Dimension Reduction and Remote Sensing Using Modern Harmonic Analysis

John J. Benedetto and Wojciech Czaja

May 13, 2013

1 Introduction

Dimension reduction (DR) refers to removing data from a given set, in such a way as to have the possibility of extricating desired intelligent messages embedded in that set, by making an appropriate analysis of the remaining smaller set. *Remote sensing* is a means of collecting data about objects without coming in contact with them. These two ideas have been quantified in many ways; and they are fundamentally related because, in many applications, data sets obtained by remote sensing are often too large for effective understanding and computation without first engaging in dimension reduction. In this regard, ideas and techniques from *modern harmonic analysis* have played a critical role in accounting for significant advances in the field. The *theme* of this chapter is to exposit the role of modern harmonic analysis in understanding the relationship between dimension reduction and remote sensing.

Typically, the setting for remote sensing is the earth's surface and atmosphere, and the concept of remote sensing is associated with garnering information dealing with geology, meteorology, oceans and glaciers, natural disasters, climate change, and the classification and detection problems associated with man-made issues. Remote sensing technology is phenomenally varied. For example, airborne photography technology, going back to Tournachon's aerial photographs in 1858 from a balloon, has given rise to more recent satellite, RADAR, and LIDAR methodologies to collect data.

However, this technology comes with certain limitations. One of them is the sheer volume of data that arises in some applications, e.g., in the analysis of hyperspectral imaging data, and which underscores the need for computationally efficient and reliable DR methods.

There is a natural *evolution* of ideas we shall highlight to develop our theme: linear DR methods (Section 2), non-linear DR methods (Section 3), the theory of frames (Section 4), compressed sensing and sparse representation (Section 5), and diffusion-based image processing (Section 6). Each of these topics can be considered in the realm of DR, and each of them is a broad, deep, highly developed theory. In each section, except Section 2, we shall provide current applications in remote sensing associated with each of the topics. The material

of Section 2 is certainly important, but, for the purpose of this chapter, it is really an introduction to the subsequent topics, especially as a comparison to the power of non-linear DR in Section 3.

We see that the topics of the sections are an evolution in the following way. There are limitations of useful applicability of linear DR methods(Section 2), and we shall see in what ways non-linear DR methods (Section 3) can apply to complicated problems. Some of the most applicable DR methods are kernel eigenmap methods in which the eigenfunctions are orthogonal and the associated eigenvalues provide a means of achieving DR. Although this is an extraordinarily useful technology for many problems, it fails to be genuinely and naturally effective in a host of classification problems. In fact, different classes that should be identified are not necessarily orthogonal, and so ambiguity arises in orthogonal eigenfunction decompositions. This limitation can be resolved by introducing the theory of frames (Section 4).

Among their inherent properties, frames can be used to give robust signal decompositions, to ensure numerical stability in such decompositions, and to provide a means to mitigate noise. By "robust", we mean that the signal decomposition can retain its accuracy even when some of the atoms or frame elements are disabled. Of course, an essential feature of effective DR is to obtain useful signal decompositions in terms of a relatively small set of atoms. This can be done by a variety of non-linear quantization methods, e.g., $\Sigma\Delta$-modulation, that can be viewed as non-linear sampling. In recent years, this type of non-linear sampling has found its most effective identity by interleaving ideas ranging from pure analysis and probability theory to intricate finite dimensional versions of the uncertainty principle. The result is called compressed sensing and sparse representation (Section 5). Section 6 is not the "next step" in the evolution so much as a powerful mathematical machine in concert with the goals of compressed sensing and sparse representation, but with fundamental applications to image processing.

2 Linear dimension reduction methods

2.1 Principal Component Analysis

Linear dimension reduction methods treat remote sensing data as a collection of vectors $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^D$ in a given D-dimensional space, and their goal is to find a lower-dimensional subspace of X, that will represent the data optimally with respect to certain measures such as minimal reconstruction error, maximal preserved variance, maximal decorellation, or highest sparsity level..

One of the oldest and most popular dimension reduction techniques in remote sensing is *Principal Component Analysis* (PCA). PCA was introduced by Pearson [81], and was subsequently developed and generalized by many other authors, with major contributions coming from Hotelling, Karhunen, and Loève. Hence, it is often known under the names of Hotelling transform or Karhunen–Loève transform.

The major principle behind PCA is to assume that our observed variables are a result of an orthogonal transformation of the unknown latent variables

that minimizes the reconstruction error, or, equivalently, that yields maximally uncorrelated output. Among the most valued aspects of PCA are its simplicity and adaptability to a broad range of problems. However, PCA also has major drawbacks. Among them the biggest problem is the assumption of the linearity of observed data. This problem has been addressed in many recent works, where it has been proved that remote sensing data cannot be simply described as a linear subspace and that manifold models are necessary, see, e.g., [2].

2.2 Linear mixing models

A traditional method for dimension reduction in remote sensing data is by means of *endmember* extraction algorithms. Endmembers are defined as a collection of the scene's constituent spectra. If $E = \{e_i : i = 1, \ldots, s\}$ is the set of endmembers for the data set X, then the *linear mixture model* is defined as,

$$x_i = \sum_{j=1}^{s} \alpha_{i,j} e_j + N_{x_i},$$

for all $x_i \in X$, where N_{x_i} denotes a noise vector associated with x_i. The set $\{\alpha_{i,j} : i = 1, \ldots, N, j = 1, \ldots, s\}$ is the set of coefficients, and it is usually assumed that they satisfy the following two conditions: $\alpha_{i,j} \geq 0$, $i = 1, \ldots, N$, $j = 1, \ldots, s$, and $\sum_{j=1}^{s} \alpha_{i,j} = 1$, $i = 1, \ldots, N$.

Let $\alpha_{i,\cdot} = (\alpha_{i,1}, \ldots, \alpha_{i,s})$ and let $\tilde{\alpha} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_s)$. Let $\|y\|_2 = \sqrt{\sum_{k=1}^{D} |y_k|^2}$ be the ℓ^2 norm in $\mathbb{R}^D$.Two common endmember coefficient sets are given by,

$$\alpha_{i,\cdot} = \arg\min_{\tilde{\alpha}} \left\| x_i - \sum_{j=1}^{s} \tilde{\alpha}_j e_j \right\|_2$$

and

$$\alpha_{i,\cdot} = \arg\min_{\tilde{\alpha}} \left\| x_i - \sum_{j=1}^{s} \tilde{\alpha}_j e_j \right\|_2 + \tau_i \|\tilde{\alpha}\|_1,$$

subject to the above constraints of non-negativity and normalization, and where τ_i denotes a positive real number. Most endmember extraction algorithms determine E as a subset of X, i.e., it is assumed that the endmembers lie within the given data set. There are several endmember extraction algorithms, including N-FINDR [90], ORASIS [18] Pixel Purity Index [17], and Support Vector Data Description (SVDD) [3].

For an overview of these and other related linear techniques in dimension reduction of remote sensing data we refer to the excellent survey [63].

3 Non-linear dimension reduction methods

3.1 Overview

Sophisticated mathematical tools for the analysis of multispectral data have been successfully applied in recent years to detect and classify objects in areas ranging from human pathology to geophysics and satellite imaging. These

analysis techniques often take the form of dimension reduction (DR) or feature classification algorithms, see, e.g., [2]. Examples of state-of-the-art DR methods include Locally Linear Embedding (LLE) [82], Hessian LLE (HLLE) [55], Isomap [87], Laplacian Eigenmaps (LE) [6], Diffusion Maps [41], and Diffusion Wavelets [42].

