Image Processing with Non-local Spectral Bases

Gabriel Peyré

HAL Id: hal-00359721

https://hal.science/hal-00359721v1

Submitted on 9 Feb 2009

IMAGE PROCESSING
WITH NON-LOCAL SPECTRAL BASES

GABRIEL PEYRÉ*

Abstract. This article studies regularization schemes that are defined using a lifting of the image pixels in a high dimensional space. For some specific classes of geometric images, this discrete set of points is sampled along a low dimensional smooth manifold. The construction of differential operators on this lifted space allows one to compute PDE flows and perform variational optimizations. All these schemes lead to regularizations that exploit the manifold structure of the lifted image. Depending on the specific definition of the lifting, one recovers several well-known semi-local and non-local denoising algorithms that can be interpreted as local estimators over a semi-local or a non-local manifold. This framework also allows one to define thresholding operators in adapted orthogonal bases. These bases are eigenvectors of the discrete Laplacian on a manifold adapted to the geometry of the image. Numerical results compare the efficiency of PDE flows, energy minimizations and thresholdings in the semi-local and non-local settings. The superiority of the non-local computations is studied through the performance of non-linear approximation in orthogonal bases.

Key words. Image processing, manifold, non-local denoising, adapted bases.

AMS subject classifications. 41A25, 42C40, 65T60

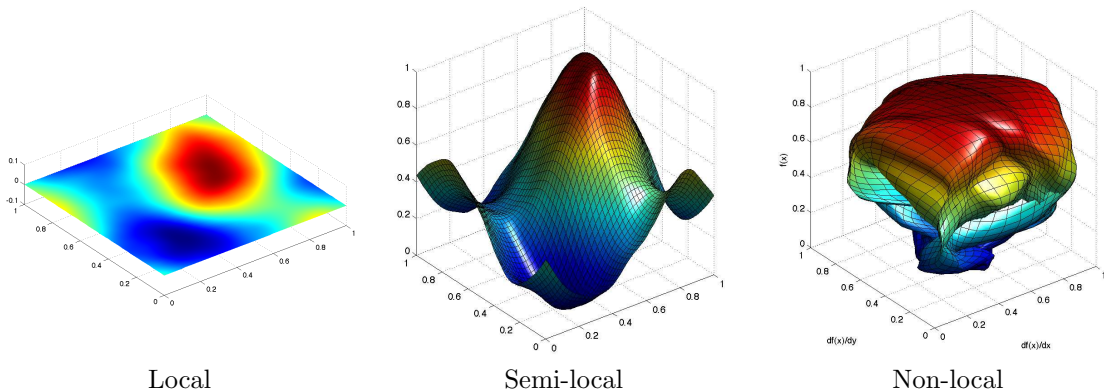|		
Local	Semi-local	Non-local

FIG. 0.1. *Representations of the discrete manifold $\mathcal{M}$ for the three computation modes for a regular image f. The color at each point $p \in \mathcal{M}$ is set according to $\tilde{f}(p)$. Left: local manifold $x \mapsto f(x)$, center: semi-local manifold $x \mapsto (x, \lambda f(x))$, right: non-local manifold $x \mapsto p_x(f)$. The non-local manifold is a subset of $\mathbb{R}^{\tau^2}$ where τ is the width of the patches $p_x(f)$, but it can be approximately embedded in $\mathbb{R}^3$ by a dimension reduction using principal components analysis.*

In order to capture the complex structure of natural images, modern processing methods need to exploit non-local interactions. The anisotropic regularity of images is a consequence of the geometric patterns one can find for instance along edges and textures. Recent works in harmonic analysis such as the grouplets of Mallat [38] show how to build highly adaptive orthogonal bases to extract elongated texture patterns. On the other hand, very effective denoising methods such as the non-local means of Buades et al. [8] lead to an adaptive filtering without the explicit construction of an orthogonal transform.

The main contribution of this paper is that it bridges the gap between adaptive filtering methods and thresholding in adapted orthogonal bases. In order to do so, it reviews recent approaches to semi-local and non-local adaptive filtering of images. The use of thresholding in non-local orthogonal bases is suggested in [50] and we further study the properties of such thresholdings. The connexions with more traditional wavelet-based and total variation algorithms is made clear together with a study of the pros and cons of each method.

1. Local, Semi-local and Non-local Manifolds in Images Processing. In the following, we consider a discrete image as a finite dimensional vector $f \in \mathbb{R}^n$, which is a mapping $f : \Lambda \to \mathbb{R}$ where the sampling grid is $\Lambda \stackrel{\text{def.}}{=} \{0, \ldots, \sqrt{n}-1\}^2$. These images are supposed to be sampled from a

---

*CNRS, Ceremade, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France (gabriel.peyre@ceremade.dauphine.fr).

continuous function defined on $[0,1]^2$ and section 2.2 briefly studies the convergence of the discrete operators to their continuous counterparts.

1.1. High Dimensional Lifting. In order to describe in a unified way several recently proposed image restoration schemes, this paper considers a lifting that maps pixels to elements of a high dimensional set

$$\varphi_f : x \in \Lambda \longmapsto \varphi_f(x) \in \mathcal{M} \subset \mathbb{R}^d.$$

This lifting should really be understood as a one-to-one association so that $\mathcal{M} = (\varphi_f(x))_{x \in \Lambda}$. Thanks to this association, one can map any function $g \in \ell^2(\Lambda)$ to a function

$$\tilde{g} : \mathcal{M} \longrightarrow \mathbb{R} \qquad \text{where} \qquad \forall\, p \in \mathcal{M}, \quad \tilde{g}(p) = g(\varphi_f^{-1}(p)). \tag{1.1}$$

In order to process $g \in \ell^2(\Lambda)$ (which might be different from f), one can make computations on the lifted mapping $\tilde{g} \in \ell^2(\mathcal{M})$. Although computations done on $\tilde{g}$ can be performed equivalently on g, it is useful to think about computations on $\tilde{g}$ as really making use of the geometry of $\mathcal{M}$. This paper explores various processing methods of elements of $\ell^2(\mathcal{M})$ using tools from harmonic analysis on manifolds and graphs.

It is also important to note that the lifting φ_f might depend on the image f, which can make even linear computations on $\tilde{f}$ actually non-linear and adaptive.

The set $\mathcal{M}$ can be thought of as being sampled along a low-dimensional manifold (a surface parameterized by a small number of relevant parameters) embedded in the high dimensional feature space $\mathbb{R}^d$. In practice, the set $\mathcal{M}$ can have a complex structure with a non-trivial topology, which prevent it from being globally a smooth manifold.

1.2. From Local to Non-Local Processing. This article considers three different kinds of such liftings φ_f, that correspond to three different computation modes.

– **Local processing** corresponds to the traditional setting where only local interactions of neighboring points in Λ are considered. The manifold is $\mathcal{M} = \Lambda$ and the corresponding lifting is trivial,

$$\forall\, x \in \mathcal{M} = \Lambda, \quad \varphi_f(x) = x. \tag{1.2}$$

In this case, the feature space dimension is $d = 2$. All classical isotropic image processing algorithms make the assumption that the regularity of f propagates equally in all directions over Λ. This includes Fourier and wavelets domain image processing [37]. Figure 0.1, left, shows this flat manifold Λ.

– **Semi-local processing** corresponds to non-linear image processing methods where the geometry of f can favor some local neighboring relationships in Λ. This is done through the following low-dimensional lifting

$$\varphi_f(x) = (x, \lambda f(x)) \in \mathcal{M} = \{(x, f(x)) \setminus x \in \Lambda\} \subset \Lambda \times \mathbb{R} \subset \mathbb{R}^3. \tag{1.3}$$

This semi-local lifting considers f as a height field surface embedded in a feature space of dimension $d = 3$. The parameter λ weights the locality of the method. This lifting is implicitly used in some recent image filtering algorithms such as Yaroslavsky's filter [57], the bilateral filter [51] and Susan filter [47]. In [18], Comaniciu and Meer introduces an iterative mean-shift procedure over the semi-local set $\mathcal{M}$ that extends bilateral filtering. This 3D lifting is used by Spira et al. [49] where a Beltrami flow over $\mathcal{M}$ allows one to perform denoising. Figure 0.1, center, shows this semi-local manifold for a smooth image f.

– **Non-local processing** takes into account neighboring relationship of pixels that can be far away in the image domain Λ. Buades et al. [8] have proposed to use the following lifting

$$\varphi_f(x) = p_x(f) \in \mathcal{M} \subset \ell^2(\Lambda_\tau) \tag{1.4}$$

where $p_x(f)$ is the patch of width $\tau > 0$ extracted from f around x

$$\forall\, t \in \Lambda_\tau, \qquad p_x(f)(t) \stackrel{\text{def.}}{=} f(x+t). \tag{1.5}$$

where the sampling grid is $\Lambda_\tau \stackrel{\text{def.}}{=} \{-(\tau-1)/2, \ldots, (\tau-1)/2\}^2$ and τ is an odd integer. The set $\mathcal{M}$ is composed of all the patches

$$\mathcal{M} \stackrel{\text{def.}}{=} \{p_x(f)\}_{x \in \Lambda} \subset \ell^2(\Lambda_\tau) \subset \mathbb{R}^{\tau^2}. \tag{1.6}$$

The dimensionality of the feature space is thus $d = \tau^2$. The set $\mathcal{M}$ can be thought as a manifold embedded in the Euclidean space $\mathbb{R}^d$, although it often has a complex structure with self intersections. Figure 0.1, right, shows this non-local manifold. For a smooth image f, this manifold can be displayed as a 2D surface embedded in $\mathbb{R}^3$ using principal component dimensionality reduction, see section 1.3 for more details.

The geometry of the local manifold $\mathcal{M} = \Lambda$ in (1.2) is flat and hence trivial. The semi-local manifold (1.3) also has a simple geometry because it has an explicit parameterization as an elevation field $x \mapsto (x, \lambda f(x))$. This paper is focussed on the study of non-local manifolds (1.6) which are embedded in the high-dimensional space $\ell^2(\Lambda_\tau)$.

1.3. The manifold structure of patches. Small patches $p_x(f)$ of width τ are extracted from an image $f \in \ell^2(\Lambda)$ as defined in equation (1.5). Each patch $p_x(f)$ is a vector of $\mathbb{R}^d$ with $d = \tau^2$. The point x is the center of the patch $p_x(f)$ and the width τ controls the size of typical features one wants to analyze. The set of patches $\mathcal{M} \subset \ell^2(\Lambda_\tau)$ can be thought as a set in a high dimensional space whose geometry reflects the features of f.

The mapping $x \mapsto p_x(f)$ gives a trivial 2D parameterization of the set $\mathcal{M}$. The set of patches in a particular function f can thus be thought locally as a 2D sub-manifold of $\mathbb{R}^d$. Excepted in simple cases, this mapping is however globally not injective and is typically badly behaved in real applications. Complex images f lead to set $\mathcal{M}$ with high curvature variations and self-intersections. A feature $p \in \mathcal{M}$ where this set intersects itself corresponds to points $x \neq y$ such that $p = p_x \approx p_y$.

For some specific classes of images, the set of patches $\mathcal{M} = \{p_x(f)\}_{x \in \Lambda}$ is close to a continuous set $\tilde{\mathcal{M}}$ having a manifold structure. In numerical processing of discretized images, this manifold model assumes that $\mathcal{M}$ is a discrete sampling of the underlying continuous manifold $\tilde{\mathcal{M}}$. After a brief review of recent developments in the field of manifold modeling of images, we give two simples examples (smooth images and cartoon images) where $\mathcal{M}$ is actually a subset of a low dimensional smooth manifold of $\mathbb{R}^d$.

Patches in natural images. The study of $\mathcal{M}$ for discretized images and a width τ of 3 pixels has been carried over by Lee et al. [36]. They report statistical evidences showing that for natural images, the set $\mathcal{M}$ restricted to high contrast patches is located around the manifold $\tilde{\mathcal{M}}$ of edge patches. These results have been refined by Carlsson et al. [11, 12] that perform a simplicial approximation of the manifold. Peyré in [42] studies several models of local image manifolds for which an explicit parameterization is available.

In computer vision, the manifold parameterization of structures such as edges and corners is an efficient way to express the salient features detection problem. Baker et al. [3] formulate this manifold search and propose efficient algorithms to solve it. Huggins and Zucker [32] propose a local principal components analysis over these features manifolds in order to locally retain only relevant quantities and speed up computations.

The geometric diffusion framework of Coifman et al. [16] can be used to process functions defined on a graph computed from $\mathcal{M}$. Szlam et al. [50] use this diffusion framework to perform various image processing tasks. In particular they perform image denoising using a linear heat diffusion on a graph as exposed in section 4.2. They also propose more general non-linear operators but they do not explore this possibility any further. We show in sections 5 and 6 how these general linear and non-linear regularization schemes on graphs relate to several recently proposed variational methods. In particular a non-linear thresholding in an adapted orthogonal basis is shown to perform better than a linear diffusion on the adapted graph.

The general study of library of images is emerging as an application of methods from learning theory. In particular, the manifold structure of certain images ensembles is emphasized by Donoho and Grimes [23] and Wakin et al. [53]. These non-linear models have applications to image synthesis in computer vision [1] and texture synthesis in computer graphics [43, 54, 40]. They share similarities with technics of local patch recopy following the original work of Efros and Leung [26]. This paper shows that the set of local patches indeed often has a manifold structure that can be exploited to perform non-local image processing such as denoising.

