Concise Signal Models

By:
Michael Wakin

Concise Signal Models

By:
Michael Wakin

Online:
< http://cnx.org/content/col10635/1.4/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: September 14, 2009
PDF generated: February 5, 2011
For copyright and attribution information for the modules contained in this collection, see p. 47.

Table of Contents

Chapter 1

Introduction to Concise Signal Models[1]

1.1 Overview

In characterizing a given problem in signal processing, one is often able to specify a model for the signals to be processed. This model may distinguish (either statistically or deterministically) classes of interesting signals from uninteresting ones, typical signals from anomalies, information from noise, etc.

Very commonly, models in signal processing deal with some notion of structure, constraint, or conciseness. Roughly speaking, one often believes that a signal has "few degrees of freedom" relative to the size of the signal. This notion of conciseness is a very powerful assumption, and it suggests the potential for dramatic gains via algorithms that capture and exploit the true underlying structure of the signal.

In these modules, we survey three common examples of concise models: linear models, sparse nonlinear models, and manifold-based models. In each case, we discuss an important phenomenon: the conciseness of the model corresponds to a low-dimensional geometric structure along which the signals of interest tend to cluster. This low-dimensional geometry again has important implications in the understanding and the development of efficient algorithms for signal processing.

We discuss this low-dimensional geometry in several contexts, including projecting a signal onto the model class (i.e., forming a concise approximation to a signal), encoding such an approximation (i.e., data compression), and reducing the dimensionality of signals and data sets. We conclude with an important and emerging application area known as Compressed Sensing (CS), which is a novel method for data acquisition that relies on concise models and builds upon strong geometric principles. We discuss CS in its traditional, sparsity-based context and also discuss extensions of CS to other concise models such as manifolds.

1.2 General Mathematical Preliminaries

1.2.1 Signal notation

We will treat signals as real- or complex-valued functions having domains that are either discrete (and finite) or continuous (and either compact or infinite). Each of these assumptions will be made clear as needed. As a general rule, however, we will use x to denote a discrete signal in $\mathbb{R}^N$ and f to denote a function over a continuous domain $\mathcal{D}$. We also commonly refer to these as discrete- or continuous-**time** signals, though the domain need not actually be temporal in nature.

[1] This content is available online at <http://cnx.org/content/m18720/1.5/>.

1.2.2 Lp and lp norms

As measures for signal energy, fidelity, or sparsity, we will employ the L_p and ℓ_p norms. For continuous-time functions, the L_p norm is defined as

$$\| f \|_{L_p(\mathcal{D})} = \left(\int_{\mathcal{D}} |f|^p\right)^{1/p}, \quad p \in (0, \infty), \tag{1.1}$$

and for discrete-time functions, the ℓ_p norm is defined as

$$\| x \|_{\ell_p} = \begin{cases} \left(\sum_{i=1}^{N} |x(i)|^p\right)^{1/p}, & p \in (0, \infty), \\ \max_{i=1,\cdots,N} |x(i)|, & p = \infty, \\ \sum_{i=1}^{N} \mathbf{1}_{x(i)\neq 0}, & p = 0, \end{cases} \tag{1.2}$$

where $\mathbf{1}$ denotes the indicator function. (While we often refer to these measures as "norms," they actually do not meet the technical criteria for norms when $p < 1$.)

1.2.3 Linear algebra

Let A be a real-valued $M \times N$ matrix. We denote the **nullspace** of A as $\mathcal{N}(A)$ (note that $\mathcal{N}(A)$ is a linear subspace of $\mathbb{R}^N$), and we denote the **transpose** of A as A^T.

We call A an **orthoprojector** from $\mathbb{R}^N$ to $\mathbb{R}^M$ if it has orthonormal rows. From such a matrix we call $A^T A$ the corresponding **orthogonal projection operator** onto the M-dimensional subspace of $\mathbb{R}^N$ spanned by the rows of A.

