SIAM J. IMAGING SCIENCES
Vol. 10, No. 2, pp. 711–750

A Tale of Two Bases: Local-Nonlocal Regularization on Image Patches with Convolution Framelets[*]

Rujie Yin[†], Tingran Gao[†], Yue M. Lu[‡], and Ingrid Daubechies[†]

Abstract. We propose an image representation scheme combining the local and nonlocal characterization of patches in an image. Our representation scheme can be shown to be equivalent to a tight frame constructed from convolving local bases (e.g., wavelet frames, discrete cosine transforms, etc.) with nonlocal bases (e.g., spectral basis induced by nonlinear dimension reduction on patches), and we call the resulting frame elements *convolution framelets*. Insight gained from analyzing the proposed representation leads to a novel interpretation of a recent high-performance patch-based image processing algorithm using the point integral method (PIM) and the low dimensional manifold model (LDMM) [S. Osher, Z. Shi, and W. Zhu, *Low Dimensional Manifold Model for Image Processing*, Tech. Rep., CAM report 16-04, UCLA, Los Angeles, CA, 2016]. In particular, we show that LDMM is a weighted ℓ_2-regularization on the coefficients obtained by decomposing images into linear combinations of convolution framelets; based on this understanding, we extend the original LDMM to a reweighted version that yields further improved results. In addition, we establish the energy concentration property of convolution framelet coefficients for the setting where the local basis is constructed from a given nonlocal basis via a linear reconstruction framework; a generalization of this framework to unions of local embeddings can provide a natural setting for interpreting BM3D, one of the state-of-the-art image denoising algorithms.

Key words. image patches, convolution framelets, regularization, nonlocal methods, inpainting

AMS subject classifications. 68U10, 68Q25

DOI. 10.1137/16M1091447

1. Introduction.

In past decades, patch-based techniques such as nonlocal means (NLM) and block-matching with 3D collaborative filtering (BM3D) have been successfully applied to image denoising and other image processing tasks [7, 8, 16, 11, 64, 30]. These methods can be viewed as instances of graph-based adaptive filtering, with similarity between pixels determined not solely by their pixel values or spatial adjacency, but also by the (weighted) ℓ^2-distance between their neighborhoods, or *patches* containing them. The effectiveness of patch-based algorithms can be understood from several different angles. On the one hand, patches from an image often enjoy sparse representations with respect to certain redundant families of vectors, or unions of bases, which motivated several dictionary- and sparsity-based

[*]Received by the editors August 29, 2016; accepted for publication (in revised form) January 31, 2017; published electronically May 16, 2017.
http://www.siam.org/journals/siims/10-2/M109144.html
Funding: The work of the first author was supported in part by the NSF under grant 1516988. The work of the third author was supported in part by the NSF under grant CCF-1319140 and by ARO under grant W911NF-16-1-0265.
[†]Department of Mathematics, Duke University, Durham, NC 27708 (rujie.yin@duke.edu, trgao10@math.duke.edu, ingrid@math.duke.edu).
[‡]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 (yuelu@seas.harvard.edu).

approaches [21, 10, 41]; on the other hand, the nonlocal characteristics of patch-based methods can be used to build highly data-adaptive representations, accounting for nonlinear and self-similar structures in the space of image patches [27, 34]. Combined with adaptive thresholding, these constructions have connections to classical wavelet-based and total variation algorithms [63, 48]. Additionally, the patch representation of signals has specific structures that can be exploited in regularization; for example, the inpainting algorithm ALOHA [29] utilized the low-rank block Hankel structure of certain matrix representations of image patches.

Among the many theoretical frameworks built to understand these patch-based algorithms, manifold models have recently drawn increased attention and have provided valuable insights in the design of novel image processing algorithms. Along with the development of manifold learning algorithms and topological data analysis, it is hypothesized that high-contrast patches are likely to concentrate in clusters and along low dimensional nonlinear manifolds; this phenomenon is very clear for cartoon images; see, e.g., [48, 49]. This intuition was made precise in [35] and was followed by more specific Klein bottle models [9, 47] on both cartoon and texture images. Adopting a point of view from diffusion geometry, [59] interprets the nonlocal mean filter as a diffusion process on the "patch manifold," relating denoising iterations to the spectral properties of the infinitesimal generator of that diffusion process; similar diffusion-geometric intuitions can also be found in [63, 50], which combined patch-based methods with manifold learning algorithms.

Recently, a new method called the low dimensional manifold model (LDMM) was proposed in [45], with strong results. LDMM is a direct regularization on the dimension of the patch manifold in a variational argument for patch-based image inpainting and denoising. The novelty of [45] includes (1) an identity relating the dimension of a manifold with L^2-integrals of ambient coordinate functions and (2) a new graph operator (which we study below) on the nonlocal patch graph obtained via the point integral method (PIM) [38, 58, 57]. The current paper is motivated by our wish to better understand the embedding of image patches in general and the LDMM construction in particular.

Typically, given an original signal $f \in \mathbb{R}^N$, patch-based methods start with explicitly building for f a redundant representation consisting of *patches* of f. The patches either start with or are centered at each pixel[1] in the domain of f and are of constant length ℓ for $1 < \ell < N$. Reshaped into row vectors stacked vertically in the natural order, these patches constitute a Hankel[2] matrix $F \in \mathbb{R}^{N\times\ell}$, which we refer to as a *patch matrix* (see Figure 1 in section 2 below). It is the patch matrix F, rather than the signal f itself, that constitutes the object of main interest in nonlocal image processing and in particular LDMM; each single pixel of image f is represented in F exactly ℓ times, a redundancy that is often beneficially exploited in signal processing tasks. (As will be made clear in Proposition 1,

[1]Possibly with a *stride* larger than 1 in many applications. In this paper though, we assume that the stride is always equal to 1 to demonstrate the key ideas. A periodic boundary condition is assumed throughout this paper.

[2]For simplicity of notation, we explain the mechanism of our approach for 1D rather than 2D signals. As shown explicitly in Appendix A, there is no real difficulty in going from 1D to 2D; the (nonessential) difference is that the patch matrix does not have the exact Hankel structure as in the 1D case due to the rearrangement of 2D patches into 1D row vectors. Nevertheless, the conversion from patch matrix vector multiplication to signal filter convolution still holds.

representing $f \in \mathbb{R}^N$ as $F \in \mathbb{R}^{N\times\ell}$ incurs an "ℓ-fold redundancy" in the sense of frame bounds.) For comparison, earlier image processing models based on total variation [53] or nonlocal regularization [23, 24] focus on regularizing the signal f directly, whereas more recent state-of-the-art image inpainting techniques such as LDMM and low-rank Hankel matrix completion [29] build upon variational frameworks for the patch matrix F and do not convert F back to f until the optimization step terminates. To the best of our knowledge, the mechanism of these regularization strategies on patch matrices has not been fully investigated.

From an approximation point of view, the patch matrix $F \in \mathbb{R}^{N\times\ell}$ has more flexibility than the original signal $f \in \mathbb{R}^N$ since one can search for efficient representation of the matrix F in either its row space or its column space. The idea of learning sparse and redundant representations for rows of F, or the patches of f, has been pursued in many papers (see, e.g., [44, 22, 32, 31, 37, 2, 21] and the references therein); this amounts to learning a redundant dictionary $\mathcal{D} \in \mathbb{R}^{\ell\times m}$, $m \geq \ell$, such that $F = A\mathcal{D}^\top$ where the rows of $A \in \mathbb{R}^{N\times m}$ are sparse. Meanwhile, each column of F can be viewed as a "coordinate function" (adopting the geometric intuition in [45]) defined on the dataset of patches and can thus be efficiently encoded using spectral bases adapted to this dataset: for example, let $\Phi : \mathbb{R} \to [0, +\infty)$ be a nonnegative smooth *kernel function* with exponential decay at infinity, and construct the following positive semidefinite kernel matrix for the dataset of patches of f:

$$\Phi_\epsilon\left(ij\right) = \Phi\left(\frac{\|F_i - F_j\|_2^2}{\epsilon}\right), \qquad 0 \leq i,\, j \leq N-1,$$

where F_i, F_j are the ith and jth rows of the patch matrix F, respectively, and $\epsilon > 0$ is a bandwidth parameter representing our confidence in the similarity between patches of f (e.g., how small L^2-distances should be to reflect the geometric similarity between patches; this is influenced, for example, by the noise level in image denoising tasks). By Mercer's theorem, Φ_ϵ admits an eigendecomposition

$$\Phi_\epsilon = \sum_{k=1}^{N} \lambda_k \phi_k \phi_k^\top,$$

where for each $1 \leq k \leq N$ the column vector $\phi_k \in \mathbb{R}^N$ is the eigenvector associated with nonnegative real eigenvalue $\lambda_k \in \mathbb{R}$. These eigenvectors constitute a basis for $\mathbb{R}^N$, with respect to which each column of the patch matrix F can be expanded as a linear combination. Though such expansions are not sparse in general, they are highly data-adaptive and result in efficient approximations when the eigenvalues have fast decay; see [34, 1] for theoretical bounds of the approximation error, [48] for empirical evidence, and [63] for applications in semisupervised learning and image denoising. By construction, the sparse representation for the rows of F relies heavily on the *local* properties of the signal f, whereas the spectral expansion for the columns of F captures more *nonlocal* information in f. We remark here that many other orthonormal or overcomplete systems can be used to produce different representations for the row and column spaces of the patch matrix F: for instance, wavelets or discrete cosine transform can be used in place of a dictionary $\mathcal{D}$, while any linear/nonlinear embedding method, dimension reduction algorithm (e.g., principal component analysis (PCA) [46], multidimensional scaling (MDS) [66, 55], Autoencoder [26], t-SNE [40]) or reproducing kernel Hilbert space technique [54] can work just as well as the kernel Φ; nevertheless, the different choices for the row (resp., column) space of F primarily read off local (resp., nonlocal) information of f. These observations motivate us to seek new representations for the patch matrix F that could reflect both local and nonlocal behavior of the signal f. This methodology is already implicit in BM3D [16], one of the state-of-the-art image denoising algorithms (see subsection 3.4 for details); we point out in this paper that such a paradigm is much more universal and can be used for a wide range of patch-based image processing tasks; we propose a regularization scheme for a signal f based on its coefficients with respect to *convolution framelets* (to be defined in section 4), a type of signal-adaptive tight frames generated from the adaptive representation of the patch matrix F.

As a first attempt at understanding the theoretical guarantees of convolution framelets, we consider the problem of determining an "optimal" local basis, in the sense of minimum linear reconstruction error, with respect to a fixed nonlocal basis (interpreted as embedding coordinate functions of the patches); convolution framelets constructed from such an "optimal" pair of local and nonlocal bases are guaranteed to have an "energy compaction property" that can be exploited to design regularization techniques in image processing. In particular, we show that when the nonlocal basis comes from MDS, right singular vectors[3] of the patch matrix F constitute the corresponding optimal local basis. The linear reconstruction framework itself—of which LDMM can be viewed as an instantiation—is general and uses variational functionals associated with nonlinear embeddings, via a linearization. This insight allows us to generalize LDMM by reformulating the manifold dimension minimization in [45] as an equivalent weighted ℓ_2-minimization on coefficients of such a convolution frame and by using more adaptive weights; for some types of images this proposed scheme leads to markedly improved results. Finally, we note that our framework is widely applicable and can be adapted to different settings, including BM3D [16] (in which case the framework needs to be extended to describe unions of local embeddings, as is done in subsection 3.4 below).

The rest of the paper is organized as follows. In section 2 we present convolution framelets as a data-adaptive redundant representation combining local and nonlocal bases for signal patch matrices. Section 3 motivates the energy compaction property of convolution framelets and establishes a guarantee for energy concentration through a linear reconstruction procedure related to (nonlinear) dimension reduction [51]. Section 4 interprets LDMM as an ℓ^2-regularization on the energy concentration of convolution framelet coefficients. This novel interpretation and insights gained from the previous section lead to improvement of LDMM by incorporating more adaptive weights in the regularization. We compare LDMM with our proposed improvement in section 5 by numerical experiments on various image processing tasks, such as inpainting and denoising. Section 6 summarizes and suggests future work. A list of notation used throughout the paper is shown in Table 1.

[3]Since the singular value decomposition (SVD) of a patch matrix is not known a priori in image reconstruction tasks, the algorithms we propose in this paper are all of an iterative nature, with the SVD basis updated in each iteration; similar strategies have previously been utilized in nonlocal image processing algorithms; see, e.g., [23, 24].

Table 1
Notations used throughout this paper.

I_k	identity matrix in $R^{k\times k}$
$\mathbb{1}_k$	vector with all one entries in R^k
$\\|\cdot\\|_{\mathrm{F}}$	matrix Frobenius norm
ℓ	dimension of *ambient space*, i.e., number of pixels in a patch
p	dimension of *embedding space*, i.e., number of coordinate functions in the nonlocal embedding
N	number of data points (patches)
X	data matrix in $\mathbb{R}^{N\times\ell}$
$\widetilde{X}$	embedded data matrix in $\mathbb{R}^{N\times p}$, "orthogonalized" s.t. $\widetilde{X} = \Phi_{\mathcal{E}}\, C_{\mathcal{E}}$
X^i (resp., $\widetilde{X}^i$)	the ith coordinate in ambient space (resp., embedding space)
$\Phi_{\mathcal{E}}$	normalized graph bases in $\mathbb{R}^{N\times p}$ from $\mathcal{E}$ and $\Phi_{\mathcal{E}}^\top\, \Phi_{\mathcal{E}} = I_p$
Φ	full orthonormal nonlocal bases extended from $\Phi_{\mathcal{E}}$
$C_{\mathcal{E}}$	diagonal matrix with entries $\\|\widetilde{X}^i\\|$
x_i	the ith row of X, i.e., the ith data point
$\tilde{x}_i$	embedding of x_i
$\mathcal{E}$	embedding function from $\mathbb{R}^\ell$ to $\mathbb{R}^p$
$\mathcal{E}_x$	affine approximation of $\mathcal{E}$ at point x
V_0	patch bases in $\mathbb{R}^{\ell\times p}$, s.t. $V_0^\top V_0 = I_p$
V	full orthonormal local bases in $O(\ell)$
C	coefficient matrix
f	1D or 2D signal, e.g., an image
F	patch matrix in $\mathbb{R}^{N\times\ell}$ generated from f, a special type of data matrix
F_i	ith patch
F^i	ith coordinate in patch space, e.g., ith pixel in all patches
$\widetilde{F}$	embedded patch matrix
Ψ	bases in $\mathbb{R}^{N\times\ell}$ from $\phi_i\, v_j^\top$
ψ	bases in $\mathbb{R}^N$ from $\phi_i * v_j(-\cdot)$
W	affinity matrix of diffusion graph with Gaussian kernel
D	degree matrix from W
L	normalized graph diffusion Laplacian
R_L	graph operator in LDMM

2. Convolution framelets.

Consider a one-dimensional (1D)[4] real-valued signal

$$f = (f[0], \dots, f[N-1])^\top \in \mathbb{R}^N$$

sampled at N points. We fix the *patch size* ℓ as an integer between 1 and N and assume a periodic boundary condition for f. For any integer $m \in [0, \dots, N-1]$, we refer to the row vector $F_m = (f[m], \dots, f[m+\ell-1]) \in \mathbb{R}^\ell$ as the *patch* of f at m of length ℓ. Construct the *patch matrix* of f, denoted as $F \in \mathbb{R}^{N\times\ell}$, by vertically stacking the patches according to their order of appearance in the original signal:

$$F = [F_0^\top, \dots, F_{N-1}^\top]^\top. \tag{1}$$

[4]The restriction to 1D signals is for notational simplicity only: the same idea can be easily generalized to signals of higher dimensions; see, e.g., Appendix A for a quick derivation for the 2D case.

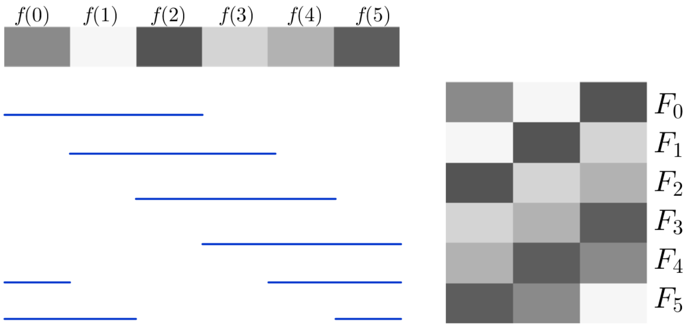

Figure 1. *Illustration of a patch matrix constructed from a 1D signal (the blue lines indicate locations of the patches F_0 through F_5 in the original 1D signal).*

See Figure 1 for an illustration. It is clear that F is a Hankel matrix, and thus f can be reconstructed from F by averaging the entries of F "along the antidiagonals", i.e.,[5]

$$f(n) = \frac{1}{\ell} \sum_{i=1}^{\ell} F_{n-i+1,\, i}, \qquad n = 0, 1, \ldots, N-1. \tag{2}$$

For simplicity, we introduce the following notations that are standard in signal processing:

- The *(circular) convolution* of two vectors v, $w \in \mathbb{R}^N$ is defined as

$$(v * w)\,[n] = \sum_{m=0}^{N-1} v[n-m]\, w[m],$$

  where periodic boundary conditions are assumed (as is done throughout this paper).
- For any $v \in \mathbb{R}^{N_1}$ and $w \in \mathbb{R}^{N_2}$ with N_1, $N_2 \leq N$, define their *convolution* in $\mathbb{R}^N$ as

$$v * w = v^0 * w^0,$$

  where $v^0 = [v^\top, \mathbf{0}_{N-N_1}^\top]^\top$, $w^0 = [w^\top, \mathbf{0}_{N-N_2}^\top]^\top$ denote the length-N zero-padded versions of v and w, respectively.
- For any $v \in \mathbb{R}^{N_1}$ with $N_1 \leq N$, define the *flip* $v(-\cdot)$ of v as $v(-\cdot)\,[n] = v^0\,[-n]$.

Using these notations, the matrix-vector product of F with any $v \in \mathbb{R}^\ell$ can be written in convolution form as

$$Fv = f * v(-\cdot). \tag{3}$$

[5]Note that in (2) the row indices start at 0, but the column indices start at 1; for instance, the entry at the upper left corner of F is denoted as F_{01}.

