the larger bandwidth Γ is not unique. In particular, any uniform shift of the coefficients in the Fourier domain corresponds to a phase multiplication in the space domain:

$$\varphi' = \varphi \cdot \exp\left(j2\pi\mathbf{k}_0^T\mathbf{x}\right); \quad \mathbf{k}_0 \in \Gamma \ominus \Lambda\,. \tag{12}$$

Since $\left|\exp\left(j2\pi\mathbf{k}_0^T\mathbf{x}\right)\right| = 1, \forall\mathbf{x}$, we can see that the zero sets of φ' are identical to that of φ.

Because the exponentials $\exp\left(j2\pi\mathbf{k}_0^T\mathbf{x}\right)$ are orthogonal to each other, the functions φ' that has the same zero set as φ lives in a subspace of dimension Γ ⊖ Λ. Here, Γ ⊖ Λ denote the set of all valid uniform shifts $\mathbf{k}_0$ of Λ, denoted by Λ + $\mathbf{k}_0$, that are contained in Γ. We will introduce the set Γ ⊖ Λ with more details in §3.2.2.

2.3.3 Minimal bandwidth representation of a surface—

We note from the previous section that the multiplication with the phase term in (10) corresponds to multiplying the trigonometric polynomial in (8) by $\mathbf{z}^{\mathbf{k}_0}$; the degree of the resulting trigonometric polynomial φ' will be greater than that of φ. In this section, we show that out of all these polynomials, the one with smallest degree is unique. More importantly, the bandwidth of the above minimal polynomial can be used as a measure of the complexity of the surface. Specifically, a more complex surface would correspond to a polynomial with a larger bandwidth.

The following result shows that for any given surface $\mathscr{S}$, there exists a unique level set function ψ, whose coefficient set $\{\mathbf{c}_\mathrm{k} : \mathbf{k} \in \Lambda\}$ has the smallest bandwidth.

Proposition 1.: *For every (hyper-)surface* $\mathscr{S}$ *given by the zero level set of* (10), *there is a unique (up to scaling) minimal trigonometric polynomial* ψ, *which satisfies* $\psi(\mathbf{x}) = 0; \forall \mathbf{x} \in \mathscr{S}$. *Any other trigonometric polynomial* ψ_1 *that also satisfies* $\psi_1(\mathbf{x}) = 0; \forall \mathbf{x} \in \mathscr{S}$ *will have* $BW(\psi_1) \supseteq BW(\psi)$. *Here,* $BW(\psi)$ *denotes the bandwidth of the function* ψ.

As seen from (10), the coefficients of ψ_1 can be the shifted version of the coefficients of ψ. Thus, the Fourier support of ψ_1 is larger than (contains) the Fourier support of ψ; the degree of the trigonometric polynomial ψ_1 is larger than the degree of the minimal polynomial ψ, which has the smallest degree or equivalently bandwidth. In this sense, the minimal polynomial ψ is unique, up to scaling. The proof of this result is given in Appendix 9.1. We refer to the ψ of the form (6) with the minimal bandwidth Λ that satisfy

$$\psi(\mathbf{x}) = 0; \quad \forall \mathbf{x} \in \mathscr{S} \tag{13}$$

as the *minimal trigonometric polynomial* of the surface $\mathscr{S}$.

In other words, when ψ is the minimal trigonometric polynomial of a surface $\mathscr{S}$, it does not have a factor with no zeros (i.e., never vanishes or vanishes only at isolated points on $[0,1)^n$). In particular, if a polynomial has a factor with no zeros in $[0,1)^n$, one can remove this factor and obtain a polynomial with a smaller bandwidth and with the same support set. Note from (8) that the minimal trigonometric polynomial will correspond to $\mathscr{P}[\psi]$ being a polynomial with the minimal degree.

As mentioned at the beginning of this section, the bandwidth Λ of the minimal polynomial of the surface $\mathscr{S}$ grows with the complexity of $\mathscr{S}$; a more oscillatory surface with a lot of details corresponds to a high bandwidth minimal polynomial, while a simple and highly smooth surface corresponds to a low bandwidth minimal polynomial. We hence consider $|\Lambda|$ as a *complexity measure* of the surface. Furthermore, we note that the surface model can approximate an arbitrary closed surface with any degree of accuracy, as long as the bandwidth is large enough [30]. One can refer to Fig.2 in [30] for illustration in 2D and see Fig. 1 for illustration in 3D. Here we illustrate this idea in 2D/3D for simplicity, but the approach is general for any dimensions.

2.3.4 Irreducible bandlimited surfaces

—We now introduce the concept of irreducible polynomials, which is important for our results. We term a surface to be irreducible if its minimal trigonometric polynomial is irreducible. A polynomial is irreducible if it cannot be factorized into smaller factors, whose zero sets are within $[0,1)^n$. Most of the irreducible surfaces are simply connected (i.e., consist of a single connected component[1]). Intuitively, a general surface may be composed of several connected components, where each connected component is irreducible. In this case, we term the above surface as the union or irreducible surfaces. The minimal polynomial of the union of irreducible surfaces will be the product of the irreducible minimal polynomials of the individual connected components. The following definitions puts the above explanations into more concrete terms:

Definition 2.: A surface is termed as irreducible, if it is the zero set of an irreducible trigonometric polynomial.

Definition 3.: *A trigonometric polynomial* $\psi(\mathbf{x})$ *is said to be irreducible, if the corresponding polynomial* $\mathscr{P}[\psi]$ *is irreducible in* $\mathbb{C}[z_1, \cdots, z_n]$. *A polynomial p is irreducible over a field of complex numbers, if it cannot be expressed as the product of two or more non-constant polynomials with complex coefficients.*

When ψ can be written as the product of several irreducible components $\psi = \prod_{i=1}^{m} \psi_i$, then $\mathscr{S}[\psi]$ is essentially the union of irreducible surfaces:

$$\mathscr{S}[\psi] = \bigcup_{i=1}^{m} \mathscr{S}[\psi_i]. \tag{14}$$

3 Lifting mapping and low-dimensional feature spaces

In this section, we show that there exists a non-linear transformation, which maps the points on an irreducible surface to a low-dimensional subspace. The transformation is intimately tied in with the specific choice of basis functions used to represent the surface. Our results show that the dimension of the subspace depends on the complexity of the surface, or equivalently the bandwidth of the minimal polynomial. We can use the rank of the feature matrix as a surrogate of the complexity of the surface to recover it, much like sparsity is used to recover signals in compressed sensing.

Consider the non-linear lifting mapping $\Phi_\Gamma : [0,1]^n \to \mathbb{C}^{|\Gamma|}$, obtained by evaluating the basis functions at $\mathbf{x}$:

$$\Phi_\Gamma(\mathbf{x}) = \begin{bmatrix} \varphi_{\mathbf{k}_1}(\mathbf{x}) \\ \vdots \\ \varphi_{\mathbf{k}_{|\Gamma|}}(\mathbf{x})) \end{bmatrix}. \tag{15}$$

We can view $\Phi_\Gamma(\mathrm{x})$ as the feature vector of the point $\mathbf{x}$, analogous to the ones used in kernel methods [45]. Here, $|\Gamma|$ denotes the cardinality of the set Γ. We denote the set

$$\mathscr{V}_\Gamma(\mathscr{S}) = \{\Phi_\Gamma(\mathbf{x}) \mid \mathbf{x} \in \mathscr{S}\} \tag{16}$$

[1] One can come up with counter examples of irreducible polynomials with multiple components. In this work, one can ignore these pathological counter examples and assume that an irreducible bandlimited surface will consist of only one connected component.

as the feature space of the surface $\mathcal{S}$. Since any point on a surface $\mathcal{S}$ satisfies (3), the feature vectors of points from $\mathcal{S}$ satisfy

$$\mathbf{c}^T \Phi_\Gamma(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in \mathcal{S}, \tag{17}$$

where **c** is the coefficients vector in the representation of ψ in (6). The above relation is illustrated in Fig. 2.

The relation (17) also implies that **c** is orthogonal to all the feature vectors of points living on $\mathcal{S}$ and hence a feature matrix constructed from points on the surface is rank deficient by one; i.e., the dimension of the feature space is at most $|\Gamma| - 1$. However, we now show that the feature matrix is often significantly low-rank depending on the geometry of the surface and the specific representations of the surface.

3.1 Shift invariant representation

We now show that if the level set function is represented by a shift invariant representation (e.g. B-splines), the dimension of the lifted feature points are dependent on the area of the surface. We consider $\varphi_{\mathbf{k}} = \beta^p\left(\frac{\mathbf{x}}{T} - \mathbf{k}\right)$ to be the p^{th}-degree tensor-product B-spline function. Note that $\beta^p(\mathbf{x})$ is support limited in $[-(p+1)/2, (p+1)/2]^n$. If the support of $\varphi_k(\mathbf{x})$ does not overlap with $\mathcal{S}$, we have $\varphi_k(\mathbf{x}) = 0; \forall \mathbf{x} \in \mathcal{S}$. Hence, we have

$$\mathbf{i}_{\mathbf{k}}^T \Phi_\Gamma(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in \mathcal{S} \tag{18}$$

where $\mathbf{i}_{\mathbf{k}}$ is the indicator vector whose k^{th} entry is one and the rest of the entries are zeros. Note that all of these indicator vectors are linearly independent. The number of basis vectors whose bandwidth does not overlap with $\mathcal{S}$ is dependent on the area of $\mathcal{S}$ as well as the support of ψ. Thus, the dimension of $\mathcal{V}_\Gamma(\mathcal{S})$ is a measure of the area of the surface $\mathcal{S}$, and satisfies

$$\dim(\mathcal{V}_\Gamma(\mathcal{S})) \le |\Gamma| - (P + 1) = A, \tag{19}$$

where P is the number of basis functions whose support does not overlap with $\mathcal{S}$.

3.2 Band-limited surface representation

We now consider the case of an arbitrary point **x** on the zero level set of $\psi(\mathbf{x})$ with bandwidth Λ. Using (10), the lifting is specified by:

$$\Phi_\Lambda(\mathbf{x}) = \begin{bmatrix} \exp(j2\pi\mathbf{k}_1^T\mathbf{x}) \\ \exp(j2\pi\mathbf{k}_2^T\mathbf{x}) \\ \vdots \\ \exp(j2\pi\mathbf{k}_{|\Lambda|}^T\mathbf{x}) \end{bmatrix}. \tag{20}$$

We note from (20) that the lifting Φ can be evaluated with a larger bandwidth $\Gamma \supset \Lambda$. When the lifting is performed with the minimal bandwidth (i.e., $\Gamma = \Lambda$), we term the corresponding lifting as the *minimal lifting.*

We now analyze $\Lambda$the dimension of the feature space $\mathscr{V}_\Lambda(\mathscr{S})$ for the minimal ($\Gamma = \Lambda$) and non-minimal lifting ($\Lambda \subset \Gamma$) cases. In both cases, we will show that the feature space is low-dimensional and is a subspace of $\mathbb{C}^{|\Lambda|}$.

3.2.1 Irreducible surface with minimal lifting ($\Gamma = \Lambda$)—We first focus on the case where ψ is an irreducible trigonometric polynomial and the bandwidth of the lifting is specified by Λ, which is the bandwidth of the minimal polynomial. The annihilation relation (17) implies that **c** is orthogonal to the feature vectors $\Phi_\Lambda(\mathbf{x})$**.** This implies that

$$\dim(\mathscr{V}_\Lambda) \le |\Lambda| - 1. \tag{21}$$

3.2.2 Irreducible surface with non-minimal lifting ($\Gamma \supset \Lambda$)—We now consider the setting where the non-linear lifting is specified by $\Phi_\Gamma(\mathbf{x})$, where $\Lambda \subset \Gamma$. Because of the annihilation relation, we have

$$\tilde{\mathbf{c}}^T \Phi_\Gamma(\mathbf{x}) = 0,$$

where $\tilde{\mathbf{c}}$ is the zero filled coefficients in (10). Since the zero set of the function $\psi_{\mathbf{k}_0}(\mathbf{x}) = \psi(\mathbf{x}) \cdot \exp\left(j2\pi \mathbf{k}_0^T \mathbf{x}\right)$ is exactly the same as that of ψ, we have

$$\sum_{\mathbf{k}} \mathbf{c}_{\mathbf{k}-\mathbf{k}_0} \exp\left(j2\pi \mathbf{k}^T \mathbf{x}\right) = 0; \quad \forall \mathbf{x} \in \mathscr{S}[\psi]\,. \tag{22}$$

This implies that any shift of $\tilde{\mathbf{c}}$ within $\Gamma \ominus \Lambda$, denoted by $\tilde{\mathbf{d}}_{\mathbf{k}} = \mathbf{c}_{\mathbf{k}-\mathbf{k}_0}$ will satisfy $\tilde{\mathbf{d}}^T \Phi_\Gamma(\mathbf{x}) = 0$. It is straightforward to see that $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{c}}$ are linearly independent for all values of $\mathbf{k}_0$. We denote the number of possible shifts such that the shifted set $\Lambda + \mathbf{k}_0$ is still within Γ (i.e., $\Lambda + \mathbf{k}_0 \subseteq \Gamma$) by $|\Gamma \ominus \Lambda|$:

$$\Gamma \ominus \Lambda = \{1 \in \Gamma \mid 1 - \mathbf{k} \in \Gamma, \forall \mathbf{k} \in \Lambda\}\,. \tag{23}$$

This set is illustrated in Fig. 3 along with Γ and Λ. Since the vectors $\mathbf{c}_{\mathbf{k}-\mathbf{k}_0}$ are linearly independent and are orthogonal to any feature vector $\Phi_\Gamma(\mathbf{x})$ on $\mathscr{S}[\psi]$, the dimension of the subspace is bounded by

$$\dim(\mathscr{V}_\Gamma) \le |\Gamma| - |\Gamma \ominus \Lambda|\,. \tag{24}$$

3.2.3 Union of irreducible surfaces with $\Gamma \supset \Lambda_i$—When $\psi = \prod_{i=1}^{m} \psi_i$, each irreducible surface $\mathcal{S}[\psi_i]$ will be mapped to a subspace of dimension $|\Gamma|$ — $|\Gamma \ominus \Lambda_i|$. This implies that the non-linear lifting transforms the union of irreducible surfaces to the well-studied union of subspace model [14,23,25].

4 Surface recovery from samples

In this section, we will use the low-rank structure of the feature maps of the points to recover the surface. As discussed in the introduction, the recovery of a surface/manifold from point clouds is an important problem in denoising, machine learning, shape recovery from point clouds, and image segmentation. For presentation purposes, we consider different cases in the increasing order of complexity. In particular, we consider irreducible (single connected component) surfaces with minimal lifting, union of irreducible components with minimal lifting, and finally the case with non-minimal lifting. Note that in practice, the bandwidth of the surface is not known apriori, and hence one has to over-estimate the bandwidth; this translates to the non-minimal lifting setting. Our results in this section show that irreducible surfaces can be recovered from very few samples, as long as the number of samples exceed a number proportional to the bandwidth. Union of irreducible surfaces can also be recovered from few samples, but each of the irreducible components need to be sampled adequately to guarantee perfect recovery.

4.1 Sampling theorems

We consider the recovery of the surface $\mathcal{S}$ from its samples $\mathbf{x}_i$; $i = 1, \cdots, N$. According to the analysis in the previous section, if the sampling point $\mathbf{x}_i$ is located on the zero level set of $\psi(\mathbf{x})$, we will then have the annihilation relation specified by (17). Notice that equation (17) is a linear equation with $\mathbf{c}$ as its unknowns. Since all the samples $\mathbf{x}_i$; $i = 1,.., N$ satisfy the annihilation relation (17), we have

$$\mathbf{c}^T \underbrace{\left[\Phi_\Gamma(\mathbf{x}_1) \quad \cdots \quad \Phi_\Gamma(\mathbf{x}_N)\right]}_{\Phi_\Gamma(\mathbf{X})} = 0. \tag{25}$$

We call $\Phi_\Gamma(\mathbf{x})$ the feature matrix of the sampling set $\mathbf{X} = \{\mathbf{x}_1, \cdots, x_N\}$. We propose to estimate the coefficients $\mathbf{c}$, and hence the surface $\mathcal{S}[\psi]$ using the above linear relation (25). Note that $\mathcal{S}[\psi]$ is invariant to the scale of $\mathbf{c}$; without loss of generality, we reformulate the estimation of the surface as the solution to the system of equations

$$\mathbf{c}^T \Phi_\Gamma(\mathbf{X}) = 0; \;\; \|\mathbf{c}\|_F = 1. \tag{26}$$

We note that without the constraint $\|\mathbf{c}\|_F = 1$, $\mathbf{c}^{\mathrm{T}} \Phi_\Gamma(\mathbf{X}) = 0$ will have a trivial solution with $\mathbf{c} = 0$. The use of the Frobenius norm constraint enables us to solve the problem using eigen decomposition. The above estimation scheme yields a unique solution, if the matrix $\Phi_\Lambda(\mathbf{X})$ has a unique null-space basis vector. We will now focus on the number of samples N and its distribution on $\mathcal{S}[\psi]$, which will guarantee the unique recovery of $\mathcal{S}[\psi]$. We will consider different lifting scenarios introduced in Section 3 separately. As we will see, in some cases considered below, the null-space has a large dimension. However, the minimal null-space vector (coefficients with the minimal bandwidth) will still uniquely identify the surface, provided the sampling conditions are satisfied.