Chapter 2

Signal Dictionaries and Representations[1]

For a wide variety of signal processing applications (including analysis, compression, noise removal, and so on) it is useful to consider the representation of a signal in terms of some dictionary [80]. In general, a **dictionary** Ψ is simply a collection of elements drawn from the signal space whose linear combinations can be used to represent or approximate signals.

Considering, for example, signals in $\mathbb{R}^N$, we may collect and represent the elements of the dictionary Ψ as an $N \times Z$ matrix, which we also denote as Ψ. From this dictionary, a signal $x \in \mathbb{R}^N$ can be constructed as a linear combination of the elements (columns) of Ψ. We write

$$x = \Psi\alpha \tag{2.1}$$

for some $\alpha \in \mathbb{R}^Z$. (For much of our notation in this section, we concentrate on signals in $\mathbb{R}^N$, though the basic concepts translate to other vector spaces.)

Dictionaries appear in a variety of settings. The most common may be the basis, in which case Ψ has exactly N linearly independent columns, and each signal x has a unique set of expansion coefficients $\alpha = \Psi^{-1}x$. The orthonormal basis (where the columns are normalized and orthogonal) is also of particular interest, as the unique set of expansion coefficients $\alpha = \Psi^{-1}x = \Psi^T x$ can be obtained as the inner products of x against the columns of Ψ. That is, $\alpha(i) =< x, \psi_i >, i = 1, 2, \cdots, N$, which gives us the expansion

$$x = \sum_{i=1}^{N} < x, \psi_i > \psi_i. \tag{2.2}$$

We also have that $\|x\|_2^2 = \sum_{i=1}^{N} < x, \psi_i >^2$.

Frames are another special type of dictionary [75]. A dictionary Ψ is a frame if there exist numbers A and B, $0 < A \leq B < \infty$ such that, for any signal x

$$A\|x\|_2^2 \leq \sum_z < x, \psi_z >^2 \leq B\|x\|_2^2. \tag{2.3}$$

The elements of a frame may be linearly dependent in general (see Figure 2.1), and so there may exist many ways to express a particular signal among the dictionary elements. However, frames do have a useful analysis/synthesis duality: for any frame Ψ there exists a dual frame $\tilde{\Psi}$ such that

$$x = \sum_z < x, \psi_z > \tilde{\psi}_z = \sum_z < x, \tilde{\psi}_z > \psi_z. \tag{2.4}$$

In the case where the frame vectors are represented as columns of the N x Z matrix Ψ, the matrix $\tilde{\Psi}$ containing the dual frame elements is simply the transpose of the pseudoinverse of Ψ. A frame is called

[1] This content is available online at <http://cnx.org/content/m18724/1.5/>.

tight if the frame bounds A and B are equal. Tight frames have the special properties of (i) being their own dual frames (after a rescaling by $1/A$) and (ii) preserving norms, i.e., $\sum_{i=1}^{N} < x, \psi_i >^2 = A\|x\|_2^2$. The remainder of this section discusses several important dictionaries.

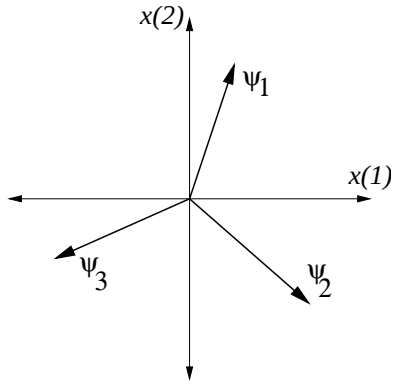


Figure 2.1: A simple, redundant frame Ψ containing three vectors that span $\mathbb{R}^2$.

2.1 The canonical basis

The standard basis for representing a signal is the canonical (or "spike") basis. In $\mathbb{R}^N$, this corresponds to a dictionary $\Psi = I_N$ (the $N \times N$ identity matrix). When expressed in the canonical basis, signals are often said to be in the "time domain."