Furthermore, it is straightforward to check for any $w \in \mathbb{R}^\ell$ and $v, s \in \mathbb{R}^N$ that

$$s^\top (v * w) = \sum_{m=0}^{N-1} s[m] \sum_{n=0}^{N-1} v[n]\, w[m-n] = \sum_{n=0}^{N-1} v[n] \sum_{m=0}^{N-1} s[m]\, w[m-n] \\ = \sum_{n=0}^{N-1} v[n] \sum_{m'=0}^{N-1} s[n+m']\, w[m'] = v^\top (s * w(-\cdot)). \tag{4}$$

Now let $\Phi \in O(N)$ and $V \in O(\ell)$ be orthogonal matrices of dimension $N \times N$ and $\ell \times \ell$, respectively; also denote the columns of Φ, V as ϕ_i, v_j correspondingly, where $1 \le i \le N$, $1 \le j \le \ell$. The outer products of the columns of Φ with the columns of V, denoted as

$$\left\{ \Psi_{ij} = \phi_i\, v_j^\top \,\middle|\, i = 1, \dots, N,\ j = 1, \dots, \ell \right\},$$

form an orthonormal basis for the space $\mathbb{R}^{N\times\ell}$ of all $N \times \ell$ matrices equipped with inner product $\langle A, B\rangle = tr\left(AB^\top\right)$. The patch matrix F can thus be written in this orthonormal basis as

$$F = \sum_{i=1}^{N}\sum_{j=1}^{\ell} tr\left(F\Psi_{ij}^\top\right)\Psi_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{\ell} C_{ij}\Psi_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{\ell} C_{ij}\phi_i v_j^\top,$$

where

$$C_{ij} := tr\left(F\Psi_{ij}^\top\right) = tr\left(F v_j \phi_i^\top\right) = \phi_i^\top F\, v_j = \phi_i^\top (f * v_j(-\cdot)) = f^\top (\phi_i * v_j),$$

and the last two equalities are due to the identities (3) and (4) given above. In other words, we have the following linear decomposition for F:

$$F = \sum_{i=1}^{N}\sum_{j=1}^{\ell} \langle f, \phi_i * v_j\rangle\, \phi_i v_j^\top. \tag{5}$$

Combining (2) and (5) leads to a decomposition of the original signal f as

$$f = \frac{1}{\ell}\sum_{i=1}^{N}\sum_{j=1}^{\ell} \langle f, \phi_i * v_j\rangle\, \phi_i * v_j, \tag{6}$$

where the convolution $\phi_i * v_j$ stems from averaging the entries of $\phi_i v_j^\top$ along the antidiagonals [cf. (2)]. Define *convolution framelets*

$$\psi_{ij} = \frac{1}{\sqrt{\ell}}\phi_i * v_j, \qquad i = 1, \dots, N, \quad j = 1, \dots, \ell; \tag{7}$$

then (6) indicates that $\{\psi_{ij} \mid 1 \le i \le N, 1 \le j \le \ell\}$ constitutes a tight frame for functions defined on $\mathbb{R}^N$. In fact, we have the following more general observation which can be derived directly from standard frame theory.

Proposition 1. *Let $V^L \in \mathbb{R}^{n\times n'}$, $V^S \in \mathbb{R}^{m\times m'}$ be such that $V^L (V^L)^\top = I_n$, $V^S (V^S)^\top = I_m$ with $m \le n$. Then $v_i^L * v_j^S$, $i = 1, \dots, n'$, $j = 1, \dots, m'$ form a tight frame for $\mathbb{R}^n$ with frame constant m.*

The proof of Proposition 1 can be found in Appendix B.

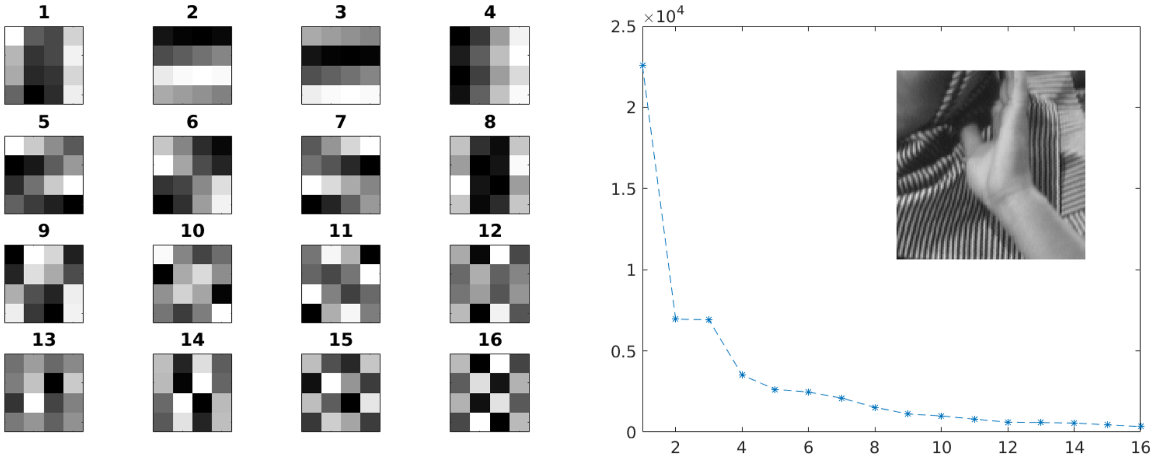

Figure 2. *Left: Right singular vectors (ordered in decreasing singular values) of the patch matrix of a cropped* BARBARA *image of size* 128×128*, with patch size* 4×4*. Right: The cropped* BARBARA *image and the singular values corresponding to the right singular vectors shown on the left. Notice the fast decay of the singular values.*

3. Approximation of functions with convolution framelets.

The construction in section 2 may seem unintuitive at a first glance. Our motivation for introducing two different bases, $\Phi \in O(N)$ and $V \in O(\ell)$, is simply to take advantage of the representability of patch matrices jointly in its row and column spaces.

3.1. Local and nonlocal approximations of a signal.

The columns of V form an orthonormal basis for $\mathbb{R}^\ell$, with respect to which the rows of F, or equivalently the length-ℓ patches of f, can be expanded; the role of V is thus similar to transforms on a localized time window, such as the *short-time Fourier transform* (STFT) or *windowed Wigner distribution function* (WWDF). For this reason, we refer to the strategy of approximating the rows of F using the columns of V as *local approximation*, and we call V a *local basis* in the construction of convolution framelets. The local basis V can be chosen as either fixed functions, e.g., Fourier or wavelet basis, or data-dependent functions, such as the right singular vectors of F. See Figure 2 for an example.

The columns of Φ, on the other hand, are treated as a basis for the columns of F. When the patch stride is set to 1, each column F is just a shifted copy of the original signal f (see Figure 1); more generally (including arbitrary patch strides), columns of F can be seen as functions defined on the set of patches $\mathscr{F} = \{F_0, \ldots, F_{N-1}\}$. When $\mathscr{F}$ is viewed as a discrete point cloud in $\mathbb{R}^\ell$, efficient representations of functions on $\mathscr{F}$ depend more on the Euclidean proximity between patches as points in $\mathbb{R}^\ell$, rather than spatial adjacency in the original signal domain, as detailed in previous work on *spectral basis* [48, 34, 27]. Therefore, it is natural to refer to the paradigm of approximating the columns of F using Φ as *nonlocal approximation*, and call Φ a *nonlocal basis* in the construction of convolution framelets.

Viewing the patch matrix F as a collection $\mathscr{F} \subset \mathbb{R}^\ell$ brings in a large class of nonlinear approximation techniques from *dimension reduction*, a field of statistics and data science dedicated to efficient data representations. Given a data matrix $X = [x_1, \ldots, x_N]^\top \in \mathbb{R}^{N \times \ell}$

consisting of N data points in an ambient space $\mathbb{R}^\ell$ (we adopt the convention that x_i's are column vectors and the ith row of X is $x_i^\top$), dimension reduction algorithms map the full data matrix X to $\widetilde{X} = [\tilde{x}_1, \ldots, \tilde{x}_N]^\top \in \mathbb{R}^{N\times p}$, where each row $\tilde{x}_i \in \mathbb{R}^p$ ($p \leq \ell$) is the image of x_i. The dissimilarity between two original data points is assumed to be given by a metric (distance) function $d(\cdot, \cdot)$ on the ambient space $\mathbb{R}^\ell$, in many applications different from the canonical Euclidean distance; one hopes that the embedding is "almost isometric" between metric spaces $\left(\mathbb{R}^\ell, d\right)$ and $\mathbb{R}^p$ equipped with the standard Euclidean distance. More precisely, let $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_p) : \mathbb{R}^\ell \to \mathbb{R}^p$ be the embedding given by p coordinate functions, and denote $\mathcal{E}(x) = [\tilde{x}^1, \ldots, \tilde{x}^p]^\top \in \mathbb{R}^p$ for any $x \in \mathbb{R}^\ell$. The embedding $\mathcal{E}$ is said to be *near-isometric* if in an appropriate sense

(P1) $d(x, x') \approx \|\mathcal{E}(x) - \mathcal{E}(x')\|_{\ell_2}$ for all x, $x' \in \mathbb{R}^\ell$.

Without loss of generality, we can assume that the coordinate functions of the embedding $\mathcal{E}$ are orthogonal on the data set $\{x_1, \ldots, x_N\}$, i.e.,

(P2) $(\widetilde{X}^s)^\top \widetilde{X}^t = 0$ for all $1 \leq s \neq t \leq p$,

where $\widetilde{X}^i$ is the ith column of $\widetilde{X}$ (and corresponds to the ith coordinate in the embedding space); for general $\widetilde{X}$ with coordinate functions nonorthogonal on the data set, we define its *orthogonal normalization* by $\widetilde{X}^O = \widetilde{X}\, O_{\widetilde{X}}$, where $O_{\widetilde{X}}$ comes from the *singular value decomposition* (SVD) of $\widetilde{X} = U_{\widetilde{X}} \Sigma_{\widetilde{X}} O_{\widetilde{X}}^\top$. Note that classical linear and nonlinear dimension reduction techniques, such as *principal component analysis* (PCA), *multidimensional scaling* (MDS), *Laplacian eigenmaps* [5], and *diffusion maps* [13], all produce embedding coordinate functions satisfying (P1) and (P2).

A standard approach in manifold learning and spectral graph theory to building basis functions on $\mathscr{F}$ is through the eigendecomposition of *graph Laplacians* for a weighted graph constructed from $\mathscr{F}$. For instance, in *diffusion geometry* [13, 14, 15], one considers the *graph random walk Laplacian* $I - D^{-1}W$, where $W \in \mathbb{R}^{N\times N}$ is the weighted adjacency matrix defined by

$$W_{ij} = \exp\left(-\|F_i - F_j\|^2/\epsilon\right) \tag{8}$$

with the bandwidth parameter $\epsilon > 0$, and $D \in \mathbb{R}^{N\times N}$ is the diagonal degree matrix with entries $D_{ii} = \sum_j W_{ij}$ for all $i = 1, \ldots, N$. If the points in $\mathscr{F}$ are sampled uniformly from a submanifold of $\mathbb{R}^\ell$, eigenvectors of $I - D^{-1}W$ converge to eigenfunctions of the Laplace–Beltrami operator on the smooth submanifold as $\epsilon \to 0$ and the number of samples tends to infinity [6, 60]. Up to a similarity transform, the random walk graph Laplacian is equivalent to the symmetric *normalized graph diffusion Laplacian*[6]

$$L = D^{1/2}(I - D^{-1}W)D^{-1/2} = I - D^{-1/2}WD^{-1/2}. \tag{9}$$

Let $L = \Phi\Lambda\Phi^\top$ be the eigendecomposition of L, where $\Phi \in O\,(N)$ and Λ is a diagonal matrix with all diagonal entries between 0 and 1. As in diffusion maps [13], the columns of $\Phi\Lambda^{1/2}$ can

[6]Note that L is different from the *normalized graph Laplacian*, which in standard spectral graph theory is constructed from an adjacency matrix with 0 or 1 in its entries, instead of the weighted adjacency matrix W in (8). One crucial difference is in the range of eigenvalues: the normalized graph Laplacian has eigenvalues in $[0, 2]$, whereas L has eigenvalues in $[0, 1]$ (see [59] or [33, section 2.2.2].)

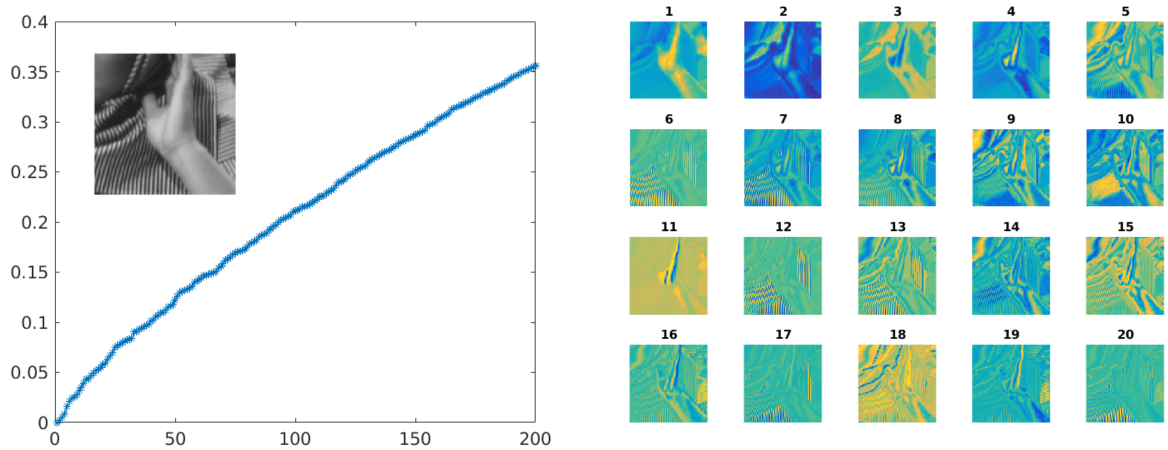

Figure 3. *Left: The same cropped* BARBARA *image of size* 128×128 *as in Figure 2, and the smallest* 200 *eigenvalues of the symmetric normalized graph diffusion Laplacian on the collection of all patches of size* $\ell = 4 \times 4$*. Right: Eigenvectors associated with the smallest* 20 *nonzero eigenvalues on the left.*

be used as coordinate functions for a spectral embedding of the patch collection into $\mathbb{R}^N$. This embedding introduces the *diffusion distance* $d(\cdot,\cdot)$ between patches F_i, F_j $(0 \le i, j \le N-1)$ by setting $d(F_i, F_j)$ as the Euclidean distance between their embedded images in $\mathbb{R}^N$, i.e., the ith and jth rows of $\Phi\Lambda^{1/2}$. If a p-dimensional embedding (with $p < \ell$) is desired, we can choose the p columns of $\Phi\Lambda^{1/2}$ corresponding to the p smallest eigenvalues of L to minimize the error of approximation in (P1). Since the columns of $\Phi\Lambda^{1/2}$ are already orthogonal, (P2) is automatically satisfied. Figure 3 is an example that illustrates a nonlocal basis obtained from eigendecomposition of a normalized graph diffusion Laplacian.

3.2. Energy concentration of convolution framelets. Convolution framelets (7) is a signal representation scheme combining both local and nonlocal bases. Advantages of local and nonlocal bases, on their own, are known for specific signal processing tasks, under a general guiding principle seeking signal representations with certain *energy concentration* patterns. Local bases such as wavelets or discrete cosine transforms (DCTs) are known to have "energy compaction" properties, meaning that real-world signals or images often exhibit a pattern of concentration of their energies in a few low-frequency components [3, 18, 42]; this phenomenon is fundamental for many image compression [67, 61] and denoising [20, 19] algorithms. On the other hand, nonlocal bases obtained from nonlinear dimension reduction or kernel PCA—viewed as coordinate functions defining an embedding of the data set—strive to capture, with only a relatively small number of basis functions, as much "variance" within the data set as possible; large portions of the variability of the data set are thus encoded primarily in the leading basis functions [36]. In the context of manifold learning, where the data points are assumed to be sampled from a smooth manifold, the number of eigenvectors corresponding to "relatively large" eigenvalues of a covariance matrix is treated as an estimate for the dimension of the underlying smooth manifold [65, 52, 5, 39].

In practice, energy concentration patterns of signal representation in specific domains have been widely exploited to design powerful regularization schemes for reconstructing signals from noisy measurements. Since convolution framelets combine local and nonlocal bases, it is reasonable to expect that convolution framelet coefficients of typical signals tend to have energy concentration properties as well. To give a motivating example, consider the case in which both local and nonlocal bases concentrate energy on their low-frequency components, and basis functions are sorted in the order of increasing frequencies: typically the coefficient matrix $C = \Phi^\top FV$ will then concentrate its energy on the upper left block storing coefficients for convolution framelets corresponding to both local and nonlocal low-frequency basis functions. As an extreme example, if Φ, V in (5) come from the full-size SVD of F, i.e.,

$$F = \Phi\Sigma V^\top, \quad F, \Sigma \in \mathbb{R}^{N\times\ell}, \Phi \in \mathbb{R}^{N\times N}, V \in \mathbb{R}^{\ell\times\ell},$$

then the only nonzero entries in the coefficient matrix $C = \Phi^\top FV = \Sigma$ lie along the diagonal of its upper $\ell \times \ell$ block. We illustrate in Figure 4 the energy concentration of several different types of convolution framelets on a 1D random signal. Figure 5 demonstrates the energy concentration of a two-dimensional (2D) example using the same cropped BARBARA image as in Figure 2 and Figure 3, in which we explore four different types of local bases V with the nonlocal basis Φ fixed as the graph Laplacian eigenvectors shown in Figure 3; notice that in this example the energy concentrates more compactly in SVD and Haar bases than in DCT and random bases.

An interesting fact to notice is the following: in order for convolution framelets to have a structured energy concentration, it is *not* strictly required that *both* local and nonlocal bases have energy concentration properties. In a sense, regularization schemes based on convolution framelets are more flexible since the energy compaction effects of a local (resp., nonlocal) basis can be amplified through coupling with a nonlocal (resp., local) basis. More specifically, given $\Phi \in \mathbb{R}^{N\times N}$ satisfying mild assumptions,[7] we can systematically construct a local basis V via minimizing a "linear reconstruction loss" such that the coefficient matrix $\Phi^\top FV$ concentrates its energy on the upper left block; this is the focus of subsection 3.3.

3.3. Energy concentration guarantee via linear reconstruction.

Throughout this subsection, we will adopt the nonlocal point of view described in subsection 3.1 and treat the patch matrix $F \in \mathbb{R}^{N\times\ell}$ of a signal $f \in \mathbb{R}^N$ as a point cloud $\mathscr{F} = \{F_0, \ldots, F_{N-1}\}$ consisting of N points in $\mathbb{R}^\ell$. Let $\mathcal{E} : \mathbb{R}^\ell \supset \mathscr{F} \to \mathbb{R}^p$ be an embedding satisfying (P1) and (P2), with $1 \le p \le \ell$. Our goal is to ensure that the dimension reduction $\mathcal{E}$ does not lose information in the original data set $\mathscr{F}$, by requiring the *approximate invertibility*[8] of $\mathcal{E}$ on its image; as will be seen in Proposition 2, the optimal L^2-reconstruction of $\mathscr{F}$ from its image $\mathcal{E}(\mathscr{F})$ leads to a local basis $V \in \mathbb{R}^\ell$. This particular local basis, paired with the nonlocal orthogonal system read off from the embedding $\mathcal{E}$, renders convolution framelets that concentrate energy on the upper left block.

Let us motivate the linear reconstruction framework by considering a linear embedding $\mathcal{E} : \mathbb{R}^\ell \to \mathbb{R}^p$ with $1 \le p \le \ell$. Assume $\widetilde{A} \in \mathbb{R}^{\ell\times p}$ is full-rank, and $\mathscr{X} = \{x_1, \ldots, x_N\} \subset$

[7]E.g., the leading columns of Φ give near-isometric embeddings of the rows of F satisfying (P1) and (P2) in subsection 3.1; when columns of Φ are not orthogonal, a QR decomposition can be applied (see subsection 3.1) as well.

[8]We remark that the "invertibility" or "reconstruction" assumptions have been widely exploited in dimension reduction techniques; see, e.g., [46, 26].