4.1.1 Case 1: Irreducible surfaces with minimal lifting—Suppose $\psi(\mathbf{x})$ is an irreducible trigonometric polynomial with bandwidth Λ. Consider the lifting which is specified by the minimal bandwidth Λ. We see from (21) that rank $(\Phi_\Lambda(\mathbf{X})) < |\Lambda| — 1$. The following result shows when the inequality is replaced by an equality.

Proposition 4.: *Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ be N independent and uniformly distributed random samples on the surface $\mathcal{S}[\psi]$, where $\psi(\mathbf{x})$ is an irreducible (minimal) trigonometric polynomial with bandwidth Λ. The feature matrix $\Phi_\Lambda(\mathbf{X})$ will have rank $|\Lambda| — 1$, if*

$$N \geq |\Lambda| - 1$$

for almost all surfaces $\mathcal{S}[\psi]$.

We note that the above results are true for almost all surfaces. This implies that the surfaces for which the above results do not hold correspond to a set of measure zero [10]. The above proposition guarantees that the solution to the system of equations specified by (26) is unique (up to scaling) when the number of samples exceeds $N = |\Lambda| — 1$ with unit probability. The proof of this proposition can be found in Appendix 9.2. With Proposition 4, we obtain the following sampling theorem.

Theorem 5 (Irreducible surfaces of any dimension).: *Let $\psi(\mathbf{x})$, $\mathbf{x} \in [0,1]^n$, $n \geq 2$ be an irreducible trigonometric polynomial whose bandwidth is given by Λ. The zero level set of $\psi(\mathbf{x})$ is denoted as $\mathcal{S}[\psi]$. If we are randomly given $N \geq |\Lambda| — 1$ samples on $\mathcal{S}[\psi]$, then almost all surfaces $\mathcal{S}[\psi]$ can be recovered.*

This theorem generalizes the results in [60] to any dimension $n \geq 2$ and is illustrated in Fig. 4 and Fig. 5.

In the theorem, when $n = 2$, then $\mathcal{S}$ is a planar curve. In this setting, if the bandwidth of ψ Λ is a rectangular region with dimension $k_1 \times k_2$. Then by this sampling theorem, we get perfect recovery with probability one, when the number of random samples on the curve exceeds $k_1 \cdot k_2 - 1$. Note that the degrees of freedom in the representation (6) is $k_1 \cdot k_2 - 1$, when we constrain $\|\mathbf{c}\|_F = 1$. This implies that if the number of samples exceed the degrees of freedom, we get perfect recovery. Note that these results are significantly less conservative than the ones in [60], which required a minimum of $(k_1 + k_2)^2$ samples. We note that the results in [60] were the worst case guarantees, and will guarantee the recovery of the curve from any $(k_1 + k_2)^2$ samples. By contrasts, our current results are high probability results; there may exist a set of $N \geq k_1 \cdot k_2 - 1$ samples from which we cannot get unique recovery.

We note that the current work is motivated by the phase transition experiments (Fig. 5) in [60], which shows that one can recover the curve in most cases when the number of samples exceeds $k_1 \cdot k_2 - 1$ rather than the conservative bound of $(k_1 + k_2)^2$. We also note that it is not straightforward to extend the proof in [60] to the cases beyond $n = 2$. Specifically, we relied on Bezout's inequality in [60], which does not generalize easily to high dimensional cases.

4.1.2 Case 2: Union of irreducible surfaces with minimal lifting—We now consider the union of irreducible surfaces $\mathcal{S}[\psi]$, where ψ has several irreducible factors $\psi(\mathbf{x}) = \psi_1(\mathbf{x})\cdots\psi_M(\mathbf{x})$. Then we have $\mathcal{S}[\psi] = \cup_{i=1}^{M} \mathcal{S}[\psi_i]$. Suppose the bandwidth of $\psi(\mathbf{x})$ is given by Λ and the bandwidth of each factor $\psi_i(\mathbf{x})$ is given by Λ_i. We have the following result for this setting.

Proposition 6.: *Let $\psi(\mathbf{x})$ be a trigonometric polynomial with M irreducible factors, i.e.,*

$$\psi(\mathbf{x}) = \psi_1(\mathbf{x})\cdots\psi_M(\mathbf{x})\,. \tag{27}$$

Suppose the bandwidth of each factor $\psi_i(\mathbf{x})$ is given by Λ_i and the bandwidth of ψ is Λ. Assume that $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ are N uniformly distributed random samples on $\mathcal{S}[\psi]$, which are chosen independently. Then with probability 1 that the feature matrix $\Phi_\Lambda(\mathbf{X})$ will be of rank $|\Lambda| - 1$ for almost all 0 if

1. each irreducible factor is randomly sampled with Ni > |Λi| — 1 points, and
2. the total number of samples satisfy N ≥ |Λ| — 1.

Similar to previous propositions, the above results are valid for almost all ψ, which implies that the set of ψ for which the above results do not hold is a set of measure zero [10]. The proof of this result can be seen in Appendix 9.2.3. Based on this proposition, we have the following sampling conditions.

Theorem 7 (Union of irreducible surfaces of any dimension).: *Let $\psi(\mathbf{x})$ be a trigonometric polynomial with M irreducible factors as in (28). If the samples $\mathbf{x}_1$,., $\mathbf{x}_n$ satisfy the conditions in Proposition 6, then the surface can be uniquely recovered by the solution of (26) for almost all ψ.*

Unlike the sampling conditions in Theorem 5 that does not impose any constraints on the sampling, the above result requires each component to be sampled with a minimum rate specified by the degrees of freedom of that component. We illustrate the above result in Fig. 6 in 2D ($n = 2$), where $\mathcal{S}$ is the union of two irreducible curves with bandwidth of 3 × 3, respectively. The above results show that if each of these simply connected curves are sampled with at least eight points and if the total number of samples is no less than 24, we can uniquely identify the union of curves. The results show that if any of the above conditions are violated, the recovery fails; by contrast, when the number of randomly chosen points satisfy the conditions, we obtain perfect recovery.

4.1.3 Case 3: Non-minimal lifting—In Section 4.1.1 and 4.1.2, we introduced theoretical guarantees for the perfect recovery of the surface in any dimensions. The sampling theorems introduced in Section 4.1.1 and 4.1.2 assume that we know exactly the

bandwidth of the surface or the union of surfaces. However, in practice, the true bandwidth of the surface is usually unknown. We now consider the recovery of the surface, when the bandwidth is over-estimated, or equivalently the lifting is performed assuming $\Gamma \supset \Lambda$. As discussed in Section 3.2.2, the dimension of $\mathscr{V}_\Gamma$ is upper bounded by $|\Gamma|$ — $|\Gamma \ominus \Lambda|$, which implies that

$$\text{rank}(\Phi_\Gamma(\mathbf{X})) \le |\Gamma| - |\Gamma \ominus \Lambda|,$$

where $\Gamma \ominus \Lambda$ represents the number of valid shifts of Λ within Γ as discussed in Section 3.2.2.

The following two propositions show when the inequality in the rank relation above can be an equality and hence we can recover the surface.

Proposition 8 (Irreducible surface with non-minimal lifting).: *Let* $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ *be N random samples on the surface* $\mathscr{S}[\psi]$, *chosen independently. The trigonometric polynomial* $\psi(\mathbf{x})$ *is irreducible whose true bandwidth is* Λ. *Suppose the lifting mapping is performed using bandwidth* $\Gamma \supset \Lambda$. *Then* $\text{rank}(\Phi_\Gamma(\mathbf{X})) = |\Gamma| - |\Gamma \ominus \Lambda|$ *for almost all* ψ, *if*

$$N \ge |\Gamma| - |\Gamma \ominus \Lambda| .$$

The proof of this proposition can be found in Appendix 9.2.4.

Proposition 9 (Union of irreducible surfaces with non-minimal lifting).: *Let* $\psi(\mathbf{x})$ *be a randomly chosen trigonometric polynomial with M irreducible factors, i.e.,*

$$\psi(\mathbf{x}) = \psi_1(\mathbf{x}) \cdots \psi_M(\mathbf{x}) . \tag{28}$$

Suppose the bandwidth of each factor $\psi_i(\mathbf{x})$ *is given by* Λ_i *and the bandwidth of* ψ *is* Λ. *Let* $\Gamma_i \supset \Lambda_i$ *be the non-minimal bandwidth of each factor* $\psi_i(\mathbf{x})$ *and* $\Gamma \supset \Lambda$ *is the bandwidth of the non-minimal lifting. Assume that* $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ *are N random samples on* $\mathscr{S}[\psi]$ *that are chosen independently. Then, the feature matrix* $\Phi_\Lambda(\mathbf{X})$ *will be of rank* $|\Gamma| - |\Gamma \ominus \Lambda|$ *for almost all f if*

1. *each irreducible factor is randomly sampled with* $N_i \ge |\Gamma| - |\Gamma_i \ominus \Lambda_i|$ *points, and*
2. *the total number of samples satisfy* $N \ge |\Gamma| - |\Gamma \ominus \Lambda|$.

We prove this result in Appendix 9.2.5. Note that in practice, when non-minimal lifting mapping is performed, we then randomly sample approximately $/\Gamma| - |\Gamma \ominus \Lambda|$ positions on $\mathscr{S}$. This random strategy ensures that the samples are distributed to the factors, roughly satisfying the conditions in Proposition 9. We further studied this proposition in Fig. 7. We considered several random surfaces obtained by choosing random coefficients, each with different bandwidth and considered their recovery from different number of samples. From which, we obtained the phase transition plot given in Fig. 7, which agrees well with the theory.

4.2 Surface recovery algorithm for the non-minimal setting

The two propositions in Section 4.1.3 show that $\Phi_\Gamma\mathbf{(X)}$ has $|\Gamma \ominus \Lambda|$ null space basis vectors $\mathbf{n}_i \leftrightarrow \mu_i$, when the non-minimal lifting with bandwidth Γ is performed. The following result from [60] shows that the null-space vectors are related to the minimal polynomial of the surface. In particular, all null-space vectors have the minimal polynomial as a factor. We will use this property to extract the surface from the null-space vectors as their greatest common divisor. We also introduce a simpler computational strategy which relies on the sum of squares of the null-space vectors.

Proposition 10 (Proposition 9 in [60]).—*The coefficients of the trigonometric polynomials of the form*

$$\theta_{\mathbf{k}}(\mathbf{x}) = \exp(j2\pi\mathbf{l}^T\mathbf{x})\psi(\mathbf{x}), \quad \forall \mathbf{k} \in \Gamma \ominus \Lambda\,.$$

is a null space vector of $\Phi_\Gamma\mathbf{(X)}$.

Note that the coefficients of $\theta_{\mathbf{k}}(\mathbf{x})$ correspond to the shifted versions of the coefficients of ψ and hence are linearly independent. We also note that any such function is a valid annihilating functions for points on $\mathcal{S}$. When the dimension of the null-space is $|\Gamma \ominus \Lambda|$, these corresponding coefficients form a basis for the null-space. Therefore, we have that any function in the null-space can be expressed as

$$\eta(\mathbf{x}) = \sum_{\mathbf{k} \in \Gamma \ominus \Lambda} \alpha_{\mathbf{k}}\psi(\mathbf{x})\exp\left(j2\pi k^T\mathbf{x}\right) \tag{29}$$

$$= \psi(\mathbf{x})\underbrace{\sum_{\mathbf{k} \in \Gamma \ominus \Lambda} \alpha_{\mathbf{k}}\exp\left(j2\pi\mathbf{k}^T\mathbf{x}\right)}_{\gamma(\mathbf{x})} = \psi(\mathbf{x})\gamma(\mathbf{x}), \tag{30}$$

where α_k and γ are arbitrary coefficients and function, respectively. Note that all of the functions obtained by the null-space vectors have ψ as a common factor.

Accordingly, we have that $\psi(\mathbf{x})$ is the greatest common divisor of the polynomials $\mu_i(x) \leftrightarrow \mathbf{n}_i$, where $\boldsymbol{n}_\mathrm{i}$ are the null-space vectors of $\Phi_\Gamma\mathbf{(X)}$, which can be estimated using singular value decomposition (SVD). Since we consider polynomials of several variables, it is not computationally efficient to find the greatest common divisor. We note that we are not interested in recovering the minimal polynomial, but are only interested in finding the common zeros of $\mu_i(\mathbf{x})$. We hence propose to recover the original surface as the zeros of the sum of squares (SoS) polynomial

$$\sigma(\mathbf{x}) = \sum_{i=1}^{|\Gamma \ominus \Lambda|} |\mu_i(\mathbf{x})|^2\,.$$

Note that rank guarantees in Propositions 9.2.4 and 9.2.5 ensure that the entire null-space will be fully identified by the feature matrix. Coupled with Proposition 10, we can conclude that the recovery using the above algorithm (SVD, followed by the sum of squares of the inverse Fourier transforms of the coefficients) will give perfect recovery of the surface under noiseless conditions. The algorithm is illustrated in Fig. 8.

4.3 Surface recovery from noisy samples

The analysis in Section 4.1.3 shows that when the bandwidth of the surface is small, the feature matrix is low rank. In practice, the sampling points are usually corrupted with some noise. We denote the noisy sampling set by $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{N}$ is the noise. We propose to exploit the low-rank nature of the feature matrix to recover it from noisy measurements. Specifically, when the sampling set $\mathbf{X}$ is corrupted by noise, the points will deviate from the original surface, and hence the features will cease to be low rank. We impose a nuclear norm penalty on the feature maps that will push the feature vectors to a subspace. Since the feature vectors are related to the original points by the exponential mapping, the original points will move to the surface. In practice it is difficult to compute the feature map. We hence rely on an iterative reweighted least-squares algorithm, coupled with the *kernel-trick,* to avoid the computation of the features. Since the cost function is non-linear (due to the non-linear kernel), we use steepest descent-like algorithm to minimize the cost function. We note that each iteration of this algorithm has similarities to non-local means algorithms, which first estimate the weight/Laplacian matrix from the patches, followed by a smoothing. We also note that this approach has conceptual similarities to kernel low-rank algorithms used in MRI and computer vision [28, 29]. These algorithms rely on explicit polynomial mappings, low-rank approximation of the features, followed by the analytical evaluation of the pre-images that is possible for polynomial kernels.

We pose the denoising as:

$$\mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{Y}\|^2 + \lambda \|\Phi(\mathbf{X})\|_* \tag{31}$$

where we use the nuclear norm of the feature matrix of the sampling set as a regularizer. Unlike traditional convex nuclear norm formulations, the above scheme is non-convex.

We adapt the kernel low-rank algorithm in [31,38] to the high dimensional setting to solve (31). This algorithm relies on an iteratively reweighted least squares (IRLS) approach [12, 26] which alternates between the following two steps:

$$\mathbf{X}^{(n)} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{Y}\|^2 + \lambda \operatorname{trace}\left[\mathscr{K}(\mathbf{X})\mathbf{P}^{(n-1)}\right], \tag{32}$$

and

$$\mathbf{P}^{(n)} = \left[\mathscr{K}\left(\mathbf{X}^{(n)}\right) + \gamma^{(n)}\mathbf{I}\right]^{-1/2} \tag{33}$$

where $\gamma^{(n)} = \frac{\gamma^{(n-1)}}{\eta}$ and $\eta > 1$ is a constant. Here, $\mathscr{K}(\mathbf{X}) = \Phi_\Gamma(\mathbf{X})^T \Phi_\Gamma(\mathbf{X})$. We use the *kernel-trick* to evaluate $\mathscr{K}(\mathbf{X})$. The kernel-trick suggests that we do not need to explicitly evaluate the features. Each entry of the matrices $\mathscr{K}(\mathbf{X})$ correspond to inner-products in feature space:

$$(\mathscr{K}(\mathbf{X}))_{(i,j)} = \underbrace{\Phi(\mathbf{x}_i)^H \Phi\left(\mathbf{x}_j\right)}_{\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right)} \tag{34}$$

which can be evaluated efficiently using the nonlinear function κ (termed as kernel function) of their inner- products in $\mathbb{R}^n$.