The example of smooth images. A patch $p_x(f)$ extracted from a smooth C^2 image f has a slowly varying intensity. In the continuous setting, such a patch is well approximated by a linear function

$$p_x(f)(t) \approx P_{(a(x),b(x))}(t) \quad \text{where} \quad P_{(a,b)}(t) = \langle t,\, a \rangle + b \quad \text{with} \quad \left\{ \begin{array}{l} a = \nabla_x f \\ b = f(b) \end{array} \right. .$$

This shows that the patch manifold $\mathcal{M}$ corresponding to a smooth image is close to the continuous 3D manifold of $\mathbb{R}^d$ composed of linear patches

$$\tilde{\mathcal{M}} = \left\{ \pi_{\Lambda_\tau} P_{(a,b)} \setminus (a,b) \in \mathbb{R}^3 \right\} \subset \ell^2(\Lambda_\tau)$$

where π_{Λ_τ} is a sampling operator that maps a continuous function to a vector defined on Λ_τ.

Figure 0.1, right, shows graphically the inclusion $\mathcal{M} \subset \tilde{\mathcal{M}}$, where the set $\mathcal{M}$ is displayed as a sampled 2D surface (the grid indicates the sampling locations) included in a box that represents a cube of $\tilde{\mathcal{M}}$ (each axis corresponds respectively to values of a and b). Even in this case, one can see that the set $\mathcal{M}$ does not samples the full manifold $\tilde{\mathcal{M}}$ (only a 2D sub-surface). When one refines the sampling of the original image f, the set $\mathcal{M}$ converges to a smooth surface embedded in $\tilde{\mathcal{M}}$ that exhibits self intersections along curves.

The example of cartoon images. As introduced in [22], a cartoon image is sampled from a function that is regular outside a set of regular edges. An even simpler model consists of images of a binary object $f = 1_B$ where the boundary ∂B is a smooth curve. Locally a patch of f is well approximated by a single straight edge

$$p_x(f)(t) \approx P_{(\theta(x),\delta(x))}(t) \qquad \text{where} \qquad P_{(\theta,\delta)}(t) = P\big(R_\theta(x - (\delta, 0)) \big),$$

where R_θ is the planar rotation of angle θ. The vertical step edge is $P(t_1, t_2) = 0$ if $t_1 < 0$ and $P(t) = 1$ otherwise. In this simple setting, the discrete set of patches $\mathcal{M}$ is close to the manifold of binary edges

$$\tilde{\mathcal{M}} = \left\{ \pi_{\Lambda_\tau} P_{(\theta,\delta)} \setminus (\theta, \delta) \in [0, 2\pi) \times \mathbb{R} \right\} \subset \ell^2(\Lambda_\tau)$$

This edge manifold $\tilde{\mathcal{M}}$ is parameterized by the orientation θ of the edge in $[0, 2\pi)$ and the distance δ of the edge from the center of the patch. Figure 1.1 shows this 2D parameterization of the manifold of step edges.
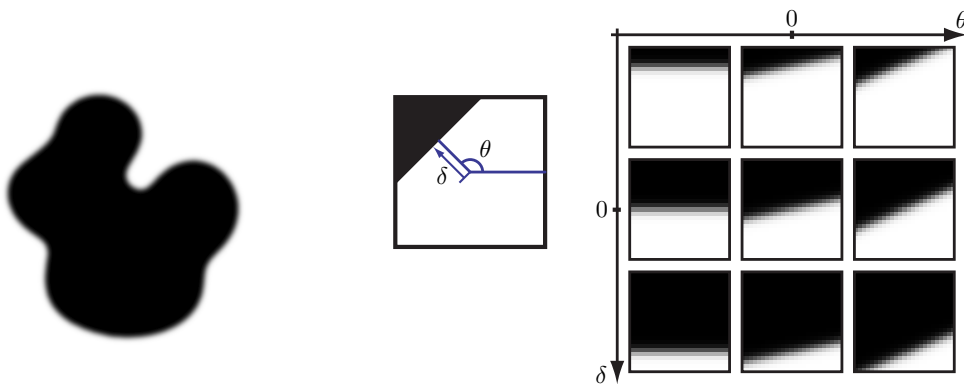

FIG. 1.1. *Left: an example of cartoon image. Right: parameterization of the manifold of edge patches and some examples.*

For this simple class of black and white images, the set $\mathcal{M}$ is thus approximately sampled along a smooth manifold $\tilde{\mathcal{M}}$ of dimension 2 embedded in $\mathbb{R}^d$ for $d = \tau^2$. In this case, the set $\mathcal{M}$ samples the whole continuous manifold $\tilde{\mathcal{M}}$, however this sampling might be highly non-uniform. The manifold $\tilde{\mathcal{M}}$ has the topology of a cylinder, and figure 1.2 shows an example of a discrete set $\mathcal{M}$ sampled on this cylinder.
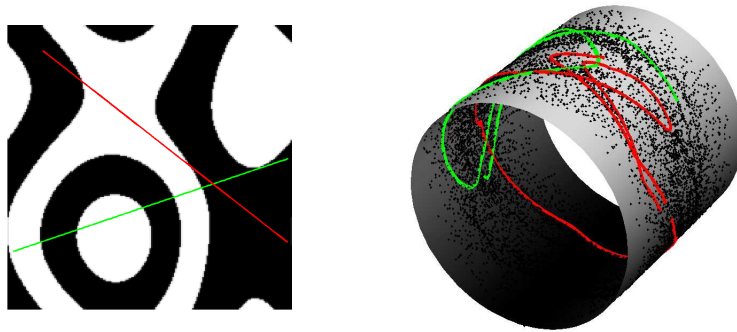
FIG. 1.2. *Left: a cartoon image. Right: 3D representation of the set $\mathcal{M}$ as samples of the edge manifold $\tilde{\mathcal{M}}$ (depicted in 3D as a cylinder). The two curves on the manifold corresponds to patches extracted along the two lines in the image.*

2. Diffusions and Laplacians on a Manifold. In order to process an image $f \in \ell^2(\Lambda)$, this paper uses tools from graph theory and calculus on manifolds in order to modify the associated mapping $\tilde{f} \in \ell^2(\mathcal{M})$ defined over the discretized manifold $\mathcal{M}$. This section reviews several linear operators that can be applied to elements of $\ell^2(\mathcal{M})$.

2.1. Discrete Diffusion Operators on Graphs. In the three processing modes, the set $\mathcal{M}$ is embedded in the Euclidean space $\mathbb{R}^d$ with $d = 2$ for local computations, $d = 3$ for semi-local computations and $d = \tau^2$ for non-local computations. As shown in section 1.3, in some special case, the set $\mathcal{M}$ is close to a low dimensional smooth manifold $\tilde{\mathcal{M}}$. It thus makes sense, for close enough features $p, q \in \mathcal{M}$ to consider their extrinsic distance $\|p - q\|$ computed over $\mathbb{R}^d$ as an approximation for the geodesic distance $d_{\tilde{\mathcal{M}}}(p, q)$ over $\tilde{\mathcal{M}}$ which is usually not available.

Diffusion kernel. In order to process a manifold mapping $g \in \ell^2(\mathcal{M})$ (such as for instance $\tilde{f}$ itself or other mappings), one defines a symmetric isotropic kernel on pairs of points of the discrete manifold

$$\forall p, q \in \mathcal{M}, \quad W_0(p, q) \stackrel{\text{def.}}{=} \exp\left(-\frac{\|p - q\|^2}{2\sigma^2}\right). \tag{2.1}$$

The parameter σ is supposed to be small enough so that the extrinsic computation of $\|p - q\|$ approximates the geodesic distance $d_{\tilde{\mathcal{M}}}(p, q)$ for close pair of points $(p, q) \in \mathcal{M}^2$. In practice, σ should be adapted for specific image processing applications.

The normalized filtering kernel is defined as

$$W(p, q) \stackrel{\text{def.}}{=} \frac{1}{D(p)} W_0(p, q) \qquad \text{where} \qquad D(p) \stackrel{\text{def.}}{=} \sum_{q \in \Lambda} W_0(p, q). \tag{2.2}$$

The kernel W defines an operator on mappings $g \in \ell^2(\mathcal{M})$

$$\forall p \in \Lambda, \quad W g(p) \stackrel{\text{def.}}{=} \sum_{p \in \mathcal{M}} W(p, q) g(q). \tag{2.3}$$

This can be thought as a low-pass filtering of g since this operator does not modify constant functions: $W1 = 1$. This filtering can be very different with respect to traditional isotropic filterings when one considers semi-local or non-local embeddings. The left images of figure 2.1 show the diffusion of a Dirac $W\delta$ located in the center of the image. This diffusion is performed using the operator W corresponding to the three computation modes (local, semi-local and non-local). One should note that although $W\delta \in \ell^2(\mathcal{M})$ is defined on the set $\mathcal{M}$, it can be equivalently displayed as a standard 2D image.

Linear operators such as W acts on elements of $\ell^2(\mathcal{M})$ which can be represented as discrete vectors in $\mathbb{R}^n$ where $n = |\mathcal{M}|$ is the number of pixels in the input image f. These operators can thus be considered as $n \times n$ matrices.

Laplacian operators. The Laplacian operator L and its symmetrized version L_0 are defined as

$$L \stackrel{\text{def.}}{=} \text{Id} - W \qquad \text{and} \qquad L_0 \stackrel{\text{def.}}{=} D^{1/2} L D^{-1/2} = \text{Id} - D^{-1/2} W_0 D^{-1/2}, \tag{2.4}$$
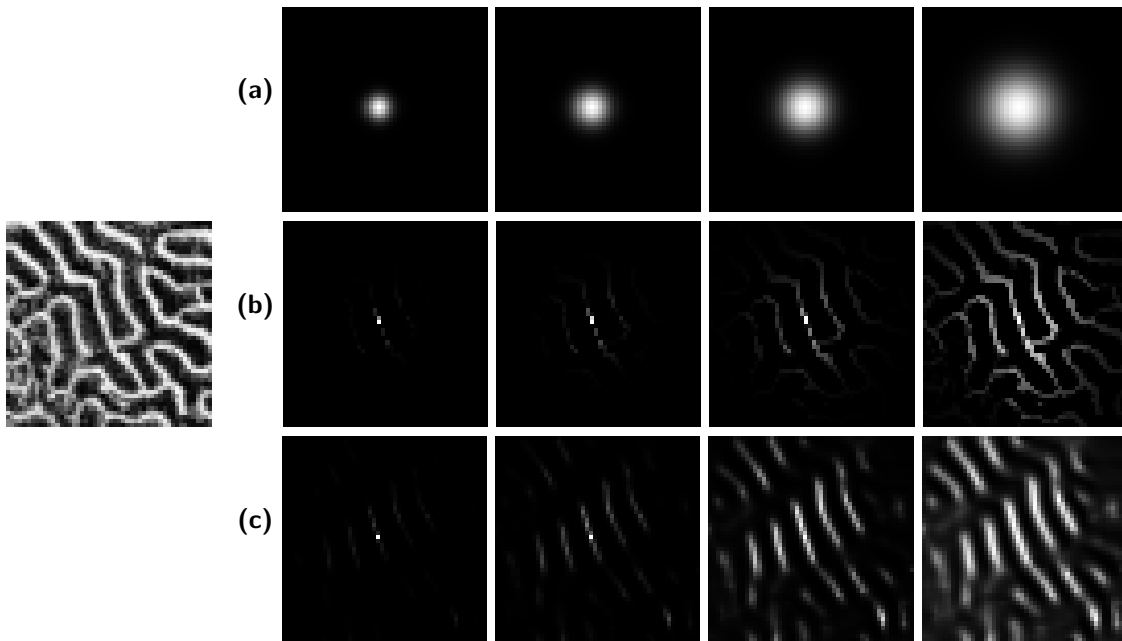
FIG. 2.1.   *Left: original image f. Right: heat diffusions with an increasing time for: (a) local embedding*
$x \mapsto x$, *(b) semi-local embedding* $x \mapsto (x, \lambda f(x))$, *(c) non-local embedding* $x \mapsto p_x(f)$.

where D is the diagonal operator $D = \operatorname{diag}_{p \in \mathcal{M}}(D(p))$.

The normalized Laplacian L_0 corresponds to a discrete graph Laplacian as defined for example
by Chung [15]. For computer graphics purposes, other discretizations of the Laplacian are available
[44, 56] that make use of a triangulation data-structure. On real image data-sets, finding such a
triangulation is however non-trivial although some methods are emerging [11].

Gradient operators. The gradient operator maps $g \in \ell^2(\mathcal{M})$ defined on the discrete set $\mathcal{M}$ to
a measure of similarity on each couple of points of $\mathcal{M} \times \mathcal{M}$

$$\forall (p, q) \in \mathcal{M} \times \mathcal{M}, \qquad (Gg)(p,q) = \sqrt{W_0(p,q)} \left(\frac{g(p)}{\sqrt{D(p)}} - \frac{g(q)}{\sqrt{D(q)}} \right). \qquad (2.5)$$

This gradient is thus a linear mapping $G : \ell^2(\mathcal{M}) \mapsto \ell^2(\mathcal{M} \times \mathcal{M})$, which corresponds to a matrix
of $n \times n^2$ elements. One can also consider point-wise gradient vectors

$$\forall p \in \mathcal{M}, \quad G_p g = ((Gg)(p,q))_{q \in \mathcal{M}}, \qquad (2.6)$$

which corresponds to vectors in $\mathbb{R}^n$.