Physical or experimental constraints often suggest that the intrinsic dimension of the multispectral dataset is much smaller than the number of bands originally used to represent the multispectral data. As such, one often assumes that the multispectral data points belong to a low-dimensional manifold $\mathcal{M}$.

In this context, we are most interested in an important subclass of the aforementioned DR methods known as *kernel eigenmap methods*. These include Kernel PCA [84], LLE [82], HLLE [55], and LE [6], among others [75]. Kernel eigenmap methods were introduced to address complexities not resolvable by linear techniques. Such methods recover the above mentioned data manifold by means of representations in terms of the significant eigenvectors of a data dependent kernel matrix. Generally speaking, these methods represent high-dimensional data in the form of a graph, with nodes formed by the data points treated as vectors, and with edges that represent the distances between pairs of such vectors. This information is stored in the so-called *adjacency* matrix, which is then modified to form the *kernel*. More specifically, the central idea underlying kernel eigenmap methods is to express correlations or similarities between vectors in the data space $X \subset \mathbb{R}^d$ in terms of kernel functions $K : X \times X \to \mathbb{R}^d$ which are symmetric and positive semi-definite. Once we have such K, we can construct a Hilbert space $\boldsymbol{K}$ and a mapping $\Phi : X \to \boldsymbol{K}$ such that

$$K(x, y) =< \Phi(x), \Phi(y) >$$

defines an inner product in $\boldsymbol{K}$. Classically, one diagonalizes this operator by the spectral theorem, and then chooses a smaller number of the most significant eigenvectors in order to achieve dimension reduction. Alternatively, we may choose the most significant and diverse eigenvectors for classification purposes.

3.2 Laplacian Eigenmaps

Many of the most popular dimension reduction techniques are built upon the mathematical foundation that is naturally expressed in the language of Laplacian Eigenmaps. These include LLE, HLLE, Diffusion Maps, Diffusion Wavelets, and Schroedinger Eigenmaps, see, e.g., [65]. Thus, for the sake of completeness, we start with a brief overview of the LE algorithm.

Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^D$. We assume that the points of X lie on a d-dimensional manifold $\mathcal{M}$, $d \leq D$, that reflects the local geometry of the data set X. The goal is to find a mapping

$$y : \mathbb{R}^N \to \mathbb{R}^n, \ d \leq n \leq D,$$

that emphasizes the local geometric structure of X. The LE algorithm [6], pioneered by Belkin and Niyogi, is divided into three steps.

1. Construction of a graph.

   Data points x_i, $1 = 1, \ldots, N$, form the nodes of the graph. An edge is drawn between x_i and x_j if x_i is among the k nearest neighbors of x_j as measured with respect to the Euclidean metric for some fixed k. This information is stored in an adjacency matrix G, with the corresponding entry $G_{i,j} = 1$. For all other pairs of nodes we let $G_{i,j} = 0$. This defines the graph, which we denote the same way as its adjacency matrix, viz., G. We assume that our graph is undirected, which is equivalent to assuming that G is symmetric .

2. Construction of the graph Laplacian.

   The edges of the graph are assigned weights and this information is stored in the weight matrix, W. A common choice of weights, motivated by the heat kernel, is the *diffusion weight matrix*:

$$W_{i,j} = \begin{cases} e^{-\frac{\|x_i - x_j\|_2^2}{\sigma}} & \text{if } G_{i,j} = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

   where $\sigma > 0$ denotes a positive constant. We then define the diagonal matrix D by $D_{i,i} = \sum_j W_{i,j}$. Finally the *graph Laplacian* is given by

$$L = D - W.$$

3. Eigendecomposition of the graph Laplacian.

   Find the mapping $y = \{y_1, \ldots, y_N\}^T$, $y_i \in \mathbb{R}^n$, by solving the minimization problem,

$$\text{argmin}_{y^T D y = I} \frac{1}{2} \sum_{i,j} \|y_i - y_j\|_2^2 W_{i,j}, \tag{3.2}$$

   which, in turn, is equivalent to solving the minimization problem,

$$\text{argmin}_{y^T D y = I} \, \text{trace}(y^T L y). \tag{3.3}$$

   Letting $z = D^{1/2} y$ in (3.3), it follows that the solution of the minimization problem (3.3) is equivalent to finding the first n solutions to the generalized eigenproblem, $Lv = \lambda D v$, sorted in increasing order of λ.

   It is clear that $v_0 = (1, 1, \ldots, 1)$ is a solution for $\lambda = 0$, see, e.g., [39]. If G is a connected graph, this solution is unique for $\lambda = 0$. Hence, the problem can be refined further by assuming that $v_0 \in \ker(y)$, i.e., we only look for eigenvectors corresponding to nonzero eigenvalues.

3.3 Schroedinger Eigenmaps

A major feature of Laplacian Eigenmaps and related methods is the preservation of local spectral distances, while reducing the overall dimension of the data represented by the aforementioned graph, cf., [39]. However, Laplacian Eigenmaps and similar DR methods lead to fully automated classification algorithms

which do not easily allow for expert input. To compensate for this deficiency, we note that the Laplace operator,

$$\Delta\Psi(x),$$

can be extended to the time-independent Schroedinger operator $\mathcal{E}$,

$$\mathcal{E}\Psi(x) = \Delta\Psi(x) + v(x)\Psi(x), \tag{3.4}$$

by adding a potential term $v(x)$.

The potential v in (3.4) is considered as a nonnegative multiplier operator. The discrete analogue of $\mathcal{E} = \Delta + v$ is the matrix $E = L + V$, where V is a nonnegative diagonal $N \times N$ potential matrix. We then replace the traditional Laplacian optimization problem (3.3) with the minimization problem,

$$\min_{y^\top Dy=I} \text{trace}(y^\top (L + \alpha V)y), \tag{3.5}$$

which, in turn, is equivalent to the minimization problem,

$$\min_{y^\top Dy=I} [\text{trace}(y^\top Ly) + \text{trace}(y^\top \alpha Vy)]. \tag{3.6}$$

The parameter $\alpha \geq 0$ is added here so that it can be used to emphasize the relative significance of the potential V with respect to the graph Laplace operator.

The minimization problem (3.6) is equivalent to the minimization problem,

$$\min_{y^\top Dy=I} \left\{ \tfrac{1}{2} \sum_{i,j} \|y_i - y_j\|^2 W_{i,j} + \alpha \sum_i V(i)\|y_i\|^2 \right\}, \tag{3.7}$$

where $V = \text{diag}(V(1), \ldots, V(N))$. The first sum in (3.7) incurs a penalty, when neighboring points x_i and x_j are mapped into points y_i and y_j, respectively, which are far apart. The second sum penalizes those points y_i, $i = 1, \ldots, N$, which are associated with large values of $V(i)$. For example, if V took only two values, 0 and 1, then the minimization (3.7) yields a dimension-reduced representation y, which forces increased clustering of the representations y_i of points associated with the value $V(i) = 1$, while attempting to ensure that close points remain close after the dimension reduction. As such we may utilize the potential V to *label* points which we would like to be identified together after the dimension reduction. Because of the built-in preservation of topology induced by the Laplacian, this labeling may be used to segment a particular class of points.