The dependence of the kernel function on the lifting is detailed in Section 5.3. Since the above problem in (32) is not quadratic, we propose to solve it using gradient descent as in [60]. We note that the cost function in (32) can be rewritten as

$$C(\mathbf{X}) = \|\mathbf{X} - \mathbf{Y}\|^2 + \lambda \sum_{i,j} \mathbf{P}_{ij}^{(n-1)} \kappa\left(\mathbf{x}_i, \mathbf{x}_j\right), \tag{35}$$

where $\mathbf{P}_{i,j}$ are the entries of the matrix $\mathbf{P}^{(n-1)}$. As will be discussed in detail in Section 5.3, the exponential kernel for a circular support as in Fig. 11.(b) can be approximated as a circularly symmetric kernel $\kappa(\mathbf{x}_i, \mathbf{x}_j) = k(\|\mathbf{x}_i - \mathbf{x}_j\|^2)$. In this case, the partial derivatives of (32) with respect to one of the vectors x_i is

$$\partial_{\mathbf{x}_i} \mathscr{C} = 2(\mathbf{x}_i - \mathbf{y}_i) + 2\lambda \sum_j \underbrace{\mathbf{P}_{ij}^{(n-1)} k'(\|\mathbf{x}_i - \mathbf{x}_j\|^2)}_{w_{i,j}} (\mathbf{x}_i - \mathbf{x}_j) \tag{36}$$

$$= 2(\mathbf{x}_i - \mathbf{y}_i) + 2\lambda \underbrace{\left(\sum_j w_{i,j}\right)}_{d_i} \mathbf{x}_i - \mathbf{W}\mathbf{X}. \tag{37}$$

Here,

$$\mathbf{W}_{ij} = \mathbf{P}_{ij}^{(n-1)} k'(\|\mathbf{x}_i - \mathbf{x}_j\|^2. \tag{38}$$

Thus, the gradient of the cost function (35) is :

$$\nabla_{\mathbf{X}} \mathscr{C} \approx 2(\mathbf{X} - \mathbf{Y}) + 2\lambda \underbrace{(\mathbf{D} - \mathbf{W})}_{\mathbf{L}} \mathbf{X}. \tag{39}$$

Here, **L** = **D** — **W** is the matrix obtained from the weights **W** and **D** is a diagonal matrix with diagonal entries $d_i = \sum_j \mathbf{W}_{ij}$.

We note that the gradient of (32) specified by (39) is also the gradient of the cost function

$$\mathscr{D} = \|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda \mathrm{trace}\left(\mathbf{X}\ \mathbf{L}\ \mathbf{X}^H\right), \tag{40}$$

which is used in approaches such as non-local means (NLM) [6] and graph regularization [48]. We note that the above optimization problem is quadratic and hence has an analytical solution. We thus alternate between the solution of (40) and updating the weights, and hence the Laplacian matrix using (38), where **P** is specified by (33). Despite the similarity to NLM, we note that NLM approaches use a fixed Laplacian unlike the iterative approach in our work. In addition, the expression of the Laplacian is also very different. We refer the readers to [60] for comparison of the proposed scheme with the above graph regularized algorithm. Once the denoised null-space matrix is obtained from the above algorithm, we can use the sum of square approach described in Section 4.1.3 to recover the surfaces. We note that the algorithm is not very sensitive to the true bandwidth of the kernel Γ, as long as it over-estimates the true bandwidth of the surface Λ.

We illustrate this approach in the context of recovering 3D shapes from noisy point clouds in Fig. 9. The data sets are obtain from AIM@SHAPE [1]. We note that the direct approach, where the null-space vector is calculated from the noisy feature matrix, often results in perturbed shapes. By contrast, the nuclear norm prior is able to regularize the recovery.

5 Recovery of functions on surfaces

As discussed in the introduction, modern machine learning algorithms pre-learn functions from given input and output data pairs [19]. For example, CNN based denoising approaches that provide state-of-the-art results essentially learn to generate noise-free pixels or patches from given training data with several noisy and noise- free patch pairs [52,58]. The problem can be formulated as estimating a nonlinear function $y = f(x)$, given input and output data pairs (x_i, y_i); $i = 1,.., N_{\text{train}}$. A challenge in the representation of such high dimensional function is the large number of parameters, which is also termed as the curse of dimensions. Kernel methods [35], random forests [46] and neural networks [57] provide a powerful class of machine learning models that can be used in learning highly nonlinear functions. These models have been widely used in many machine learning tasks [16].

We now show that the results shown in the previous sections provide an attractive option to compactly represent functions, when the data lie on a smooth surface or manifold in high dimensional spaces. We note that the manifold assumption is widely assumed in a range of machine learning problems [11, 13]. We now show that if the data lie on a smooth surface in high dimensional space, one can represent the multidimensional functions very efficiently using few parameters.

We model the function using the same basis functions used to represent the level set function. In our case[2], we model it as a band-limited multidimensional function:

[2] We note that similar results can be obtained when the function f and the level set function are represented as a linear combination of shift-invariant functions or polynomials.

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \Gamma} \beta_{\mathbf{k}} \exp\left(j2\pi \mathbf{k}^T \mathbf{x}\right) = \beta^T \Phi_{\Gamma}(\mathbf{x}), \tag{41}$$

where $\mathbf{x} \in \mathbb{R}^n$. The number of free parameters in the above representation is $|\Gamma|$, where $\Gamma \subset \mathbb{Z}^n$ is the bandwidth of the function. Note that $|\Gamma|$ grows rapidly with the dimension n. The large number of parameters needed for such a representation makes it difficult to learn such functions from few labeled data points. We now show that if the points lie on the union of irreducible surfaces as in (14), where the bandwidth of ψ is given by $\Lambda \subset \Gamma$, we can represent functions of the form (41) efficiently.

5.1 Compact representation of features using anchor points

We use the upper bound of the dimension of the feature matrix in (24) to come up with an efficient representation of functions of the form 41. The dimension bound (24) implies that the features of points on $S[\psi]$ lie in a subspace of dimension $r = |\Gamma| — |\Gamma \ominus \Lambda|$, which is far smaller than $|\Gamma|$ especially when the dimension n is large. We note that kernel methods often approximate the feature space using few eigen vectors of kernel PCA. However, there is no guarantee that these basis vectors are mappings of some points on S. Hence, it is a common practice to consider all the training samples to capture the low-dimensional feature vectors in kernel PCA. We now show that it is possible to find a set of $N \geq r$ anchor points $\mathbf{a}_1, \cdots, \mathbf{a}_N \in \mathcal{S}[\psi]$, such that the feature space $\mathscr{V}_{\Gamma}(\mathcal{S})$ is in span$\{\Phi_{\Gamma}(\mathrm{a}_1), \cdots, \Phi_{\Gamma}(\mathrm{a}_N)\}$. This result is a Corollary of Proposition 9.

Corollary 11.—*Let* $\psi(\mathbf{x})$ *be a randomly chosen trigonometric polynomial with M irreducible factors as in* (28). *Suppose* $\Gamma_i \supset \Lambda_i$ *is the non-minimal bandwidth of each factor* $\psi_i(\mathrm{x})$ *and* $\Gamma \supset \Lambda$ *is the total bandwidth. Let* $\{\mathrm{a}_1, \cdots, \mathrm{a}_N\}$ *be N randomly chosen anchor points on* $\mathcal{S}[\psi]$ *satisfying*

1. *each irreducible factor* $\mathcal{S}[\psi_i]$ *is sampled with* $N_i \geq |\Gamma_i| — |\Gamma_i \ominus \Lambda_i|$ *points, and*

2. *the total number of samples satisfy* $N \geq |\Gamma| — |\Gamma \ominus \Lambda|$.

Then,

$$\mathscr{V}_{\Gamma}(\mathcal{S}) \subseteq \mathrm{span}\left\{\Phi_{\Gamma}(\mathbf{a}_i); i = 1, \cdots, N\right\} \tag{42}$$

with probability 1.

As discussed in Section 4.1.3, if we randomly choose $N \geq /\Gamma/ — /\Gamma \ominus \Lambda/$points on $\mathcal{S}[\psi]$, the feature matrix will satisfy the conditions in Corollary 11 and hence (42) with unit probability. This relation implies that the feature vector of any point $\mathbf{x} \in \mathcal{S}[\psi]$ can be expressed as the linear combination of the features of the anchor points $\Phi_{\Gamma}(\mathrm{a}_i)$; $i = 1, \cdots,$ N:

$$\Phi_{\Gamma}(x) = \sum_{i=1}^{N} \alpha_i(x) \Phi_{\Gamma}(\mathbf{a}_i) \tag{43}$$

$$= \underbrace{\left[\Phi_\Gamma(\mathbf{a}_1)\ \cdots\ \Phi_\Gamma(\mathbf{a}_N)\right]}_{\Phi(\mathbf{A})} \underbrace{\begin{bmatrix}\alpha_1(\mathbf{x})\\ \vdots\\ \alpha_N(\mathbf{x})\end{bmatrix}}_{\boldsymbol{\alpha}(\mathbf{x})} \tag{44}$$

Here, $\alpha_i(\mathbf{x})$ are the coefficients of the representation. Note that the complexity of the above representation is dependent on N, which is much smaller than $|\Gamma|$, when the surface is highly band-limited. We note that the above compact representation is exact only for $\mathbf{x} \in \mathcal{S}[\psi]$ and not for arbitrary $\mathbf{x} \in \mathbb{R}^n$; the representation in (44) will be invalid for $\mathbf{x} \notin \mathcal{S}[\psi]$.

However, this direct approach requires the computation of the high dimensional feature matrix, and hence may not be computationally feasible for high dimensional problems. We hence consider the normal equations and solve for $\boldsymbol{\alpha}(\mathbf{x})$ as

$$\alpha(\mathbf{x}) = \left(\underbrace{\Phi(\mathrm{A})^H\Phi(\mathrm{A})}_{\mathscr{K}(\mathbf{A})}\right)^\dagger \underbrace{\left(\Phi(\mathrm{A})^H\Phi_\Gamma(\mathbf{x})\right)}_{\mathbf{k}_\mathbf{A}(\mathbf{x})}, \tag{45}$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse.

5.2 Representation and learning of functions

Using (41), (44), and (45), the function $\mathrm{f}:\mathbb{R}^n \to \mathbb{R}^m$ can be written as

$$\mathrm{f}(\mathbf{x}) = \beta^T\Phi(\mathrm{A})\mathscr{K}(\mathrm{A})^\dagger\mathbf{k}_\mathrm{A}(\mathbf{x}) \tag{46}$$

$$= \underbrace{\left[\overbrace{\beta^T\Phi_\Gamma(\mathbf{a}_1)}^{\mathbf{f}(\mathbf{a}_1)}, \ldots, \overbrace{\boldsymbol{\beta}^T\Phi_\Gamma(\mathbf{a}_N)}^{\mathbf{f}(\mathbf{a}_N)}\right]}_{\mathbf{F}} \underbrace{\mathscr{K}(\mathbf{A})^\dagger\mathbf{k}_\mathbf{A}(\mathbf{x})}_{\boldsymbol{\alpha}(\mathbf{x})} \tag{47}$$

Here, f(x) is an $M \times 1$ vector, while $\mathbf{F}$ is an $M \times$ N matrix. $\mathscr{K}(\mathbf{A})$ is an $N \times N$ matrix and $\mathbf{k}_\mathbf{A}$(x) is an $N \times 1$ vector. Thus, if the function values at the anchor points, specified by f(a_i); $i = 1, \cdots, N$ are known, one can compute the function for any point $\mathbf{x} \in \mathcal{S}[\psi]$.

We note that the direct representation of a function $f:\mathbb{R}^n \to \mathbb{R}$ in (41) requires $|\Gamma|$ parameters, which can be viewed as the area of the green box in Fig. 3. By contrast, the above representation only requires $|\Gamma| \ominus |\Gamma : \Lambda|$ anchor points, which can be viewed as the area of the gray region in Fig. 3. The more efficient representation allows the learning of complex functions from few data points, especially in high dimensional applications.

We demonstrate the above local function representation result in a 2D setting in Fig. 10. Specifically, the original band-limited function is with bandwidth 13 × 13. The direct representation of the function has 13 × 13 = 169 degrees of freedom. Now, if we only care about points on a curve which is with bandwidth 3 × 3, then the same function living on the curve can be represented exactly using 48 anchor points, thus significantly reducing the degrees of freedom. However, note that the above representation is only exact on the curve. We note that the function goes to zero as one moves away from the curve.

The choice of anchor points depends on the geometry of the surface, including the number of irreducible components. For arbitrary training samples, we can estimate the unknowns **F** in (47) from the linear relations

$$\underbrace{\left[\mathbf{y}_1, \ldots \mathbf{y}_P\right]}_{\mathbf{Y}} = \mathbf{F}\underbrace{\left[\alpha(\mathbf{x}_1), \ldots, \alpha(\mathbf{x}_P)\right]}_{\mathbf{Z}} \tag{48}$$

as $\mathbf{F} = \mathbf{Y}\mathbf{Z}^H(\mathbf{Z}\mathbf{Z}^H)^\dagger$ The above recovery is exact when we have $N = r$ achor points because **Z** has full column rank in this case. The reason why **Z** has full column rank is due to (44) and (45). Equation (44) suggests that rank($\mathbf{Z}$) ≥ N, while equation (45) shows rank($\mathbf{Z}$) < N. Therefore, we have rank($\mathbf{Z}$) = N, indicating that **Z** has full rank in this case. When $N > r$, the **F** is obtained using the pseudo-inverse, which is based on the least square approximation.

5.3 Efficient computation using *kernel trick*

We use the *kernel-trick* to evaluate $\mathscr{K}(\mathrm{A})$ and $\mathbf{k}_\mathrm{A}$ ($\mathbf{x}$), thus eliminating the need to explicitly evaluating the features of the anchor points and **x**. Each entry of the matrix $\mathscr{K}(\mathrm{A})$ is computed as in (34), while the vector $\mathbf{k}_\mathrm{A}$ ($\mathbf{x}$) is specified by:

$$(\mathbf{k}_\mathbf{A}(\mathbf{x}))_i = \underbrace{\Phi_\Gamma(\mathbf{a}_i)^H \Phi_\Gamma(\mathbf{x})}_{\kappa(\mathbf{a}_i, \mathbf{x})}, \tag{49}$$

which can be evaluated efficiently as nonlinear function κ (termed as kernel function) of their inner-products in $\mathbb{R}^n$. We now consider the kernel function κ for specific choices of lifting.

Using the lifting in (20), we obtain the kernel as

$$\kappa(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{k} \in \Gamma} \exp\left(j2\pi\mathbf{k}^T(\mathbf{y} - \mathbf{x})\right).$$

Note that the kernel is shift invariant in this setting. Since $\kappa : \mathbb{R}^n \to \mathbb{R}$ is an n dimensional function, evaluating and storing it is often challenging in multidimensional applications. We now focus on approximating the kernel efficiently for fast computation. We consider the impact of the shape of the bandwidth set Γ on the shape of the kernel. Specifically, we consider sets of the form

$$\Gamma = \left\{\mathbf{k} \in \mathbb{Z}^n, \|\mathbf{k}\|_q \le d\right\}, \tag{50}$$

where d denotes the size of the bandwidth. The integer q specifies the shape of Γ [55], which translates to the shape of the kernel

$$k_{d,n}^{q}(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbb{Z}^{n}, \|\mathbf{k}\|_{q} \le d} \exp\left(j2\pi \mathbf{k}^{T}\mathbf{x}\right). \tag{51}$$

We term the q =1 case as the diamond Dirichlet kernel. If $q = 2$, we call it the circular Dirichlet kernel. We call the Dirichlet kernel the cubic Dirichlet kernel if $q = \infty$. See Figure 11 for the bandwidth and Figure 12 to see the associated kernel.

We note from the above figures that the circular Dirichlet kernel ($q = 2$) is roughly circularly symmetric, unlike the triangular or diamond kernels. This implies that we can safely approximate it as

$$\kappa(\mathbf{x}, \mathbf{y}) \approx g\left(\|\mathbf{x} - \mathbf{y}\|^{2}\right) \tag{52}$$

where $g: \mathbb{R}_{+} \to \mathbb{R}$. We note that this approximation results in significantly reduced computation in the multidimensional case. The function g may be stored in a look-up table or computed analytically. We use this approach to speed up the computation of multidimensional functions in Section 6.