This normalized gradient is different from the more classical un-normalized gradient, as intro-
duced for instance in [30, 29]

$$\forall (p, q) \in \mathcal{M} \times \mathcal{M}, \qquad (\tilde{G}g)(p,q) = \sqrt{W_0(p,q)} \, (g(p) - g(q)). \qquad (2.7)$$

Section 7.1 and in particular figure 7.4 studies the advantages of using the normalized formulation.

The Laplacian L_0 and its un-normalized version $D - W_0$ are symmetric operators than can be
decomposed as

$$L_0 = G^{\mathrm{T}} G \qquad \text{and} \qquad D - W_0 = \tilde{G}^{\mathrm{T}} \tilde{G} \qquad (2.8)$$

where G^{T} is the transposed matrix.

Approximate Operators. The diffusion and Laplacian kernels W_0 and L_0 are linear operators on
$\ell^2(\mathcal{M}) \simeq \mathbb{R}^n$ and are computed as matrices of size $n \times n$ acting on $\mathbb{R}^n$. These matrices are defined
using the gaussian weights (2.1) and are thus full and difficult to use for intensive computations. In
order to speed up both filtering computations and eigenvectors extraction, one typically uses either

a k-nearest-neighbors or an ε-distance approximation of the gradient defined in equation (2.5)

$$(G^k g)(p,q) \overset{\text{def.}}{=} \begin{cases} (Gg)(p,q) & \text{if} \quad q \in \mathrm{NN}_k(p), \\ 0 & \text{otherwise}, \end{cases} \tag{2.9}$$

$$\text{and} \quad (G^\varepsilon g)(p,q) \overset{\text{def.}}{=} \begin{cases} (Gg)(p,q) & \text{if} \quad \|p-q\| \leqslant \varepsilon, \\ 0 & \text{otherwise}, \end{cases} \tag{2.10}$$

where q belongs to the set $\mathrm{NN}_k(p)$ of k-nearest neighbors of p if

$$\# \left\{ r \in \mathcal{M} \setminus \|p-r\| < \|p-q\| \right\} < k.$$

The approximated Laplacians are defined as

$$L_0^k \overset{\text{def.}}{=} G^{k\mathrm{T}} G^k \quad \text{and} \quad L_0^\varepsilon \overset{\text{def.}}{=} G^{\varepsilon\mathrm{T}} G^\varepsilon. \tag{2.11}$$

These reduced operators can be conveniently stored as sparse matrices which leads to efficient algorithms for the iterative resolution of partial differential equations and the extraction of eigenvectors. The convergence of L_0^k and L_0^ε toward L_0 when $k \to n$ and $\varepsilon \to 0$ is fast whenever the set of points $\mathcal{M}$ is close to a smooth manifold $\tilde{\mathcal{M}}$. The following section studies the convergence of these finite dimensional operators to their continuous counterparts when $n \to +\infty$.

2.2. Convergence Analysis. In order to understand the behavior of the denoising algorithms proposed in this paper, it is important to study the connexion between the discrete operators introduced in section 2.1 and their continuous counterparts defined using differential calculus on manifolds.

As explained in section 1.3, the set $\mathcal{M}$ can sometimes be well approximated by a low dimensional manifold $\tilde{\mathcal{M}}$. In order to study the convergence properties of operators such as W and L_0, the discrete set $\mathcal{M}$ is supposed to be sampled on a continuous manifold $\tilde{\mathcal{M}} \subset \mathbb{R}^d$ of dimension $\tilde{d} < d \ll n$. The Riemannian metric of $\tilde{\mathcal{M}}$ is defined thanks to the embedding of $\tilde{\mathcal{M}}$ in $\mathbb{R}^d$.

For such a smooth manifold $\tilde{\mathcal{M}}$, the Laplace Beltrami $\Delta_{\tilde{\mathcal{M}}}$ operates on smooth functions $\tilde{g} \in \mathrm{L}^2(\tilde{\mathcal{M}})$. It can be defined as the limit of a local averaging

$$\forall p \in \tilde{\mathcal{M}}, \quad \frac{2}{\sigma^2} \left(\tilde{g}(p) - (\mathcal{W}\tilde{g})(p) \right) \quad \overset{\sigma \to 0}{\longrightarrow} \quad \Delta_{\tilde{\mathcal{M}}} \tilde{g}(p) \tag{2.12}$$

where $\mathcal{W}$ is a local geodesic averaging

$$\forall p \in \tilde{\mathcal{M}}, \quad (\mathcal{W}\tilde{g})(p) = \frac{1}{|B(p)|} \int_{B(p)} \tilde{g}(q)\mathrm{d}q \quad \text{where} \quad B(p) \overset{\text{def.}}{=} \left\{ q \in \tilde{\mathcal{M}} \setminus d_{\tilde{\mathcal{M}}}(p,q) \leqslant \sigma \right\} \tag{2.13}$$

where $d_{\tilde{\mathcal{M}}}(p,q)$ is the geodesic distance along the manifold $\tilde{\mathcal{M}}$, the length of the shortest curve traced on $\tilde{\mathcal{M}}$ between p and q.

Equation (2.12) is a continuous version of the definition (2.4) of the discrete Laplacians L and L_0. The width σ of the averaging in (2.13) is the equivalent of the width of the gaussian kernel in (2.1). The discrete sum used in the avergaging (2.2) is replaced by a continuous averaging over a geodesic ball in (2.13).

One can make this connection more precise since if the sampling $\mathcal{M}$ of $\tilde{\mathcal{M}}$ is uniform, both L and L_0 converge to $\Delta_{\tilde{\mathcal{M}}}$. This convergence is proved by Belkin [5] and Lafon [16] when one first take an infinite number of points and then $\sigma \to 0$. The following bound on both n and σ is derived by Singer [46] and refines a result of Hein et al. [31] in the case where $\mathcal{M}$ is drawn uniformly at random on $\tilde{\mathcal{M}}$

$$\forall p \in \mathcal{M}, \quad \frac{2}{\sigma^2}(Lg)(p) = (\Delta_{\tilde{\mathcal{M}}}\tilde{g})(p) + O\left(n^{-1/2}\sigma^{-(2+\tilde{d}/2)}, \sigma \right).$$

where the discrete $g \in \ell^2(\mathcal{M})$ is obtained by sampling a smooth function $\tilde{g} \in \mathrm{L}^2(\tilde{\mathcal{M}})$. The original bound [46, 31] is derived for L but the result also applies to L_0. This suggests an optimal asymptotic choice for the gaussian width $\sigma \overset{\text{def.}}{=} C(\tilde{\mathcal{M}})n^{-1/(6+\tilde{d})}$, where $C(\tilde{\mathcal{M}})$ is a constant that depends on

geometric properties of $\tilde{\mathcal{M}}$. For this well chosen σ, the discrete Laplacians converges with a speed σ toward $\Delta_{\tilde{\mathcal{M}}}$.

The main issue is thus to know wether, in applications to image processing, the set $\mathcal{M}$ is an uniform sampling of some continuous manifold $\tilde{\mathcal{M}}$. Indeed, this sampling is constrained by the lifting $x \in \Lambda \mapsto \varphi_f(x) \in \mathcal{M}$, which might not be uniform in practical situations. To be more precise, one needs to make the distinction between the three modes of computation introduced in section 1.1.

− *Local lifting:* in this case, the embedding is trivial $\varphi_f(x) = x$ and thus L converges to

$$\Delta_{\tilde{\mathcal{M}}} = \Delta = -\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2},$$

the traditional Laplacian over $\mathbb{R}^2$.

− *Semi-local lifting:* in this case, $\mathcal{M}$ corresponds to a height field and L is close to $\mathcal{L}$. If λ in equation (1.3) is small, the sampling of the manifold is close to uniform. This approximate convergence is related to the convergence of neighboring filters to non-linear partial differential equations, see for instance [4, 27, 48]

− *Non-local lifting:* in this case, $\mathcal{M}$ might be a complicated set of points. Indeed, the examples of smooth images and cartoon images exposed in section 1.3 show that even in simple cases, this sampling might be far from uniform. In this case L_0 is very different from $\mathcal{L}$ and the convergence of L_0 to $\Delta_{\tilde{\mathcal{M}}}$ is not true in general. In some restricted cases one can however make an asymptotic derivation for non-local averaging operators, see for instance the analysis of the non-local means filter [9].

2.3. Manifold Spectral Basis. The symmetric Laplacian L_0 defined in equation (2.4) is $L_0 = \mathrm{Id} - \tilde{W}$ where $\tilde{W} = D^{-1/2}W_0 D^{-1/2}$. Using the decomposition (2.8), the inner product associated to this symmetric operator can be written for $g \in \ell^2(\mathcal{M})$

$$\langle L_0 g, g \rangle = \langle G^{\mathrm{T}} G g, g \rangle = \langle Gg, Gg \rangle = \sum_{p,q \in \mathcal{M}} W_0(p,q) \left(\frac{g(p)}{\sqrt{D(p)}} - \frac{g(q)}{\sqrt{D(q)}} \right)^2.$$

This implies that L_0 is a positive semi-definite operator with only the constant vectors in its kernel as soon as the graph is connected. One can thus define the orthogonal basis $(u_\omega)_{\omega=0}^{n-1}$ of $\ell^2(\mathcal{M})$

$$L_0 u_\omega = \lambda_\omega u_\omega \qquad \text{with} \qquad 0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n.$$

These eigenvectors can be conveniently stored in an orthogonal matrix $U_0 \in \mathbb{R}^{n \times n}$, where one uses the identification $\ell^2(\mathcal{M}) \simeq \mathbb{R}^n$

$$U_0(p, \omega) = u_\omega(p) \quad \Longrightarrow \quad L_0 = U_0 \Lambda U_0^{\mathrm{T}} \quad \text{where} \quad \Lambda = \{\lambda_\omega\}_\omega.$$

The parameter ω plays the role of a frequency parameter with small ω corresponding to low-frequencies. Indeed, in the case of the local lifting $\varphi_f(x) = x$, the eigenvectors of L_0 are exactly the Fourier basis, which can be written for the periodic 1D case as

$$\forall\, p \in \mathbb{Z}/n\mathbb{Z}, \quad \begin{cases} u_{2\omega}(p) &= \frac{2}{\sqrt{2n}} \cos\left(\frac{2\pi p \omega}{n}\right), \\ u_{2\omega+1}(p) &= \frac{2}{\sqrt{2n}} \sin\left(\frac{2\pi p \omega}{n}\right), \end{cases} \quad \text{with} \quad \lambda_{2\omega} = \lambda_{2\omega+1} = \sin^2\left(\frac{\omega\pi}{n}\right). \quad (2.14)$$

The eigendecomposition (2.3) corresponds to the diffusion geometries expansion introduced by Coifman et al. [16] which has been applied to image graphs by Szlam et al. [50].

Figure 2.2 shows some u_ω on a cartoon image. It compares the eigenvectors of the local, semi-local and non-local Laplacians. The eigenvectors for the local Laplacian are the Fourier vectors that extend those of (2.14) to 2D. The eigenvectors for the semi-local Laplacian are similar excepted that the oscillations are localized either inside or outside the boundary. The eigenvectors of the non-local Laplacian can be classified in three groups.

− The first two eigenvectors are constant on each side of the shape B. The projection on these two vectors recovers perfectly a binary cartoon image.

− The low frequency eigenvectors for small ω are located along the boundary ∂B and are highly oscillating. The projection on these eigenvectors recovers sharp steps along ∂B.

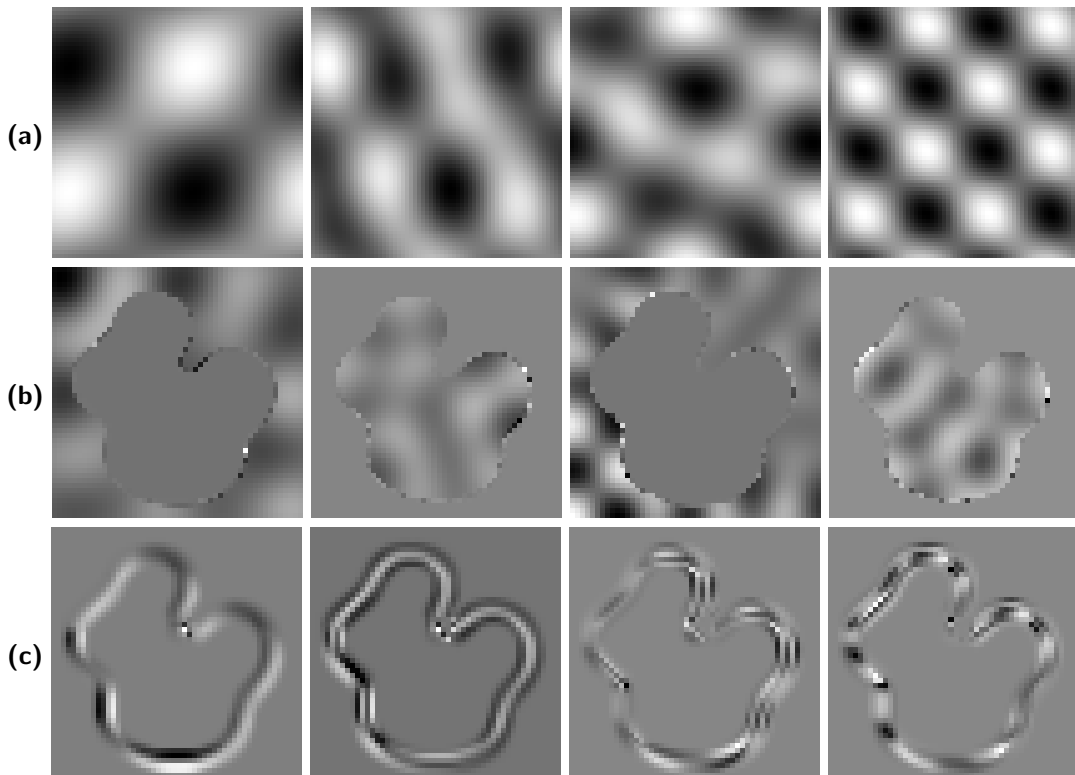− The remaining higher frequencies are irrelevant and capture noise.