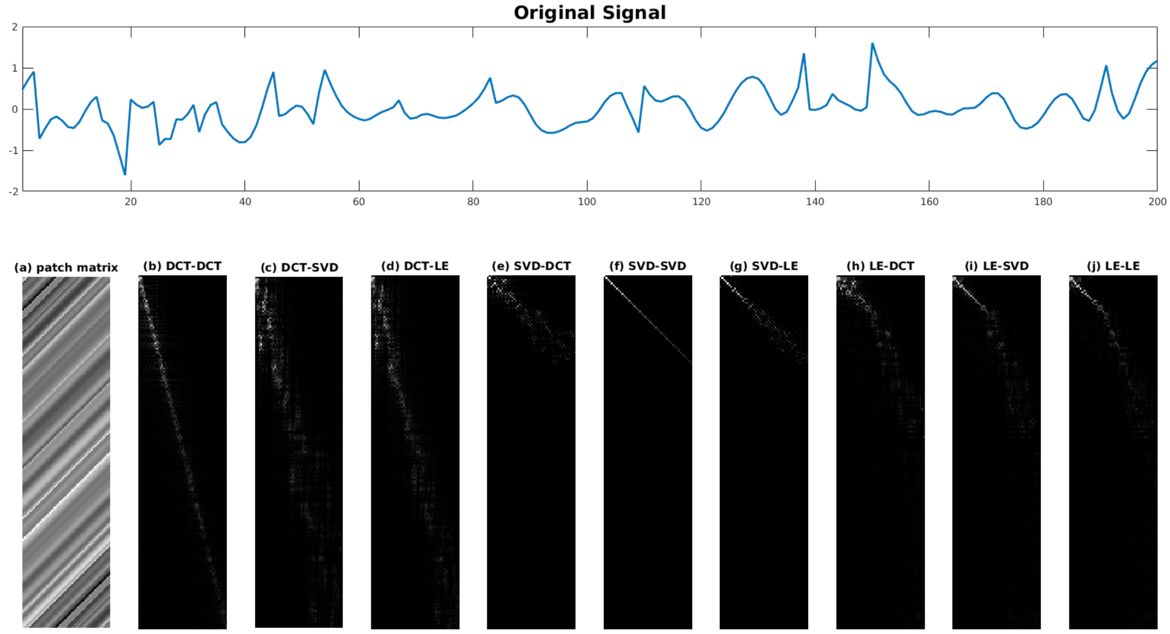

Figure 4. *A* 1*D signal of length* $N = 200$ *and several convolution framelet coefficient matrices with fixed patch size* $\ell = 50$*. Top: A piecewise smooth* 1*D signal* f *randomly generated from the stochastic model proposed in* [12]*. Bottom:* (a) *The patch matrix* F *of the signal* f *on the top panel.* (b)–(j) *Energy concentration patterns of the coefficients of* f *in several convolution framelets. The titles of subplots* (b)–(j) *indicate the different choices (discrete cosine transform (DCT), singular value decomposition (SVD), Laplacian eigenmaps (LE)) for the nonlocal basis* Φ *(appearing before the dash) and the local basis* V *(appearing after the dash). These plots suggest that data-adaptive nonlocal bases (SVD or LE) tend to concentrate more energy on the upper left part of the coefficient matrix* C *than DCT does.*

range$\,(\widetilde{A}) \subset \mathbb{R}^\ell$, i.e., points in $\mathscr{X}$ are sampled from the p-dimensional linear subspace of $\mathbb{R}^\ell$ spanned by the columns of $\widetilde{A}$. Denote $X \in \mathbb{R}^{N\times\ell}$ for the data matrix storing the coordinates of x_j in its jth row, and $X\widetilde{A} = \widetilde{\Phi}\Sigma V_0^\top$ for the reduced SVD of $X\widetilde{A}$ (thus $\widetilde{\Phi} \in \mathbb{R}^{N\times p}$, $V_0 \in \mathbb{R}^{p\times p}$, and $\Sigma \in \mathbb{R}^{p\times p}$ contains the singular values of $X\widetilde{A}$ along the diagonal and zeros elsewhere). Define $A := \widetilde{A}V_0 \in \mathbb{R}^{\ell\times p}$, and consider the linear embedding $\mathcal{E} : \mathbb{R}^\ell \to \mathbb{R}^p$ given by

$$\mathcal{E}\,(x) = x^\top A \quad \forall x \in \mathbb{R}^\ell.$$

In matrix notation, the image of $\mathscr{X}$ under $\mathcal{E}$ is XA. Note that (P2) is automatically satisfied because the columns of $XA = \widetilde{\Phi}\Sigma$ are orthogonal.

Now that $\mathscr{X}$ is in range$\,(\widetilde{A})$ and $V_0 \in \mathbb{R}^{p\times p}$ is orthonormal, we also have $\mathscr{X} \subset$ range$\,(A)$ and thus can write $X^\top = AB$ for some $B \in \mathbb{R}^{p\times N}$. It follows that $\mathcal{E}$ is invertible on $\mathcal{E}\,(\mathscr{X})$ since

$$X^\top = AB = AA^\dagger AB = AA^\dagger X^\top = \left(AA^\dagger\right)^\top X^\top = \left(A^\dagger\right)^\top A^\top X^\top,$$

where $A^\dagger = \left(A^\top A\right)^{-1} A^\top$ is the Moore–Penrose pseudoinverse of A, or equivalently

$$X = XAA^\dagger. \tag{10}$$

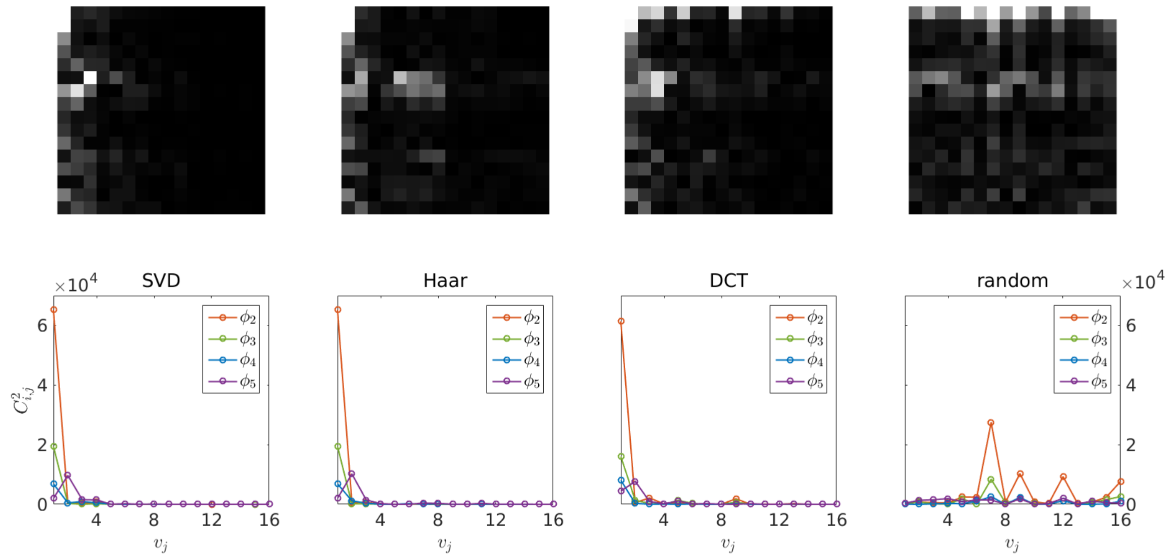

Figure 5. *Energy concentration of convolution framelet coefficient matrices of the same cropped* 128×128 BARBARA *image shown in Figure 2 and Figure 3, with fixed nonlocal basis* Φ *(Laplacian eigenfunctions) and different local bases* V*. The patch size is fixed as* $\ell = 4 \times 4$*. Top: The top* 16×16 *blocks (corresponding to convolution framelets* $\phi_i * v_j$ *with* $1 \le i, j \le 16$*) of the convolution framelet coefficient matrices using (from left to right) SVD basis, Haar basis, DCT basis, and random orthonormal basis. Bottom: Squared coefficients* C_{ij}^2 *for the second to the fifth rows in each coefficient block on the top panel. For each type of local basis, the line corresponding to* ϕ_i $(i = 2, 3, 4, 5)$ *depicts* C_{ij}^2 *with* i *fixed and* j *ranging from 1 to 16.*

In other words, in this case the dimension reduction $\mathcal{E}$ is "lossless" in the sense that we can perfectly reconstruct $\mathscr{X} \subset \mathbb{R}$ from its embedded image $\mathcal{E}(\mathscr{X})$ in a space of lower dimension. Using a Gram–Schmidt process, we can write $A^\dagger = R\widetilde{V}^\top$, where $R \in \mathbb{R}^{p\times p}$ is upper-triangular and $\widetilde{V} \in \mathbb{R}^{p\times \ell}$ is orthogonal. This transforms (10) into

$$X = XAR\widetilde{V}^\top = \widetilde{\Phi}\Sigma R\widetilde{V}^\top = \sum_{1\le i\le j\le p} \Sigma_{ii}\, R_{ij}\, \phi_i\, v_j^\top, \tag{11}$$

where we invoked $XA = \widetilde{\Phi}\Sigma$ and denoted $\{\phi_i \mid 1 \le i \le p\}$, $\{v_j \mid 1 \le j \le p\}$ for the columns of $\widetilde{\Phi}$, $\widetilde{V}$, respectively; note that the coefficient matrix ΣR is upper-triangular. Let $\Phi \in O(N)$ and $V \in O(\ell)$ be orthonormal matrices, the columns of which extend $\{\phi_i \mid 1 \le i \le p\}$ and $\{v_j \mid 1 \le j \le p\}$ to complete bases on $\mathbb{R}^N$, $\mathbb{R}^\ell$, respectively. Following section 2, denote the outer products of columns of Φ with columns of V as

$$\Psi_{ij} = \phi_i\, v_j^\top \quad \text{for } 1 \le i \le N,\ 1 \le j \le \ell.$$

The expression (11), now understood as an expansion of X in orthogonal system $\{\Psi_{ij}\}$, uses only $p\,(p+1)/2$ out of a total number of $N \times \ell$ basis functions. It is clear that in this case the energy of X concentrates on (the upper triangular part of) the upper left block of the coefficient matrix $\Phi^\top XV$, or equivalently on components corresponding to $\{\Psi_{ij} \mid 1 \le i \le j \le p\}$. This establishes Proposition 2 below for all linear embeddings satisfying (P2).

A similar argument can be applied to general nonlinear embeddings satisfying (P2); all nonlinear dimension reduction methods based on kernel spectral embedding, such as MDS, Laplacian eigenmaps, and diffusion maps, belong to this category. In these cases we generally cannot expect a perfect reconstruction of type (10), but we can still seek a linear reconstruction in the form of $R\widetilde{V}^\top$, with upper triangular R and orthogonal $\widetilde{V}$, that reduces the reconstruction error between $\widetilde{X}R\widetilde{V}^\top$ and the original X as much as possible.

Proposition 2. *Let $\mathscr{X} = \{x_1, \ldots, x_N\}$ be a point cloud in $\mathbb{R}^\ell$, $1 \le p \le \ell$, and $\mathcal{E} : \mathbb{R}^\ell \supset \mathscr{X} \to \mathbb{R}^p$ an embedding satisfying* (P2). *Let $X \in \mathbb{R}^{N\times\ell}$ be the matrix storing the coordinates of x_j in its jth row, and let $\widetilde{X} \in \mathbb{R}^{N\times p}$ be the matrix storing the coordinates of $\mathcal{E}(x_j)$ in its jth row ($1 \le j \le N$). For $V_{\mathcal{E}}$ given by*

$$(V_{\mathcal{E}}, R_{\mathcal{E}}) = \underset{\substack{\widetilde{V}^\top\widetilde{V}=I_p,\, \widetilde{V}\in\mathbb{R}^{\ell\times p} \\ \widetilde{R}_{ij}=0,\, 1\le j<i\le p}}{\arg\min} \|\widetilde{X}\,\widetilde{R}\,\widetilde{V}^\top - X\|_{\mathrm{F}}^2, \tag{12}$$

construct V that extends $V_{\mathcal{E}}$ to a complete orthonormal basis in $\mathbb{R}^l$; for $\Phi_{\mathcal{E}}$ derived from the decomposition

$$\widetilde{X} = \Phi_{\mathcal{E}}C_{\mathcal{E}}, \quad \Phi_{\mathcal{E}} \in \mathbb{R}^{N\times p} \text{ orthonormal}, C_{\mathcal{E}} \in \mathbb{R}^{p\times p} \text{ diagonal}, \tag{13}$$

also construct Φ that extends $\Phi_{\mathcal{E}}$ to a complete orthonormal basis in $\mathbb{R}^N$. Then $C = \Phi^\top XV \in \mathbb{R}^{N\times\ell}$ concentrates its energy on the upper triangle part of its upper left $p \times p$ block.

Proof. Let $\Phi_{\mathcal{E}}$, Φ be defined as in the statement of Proposition 2, $\widetilde{V}_0 \in \mathbb{R}^{\ell\times p}$ an arbitrary matrix with orthonormal columns, and V_0 an arbitrary extension of $\widetilde{V}_0$ to an orthonormal basis on $\mathbb{R}^\ell$. The first term within the Frobenius norm of (12) can be rewritten as

$$\begin{aligned}\widetilde{X}\,\widetilde{R}\,\widetilde{V}_0^\top &= \Phi_{\mathcal{E}}\,C_{\mathcal{E}}\,[\widetilde{R},\, \mathbf{0}_{\ell,\,\ell-p}]\,V_0^\top \\ &= \Phi\begin{bmatrix} C_{\mathcal{E}} \\ \mathbf{0}_{\,N-p,\,p}\end{bmatrix}[\widetilde{R},\, \mathbf{0}_{\,\ell,\,\ell-p}]\,V_0^\top = \Phi\begin{bmatrix} C_{\mathcal{E}}\,\widetilde{R} & \mathbf{0}_{\,\ell,\,\ell-p} \\ \mathbf{0}_{\,N-p,\,p} & \mathbf{0}_{\,N-p,\,\ell-p}\end{bmatrix}V_0^\top.\end{aligned} \tag{14}$$

The minimization problem in (12) can thus be reformulated as

$$\min_{\substack{\widetilde{V}\in O(\ell), \\ \widetilde{R}_{ij}=0,\,1\le j<i\le p}} \left\| \begin{bmatrix} C_{\mathcal{E}}\,\widetilde{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} - C \right\|_{\mathrm{F}}^2, \quad \text{where } C = \Phi^\top XV_0. \tag{15}$$

For any fixed orthonormal $V_0 \in \mathbb{R}^{\ell\times\ell}$ (which also fixes C since Φ and X are already given), the optimal upper triangular matrix $\widetilde{R}^*$ is clearly characterized by $\widetilde{R}^*_{ij} = \left(C_{\mathcal{E}}^{-1}C\right)_{ij}$ for all $1 \le i \le j \le p$. In fact, if we partition the matrix C into blocks compatible with the block structure in (15), denoted as

$$C = \begin{bmatrix} C_{LT} & C_{RT} \\ C_{LB} & C_{RB}\end{bmatrix},$$

then $C_{\mathcal{E}}\widetilde{R}$ must cancel out with the upper triangle part of C_{LT} in order to achieve the minimum of the minimization problem in (15). The optimization problem in (15) is thus equivalent to minimizing the L^2 energy of the remaining strictly lower triangular part of C_{LT} together with the L^2 energy of the other three blocks C_{RT}, C_{LB}, and C_{RB}. In addition, since $\|C\|_{\mathrm{F}}^2 = \|X\|_{\mathrm{F}}^2$ is constant, this is further equivalent to maximizing the L^2 energy of the upper triangular part of C_{LT} (which gets canceled out with $C_{\mathcal{E}}\widetilde{R}^*$ anyway). Simply put, we have

$$\operatorname*{arg\,min}_{\widetilde{V}\in O(\ell)} \sum_{\substack{1\leq j<i\leq p,\\ \text{or } i,j>p}} C_{ij}^2 \quad = \quad \operatorname*{arg\,max}_{\widetilde{V}\in O(\ell)} \sum_{1\leq i\leq j\leq p} C_{ij}^2. \tag{16}$$

This indicates that the optimal local basis $V_{\mathcal{E}}$, and consequently its extension V to a complete orthonormal basis on $\mathbb{R}^\ell$, must concentrate as much energy of the coefficient matrix C as possible on the upper triangular part[9] of the upper left $p \times p$ block. ■

Remark 3. The core idea behind Proposition 2 is to approximate the inverse of an arbitrary (possibly nonlinear) dimension reduction embedding $\mathcal{E}$ using a global linear function

$$\mathcal{E}^{-1}(\widetilde{X}) \approx \widetilde{X}\, R\, \widetilde{V}^\top, \tag{17}$$

where the upper triangular matrix $R \in \mathbb{R}^{p\times p}$ and the orthonormal matrix $\widetilde{V} \in \mathbb{R}^{\ell\times p}$ together play the role of $A^\dagger$ in (10) for linear embeddings. Note that it is straightforward to incorporate a bias correction in the linear reconstruction (17) by considering $\mathcal{E}^{-1}(\widetilde{X}) \approx \widetilde{X}\, R\, \widetilde{V}^\top - B$, where $B \in \mathbb{R}^{N\times\ell}$ is a "centering matrix"; we assume $B = 0$ in Proposition 2 for simplicity, but the argument can be easily extended to $B \neq 0$.

Remark 4. From the perspective of basis selection in convolution framelets, Proposition 2 states that, given a fixed nonlocal basis $\Phi_{\mathcal{E}}$ that induces a nonlinear embedding $\widetilde{X}$ of the patch matrix X, the "optimal" local basis $V_{\mathcal{E}}$ is the minimizer of the optimization problem (12) that provides the best linear reconstruction of X from $\widetilde{X}$. In practice, we have only an approximation of X, its corresponding manifold, and the induced nonlinear basis $\Phi_{\mathcal{E}}$; although one could optimize with respect to these approximations, this would not lead to a truly optimal basis.

Remark 5. As will be seen in section 4, LDMM [45] implicitly exploits the energy concentration pattern characterized in Proposition 2. More systematic exploitation of the energy concentration pattern leads to our improved design of *reweighted LDMM*; see subsection 4.2.

Example: Optimal local basis for multidimensional scaling (MDS). When $\mathcal{E}$ is given by MDS, the optimal local basis V in the sense of (12) consists of the right singular vectors of the centered data matrix X. To see this, first recall that in MDS the eigendecomposition is performed on the doubly centered distance matrix $K = \frac{1}{2}HD^2H$, where $(D^2)_{ij} = d^2(X_i, X_j)$ and $H = I_N - \frac{1}{N}\mathbb{1}_N \mathbb{1}_N^\top$; coordinate functions for the low dimensional embedding are then chosen as the eigenvectors of K corresponding to the largest eigenvalues, weighted by the square roots of their corresponding eigenvalues. In particular, when $d(\cdot,\cdot)$ is the Euclidean distance on $\mathbb{R}^\ell$, one has $K = -HXX^\top H$, and the eigenvectors of K correspond to the left singular vectors of the centered data matrix HX. (Here the centering matrix is $B = HX - X$;

[9] One could also require that the energy concentrates on the lower triangle. Yet this is equivalent to changing V to PVP, where $P = \left[\begin{smallmatrix} J_p & 0 \\ 0 & I_{\ell-p}\end{smallmatrix}\right]$, and J_p is antidiagonal with nonzero entries all equal to one.

see Remark 3.) Let $HX = U_X \Sigma_X V_X^\top$ be the reduced SVD of HX as computed in the standard MDS procedure. Then the optimal V for (12) is exactly V_X, and the corresponding matrix basis has the sparsest representation of X. The proof of this statement can be found in Appendix D.

3.4. Connection with nonlocal transform-domain image processing techniques. Under certain circumstances, the framework of convolution framelets can be interpreted as a nonlocal method applied to signal representation in a transform domain. For instance, if we use wavelets for the local basis V, and eigenvectors of the normalized graph diffusion Laplacian L (see (9)) for the nonlocal basis Φ, then L can be seen as defined on the wavelet coefficients since

$$W_{ij} = \exp\left(-\|F_i - F_j\|^2/\epsilon\right) = \exp\left(-\|F_i V - F_j V\|^2/\epsilon\right) \quad \forall 1 \le i, j \le N.$$

Convolution framelets thus have the potential to serve as a natural framework for other nonlocal transform domain image processing techniques. As an example, we show in what follows that BM3D [16, 17], a widely accepted state-of-the-art image denoising algorithm based on nonlocal filtering in a transform domain, may also be interpreted through our convolution framelet framework, with a slightly extended notion of "nonlocal basis."