2.2 Fourier dictionaries

The frequency domain provides one alternative representation to the time domain. The Fourier series and discrete Fourier transform are obtained by letting Ψ contain complex exponentials and allowing the expansion coefficients α to be complex as well. (Such a dictionary can be used to represent real or complex signals.) A related "harmonic" transform to express signals in $\mathbb{R}^N$ is the discrete cosine transform (DCT), in which Ψ contains real-valued, approximately sinusoidal functions and the coefficients α are real-valued as well.

2.3 Wavelets

Closely related to the Fourier transform, wavelets provide a framework for localized harmonic analysis of a signal [80]. Elements of the discrete wavelet dictionary are local, oscillatory functions concentrated approximately on dyadic supports and appear at a discrete collection of scales, locations, and (if the signal dimension $D > 1$) orientations.

2.3.1 Scale

In wavelet analysis and other settings, we will frequently refer to a particular **scale** of analysis for a signal. Consider, for example, continuous-time functions f defined over the domain $\mathcal{D} = [0, 1]^D$. A **dyadic**

hypercube $X_j \subseteq [0,1]^D$ at scale $j \in \mathbb{N}$ is a domain that satisfies

$$X_j = \left[\beta_1 2^{-j}, (\beta_1 + 1)\, 2^{-j}\right] \times \cdots \times \left[\beta_D 2^{-j}, (\beta_D + 1)\, 2^{-j}\right] \tag{2.5}$$

with $\beta_1, \beta_2, \cdots, \beta_D \in \{0, 1, \cdots, 2^j - 1\}$. We call X_j a **dyadic interval** when $D = 1$ or a **dyadic square** when $D = 2$ (see Figure 2.2). Note that X_j has sidelength 2^{-j}.

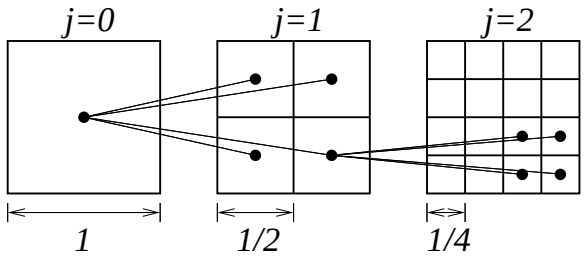


Figure 2.2: Dyadic partitioning of the unit square at scales $j = 0, 1, 2$. The partitioning induces a coarse-to-fine parent/child relationship that can be modeled using a tree structure.

For discrete-time functions the notion of scale is similar. We can imagine, for example, a "voxelization" of the domain $[0,1]^D$ ("pixelization" when $D = 2$), where each voxel has sidelength 2^{-B}, $B \in \mathbb{N}$, and it takes 2^{BD} voxels to fill $[0,1]^D$. The relevant scales of analysis for such a signal would simply be $j = 0, 1, \cdots, B$, and each dyadic hypercube X_j would refer to a collection of voxels.

2.3.2 Wavelet fundamentals

The wavelet transform offers a multiscale decomposition of a function into a nested sequence of scaling spaces $V_0 \subset V_1 \subset \cdots \subset V_j \subset \cdots$. Each scaling space V_j is spanned by a discrete collection of dyadic translations of a lowpass scaling function φ_j, and the difference between adjacent scaling spaces V_j and V_{j+1} is spanned by a discrete collection of dyadic translations of a bandpass wavelet function ψ_j. Figure 2.3 shows an example of this multiscale organization in the case of the Haar wavelet dictionary. Each wavelet function at scale j is concentrated approximately on some dyadic hypercube X_j, and between scales, both the wavelets and scaling functions are "self-similar," differing only by rescaling and dyadic dilation. When $D > 1$, the difference spaces are partitioned into $2^D - 1$ distinct orientations (when $D = 2$ these correspond to vertical, horizontal, and diagonal directions). The wavelet transform can be truncated at any scale j. We then let the basis Ψ consist of all scaling functions at scale j plus all wavelets at scales j and finer.