3.4 Properties

It is known that the rescaled graph Laplacian on G converges to the Laplace-Beltrami operator on the underlying manifold $\mathcal{M}$, e.g., [7], when we assume that all data points are connected. The Laplacian map on $X \subset \mathbb{R}^D$ is given by

$$L_N^\sigma f(x_i) = f(x_i) \sum_j e^{-\frac{\|x_i - x_j\|^2}{\sigma}} - \sum_j f(x_j) e^{-\frac{\|x_i - x_j\|^2}{\sigma}}.$$

Replacing the x_i by an arbitrary $x \in \mathbb{R}^D$ yields

$$L_N^\sigma f(x) = f(x) \sum_j e^{-\frac{\|x-x_j\|^2}{\sigma}} - \sum_j f(x_j) e^{-\frac{\|x-x_j\|^2}{\sigma}}.$$

It turns out that there exists a positive constant C such that for i.i.d. uniformly sampled data points $\{x_1, \ldots, x_N\}$ and $\sigma_N = N^{-\frac{1}{n+2+s}}$, where $s > 0$, and $f \in \mathcal{C}^\infty(\mathcal{M})$, we have the convergence,

$$\lim_{N\to\infty} C \frac{(\pi\sigma_N)^{-\frac{n+2}{2}}}{N} L_N^{\sigma_N} f(x) = \Delta_{\mathcal{M}} f(x), \tag{3.8}$$

in probability, see [7].

This convergence carries over to the Schroedinger operator as we shall see next. Let v be a given potential on the manifold $\mathcal{M}$. The associated matrix V acting on a discrete m-point set is defined as $V = \mathrm{diag}(v(x_1), \ldots, v(x_N))$. Since the potential does not depend on σ, we may replace x_i by an arbitrary $x \in \mathbb{R}^D$ to obtain the map,

$$V_m f(x) = v(x) f(x).$$

Clearly this extension coincides with the continuous potential on the manifold. As such, adding the discrete potential V_N to the discrete Laplacian does not impede the convergence in (3.8). Consequently, the term

$$C \frac{(\pi\sigma_N)^{-\frac{n+2}{2}}}{N} L_N^{\sigma_N} f(x) + V_N f(x) \tag{3.9}$$

converges for $N \to \infty$ to

$$\Delta_{\mathcal{M}} f(x) + v(x) f(x),$$

in probability, see [44], [68]. We note that (3.9) induces a specific choice of the parameter α. Indeed, in order to consider $E = L + \alpha V$, rescaling of $L_N^{\sigma_N}$ in (3.9) implies the need to reversely rescale V. This can be done by means of multiplication with $\alpha = \frac{1}{C} N (\pi\sigma_N)^{\frac{n+2}{2}}$, which converges to infinity as $N \to \infty$.

3.5 Applications to remote sensing

Although a mathematical novelty, kernel eigenmap methods associated with diffusion-based kernels have been used in various roles in applications to remote sensing. The non-linear nature of hyperspectral satellite imagery data (HSI), has been analyzed and verified by Bachmann et al. in [2]. This led to a wide range of applications of non-linear techniques. LLE has been employed for dimensionality reduction and vector segmentation by Mohan et al. in [79]. Castrodad used iterative schemes to perform semi-supervised multi-class classification and segmentation on HSI data, see [29]. Kernel fusion based on LE-derived kernels has been exploited for the purpose of spatial-spectral integration in [10], [58]. Kernel methods in remote sensing have also been combined with other mathematical techniques, such as Randomized Anisotropic Transforms, [37], or approximate graph constructions and randomized projections, [68]. For more

in-depth discussion and additional examples of kernel DR methods in remote sensing we refer to the work of Chui and Wang [38].

The *Schroedinger Eigenmaps algorithm* [43] allows experts, e.g., trained analysts, to introduce their input in the form of additional, labeled information to improve the detection and classification processes. This labeled information can take the form of a barrier potential for the associated Schroedinger operator on a graph. This added potential steers the diffusion process, induced by the Schroedinger operator on the data-dependent graph, according to both the dynamics of the labeled data and the geometry of the underlying graph. This is the major difference from the case of Laplacian Eigenmaps, where the diffusion process is determined by the geometry of the data alone, cf., [40]. In [9], the impact of Schroedinger Eigenmaps on classification is analyzed on multispectral and hyperspectral imagery. Efficient methods for building the potentials are based on expert ground-truth data and on automated clustering techniques, and it is shown that they lead to significant improvements in class separation [9], see Figure 1.

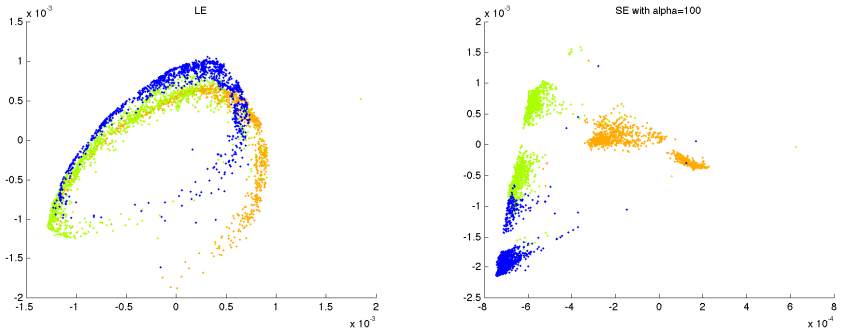

Figure 1: The effect of using Schroedinger potentials for decorrelating clusters as compared to traditional Laplacian embedding.

4 The theory of frames

4.1 Overview

Frames were introduced by Duffin and Schaeffer in 1952 [57]. However, their practical potential was not recognized until the 1990s. We refer the interested reader to other works for a more in-depth treatment of frames and their constructions and applications [14, 8, 27, 35, 71, 72]. Since then, frames were both generalized and specialized to allow for constructions of appropriately designed representation systems with varied features adapted to specific applications. Among the generalizations of frames, many ideas have been proposed in the recent years, e.g., frames of subspaces [28], pseudo-frames [76], fusion frames, oblique frames [36], outer frames [1], and multiplicative frames [15]. Finally, many of these constructions have been unified by an operator-based approach

called g-frames [86].

4.2 Frames

A *frame* for a Hilbert space $\mathcal{H}$ is a collection $\{x_i : i \in I\} \subset \mathcal{H}$ of vectors such that there exist constants $0 < A \leq B < \infty$ so that, for each $y \in \mathcal{H}$,

$$A\|y\|^2 \leq \sum_{i \in I} |\langle x_i, y\rangle|^2 \leq B\|y\|^2. \tag{4.1}$$

Constants A and B, which satisfy (4.1), are called *frame bounds* of $\{x_i : i \in I\}$. Optimally chosen values A and B are referred to as the *optimal frame bounds* of the frame. When $A = B$, the frame $\{x_i : i \in I\}$ is referred to as a *tight frame.*

As an example of a frame one may choose an orthonormal basis - it is in fact a tight frame with constants $A = B = 1$. A union of any two orthonormal bases is a tight frame with constants $A = B = 2$, etc. A union of an orthonormal basis with N arbitrary unit norm vectors is a frame with, not necessarily optimal, bounds $A = 1$ and $B = N + 1$ and, in general, it is not a tight frame.