The basic algorithmic paradigm of BM3D can be roughly summarized in three steps. First, for a given image decomposed into N patches of size ℓ, denoted as $F_1, \ldots, F_N$, a block-matching process groups all patches similar to F_i in a set S_i and forms matrix $F_{S_i} \in \mathbb{R}^{|S_i| \times \ell}$ consisting of patches in S_i; denote $\sigma = \sum_{i=1}^N |S_i|$. Second, let $V_{S_i} \in \mathbb{R}^{\ell \times \ell}$ be a local basis,[10] let $\Phi_{S_i} \in \mathbb{R}^{|S_i| \times |S_i|}$ be a nonlocal basis for S_i, and calculate coefficient matrix $C_{S_i} = \Phi_{S_i}^\top F_{S_i} V_{S_i}$ for group S_i; the matrix F_{S_i} is then denoised by hard-thresholding (or Wiener filtering) C_{S_i} and estimating $\widehat{F}_{S_i} = \Phi_{S_i} \widehat{C}_{S_i} V_{S_i}^\top$ from the resulting coefficient matrix $\widehat{C}_{S_i}$. In matrix form, this can be written as

$$\begin{bmatrix} \widehat{F}_{S_1} \\ \vdots \\ \widehat{F}_{S_N} \end{bmatrix} = \begin{bmatrix} \Phi_{S_1} & & \\ & \ddots & \\ & & \Phi_{S_N} \end{bmatrix} \begin{bmatrix} \widehat{C}_{S_1} & & \\ & \ddots & \\ & & \widehat{C}_{S_N} \end{bmatrix} \begin{bmatrix} V_{S_1} & & \\ & \ddots & \\ & & V_{S_N} \end{bmatrix}^\top . \tag{18}$$

In the third and last step, pixel values at each location of the image are reconstructed using a weighted average of all patches covering that location in the union of all estimated $\widehat{F}_{S_i}$'s; the contribution of an estimated patch contained in $\widehat{F}_{S_i}$ is proportional to $w_i := \|C_{S_i}\|_{\ell_0}^{-1}$, i.e., inversely proportional to the sparsity of C_{S_i}. If we set $A_F \in \mathbb{R}^{N \times \sigma}$ to be a weighted incidence matrix defined by

$$(A_F)_{kq} = \begin{cases} w_q & \text{if patch } F_k \text{ is contained in } S_q, \\ 0 & \text{otherwise} \end{cases}$$

and let $D \in \mathbb{R}^{N \times N}$ be a diagonal matrix with

$$D_{kk} = \sum_{q=1}^{\sigma} (A_F)_{kq} ,$$

[10] In the original BM3D [16], V_{S_i} is set as DCT, DFT, or wavelet, and Φ_{S_i} is set as the 1D Haar transform; in BM3D-SAPCA [17], V_{S_i} is set to the principal components of F_{S_i} when $|S_i|$ is large enough.

then the patch matrix of the original noise-free image is estimated via

(19)
$$\widehat{F} = D^{-1}A_F \begin{bmatrix} \widehat{F}_{S_1} \\ \vdots \\ \widehat{F}_{S_N} \end{bmatrix} = D^{-1}A_F \begin{bmatrix} \Phi_{S_1} & & \\ & \ddots & \\ & & \Phi_{S_N} \end{bmatrix} \begin{bmatrix} \widehat{C}_{S_1} & & \\ & \ddots & \\ & & \widehat{C}_{S_N} \end{bmatrix} \begin{bmatrix} V_{S_1} & & \\ & \ddots & \\ & & V_{S_N} \end{bmatrix}^{\top}.$$

The denoised image $\hat{f}$ is finally constructed from $\widehat{F}$ by taking a weighted average along antidiagonals of $\widehat{F}$, with adaptive weights depending on the pixels.

In this three-step procedure, if we define

$$\Phi = D^{-1}A_F \begin{bmatrix} \Phi_{S_1} & & \\ & \ddots & \\ & & \Phi_{S_N} \end{bmatrix}, \tag{20}$$

$$V = \begin{bmatrix} V_{S_1} & & \\ & \ddots & \\ & & V_{S_N} \end{bmatrix}, \tag{21}$$

then Φ, V together define a tight frame similar to our construction of convolution framelets in section 2. The main difference here is that our energy concentration intuition described in subsection 3.2 would not carry through to this setup, because in general every patch appears in multiple F_{S_i}'s and it is difficult to conceive that Φ consistently defines an embedding $\mathcal{E}$ for the patches of the image. This technicality, however, can be easily remedied if we extend our framework from a global embedding over the entire data set $\mathscr{X}$ to a union of "local embeddings" on "local charts" of $\mathscr{X}$, i.e.,

$$\mathcal{E}_{S_i} : \mathbb{R}^{\ell} \supset S_i \to \mathbb{R}^{p_i}, \quad i = 1, \ldots, N,$$

where $\mathscr{X}$ is covered by the unions of all S_i's; note that the target spaces $\mathbb{R}^{p_i}$ do not even have to be of the same dimension (assuming $p_i \leq \ell$ for simplicity). For each embedding $\mathcal{E}_{S_i}$, $\Phi_{S_i} \in \mathbb{R}^{|S_i|\times|S_i|}$ and $V_{S_i} \in \mathbb{R}^{\ell\times\ell}$ define nonlocal and local orthonormal bases, respectively. It can be expected that the energy concentration of convolution framelet coefficients in this setup will be more involved since both concentration patterns within and across local embedding spaces will be intertwined. We will further explore these interactions in a future work.

4. LDMM as a regularization on convolution framelet coefficients.

In this section, we connect the discussion on convolution framelets with the recent development of the low dimensional manifold model (LDMM) [45] for image processing. We explain in subsection 4.1 that the objective function in the optimization framework of LDMM can be reinterpreted as a "weighted energy" of the convolution framelet coefficients; we push this intuition further in subsection 4.2 and propose a "reweighted version" of LDMM that utilizes more thoroughly the energy concentration pattern explained in subsection 3.2 and subsection 3.3.

We begin with a brief sketch of the fundamental ideas underlying LDMM; interested readers are referred to [45] for a detailed exposition. The basic assumption in LDMM is that the collection of all patches of a fixed size from an image live on a low dimensional smooth manifold isometrically embedded in a Euclidean space. If we denote f for the image and write $\mathcal{M}(f) = \mathcal{M}_\ell(f)$ for the manifold of all patches of size ℓ from f, then the image f can be reconstructed from its (noisy) partial measurements y by solving the optimization problem

$$\arg\min_f \quad \dim(\mathcal{M}(f)) + \mu \|y - Sf\|_2^2, \tag{22}$$

where μ is a parameter and S is the measurement (sampling) matrix. In other words, LDMM utilizes the dimension of the "patch manifold" $\mathcal{M}(f)$ as a regularization term in a variational framework. It is shown[11] in [45] that

$$\dim(\mathcal{M}(f)) = \sum_{j=1}^{\ell} |\nabla_{\mathcal{M}} \alpha_j(x)|^2, \tag{23}$$

where α_j is the jth coordinate function on $\mathcal{M}(f)$, i.e.,

$$x = (\alpha_1(x), \ldots, \alpha_\ell(x)) \quad \forall x \in \mathcal{M}(f) \subset \mathbb{R}^\ell,$$

and $\nabla_{\mathcal{M}}$ is the gradient operator on the Riemannian manifold $\mathcal{M}$. Note that α_j corresponds exactly to the jth column of the patch matrix F of f; see (1) and Figure 1.

With the right-hand side of (23) substituted for $\dim(\mathcal{M}(f))$ in (22), a split Bregman iterative scheme can be applied to the optimization problem (22), casting the latter into subproblems that iteratively optimize the dimension regularization with respect to each coordinate function α_j and the measurement fidelity term. In the nth iteration, the subproblem of dimension regularization decouples into the following optimization problems on each coordinate function:

$$\min_{\alpha_j \in H^1(\mathcal{M}_{(n-1)})} \|\nabla \alpha_j\|^2_{L^2(\mathcal{M}_{(n-1)})} + \mu \sum_{x \in \mathcal{M}_{(n-1)}} |\alpha_j(x) - e_j(x)|^2, \quad j = 1, \ldots, \ell, \tag{24}$$

where $\mathcal{M}_{(n-1)} = \mathcal{M}(f_{(n-1)})$ is the patch manifold associated with the reconstruction $f_{(n-1)}$ from the $(n-1)$th iteration, μ is a penalization parameter, and e_j is a function on this manifold originating from the split Bregman scheme. The Euler–Lagrange equations of the minimization problems in (24) are cast into integral equations by the point integral method (PIM) and then discretized as

$$\left[D_{(n-1)} - W_{(n-1)} + \mu W_{(n-1)}\right] F^j = \mu W_{(n-1)} E^j_{(n-1)}, \quad j = 1, \ldots, \ell, \tag{25}$$

where F^j, $E^j_{(n-1)}$ are the jth columns of the patch matrix F and the matrix $E_{(n-1)}$, corresponding to α_j and e_j in (24), respectively; the weighted adjacency matrix $W_{(n-1)}$ and the diagonal degree matrix $D_{(n-1)}$, both introduced by PIM, are updated in each iteration after building the patch matrix $F_{(n-1)}$ from $f_{(n-1)}$. We refer interested readers to [45] for more details.

[11] We give a simplified proof of identify (23) in Appendix C.

The rest of this section presents a connection we discovered between solving (25) and an ℓ_2-regularization problem on the convolution framelet coefficients of f. We will establish in subsection 4.1 the equivalence between (22) and an optimization problem of the form

$$\min_f \quad \sum_{i,j} \widetilde{\lambda}_i \left|\langle f, \psi_{ij}\rangle\right|^2 + \text{“fidelity of } f\text{,”} \tag{26}$$

where $\{\psi_{ij}\}$ is a system of convolution framelets associated with f, and $0 \le \widetilde{\lambda}_1 \le \widetilde{\lambda}_2 \le \cdots$ is a nondecreasing sequence of nonnegative numbers depending on the convolution framelets. An explicit and more formal expression for (26) is given in (33).

4.1. Dimension regularization in convolution framelets. The low dimension assumption in LDMM is reflected in the minimization of a quadratic form derived from (23) for the column vectors of the patch matrix F associated with image f. From a manifold learning point of view, (23) is not the only approach to imposing dimension regularization. Since the columns of F are understood as coordinate functions in $\mathbb{R}^\ell$, F is indeed a data matrix representing a point cloud in $\mathbb{R}^\ell$ (see subsection 3.3). If this point cloud is sampled from a low dimensional submanifold of $\mathbb{R}^\ell$, then one can attempt to embed the point cloud into a Euclidean space of lower dimension without significantly distorting pairwise distances between points. As we have seen in Proposition 2, if there exists a good low dimensional embedding Φ for the data matrix F, the energy of convolution framelet coefficients will concentrate on a small triangular block on the upper left part of the coefficient matrix, provided that an appropriate local basis V is chosen to pair with Φ; a lower intrinsic dimension corresponds to a smaller upper left block and thus more compact energy concentration. Therefore, as an alternative to (23), one can impose regularization on convolution framelet coefficients to push more energy into the upper left block of the coefficient matrix; see details below.

We start by reformulating the optimization problem (22) proposed in [45] as an ℓ_2-regularization problem for convolution framelet coefficients, where the convolution framelets themselves will be estimated along the way since they are adaptive to the data set. For simplicity of notation, we drop the subindex $(n-1)$ in (25) as W, D, and E are fixed when updating the patch matrix within each iteration. To distinguish from the notation $F_0, \ldots, F_{N-1}$, which stands for the rows of matrix F, we use superindices $F^1, \ldots, F^\ell$ to denote the columns of F. Let $F_D^j = D^{1/2}F^j$ and $E_D^j = D^{1/2}E^j$; then the linear systems (25) can be rewritten as

$$(D-W)\,D^{-1/2}\,F_D^j + \mu W\,D^{-1/2}(F_D^j - E_D^j) = 0, \quad j = 1, \ldots, \ell. \tag{27}$$

This system can be instantiated as the Euler–Lagrange equations of a different variational problem. Multiplying both sides of (27) by $(W\,D^{-1/2})^{-1} = D^{1/2}\,W^{-1}$ from the left,[12] we have the equivalent linear system

$$D^{1/2}W^{-1}(D-W)D^{-1/2}F_D^j + \mu(F_D^j - E_D^j) = 0, \quad j = 1, \ldots, \ell. \tag{28}$$

Notice that

$$D^{1/2}W^{-1}(D-W)D^{-1/2} = D^{1/2}W^{-1}D^{1/2} - I = (I-L)^{-1} - I,$$

[12] The random walk matrix $D^{-1}W$ is invertible since all of its eigenvalues are positive; thus W is also invertible.

where L is the normalized graph diffusion Laplacian defined in (9). Therefore, solving (28) is equivalent to minimizing the following objective function:

$$\sum_{j=1}^{\ell} \left[(F_D^j)^\top \left((I-L)^{-1} - I \right) F_D^j + \mu \| F_D^j - E_D^j \|^2 \right].$$

This is also equivalent to determining

$$\underset{F \in \mathbb{R}^{N \times \ell}}{\arg\min} \sum_{j=1}^{\ell} (F^j)^\top \, R_L \, F^j + \mu \| F - E \|^2_{\mathrm{F},\, D^{1/2}}, \tag{29}$$

where $\| \cdot \|_{\mathrm{F},\, A} = \| A \cdot \|_{\mathrm{F}}$ is the A-weighted Frobenius norm, and

$$R_L = D\, W^{-1}(D - W) = D^{1/2}((I-L)^{-1} - I)D^{1/2}. \tag{30}$$

The first term in (29) corresponds to the manifold dimension regularization term proposed in [45], whereas the second term promotes data fidelity. By the equivalence between (27) and (28), it suffices to focus on (29) hereafter.

To motivate our approach to analyzing (29), let us briefly investigate a similar but simpler regularization term based on nonlocal graph Laplacian $\sum_j (F^j)^\top L F^j$, which differs from the dimension regularization term in (29) only in that the graph Laplacian L replaces R_L. If we let $L = \Phi \Lambda \Phi^\top$ be the eigendecomposition of L with eigenvalues $\lambda_1, \ldots, \lambda_N$ on the diagonal of Λ in ascending order, and pick any matrix $\widetilde{V} \in \mathbb{R}^{\ell \times \ell'}$ satisfying $\widetilde{V}\,\widetilde{V}^\top = I_\ell$, then

$$\sum_{j=1}^{\ell} (F^j)^\top L F^j = tr\left(F^\top \Phi \Lambda \Phi^\top F \right) = tr\left((\Phi^\top F \widetilde{V})^\top \Lambda (\Phi^\top F \widetilde{V}) \right) = \sum_{i=1}^{N} \sum_{j=1}^{\ell'} \lambda_i \, C_{ij}^2, \tag{31}$$

where C_{ij} is the (i,j)-entry of $C = \Phi^\top F \widetilde{V}$. Minimizing this quadratic form will thus automatically regularize the energy concentration pattern by pushing more energy to the left part of C where the columns correspond to smaller eigenvalues λ_i. Note that the only assumption we put on $\widetilde{V}$ is that its columns constitute a frame; by Proposition 1, $\widetilde{V}$ being a frame in the patch space already suffices for constructing a convolution framelet system with Φ.

Now we consider the minimization problem (29) with $R_L = D^{1/2}((I-L)^{-1} - I)D^{1/2}$ in the manifold dimension regularization term. Using $L = \Phi \Lambda \Phi^\top$, the operator R_L can be written as $D^{1/2} \Phi \, \widetilde{\Lambda} \, \Phi^\top D^{1/2}$, where $\widetilde{\Lambda} = (I - \Lambda)^{-1} - I$ is a diagonal matrix. Similar to (31), we have

$$\sum_{j=1}^{\ell} (F^j)^\top \, R_L \, F^j = tr\,(F^\top D^{1/2} \Phi \widetilde{\Lambda} \Phi^\top D^{1/2} F) = tr\,(F^\top \widetilde{\Phi} \widetilde{\Lambda} \widetilde{\Phi}^\top F) = \sum_{i=1}^{N} \sum_{j=1}^{\ell} \widetilde{\lambda}_i \widetilde{C}_{ij}^2, \tag{32}$$

where $\widetilde{\Phi} = D^{1/2}\Phi$, $\widetilde{\lambda}_i = \lambda_i / (1 - \lambda_i)$ for $i = 1, \ldots, N$, and $\widetilde{C} = \widetilde{\Phi}^\top F \widetilde{V}$ is the convolution framelet coefficient matrix. The optimization problem (29) can thus be recast as

$$\begin{aligned} \min_{F \in \mathbb{R}^{N \times \ell}} \quad & \sum_{i=1}^{N} \sum_{j=1}^{\ell} \widetilde{\lambda}_i \widetilde{C}_{ij}^2 + \mu \| F - E \|^2_{\mathrm{F},\, D^{1/2}} \\ \text{s.t.} \quad & \widetilde{C} = \widetilde{\Phi}^\top F \widetilde{V}. \end{aligned} \tag{33}$$

Ideally, if the first p columns of $\widetilde{\Phi}$ provide a low dimensional embedding of the patch manifold $\mathcal{M}$ with small isometric distortion, then $\lambda_i \approx 1$ for all $i > p$, which corresponds to large $\widetilde{\lambda}_i$ and forces $\widetilde{C}_{ij}$ for the optimal $\widetilde{C}$ to be close to 0 for all $j \geq 1$ and all $i > p$. Intuitively, since $0 \leq \widetilde{\lambda_1} \leq \widetilde{\lambda_2} \leq \cdots \leq \widetilde{\lambda_N}$, the coefficient matrix $\widetilde{C}$ of the minimizer of (33) will likely concentrate most of its $\|\widetilde{C}\|_{\mathrm{F}}^2$ energy on its top few rows corresponding to the smallest eigenvalues. Note that (32) imposes a much stronger regularization on the lower part of $\widetilde{C}$ than (31) does, since λ_j's are bounded from above by 1 but $\widetilde{\lambda}_j$ can grow to $+\infty$ as subindex j increases.

Remark 6. In a recent work [56], the original LDMM framework was extended by replacing the original patches by *semilocal patches*, i.e., patches with attached spatial (x, y)-coordinates. As reported in [56], this led to enhanced reconstruction quality as well as a reduced number of iterations. This extension can be incorporated as well into the optimization paradigm established in this section, if the combined spatial and pixel-value distances are used to compute pairwise patch similarities in the construction of weighted graph adjacency matrix W, leading to a new nonlocal spectral basis. The missing pixel values are solved by a linear equation similar to (25), so that our optimization formulation on the patch matrix still carries through, but now with respect to a different set of convolution framelets generated by the new nonlocal basis.

4.2. Reweighted LDMM. As explained in subsection 4.1, LDMM regularizes the energy concentration of convolution framelet coefficients by pushing the energy to the upper part of the coefficient matrix. This is clearly suboptimal from the point of view of Proposition 2: the energy $\|\widetilde{C}\|_{\mathrm{F}}^2$ should actually concentrate on the *upper left* part as opposed to merely on the upper part of C, at least when an appropriate local basis is chosen. This observation motivates us to modify the objective function in (33) to reflect the stronger pattern of energy concentration pointed out in Proposition 2. We refer to the modified optimization problem as *reweighted LDMM* (rw-LDMM) since it differs from the original LDMM mainly in the weights in front of each $\widetilde{C}_{ij}^2$ in (33).