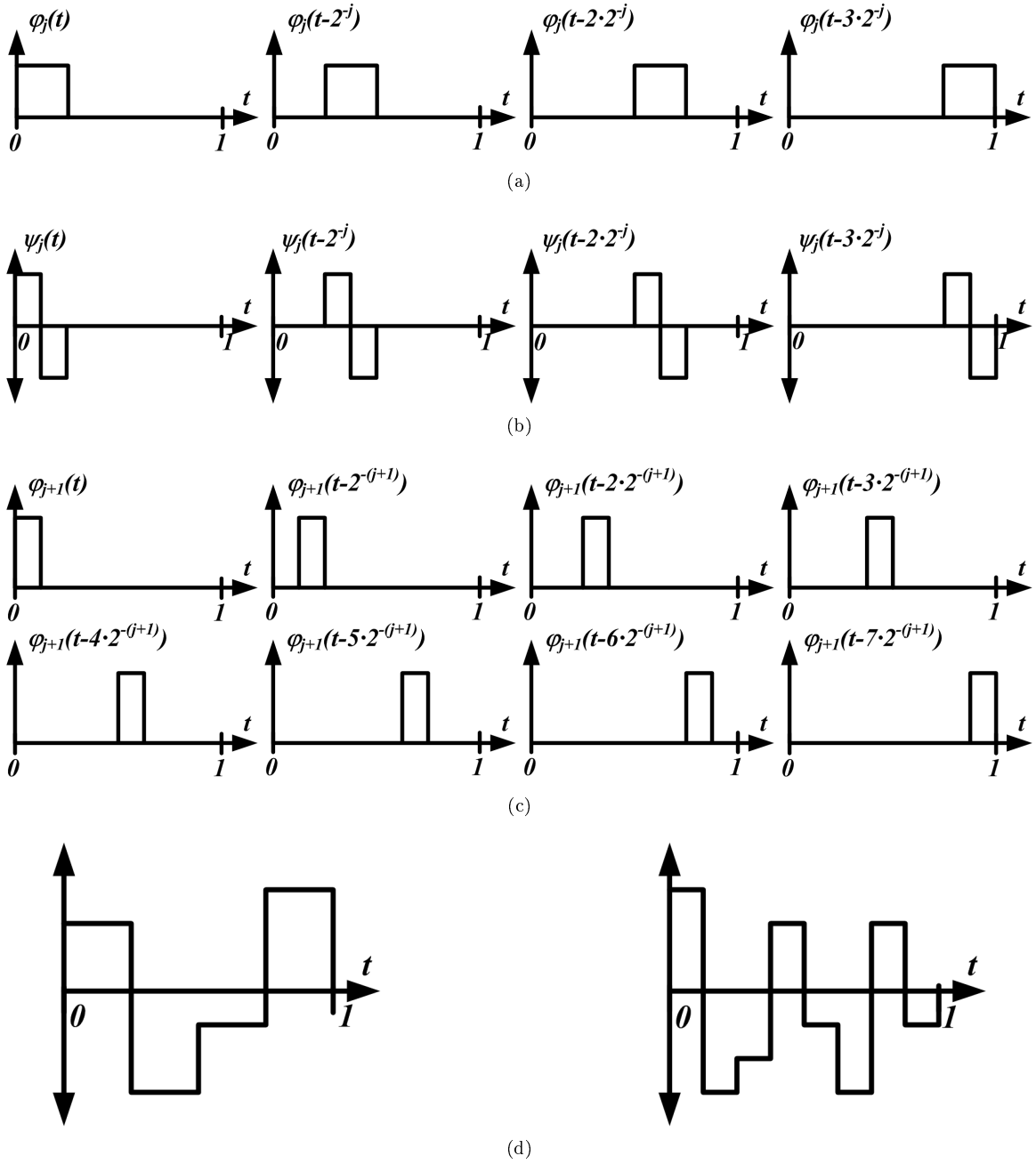


Figure 2.3: Multiscale wavelet representations on the interval $[0, 1]$. (a) Haar scaling functions spanning V_j with $j = 2$. (b) Haar wavelet functions spanning the difference space between V_j and V_{j+1}. (c) Haar scaling functions spanning V_{j+1}. (d) Two example functions belonging to the spaces (left) V_j and (right) V_{j+1}.