FIG. 2.2. *Some eigenvectors u_ω of Laplacians for (a) the local Laplacian (Fourier basis), (b) the semi-local Laplacian, (c) the non-local Laplacian. The image f used to compute the discrete manifold is shown on figure 1.1, left.*

3. Denoising: PDE Flows, Variational Minimization and Thresholding. Denoising is modeled in a probabilistic way as an inverse problem where one wishes to recover an image f from a noisy observation $f = f^0 + \varepsilon$ where ε is a white noise of variance $|\varepsilon|^2$. A denoising algorithm builds an estimator $\bar{f} \in \mathbb{R}^n$ of the true data f^0 that depends only on the observed f. This estimator is a random vector that depends on the gaussian noise ε and its efficiency is measured using the expectation of the error $E(\|f^0 - \bar{f}\|^2)$.

This section reviews three important classes of estimators, which are all based on taking advantage of a well chosen energy $\mathcal{E}_f(g)$ that should be small for typical images one wishes to recover. It is important to note that the energy $\mathcal{E}_f$ might depend on the noisy input f itself, which makes some of the proposed methods adaptive to the content of the image.

Sections 4, 5 and 6 specialize these three kinds of estimators to computations on a discrete manifold $\mathcal{M}$. They use either operators such as the gradient and the Laplacian on $\mathcal{M}$ (defined in section 2.1) or the manifold spectral basis (introduced in section 2.3) to define the energy $\mathcal{E}_f$.

Two approaches are usually considered in the literature. The first one uses tools from variational calculus and partial differential equations, while the other one exploits decompositions of harmonic analysis with thresholdings in orthogonal bases. All the proposed estimators $\bar{f} = f_t$ depend on some scale $t > 0$ that controls the degree of regularization one imposes on the solution. This scale is adapted to the noise level more or less automatically.

3.1. PDE Flows. A PDE-flow regularizes the original noisy input f by a gradient descent of the energy

$$\forall\, t > 0, \quad \frac{\partial f_t}{\partial t} = -\mathrm{grad}_{f_t}(\mathcal{E}_f) \qquad \text{with} \qquad f_0 = f. \tag{3.1}$$

If the original energy $\mathcal{E}_f$ is quadratic, this leads to a linear differential equation, otherwise it can be non-linear. For instance, the heat equation and the total variation flow are derived formally from

$$\mathcal{E}_f(f) = \begin{cases} \frac{1}{2}\int |\nabla_x f|^2 \mathrm{d}x, \\ \int |\nabla_x f| \mathrm{d}x \end{cases} \implies \frac{\partial f_t^{\mathrm{h}}}{\partial t}(x) = \begin{cases} \Delta f_t \\ \mathrm{div}\left(\frac{\nabla f_t}{\|\nabla f_t\|}\right) \end{cases}$$

and do not depend on f. Other non-linear flows such as the anisotropic diffusion of Perona and Malik [41] can be considered.

It is also possible to compute quadratic energies that depend on f, such as an anisotropic diffusion along a tensor field estimated from f, see [55]. In this case, although the flow $t \mapsto f_t$ is linear, the estimator computation $f \mapsto f_t$ is actually non-linear because $\mathcal{E}_f$ depends on the noisy input f. Section 4 derives PDE flows using an energy $\mathcal{E}_f$ computed using operators on the discrete manifold $\mathcal{M}$. Such energies depends on f and makes the whole computation adaptive and non-linear.

Time-dependant manifold. The flow of equation (3.1) is computed from the energy E_f which depends only on the noisy input image. The discrete manifold $\mathcal{M}$ is computed once for all from the original input and does not change during the computations. Another option is to modify the manifold and actually compute it from the current estimate f_t at time t

$$\forall\, t > 0, \quad \frac{\partial f_t}{\partial t}(x) = -\mathrm{grad}_{f_t}(\mathcal{E}_{f_t}) \qquad \text{with} \qquad f_0 = f. \tag{3.2}$$

The overall process becomes even more non-linear than the original PDE.

Several works on adaptive filtering use such time-dependant regularization, see for instance anisotropic diffusion along time-varying tensors fields [52] and diffusion on triangulated surfaces [21].

3.2. Variational Regularization. A variational regularization defines an estimator f_t as a minimizer of a penalized functional

$$\forall\, t \geqslant 0, \qquad f_t = \operatorname*{argmin}_{g} \frac{1}{2}\|f - g\|_{\ell^2}^2 + t\mathcal{E}_f(g), \tag{3.3}$$

The Rudin-Osher-Fatemi denoising method [45] for instance uses the total variation energy $\mathcal{E}(g) \overset{\text{def.}}{=} \int |\nabla g|$ which does not depend on f.

3.3. Thresholding in an Orthogonal Basis. Another line of research, initiated by Donoho and Johnstone [24], solves the estimation problem using operators that manipulate independently each coefficient of the decomposition of f in an orthogonal basis.

ℓ^2, ℓ^1 *and* ℓ^0 *regularization.* Thresholding operators in an orthogonal basis $\{u_\omega\}_\omega$ of $\mathbb{R}^n$ are defined as

$$f_t^p \overset{\text{def.}}{=} \sum_\omega s_t^p(\langle f,\, u_\omega \rangle)\, u_\omega \qquad \text{for} \qquad p \in \{0, 1, 2\}. \tag{3.4}$$

The pointwise functional s_t^p modifies independently each coefficient and is defined by

$$s_t^2(x) = \frac{x}{1 + t\lambda_\omega}, \qquad s_t^1(x) \overset{\text{def.}}{=} \begin{cases} x - \mathrm{sign}(x)t & \text{if } |x| > t, \\ 0 & \text{if } |x| \leqslant t, \end{cases}$$

$$\text{and} \qquad s_t^0(x) \overset{\text{def.}}{=} \begin{cases} x & \text{if } |x| > t, \\ 0 & \text{if } |x| \leqslant t. \end{cases}$$

The estimator f_t^2 is linear and depends on a set of weights $\lambda_\omega > 0$ whereas the soft thresholding f_t^1 and hard thresholding f_t^0 perform a non-linear denoising.

In order to achieve an optimal decay rate of the risk $E(\|f^0 - f_t\|)$, Donoho and Johnstone [24] shows that one can use the universal threshold value $t \overset{\text{def.}}{=} |\varepsilon|\sqrt{2\log(n)}$. In practice however this threshold value can be set approximately to $t = 3|\varepsilon|$.

These estimators enjoy several optimality results when the function f^0 to recover belongs to a smoothness class that can be efficiently approximated linearly (for $p = 2$) or non-linearly (for $p = 1$ and $p = 0$). Donoho and Johnstone [24] show that f_t^0 and f_t^1, for the universal threshold $t \overset{\text{def.}}{=} |\varepsilon|\sqrt{2\log(n)}$, are asymptotically minimax for the estimation of piecewise regular 1D signals when one uses the wavelet basis [37].

Variational interpretation of thresholding. As shown by [14], these thresholding operators are also minimizers of variational energies of the form (3.3)

$$f_t^p = \underset{g}{\operatorname{argmin}} \frac{1}{2}\|f - g\|_{\ell^2}^2 + t\mathcal{E}_f^p(g) \quad \text{where} \quad \begin{cases} \mathcal{E}_f^2(g) = \frac{1}{2}\sum_\omega \lambda_\omega |\langle g, u_\omega\rangle|^2, \\ \mathcal{E}_f^1(g) = \sum_\omega |\langle g, u_\omega\rangle|, \\ \mathcal{E}_f^0(g) = \frac{t}{2}\#\{\omega \setminus \langle f, u_\omega\rangle \neq 0\}. \end{cases} \tag{3.5}$$

This bridges the gap between the construction of variational denoisers and the use of orthogonal bases.

If the orthogonal basis $(u_\omega)_\omega$ is a fixed basis (such as Fourier or wavelets) then the energy $\mathcal{E}_f^p$ do not depends on the input f and the estimator f_t is non adaptive. Section 6 uses manifold spectral bases computed from f, which makes the thresholding estimators adaptive to the input.

4. Filtering and Diffusion on Manifolds. This section explains how PDE flows introduced in section 3.1 can be extended by using energies $\mathcal{E}_f$ derived from operators on the manifold set $\mathcal{M}$. Such energies are used to process the function $\widetilde{f} \in \ell^2(\mathcal{M})$ defined from the noisy input $f \in \ell^2(\Lambda)$, see equation (1.1). Since the set $\mathcal{M} = (\varphi_f(x))_x$ is computed from the noisy input f, the corresponding flow $t \mapsto f_t$ is adaptive and non-linear even if $\mathcal{E}_f$ is a quadratic energy. One recovers this way several recently proposed semi-local and non-local denoising algorithms.

4.1. Non-iterative Manifold Smoothing. Before introducing linear and non-linear PDE flows, a simpler approach to denoise f is to perform a 1-step diffusion of $\widetilde{f} \in \ell^2(\mathcal{M})$

$$f_t \overset{\text{def.}}{=} W\widetilde{f}, \tag{4.1}$$

using the averaging operator defined in equation (2.2). The scale parameter t of this estimator corresponds to the scale $t = \sigma$ used in the definition of the smoothing kernel (2.1).

To minimize the average risk $E(\|f_t - \widetilde{f}^0\|^2)$, one needs to correctly define this scale parameter t that should grow in proportion to the noise level $|\varepsilon|$. This non-iterative denoising approach corresponds to various filtering operators.

− An isotropic linear filtering for the local embedding $\varphi_f(x) = x$: the operator W is a convolution against a gaussian kernel.

− A spatially varying filter for the semi-local embedding $\varphi_f(x) = (x, \lambda f(x))$: the degree of spatial adaptivity is controlled by λ and the corresponding filter $y \mapsto W(x, y)$ is compactly supported around x. In a general setting, the corresponding estimator can be written as

$$\bar{f}(x) = \frac{1}{D(x)} \sum_y w_s\left(\frac{\|x - y\|}{\sigma_s}\right) w_v\left(\frac{\|f(x) - f(y)\|}{\sigma_v}\right) f(y)$$

where $D(x)$ is the usual normalizing term. Our choice of gaussian weight (2.1) corresponds to choosing $w_s = w_v$ as a gaussian kernel and $\sigma_s/\sigma_v = \lambda$ is the embedding constant. Depending on specific choice for w_s and w_v, one retrieve several neighborhood filters such as Yaroslavsky's filter [57], the bilateral filter [51] or Susan filter [47].

− A spatially non-local filter for the non-local embedding $\varphi_f(x) = p_x(f)$: this corresponds to the non-local means of Buades et al. [8]. In a general setting, the corresponding non-local estimator can be written as

$$\bar{f}(x) = \frac{1}{D(x)} \sum_y w\left(\frac{\|p_x(f) - p_y(f)\|}{\sigma}\right) f(y)$$

where $D(x)$ is the usual normalizing term. Our choice of gaussian weight (2.1) corresponds to the original choice of w, see [8], but other kernels can be used. More advanced non-local filters have been also proposed, see for instance [19].

It is important to note that, although the filtering $\widetilde{f} \mapsto W\widetilde{f}$ is linear, the operator W depends in a non-linear way on the input f, making the overall process non-linear.

4.2. Manifold Heat Diffusion. Instead of performing a fixed amount of diffusion, one can consider a continuous diffusion process defined through the normalized gradient L defined in equation (2.5). This heat diffusion corresponds to the PDE flow (3.1) of the following manifold quadratic energy

$$\forall\, g \in \ell^2(\mathcal{M}), \quad \mathcal{E}_f^{\mathrm{q}}(g) = \frac{1}{2}\|Gg\|^2, \tag{4.2}$$

which depends on the input f since the operator G is computed on the discrete manifold $\mathcal{M}$.

To diffuse a noisy image f, one thus looks for a function $(x,t) \mapsto f_t^{\mathrm{h}}(x)$ that satisfies

$$\forall\, t > 0, \quad \frac{\partial f_t^{\mathrm{h}}}{\partial t}(x) = -L f_t^{\mathrm{h}}(x) \qquad \text{with} \qquad f_0^{\mathrm{h}} = \tilde{f}. \tag{4.3}$$

The estimator of f^0 from the noisy data f is then defined as f_t^{h} where the stopping time t should be carefully chosen to match the noise level $|\varepsilon|$.

It is important to note that in this application, the width σ of the smoothing kernel W_0, equation (2.4), is independent of the noise level $|\varepsilon|$. This width σ should be set to account for the sampling $\mathcal{M}$ of the true underlying continuous manifold. This is different from the non-iterative diffusion (4.1) where σ depends on $|\varepsilon|$.

The PDE (4.2) is applied to a discrete manifold in the geometric diffusion framework [16]. It has been used to process images with semi-local and non-local embeddings by Szlam et al. [50]. The mean-shift procedure with a gaussian kernel applied to the set $\mathcal{M}$ as done [18] is closely related to (4.2). This diffusion is also related to the short time evaluation of the Laplace Beltrami flow of Spira et al. [49], although these procedures correspond to a time-dependant embedding and a PDE diffusion flow (see the end of section 3.1).