Given a frame $\{x_i : i \in I\}$, a *dual frame* is a collection $\{x_i^* : i \in I\} \subset \mathcal{H}$ of vectors such that for all $x \in \mathcal{H}$, we have the reconstruction formula,

$$x = \sum_{i \in I} \langle x, x_i\rangle x_i^*.$$

Every frame possesses a dual frame. In order to obtain a dual frame to a given frame, we shall define the frame operator.

4.3 Frame operator

Let $\ell^2(I)$ denote the space of square summable sequences indexed by I. Given a frame $\{x_i : i \in I\}$, the *analysis operator* $\theta : \mathcal{H} \to \ell^2(I)$ is defined as

$$\theta(x) = \big(\langle x, x_i\rangle\big)_{i \in I}.$$

The adjoint of the analysis operator θ^* is called the *synthesis operator,* and $S = \theta^*\theta$ is the *frame operator.* The synthesis operator satisfies the equation,

$$\theta^*(c) = \sum_{i \in I} c_i x_i,$$

where c is any finitely supported sequence in $\ell^2(I)$. The following results are well known, e.g., [35].

Theorem 4.1. *Let $\{x_i : i \in I\} \subset \mathcal{H}$ be a frame for $\mathcal{H}$. Then the following are satisfied:*
a. *θ is a bounded operator from $\mathcal{H}$ into $\ell^2(I)$.*
b. *θ^* extends to a bounded operator from $\ell^2(I)$ into $\mathcal{H}$.*
c. *θ and θ^* are adjoint operators of each other.*

Theorem 4.2. *Let $\{x_i : i \in I\} \subset \mathcal{H}$ be a frame for $\mathcal{H}$. The frame operator $S = \theta^*\theta$ maps $\mathcal{H}$ onto $\mathcal{H}$ and is a positive invertible operator satisfying $A \cdot Id \leq S \leq B \cdot Id$ and $B^{-1} \cdot Id \leq S^{-1} \leq A^{-1} \cdot Id$. In particular, $\{x_i : i \in I\}$ is a tight frame if and only if $S = A \cdot Id$.*

The sequence $\{S^{-1}x_i : i \in I\}$ of vectors in $\mathcal{H}$ is called the *canonical dual frame*, and it is a dual frame for $\{x_i : i \in I\}$, i.e., we have

$$x = \sum_{i \in I} \langle x, S^{-1}(x_i)\rangle x_i$$

and

$$x = \sum_{i \in I} \langle x, x_i\rangle S^{-1}(x_i),$$

where both sums converge unconditionally in $\mathcal{H}$. We note here that dual frames are not in general unique and this underlies the importance of the canonical dual frame. On the other hand, there are significant applications of frames where dual frames other than the canonical dual are critical, [74].

4.4 Parseval frames

For a particular given frame, it may not be easy to apply the procedure in the preceding paragraph to obtain a dual frame. One special case in which it is easy is that of Parseval frames. A *Parseval frame* is a tight frame consisting of unit norm vectors. For Parseval frames, we have that, for every $x \in \mathcal{H}$,

$$x = \sum_{i \in I} \langle x, x_i\rangle x_i. \tag{4.2}$$

In particular, Parseval frames are dual frames of themselves. For this reason, among others, Parseval frames are the best behaved of frames. The following theorem, which goes back to Naimark, who used different terminology, is the source of most of the basic, general properties of Parseval frames.

Theorem 4.3. *A collection $\{x_i : i \in I\} \subset \mathcal{H}$ of vectors in $\mathcal{H}$ is a Parseval frame for H if and only if there exist a Hilbert space $\mathcal{K}$ containing $\mathcal{H}$ as a closed subspace and an orthonormal basis $\{e_i : i \in I\}$ for $\mathcal{K}$ such that, for all $i \in I$, $Pe_i = x_i$, where P is the orthogonal projection onto $\mathcal{H}$.*

In finite dimensional Hilbert vector spaces, the notion of a frame becomes intuitively simple. Let $N \geq d$; $\{x_i : i = 1, \ldots, N\}$ be a frame for $\mathbb{F}^d$, where $\mathbb{F}$ denotes the field of real or complex numbers, if and only if it is a spanning system for $\mathbb{F}^d$. State of the art mathematical algorithms construct frames through minimization of frame potential energy functions on complex manifolds [13, 85].

4.5 Applications to remote sensing

Frame theoretic techniques are relatively new in remote sensing processing. However, as typical data collected in remote sensing experiments is far from being orthogonal, see Figure 2, these techniques find novel applications. In [12], [69] and [89], kernel eigenmap methods were used to map the high dimensional space X to a low dimensional feature space Y, and then a frame is constructed for Y, which plays the same role as endmembers play in linear mixture models.

An original idea for constructing 2D tight frames that provide a new way to analyze, visualize, and process data at multiple scales and directions was

proposed by Bosch et al. in [19]. This is achieved by employing the proper choice of functional components in directional signal representations to isolate directional information, while, at the same time, effectively characterizing the underlying variation. This new family of frames has been shown to be suitable for a range of geostatistical applications, including super resolution and image inpainting.

Examples of specifically constructed frames with built-in features have been utilized in remote sensing data processing, see, e.g., [11], [20], [62]. Olshausen and his collaborators [34] also used learned dictionaries which are effectively frames, in their work on improving the performance of supervised classification algorithms for HSI data.

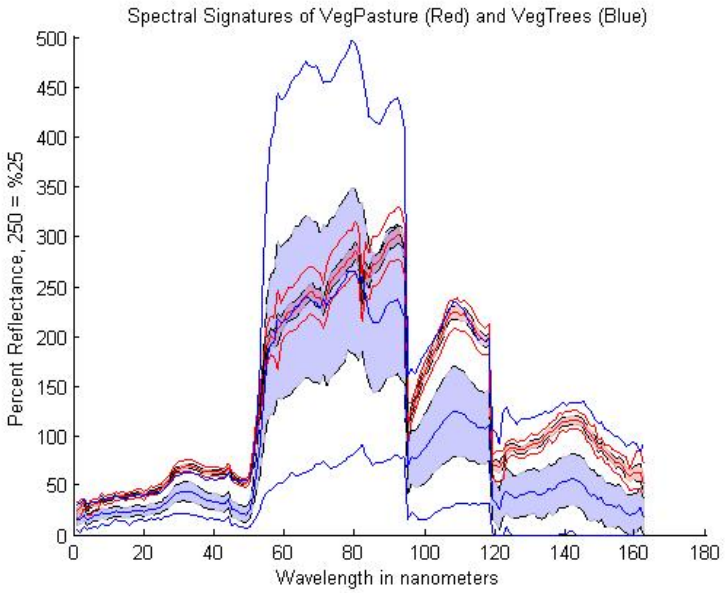

Figure 2: Spectral classes in HSI are typically non-orthogonal.

5 Sparse representation and compressed sensing

5.1 Overview

Since the introduction of multi-scale techniques in image analysis, there has been a strong motivation to provide sparse representations with the ability to detect edges, in particular, and directional content in general. As wavelets became a popular method to analyze multidimensional data, they fail to provide optimal n-term approximation rates for images with C^2 edges, see [22]. Therefore, a number of new representations have been introduced in an attempt to solve this problem. A few examples of these constructions are contourlets [48], curvelets [22], brushlets [78], wedgelets [53], shearlets [73], and composite wavelets [67].