Wavelets are essentially bandpass functions that detect abrupt changes in a signal. The scale of a wavelet, which controls its support both in time and in frequency, also controls its sensitivity to changes in the signal. This is made more precise by considering the wavelet analysis of smooth signals. Wavelet are often characterized by their number of **vanishing moments**; a wavelet basis function is said to have H vanishing moments if it is orthogonal to (its inner product is zero against) any H-degree polynomial. Sparse (Nonlinear) models (Section 4.2: Sparse (nonlinear) models) discusses further the wavelet analysis of smooth and piecewise smooth signals.

The dyadic organization of the wavelet transform lends itself to a multiscale, tree-structured organization of the wavelet coefficients. Each "parent" function, concentrated on a dyadic hypercube X_j of sidelength 2^{-j}, has 2^D "children" whose supports are concentrated on the dyadic subdivisions of X_j. This relationship can be represented in a top-down tree structure, as demonstrated in Figure 2.2. Because the parent and children share a location, they will presumably measure related phenomena about the signal, and so in general, any patterns in their wavelet coefficients tend to be reflected in the connectivity of the tree structure. Figure 2.4 and Figure 2.5 show an example of the wavelet transform applied to the Cameraman test image; since the dimension $D = 2$, each scale is partitioned into vertical, horizontal, and diagonal wavelet analysis, and each parent coefficient has $2^D = 4$ children.

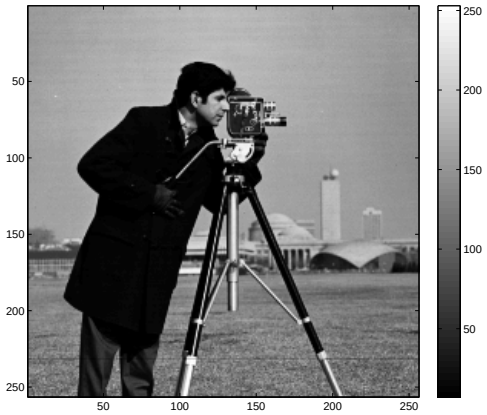
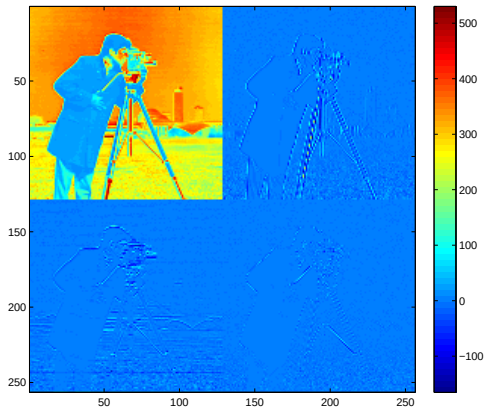
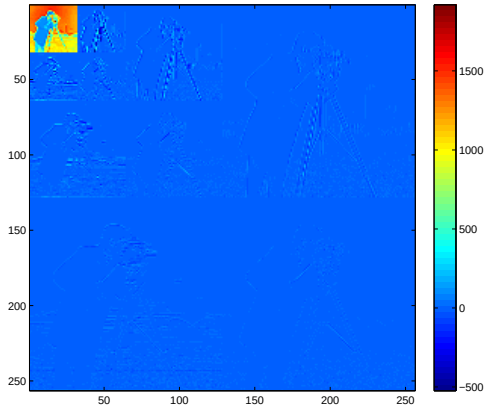


Figure 2.4: Cameraman test image (size 256×256) for use in wavelet decomposition and approximation examples.