Iterative resolution. The equation (4.3) can be solved using an explicit discretization in time with time-steps η. This leads to the following iterations

$$f_{k+1}^{\mathrm{h}} = (1 - \eta) f_k^{\mathrm{h}} + \eta W f_k^{\mathrm{h}} \qquad \text{with} \qquad f_0^{\mathrm{h}} = \tilde{f}. \tag{4.4}$$

If one chooses $\eta = t/k$, one has $f_k^{\mathrm{h}} \to f_t^{\mathrm{h}}$ when $k \to +\infty$. To ensure numerical stability, n and k should grow according to the CFL condition. The non-iterative denoising, equation (4.1), corresponds to only one iteration of this process with a large time step $\eta = 1$.
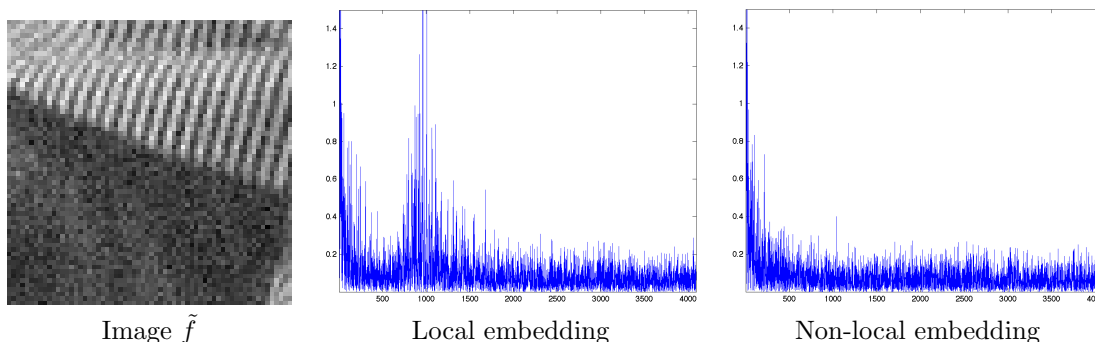

Image $\tilde{f}$ Local embedding   Non-local embedding

FIG. 4.1.   *The coefficients $|\langle \tilde{f},\, u_\omega \rangle|$ of a noisy image f. Center: using the local spectral basis $(u_\omega)_\omega$, which is the Fourier basis, right: using the non-local spectral basis.*

Spectral resolution. The normalized Laplacian introduced in equation (2.4) is related to the symmetric Laplacian through $L = D^{-1/2} L_0 D^{1/2}$. The eigenvectors of L are thus $U = D^{-1/2} U_0$ and one can compute the eigendecomposition of some vector g on U as

$$\hat{g}(\omega) = (U^{-1}g)(\omega) = \langle u_\omega,\, D^{1/2}g \rangle.$$

This spectral decomposition $\hat{g}$ allows one to solve the heat equation in the transformed domain since

$$\forall\, \omega, \quad \frac{\mathrm{d}}{\mathrm{d}t}\hat{f}_t^{\mathrm{h}}(\omega) = -\lambda_\omega \hat{f}_t^{\mathrm{h}}(\omega) \qquad \Longrightarrow \qquad \hat{f}_t^{\mathrm{h}}(\omega) = e^{-\lambda_\omega t}\hat{\tilde{f}}(\omega). \tag{4.5}$$

This equation allows one to solve exactly for f_t^{h} and offers a non-iterative alternative to (4.4) to compute the solution of the heat equation at a fixed time t. The diffused function f_t^{h} at a time t has its spectral coefficients reduced by the factor $e^{-\lambda_\omega t}$ which is small for large t and for high-frequencies ω.

On figure 4.1, one can see the magnitude of noisy coefficients $\langle \tilde{f}, u_\omega \rangle$ for the local and non-local expansions. The energy of the image is more concentrated on the low frequency part of the spectrum for the non-local Laplacian. This is a result of the adaptivity of the non-local Laplacian to the geometric content of the image. The high frequency residual is mostly the projection of the noise $\langle \tilde{\varepsilon}, u_\omega \rangle$. This compaction of the energy makes the heat diffusion over the non-local manifold much more efficient that over the local one, since the spectral attenuation efficiently removes the high frequency noise.

Diffusion examples. Figure 2.1 shows the filtering f_t^{h} of an impulse Dirac distribution $f_0^{\mathrm{h}} = \delta$. On figure 4.2 one can see the time evolution of f_t^{h} for the three computation modes, starting from the noisy input $f_0^{\mathrm{h}} = \tilde{f}$.

– The local embedding leads to the traditional Euclidean heat equation. This diffusion blurs the image since the embedding does not take into account the geometric features of f.

– Both the semi-local and the non-local embeddings correspond to a non-linear processing of f since the Laplacian L takes into account the structures of the image. This leads to a diffusion that does not blur the geometrical features. On geometrical images, these two diffusions give similar results. On complex natural images, non-local diffusion often surpasses semi-local diffusion as reported by Buades et al. [8], see section 7.
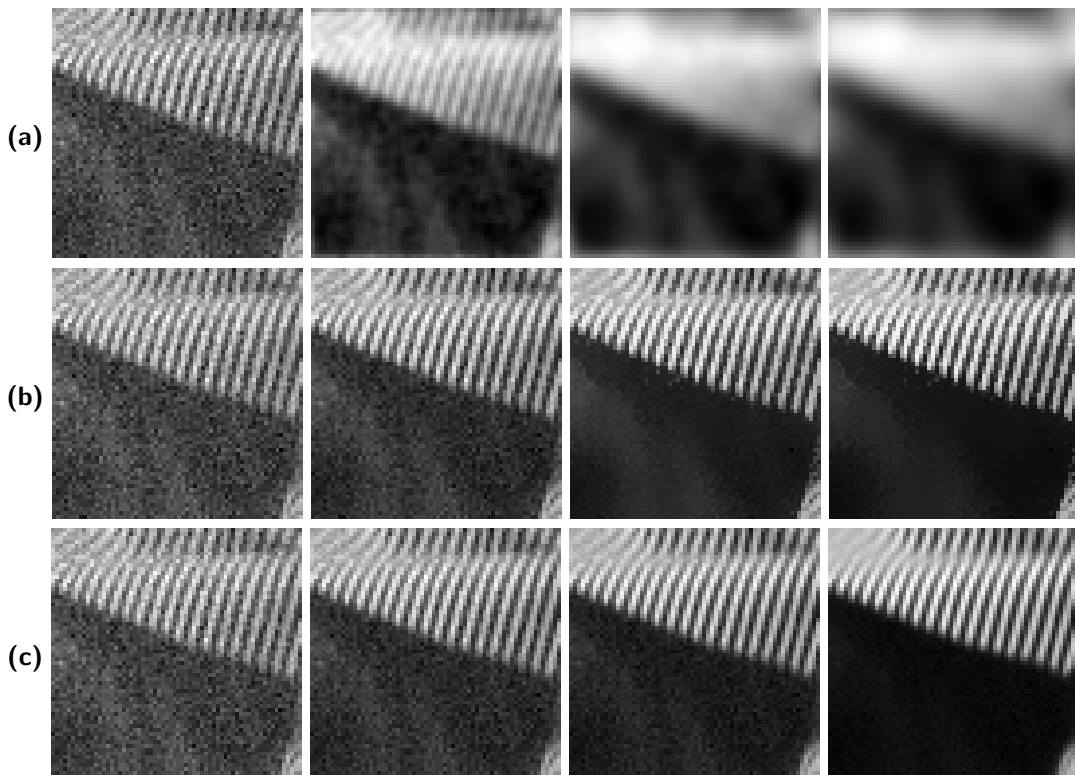

FIG. 4.2. *Heat flow f_t^h with an increasing t for: (a) local embedding (traditional heat equation) (b) semi-local embedding, (c) non-local embedding.*

4.3. Time-dependant Manifold Diffusion. The manifold diffusion (4.3) use a fixed discrete manifold $\mathcal{M}$ computed from the noisy input $\tilde{f}$. As defined in equation (3.2), it is possible to consider a flow where the manifold is continuously updated during the diffusion

$$\forall t > 0, \quad \frac{\partial f_t^{\tilde{\mathrm{h}}}}{\partial t}(x) = -L_{\mathcal{M}_t} f_t^{\tilde{\mathrm{h}}}(x) \qquad \text{with} \quad \left\{ \begin{array}{l} f_0^{\tilde{\mathrm{h}}} = \tilde{f}, \\ \mathcal{M}_t = \left(\varphi_{f_t^{\tilde{\mathrm{h}}}}(x) \right)_{x \in \Lambda} \end{array} \right. \qquad (4.6)$$

In this equation, the manifold $\mathcal{M}_t$ is computed from the solution of flow at time t, and the dependancy of the Laplacian operator $L = L_{\mathcal{M}_t}$ on the manifold is made explicit. The resulting flow $t \mapsto f_t^{\tilde{\mathrm{h}}}$ depends non-linearly from the initial condition $f_0^{\tilde{\mathrm{h}}}$, which makes the overall process radically different from the original heat flow f_t^{h}.

Such a time-dependant flow has been defined, in the semi-local setting, by Spira et al. [49], where the Laplacian is an accurate discretizeation of the Laplace Beltrami $\Delta_{\tilde{\mathcal{M}}}$ of the height field $\tilde{\mathcal{M}}$. In the non-local setting, this flow shares similarities with iterated non-local means filterings introduced by Brox and Cremers [7]. This iterated non-local means however does not converge to $f_t^{\tilde{\mathrm{h}}}$ but rather to a stationary point of a filtering process.

The non-linear PDE (4.6) can be solved explicitly in time by iterating the following two steps

$$\begin{cases} \mathcal{M}_k = \left(\varphi_{f_k^{\tilde{\mathrm{h}}}}(x) \right)_{x \in \Lambda}, \\ f_{k+1}^{\tilde{\mathrm{h}}} = (1-\eta) f_k^{\tilde{\mathrm{h}}} + \eta W_{\mathcal{M}_k} f_k^{\tilde{\mathrm{h}}}, \end{cases}$$

with the initialization $f_0^{\tilde{\mathrm{h}}} = \tilde{f}$ and where the averaging $W_{\mathcal{M}_k}$ is computed using the discrete manifold $\mathcal{M}_k$.

5. Variational Regularization on Manifolds. As explained in section 3.2, one can replace the PDE flow by a variational optimization that takes into account both a regularization through the energy $\mathcal{E}_f$ and an attach to the original noisy data.

This section first details the quadratic energy (4.2) in the case of a discrete manifold energy $\mathcal{E}_f$ and then proposes other non-linear alternatives. These estimators have been proposed in recent works using a slightly different formalism. We show how they fit into the regularization framework introduced in section 3.2 and detail several algorithms to compute efficiently these flows.

5.1. Quadradic Manifold Regularization. The quadratic energy (4.2) used to define the heat equation leads to the following quadratic regularization

$$\forall t \geqslant 0, \qquad f_t^{\mathrm{q}} = \operatorname*{argmin}_{g \in \ell^2(\mathcal{M})} \|\tilde{f} - g\|_{\ell^2}^2 + t\|Gg\|_{\ell^2}^2 \tag{5.1}$$

where the manifold gradient G is defined in equation (2.5).

A similar energy is studied by Gilboa and Osher for image restoration and segmentation [30]. They use a continuous formulation of the non-normalized gradient $\tilde{G}$ as defined in equation (2.7), whereas in this paper we use a normalized operator G defined in equation (2.4). For semi-supervised learning applications, Belkin et al. [6] have introduced a manifold regularization framework that performs interpolation by a quadratic penalization similar to (5.1).

Once the operator G is fixed, the solution of minimization (5.1) depends linearly on the input f and requires the inversion of the graph laplacian $L_0 = G^{\mathrm{T}}G$

$$\forall t \geqslant 0, \qquad f_t^{\mathrm{q}} = (\mathrm{Id} + tG^{\mathrm{T}}G)^{-1}\tilde{f} = (\mathrm{Id} + tL_0)^{-1}\tilde{f}.$$

It is important to remember that the mapping $f \mapsto f_t^{\mathrm{q}}$ is however non-linear because G actually depends on f in a non-linear way.

One can take advantage of the spectral basis introduced in equation (2.3) and compute the quadratic regularized solution over the transformed domain

$$\forall t \geqslant 0, \qquad \langle f_t^{\mathrm{q}}, u_\omega \rangle = \frac{1}{1 + t\lambda_\omega} \langle \tilde{f}, u_\omega \rangle.$$

One should compare this spectral attenuation with the one obtained in (4.5) for the heat equation. The diagonal kernel for the quadratic regularization is $(1 + t\lambda_\omega)^{-1}$ which is similar to the low-pass kernel $e^{-\lambda_\omega t}$ of the heat diffusion.

This quadratic regularization (4.5) can also be interpreted as an estimator f_t^2 as described in equation (3.5) using the spectral basis $\{u_\omega\}_\omega$ and the eigenvalues λ_ω as quadratic weights.