A very different approach to induce sparsity in representation can be achieved by combining compression with sampling. This approach led to the introduction of one of the most fundamental models in data complexity reduction, that has been the focus of much recent attention - Compressed Sensing (CS). At its foundation is the concept of sparse signals. Given a basis for the ambient (potentially high-dimensional) space $\mathbb{R}^D$, a signal is called K-sparse if it can be represented using at most K nonzero coefficients. The theory of CS [25, 26, 23, 24, 54, 56] exploits this model in order to maintain a low-dimensional representation of the signal from which a faithful approximation to the original signal can be recovered efficiently. Dimensionality reduction in CS is linear and nonadaptive, i.e., the mapping does not depend on the data.

5.2 K-sparse signals

CS theory states that K-sparse signals $x \in \mathbb{R}^D$ can be recovered from $n < D$ linear measurements $y = \Phi x$, where Φ represents an $n \times D$ measurement matrix. This can be achieved via the following recovery algorithm,

$$\min_{\tilde{x} \in \mathbb{R}^D} \|\tilde{x}\|_1 \quad \text{subject to} \quad \Phi(\tilde{x}) = y, \tag{5.1}$$

where $\|x\|_1 = \sum_{j=1}^{D} |x_j|$.

Naturally, as $n < N$, this recovery cannot be obtained with just any sensing matrix Φ. Hence, we consider here matrices Φ which satisfy the *Restricted Isometry Property (RIP) of order K*, that is, matrices for which there exists a constant $\delta_K \in (0, 1)$, such that

$$(1 - \delta_K)\|x\|_2^2 \leq \|\Phi(x)|_2^2 \leq (1 + \delta_K)\|x\|_2^2,$$

for every K-sparse vector $x \in \mathbb{R}^D$. Under this assumptions, Candès proved the following theorem in [21].

Theorem 5.1. *Let x^* be the solution of the minimization problem (5.1). Let x_K denote the best K-sparse approximation to $x \in \mathbb{R}^D$. Assume that $\delta_{2K} < \sqrt{2} - 1$. Then, there exists $C > 0$ such that,*

$$\|x^* - x\|_1 \leq C\|x - x_K\|_1$$

and

$$\|x^* - x\|_2 \leq CK^{-1/2}\|x - x_K\|_1.$$

Theorem 5.1 clearly implies that for K-sparse vectors x the recovery in (5.1) is exact. However, two types of questions now follow. One is how to find matrices Φ which satisfy RIP. The other is what is the fewest number of measurements we can afford to take and still recover the signal. While it is quite difficult to satisfactorily answer these questions in a deterministic manner, statistical concepts proved to be much easier to deal with. As such, one can assert that, with high probability, every K-sparse signal $x \in \mathbb{R}^D$ can be recovered from just $n = O(K \log(D/K))$ measurements $y = \Phi x$, and the measurement matrix Φ is

an $n \times D$ measurement matrix drawn randomly from an acceptable probabilistic distribution. This includes random sampling matrices Φ which have i.i.d. Bernoulli, Gaussian, or uniform entries, see Section 5.3. We note that RIP is just one of several ways to provide such conditions, see [88]. We note that the number of samples n is linear at the "information level", i.e., with respect to K, and is logarithmic in terms of the ambient dimension D. The number of random samples n is typically taken large enough to ensure that all K-sparse signals remain well-separated when embedded in $\mathbb{R}^n$. CS theory applies equally well to signals that are not strictly sparse but compressible, i.e., if the coefficients in the signal's representation decay fast enough. Furthermore, near optimal recovery is guaranteed even in the presence of noise, e.g., [21].

5.3 Random projections

The notion of using a random projection for dimensionality reduction is not new. In fact, it can be traced back to long before the present wave of interest in CS. One fundamental result in which this type of dimension reduction manifests itself is the *Johnson-Lindenstrauss Lemma* (JL) [70], cf., [46], where one can use a random projection for a stable embedding of a finite data set, effectively providing dimension reduction.

Lemma 5.2 (Johnson-Lindenstrauss)**.** *Given $0 < \epsilon < 1$, a set X of N points in $\mathbb{R}^D$, and a number $n \geq O(\ln D)/\epsilon^2$, there is a Lipschitz function $f : \mathbb{R}^D \to \mathbb{R}^n$ such that, for all $u, v \in X$,*

$$(1-\epsilon)\|u-v\| \leq \|f(u)-f(v)\| \leq (1+\epsilon)\|u-v\|.$$

The statement of JL is completely deterministic. Surprisingly, probability enters in two different but related ways. On the one hand, the technique of proof of JL depends on concentration of measure inequalities, see [46], [4]. On the other hand, the result itself is very useful from the perspective of CS. In [5] a fundamental connection was identified between CS theory and JL, despite the fact that the former allows for the embedding of an uncountable number of points. This connection allows us, in particular, to answer the question about construction of matrices which satisfy RIP due to the following theorem [5].

Theorem 5.3. *Let D, n, and δ be given. Assume that entries of the matrix Φ are independent realizations of a probability distribution satisfying*

$$\Pr(|\|\Phi(x)\|_{\ell_2^n}^2 - \|x\|_{\ell_2^D}^2| \geq \epsilon\|x\|_{\ell_2^D}^2) \leq 2e^{-nC(\epsilon)}, \quad \epsilon \in (0,1).$$

Then, there exist constants $C_1, C_2 > 0$ such that Φ satisfies RIP with δ and any $K \leq C_1 n/\log(D/K)$ with probability exceeding $1 - 2e^{-C_2 n}$.

We note that computing random projections is relatively inexpensive: projecting N points from D to n dimensions costs $O(DnN)$.

Manifold models generalize the notion of sparsity beyond bases. These models arise whenever a signal in $\mathbb{R}^D$ is a continuous function of a K-dimensional parameter. For example, a pure sinusoid is completely determined by its amplitude, phase, and frequency. So a class of signals consisting of pure sinusoids

would form a three-dimensional manifold in $\mathbb{R}^D$. The dimension of the manifold under this model is analogous to the sparsity level in the CS model. In [4] the authors extend CS theory by demonstrating that random linear projections can be used to map the high-dimensional manifold-modeled data to a low-dimensional space while, with high probability, approximately preserving all pairwise distances between the points.

5.4 Applications to remote sensing

CS has many promising applications in signal acquisition, compression, and medical imaging. Among them, new possibilities arise for using CS in remote sensing.

Based on CS concepts, Patel et al. introduced a new synthetic aperture radar (SAR) imaging modality which can provide a high-resolution map of the spatial distribution of targets and terrain, while using a significantly reduced number of waveforms [80].

Deloye et al. [47] analyze the coded aperture snapshot spectral imager (CASSI) system, which is a class of imaging spectrometers that provide implementation of compressive sensing ideas for hyperspectral imaging.

In the context of HSI data, Greer [66] shows that standard theoretical guarantees do not apply to the performance of classical CS reconstruction algorithms such as orthogonal matching pursuit (OMP) and basis pursuit (BP). He introduces a new algorithm, sparse demixing (SD), and proves its optimality in reconstruction sparsity and accuracy.