(a)



(b)

Figure 2.5: Wavelet analysis of the Cameraman test image. (a) One-level wavelet transform, where the N-pixel image is transformed into four sets of $N/4$ coefficients each. The top left quadrant represents the scaling coefficients at the next coarser scale (relative to the scale of pixelization). The remaining quadrants represent the wavelet coefficients from the difference spaces, partitioned into the vertical, horizontal, and diagonal subbands. (b) Three-level wavelet transform, where the wavelet decomposition has been iterated twice more on the scaling coefficients. The multiple scales of wavelet coefficients exhibit a parent-child dependency. The largest coefficients tend to concentrate at the coarsest scales and around high-frequency features such as edges in the image.

In addition to their ease of modeling, wavelets are computationally attractive for signal processing; using a filter bank, the wavelet transform of an N-voxel signal can be computed in just $O(N)$ operations.

2.4 Other dictionaries

A wide variety of other dictionaries have been proposed in signal processing and harmonic analysis. As one example, complex-valued wavelet transforms have proven useful for image analysis and modeling [72], [73], [94], [65], [102], [91], [66], thanks to a phase component that captures location information at each

scale. Just a few of the other harmonic dictionaries popular in image processing include wavelet packets [80], Gabor atoms [80], curvelets [29], [18], and contourlets [50], [51], all of which involve various space-frequency partitions. We mention additional dictionaries in Compression (Chapter 6) .

Chapter 3

Manifolds[1]

As we will soon discuss, manifold models can provide an alternative to signal dictionaries as a framework for concise signal modeling. In this module, we present a minimal set of definitions and terminology from differential geometry and topology that serve as an introduction to manifolds. We refer the reader to the introductory and classical texts [90], [86], [68], [11] for more depth and technical precision.

3.1 General terminology

A K-dimensional manifold $\mathcal{M}$ is a topological space[2] that is locally homeomorphic[3] to $\mathbb{R}^K$ [68]. This means that there exists an open cover of $\mathcal{M}$ with each such open set mapping homeomorphically to an open ball in $\mathbb{R}^K$. Each such open set, together with its mapping to $\mathbb{R}^K$ is called a **chart**; the set of all charts of a manifold is called an **atlas**.

The general definition of a manifold makes no reference to an ambient space in which the manifold lives. However, as we will often be making use of manifolds as models for sets of signals, it follows that such "signal manifolds" are actually subsets of some larger space (for example, of $L_2(\mathbb{R})$ or $\mathbb{R}^N$). In general, we may think of a K-dimensional submanifold embedded in $\mathbb{R}^N$ as a nonlinear, K-dimensional "surface" within $\mathbb{R}^N$.

3.2 Examples of manifolds

One of the simplest examples of a manifold is simply the circle in $\mathbb{R}^2$. A small, open-ended segment cut from the circle could be stretched out and associated with an open interval of the real line (see Figure 3.1). Hence, the circle is a 1-D manifold. (We note that at least two charts are required to form an atlas for the circle, as the entire circle itself cannot be mapped homeomorphically to an open interval in $\mathbb{R}^1$.)

[1] This content is available online at <http://cnx.org/content/m18722/1.4/>.

[2] A **topological space** is simply a set X, together with a collection T of subsets of X called open sets, such that: (i) the empty set belongs to T, (ii) X belongs to T, (iii) arbitrary unions of elements of T belong to T, and (iv) finite intersections of elements of T belong to T.

[3] A **homeomorphism** is a function between two topological spaces that is one-to-one, onto, continuous, and has a continuous inverse.