5.2. Total Variation Manifold Regularization. In order to turn the quadratic regularization (5.1) into a non-linear penalty that favors emergence of large gradients near discontinuities, one can consider the following non-linear TV-like energy

$$\forall\, g \in \ell^2(\mathcal{M}), \quad \mathcal{E}_f^{\mathrm{tv}}(g) = \sum_{p \in \mathcal{M}} \|G_p g\|_{\ell^2},$$

where the gradient vector $G_p g$ is defined in equation (2.6). The estimator is defined as

$$f_t^{\mathrm{tv}} = \operatorname*{argmin}_{g \in \ell^2(\mathcal{M})} \frac{1}{2}\|\tilde{f} - g\|_{\ell^2}^2 + t \sum_{p \in \mathcal{M}} \|G_p g\|. \tag{5.2}$$

A similar formulation is introduced by Gilboa et al. [29] in the continuous setting. They use the un-normalized Laplacian introduced in equation (2.7) whereas our formulation (5.2) uses the normalized gradient (2.5). Section 7.1 and in particular figure 7.4 shows the superiority of the normalized regularization over regular areas of the image. A related non-linear energy is used in Kindermann et al. [33] to solve more general inverse problems. An alternative non-linear regularization is proposed by Zhou and Schölkopf [58] and Elmoataz et al. [28] that consider the p-laplacian which is non-linear for $p \neq 2$.

The estimator f_t^{tv} depends non-linearly on the input f both because the energy $\mathcal{E}_f^{\mathrm{tv}}$ is not quadratic and because the operator G depends non-linearly on f. The computation of f_t^{tv} requires some special care for points p where $G_p f_t^{\mathrm{tv}}$ is small. A simple fix consists in replacing $\mathcal{E}_f^{\mathrm{tv}}$ by a regularized functional such as

$$\mathcal{E}_f^{\mathrm{tv},\eta}(g) = \sum_{p \in \mathcal{M}} (\eta^2 + \sum_q |(Gg)(p,q)|^2)^{1/2}$$

for a small enough η. A more accurate approach solves the discrete optimization problem with the help of the graph-cut algorithm [29].

In this paper we propose to solve the non-linear problem (5.2) using an iterative scheme derived from Chambole's algorithm [13] originally designed to solve the classical Rudin-Osher-Fatermi flow [45]. The algorithm iteratively computes gradient fields $\theta^k = (\theta^k(p))_p$, where each $\theta^k(p)$ is a vector in $\mathbb{R}^n$. The algorithm initially set

$$\forall\, p \in \mathcal{M}, \quad \theta^0(p) = 0 \in \mathbb{R}^n$$

and iterates

$$\forall\, p \in \mathcal{M}, \quad \theta^{k+1}(p) = \frac{\theta^k(p) - \eta g(p)}{1 + \eta\|g(p)\|} \quad \text{where} \quad g = G(G^{\mathrm{T}}\theta^k - \tilde{f}/t).$$

One can then prove [13] that if η is small enough,

$$\tilde{f} - tG^{\mathrm{T}}\theta^k \xrightarrow{k \to +\infty} f_t^{\mathrm{tv}}.$$

The non-linear regularization (5.2) is more involved to solve numerically than the quadratic regularization (5.1). This is due not only to the non-quadratic and non-smooth nature of $\mathcal{E}_f^{\mathrm{tv}}$ but also the the high dimensionality of $Gg \in \ell^2(\mathcal{M} \times \mathcal{M})$. If one uses the k-nearest neighbor approximation G_k of G, equation (2.9), the dimension of the range of G_k is reduced to kn, which is simpler to handle. The next section proposes alternative non-linear regularizations that involve the ℓ^0 and ℓ^1 norms over a transformed domain of dimension n, which can be solved with a fast thresholding.

6. Thresholding in a Manifold Spectral Basis. Instead of using a PDE flow or a variational minimization, an alternative for the computation of estimators is to use thresholding in orthogonal bases as explained in section 3.3. One can use the manifold spectral basis defined in section 2.3 to define these thresholding estimators.

These thresholding operators are mentioned originally in [50] without a detailed study. The authors of [50] do not go further in the analysis of such estimators and exploit the heat flow (4.3) for

image processing purpose. We detail here this construction and give in the next section numerical evidences of its efficiency.

The spectral orthogonal basis $(u_\omega)_\omega$ of $\ell^2(\mathcal{M})$ is computed using the Laplacian L_0 of the noisy input f. One then defines

$$f_t^p \overset{\text{def.}}{=} \sum_\omega s_t^p(\langle \tilde{f}, u_\omega \rangle)\, u_\omega \in \ell^2(\mathcal{M}) \tag{6.1}$$

for $p = 0, 1$, where the thresholding operators are defined in equation (3.4).

Numerical computation of the thresholding estimators. The main difficulty in computing the estimators of equation (6.1) is the need to extract the eigenvectors $(u_\omega)_\omega$ of the Laplacian matrix L_0. In numerical simulations, the sparse Laplacian L_0^k defined in equation (2.11) is used. The sparsity of the operator makes the use of an inverse power method convenient for the progressive extraction of eigenvectors, see [2]. This allows to build estimators for images of approximately $n = 100 \times 100$ pixels. For larger images, one can apply the estimator to overlapping sub-blocks extracted from the noisy image.

Connexion with manifold BV. As explained in equation (3.5), the soft thresholding estimator f_t^1 minimizes the variational problem

$$f_t^1 = \underset{g \in \ell^2(\mathcal{M})}{\operatorname{argmin}} \frac{1}{2}\|\tilde{f} - g\|^2 + t \sum_\omega |\langle g, u_\omega \rangle|$$

This is related to the manifold total variation regularization introduced in equation (5.2), that uses the ℓ^1 norm of the gradient. A notable difference is that the scale parameter t of the TV-regularization is not directly related to the noise level $|\varepsilon|$ that corrupts f. In contrast, the threshold t of the estimator f_t^1 is set proportionally to $|\varepsilon|$ and $t = 3|\varepsilon|$ works well in practice.

Contrary to the soft thresholding estimator f_t^1, the hard thresholding f_t^0 in a spectral basis has no equivalent in the manifold total variation regularization. This thresholding enforces the strict sparsity of the decomposition of f on the adapted basis $(u_\omega)_\omega$.

7. Numerical Results. In our numerical experiments, we have compared the various estimators proposed in this article, in both their semi-local and non-local versions. This includes:
– Non-iterative smoothing, section 4.1, which corresponds to the bilateral filter [51] for the semi-local kernel and to non-local means [8] for the non-local kernel.
– Heat diffusion, section 4.2, which corresponds to the diffusion maps framework [50].
– Time-denpendant heat diffusion, section 4.3.
– Quadratic regularization, 5.1, which is closely related to the work of [30].
– ℓ^1 regularization, section 5.2, with the normalized gradient G as defined in equation (2.5).
– ℓ^1 regularization, with the un-normalized gradient $\tilde{G}$ as defined in equation (2.7), which is equivalent to the non-local total-variation [29].
– Hard and soft thresholding in the manifold spectral basis, which is the new class of estimators introduced in section 6.
– Hard thresholding in a 7/9 translation invariant wavelet basis, see [37].
The efficiency of an estimator $\bar{f} = f_t$ is measured using the PSNR of the recovery error

$$R(\bar{f}) \overset{\text{def.}}{=} -20 \log_2(\|\bar{f} - f^0\|/\|f^0\|_\infty)$$

expressed in dB. This efficiency is measured for various noise levels $|\varepsilon|$, which is reported as the noisy PSNR $R(f)$.

For all of these estimators, the semi-local parameter λ of equation (1.3), the width of the patches τ of equation (1.5), the scale σ of the weights in equation (2.1) and the scale/threshold t of the estimators f_t are determined in an oracle manner in order to minimize $\|f^0 - f_t\|$. This leads to a fair comparison between all the estimators, although in practice determining t is easier for thresholding denoisers since $t \approx 3|\varepsilon|$. In order to speed-up the computation, all the algorithms use sparse approximate gradient G^k and laplacian L_0^k as introduced in equations (2.9) and (2.11). The number of nearest neighbors is set to $k = 30$ in all the experiments.
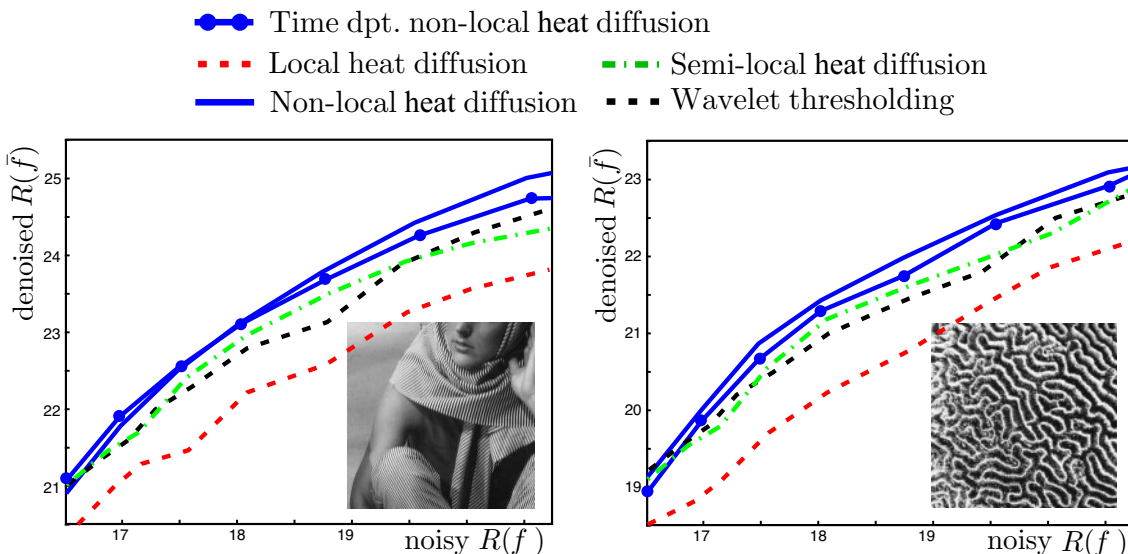
FIG. 7.1.   *Recovery error $R(\bar{f})$ of local, semi-local and non-local estimators $\bar{f}$ for a natural image and a texture for various noise levels $R(f)$.*

7.1. Numerical Comparison of the Denoising Methods. The numerical results compare the various algorithms in denoising situations with various noise levels. The tests are done on two images of 256×256 pixels, the first one (figure 7.1, left) contains smooth areas, edges and parallel stripes and the second one (figure 7.1, right) is a purely oscillating texture with complex patterns.

Comparison of the local and non-local kernels. In order to study the advantages of using non-local computation over the more classical local and semi-local settings, we compare the denoising efficiency of the heat diffusion, exposed in section 4.2. Figure 7.1 shows a graphical comparison of the evolution of the recovery error $R(\bar{f})$ with the noise level. We also report the results of the wavelet estimator, which constitutes a good benchmark. Figure 7.2 shows a visual comparison of the denoising methods for a noisy PSNR of 18.3dB. It is clear from these experiments that non-local methods tend to outperform both semi-local filterings and wavelet denoising. On more structured textures with periodic patterns such as the one on figure 7.4, the non-local technics perform even better. The time dependent heat flow, introduced in section 4.3, does not seem to improve over a fixed manifold flow (such as the manifold heat equation). It is an open question to further understand the behavior of such a highly non-linear flow on images.

Comparison of the non-local denoising algorithms. Figure 7.3 compares the efficiency of the non-local denoising methods using the recovery error $R(\bar{f})$. It shows that the estimators based on non-local graphs behave similarly, with a slight improvement for the threshold-based estimators. We note that in all the tests, the soft thresholding consistently performed better than the hard thresholding estimator.

Figure 7.4 shows a visual comparison of the denoising methods for a noisy PSNR of 18.3dB. One can notice the importance of normalizing the gradient operator as done in the estimator introduced in 5.2. The regularization with the un-normalized gradient tends to perform poorly over uniform areas where lots of patches share similar values and thus high values of $D(x)$ are expected. On this example, the thresholding operators give superior results and in particular the soft thresholding outperforms the non-local TV flow by 0.3dB. One should not however draw final conclusion from this image containing highly repetitive patterns. On more complex textures such as the one depicted on figure 7.3, right, both methods give similar results.

Overall Comparison and Numerical complexity. These experiments tend to support the following conclusions:

– Non-local methods gives superior results than semi-local ones in particular over repetitive patterns.

– Iterative non-local algorithms (like heat-diffusion, section 4.2, or quadratic regularization, section 5.1) bring some improvement over non-iterated methods like non-local means (section 4.1) and wavelet thresholding [37].