An additional simplification is to assume that **x** and **y** are unit-norm vectors. In this case, we can approximate

$$g(\|\mathbf{x}_{i} - \mathbf{y}_{i}\|_{2}^{2}) = g(\|\mathbf{x}_{i}\|_{2}^{2} + \|\mathbf{y}_{i}\|_{2}^{2} - 2\langle \mathbf{x}_{i}, \mathbf{y}_{i}\rangle) \approx g(2 - 2\langle \mathbf{x}, \mathbf{y}\rangle) = :\gamma(\langle \mathbf{x}, \mathbf{y}\rangle), \tag{53}$$

where $\boldsymbol{\gamma}(z) = \mathrm{g}(1 - z/2)$. Here, we term $\boldsymbol{\gamma}$ as the activation function. While we do not make this simplifying assumption in our computations, it enables us to show the similarity of the computational structure of (46) to current neural network. The plot of this activation function, along with commonly used activation functions, is shown in Figure 12 (d).

With the aforementioned analysis, we can then rewrite (46) as

$$\mathbf{f}(\mathbf{x}) = \underbrace{[\mathbf{f}_{1}, \ldots, \mathbf{f}_{N}]}_{\mathbf{F}} \mathscr{K}(\mathbf{A})^{\dagger} \underbrace{\begin{bmatrix} g(\|\mathbf{x} - \mathbf{a}_{1}\|^{2}) \\ \vdots \\ g(\|\mathbf{x} - \mathbf{a}_{N}\|^{2}) \end{bmatrix}}_{\mathbf{k}_{\mathbf{A}}(\mathbf{x})} \tag{54}$$

$$\approx \underbrace{\mathbf{F}\mathscr{K}(\mathbf{A})^{\dagger}}_{\widetilde{\mathbf{F}}} \underbrace{\begin{bmatrix} \gamma(\langle \mathbf{x}, \mathbf{a}_{1}\rangle) \\ \vdots \\ \gamma(\langle \mathbf{x}, \mathbf{a}_{N}\rangle) \end{bmatrix}}_{\Gamma_{\mathbf{A}}(\mathbf{x})} \tag{55}$$

In the second step, we used the approximation in (53).

5.4 Optimization of the anchor points and coefficients

The above results show the existence of a computational structure of the form (55) with N anchor points $a_1,.., a_N$ on the surface and the corresponding coefficients $\tilde{\mathbf{f}}_1, .., \tilde{\mathbf{f}}_N$ that can represent the function exactly. We note that the anchor points need not to be selected as a subset of the training data. We note that Corollary 11 guarantees $\mathscr{K}(\mathbf{A})$ to have full column rank as $N = r$. However, the condition number of this matrix may be poor, depending on the choice of the anchor points. It may be worthwhile to choose the anchors such that the condition number of $\mathscr{K}(\mathbf{A})$ is low, which will reduce the noise amplification in (45).

We hence propose to solve for the anchor points $\mathbf{A}$ and the corresponding coefficients $\tilde{\mathbf{F}}$ such that it minimizes the least square error evaluated on the training data:

$$\tilde{\mathbf{F}}^{*}, \mathbf{A}^{*} = \underset{\tilde{\mathbf{F}}, \mathbf{A}}{\operatorname{argmin}} \sum_{i=1}^{N_{\text{train}}} \|\tilde{\mathbf{F}}\, \boldsymbol{\Gamma}_{\mathbf{A}}(\mathbf{x}_i) - \mathbf{y}_i\|^2 \tag{56}$$

We propose to minimize the above expression using stochastic gradient descent. This approach will allow the choice of the anchor points $\mathbf{a}_1,.., \mathbf{a}_N$.

6 Relation to neural networks

We now briefly discuss the close relation of the proposed framework with neural networks. We consider the function learning setting, which is considered in Section 5 and show that the computational structure closely mimics a neural network with one hidden layer. We discuss briefly the benefits of depth in improving the representation. We also show that the above framework can be used to approximate the learning of a manifold from data, which can be viewed as a signal subspace alternative to the null-space approach considered in Section 4. We also show that the computational structure closely mimics an auto-encoder.

6.1 Task/function learning from input output pairs

We now focus on the learning of a function (54) from training data pairs and will show its equivalence with neural networks. Note that the computation involves the inner product of the input signal $\mathbf{x}$ with templates $\mathbf{a}_i$; $i = 1,.., N$, followed by the non-linear activation function γ to obtain $\mathbf{k}A(\mathbf{x})$. These terms are then weighted by the fully connected layer $\mathscr{K}(\mathbf{A})^{\dagger}$ A followed by weighting by the second fully connected layer $\tilde{\mathbf{F}}$. See Fig. 13 for the visual illustration.

As noted above, the representation using anchor points to reduce the degrees of freedom significantly compared to the direct representation. However, we note that the number of parameters needed to represent a high bandwidth function in high dimensions is still high. We now provide some intuition on how the low-rank tensor approximation of functions and composition can explain the benefit of common operations in deep networks.

We now consider the case when the band-limited multidimensional function $f: \mathbb{R}^n \to \mathbb{R}$ in (41) can be approximated as

$$f(\mathbf{x}) = \left(\sum w_i\ f_i(\mathbf{x})\right)^2. \qquad (57)$$

Clearly, the bandwidth of f is almost twice that of $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, showing the benefit of adding layers. While an arbitrary function with the same bandwidth as f cannot be represented as in (57), one may be able to approximate it closely. The new layer will have a quadratic non-linearity Q, if the function has the form (57). Note that one may use arbitrary non-linearity in place of the quadratic one in (57).

Similarly, one may perform a low-rank tensor approximation of an arbitrary N dimensional function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Specifically, the approximation involves the sum of products of 1-D functions.

$$f(x_1, .., x_N) \approx \sum_{i=1}^{r} h_1^{(i)}(x_1) \cdot h_2^{(i)}(x_2) \ldots h_N^{(i)}(x_N), \qquad (58)$$

where $h_i : \mathbb{R} \rightarrow \mathbb{R}$. The above sum of products can also be realized by taking weighted linear combination of 1-D functions, followed by a non-linearity as in (57). This allows one to have a hierarchical structure, where lower dimensional functions are pooled together to represent a multidimensional function.

In image processing applications, the functions to be learned are shift-invariant. This allows one to learn functions of small image patches (e.g. 3×3) of a specified dimension at each layer. The functions on nearby pixels in the output thus correspond to information from different 3 × 3 neighborhoods. The low-dimensional functions from non-overlapping 3 × 3 neighborhoods could be combined with downsampling as in (58) to represent a high dimensional function (e.g. 9 × 9) neighborhoods. The process can be repeated to improve the efficiency of representation.

6.2 Relation to auto-encoders

We note that the space of band-limited functions of the form (41) can reasonably approximate lower order polynomials in $\mathbb{R}^n$ for sufficiently high bandwidth Γ [49]. In particular, let us assume that there exists a set of coefficients β such that

$$\mathbf{x} \approx \widetilde{\mathbf{x}} = \sum_{\mathbf{k} \in \Gamma} \beta_{\mathbf{k}} \exp\left(j2\pi \mathbf{k}^T \mathbf{x}\right) \qquad (59)$$

In this case, the above results imply that one can represent any point on the surface $\mathcal{S}[\psi]$ as

$$\mathbf{x} \approx \underbrace{[\mathbf{a}_1, \ldots, \mathbf{a}_n]}_{\mathbf{A}} \underbrace{\mathscr{K}(\mathbf{A})^{\dagger} \mathbf{k}_{\mathbf{A}}(\mathbf{x})}_{\boldsymbol{\alpha}(\mathbf{x})} \qquad (60)$$

We note that the resulting network is hence essentially an auto-encoder. Specifically, the inner-products between the feature vectors of $\mathbf{x}$ and the anchor point a_i denoted by $\alpha(\mathbf{x})$ can be viewed as the latent features or compact code. As described previously, the coefficients $\alpha = \mathscr{K}(\mathrm{A})^{\dagger}\mathbf{k}_{\mathbf{A}}(\mathbf{x})$ captures the geometry of the surface, while the top layer **A** is the decoder that recover the signal from its latent vectors.

We note that the surface recovery algorithms in Section 4 follow a null-space approach, where we identify the null-space of the feature space or equivalently the annihilation functions from the samples of the surface. Specifically, the sum of squares of the null-space functions in Section 4.2 provides a measure of the error in projecting the feature vector to the null-space of the feature matrix.

$$\gamma(\mathbf{x}) = \sum_{i=1}^{|\Gamma \ominus \Lambda|} |\mu_i(\mathbf{x})|^2 = \sum_{i=1}^{|\Gamma \ominus \Lambda|} \left|\mathbf{n}_i^T \Phi_{\Gamma}(\mathbf{x})\right|^2 \tag{61}$$

$$= \|\mathbf{N}\ \Phi_{\Gamma}(\mathbf{x})\|^2 \tag{62}$$

where $\mathbf{n}_{\boldsymbol{i}}$ are the null-space vectors. The projection energy is zero if the point **x** is on $\mathscr{S}$ and is high when it is far from it.

By contrast, the auto-encoder approach can be viewed as a signal subspace approach, where we project the samples to the basis vectors specified by the feature vectors of the anchors $\Phi_{\Gamma}(\mathrm{a}_{\boldsymbol{i}})$. Specifically, we use the non-linearity specified by (53) and trained the network parameters (**A** as well as the weights of the inner-products) using stochastic gradient descent. The training data corresponds to randomly drawn points on the surface. To ensure that the network learns a projection, we trained the network as a denoising auto-encoder; the inputs correspond to samples on the surface corrupted with Gaussian noise, while the labels are the true samples. Once the training is complete, we plot the approximation error

$$E(\mathbf{x}) = \|\mathbf{x} - \mathrm{F}\mathscr{K}(\mathrm{A})^{\dagger}\mathbf{k}_{\mathbf{A}}(\mathbf{x})\|^2 = \|\underbrace{(\mathbf{I} - \mathbf{F}\mathscr{K}(\mathrm{A})^{\dagger}\mathbf{k}_{\mathrm{A}})}_{\mathscr{R}}(\mathbf{x})\|^2 \tag{63}$$

as a function of the input point in Fig. 14.

We trained the network using the exemplar curve shown in Fig. 8. We randomly choose 1000 points on the curve as the training data and 250 features are chosen in the middle layer. The bandwidth of the Dirichlet kernel is chosen to be 15. The trained network is then used to learn the curve. The learned results are shown in Fig. 14. From which one can see that the proposed learning framework performs well. We note that the projection error is close to zero on the surface, while it is high if it is away from the surface. Note that this closely mimics the plot in Fig. 8. Once trained, the surface can be estimated in low-dimensional settings as the zero set of the projection error as shown in Fig. 14.(b), which closely approximates the true curve in (c). We note that $\mathscr{R}$ can be viewed as a residual denoising auto-encoder. Once trained, this network can be used as a prior in inverse problems as in [2], where we have used the null-space network in Section 6. We have also used the null-space prior (61) in our prior work [39], where the null-space basis was learned as described in Section 4.3.

7 Illustration in denoising

We now illustrate the preliminary utility of the proposed network in image denoising. Specifically, we consider the learning of a function $f:\mathbb{R}^{p^2} \to \mathbb{R}$, which predicts the denoised center pixel of a patch from the noisy $p \times p$ patch. The function f in p^2 dimensional space is associated with a large number of free parameters; learning of these unknowns are challenging due to the curse of dimensionality. Then the result in the previous section offers a work-around, which suggests that the function can be expressed as the linear combination of the features of "anchor-patches", weighted by $\mathbf{p}$.

We propose to learn the anchor patches $\mathbf{a}_i$ and the function values $f(\mathbf{a}_i)$ from exemplar data using stochastic gradient descent to minimize (56). Note that the learned representation is valid for any patch, and hence the proposed scheme is essentially a convolutional neural network. The difference of our structure in (55) with the commonly used convolutional neural networks (CNN) structure is the activation function $\boldsymbol{\gamma}$. We replaced the ReLU non-linearity in a network with the proposed function $\boldsymbol{\gamma}$ in a single layer network. For the two-layer network, we replaced the ReLU non-linearity with $\boldsymbol{\gamma}$ and Q as indicated in (57).

We first tested the performance of the network on the MNIST dataset [20]. In the experiments, we choose the patch size to be 7 × 7 and $d = 7$ in (51). We also trained a ReLU network with the same parameters for comparison. Besides, we compared the proposed scheme against non-local means (NLM) and dictionary learning (DL) [9]. All algorithms, except for NLM were trained using the MNIST training set provided in TensorFlow. For the proposed network and the ReLU network, they are trained using 300 epoches and for the dictionary learning method, 500 iterations are used to learn the dictionaries. The comparison of the testing results is shown in Figure 15. The comparison of the PSNR is reported in the caption. The results show that the neural network based approaches offer improved performance compared to dictionary learning and non-local methods. Our results also show that the proposed networks provide comparable, if not slightly better performance, compared to the ReLU networks. The results also show the slight improvement in performance offered by the proposed two-layer networks over single layer networks.

The size of the image in the MNIST dataset is small. To better demonstrate the performance of the proposed network, we also applied the proposed scheme to the denoising of natural images. The algorithm was trained on the images of Hill, Cameraman, Couple, Bridge, Barbara and Boat at three different noise settings. We assume the noise is Gaussian white noise in the natural images setting. We compared the proposed scheme against dictionary learning (DL), non-local means (NLM) and transform learning (TL) [42]. In the experiments for natural images, the patch size is chosen as 9 × 9 and $d = 7$ in (51). For the proposed network and the ReLU network, they are trained using 300, 400, 450 epoches corresponding to the noise level $\sigma = 10$, 20,100, and for the dictionary learning method, 500 iterations are used to learn the dictionaries. We then tested the denoising performance on two natural images: Man and Lighthouse.

The quantitative results (PSNR) of the algorithm are shown in Table 1, while the results on Man and Lighthouse with noise of standard deviation $\sigma = 20$ are shown in Fig. 16 and Fig.

17. In Table 1, Fig. 16 and Fig. 17, “ReLU1” and “ReLU2” represent one-layer ReLU network and two-layer ReLU network, while “Proposed1” and “Proposed2” stand for the proposed one-layer network and proposed two-layer network. The results show that the performance of the neural network schemes is superior to classical methods and the proposed networks provide comparable or slightly better performance than the ReLU networks.

(8) Conclusion

In this work, we considered a data model, where the signals are localized to a surface that is the zero level set of a band-limited function ψ. The bandwidth of the function can be seen as a complexity measure of the surface. We show that the non-linear features of the samples, obtained by an exponential lifting, satisfy an annihilation relation. Using the annihilation relation, we developed theoretical sampling guarantees for the unique recovery of the surface. Our main contribution here is to prove that with probability 1, the surface can be uniquely recovered using a collection of samples, whose number is equal to the degrees of freedom of the representation. When the true bandwidth of the surface is unknown, which is usually the case, we introduced a method using the SoS polynomial to specify the surface. We also introduced the way to get back the samples when the original samples are corrupted by noise.

We then use this model to efficiently represent arbitrary band-limited functions f living on the surface. We show that the exponential features of the points on the surface live in a low-dimensional subspace. This subspace structure is used to represent the f efficiently using very few parameters. We note that the computational structure of the function evaluation mimics a single-layer neural network. We applied the proposed computational structure to the context of image denoising.

Acknowledgments

The authors would like to thank Dr. Greg Ongie for the valuable discussion about the proofs of the results in this paper. The first author would also like to thank Prof. Theodore Shifrin for his help during the discussion on the intersection of two surfaces and Mr. Biao Ma for the discussion on Lemma 15 and Corollary 16.

9: Appendix

9.1 Proof of Proposition 1

As we mentioned in Section 2.3.3, if we have a (hyper-)surface $\mathcal{S}$ which is given by the zero level set of a trigonometric polynomial, then there will be a minimal polynomial which defines $\mathcal{S}$ (Proposition 1). To prove this result, we need the following famous result.

Lemma 12 (Hilbert’s Nullstellensatz [3]).