Note that the objective function in the optimization problem (33) is invariant to the choices of $\widetilde{V}$—this is consistent with the interpretation of the regularization term as an estimate for the manifold dimension (the dimension of a manifold is basis-independent); but we can modify the objective function by incorporating patch bases as well. Consider a matrix $V = [v_1, \ldots, v_l]$ consisting of basis vectors for the ambient space $\mathbb{R}^\ell$ where the patches live, and define s_j, the *energy filtered by* v_j of the signal, as

$$s_j = \|Fv_j\|^2 = \|f * v_j(-\cdot)\|^2 = \sum_{i=1}^{N} C_{ij}^2. \tag{34}$$

Note that s_j is precisely the jth singular value of the patch matrix F when v_j is chosen as the jth right singular vector of F. If s_j decays fast enough as j increases, the patches on average will be approximated efficiently using a few v_j's with large s_j values. As discussed in subsection 3.2, natural candidates of V include DCT bases, wavelet bases, or even SVD bases of F (which are optimal low-rank approximations of F in the L^2-sense; when the true F is unknown, as in the case of signal reconstruction, we can also consider using right singular vectors obtained from an estimated patch matrix).[13]

After choosing such a basis V, the energy of the optimal coefficients matrix C with respect to convolution framelets $\{\psi_{ij}\}$ concentrate mostly within the upper left $p \times r$ block, where r depends on the decay rate of s_j. For this purpose, instead of using weights $\widetilde{\lambda}_i$ in (33) alone, we propose using weights $\widetilde{\lambda}_i \gamma_j$, where γ_j is a weight associated to v_j such that γ_j increases as s_j decreases; one such example[14] is to set $\gamma_j = 1 - s_1^{-1} s_j \in [0, 1]$. In other words, we *reweight* the penalties $\widetilde{\lambda}_i$ to fine-tune the regularization. With this modification, the quadratic form (32) becomes[15]

$$tr\Big((FV\,\Gamma^{1/2})^\top \, R_L \, (FV\,\Gamma^{1/2}) \Big) = \sum_{i=1}^{N} \sum_{j=1}^{\ell} \widetilde{\lambda}_i \gamma_j \, \widetilde{C}_{ij}^2. \tag{35}$$

Substituting this new quadratic energy for the original quadratic energy in (33) and (29) yields the following optimization problem:

$$\begin{aligned} &\operatorname*{arg\,min}_{F \in \mathbb{R}^{N\times \ell}} \sum_{j=1}^{\ell} \gamma_j (Fv_j)^\top R_L \, (Fv_j) + \mu \| F - E \|_{\mathrm{F},\, D^{1/2}}^2 \\ \Leftrightarrow \qquad &\operatorname*{arg\,min}_{F \in \mathbb{R}^{N\times \ell}} tr\Big((FV\,\Gamma^{1/2})^\top \, R_L \, (FV\,\Gamma^{1/2}) \Big) + \mu \| F - E \|_{\mathrm{F},\, D^{1/2}}^2. \end{aligned} \tag{36}$$

Using PIM, the Euler–Lagrange equations of (36) turn into the corresponding linear systems:

$$(\gamma_j (D - W) + \mu W) F v_j = \mu W E v_j, \quad j = 1, \dots, \ell. \tag{37}$$

We shall refer to the optimization problem (36) (sometimes also the linear system (37) when the context is clear) as *reweighted LDMM* (rw-LDMM).

In practice, we observed that it often suffices to reweight the penalties only for the coefficients in the leading columns, i.e., keep the γ_j's in (35) only for $1 \le j \le r$, where r is a relatively small number compared with ℓ. This can be done by first noting that the quadratic energy in (31) equals

$$tr\left(V^\top F^\top \Phi \Lambda \Phi^\top F V \right) = tr\left(V_r^\top F^\top \Phi \Lambda \Phi^\top F V_r \right) + tr\left((F V_r^c)^\top \, \Phi \Lambda \Phi^\top F V_r^c \right), \tag{38}$$

[13]As mentioned in Remark 4, one could construct an "optimal" local basis V by solving (12) using estimates of $F (= X)$ and its spectral basis $\Phi (= \Phi_{\mathcal{E}})$, i.e., finding a local basis by performing a QR decomposition on the estimate of $F^\top \Phi$; however, this would not lead to a truly optimal basis. When we tried this in examples, the performance was slightly inferior to that of using the right singular vectors of the patch matrix as the local basis, yet the computational cost was significantly higher due to the explicit computation of Φ.

[14]We have also experimented with other forms of γ_j—for instance, $\gamma_j = s_1 s_j^{-1} - 1$, which sends γ_j to $+\infty$ when s_j is close to 0 and is thus a stronger regularization than the one used in rw-LDMM (which only sends γ_j to 1 as $s_j \to 0$). We do not use such stronger regularization weights since in practice they tend to produce oversmoothed results for reconstruction. This is not surprising, as natural images may contain intricate details that are encoded in convolution framelet components corresponding to small s_j's; these details are likely smoothed out if γ_j overregularizes the convolution framelet coefficients.

[15]The reweighted quadratic form (35), as well as (39) below, depends on V only through $\Gamma^{1/2}$. In fact, as long as $VV^\top = I_\ell$, there holds $\| x - y \|_{\ell_2} = \| V^\top x - V^\top y \|_{\ell_2}$, and thus W—the weighted adjacency matrix constructed using a Gaussian RBF—is V-invariant; consequently R_L is V-invariant as well.

where $V_r \in \mathbb{R}^{\ell\times r}$ consists of the left r columns of V, and V_r^c consists of the remaining columns. We can then reweight only the first term in the summation on the right-hand side of (38), i.e., replace (35) with

(39)
$$tr\left((FV_r\Gamma_r^{1/2})^\top R_L(FV_r\Gamma_r^{1/2})\right) + tr\left((FV_r^c)^\top \Phi\Lambda\Phi^\top FV_r^c\right) = \sum_{i=1}^{N}\widetilde{\lambda}_i\left(\sum_{j=1}^{r}\gamma_j C_{ij}^2 + \sum_{j=r+1}^{\ell} C_{ij}^2\right).$$

The linear systems (37) change accordingly to

(40)
$$\begin{aligned}(\gamma_j(D-W)+\mu W)Fv_j &= \mu WEv_j, \quad j=1,\ldots,r,\\ ((D-W)+\mu W)Fv_j &= \mu WEv_j, \quad j=r+1,\ldots,\ell.\end{aligned}$$

In all numerical experiments presented in section 5, we set $r \approx 0.2\ell$; i.e., only coefficients in the left 20% columns are reweighted in the regularization. We did not observe serious changes in performance when this economic reweighting strategy was adopted, but the improvement in computational efficiency is significant: for example, when right singular vectors of F are used as a local basis, solving (40) with partial SVD in each iteration is much faster than the full SVD required in (37). One can avoid explicitly computing $v_{r+1},\ldots,v_\ell$ by converting (40) into

(41)
$$\begin{aligned}(\gamma_j(D-W)+\mu W)Fv_j &= \mu WEv_j, \quad j=1,\ldots,r,\\ ((D-W)+\mu W)F\left(I_N - V_rV_r^\top\right) &= \mu WE\left(I_N - V_rV_r^\top\right);\end{aligned}$$

see Algorithm 1 for more details.[16] Variants of Algorithm 1 with other choices of V, such as DCT or wavelet basis, are just simplified versions of Algorithm 1 where V is a fixed input. Regardless of the choice for local basis, rw-LDMM yields consistently better inpainting results than LDMM in all of our numerical experiments; see details in subsection 5.2.

The total cost of Algorithm 1 is $O\left(TN^2\ell\right)$ flops, where T is the number of iterations, N is the number of pixels in the image, and ℓ is the number of pixels in each patch. Within each iteration, the partial SVD in line 7 of Algorithm 1 is performed using the randomized PCA algorithm proposed in [25] at a cost of $O\left(N\ell\log\ell\right)$; constructing the weighted adjacency matrix in line 9 and constructing the diagonal degree matrix in line 10 both cost $O\left(N^2\ell\right)$; each of the ℓ linear systems in lines 15 and 17 is solved using a GMRES routine (with a prefixed number of iterations) in $O\left(N^2\right)$ flops, incurring a total of $O\left(N^2\ell\right)$; the matrix product in line 18 costs $O\left(N\ell^2\right)$. All these sum up to $O\left(N^2\ell\right)$ for each of a total number of T iterations, leading to a total of $O\left(TN^2\ell\right)$ flop counts. Such a computational cost is relatively expensive. As the main focus of this paper is to present the novel local-nonlocal regularization framework based on convolution framelets, rather than designing efficient algorithmic pipelines for LDMM and its variants, we defer the exploration of faster algorithms to future work.

5. Numerical results.

[16] The linear systems in Algorithm 1 actually produce FV_r and $F\left(I_N - V_rV_r^\top\right)$ separately; the two matrices are combined together to reconstruct F through $F = FV_rV_r^\top + F\left(I_N - V_rV_r^\top\right)$.

Algorithm 1. Inpainting using Reweighted LDMM with Local SVD Basis.

procedure RW-LDMM-SVD$(f_{(0)}, \ell)$ ▷ subsampled image $f_{(0)} \in \mathbb{R}^N$, patch size $\ell \in \mathbb{Z}^+$
  $f_{(0)} \leftarrow$ randomly assign values to missing pixels in $f_{(0)}$
  $n \leftarrow 0$, $r \leftarrow \lceil 0.2\ell \rceil$ ▷ reweight only the first r columns
  $d_{(0)} \leftarrow 0 \in \mathbb{R}^{N\times\ell}$
  $F_{(0)} \leftarrow$ patch matrix of $f^{(0)}$ ▷ $F_{(n)} \in \mathbb{R}^{N\times\ell}$
  while not converge **do**
    $s_1, \dots, s_r, V_{(n)}^1, \dots, V_{(n)}^r \leftarrow$ partial SVD of $F_{(n)}$ ▷ $s_i \in \mathbb{R}^+, V_{(n)}^i \in \mathbb{R}^{\ell\times 1}$
    $V_{(n)} \leftarrow \left[V_{(n)}^1, \dots, V_{(n)}^r\right]$ ▷ $V_{(n)} \in \mathbb{R}^{\ell\times r}$
    $W_{(n)} \leftarrow$ weighted adjacency matrix constructed from $F_{(n)}$ ▷ $W_{(n)} \in \mathbb{R}^{N\times N}$

$$W_{(n)}(i,j) = \exp\left(-\frac{\|F_i - F_j\|^2}{\epsilon}\right), \quad 1 \le i,\, j \le N,$$

    $D_{(n)} \leftarrow$ diagonal matrix containing row sums of $W_{(n)}$ ▷ $D_{(n)} \in \mathbb{R}^{N\times N}$

$$D_{(n)}(i,i) = \sum_{j=1}^{N} W_{(n)}(i,j), \quad 1 \le i \le N,$$

    $E_{(n)} \leftarrow F_{(n)} - d_{(n)}$ ▷ $E_{(n)} \in \mathbb{R}^{N\times\ell}$
    $H_{(n)} \leftarrow 0 \in \mathbb{R}^{N\times r}$, $U_{(n)} \leftarrow 0 \in \mathbb{R}^{N\times(\ell-r)}$ ▷ $H_{(n)} \in \mathbb{R}^{N\times r}$, $U_{(n)} \in \mathbb{R}^{N\times(\ell-r)}$
    for $i \leftarrow 1, r$ **do**
      $\gamma_i \leftarrow 1 - s_1^{-1} s_i$
      $H_{(n)}^i \leftarrow$ solution of the linear system ▷ $H_{(n)}^i \in \mathbb{R}^{N\times 1}$

$$\left(\gamma_i \left(D_{(n)} - W_{(n)}\right) + \mu W_{(n)}\right) H_{(n)}^i = \mu W_{(n)} E_{(n)} V_{(n)}^i$$

    end for
    $U_{(n)} \leftarrow$ solution of the linear systems

$$\left(D_{(n)} - W_{(n)} + \mu W_{(n)}\right) U_{(n)} = \mu W_{(n)} E_{(n)} \left(I_N - V_{(n)} V_{(n)}^\top\right)$$

    $\widetilde{F}_{(n+1)} \leftarrow H_{(n)} V_{(n)}^\top + U_{(n)} + d_{(n)}$
    $\widetilde{f}_{(n+1)} \leftarrow$ average out entries of $\widetilde{F}_{(n+1)}$ according to (2)
    $f_{(n+1)} \leftarrow$ reset subsampled pixels to their known values
    $F_{(n+1)} \leftarrow$ patch matrix of $f_{(n+1)}$
    $d_{(n+1)} \leftarrow \widetilde{F}_{(n+1)} - F_{(n+1)}$
    $n \leftarrow n + 1$
  end while
  return $f_{(n)}$
end procedure

5.1. Linear and nonlinear approximation with convolution framelets. For an orthogonal system $\{e_n\}_{n\geq 0}$, the N-term *linear* approximation of a signal f is

$$f_N = \sum_{j=0}^{N-1} \langle f, e_j \rangle \, e_j,$$

whereas the N-term *nonlinear* approximation of f uses the N terms with largest coefficients in magnitude, i.e.,

$$\widetilde{f}_N = \sum_{j\in\mathcal{I}_N} \langle f, e_j \rangle \, e_j,$$

where

$$\mathcal{I}_N \subset \mathbb{N}, |\mathcal{I}_N| = N, \text{ and } |\langle f, e_i \rangle| \geq |\langle f, e_k \rangle| \quad \forall i \in \mathcal{I}_N, \, k \notin \mathcal{I}_N.$$

We compare in this section linear and nonlinear approximations of images using different convolution framelets $\left\{\psi_{ij} = \ell^{-1/2}\phi_i * v_j\right\}$. To make sense of linear approximation, which requires a predetermined ordering of the basis functions, we fix the nonlocal basis $\{\phi_i\}$ to be the eigenfunctions of the normalized graph diffusion Laplacian L (see (9)); ψ_{ij}'s are then ordered according to descending magnitudes $|(1-\lambda_i)\, s_j|$, where λ_i is the ith eigenvalue of L (which lies in $[0,1]$) and s_j is the energy of the function filtered by v_j (see (34)). We take a cropped BARBARA image of size 128×128, as shown in Figure 3, subtract the mean pixel value from all pixels, and then perform linear and nonlinear approximations for the resulting image. Figure 6 presents the N-term linear and nonlinear approximation results with $N = 8$, patch size $\ell = 16$ (4×4 patches), and local basis V chosen as patch SVD basis (right singular vectors of the patch matrix), Haar wavelets, DCT basis, and—as a baseline—randomly generated orthonormal vectors. In terms of visual quality, nonlinear approximation produces consistently better results here than linear approximation; as we also expected, SVD basis, Haar wavelets, and DCT basis all outperform the baseline using random local basis.

The superiority of nonlinear over linear approximation is also justified in terms of the peak signal-to-noise ratio (PSNR) of the reconstructed images. In Figure 7, we plot PSNR as a function of the number of terms used in the approximations. Except for random local basis, PSNR curves for all types of nonlinear approximation are higher than the curves for linear approximation, suggesting that sparsity-based regularization on convolution framelet coefficients may lead to stronger results than ℓ_2-regularization. When the number of terms is large, even nonlocal approximation with random local basis outperforms linear approximation with SVD, wavelets, or DCT basis. Figure 8 shows several convolution framelet components with the largest coefficients in magnitude for each choice of local basis.

5.2. Inpainting with rw-LDMM. We first compare rw-LDMM with LDMM in the same setup as in [45] for image inpainting: given the randomly subsampled original image with only a small portion (e.g., 5% to 20%) of the pixels retained, we reconstruct the image from an initial guess that fills missing pixels with Gaussian random numbers. The mean and variance of the pixel values filled in the initialization match those of the retained pixels. In our numerical experiments, rw-LDMM outperforms LDMM whenever the same initialization is provided. For LDMM, we use the MATLAB code and hyperparameters provided by the

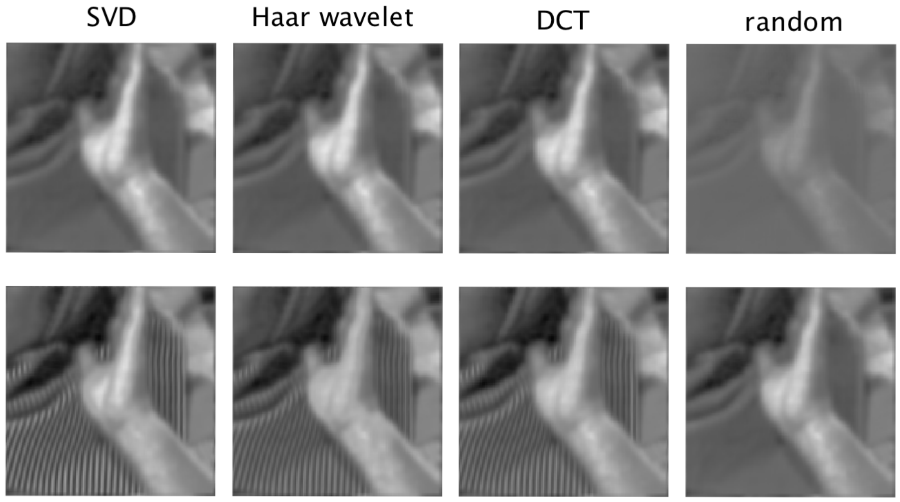

Figure 6. *Linear (top) and nonlinear (bottom) 8-term convolution framelet approximation of the 128×128 cropped* BARBARA *image shown in Figure 3 using 4×4 patches. Except for the last column corresponding to random local basis, nonlinear approximation captures much more texture on the scarf than linear approximation does.*

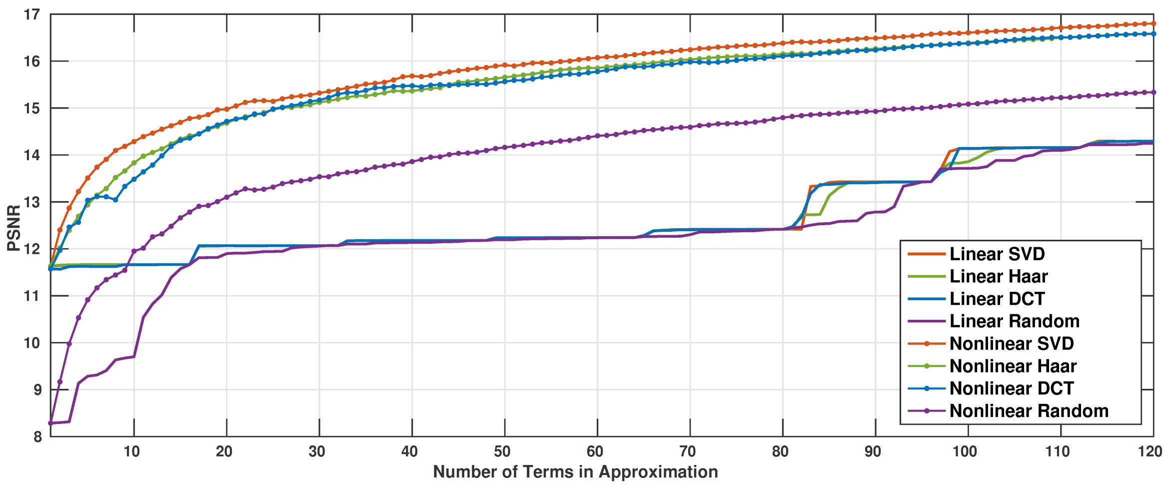

Figure 7. *PSNR as a function of the number of approximation terms in linear and nonlinear approximations of the 128×128 cropped* BARBARA *image in Figure 3 using 4×4 patches. Except for random local basis, the PSNR curves for linear approximation are almost identical.*

authors of [45]; for rw-LDMM, we experiment with both SVD and DCT basis as local basis and reweight only the leading 20% functions in the local basis. We run both LDMM and rw-LDMM for 100 iterations on images of size 256×256, and the patch size is always fixed