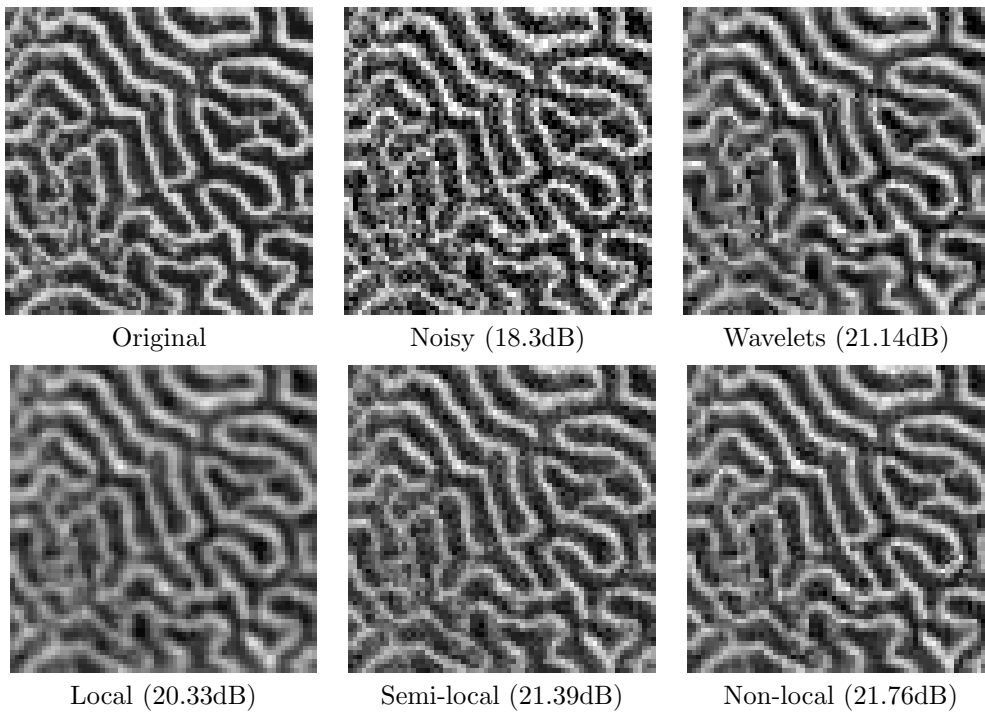– Non-linear regularization like non-local TV (section 5.2) is slightly superior to quadratic regu-

| Original | Noisy (18.3dB) | Wavelets (21.14dB) |
| Local (20.33dB) | Semi-local (21.39dB) | Non-local (21.76dB) |

FIG. 7.2. *Visual comparison of local, semi-local and non-local estimators for a noisy PSNR of* 18.3dB.
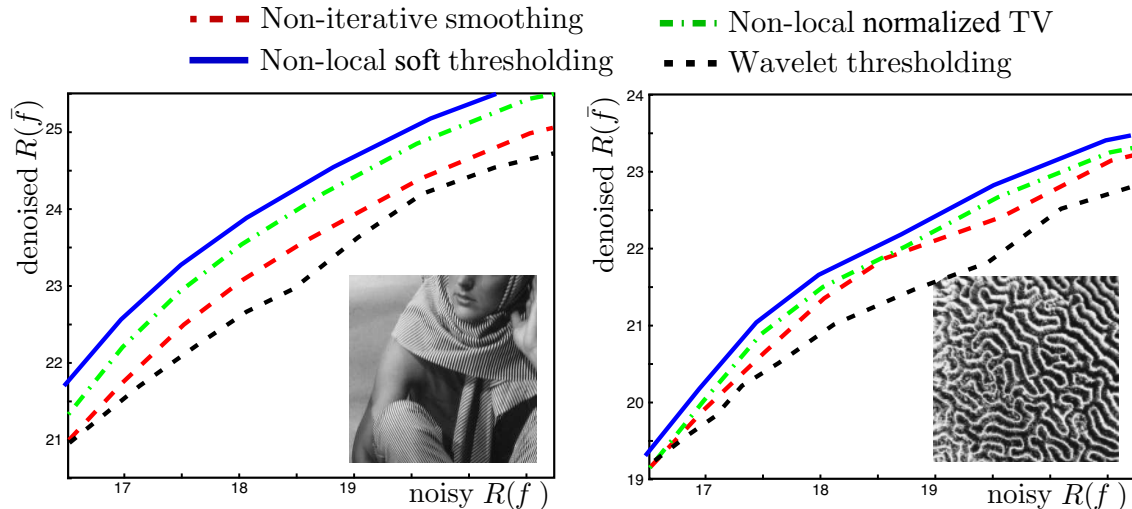

FIG. 7.3. *Risk $R(\bar{f})$ of several estimators $\bar{f}$ for a natural image and a texture for various noise levels $R(f)$.*

larization.

– Non-linear soft thresholding (section 6) is similar or better than non-local TV.

It is however important to note that all of these improvements come with an important computational overhead.

Iterative methods such as the manifold heat diffusion, section 4.2, require typically of the order of 20 iterations to give satisfactory results and non-linear regularization such as manifold BV, section 5.2, requires even more iterations. Thus, These methods are much slower than a single filtering with non-local means.

A major bottleneck of the spectral thresholding described in section 6 with respect to the other approaches is the need to extract eigenvectors from the Laplacian operator. This can be somehow alleviated by applying the method to overlapping sub-blocks in the image and using sparse operators, as detailed in section 6. The main advantage of thresholding over PDE and regularization methods of sections 4 and 5 is that it does not require to finely tune a stopping time
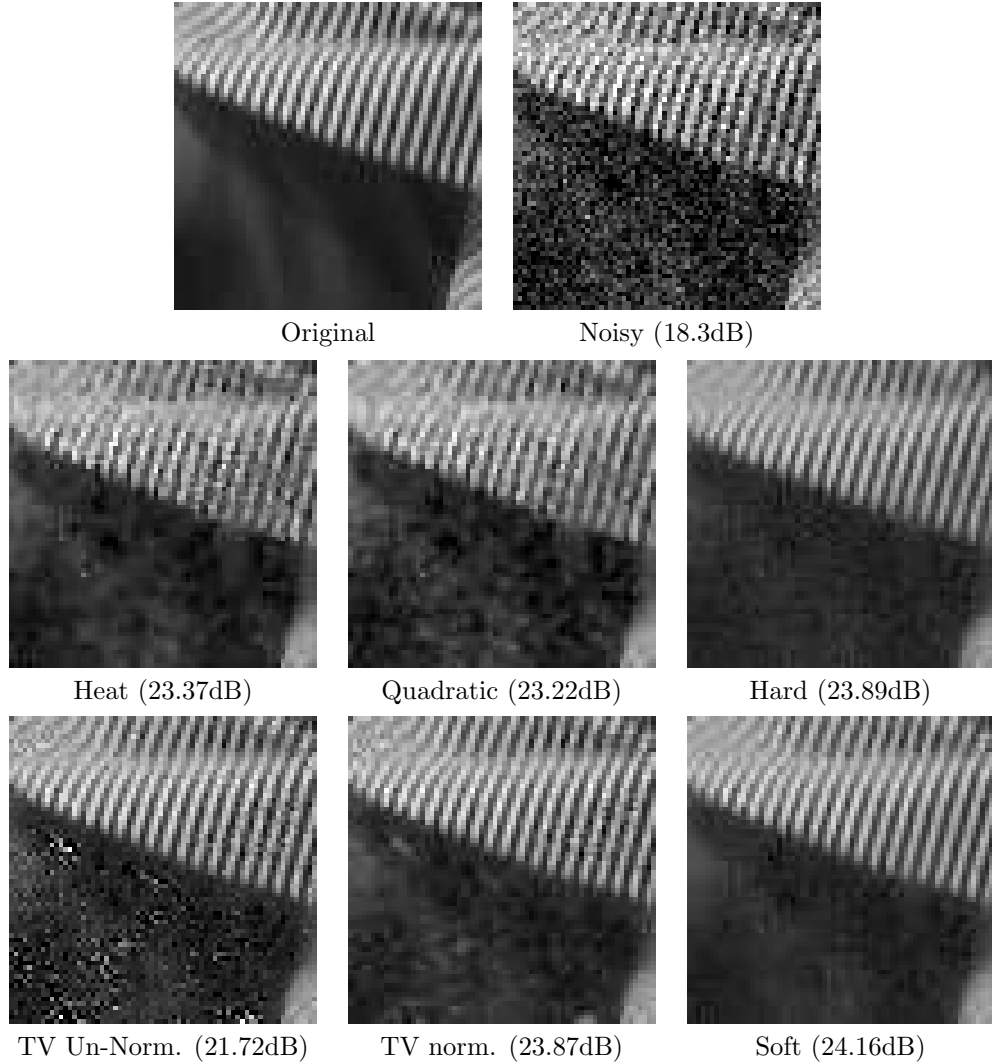
Original   Noisy (18.3dB)

Heat (23.37dB)   Quadratic (23.22dB)   Hard (23.89dB)

TV Un-Norm. (21.72dB)   TV norm. (23.87dB)   Soft (24.16dB)

FIG. 7.4. *Risk $R(\bar{f})$ of several estimators $\bar{f}$ for a natural image and a texture for various noise levels $R(f)$. The soft and hard thresholding are applied over the non-local manifold spectral basis.*

or a regularization parameter t.

7.2. Non-linear Approximation. To further understand the superiority of the non-local denoising process, we exploit the connexion between non-linear estimation and approximation in an orthogonal basis. This property of orthogonal decomposition is used extensively to obtain minimax optimality results for thresholding estimators, see [24]. Here we also exploit this feature of the spectral estimator to explore the behavior of non-local approximations and estimations.

M-terms approximation. The non-linear approximation of an image f with M coefficients in an orthogonal basis $\{u_\omega\}_\omega$ is written as

$$f_M \overset{\text{def.}}{=} \sum_{|\langle f, u_\omega \rangle| > T} \langle f, u_\omega \rangle u_\omega \qquad \text{with} \qquad M \overset{\text{def.}}{=} \text{Card}\left\{\omega \setminus |\langle f, u_\omega \rangle| > T\right\},$$

The approximation error is then:

$$\|f - f_M\|^2 = \sum_{|\langle f, u_\omega \rangle| \leqslant T} |\langle f, u_\omega \rangle|^2.$$

One wishes to use a basis $\{u_\omega\}_\omega$ that gives the best representation of data f drawn from an image ensemble $f \in \Theta \subset \ell^2(\Lambda)$. Optimizing the representation is then equivalent to maximizing the decay of the error $\|f - f_M\|^2$ when M increases, for all $f \in \Theta$. Asymptotically, one looks for bases $\{u_\omega\}_\omega$ such that $\|f - f_M\|^2 = O(M^{-\alpha})$ for the largest possible α.

Non-linear estimation and approximation. Donoho and Johnstone [24] proved that if the threshold is set to $T \stackrel{\text{def.}}{=} |\varepsilon| \sqrt{2 \log(n)}$ then the decay of the average risk $E(\|f^0 - \bar{f}\|^2)$ is linked to the decay of the non-linear approximation in the basis $\{u_\omega\}_\omega$. More precisely they show that

$$\|f^0 - f_M^0\|^2 \leqslant C_1 \, M^{-\alpha} \qquad \Longrightarrow \qquad E(\|f^0 - \bar{f}\|^2) \leqslant C_2 \, |\varepsilon|^{\frac{2\alpha}{\alpha+1}}. \qquad (7.1)$$

where C_1 and C_2 are constants that depend only on f^0. This result is valid only for a fixed basis that does not depend on the noisy input f.

Extension to orthogonal dictionaries. Results similar to (7.1) can be proved for dictionaries of orthogonal bases such as the wavelet and cosine packets [17] or the bandlets [35]. Such a dictionary is written in abstract form as $\mathcal{D} = \{\mathcal{B}(\lambda)\}_{\lambda \in \Gamma}$ where λ is a parameter that specifies the orthogonal basis $\mathcal{B}(\lambda)$ of $\mathbb{R}^n$ one wants to use to process a given function. In practice, one looks for a best basis $\mathcal{B}(\lambda^\star)$ in order to denoise or compress some input f. This is usually done by minimizing some Lagrangian cost function [17] that enforces both good approximation in $\mathcal{B}(\lambda^\star)$ and sparsity of the expansion in $\mathcal{B}(\lambda^\star)$. In order for the equivalence (7.1) between estimation and approximation to hold in this more general setting, one needs to have a control over the total number of basis vectors in the dictionary $\mathcal{D}$. Although the total number of bases $\#\Gamma$ might grow exponentially with the dimension n, the total number of vectors has to grow only in a polynomial manner, i.e.

$$\# \bigcup_{\lambda \in \Gamma} \mathcal{B}(\lambda) \leqslant C n^\beta \qquad (7.2)$$

for some constants C and β, see [34, 25] for studies of denoising in a dictionary of orthogonal bases.

The spectral bases constructed in section 2.3 also form a dictionary $\mathcal{D} = \{\mathcal{B}(\mathcal{M})\}_{\mathcal{M} \in \Gamma}$ of orthogonal bases, where each $\mathcal{M}$ is a discrete manifold (either non-local or semi-local). In practice, it is difficult to control the number of such graphs $\#\Gamma$ for a set of natural images. This number can be very large if one does not enforce strong constraints on the regularity of typical images, but the averaging implied by the width τ of the patch somehow alleviates this problem. Taking large patches is implicitly adding a strong regularization to the computation of the graph which avoids over-fitting and leads to good denoising performance. Our construction of adaptive non-local dictionaries thus does not come with constraint such as (7.2) but it is an important avenue for future work to explore the theoretical behavior of these estimators on realistic image models such as those considered below.

Theoretical bounds on non-linear approximation. Equation (7.1) shows that assessing the quality of the estimation process $f \mapsto \bar{f}$ for the different kinds of orthogonal bases amounts to quantify the decay of the approximation error $\|f - f_M\|^2$ for the images ensembles one would like to study.

– For uniformly smooth images, both Fourier and wavelet bases behave similarly and the approximation error for a C^α image satisfies $\|f - f_M\|^2 = O(M^{-\alpha})$. Since the semi-local and non-local basis vectors are in fact very similar to the Fourier vectors in this case, all the approximations behave similarly.