Let $\mathbb{K}$ be an algebraically closed field (for example $\mathbb{C}$). Suppose $I \subset \mathbb{K}[x_1, \cdots, x_n]$ is an ideal of polynomials, and $\mathcal{Z}(I)$ denotes the set of common zeros of all the polynomials in I. $\mathcal{I}(\mathcal{Z}(I))$ represents the ideal of polynomials in $\mathbb{K}[x_1, \cdots, x_n]$ vanishing on $\mathcal{Z}(I)$. Then, we have

$$\mathcal{I}(\mathcal{Z}(I)) = \sqrt{I},$$

where $\sqrt{I}$ denotes the radical of I, specified by the set

$$\sqrt{I} = :\left\{p \mid p^n \in I,\ for\ some\ n \in \mathbb{Z}^+\right\} \tag{64}$$

Remark 1.—We say a set $I \subset K[x_1, \cdots, x_n]$ is an ideal, if I is closed under the addition operation (e.g. addition“+”), satisfies the associative property, has a unit element 0, and a valid inverse for every element in I. For the operation multiplication (e.g. “·”), we have $r \cdot p \in I$ and $p \cdot r \in I$ for any $r \in K[x_1, \cdots, x_n]$

Remark 2.—An important property of the radical of the ideal I is that $I \subset \sqrt{I}$. Note that setting $n = 1$ in (64) will yield I.

Remark 3.—The above lemma states that the set of all polynomials that vanish on the common zeros $\mathcal{Z}(I)$ of the polynomials in I is given by $\sqrt{I} \supset I$. Specifically, if we are given another polynomial $\eta(\mathbf{x})$ that also vanishes on the common zero set $\mathcal{Z}(I)$, then there must be positive integer n such that $\eta^n(\mathbf{x}) \in I$.

We denote the ideal generated by a function f by $(f) = \{\mu | \mu = f\gamma\}$, where γ is an arbitrary polynomial. The identity in this ideal is the zero polynomial. In particular, (f) is the family of all functions that have f as a factor. We note that the set of common zeros of all the functions in (f), denoted by $\mathcal{Z}[(f)]$ is the same as the zero set of f, denoted by $Z[f]$.

Lemma 13.

Let f, g be two polynomials in $\mathbb{C}[x_1, \cdots, x_n]$ with the same zero set. Then the two polynomials must have (up to scaling) the same factors.

Proof.—Suppose $Z[f] = Z[g] = Z$ is the zero set of f and g. Since $Z[f] = \mathcal{Z}[(f)]$, we have $\mathcal{Z}[(f)] = \mathcal{Z}[(g)] = Z$. By the Hilbert’s Nullstellensatz, we have

$$\mathcal{I}(\mathcal{Z}(f)) = \sqrt{(f)}, \qquad \mathcal{I}(\mathcal{Z}(g)) = \sqrt{(g)}.$$

Since Z(f) = Z(g), we then have $\mathcal{I}(\mathcal{Z}(f)) = \mathcal{I}(\mathcal{Z}(g))$ and hence $\sqrt{(f)} = \sqrt{(g)}$ As mentioned above, we have $I \subset \sqrt{I}$ for any ideal I. Therefore, we have $(f) \subset \sqrt{(f)}$ and $(g) \subset \sqrt{(g)}$. This implies that $f \in \sqrt{(f)}$ and $g \in \sqrt{(g)}$. Because we have $\sqrt{(f)} = \sqrt{(g)}$, we can obtain that $f \in \sqrt{(g)}$ and $g \in \sqrt{(f)}$. By which we have that there exist m, $m, n \in \mathbb{Z}$ and $p, q \in \mathbb{C}[x_1, \cdots, x_n]$ such that

$$f^n = p \cdot g, \qquad g^m = q \cdot f.$$

Therefore, we can obtain that the irreducible factors of g are of f as well and vice versa, which proves the desired conclusion. □

With this conclusion, we can now prove Proposition 1.

Proof of Proposition 1.—The proof of the existence and uniqueness about ψ is same as the proof of Proposition A.3 in [30] and thus we omit them here.

In this proof, we show that $BW(\psi) \subseteq BW(\psi_1)$. Note that the algebraic surface $X = \{p = \mathscr{P}[\psi] = 0\}$ is the union of irreducible surfaces $X_j = \{p_{i_j} = 0\} \subset \mathbb{C}^n$. Define

$$\nu(x_1, \cdots, x_n) = \left(e^{j2\pi x_1}, \cdots, e^{j2\pi x_n}\right).$$

Let $\mathscr{S}_j = \nu^{-1}\left(X_j \cap \mathbb{T}^n\right)$. Then we have a decomposition of $\mathscr{S}$ as the union of surfaces $\mathscr{S}_j$. If ψ_1 is another trigonometric polynomial with $\mathscr{S}$ as the zero level set as well. Then ψ_1 vanishes on each $\mathscr{S}_j$. Let $q = \mathscr{P}[\psi_1]$. Then we have $q = 0$ on the infinite set $\nu\left(\mathscr{S}_j\right)$, by which we can infer that q and p will have the same zero set using Theorem 14. Then by Lemma 13, we have $p \mid q$, which implies that $BW(\psi) \subseteq BW(\psi_1)$. □

9.2 Proof of results in Section 4

The key property of surfaces that we exploit is that the dimension of the intersection of two band-limited surfaces of dimension k is strictly lower than k, provided their level set functions do not have any common factors. Hence, if we randomly sample one of the surfaces, the probability that the samples fall on the intersection of the two surfaces is zero. This result enables us to come up with the sampling guarantees. We will now show the results about the intersections of the zero sets of two trigonometric surfaces.

9.2.1 Intersection of surfaces

We will first state a known result about the intersection of the zero sets of two polynomials (non-trigonometric) whose level set functions do not have a common factor.

Theorem 14 ([15],pp.115, Theorem 14).—*Let $\mathscr{S}[\psi]$ and $\mathscr{S}[\eta]$ be two surfaces of dimension n–1 over a field $\mathbb{K}$, which are the zero sets of the polynomials $\psi: \mathbb{K}^n \to \mathbb{K}$ and $\eta: \mathbb{K}^n \to \mathbb{K}$, respectively. If ψ and η do not have a common factor, then*

$$\dim(\mathscr{S}[\psi] \cap \mathscr{S}[\eta]) < n - 1.$$

The above result is a generalization of the two dimensional case $\left(\mathbb{C}^2\right)$ in [30], where Bézout's inequality was used to prove the result. Specifically, the result in [30] suggests that the intersection of two curves consists of a set of isolated points, if their potential function does not have any common factor. Theorem 14 generalizes the above result to $n > 2$; it suggests that the intersection of two surfaces with dimension k is another surface, whose dimension is strictly less than k. For instance, the intersection of two 3-D surfaces which are given by the zero level set of some polynomials, could yield 2D curves or isolated points. We now extend Theorem 14 to trigonometric polynomials using the mapping ν specified by (7).

Lemma 15.—*Let* $\mathcal{S}[\psi]$ *and* $\mathcal{S}[\eta]$ *within* $[0,1]^n \subset \mathbb{R}^n$ *be two surfaces of dimension* $n-1$ *over* $\mathbb{R}$*, which are the zero level sets of the trigonometric polynomials* ψ *and* η*. Suppose* ψ *and* η *do not have a common factor, then*

$$\dim(\mathcal{S}[\psi] \cap \mathcal{S}[\eta]) < n-1.$$

Proof. Let $\nu = (\nu_1, \cdots, \nu_n)$ be defined by (7). We now would like to prove the result by way of contradiction. Suppose

$$\dim(\mathcal{S}[\psi] \cap \mathcal{S}[\eta]) = \dim(\mathcal{S}[\psi]) = \dim(\mathcal{S}[\eta]) = n-1.$$

This implies that $\nu(\mathcal{S}[\psi] \cap \mathcal{S}[\eta])$ will have the same dimension of $\nu(\mathcal{S}[\psi])$ and $\nu(\mathcal{S}[\eta])$. However, this is impossible according to Theorem 14. Therefore, we have the desired result. □

Based on this lemma, we can directly have the following Corollary.

Corollary 16.—*Suppose* $\psi(\mathbf{x})$, $\eta(\mathbf{x})$, $\mathbf{x} \in [0,1]^n$ *are two trigonometric polynomials as in Lemma 15. Consider the* $n-1$ *dimensional Lebesgue measure on* $\mathcal{S}[\psi]$*. Then this Lebesgue measure of the intersection of the zero level sets of the trigonometric polynomials is zero, i.e.,*

$$m(\mathcal{S}[\psi] \cap \mathcal{S}[\eta]) = 0.$$

The Lebesgue measure can be viewed as the area of the n−1 dimensional surface. For example, when $n = 3$, $\mathcal{S}[\psi]$ and $\mathcal{S}[\eta]$ are 2-D surfaces, while their intersection is a 1-D curve or a set of isolated points with zero area.

9.2.2 Proof of Proposition 4

Proof.—We note that $N \geq |\Lambda| - 1$ is a necessary condition for the matrix to have a rank of $|\Lambda| - 1$. We now assume that the surface is sampled with $N \geq |\Lambda| - 1$ random samples, chosen independently, denoted by $\mathbf{x}_i; i = 1, \cdots, N \in \mathcal{S}[\psi]$. Since $\mathbf{c} \leftrightarrow \psi$ is a valid non-trivial null-space vector for the feature matrix $\Phi_\Lambda(\mathbf{X})$ formed from these samples, we have rank$(\Phi_\Lambda(\mathbf{X})) \leq |\Lambda| - 1$. The polynomial $\psi(\mathbf{x}) = \mathbf{c}^T\Phi_\Lambda(\mathbf{x})$ is the minimal irreducible polynomial that defines the surface.

We now prove the desired result by contradiction. Assume that these exists another linearly independent null-space vector $\mathbf{d} \leftrightarrow \eta$ or equivalently the rank of $\Phi_\Lambda(\mathbf{X})$ is strictly less than $|\Lambda|-1$. Since $\mathbf{c}$ and $\mathbf{d}$ are linearly independent and $\psi(\mathbf{x})$ is the minimal polynomial, we know that $\psi(\mathbf{x})$ and $\eta(\mathbf{x})$ will not share a common factor. Also note that $\mathbf{x}_i \in \mathcal{S}[\psi] \cap \mathcal{S}[\eta]$. However, since $\psi(\mathbf{x})$ and $\eta(\mathbf{x})$ do not share a common factor, the probability of each sample to be at the intersection of the two polynomials $(\mathbf{x}_i \in \mathcal{S}[\psi] \cap \mathcal{S}[\eta])$ is zero by Corollary 16. Therefore, with probability 1 that such $\mathbf{d}$ does not exist, meaning that with probability 1 that the feature matrix will be of rank $|\Lambda| - 1$ when $N \geq |\Lambda| - 1$. □

9.2.3 Proof of Proposition 6

Proof.—We note that $N \geq |\Lambda| - 1$ is a necessary condition for the matrix to have a rank of $|\Lambda| - 1$. We now assume that the surface is sampled with N random samples $\mathbf{x}_i$; $i = 1, \cdots, N$ satisfying the conditions in Proposition 6. The minimal polynomial $\psi(\mathbf{x}) = \mathbf{c}^T\Phi_\Lambda(\mathbf{x})$ that defines the surface can be factorized as $\psi(\mathbf{x}) = \psi_1(\mathbf{x}) \cdot \psi_2(\mathbf{x}) \cdots \psi_M(\mathbf{x})$.

We will prove the result by contradiction. Assume that these exists another linearly independent null-space vector $\mathbf{d} \leftrightarrow \eta$, or equivalently the rank of $\Phi_\Lambda(\mathbf{X})$ is less than $|\Lambda| - 1$. Since $\mathbf{c}$ and $\mathbf{d}$ are linearly independent, ψ and η should differ by at least one factor. Without loss of generality, let us assume that $\eta(\mathbf{x}) = \mu(\mathbf{x})\prod_{i=1}^{M-1}\psi_i(\mathbf{x})$, where μ is an arbitrary polynomial of bandwidth Λ_M. Besides, μ and ψ_M does not share a factor. Using the result of Proposition 4, we see that the probability of μ and an irreducible ψ_M vanish at $|\Lambda_i| - 1$ independently drawn random locations is zero. If multiple factors are shared, the same argument can be extended to each one of the factors independently. □

9.2.4 Proof of Proposition 8

Proof.—We note that $N \geq |\Gamma| - |\Gamma \ominus \Lambda|$ is a necessary condition for the matrix to have the specified rank. We now assume that the surface is sampled with $N \geq |\Gamma| - |\Gamma \ominus \Lambda|$ random samples, chosen independently. We note that $\mathbf{c} \leftrightarrow \psi$ specified by (10), as well as the $|\Gamma \ominus \Lambda|$ translates of $\mathbf{c}$ within Γ, are valid linearly independent null-space vectors of $\Phi_\Lambda(\mathbf{X})$. We thus have

$$\operatorname{rank}(\Phi_\Lambda(\mathbf{X})) \leq |\Gamma| - |\Gamma \ominus \Lambda| \tag{65}$$

We will show that the rank condition can be satisfied with probability 1 by contradiction. Assume that these exists another linearly independent null-space vector $\mathbf{d} \leftrightarrow \eta$ or equivalently the rank of $\Phi_\Lambda(\mathbf{X})$ is less than $|\Gamma| - \Gamma \ominus \Lambda|$. Since $\mathbf{d}$ are linearly independent with $\mathbf{c}$ and its translates within Γ, we cannot express $\mathbf{d}$ as the linear combinations of the the other null-space vectors. Specifically, we have

$$\eta(\mathbf{x}) \neq \sum_{\mathbf{k} \in \Gamma \ominus \Lambda} \alpha_{\mathbf{k}}\psi(\mathbf{x})\exp\left(j2\pi\mathbf{k}^T\mathbf{x}\right) \tag{66}$$

$$= \psi(\mathbf{x})\underbrace{\sum_{\mathbf{k} \in \Gamma \ominus \Lambda} \alpha_{\mathbf{k}}\exp\left(j2\pi\mathbf{k}^T\mathbf{x}\right)}_{\gamma(\mathbf{x})} = \psi(\mathbf{x})\gamma(\mathbf{x})\,. \tag{67}$$

Here $\alpha_{\mathbf{k}}$ is an arbitrary coefficients and hence γ is an arbitrary polynomial. The linear independence property implies that $\eta(\mathbf{x})$ cannot have $\psi(\mathbf{x})$ as a factor. Since $\psi(\mathbf{x})$ is the minimal polynomial, this also means that η and ψ does not have any common factor.

Consider now the random sampling set $\mathbf{x}_i; i = 1..|\Gamma| - |\Gamma \ominus \Lambda|$. We have

$$\mathbf{c}^T\Phi_\Lambda(\mathbf{x}_i) = \mathbf{d}^T\Phi_\Lambda(\mathbf{x}_i) = 0,\ i = 1,\cdots,\left|\Lambda\right| - 1.$$

This implies that $\mathbf{x}_i \in \mathcal{S}[\psi] \cap \mathcal{S}[\eta]$. However, since $\psi(\mathbf{x})$ and $\eta(\mathbf{x})$ do not share a common factor, the probability of each sample to be at the intersection of the two polynomials ($\mathbf{x}_i \in \mathcal{S}[\psi] \cap \mathcal{S}[\eta]$) is zero by Corollary 16. Therefore, we have rank $(\Phi_\Lambda(\mathbf{X})) = |\Gamma| - |\Gamma \ominus \Lambda|$ with probability one.

□

9.2.5 Proof of Proposition 9

Proof.—We note that $N \geq |\Gamma| - \Gamma \ominus \Lambda|$ is a necessary condition for the matrix to have the specified rank. We now assume that the surface is sampled with N random samples satisfying the sampling conditions in Proposition 9. The minimal polynomial $\psi(\mathbf{x}) = \mathbf{c}^T\Phi_\Lambda(\mathbf{x})$ that defines the surface can be factorized as $\psi(\mathbf{x}) = \psi_1(\mathbf{x}) \cdot \psi_2(\mathbf{x})\cdots\psi_M(\mathbf{x})$.

Assume that there exists another linearly independent null-space vector $\mathbf{d} \leftrightarrow \eta$ or equivalently the rank of $\Phi_\Lambda(\mathbf{X})$ is less than $|\Gamma| - \Gamma \ominus \Lambda|$. Similar to the above arguments, if η and ψ does not have any common factors, the rank condition is satisfied with probability 1. Similar to Section 9.2.3, linear independence implies that $\eta(\mathbf{x})$ cannot be a factor of ψ; there is at least one factor ψ_i that is distinct. Based on Proposition 8, these factors cannot vanish on more than $|\Gamma_i| - |\Gamma_i \ominus \Lambda_i|$ common samples. □

References

[1]. Aim@shape, digital shape workbench. http://www.infra-visionair.eu/.

[2]. Aggarwal HK, Mani MP, and Jacob M, Modi: Model-based deep learning architecture for inverse problems, IEEE transactions on medical imaging, 38 (2018), pp. 394–405. [PubMed: 30106719]

[3]. Atiyah M and MacDonald I, Introduction To Commutative Algebra, Addison-Wesley series in mathematics, Avalon Publishing, 1994.