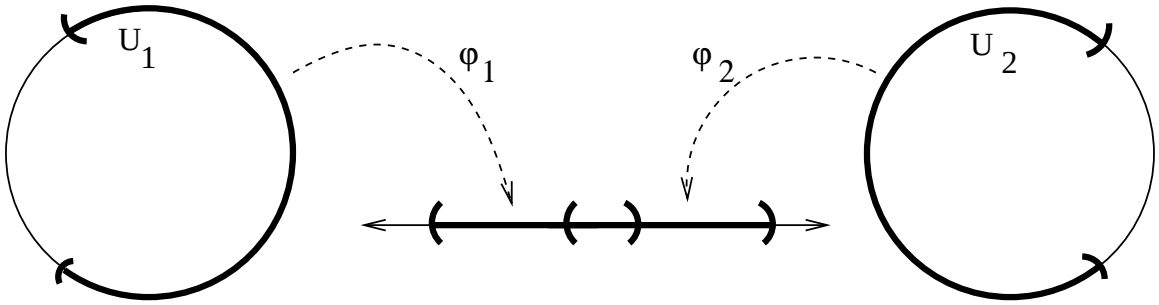


Figure 3.1: A circle is a manifold because there exists an open cover consisting of the sets U_1, U_2, which are mapped homeomorphically onto open intervals in the real line via the functions φ_1, φ_2. (It is not necessary that the intervals intersect in $\mathbb{R}$.)

We refer the reader to [92] for an excellent overview of several manifolds with relevance to signal processing, including the rotation group $SO(3)$, which can be used for representing orientations of objects in 3-D space, and the Grassman manifold $G(K, N)$, which represents all K-dimensional subspaces of $\mathbb{R}^N$. (Without working through the technicalities of the definition of a manifold, it is easy to see that both types of data have a natural notion of neighborhood.)

3.3 Tangent spaces

A manifold is **differentiable** if, for any two charts whose open sets on $\mathcal{M}$ overlap, the composition of the corresponding homeomorphisms (from $\mathbb{R}^K$ in one chart to $\mathcal{M}$ and back to $\mathbb{R}^K$ in the other) is differentiable. (In our simple example, the circle is a differentiable manifold.)

To each point x in a differentiable manifold, we may associate a K-dimensional **tangent space** Tan_x. For signal manifolds embedded in L_2 or $\mathbb{R}^N$, it suffices to think of Tan_x as the set of all directional derivatives of smooth paths on $\mathcal{M}$ through x. (Note that Tan_x is a linear subspace and has its origin at 0, rather than at x.)

3.4 Distances

One is often interested in measuring distance along a manifold. For abstract differentiable manifolds, this can be accomplished by defining a Riemannian metric on the tangent spaces. A Riemannian metric is a collection of inner products $<,>_x$ defined at each point $x \in \mathcal{M}$. The inner product gives a measure for the "length" of a tangent, and one can then compute the length of a path on $\mathcal{M}$ by integrating its tangent lengths along the path.

For differentiable manifolds embedded in $\mathbb{R}^N$, the natural metric is the Euclidean metric inherited from the ambient space. The length of a path $\gamma : [0,1] \mapsto \mathcal{M}$ can then be computed simply using the limit

$$\mathrm{length}(\gamma) = \lim_{j \to \infty} \sum_{i=1}^{j} \|\gamma(i/j) - \gamma((i-1)/j)\|_2. \tag{3.1}$$

The **geodesic distance** $d_{\mathcal{M}}(x, y)$ between two points $x, y \in \mathcal{M}$ is then given by the length of the shortest path γ on $\mathcal{M}$ joining x and y.

3.5 Condition number

To establish a firm footing for analysis, we find it helpful assume a certain regularity to the manifold beyond mere differentiability. For this purpose, we adopt the condition number defined recently by Niyogi et al. [87].

Definition 3.1:
[87] Let $\mathcal{M}$ be a compact submanifold of $\mathbb{R}^N$. The **condition number** of $\mathcal{M}$ is defined as $1/\tau$, where τ is the largest number having the following property: The open normal bundle about $\mathcal{M}$ of radius r is imbedded in $\mathbb{R}^N$ for all $r < \tau$.