– A good model for images with regular edges is composed of images that are C^α outside a set of C^α edge curves, see [22]. This is close to the cartoon image model considered in section 1.3. For this ensemble of geometric images, the Fourier expansion behaves baddly because of the presence of step discontinuities and satisfies $\|f - f_M\|^2 = O(M^{-1/2})$. In contrast, the wavelet expansion is able to make use of the bounded total variation of such a function and its error satisfies $\|f - f_M\|^2 = O(M^{-1})$, see [37]. This is however sub-optimal and other bases such as wedgelets [22], curvelets [10] and bandlets [35, 39] are able to improve this approximation rate by exploiting the regularity of edge curves. One might expect that non-local adapted bases also capture this asymptotic decay of the error, thus providing an optimal tool to denoise cartoon images.

– A simple model for turbulent textures is an oscillating function with locally parallel level sets. Such a function can be written as $f(x) = \cos(\Phi(x))$ where Φ is a slowly varying phase function, see [20]. For this ensemble of locally parallel textures, the wavelets expansion fails to capture the oscillating regularity. Oscillating atoms such as the wave atoms [20] are able to capture such a behavior and enjoy a better approximation rate. Non-local processing methods also seems to perform well on these locally parallel textures, as shown on figure 7.2.

– For more complex images ensembles containing textures, there is no available scheme with provable approximation rate. Non-local denoising algorithms seems however to process efficiency several kinds of repetitive texture patterns [8].

In the following we perform a numerical estimation of the approximation rate in order to have information on the relative efficiency of processing methods.

Numerical estimation of the approximation error. Although our non-local dictionary of spectral bases does not satisfy the polynomial grow (7.1), we use the approximation error $\|f - f_M\|$ as a measure of the efficiency of this adaptive estimation scheme. In order to have a fair comparison, the non-local bases are computed from a noisy image $f + \varepsilon$. The noise level $|\varepsilon|^2$ is set to $n^{-1}\|f - f_M\|^2$ where f_M is the M-terms approximation of f in the wavelet basis.

Figure 7.5 shows the approximation error $\|f - f_M\|^2$ of f in various orthogonal bases. It is important to remember that the adapted basis for the semi-local and non-local Laplacians are estimated from a noisy observation $f + \varepsilon$ and not from the original image f that is not available in a denoising scenario. One can see that the non-local approximation is consistently better than the other approximation schemes. This reveals the superiority of non-local denoising methods to process images whose structures are not well captured by classical expansions such as Fourier or wavelets. Note however that both the semi-local and the non-local expansions are not suitable for applications to image compression since the adapted basis $\{u_\omega\}_\omega$ is not available for the decoder.
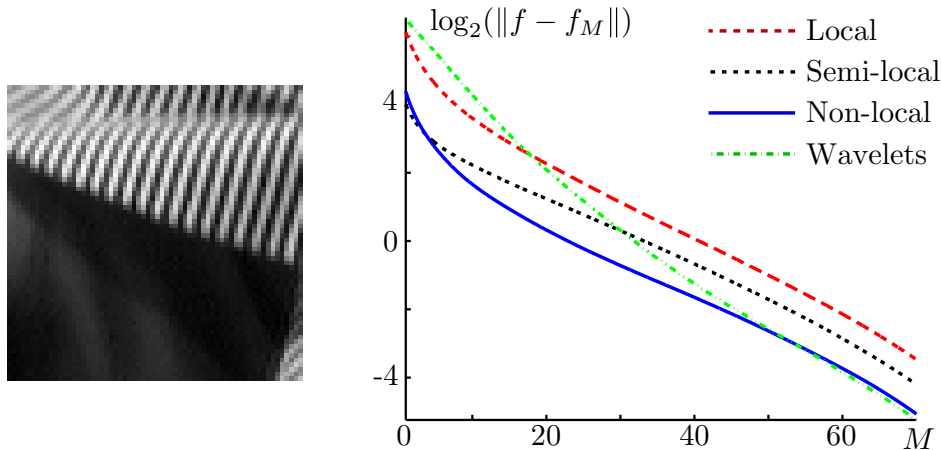

FIG. 7.5. *Left: original image f. Right: error decay $\log_2(\|f - f_M\|)$ of the approximation f_M of f with M coefficients (expressed in % of the total number of coefficients) in various orthogonal bases.*

Conclusion. This article reviewed several recent approaches to image estimation using a regularizer computed on a discrete manifold. This allows one to consider both local, semi-local and non-local methods in an unifying framework. The spectral bases associated to the Laplacians on these discrete manifolds define adaptive orthogonal decompositions that are exploited to build nonlinear thresholding operators. The numerical experiments show that this new class of estimators performs at least as good as state of the art diffusions and regularizations on non-local graphs. It also enables an exploration of semi-local and non-local methods using approximation in adaptive orthogonal bases.

REFERENCES

[1] M. AHARON AND R. KIMMEL, *Representation analysis and synthesis of lip images using dimensionality reduction*, Accepted to International Journal of Computer Vision, 67 (2006).

[2] Z. BAI, J. DEMMEL, J. DONGARRA, A. RUHE, AND H. VAN DER VORST, eds., *Templates for the solution of algebraic eigenvalue problems*, SIAM, Philadelphia, PA, 2000.

[3] S. BAKER, S.K. NAYAR, AND H. MURASE, *Parametric feature detection*, International Journal of Computer Vision, 27 (1998), pp. 27–50.

[4] D. BARASH, *A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation*, IEEE Trans. Pattern Analysis and Machine Intelligence, 24 (2002), pp. 844–847.

[5] M. BELKIN AND P. NIYOGI, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Computation, 15 (2003), pp. 1373–1396.

[6] M. Belkin, P. Niyogi, and V. Sindhwani, *Manifold regularization: A geometric framework for learning from labeled and unlabeled examples*, J. Mach. Learn. Res., 7 (2006), pp. 2399–2434.

[7] T. Brox and D. Cremers, *Iterated nonlocal means for texture restoration*, in Proc. International Conference on Scale Space and Variational Methods in Computer Vision, F. Sgallari, A. Murli, and N. Paragios, eds., vol. 4485 of LNCS, Ischia, Italy, May 2007, Springer, pp. 13–24.

[8] A. Buades, B. Coll, and J. M. Morel, *On image denoising methods*, SIAM Multiscale Modeling and Simulation, 4 (2005), pp. 490–530.

[9] ———, *Neighborhood filters and PDE's*, Numerische Mathematik, 105 (2006), pp. 1–34.

[10] E. Candès and D. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Comm. Pure Appl. Math., 57 (2004), pp. 219–266.

[11] G. Carlsson and V. de Silva, *Topological approximation by small simplicial complexes*, Preprint, (2006).

[12] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *On the local behavior of spaces of natural images*, to appear in Int. Journal of Comp. Vision, (2007).

[13] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.

[14] A. Chambolle, R.A. Devore, N.Y. Lee, and B.J. Lucier, *Nonlinear wavelet image-processing: Variational-problems, compression, and noise removal through wavelet shrinkage*, IEEE Tr. on Image Proc., 7 (1998), pp. 319–335.

[15] F. R. K. Chung, *Spectral graph theory*, Regional Conference Series in Mathematics, American Mathematical Society, 92 (1997), pp. 1–212.

[16] 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*, Proceedings of the National Academy of Science, 102 (2005), pp. 7426–7431.

[17] R. R. Coifman and M. V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Transactions on Information Theory, 38 (1992), pp. 713–718.

[18] D. Comaniciu and P. Meer, *Mean shift: A robust approach toward feature space analysis*, IEEE Trans. Pattern Analysis and Machine Intelligence, 24 (2002), pp. 603–619.

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse 3d transform-domain collaborative filtering*, IEEE Transaction on Image Processing, 16 (2007).

[20] L. Demanet and L. Ying, *Wave atoms and sparsity of oscillatory patterns*, to appear in Applied and Computational Harmonic Analysis, (2007).

[21] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr, *Implicit fairing of irregular meshes using diffusion and curvature flow*, in Proc. of SIGGRAPH '99, New York, NY, USA, 1999, ACM Press/Addison-Wesley Publishing Co., pp. 317–324.

[22] D. Donoho, *Wedgelets: Nearly-minimax estimation of edges*, Ann. Statist, 27 (1999), pp. 353–382.

[23] D.L. Donoho and C. Grimes, *Image manifolds isometric to euclidean space*, J. Math. Imaging Computer Vision, 23 (2005).

[24] D. Donoho and I. Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[25] C. Dossal, E. Le Pennec, and S. Mallat, *Denoising with bandelets*, Preprint, (2007).

[26] A. A. Efros and T. K. Leung, *Texture synthesis by non-parametric sampling*, in ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2, IEEE Computer Society, 1999, p. 1033.

[27] M. Elad, *On the origin of the bilateral filter and ways to improve it*, IEEE Trans. Image Processing, 11 (2002), pp. 1141–1151.

[28] A. Elmoataz, O. Lezoray, and S. Bougleux, *Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing*, Preprint, (2007).

[29] G. Gilboa, J. Darbon, S. Osher, and T.F. Chan, *Nonlocal convex functionals for image regularization*, UCLA CAM Report 06-57, (2006).

[30] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, SIAM Multiscale Modeling and Simulation, 6 (2007), pp. 595–630.

[31] M. Hein, J-Y. Audibert, and U. von Luxburg, *From graphs to manifolds – weak and strong pointwise consistency of graph Laplacians*, in Learning Theory: 18th Annual Conference on Learning Theory, COLT 2005, Bertinoro, Italy, June 27-30, 2005, Proceedings, 2005, pp. 470–485.

[32] P. S. Huggins and S. W. Zucker, *Representing edge models via local principal component analysis*, Proc. of ECCV'02, Springer LNCS 2350, (2002), pp. 384–398.

[33] S. Kindermann, S. Osher, and P. W. Jones, *Deblurring and denoising of images by nonlocal functionals*, SIAM Multiscale Modeling and Simulation, 4 (2005), pp. 1091–1115.

[34] H. Krim, D. Tucker, S. Mallat, and D. Donoho, *On denoising and best signal representation*, IEEE Tr. on Information Theory, 45 (1999), pp. 319–335.

[35] E. Le Pennec and S. Mallat, *Bandelet Image Approximation and Compression*, SIAM Multiscale Modeling and Simulation, 4 (2005), pp. 992–1039.

[36] A.B. Lee, K.S. Pedersen, and D. Mumford, *The nonlinear statistics of high-contrast patches in natural images*, International Journal of Computer Vision, 54 (2003), pp. 83–103.

[37] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.

[38] S. Mallat, *Geometrical grouplets*, to appear in Applied and Computational Harmonic Analysis, (2008).

[39] S. Mallat and G. Peyré, *Orthogonal bandelets for geometric image approximation*, Preprint, (2006).

[40] W. Matusik, M. Zwicker, and F. Durand, *Texture design using a simplicial complex of morphable textures*, ACM Trans. Graph., 24 (2005), pp. 787–794.

[41] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1990), pp. 629–639.

[42] G. PEYRÉ, *Manifold models for signals and images*, preprint, (2007).

[43] ———, *Texture synthesis and modification with a patch-valued wavelet transform*, Proc. SSVM 2007, (2007).

[44] K. POLTHIER, *Computational aspects of discrete minimal surfaces*, Proc. of the Clay Summer School on Global Theory of Minimal Surfaces, (2002).

[45] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[46] A. SINGER, *From graph to manifold laplacian: the convergence rate*, Applied and Computational Harmonic Analysis, 21 (2006), pp. 135–144.

[47] S. M. SMITH AND J. M. BRADY, *SUSAN - a new approach to low level image processing*, International Journal of Computer Vision, 23 (1997), pp. 45–78.

[48] N. A. SOCHEN, R. KIMMEL, AND A. M. BRUCKSTEIN, *Diffusions and confusions in signal and image processing*, Journal of Mathematical Imaging and Vision, 14 (2001), pp. 195–209.

[49] A. SPIRA, R. KIMMEL, AND N. SOCHEN, *A short time beltrami kernel for smoothing images and manifolds*, IEEE Trans. Image Processing, 16 (2007), pp. 1628–1636.

[50] A. SZLAM, M. MAGGIONI, AND R. R. COIFMAN, *A general framework for adaptive regularization based on diffusion processes on graphs*, Yale technichal report, (2006).

[51] C. TOMASI AND R. MANDUCHI, *Bilateral filtering for gray and color images*, in ICCV '98: Proceedings of the Sixth International Conference on Computer Vision, 1998, p. 839.

[52] D. TSCHUMPERLÉ AND R. DERICHE, *Vector-valued image regularization with PDEs: Acommon framework for different applications*, IEEE Trans. Pattern Anal. Mach. Intell, 27 (2005), pp. 506–517.

[53] M. WAKIN, D. DONOHO, H. CHOI, AND R. BARANIUK, *The Multiscale Structure of Non-Differentiable Image Manifolds*, in Optics and Photonics, San Diego, CA, July 2005.

[54] J. WANG, X. TONG, S. LIN, M. PAN, C. WANG, H. BAO, B. GUO, AND H-Y. SHUM, *Appearance manifolds for modeling time-variant appearance of materials*, ACM Trans. Graph., (2006).

[55] J. WEICKERT, *Coherence-enhancing diffusion filtering*, Int. J. Comput. Vision, 31 (1999), pp. 111–127.

[56] G. XU, *Discrete laplace-beltrami operators and their convergence*, Comput. Aided Geom. Des., 21 (2004), pp. 767–784.

[57] L. P. YAROSLAVSKY, *Digital Picture Processing — an Introduction*, Springer, Berlin, 1985.

[58] D. ZHOU AND B. SCHOLKOPF, *Regularization on discrete spaces*, in German Pattern Recognition Symposium, 2005, p. 361.