6 Diffusion-based image processing

6.1 Overview

Diffusion-based methods have been extensively used in image analysis, both as self-contained techniques [61], [16], and as tools for approximating the total variation (TV) functional, see, e.g., [30] and [83]. Further, even earlier, wavelet-type systems appeared in the context of variational problems, e.g., [33], [60]. In particular, [33] discusses the TV minimization in the wavelet domain that successfully reproduces lost coefficients. A related work, [60], describes an algorithm for filling in holes in overlapping texture and cartoon image layers by means of a direct extension of the image decomposition method called Morphological Component Analysis (MCA). The relationship between wavelet-based image processing algorithms and variational problems is analyzed in [31]. Shearlet-based TV minimization utilized for image denoising is studied in [59].

6.2 Ginzburg-Landau energy

The *Ginzburg-Landau (GL) energy functional* [64], [32],

$$GL(u) = \frac{\epsilon}{2}\int |\nabla u(x)|^2 dx + \frac{1}{4\epsilon}\int W(u)dx, \tag{6.1}$$

$$W(u) = (u^2-1)^2,$$

may be considered as a diffuse interface approximation to the TV functional $\int |\nabla u| dx$ for the case of binary images. Based on this concept, several efficient

algorithms for deconvolution, image inpainting, superresolution, and other applications have been proposed, see, e.g., [51], [49], [52]. The common approach has been to use modifications of the GL functional as the primary regularizer for the solution of an ill-posed problem. In those works $\int |\nabla u(x)|^2 dx$ has been replaced with wavelet-based, shearlet-based, and related semi-norms, with the goal of removing the "fuzzy" diffuse interface features and utilizing advantages of sparse directional representations. Characterizing signal regularity in terms of the decay of wavelet coefficients via a Besov semi-norm, cf., [77], allows one to construct a method with properties similar to PDE-based methods but without an ϵ-scale blur.

6.3 Composite wavelet Ginzburg-Landau energy

Our goal is to explore the possibilities arising from constructions of frames with directional content, such as, e.g., curvelets, shearlets, or, more generally, composite wavelets. For this purpose, let $\psi_{i,\gamma} = D_{a^{-i}} L_\gamma \psi$ denote the dilation D and the generalized shift L_γ, associated with a composite wavelet construction, for parameters $i \in \mathbb{Z}$ and $\gamma \in \Gamma$. We do not specify the choice of the parameter set Γ, as it depends on the choice of the directional multiscale representation, and we refer the reader to [45] for details. For more information on the structure of composite wavelets and their theoretical underpinnings we refer to [67]. For any $u \in L^2([0,1]^2)$ we define the *composite wavelet seminorm* as

$$|u|_{CW}^2 = \sum_{i=0}^{\infty} |\det a|^i \sum_{\gamma \in \Gamma} |\langle u, \psi_{i,\gamma} \rangle|^2.$$

We define the *composite wavelet GL energy* (CWGL),

$$CWGL(u) = \frac{\epsilon}{2} |u|_{CW}^2 + \frac{1}{4\epsilon} \int W(u) dx.$$

Then the "composite wavelet Allen-Cahn equation", i.e., the gradient descent minimization equation for $CWGL$, is

$$u_t = \epsilon \Delta_{CW} u - \frac{1}{\epsilon} W'(u),$$

where

$$\Delta_{CW} u = -\sum_{i=0}^{\infty} |\det a|^i \sum_{\gamma \in \Gamma} \langle u, \psi_{i,\gamma} \rangle \psi_{i,\gamma}$$

is the "composite wavelet Laplace" operator. This means that we are replacing the regular Laplace operator that appears in the classical Allen-Cahn equation and that expresses the gradient descent minimization of GL energy.

On the one hand, the minimizers of CWGL energy exhibit properties similar to those in the classical diffuse interface model. On the other hand, the dependence on ϵ is different for CWGL. In fact, ϵ defines the dominant wavelet scales in the decomposition of the minimizer, see [51]; and, since the wavelet functions are well-localized, increasing ϵ causes the low phase-transition blur, which provides advantages in reconnecting edges over large gaps without losing the image sharpness.

6.4 Applications to remote sensing

Consider an image f as a function on $[0,1]^2$, and let $\Omega \subset [0,1]^2$ be the domain where we assume this image to be known. The goal of *image inpainting* (data recovery) is to recover the values $f(x)$ for $x \in \Omega^C$, the complement of Ω. In this model, the GL-type energy term plays the role of a regularizer, while the forcing term is expressed as the L^2 norm between the minimizer u and the known image f on the known domain. In the following formula we consider composite wavelet GL energy and seminorm; however, without loss of generality, these can be replaced with other types of GL energy functionals. Define

$$E(u) = CWGL(u) + \frac{\mu}{2}\|u - f\|^2_{L^2(\Omega)}.$$

One recovers the complete image as the minimizer of this modified functional E. In order to find this minimizer, it is necessary to consider it as a stable state solution of the respective gradient descent equation,

$$u_t = \epsilon \Delta_{CW} u - \frac{1}{\epsilon} W'(u) - \mu \mathbf{1}_\Omega (u - f),$$

where $\mathbf{1}_\Omega$ is the characteristic function of the known portion of the domain.

In a series of works [49], [50], [51] Bertozzi and Dobrosotskaya studied a wavelet analogue of the Ginzburg-Landau energy with an additional edge-preserving forcing term. Their applications include inpainting, superresolution, segmentation, denoising, and contour detection, and they have been used for example in partial road classification and inpainting for satellite imagery.

In [52] this work has been extended to allow for utilizing the directional content based on shearlet representations.

In [45] the above approach is utilized in an algorithm to recover missing data, due to sparsity of composite wavelet representations, especially when compared to inpainting algorithms induced by traditional wavelet representations, cf., Figure 3.

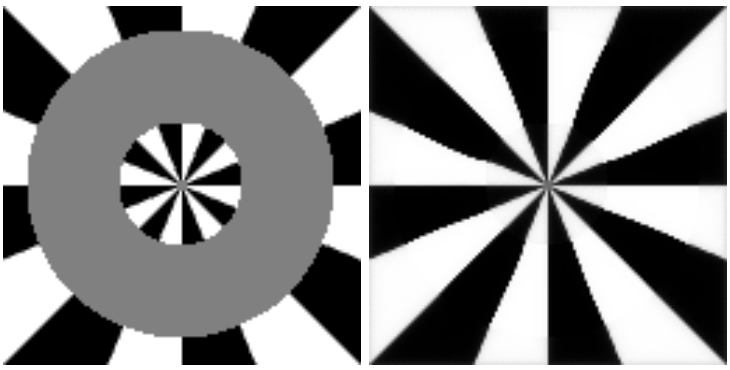

Figure 3: Data recovery by means of CWGL algorithm with post processing.

References

[1] A. Aldroubi, C. Cabrelli, and U. Molter, "Wavelets on irregular grids with arbitrary dilation matrices and frame atoms for $L^2(\mathbb{R}^d)$, Appl. Comput. Harmon. Anal., vol. 17, no. 2, pp. 11–140, 2004.

[2] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pp. 441–454, 2005.

[3] A. Banerjee, P. Burlina, and J. Broadwater, "A machine learning approach for finding hyperspectral endmembers," IEEE International Geoscience and Remote Sensing Symposium, 2007, pp. 3817–3820, 2007.

[4] R. G. Baraniuk and M. B. Wakin, "Random Projections of Smooth Manifolds, Foundations of Computational Mathematics," vol. 9, no. 1, pp. 51–77, 2009.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices, Constructive Approximation," vol. 28, no. 3, pp. 253–263, 2008.