[4]. Belkin M, Niyogi P, and Sindhwani V, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, Journal of machine learning research, 7 (2006), pp. 2399–2434.

[5]. Bernard O, Friboulet D, ThÉvenaz P, and Unser M, Variational b-spline level-set: a linear filtering approach for fast deformable model evolution, IEEE Transactions on Image Processing, 18 (2009), pp. 1179–1191. [PubMed: 19403364]

[6]. Buades A, Coll B, and Morel J-M, Non-local means denoising, Image Processing On Line, 1 (2011), pp. 208–212.

[7]. Cho Y and Saul LK, Kernel methods for deep learning, in Advances in neural information processing systems, 2009, pp. 342–350.

[8]. Dabov K, Foi A, Katkovnik V, and Egiazarian K, Image denoising with block-matching and 3D filtering, in Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, N. M. Nasrabadi, S. A. Rizvi, E. R. Dougherty, J. T. Astola, and K. O. Egiazarian, eds., vol. 6064, International Society for Optics and Photonics, SPIE, 2006, pp. 354–365.

[9]. Elad M and Aharon M, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Transactions on Image processing, 15 (2006), pp. 3736–3745. [PubMed: 17153947]

[10]. Federer H, Geometric measure theory, Springer, 2014.

[11]. Fefferman C, Mitter S, and Narayanan H, Testing the manifold hypothesis, Journal of the American Mathematical Society, 29 (2016), pp. 983–1049.

[12]. Fornasier M, Rauhut H, and Ward R, Low-rank matrix recovery via iteratively reweighted least squares minimization, SIAM Journal on Optimization, 21 (2011), pp. 1614–1640.

[13]. Gallese V, The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity, Psychopathology, 36 (2003), pp. 171–180.

[14]. Gedalyahu K and Eldar YC, Time-delay estimation from low-rate samples: A union of subspaces approach, IEEE Transactions on Signal Processing, 58 (2010), pp. 3017–3031.

[15]. Gunning RC AND Rossi H, Analytic functions of several complex variables, vol. 368, American Mathematical Soc, 2009.

[16]. HASTIE T, TlBSHIRANI R, Friedman J, and Franklin J, The elements of statistical learning: data mining, inference and prediction, The Mathematical Intelligencer, 27 (2005), pp. 83–85.

[17]. Jackson AS, Bulat A, Argyriou V, and Tzimiropoulos G, Large pose 3d face reconstruction from a single image via direct volumetric cnn regression, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1031–1039.

[18]. Jacob M, Blu T, and Unser M, Efficient energies and algorithms for parametric snakes, IEEE transactions on image processing, 13 (2004), pp. 1231–1244. [PubMed: 15449585]

[19]. KoULAMAs C AND Kyparisis GJ, Single-machine and two-machine flowshop scheduling with general learning functions, European Journal of Operational Research, 178 (2007), pp. 402–407.

[20]. LeCun Y, Cortes C, AND Burges C, Mnist handwritten digit database, ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist,2 (2010).

[21]. Li C, Xu C, Gui C, and Fox MD, Distance regularized level set evolution and its application to image segmentation, IEEE transactions on image processing, 19 (2010), pp. 3243–3254. [PubMed: 20801742]

[22]. Li T, Krupa A, and Collewet C, A robust parametric active contour based on fourier descriptors, in 2011 18th IEEE International Conference on Image Processing, IEEE, 2011, pp. 1037–1040.

[23]. Lu YM AND Do MN, A theory for sampling signals from a union of subspaces, IEEE transactions on signal processing, 56 (2008), pp. 2334–2345.

[24]. Mairal J, Koniusz P, Harchaoui Z, AND Schmid C, Convolutional kernel networks, in Advances in neural information processing systems, 2014, pp. 2627–2635.

[25]. MisHALI M, Eldar YC, and Elron AJ, Xampling: Signal acquisition and processing in union of subspaces, IEEE Transactions on Signal Processing, 59 (2011), pp. 4719–4734.

[26]. MoHAN K and Fazel M, Iterative reweighted algorithms for matrix rank minimization, The Journal of Machine Learning Research, 13 (2012), pp. 3441–3473.

[27]. Mohsin YQ, Lingala SG, DiBella E, and Jacob M, Accelerated dynamic mri using patch regularization for implicit motion compensation, Magnetic resonance in medicine, 77 (2017), pp. 1238–1248. [PubMed: 27091812]

[28]. Muller K-R, Mika S, Ratsch G, Tsuda K, and Scholkopf B, An introduction to kernel-based learning algorithms, IEEE transactions on neural networks, 12 (2001), pp. 181–201. [PubMed: 18244377]

[29]. NAKARMI U, Wang Y, Lyu J, Liang D, and Ying L, A kernel-based low-rank (klr) model for low-dimensional manifold recovery in highly accelerated dynamic mri, IEEE transactions on medical imaging, 36 (2017), pp. 2297–2307. [PubMed: 28692970]

[30]. Ongie G AND Jacob M, Off-the-grid recovery of piecewise constant images from few fourier samples, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1004–1041. [PubMed: 29973971]

[31]. Ongie G, Willett R, Nowak RD, and Balzano L, Algebraic variety models for high-rank matrix completion, arXiv preprint arXiv:1703.09631, (2017).

[32]. OsHER S AND Fedkiw RP, Level set methods: an overview and some recent results, Journal of Computational physics, 169 (2001), pp. 463–502.

[33]. Pan H, Blu T, AND Dragotti PL, Sampling curves with finite rate of innovation, IEEE Transactions on Signal Processing, 62 (2013), pp. 458–471.

[34]. PeyrÉ G AND Mallat S, Surface compression with geometric bandelets, ACM Transactions on Graphics (TOG), 24 (2005), pp. 601–608.

[35]. Pillonetto G, Dinuzzo F, Chen T, De Nicolao G, and Ljung L, Kernel methods in system identification, machine learning and function estimation: A survey, Automatica, 50 (2014), pp. 657–682.

[36]. PODDAR S AND JACOB M, Dynamic mri using smoothness regularization on manifolds (storm), IEEE transactions on medical imaging, 35 (2015), pp. 1106–1115. [PubMed: 26685228]

[37]. ______, Recovery of noisy points on band,limited surfaces: Kernel methods re-explained, in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 4024–4028.

[38]. _____, Recovery of point clouds on surfaces: Application to image reconstruction, in Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE, 2018, pp. 1272–1275.

[39]. Poddar S, Mohsin YQ, Ansah D, Thattaliyath B, Ashwath R, and Jacob M, Manifold recovery using kernel low-rank regularization: application to dynamic imaging, IEEE Transactions on Computational Imaging, 5 (2019), pp. 478–491. [PubMed: 33768137]

[40]. Potts D and Steidl G, Fourier reconstruction of functions from their nonstandard sampled radon transform, Journal of Fourier Analysis and Applications, 8 (2002), pp. 513–534.

[41]. Rauth M AND Strohmer T, Smooth approximation of potential fields from noisy scattered data, Geophysics, 63 (1998), pp. 85–94.

[42]. Rayishankar S and Bresler Y, Learning doubly sparse transforms for images, IEEE Transactions on Image Processing, 22 (2013), pp. 4598–4612. [PubMed: 23893720]

[43]. Rousson M AND Paragios N, Shape priors for level set representations, in European Conference on Computer Vision, Springer, 2002, pp. 78–92.

[44]. SAJDA P, Laine A, and Zeeyi Y, Multi-resolution and wavelet representations for identifying signatures of disease, Disease markers, 18 (2002), pp. 339–363. [PubMed: 14646044]

[45]. SCHÖLKOPF B and a. J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2002.

[46]. SCHROFF F, Criminisi A, and a. Zisserman, Object class segmentation using random forests, in BMVC, 2008, pp. 1–10.

[47]. Shao H, Kumar A, and Thomas Fletcher P, The riemannian geometry of deep generative models, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 315–323.

[48]. Smola AJ AND Kondor R, Kernels and regularization on graphs, in Learning theory and kernel machines, Springer, 2003, pp. 144–158.

[49]. T. S0REVIK AND M. A. Nome, Trigonometric interpolation on lattice grids, BIT Numerical Mathematics, 56 (2016), pp. 341–356.

[50]. STROHMER T, Computationally attractive reconstruction of bandlimited images from irregular samples, IEEE Transactions on image processing, 6 (1997), pp. 540–548. [PubMed: 18282947]

[51]. STROHMER T, Binder T, and Sussner M, How to recover smooth object boundaries in noisy medical images, in Proceedings of 3rd IEEE International Conference on Image Processing, vol. 1, IEEE, 1996, pp. 331–334.

[52]. Tian C, Xu Y, Fei L, Wang J, Wen J, and Luo N, Enhanced cnn for image denoising, CAAI Transactions on Intelligence Technology, 4 (2019), pp. 17–23.

[53]. Tsakiris MC AND Vidal R, Algebraic clustering of affine subspaces, IEEE transactions on pattern analysis and machine intelligence, 40 (2017), pp. 482–489. [PubMed: 28287957]

[54]. Wang S and Wang MY, Radial basis functions and level set method for structural topology optimization, International journal for numerical methods in engineering, 65 (2006), pp. 2060–2090.

[55]. WEISZ F, Summability of multi-dimensional trigonometric fourier series, Surv. Approx. Theory, 7 (2012), pp. 1–179.

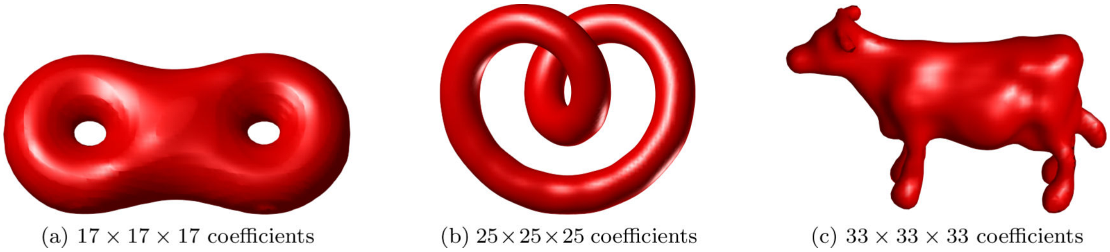
[56]. Yang Z and Jacob M, Nonlocal regularization of inverse problems: a unified variational framework, IEEE Transactions on Image Processing, 22 (2012), pp. 3192–3203. [PubMed: 23014745]

[57]. ZHANG J, Walter GG, Miao Y, and Lee WNW, Wavelet neural networks for function learning, IEEE transactions on Signal Processing, 43 (1995), pp. 1485–1497.

[58]. Zhang K, w. Zuo, Y. Chen, D. Meng, and L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, IEEE Transactions on Image Processing, 26 (2017), pp. 3142–3155. [PubMed: 28166495]

[59]. Zou Q AND Jacob M, Sampling of surfaces and learning functions in high dimensions, in ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 8354–8358.

[60]. Zou Q, Poddar S, and Jacob M, Sampling of planar curves: Theory and fast algorithms, IEEE Transactions on Signal Processing, 67 (2019), pp. 6455–6467.


(a) $17 \times 17 \times 17$ coefficients

(b) $25 \times 25 \times 25$ coefficients

(c) $33 \times 33 \times 33$ coefficients

Figure 1:

Illustration the fertility of our level set representation model in 3D. The three examples show that our model is capable to capture the geometry of the shape even though the shape has complicated topologies, which demonstrated that the representation is not restrictive.

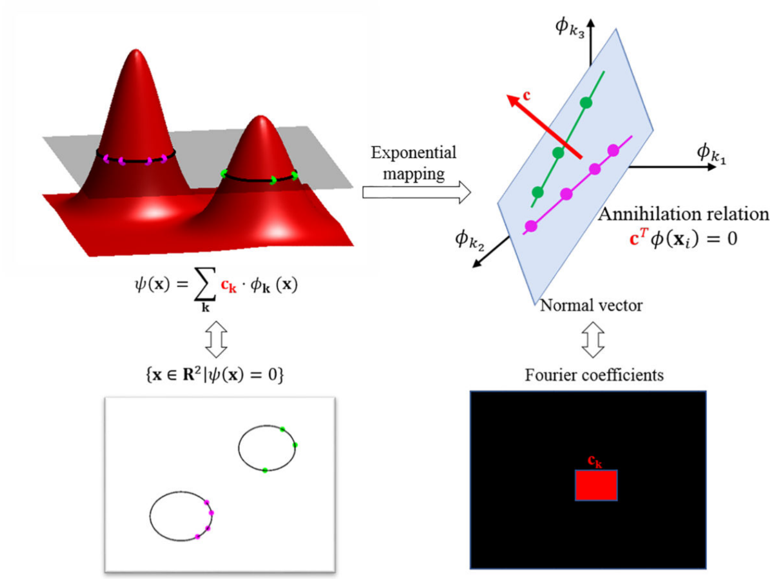

ϕ_{k_3}
$\mathbf{c}$
Exponential mapping
ϕ_{k_1}
Annihilation relation
$\mathbf{c}^T\phi(\mathbf{x}_i) = 0$
ϕ_{k_2}
Normal vector
$\psi(\mathbf{x}) = \sum_{\mathbf{k}} \mathbf{c_k} \cdot \phi_{\mathbf{k}}(\mathbf{x})$
$\{\mathbf{x} \in \mathbf{R}^2 | \psi(\mathbf{x}) = 0\}$
Fourier coefficients
$\mathbf{c_k}$

Figure 2:

Illustration of the annihilation relations (17) in 2D. Assume that the curve is the zero set of a band-limited function $\psi(\mathbf{x})$, shown in the top left. The Fourier coefficients of ψ, denoted by $\mathbf{c}$, are bandlimited in Λ, denoted by the red square in the bottom right. Each point on the curve satisfies $\psi(\mathbf{x}_i) = 0$. Using the representation (3), we have $\mathbf{c}^T\Phi_\Lambda(\mathbf{x}_i) = 0$. This means that the feature map will lift each point in the level set to a $|\Lambda|$ dimensional subspace whose normal vector is specified by $\mathbf{c}$, as illustrated by the plane and the red vector c in the top right. Note that if more than one closed curve are presented, each curve will be lifted to a lower dimensional subspace in the feature space, as shown by the two lines in the plane, and the lower dimensional spaces will span the $|\Lambda|$ dimensional subspace. (Figure courtesy of Q. Zou, reprint from [60] with permission from IEEE).

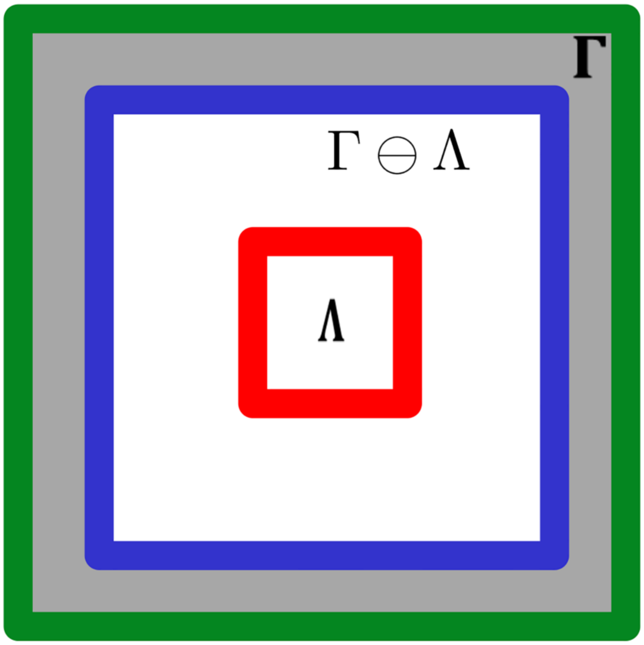


Figure 3:
The non-minimal filter bandwidth Γ (green) is illustrated along with the minimal filter bandwidth Λ (red). The set $\Gamma \ominus \Lambda$ (blue) contains all indices at which Λ can be centered, while remaining inside Γ.