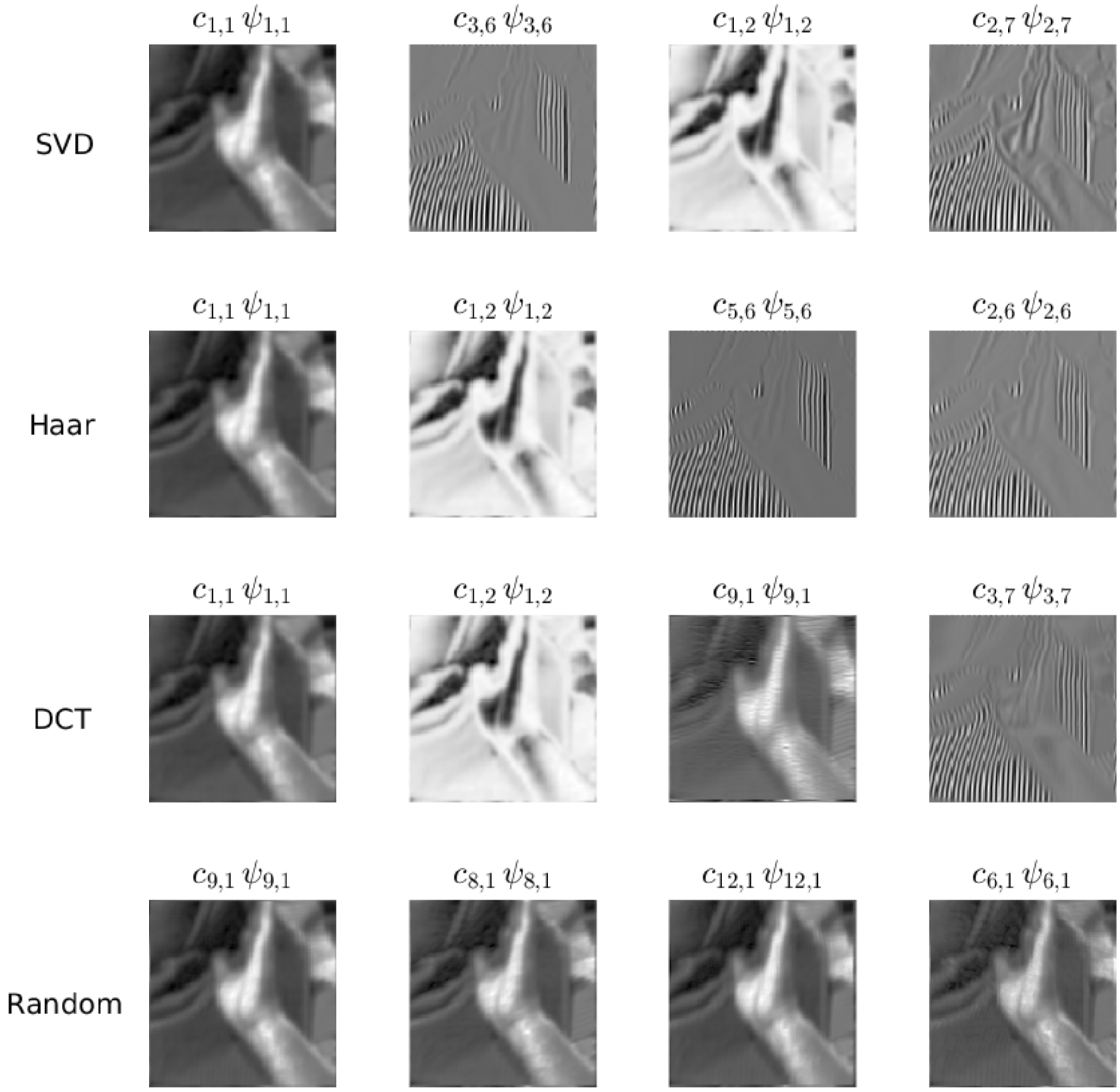

Figure 8. *The first four terms in each type of nonlocal convolution framelet approximation. Components in each row correspond to the four convolution framelet coefficients with the largest magnitudes. The cropped* 128×128 BARBARA *image is the same that as shown in Figure 3 using* 4×4 *patches.*

as 10×10. PSNRs[17] of the reconstructed images obtained after the 100th iteration[18] are

[17] $\mathrm{PSNR}(f, f') \doteq 20 \log_{10}(\mathrm{MAX}(f)) - 10 \log_{10}(\mathrm{MSE}(f, f'))$.

[18] The number of iterations is also a hyperparameter to be determined. We use 100 iterations to make fair comparisons between our results and those in [45]. In case the reconstruction degenerates after too many iterations due to overregularization, one may—for the purpose of comparison only—also look at the highest PSNR within a fixed number of iterations for each algorithm. We include those comparisons in M109144_01.pdf [local/web 5.79MB] as well.

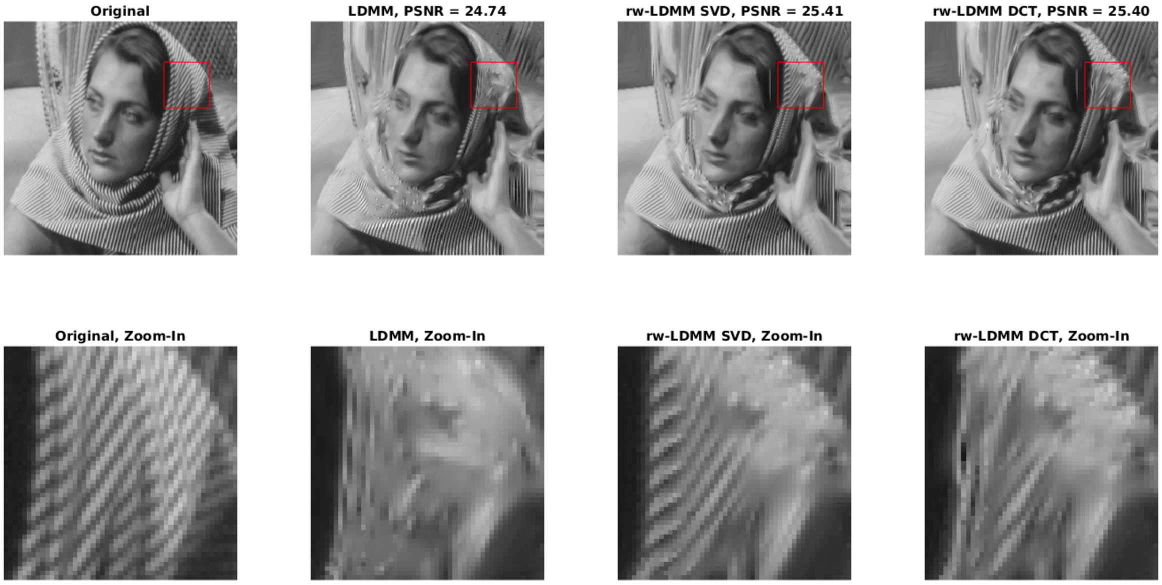

Figure 9. *Reconstructed* 256×256 BARBARA *images from* 10% *randomly subsampled pixels using LDMM and rw-LDMM both with patch size* 10×10 *(fixed for all numerical comparisons in this paper). The same random initialization for missing pixels was used for LDMM and both SVD and DCT versions of rw-LDMM. Top: Both rw-LDMM algorithms outperform LDMM in terms of PSNR. Bottom: Zoom-in views of the* 50×50 *blocks enclosed by red boxes on each reconstructed image illustrate better texture restoration by rw-LDMM.*

used to measure the inpainting quality. Figure 9 compares the three algorithms for a cropped BARBARA image of size 256×256; Figure 10 plots PSNR as a function of the number of iterations and indicates that rw-LDMM outperforms LDMM consistently for a wide range of iteration numbers. More numerical results and comparisons can be found in M109144_01.pdf [local/web 5.79MB].

For image inpainting with randomly sampled pixel values at rate r, a patch size of order $O(1/r)$ leads each patch to contain $O(1)$ known pixel values (in expectation). For example, if $r = 10\%$, then the 10×10 patch size results in an expected number of 10 pixels per patch; on the other hand, the dimension of the patch manifold is likely to increase as the patch size increases, while at the same time the number of available patches (samples) is kept the same as the number of pixels in the image; hence the effective sampling rate of the patch manifold decreases. Therefore, it is advisable to choose a relatively small patch size as long as each patch contains at least a few pixels whose values are known.

We also compare LDMM and rw-LDMM with ALOHA (annihilating filter-based low-rank Hankel matrix) [29], a recent patch-based inpainting algorithm using a low-rank block-Hankel structured matrix completion approach. The central observation underlying ALOHA is that image patches admit annihilating filters, as a consequence of their localized frequencies in the Fourier domain. By a commutative property, for any individual image patch, the frequency localization can be translated into the low-rank property of a particular block Hankel matrix associated with that patch; this low-rank property can be further utilized as a regularization in the process of filling in the missing pixels for image inpainting tasks. The main algorithm

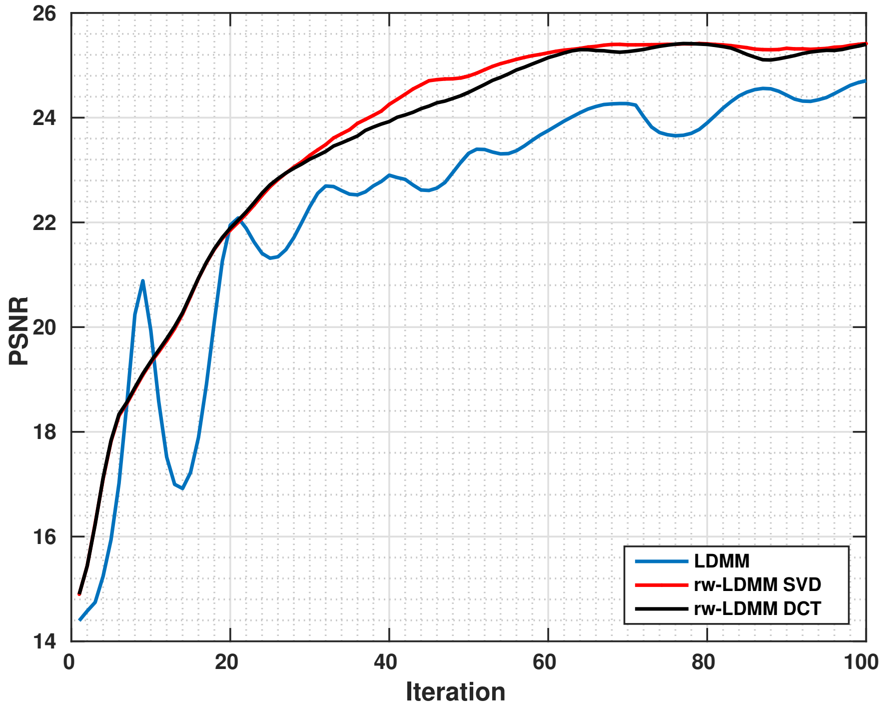

Figure 10. *PSNR of the reconstructed* 256 × 256 BARBARA *image (see Figure 9) at each iteration of LDMM, rw-LDMM with SVD, and rw-LDMM with DCT. After the* 20*th iteration, both rw-LDMM algorithms always achieve higher PSNR than the original LDMM.*

of ALOHA thus builds upon performing matrix completion for such block Hankel matrices, each constructed from an image patch. This is fundamentally different from the construction of convolution framelets, which relies on the the full patch matrix consisting of all patches in the same image. Further examination shows that each block Hankel matrix can be viewed as a submatrix of the full patch matrix up to row and column permutations. Therefore, ALOHA explores substructures in the full patch matrix and thus enforces a stronger regularization on the local consistency of neighboring patches. On the other hand, it does not incorporate similarity between nonlocal patches in an image as convolution framelets does. For some test images with strong texture patterns (e.g., BARBARA, FINGERPRINT, CHECKERBOARD, SWIRL), restoration from 10% random subsamples by ALOHA reaches higher PSNR than LDMM and rw-LDMM; see M109144_01.pdf [local/web 5.79MB] for more details. However, we observe that the reconstruction by ALOHA sometimes contains artifacts that are not present in those obtained by rw-LDMM and LDMM, even though the ALOHA results can have higher PSNR (see, e.g., Figure 11 and Figure 12). Intuitively, this effect suggests different inpainting mechanisms underlying LDMM/rw-LDMM and ALOHA: LDMM and rw-LDMM, as indicated in [45], “spread out” the retained subsamples to missing pixels, whereas ALOHA exploits the intrinsic (rotationally invariant) low-rank property of the block Hankel structure for each image patch. Numerical results with critically low subsample rates (2% and 5%) are in accordance with this intuition; see Figure 13, Figure 14, and more examples in M109144_01.pdf

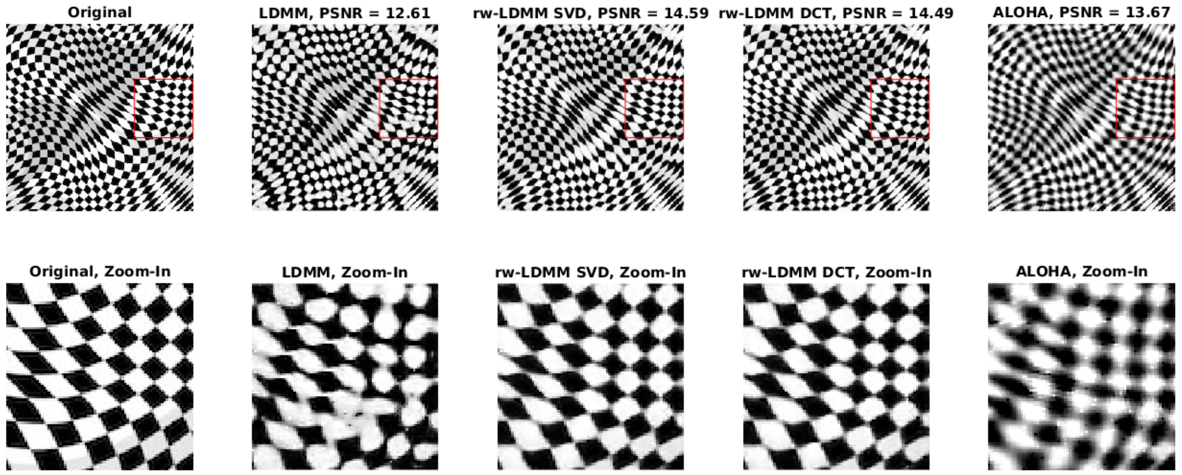

Figure 11. *Reconstructed* 256×256 CHECKERBOARD *images from* 10% *randomly subsampled pixels using LDMM, rw-LDMM, and ALOHA. Top: Restored images. Bottom: Zoom-in views of the* 80×80 *blocks enclosed by red boxes. Compared with LDMM and ALOHA, the proposed rw-LDMM reconstructs images with higher PSNR and fewer visual artifacts.*

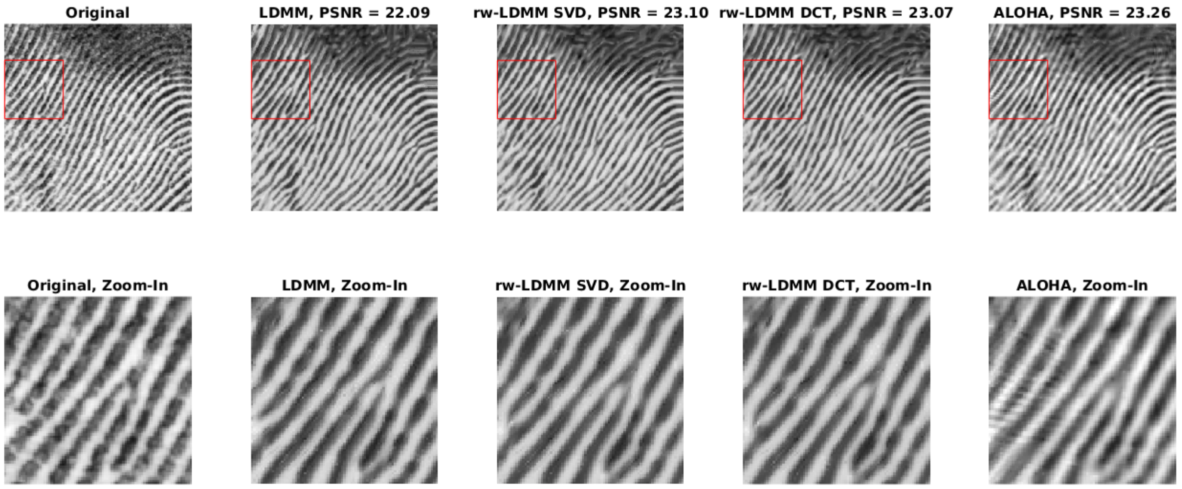

Figure 12. *Reconstructed* 256×256 FINGERPRINT *images from* 10% *randomly subsampled pixels using LDMM, rw-LDMM, and ALOHA. Top: Restored images. Bottom: Zoom-in views of the* 80×80 *blocks enclosed by red boxes. Compared with LDMM and ALOHA, the proposed rw-LDMM reconstructs images with comparable or higher PSNR and fewer visual artifacts.*

[local/web 5.79MB].

5.3. Other image processing applications. In this subsection, we compare results of LDMM and rw-LDMM applied to other image processing tasks.

We first test LDMM, rw-LDMM, and ALOHA on several different types of image inpainting tasks, beyond the regime of inpainting from randomly sampled pixel values in subsection 5.2. Figure 15 illustrates the reconstructed BOAT image from overlaying texts, scratch-

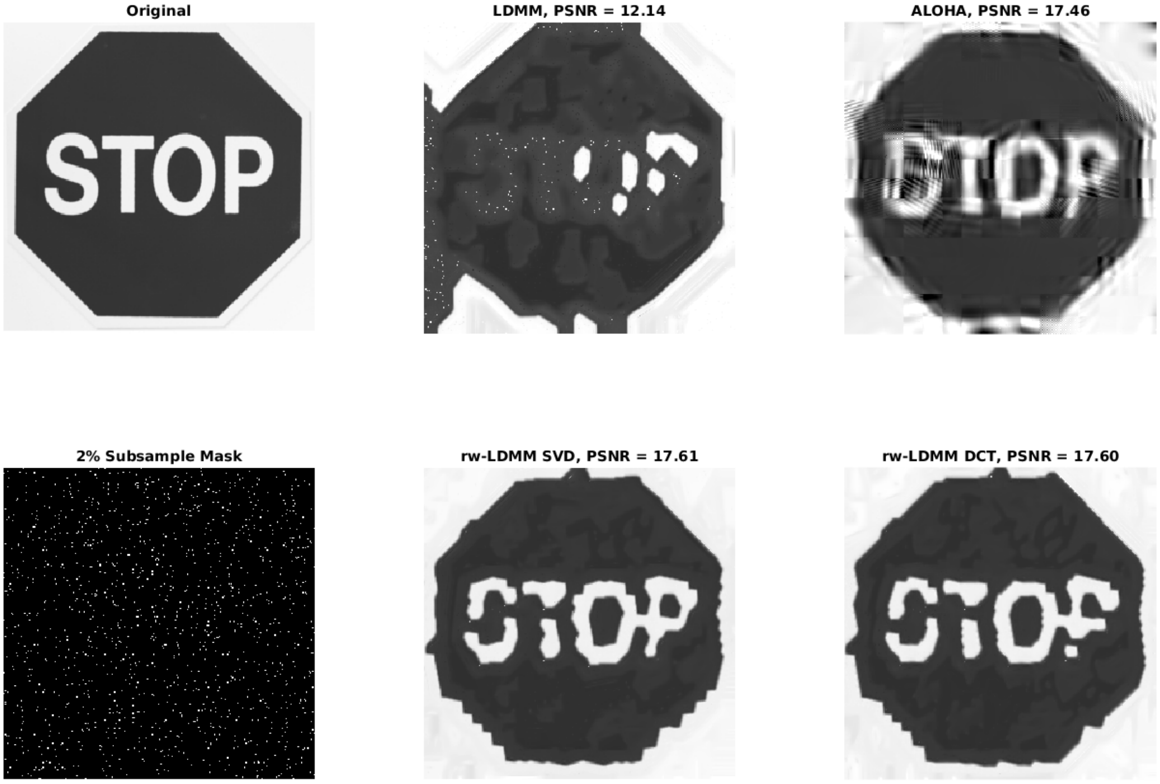

Figure 13. *Reconstructions of a* 256×256 *traffic sign from* 2% *randomly subsampled pixels using LDMM, rw-LDMM, and ALOHA. Both ALOHA and rw-LDMM restore legible letters even under such a critically low subsample rate; rw-LDMM methods also achieve higher PSNR with fewer visual artifacts.*

type damages, and a drip-painting-like corruption in the style of Jackson Pollock. Both LDMM and rw-LDMM slightly outperform ALOHA in these inpainting experiments with nonrandomly subsampled pixels.