[6] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for dimensionality reduction and data representation," Neural. Comput., vol. 15, no. 6, pp. 1373–1396, 2003.

[7] M. Belkin and P. Niyogi, "Towards a theoretical foundation for Laplacian-based manifold methods," J. Comput. System Sci., vol. 74, no. 8, pp. 1289–1308, 2008.

[8] J. J. Benedetto, "Frame decompositions, sampling and uncertainty principle inequalities," in Wavelets: Mathematics and Applications, J. Benedetto, M. Frazier, Eds., CRC Press, Boca Raton, FL, 1994, pp. 247–304.

[9] J. J. Benedetto, W. Czaja, J. Dobrosotskaya, T. Doster, K. Duke, D. Gillis, "Semi-supervised learning of heterogeneous data in remote sensing imagery," in Independent Component Analyses, Compressive Sampling, Wavelets, Neural Net, Biosystems, and Nanoengineering X, Proc. SPIE, vol. 8401, 8401-03, 2012.

[10] J. J. Benedetto, W. Czaja, J. Dobrosotskaya, T. Doster, K. Duke, D. Gillis, "Integration of heterogeneous data for classification in hyperspectral satellite imagery," in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, Proc. SPIE, vol. 8390, 8390-78, 2012.

[11] J. J. Benedetto, W. Czaja, M. Ehler, C. Flake, and M. Hirn, "Wavelet packets for multi- and hyper-spectral imagery", in Wavelet Applications in Industrial Processing VII, Proc. SPIE, vol. 7535, 7535-08, 2010.

[12] J. J. Benedetto, W. Czaja, J. C. Flake, and M. Hirn, "Frame based kernel methods for automatic classification in hyperspectral data," in IEEE IGARSS, Cape Town, South Africa, 2009.

[13] J. J. Benedetto and M. Fickus, "Finite normalized tight frames," Adv. Comp. Math., vol. 18, pp. 357–385, 2003.

[14] J. J. Benedetto and D. Walnut, "Gabor frames for L^2 and related spaces," in Wavelets: Mathematics and Applications, J. Benedetto, M. Frazier, Eds., CRC Press, Boca Raton, FL, 1994, pp. 97–162.

[15] J. J. Benedetto and M. Dellomo,"Reactive sensing and multiplicative frames," preprint, 2013.

[16] A. Bertozzi, S. Esedoglu, and A. Gillette, "Analysis of a two-scale Cahn-Hilliard model for image inpainting," Multiscale Modeling and Simulation, vol. 6, no. 3, pp. 913–936, 2007.

[17] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of aviris data", Fifth JPL Airborne Earth Science Workshop, vol. 1 of JPL Publication 95-1, pp. 23–26, 1995.

[18] J. Bowles, P. Palmadesso, J. Antoniades, M. Baumbeck, and L. Rickard, "Use of filter vectors in hyperspectral data analysis", Proceedings of SPIE, vol. 2553, pp. 148–157, 1995.

[19] E. H. Bosch, A. Castrodad, J. S. Cooper, W. Czaja, and J. Dobrosotskaya, "Multiscale and multidirectional tight frames for image analysis," to appear in Proc. SPIE, vol. 8750, 2013.

[20] E. H. Bosch, A. González, J. Vivas, and G. Easley, "Directional wavelets and a wavelet variogram for two-dimensional data", Mathematical Geosciences, vol. 41, no. 6, 611–641, 2009.

[21] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," Compte Rendus de l'Academie des Sciences, vol. 346, pp. 589–592, 2008.

[22] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise-C^2 singularities," Comm. Pure Appl. Math., vol. 57, pp. 219–266, 2002.

[23] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inform. Theory, vol. 52, pp. 489–509, 2006.

[24] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comm. Pure Appl. Math., vol. 59, pp. 1207–1223, 2006.

[25] E. J. Candès and T. Tao, "Decoding by linear programming," IEEE Trans. Inform. Theory, vol. 51, pp 4203–4215, 2005.

[26] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies," IEEE Trans. Inform. Theory, vol. 52, pp. 5406–5425, 2006.

[27] P. Casazza, "The art of frame theory, arXiv preprint math/9910168, 1999.

[28] P. Casazza and G. Kutyniok, "Frames of subspaces, in Wavelets, Frames and Operator Theory, Contemp. Math., vol. 345, pp. 87–113, 2003.

[29] A. Castrodad, "Graph-based denoising and classification of hyperspectral imagery using nonlocal operators," in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV, Proc. SPIE, vol. 7334, 7334-0E, 2009.

[30] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," Numerische Mathematik, vol. 76, pp. 167–188, 1997.

[31] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," IEEE Transactions on Image Processing, vol. 7, no. 3, pp. 319–333, 1998.

[32] T. F. Chan and J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods, SIAM, 2005.

[33] T. F. Chan, J. Shen, and H.-M. Zhou, "Total variation wavelet inpainting," Journal of Mathematical Imaging and Vision vol. 25, 2006.

[34] A. S. Charles, B. A. Olshausen, C. J. Rozell, "Learning Sparse Codes for Hyperspectral Imagery", Selected Topics in Signal Processing, vol. 5, no., 5, pp. 963–978, 2011.

[35] O. Christensen, "An Introduction to Frames and Riesz Bases," Birkhauser, Boston, MA, 2003.

[36] O. Christensen and Y. Eldar, "Oblique dual frames and shift-invariant spaces, Appl. Comput. Harmon. Anal., vol. 17, no. 1, pp. 48–68, 2004.

[37] C. K. Chui and J. Wang, "Randomized anisotropic transform for nonlinear dimensionality reduction," *International Journal on Geomathematics*, vol. 1, no. 1, pp. 23–50, 2010.

[38] C. K. Chui and J. Wang, "Dimensionality reduction of hyper-spectral imagery data for feature classification," in Handbook of Geomathematics, Volume 1, W. Freeden, M. Z. Nashed, and T. Sonar, Eds., pp. 1005–1048, 2010.

[39] F. R. K. Chung, "Spectral Graph Theory," CBMS Regional Conference Series in Mathematics, vol. 92, 1997.

[40] R. R. Coifman and S. Lafon, "Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions," Appl. Comput. Harmon. Anal., vol. 21, no. 1, pp. 31–52, 2006.

[41] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. J. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data. part i: Diffusion maps," Proc. Nat. Acad. Sci., vol. 102, pp. 7426–7431, 2005.

[42] R. R. Coifman and M. Maggioni, "Diffusion wavelets," Appl. Comput. Harmon. Anal., vol. 21, no. 1, pp. 53–94, 2006.

[43] W. Czaja and M. Ehler, "Schroedinger eigenmaps for the analysis of biomedical data", IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 5, pp. 1274–1280, 2013.

[44] W. Czaja and A. Halevy, "On convergence of Schroedinger eigenmaps," preprint, 2011.

[45] W. Czaja, J. Dobrosotskaya, and B. Manning, "Composite wavelet representations for reconstruction of missing data", to appear in Proc. SPIE, vol. 8750, 2013.

[46] S. Dasgupta and A. Gupta, "An elementary proof of the Johnson-Lindenstrauss lemma," Technical report 99-006, UC Berkeley, 1999.