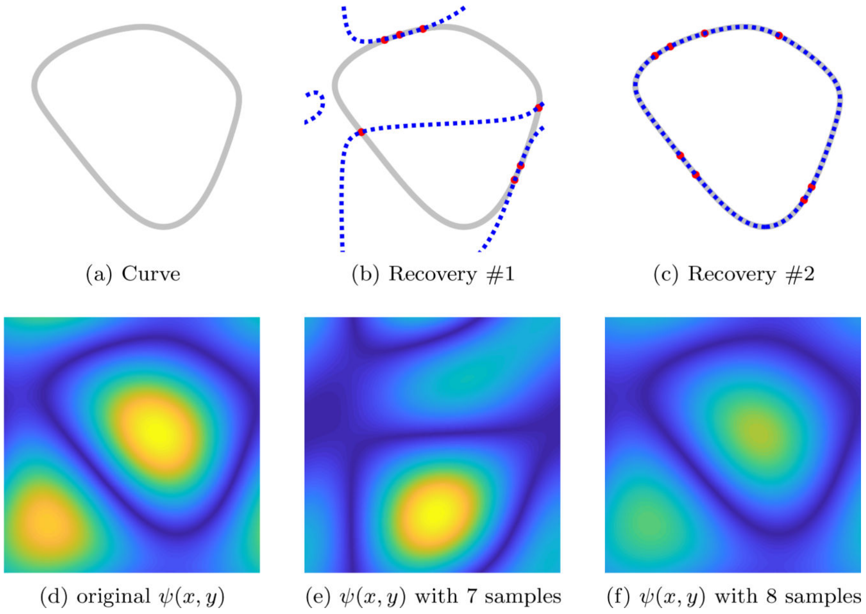


Figure 4:

Illustration of Theorem 5 in 2D. The irreducible curve given by (a) is the original curve, which is obtained by the zero level set of a trigonometric polynomial whose bandwidth is 3 × 3. According to Theorem 5, we will need at least 8 samples to recover the curve. In (b), we randomly choose 7 samples (the red dots) on the original curve (the gray curve). The blue dashed curve shows the recovered curve from this 7 samples. Since the sampling condition is not satisfied, the recovery failed. In (c), we randomly choose 8 points (the red dots). From (c), we see that the blue dashed curve (recovered curve) overlaps the gray curve (the original curve), meaning that we recover the curve perfectly. In (d) - (f), we showed the original trigonometric polynomial, the polynomial obtained from 7 samples and the polynomial obtained from 8 samples.

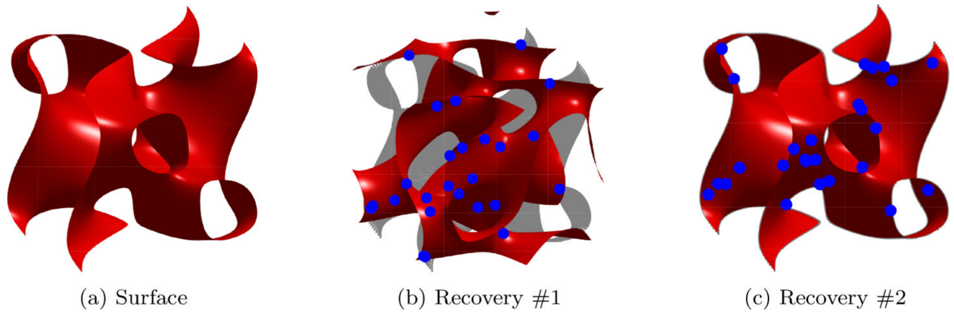
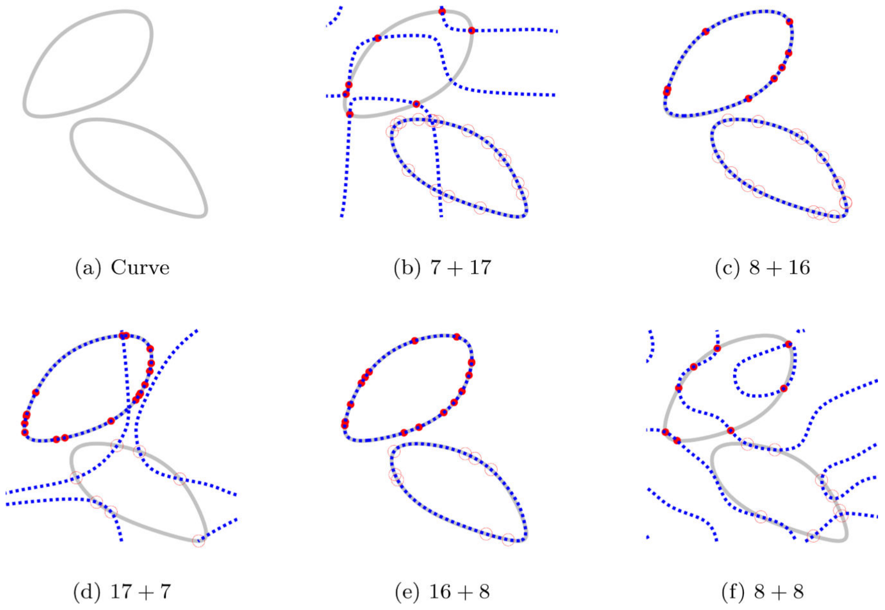


Figure 5:

Illustration of Theorem 5 in 3D. The irreducible surface given by (a) is the original surface, which is given by the zero level set of a trigonometric polynomial whose bandwidth is 3 × 3 × 3. According to Theorem 5, we will need at least 26 samples to recover the surface. In (b), we randomly choose 25 samples (the blue dots) on the original surface (the gray part). The red surface is what we recovered from the 25 samples. Since the sampling condition is not satisfied, the recovery failed. In (c), we randomly choose 26 points (the blue dots). From (c), we see that the red surface (recovered surface) overlaps the gray surface (the original surface), meaning that we recover the surface perfectly.


(a) Curve

(b) $7 + 17$

(c) $8 + 16$

(d) $17 + 7$

(e) $16 + 8$

(f) $8 + 8$

Figure 6:

Illustration of Theorem 7. The original curve (a) is given by the zero set of a reducible trigonometric polynomial with bandwidth 5 × 5, which is the product of two trigonometric polynomials with bandwidth 3 × 3. According to the sampling theorem, we totally need at least 24 samples and each of the components needs to be sampled for at least 8 samples. We first choose 7 samples (red dots) on the first component and 17 samples (red circles) on the second one. The gray curve in (b) is the original curve and the blue dashed curve is what we recovered from the 7+ 17 = 24 samples. Since the sampling condition is not satisfied, the recovery failed. In (c), we choose 8 samples (red dots) on the first component and 16 samples (red circles) on the second one. From (c), we see that the gray curve (the original curve) overlaps the blue dashed curve (recovered curve), meaning that we recovered the curve successfully. In (d), we choose 17 samples on the first component and 7 samples on the other one. From (d), we see that the recovery is not successful. In (e), we have 16 samples on the first component and 8 samples on the second one. The original curve overlaps the recovered one. So we recovered it perfectly. Lastly, we choose 8 samples on each of the component and we failed to recover the curve as shown in (f). Note that the recovered curves pass through the samples in all cases.

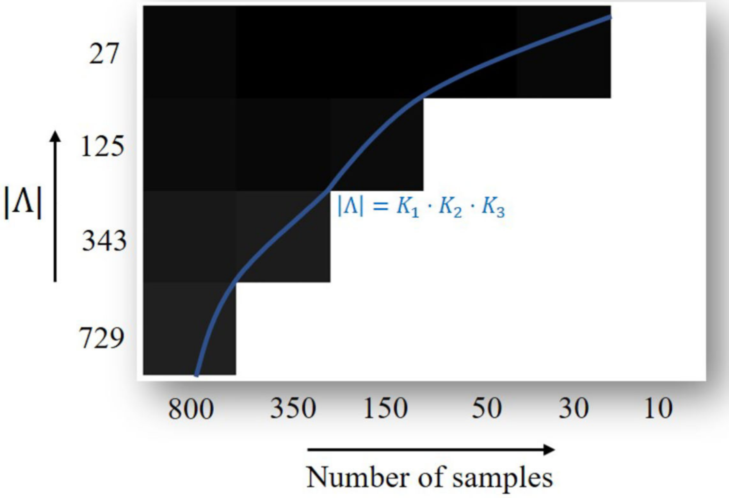
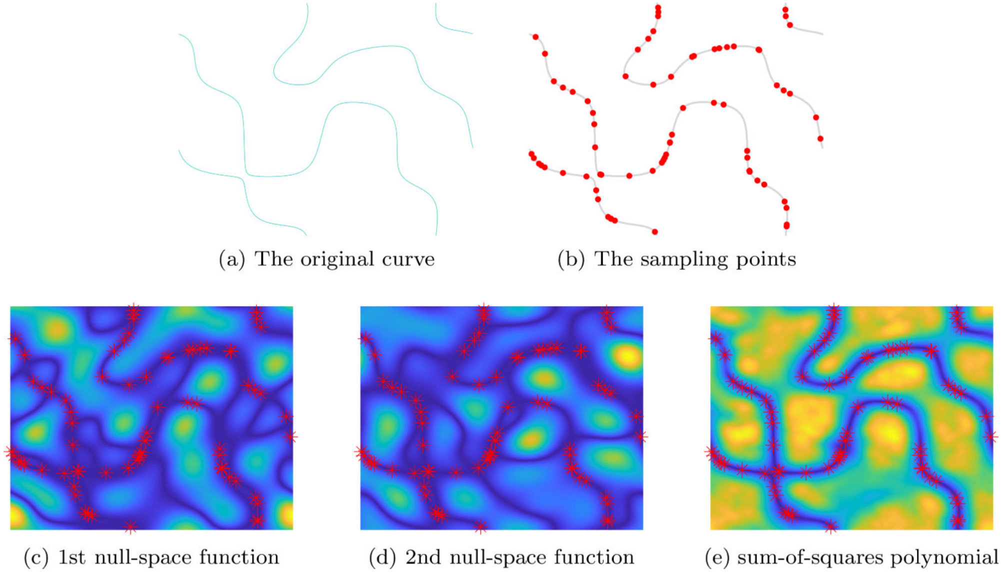


Figure 7:

Effect of number of sampled points on surfaces reconstruction error. We randomly generated several surfaces with different bandwidths and number of sampled points, and tried to recover the surfaces from these samples. The reconstruction errors of the surfaces averaged over several trials are shown in the above phase transition plot, as a function of bandwidth and number of sampled entries. the color black indicates that the true surfaces can be recovered in any of the experiments, while the color white represents that the true surfaces are not recovered in all the experiments. It is seen that we can almost recover the surfaces with $|\Lambda| = k_1 \cdot k_2 \cdot \mathrm{k}_3$ samples.



(a) The original curve (b) The sampling points

(c) 1st null-space function (d) 2nd null-space function (e) sum-of-squares polynomial

Figure 8:

Illustration of the sampling fashion for non-minimal bandwidth. We consider the curve as shown in (a), which is given by the zero level set of a trigonometric polynomial of bandwidth 5 × 5. We choose the non-minimal bandwidth Γ as 11 × 11. According to the sampling condition for non-minimal bandwidth, we sampled on 72 random locations. We randomly chose two null-space vectors for the feature matrix of the sampling set, which gave us functions (c) and (d). We can see that all of these functions have zeros on the original zero set, in addition to processing several other zeros. The sum of squares function is shown in (e), showing the common zeros, which specifies the original curve.

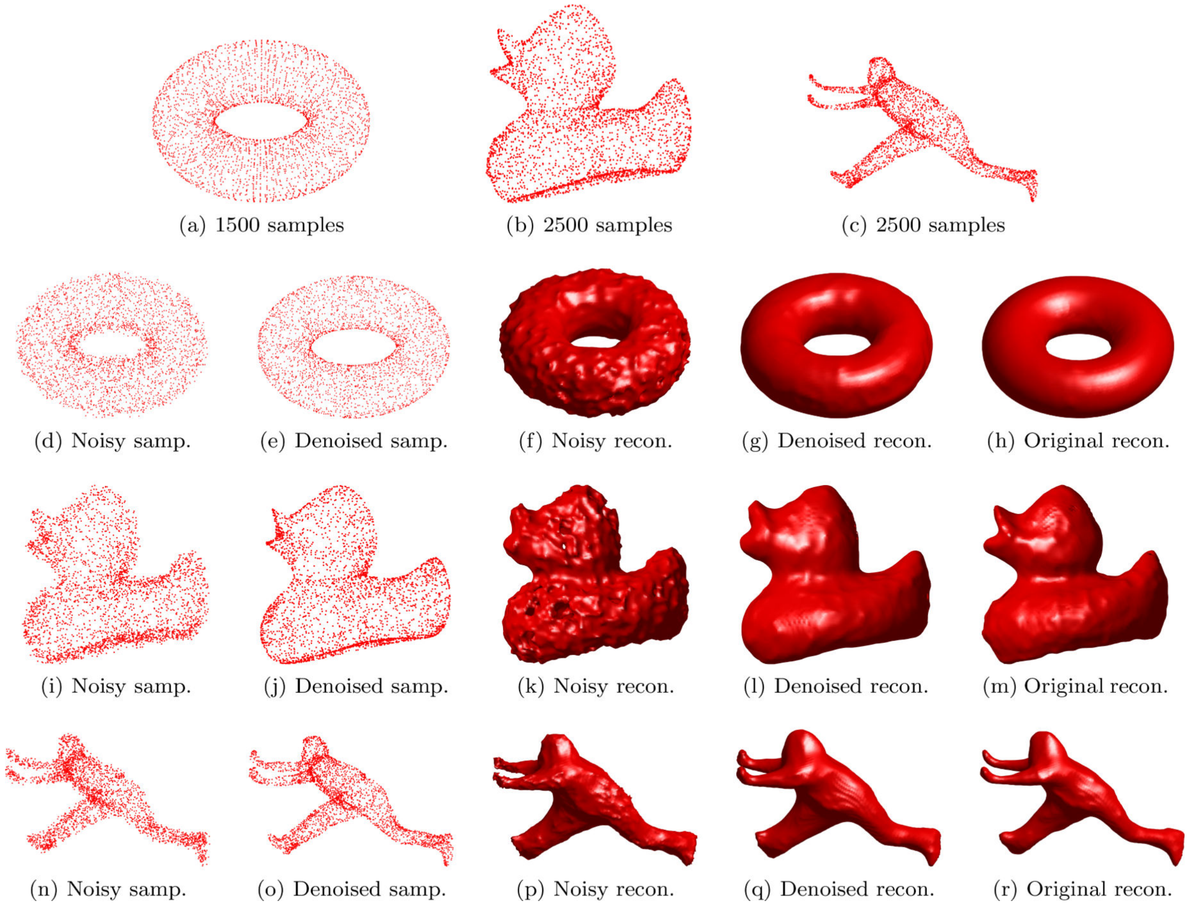
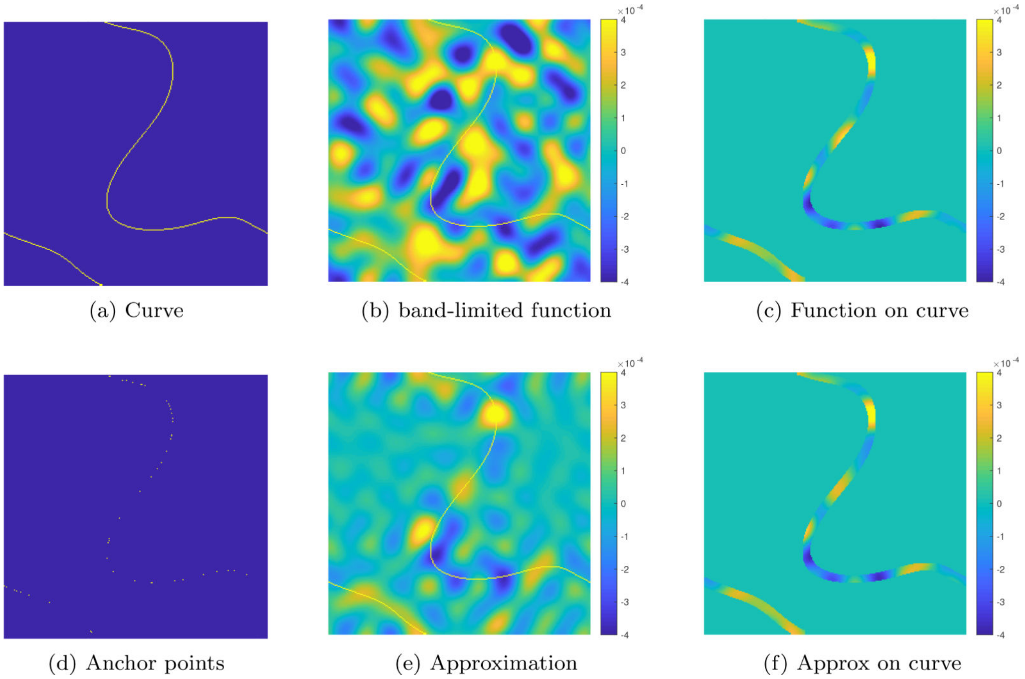


Figure 9:

Illustration of the points cloud denoising algorithm and surface recovery algorithm with unknown bandwidth. The first row shows the samples drawn from three surfaces. Noise is added to the samples (see (d), (i), (n)). Then we use the proposed algorithm to denoise the points. The parameter λ in (31) is chosen as 1.4 for the denoising algorithm. The number of iterations for the denoising algorithm is 30. The surfaces that are recovered from noisy samples and denoised samples are also presented for comparison. The bandwidth was chosen as 31 × 31 × 31 for all the experiments.