We then apply LDMM and rw-LDMM to the task of virtually removing cracks in paint surface present in digital images of real art paintings. The top left panel of Figure 16 is a 256×256 subimage downsampled to half-size from a high-resolution digital photo[19] of the painting *St. John the Evangelist Reproving the Philosopher Crato* (circa. 1370–1380) by Francescuccio Ghissi. The mask of pixels identified as cracks is provided by art conservators from the North Carolina Museum of Art (NCMA); see the top center panel in Figure 16. Since no "crack-free" image is available as ground truth, we compare the LDMM and rw-LDMM results with a reference inpainting result (the top right panel in Figure 16) used by NCMA art conservators in the *Reunited: Francescuccio Ghissi's St. John Altarpiece* exhibition.[20] As demonstrated in the bottom row of Figure 16, all LDMM and rw-LDMM inpainting algorithms recover visually comparable crack-free images to the reference result.

[19]The Ghissi image appears here with the permission of the North Carolina Museum of Art (NCMA).
[20]https://dukeipai.org/projects/ghissi/

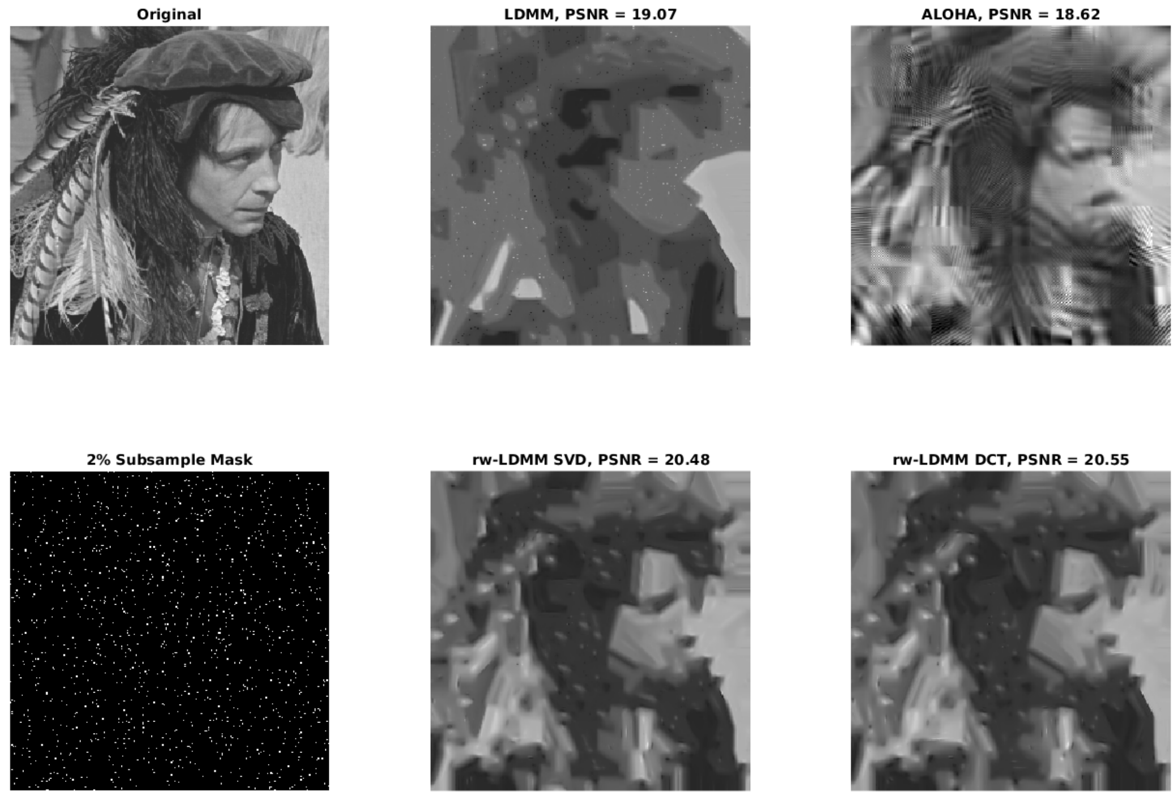

Figure 14. *Reconstructed* 256×256 MAN *images from* 2% *randomly subsampled pixels using LDMM, rw-LDMM, and ALOHA. The proposed rw-LDMM methods restore recognizable human shapes as well as color patterns on the hair and the hat decoration even under such a critically low subsample rate.*

Finally, we compare the denoising performance of LDMM and rw-LDMM on the SWIRL image; Figure 17 demonstrates that LDMM-type algorithms can effectively estimate the original image from contaminations with moderate levels of noise, but their performances are still not compatible to BM3D [16], which is the current state of the art in image denoising.

6. Conclusion and future work. In this paper, we present convolution framelets, a patch-based representation that combines local and nonlocal bases for image processing. We show the energy compaction property of these convolution framelets in a linear reconstruction framework motivated by nonlinear dimension reduction; i.e., the L^2-energy of a signal concentrates on the upper left block of the coefficient matrix with respect to convolution framelets. This energy concentration property is exploited to improve LDMM by incorporating "near-optimal" local patch bases into the regularization mechanism for the purpose of strengthening the energy concentration pattern. Numerical experiments suggest that the proposed reweighted LDMM algorithm performs better than the original LDMM in inpainting problems, especially for images containing high-contrast nonregular textures.

One avenue we would like to explore in future work is comparing the ℓ_2-regularization with other regularization frameworks. In fact, our numerical experiments suggest that nonlinear

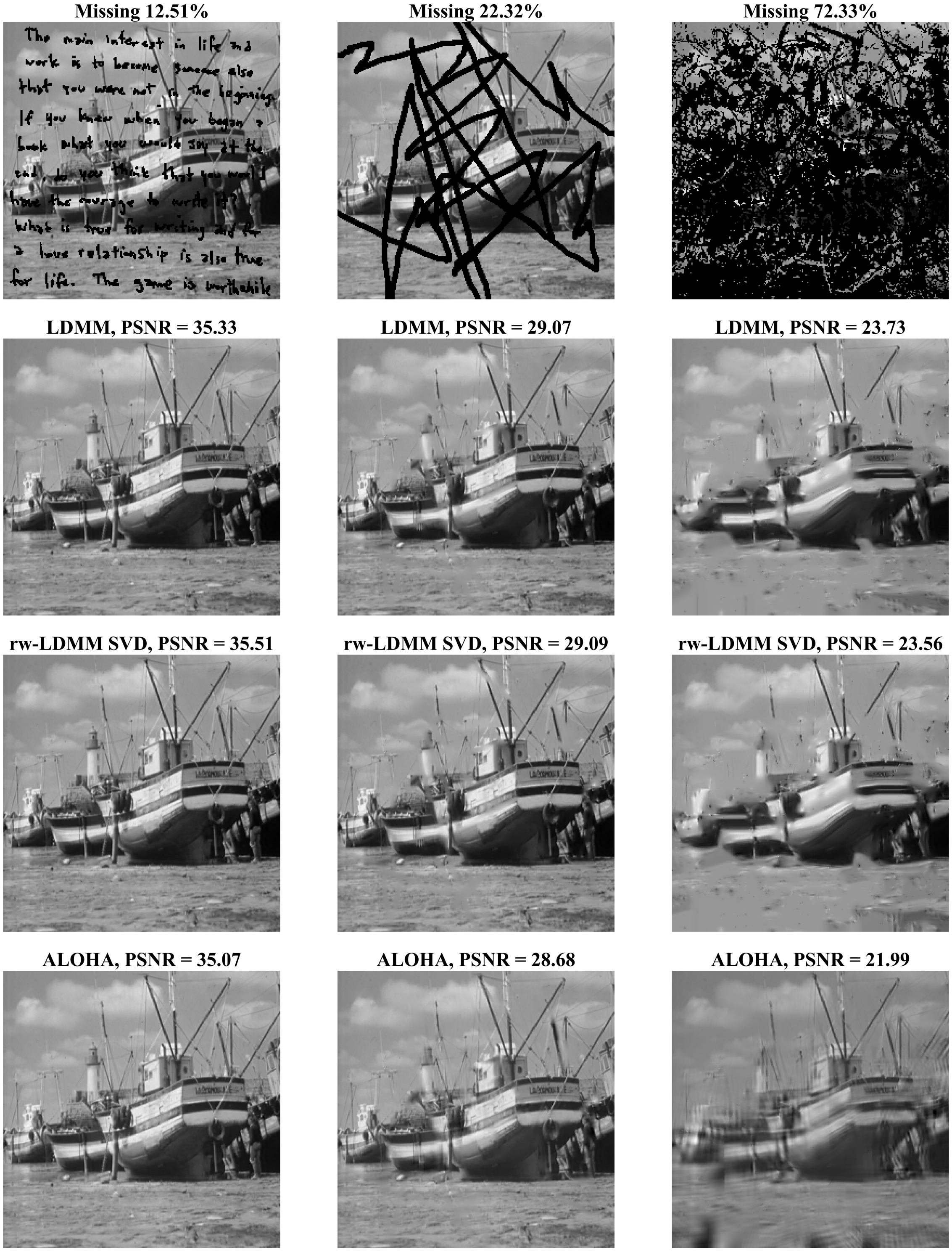

Figure 15. *Reconstructed* 256×256 BOAT *images from several types of nonrandom subsamplings. Columns from left to right: Removing overlaying texts; removing scratches; removing a drip-painting-like corruption (mask extracted from Jackson Pollock's* Number 32, 1950, *with reverted black and white pixel values).*

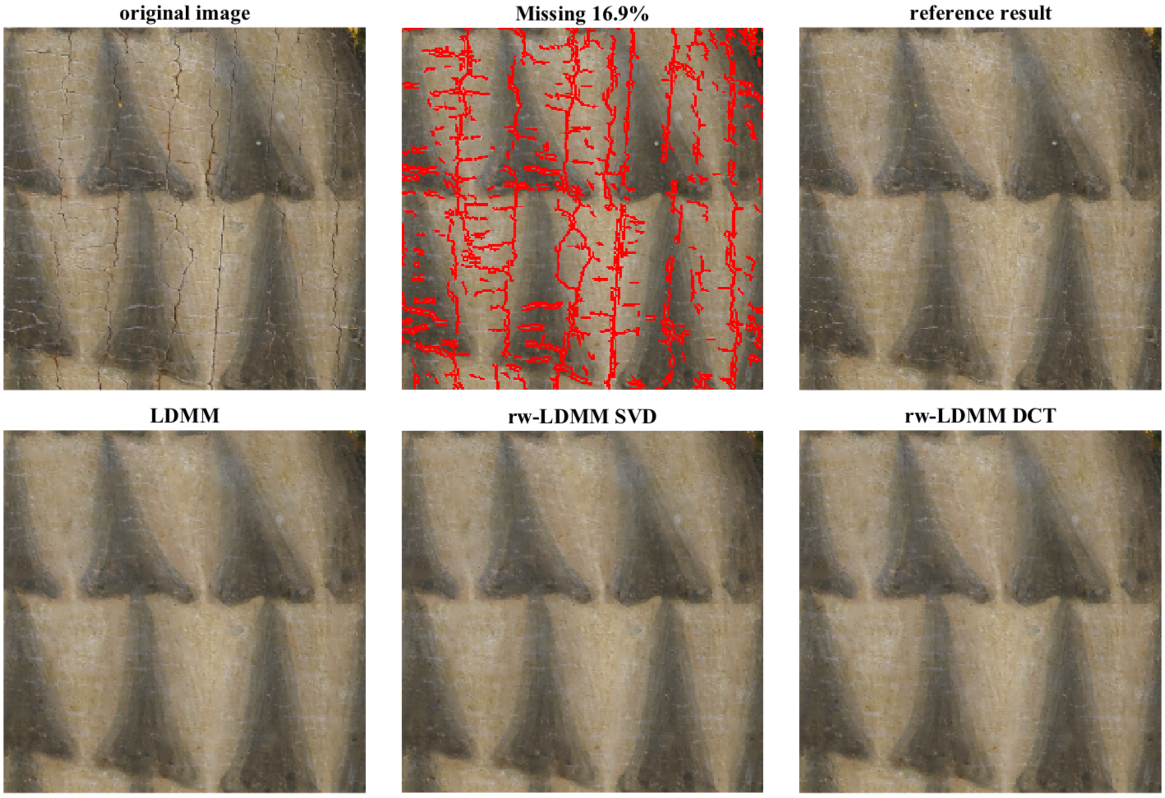

Figure 16. *Crack removal from high-resolution digitization of a real painting. Top left: A* 256×256 *detail from* St. John the Evangelist Reproving the Philosopher Crato *(circa.* 1370–1380*), by Francescuccio Ghissi, in the North Carolina Museum of Art (image appears here with museum permission). Top center: Input mask of the pixels to be inpainted (highlighted in red), corresponding to the "dark crack pixels." Top right: Reference result used by art conservators, based on a different nonlocal patch-based image inpainting algorithm* [43]. *Bottom: Inpainting results using LDMM and rw-LDMM.*

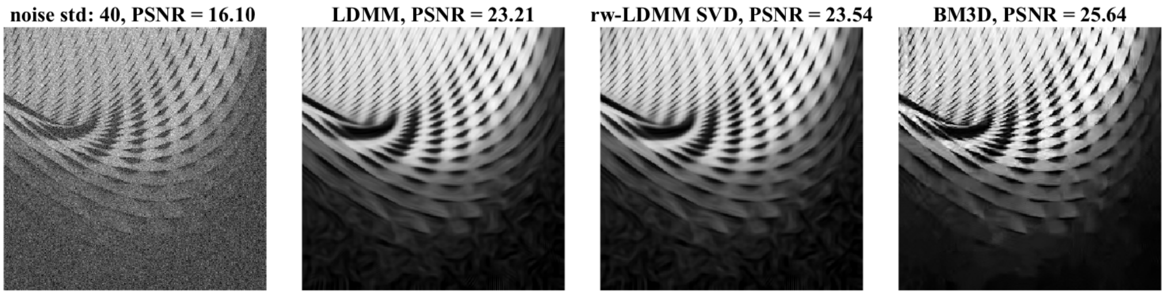

Figure 17. *Denoised* 256×256 SWIRL *images using LDMM and rw-LDMM.*

approximation of signals with convolution framelets could outperform linear approximation; hence regularization techniques based on ℓ_1- and ℓ_0-norms have the potential to further improve the reconstruction performance. Furthermore, although we established an energy concentration guarantee in subsection 3.3, it remains unclear in concrete scenarios which local patch basis exactly attains the optimality condition in Proposition 2. We made a first attempt in this direction for specific linear embedding in subsection 3.3, but similar results for nonlinear embeddings, as well as further extensions of the framework to unions of local embeddings (which we expect will also provide insights for other nonlocal transform-domain techniques, including BM3D), are also of great interest.

Another direction we intend to explore is the influence of the patch size ℓ. Throughout this work, as well as in most patch-based image processing algorithms, the patch size is a hyperparameter to be chosen empirically and fixed; however, historically neuroscience experiments [62] and fractal image compression techniques [4, 28] provide evidence for the importance of perceiving patches of different sizes simultaneously in the same image. Since patch matrices corresponding to varying patch sizes of the same image are readily available, we can potentially combine convolution framelets across different scales to build multiresolution convolution framelets.

From an application perspective, we are interested in investigating LDMM-type algorithms for a wider range of image processing tasks. Though we demonstrated in subsection 5.3 the capability of LDMM and rw-LDMM for several different types of image inpainting tasks beyond the regime of random subsamples, we were unable to extend the algorithmic framework to achieve satisfactory performance on object removal tasks (which amounts to completely removing undesirable objects from an image and then filling up the "large hole" in a visually unconscious manner). One plausible explanation for this limitation is that larger patches are preferable for inpainting sizable missing objects, but the low-dimensional manifold assumption is practical only for small to medium-sized patches; for instance, the impressive object removal performance of ALOHA reported in [29] relies on handling patches of size 120-by-120 (or even larger). Studying the theoretical and practical applicability of the low dimensional assumption of patch manifolds is yet another intriguing future direction.