[47] C. J. Deloye, J. C. Flake, D. Kittle, E. H. Bosch, R. S. Rand, D. J. Brady, "Exploitation Performance and Characterization of a Prototype Compressive Sensing Imaging Spectrometer", in Excursions in Harmonic Analysis, vol. 1, Applied and Numerical Harmonic Analysis, pp. 151–171, 2013.

[48] M. N. Do and M. Vetterli, "Contourlets: A Directional Multiresolution Image Representation," Proc. of IEEE International Conference on Image Processing (ICIP), Rochester, 2002.

[49] J. Dobrosotskaya and A. Bertozzi, "A Wavelet-Laplace variational technique for image deconvolution and inpainting," IEEE Transactions on Image Processing," vol. 17, no. 5, 2008.

[50] J. Dobrosotskaya and A. Bertozzi, "Wavelet analogue of the Ginzburg-Landau energy and its Gamma-convergence," Interfaces and Free Boundaries, vol. 12, no. 2, pp. 497–525, 2010.

[51] J. Dobrosotskaya and A. Bertozzi, "Analysis of the wavelet Ginzburg-Landau energy in image applications with edges," SIAM J. Imaging Sci., vol. 6, no. 1, pp. 698–729, 2013.

[52] J. Dobrosotskaya and W. Czaja, "Shearlet Ginzburg-Landau energy, its Gamma convergence and applications," preprint, 2013.

[53] D. L. Donoho, "Wedgelets: nearly minimax estimation of edges," Ann. Statist., vol. 27, no. 3, pp. 859–897, 1999.

[54] D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[55] D. L. Donoho and C. Grimes, "Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data," Proc. Nat. Acad. Sci., vol. 100, pp. 5591–5596, 2003.

[56] D. Donoho and J. Tanner, "Sparse nonnegative solutions of underdetermined linear equations by linear programming," Proc. National Academy of Sciences, vol. 102, no. 27, pp. 9446–9451, 2005.

[57] R. J. Duffin and A. C. Schaeffer , "A class of nonharmonic Fourier series," Trans. Amer. Math. Soc., vol. 72, pp. 341–366, 1952.

[58] K. Duke, "A Study of the Relationship Between Spectrum and Geometry Through Fourier Frames and Laplacian Eigenmaps", Ph.D. Thesis, University of Maryland College Park, 2012.

[59] G. R. Easley, D. Labate, and F. Colonna, "Shearlet based total variation for denoising," IEEE Trans. Image Process., vol. 18, no. 2, pp. 260–268, 2009.

[60] M. Elad, J. L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon texture image inpaitning using morphological component analysis," Appl. Comput. Harmon. Anal., vol. 19, pp. 340–358, 2005.

[61] H. Emmerich, Diffuse Interface Approach in Materials Science Thermodynamic Concepts and Applications of Phase-Field Models, Springer, 2003.

[62] J. C. Flake, "The Multiplicative Zak Transform, Dimension Reduction, and Wavelet Analysis of LIDAR Data", Ph.D. Thesis, University of Maryland College Park, 2010.

[63] D. Gillis and J. Bowles, "An Introduction to Hyperspectral Image Data Modeling", in Excursions in Harmonic Analysis, vol. 1, Applied and Numerical Harmonic Analysis, pp. 173–194, 2013.

[64] V. L. Ginzburg and L. D. Landau, Zh. Eksp. Teor. Fiz. vol. 20 (1064), 1950.

[65] Y. Goldberg, A. Zakai, D. Kushnir, and Y. Ritov, "Manifold Learning: The Price of Normalization," Journal of Machine Learning Research, vol. 9, pp. 1909–1939, 2008.

[66] J. B. Greer, "Hyperspectral Demixing: Sparse Recovery of Highly Correlated Endmembers", in Excursions in Harmonic Analysis, vol. 1, Applied and Numerical Harmonic Analysis, pp. 195–210, 2013.

[67] K. Guo, D. Labate, W.-Q. Lim, G. Weiss, and E. Wilson, "The theory of wavelets with composite dilations," in Harmonic Analysis and Applications, C. Heil, Ed., Applied and Numerical Harmonic Analysis, pp. 231–250, Birkhauser Boston, 2006.

[68] A. Halevy, "Extensions of Laplacian Eigenmaps for Manifold Learning", Ph.D. Thesis, University of Maryland College Park, 2011.

[69] M. Hirn, "Enumeration of Harmonic Frames and Frame Based Dimension Reduction", Ph.D. Thesis, University of Maryland College Park, 2009.

[70] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," Contemp. Math., vol. 26, pp. 189–206, 1984.

[71] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Parts I and II)" IEEE Signal Processing Magazine, vol. 24, no. 4, pp. 86–104, and vol. 24, no. 5, pp. 115–125, 2007.

[72] J. Kovačević and A. Chebira, "Introduction to frames," Foundations and Trends in Signal Processing, Now Publishers, vol. 2, no. 1, 2008.

[73] D. Labate, W. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," in Wavelets XI, SPIE Proc., vol. 5914, pp. 254–262, 2005.

[74] M. Lammers, A. Powell, and Ö. Yilmaz, "Alternative dual frames for digital-to-analog conversion in sigma-delta quantization," Advances in Computational Mathematics, vol. 32, no. 1, pp. 73–102, 2009.

[75] J. A. Lee and M. Verleysen, Nonlinear Dimensionality Reduction, Springer, 2007.

[76] S. Li and H. Ogawa, "Pseudoframes for subspaces with applications, Journal of Fourier Analysis and Applications, vol. 10, no. 4, pp. 409–431, 2004.

[77] S. Mallat, Wavelet Tour of Signal Processing, Academic Press, 1999.

[78] F. Meyer and R. Coifman, "Brushlets: A Tool for Directional Image Analysis and Image Compression", Appl. Comput. Harmon. Anal., vol. 4, pp. 147–187, 1997.

[79] A. Mohan, G. Sapiro, and E. Bosch, "Spatially Coherent Nonlinear Dimensionality Reduction and Segmentation of Hyperspectral Images," IEEE Geoscience and Remote Sensing Letters, vol. 4, no. 2, pp. 206–210, 2007.

[80] V. M. Patel, G. R. Easley, D. M. Healy, Jr. and R. Chellappa, "Compressed synthetic aperture radar," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 244–254, 2010.

[81] K. Pearson, "On lines and planes of closest fit to systems of points in space," Philosophical Magazine, vol. 2, no. 6, pp. 559–572, 1901.

[82] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[83] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation based noise removal algorithms",' Physica D., vol. 60, pp. 259–268, 1992.

[84] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, vol. 10, no., 5, pp. 1299–1319, 1998.

[85] N. Strawn, "Geometric Structures and Optimization on Spaces of Finite Frames", Ph.D. Thesis, University of Maryland College Park, 2011.

[86] W. Sun, "G-frames and g-Riesz bases," Journal of Mathematical Analysis and Applications, vol. 322, no. 1, pp. 437–452, 2006.

[87] V. Tenenbaum, J. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science, vol. 209, pp. 2319–2323, 2000.

[88] R. Wang, "Global Geometric Conditions on Dictionaries for the Convergence of L^1 Minimization Algorithms", Ph.D. Thesis, University of Maryland College Park, 2013.

[89] D. Widemann, "Dimensionality Reduction for Hyperspectral Data", Ph.D. Thesis, University of Maryland College Park, 2008.

[90] M. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data", Proc. SPIE, vol. 3753, pp. 266–275, 1999.