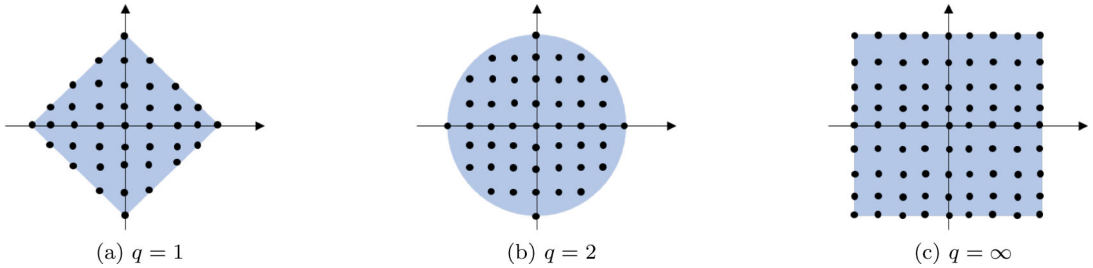


(a) Curve (b) band-limited function (c) Function on curve

(d) Anchor points (e) Approximation (f) Approx on curve

Figure 10:

Illustration of the local representation of functions in 2D. We consider the local approximation of the band-limited function in (b) with a bandwidth of 13 × 13, living on the band-limited curve shown in (a). The bandwidth of the curve is 3 × 3. The curve is overlaid on the function in (b) in yellow. The restriction of the function to the vicinity of the curve is shown in (c). Our results suggest that the local function approximation requires $13^2 - 11^2 = 48$ anchor points. We randomly select the points on the curve, as shown in (d). The interpolation of the function values at these points yields the global function shown in (e). The restriction of the function to the curve in (f) shows that the approximation is good.

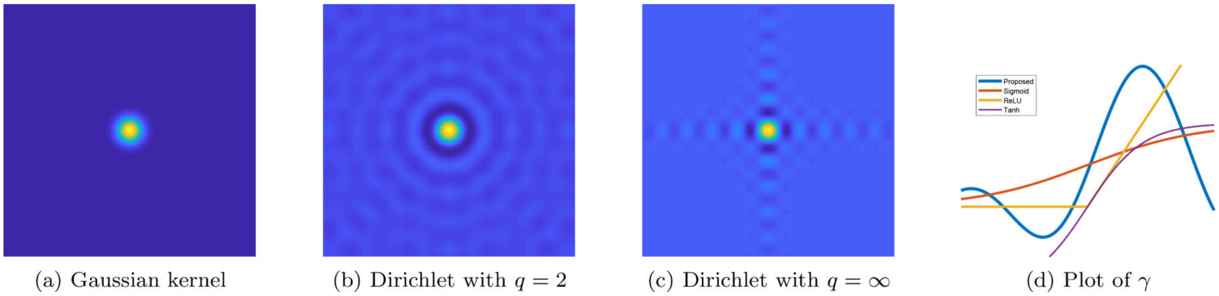

(a) $q = 1$

(b) $q = 2$

(c) $q = \infty$

Figure 11:

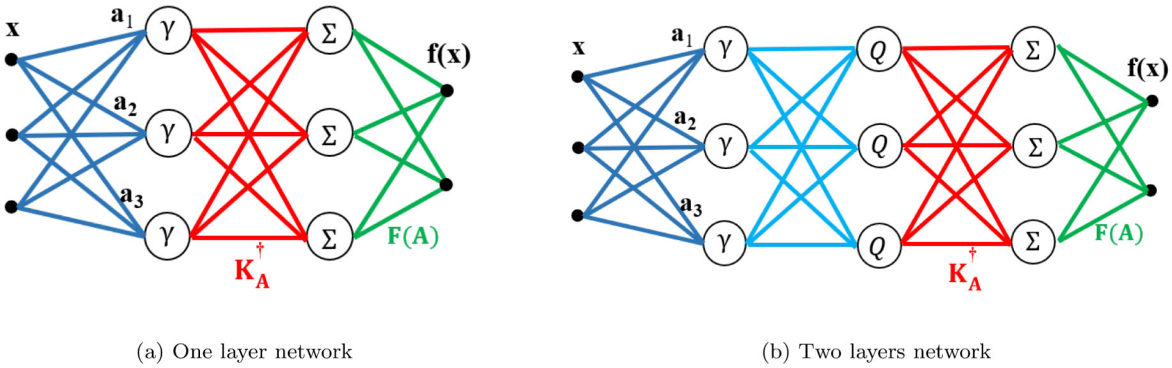
bandwidth of the set Λ with different *q* values.



(a) Gaussian kernel (b) Dirichlet with $q = 2$ (c) Dirichlet with $q = \infty$ (d) Plot of γ

Figure 12:

Visualization of kernels in $\mathbb{R}^2$ and the non-linear function $\boldsymbol{\gamma}$ with some commonly used activation functions.

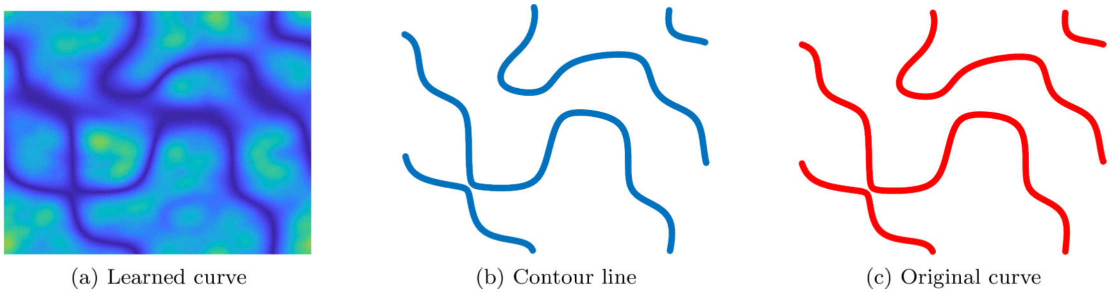


(a) One layer network

(b) Two layers network

Figure 13:

Computational structure of function evaluation. (a) corresponds to (46) to compute the band-limited multidimensional function **f** on $\mathscr{S}[\psi]$. The inner-product between the input vector **x** and the anchor templates on the surface are evaluated, followed by non-linear activation functions $\boldsymbol{\gamma}$ to obtain the coefficients $a_i(\mathbf{x})$. These coefficients are operated with the fully connected linear layers $\mathbf{K}_\mathbf{A}^\dagger$ and $\mathbf{F}(\mathbf{A})$. The fully connected layers can be combined to obtain a single fully connected layer $\widetilde{\mathbf{F}}$. Note that this structure closely mimics a neural network with a single hidden layer. (b) uses an additional quadratic layer, which combines functions of a lower bandwidth to obtain a function of a higher bandwidth.


(a) Learned curve

(b) Contour line

(c) Original curve

Figure 14:

Illustration of the surface learning network using the curve in Fig. 8. (a) and (b) are the learned results. We compared the learned curve (blue curve) with the original curve (red curve) in (c). From which we see that the two curves are almost the same, indicating that the learned network performs well.

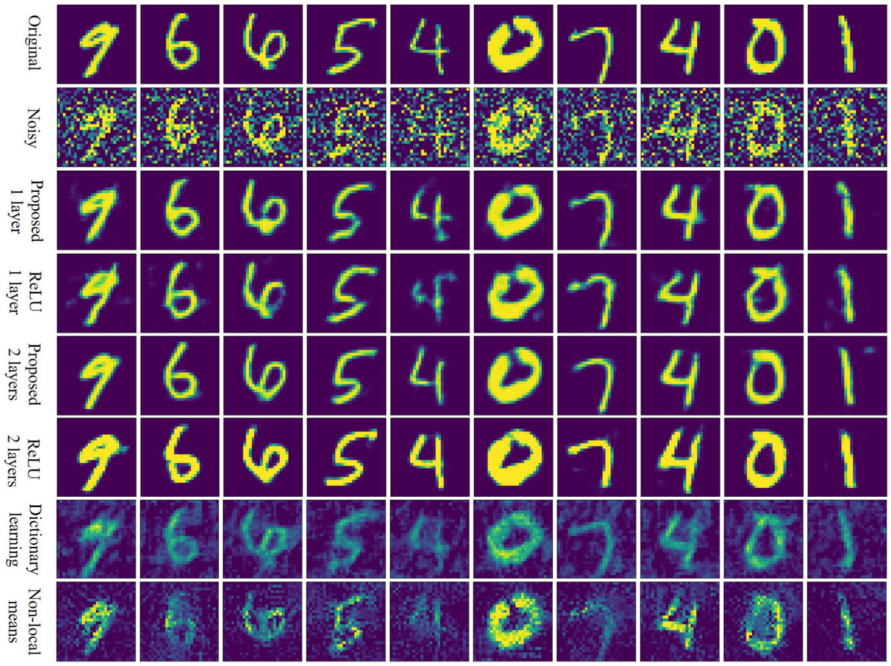


Figure 15:

Comparison of our learned denoiser using the proposed activation function and the ReLU activation function. The testing results show that the denoising performance using the proposed activation function is comparable to the performance using ReLU. The eight rows in the figure correspond to the original images, the noisy images, the denoised images using the proposed one-layer network, the denoised images using one layer ReLU network, the denoised images using the proposed two-layer network, the denoised images using two-layer ReLU network, the denoised images using dictionary learning and the denoised images using non-local means. The averaged PSNR of the denoised images using the proposed one-layer network, one layer ReLU network, proposed two-layer network, two-layer ReLU network, dictionary learning and non-local means are 19.68 dB, 20.03 dB, 20.86 dB, 17.48 dB, 14.76 dB and 14.28 dB respectively. From the quantitative results, we can see that our proposed one-layer network performs comparable to the one-layer ReLU network. For the proposed two-layer network, the performance is getting better from both quantitative and visual points of view. For the two-layer ReLU network, visually the performance is better than that of the one-layer ReLU network. But the PSNR is getting worse. The main reason that causes the low PSNR for the two-layer ReLU network is the change of the pixel values on each hand-written digit.

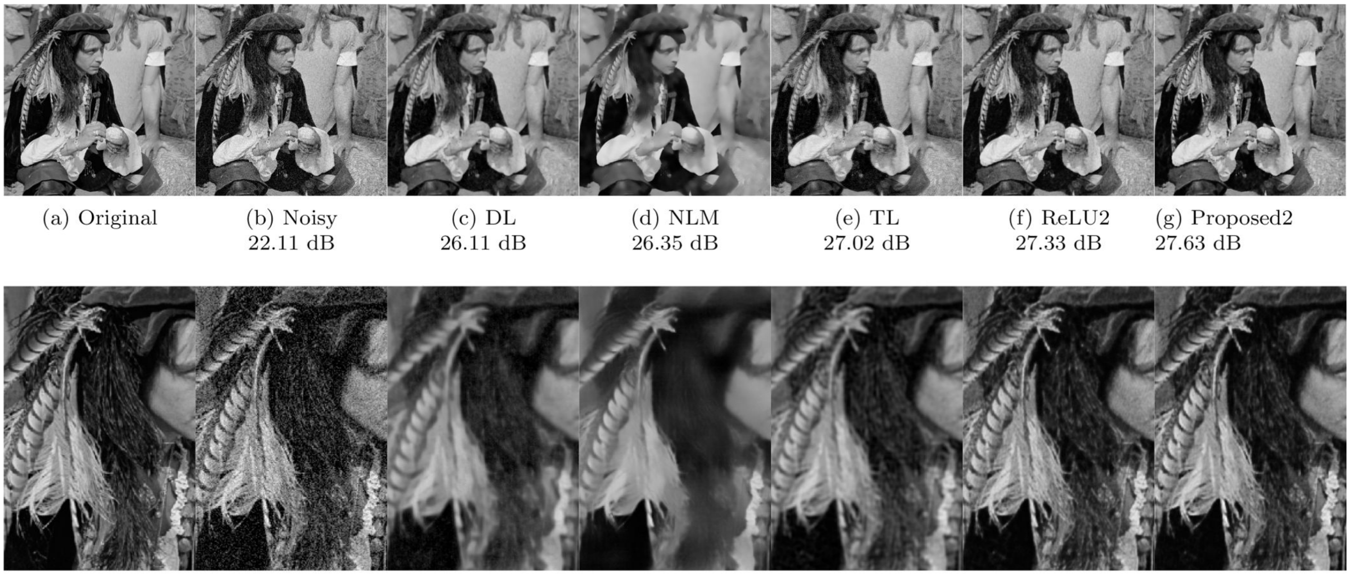
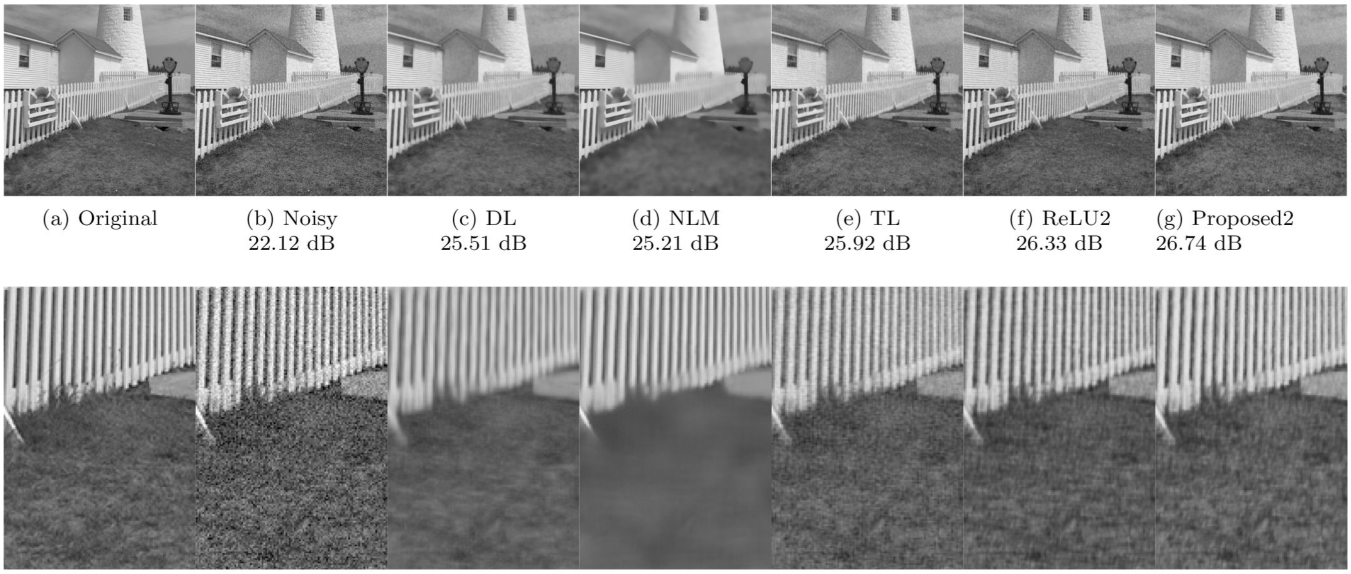


Figure 16:

Comparison of the proposed denoising algorithms on the image \Man” with $\sigma = 20$.



(a) Original

(b) Noisy 22.12 dB

(c) DL 25.51 dB

(d) NLM 25.21 dB

(e) TL 25.92 dB

(f) ReLU2 26.33 dB

(g) Proposed2 26.74 dB

(h) Original

(i) Noisy

(j) DL

(k) NLM

(l) TL

(m) ReLU 2

(n) Proposed 2

Figure 17:

Comparison of the proposed denoising algorithms on the image “Lighthouse” with $\sigma = 20$.

Table 1:

The PSNR (dB) of the denoised results for the two testing natural images with different noise level.

Img.	σ	DL	NLM	TL	ReLU1	ReLU2	Proposed1	Proposed2
Man	10	26.63	26.64	27.41	30.29	31.11	30.99	**31.19**
	20	26.11	26.35	27.02	27.47	27.33	27.25	**27.63**
	100	19.69	20.95	21.65	21.85	**22.11**	21.91	22.06
Lighthouse	10	27.08	29.08	28.71	28.88	29.27	30.05	**30.28**
	20	25.51	25.21	25.92	26.25	26.33	26.69	**26.74**
	100	19.14	20.14	20.15	20.21	20.46	20.35	**20.47**