Appendix A. Convolution framelets in 2D. We briefly explain how the 1D theory established in section 2 easily carries through to the 2D case, which is of central interest in image processing. For simplicity, let f be a 2D signal defined on a square lattice $\{(i_1, i_2) \mid 1 \leq i_1, i_2 \leq N\}$, and consider square patches of size $\ell \times \ell$, where $1 \leq \ell \leq N$; essentially the same argument applies to rectangular lattices and patches. Let $\{v_{\ell_1,\ell_2} \mid 1 \leq \ell_1, \ell_2 \leq \ell\}$ be an orthonormal local basis supported inside the square sublattice $\{(\ell_1, \ell_2) \mid 1 \leq \ell_1, \ell_2 \leq \ell\}$, and let $\{\phi_{j_1,j_2} \mid 1 \leq j_1, j_2 \leq N\}$ be an orthonormal global basis. We assume f extends by periodicity to the infinite 2D integer grid. Note that

$$
\begin{aligned}
& f\left[i_1+\ell_1, i_2+\ell_2\right] = \sum_{j_1,j_2=1}^{N} \sum_{\iota_1,\iota_2=1}^{N} f\left[\iota_1+\ell_1, \iota_2+\ell_2\right] \phi_{j_1,j_2}\left[\iota_1, \iota_2\right] \phi_{j_1,j_2}\left[i_1, i_2\right] \\
&= \sum_{j_1,j_2=1}^{N} \sum_{\iota_1,\iota_2=1}^{N} \sum_{k_1,k_2=1}^{\ell} \sum_{\kappa_1,\kappa_2=1}^{N} f\left[\iota_1+\kappa_1, \iota_2+\kappa_2\right] v_{k_1,k_2}\left[\kappa_1, \kappa_2\right] v_{k_1,k_2}\left[\ell_1, \ell_2\right] \phi_{j_1,j_2}\left[\iota_1, \iota_2\right] \phi_{j_1,j_2}\left[i_1, i_2\right] \\
&= \sum_{j_1,j_2=1}^{N} \sum_{k_1,k_2=1}^{\ell} C_{(j_1,j_2),(k_1,k_2)}\ \phi_{j_1,j_2}\left[i_1, i_2\right] v_{k_1,k_2}\left[\ell_1, \ell_2\right],
\end{aligned}
$$

where

$$\begin{aligned}C_{(j_1,j_2),(k_1,k_2)} &= \sum_{\iota_1,\iota_2=1}^{N} \sum_{\kappa_1,\kappa_2=1}^{N} f\left[\iota_1+\kappa_1,\iota_2+\kappa_2\right] v_{k_1,k_2}\left[\kappa_1,\kappa_2\right] \phi_{j_1,j_2}\left[\iota_1,\iota_2\right] \\ &= \sum_{m_1,m_2=1}^{N} f\left[m_1,m_2\right] v_{k_1,k_2} * \phi_{j_1,j_2}\left[m_1,m_2\right] =: \left\langle f, \phi_{j_1,j_2} * v_{k_1,k_2}\right\rangle .\end{aligned}$$

Moreover, for any (I_1, I_2) in the $N \times N$ square lattice,

$$\begin{aligned}f\left[I_1,I_2\right] &= \frac{1}{\ell^2} \sum_{i_1+\ell_1=I_1} \sum_{i_2+\ell_2=I_2} f\left[i_1+\ell_1,i_2+\ell_2\right] \\ &= \frac{1}{\ell^2} \sum_{j_1,j_2=1}^{N} \sum_{k_1,k_2=1}^{\ell} \left\langle f,\phi_{j_1,j_2} * v_{k_1,k_2}\right\rangle \sum_{i_1+\ell_1=I_1} \sum_{i_2+\ell_2=I_2} \phi_{j_1,j_2}\left[i_1,i_2\right] v_{k_1,k_2}\left[\ell_1,\ell_2\right] \\ &= \frac{1}{\ell^2} \sum_{j_1,j_2=1}^{N} \sum_{k_1,k_2=1}^{\ell} \left\langle f,\phi_{j_1,j_2} * v_{k_1,k_2}\right\rangle \left(\phi_{j_1,j_2} * v_{k_1,k_2}\right)\left[I_1,I_2\right],\end{aligned}$$

which is exactly the 2D analogy of (6).

Appendix B. Proof of Proposition 1.

Lemma 7. *Let $\widetilde{V} \in \mathbb{R}^{l\times p}$, s.t. $\widetilde{V}\,\widetilde{V}^\top = I_\ell$; then for all $f \in \mathbb{R}^N$, $N \geq \ell$,*

$$f = \frac{1}{\ell} \sum_{i=1}^{p} f * \widetilde{v}_i * \widetilde{v}_i(-\cdot).$$

Proof of Lemma 7. By definition

$$\widetilde{v}_i * \widetilde{v}_i(-\cdot)[n] = \sum_{m=0}^{N-1} \widetilde{v}_i[n-m]\widetilde{v}_i[-m] = \sum_{m'=0}^{l-1} \widetilde{v}_i[n+m']\widetilde{v}_i[m'],$$

since $\widetilde{v}_i[m] = 0$ for all $\ell \leq m \leq N$, $i = 1, \ldots, p$. Therefore,

$$\sum_{i=1}^{p} \widetilde{v}_i * \widetilde{v}_i(-\cdot)[n] = \sum_{i=1}^{p} \sum_{m=0}^{\ell-1} \widetilde{v}_i[n+m]\widetilde{v}_i[m],$$

and if we change the order of summation, we have $\sum_{i=1}^{p} \widetilde{v}_i[m+n]\widetilde{v}_i[m] = \delta(n)$, which follows from $\widetilde{V}\widetilde{V}^\top = I_\ell$. In sum, $\sum_{i=1}^{p} \widetilde{v}_i * \widetilde{v}_i(-\cdot)[n] = \ell \cdot \delta(n)$; hence $f = \frac{1}{\ell}\sum_{i=1}^{p} f * \widetilde{v}_i * \widetilde{v}_i(-\cdot)$. ■

Proof of Proposition 1. By Lemma 7,

$$\begin{aligned}f &= \frac{1}{m} \sum_{i=1}^{m'} f * v_i^S * v_i^S(-\,\cdot) = \frac{1}{m} \sum_{i=1}^{m'} \left(\sum_{j=1}^{n'} \langle f * v_i^S(-\cdot),\, v_j^L\rangle\, v_j^L \right) * v_i^S \\ &= \sum_{i,j} \left\langle f,\, \frac{1}{\sqrt{m}}\, v_j^L * v_i^S \right\rangle \frac{1}{\sqrt{m}}\, v_j^L * v_i^S \doteq \sum_{i,j} c_{ij}\, \psi_{ij}, \quad \text{where } \psi_{ij} = \frac{1}{\sqrt{m}}\, v_j^L * v_i^S. \qquad ■\end{aligned}$$

Appendix C. A simplified proof of the dimension identity (23).

Proposition 8. *Assume a d-dimensional Riemannian manifold $\mathcal{M}$ is isometrically embedded into $\mathbb{R}^\ell$, with coordinate functions $\{\alpha_j \mid 1 \le j \le \ell\}$. Then at any point $x \in \mathcal{M}$*

$$d = \dim(\mathcal{M}) = \sum_{j=1}^{\ell} |\nabla_{\mathcal{M}} \alpha_j(x)|^2,$$

where $\nabla_{\mathcal{M}} : C^\infty(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ is the gradient operator on $\mathcal{M}$.

Proof. Let $\nabla : C^\infty(\mathbb{R}^\ell) \to \mathfrak{X}(\mathcal{M})$ be the gradient operator on $\mathbb{R}^\ell$. For any $f \in C^\infty(M)$, if f is the restriction to $\mathcal{M}$ of a smooth function $\bar{f} \in C^\infty(\mathbb{R}^\ell)$, then $\nabla_{\mathcal{M}} f(x)$ is the projection of $\nabla \bar{f}$ to $T_x M$, the tangent space of $\mathcal{M}$ at $x \in \mathcal{M}$. Now, fix an arbitrary point $x \in \mathcal{M}$, and let $E_1(x), \ldots, E_d(x)$ be an orthonormal basis for $T_x\mathcal{M}$. We have for any $1 \le j \le \ell$

$$\nabla_{\mathcal{M}} \alpha_j(x) = \sum_{k=1}^{d} \langle \nabla \alpha_j(x), E_k(x) \rangle E_k(x),$$

and thus

$$|\nabla_{\mathcal{M}} \alpha_j(x)|^2 = \sum_{k=1}^{d} |\langle \nabla \alpha_j(x), E_k(x) \rangle|^2.$$

Note that $\nabla \alpha_j$ is a constant vector in $\mathbb{R}^\ell$ with 1 at the jth entry and 0 elsewhere. Consequently, inner product $\langle \nabla \alpha_j(x), E_k(x) \rangle$ simply picks out the jth coordinate of $E_k(x)$. Therefore,

$$\sum_{j=1}^{\ell} |\nabla_{\mathcal{M}} \alpha_j(x)|^2 = \sum_{j=1}^{\ell} \sum_{k=1}^{d} |\langle \nabla \alpha_j(x), E_k(x) \rangle|^2 = \sum_{k=1}^{d} \left(\sum_{j=1}^{\ell} |\langle \nabla \alpha_j(x), E_k(x) \rangle|^2 \right)$$
$$= \sum_{k=1}^{d} |E_k(x)|^2 = \sum_{k=1}^{d} 1 = d,$$

which completes the proof. ■

Appendix D. Proof of the optimality and the sparsity of SVD local basis with respect to MDS nonlocal basis.

Proposition 9. *Let $HX = U_X \Sigma_X V_X^\top$ be the reduced SVD of the centered data matrix $X \in \mathbb{R}^{N \times \ell}$, where $H = I_N - \frac{1}{N} \mathbb{1}_N \mathbb{1}_N^\top$ is the centering matrix. The optimal V for (12) is exactly V_X; the corresponding matrix basis has the sparsest representation of X.*

Proof. Without loss of generality, assume $X = HX$. In MDS, $\Phi_{\mathcal{E}} = U_X$ with $p = \ell - 1$. The entries of the coefficient matrix $C = \Phi^\top X \widetilde{V}$ can be explicitly computed as

$$C_{ij} = \phi_i^\top X \, \tilde{v}_j = u_{X,i}^\top X \, \tilde{v}_j = u_{X,i}^\top U_X \, \Sigma_X \, V_X^\top \, \tilde{v}_j = \sigma_{X,i} \, v_{X,i}^\top \, \tilde{v}_j,$$

where $u_{X,i}$, $\tilde{v}_j$ are the columns of U_X and $\widetilde{V}$, respectively, and $\sigma_{X,i}$ is the ith diagonal entry of Σ_X. According to (16), the optimal $\widetilde{V}$ should satisfy $v_{X,i}^\top \tilde{v}_j = 0$ for all $i > j$, which is achieved by setting $\tilde{v}_i = v_{X,i}$. Moreover, since $\mathrm{rank}(C) = \ell - 1$, C has at least $(\ell - 1)$ nonzero entries; it follows from $v_{X,i}^\top v_{X,j} = \delta_{i,j}$ that $C = U_X^\top X V_X$ has exactly $(\ell - 1)$ nonzero entries and is thus the sparsest representation. ■

Acknowledgment. We thank the authors of LDMM [45] for providing us with their code.

REFERENCES

[1] Y. Aflalo, H. Brezis, and R. Kimmel, *On the optimality of shape and data representation in the spectral domain*, SIAM J. Imaging Sci., 8 (2015), pp. 1141–1160, https://doi.org/10.1137/140977680.

[2] M. Aharon, M. Elad, and A. M. Bruckstein, *On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them*, Linear Algebra Appl., 416 (2006), pp. 48–67.

[3] N. Ahmed, T. Natarajan, and K. R. Rao, *Discrete cosine transform*, IEEE Trans. Comput., C-23 (1974), pp. 90–93, https://doi.org/10.1109/T-C.1974.223784.

[4] M. F. Barnsley and A. D. Sloan, *Methods and Apparatus for Image Compression by Iterated Function System*, US Patent 4,941,193, July 10, 1990.

[5] M. Belkin and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Comput., 15 (2003), pp. 1373–1396.

[6] M. Belkin and P. Niyogi, *Convergence of Laplacian eigenmaps*, Adv. Neural Inform. Process. Syst., 19 (2007), p. 129.

[7] A. Buades, B. Coll, and J.-M. Morel, *A non-local algorithm for image denoising*, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, IEEE, Washington, DC, 2005, pp. 60–65.

[8] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530, https://doi.org/10.1137/040616024.

[9] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian, *On the local behavior of spaces of natural images*, Int. J. Comput. Vision, 76 (2008), pp. 1–12.

[10] P. Chatterjee and P. Milanfar, *Clustering-based denoising with locally learned dictionaries*, IEEE Trans. Image Process., 18 (2009), pp. 1438–1451.

[11] P. Chatterjee and P. Milanfar, *Patch-based near-optimal image denoising*, IEEE Trans. Image Process., 21 (2012), pp. 1635–1649.

[12] A. Cohen and J.-P. D'Ales, *Nonlinear approximation of random functions*, SIAM J. Appl. Math., 57 (1997), pp. 518–540, https://doi.org/10.1137/S0036139994279153.

[13] R. R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 5–30.

[14] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 7426–7431, https://doi.org/10.1073/pnas.0500334102.

[15] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 7432–7437, https://doi.org/10.1073/pnas.0500896102.

[16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse 3-D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16 (2007), pp. 2080–2095.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *BM3D image denoising with shape-adaptive principal component analysis*, in SPARS'09, Signal Processing with Adaptive Sparse Structured Representations, Saint Malo, France, 2009.

[18] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 61, SIAM, Philadelphia, 1992.

[19] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, *Wavelet shrinkage: Asymptopia?*, J. Roy. Statist. Soc. Ser. B, 57 (1995), pp. 301–369.

[20] D. L. Donoho and J. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455, https://doi.org/10.1093/biomet/81.3.425.

[21] M. Elad and M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process., 15 (2006), pp. 3736–3745.

[22] K. Engan, S. O. Aase, and J. Hakon Husoy, *Method of optimal directions for frame design*, in Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, IEEE, Washington, DC, 1999, pp. 2443–2446.
[23] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, Multiscale Model. Simul., 6 (2007), pp. 595–630, https://doi.org/10.1137/060669358.
[24] G. Gilboa and S. Osher, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028, https://doi.org/10.1137/070698592.
[25] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288, https://doi.org/10.1137/090771806.
[26] G. E. Hinton and R. R. Salakhutdinov, *Reducing the dimensionality of data with neural networks*, Science, 313 (2006), pp. 504–507, https://doi.org/10.1126/science.1127647.
[27] R. Izbicki and A. B. Lee, *Nonparametric conditional density estimation in a high-dimensional regression setting*, J. Comput. Graph. Statist., 25 (2016), pp. 1297–1316, https://doi.org/10.1080/10618600.2015.1094393.
[28] A. E. Jacquin, *Image coding based on a fractal theory of iterated contractive image transformations*, IEEE Trans. Image Process., 1 (1992), pp. 18–30, https://doi.org/10.1109/83.128028.
[29] K. H. Jin and J. C. Ye, *Annihilating filter-based low-rank Hankel matrix approach for image inpainting*, IEEE Trans. Image Process., 24 (2015), pp. 3498–3511.
[30] A. Kheradmand and P. Milanfar, *A general framework for regularized, similarity-based image restoration*, IEEE Trans. Image Process., 23 (2014), pp. 5136–5151.
[31] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. S. Lee, and T. J. Sejnowski, *Dictionary learning algorithms for sparse representation*, Neural Comput., 15 (2003), pp. 349–396.
[32] K. Kreutz-Delgado and B. D. Rao, *FOCUSS-based dictionary learning algorithms*, in International Symposium on Optical Science and Technology, International Society for Optics and Photonics, Bellingham, WA, 2000, pp. 459–473.
[33] S. S. Lafon, *Diffusion Maps and Geometric Harmonics*, Ph.D. thesis, Yale University, New Haven, CT, 2004.
[34] A. B. Lee and R. Izbicki, *A spectral series approach to high-dimensional nonparametric regression*, Electron. J. Stat., 10 (2016), pp. 423–463.
[35] A. B. Lee, K. S. Pedersen, and D. Mumford, *The nonlinear statistics of high-contrast patches in natural images*, Internat. J. Comput. Vision, 54 (2003), pp. 83–103.
[36] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*, Springer, New York, 2007.
[37] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, *Learning unions of orthonormal bases with thresholded singular value decomposition*, in Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, Washington, DC, 2005, pp. v–293.
[38] Z. Li, Z. Shi, and J. Sun, *Point Integral Method for Solving Poisson-Type Equations on Manifolds from Point Clouds with Convergence Guarantees*, preprint, https://arxiv.org/abs/1409.2623, 2014.
[39] A. V. Little, M. Maggioni, and L. Rosasco, *Multiscale geometric methods for data sets* I: *Multiscale SVD, noise and curvature*, Appl. Comput. Harmon. Anal., to appear, https://doi.org/10.1016/j.acha.2015.09.009.
[40] L. v. d. Maaten and G. Hinton, *Visualizing data using t-SNE*, J. Mach. Learn. Res., 9 (2008), pp. 2579–2605.
[41] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, *Non-local sparse models for image restoration*, in Proceedings of the 12th IEEE International Conference on Computer Vision, IEEE, Washington, DC, 2009, pp. 2272–2279.
[42] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, New York, 2008.
[43] A. Newson, A. Almansa, Y. Gousseau, and P. Pérez, *Non-local patch-based image inpainting*, Image Processing On Line (IPOL), preprint, 2015.
[44] B. A. Olshausen and D. J. Field, *Sparse coding with an overcomplete basis set: A strategy employed by V1?*, Vision Res., 37 (1997), pp. 3311–3325.
[45] S. Osher, Z. Shi, and W. Zhu, *Low Dimensional Manifold Model for Image Processing*, Tech. Rep., CAM report 16-04, UCLA, Los Angeles, CA, 2016.

[46] K. Pearson, *On lines and planes of closest fit to systems of point in space*, Philos. Mag., 2 (1901), pp. 559–572.

[47] J. A. Perea and G. Carlsson, *A Klein-bottle-based dictionary for texture representation*, Internat. J. Comput. Vision, 107 (2014), pp. 75–97.

[48] G. Peyré, *Image processing with nonlocal spectral bases*, Multiscale Model. Simul., 7 (2008), pp. 703–730, https://doi.org/10.1137/07068881X.

[49] G. Peyré, *A review of adaptive image representations*, IEEE J. Sel. Topics Signal Process., 5 (2011), pp. 896–911.

[50] X. Qi, *Vector Nonlocal Mean Filter*, Ph.D. thesis, University of Toronto, Toronto, Canada, 2015.

[51] L. Rosasco, *Data Representation: From Signal Processing to Machine Learning*, Mini-Tutorial, SIAM Conference on Image Science, Albuquerque, NM, 2016, http://meetings.siam.org/sess/dsp_programsess.cfm?SESSIONCODE=22699&_ga=1.165936247.87958178.1459945433.

[52] S. T. Roweis and L. K. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science, 290 (2000), pp. 2323–2326, https://doi.org/10.1126/science.290.5500.2323.

[53] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[54] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, 2002.

[55] R. N. Shepard, *The analysis of proximities: Multidimensional scaling with an unknown distance function.* I, Psychometrika, 27 (1962), pp. 125–140.

[56] Z. Shi, S. Osher, and W. Zhu, *Low Dimensional Manifold Model with Semi-local Patches*, Technical Report, CAM report 16-63, UCLA, Los Angeles, CA, 2016.

[57] Z. Shi and J. Sun, *Convergence of the Point Integral Method for the Poisson Equation with Dirichlet Boundary on Point Cloud*, preprint, https://arxiv.org/abs/1312.4424, 2013.

[58] Z. Shi and J. Sun, *Convergence of the Point Integral Method for Poisson Equation on Point Cloud*, preprint, https://arxiv.org/abs/1403.2141, 2014.

[59] A. Singer, Y. Shkolnisky, and B. Nadler, *Diffusion interpretation of nonlocal neighborhood filters for signal denoising*, SIAM J. Imaging Sci., 2 (2009), pp. 118–139, https://doi.org/10.1137/070712146.

[60] A. Singer and H.-t. Wu, *Spectral Convergence of the Connection Laplacian from Random Samples*, preprint, https://arxiv.org/abs/1306.1587, 2013.

[61] A. Skodras, C. Christopoulos, and T. Ebrahimi, *The JPEG* 2000 *still image compression standard*, IEEE Signal Process. Mag., 18 (2001), pp. 36–58.

[62] S. Smale, T. Poggio, A. Caponnetto, and J. Bouvrie, *Derived distance: Towards a mathematical theory of visual cortex*, Artificial Intelligence (2007).

[63] A. D. Szlam, M. Maggioni, and R. R. Coifman, *Regularization on graphs with function-adapted diffusion processes*, J. Mach. Learn. Res., 9 (2008), pp. 1711–1739.

[64] H. Talebi and P. Milanfar, *Global image denoising*, IEEE Trans. Image Process., 23 (2014), pp. 755–768.

[65] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2323, https://doi.org/10.1126/science.290.5500.2319.

[66] W. S. Torgerson, *Multidimensional scaling:* I. *Theory and method*, Psychometrika, 17 (1952), pp. 401–419.

[67] G. K. Wallace, *The JPEG still picture compression standard*, IEEE Trans. Consumer Electron., 38 (1992), pp. xviii–xxxiv.