The open normal bundle of radius r at a point $x \in \mathcal{M}$ is simply the collection of all vectors of length $< r$ anchored at x and with direction orthogonal to Tan_x.

In addition to controlling local properties (such as curvature) of the manifold, the condition number has a global effect as well, ensuring that the manifold is self-avoiding. These notions are made precise in several lemmata, which we repeat below for completeness.

Lemma 3.1:
[87] If $\mathcal{M}$ is a submanifold of $\mathbb{R}^N$ with condition number $1/\tau$, then the norm of the second fundamental form is bounded by $1/\tau$ in all directions.

This implies that unit-speed geodesic paths on $\mathcal{M}$ have curvature bounded by $1/\tau$. The second lemma concerns the twisting of tangent spaces.

Lemma 3.2:
[87] Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^N$ with condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points with geodesic distance given by $d_{\mathcal{M}}(p,q)$. Let θ be the angle between the tangent spaces Tan_p and Tan_q defined by $cos(\theta) = min_{u \in \text{Tan}_p} max_{v \in \text{Tan}_q} | < u, v > |$. Then $cos(\theta) > 1 - \frac{1}{\tau} d_{\mathcal{M}}(p,q)$.

The third lemma concerns self-avoidance of $\mathcal{M}$.

Lemma 3.3:
[87] Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^N$ with condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points such that $\|p - q\|_2 = d$. Then for all $d \leq \tau/2$, the geodesic distance $d_{\mathcal{M}}(p,q)$ is bounded by $d_{\mathcal{M}}(p,q) \leq \tau - \tau\sqrt{1 - 2d/\tau}$.

From Lemma 3.3, p. 13 we have an immediate corollary.

Corollary 3.1:
Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^N$ with condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points such that $\|p - q\|_2 = d$. If $d \leq \tau/2$, then $d \geq d_{\mathcal{M}}(p,q) - \frac{(d_{\mathcal{M}}(p,q))^2}{2\tau}$.

Chapter 4

Low-Dimensional Signal Models[1]

We now survey some common and important models in signal processing, each of which involves some notion of conciseness to the signal structure. We see in each case that this conciseness gives rise to a low-dimensional geometry within the ambient signal space.

4.1 Linear models

Some of the simplest models in signal processing correspond to **linear subspaces** of the ambient signal space. Bandlimited signals are one such example. Supposing, for example, that a 2π-periodic signal f has Fourier transform $F(\omega) = 0$ for $|\omega| > B$, the Shannon/Nyquist sampling theorem [81] states that such signals can be reconstructed from $2B$ samples. Because the space of B-bandlimited signals is closed under addition and scalar multiplication, it follows that the set of such signals forms a $2B$-dimensional linear subspace of $L^2([0, 2\pi))$.

Linear signal models also appear in cases where a model dictates a **linear constraint** on a signal. Considering a discrete length-N signal x, for example, such a constraint can be written in matrix form as

$$Ax = 0 \tag{4.1}$$

for some $M \times N$ matrix A. Signals obeying such a model are constrained to live in $\mathcal{N}(A)$ (again, obviously, a linear subspace of $\mathbb{R}^N$).

A very similar class of models concerns signals living in an affine space, which can be represented for a discrete signal using

$$Ax = y. \tag{4.2}$$

The class of such x lives in a shifted nullspace $\hat{x} + \mathcal{N}(A)$, where $\hat{x}$ is any solution to the equation $A\hat{x} = y$.

Revisiting the dictionary setting (see Signal Dictionaries and Representations (Chapter 2)), one last important linear model arises in cases where we select K specific elements from the dictionary Ψ and then construct signals using linear combinations of only these K elements; in this case the set of possible signals forms a K-dimensional hyperplane in the ambient signal space (see Figure 4.1(a)).

[1] This content is available online at <http://cnx.org/content/m18726/1.4/>.

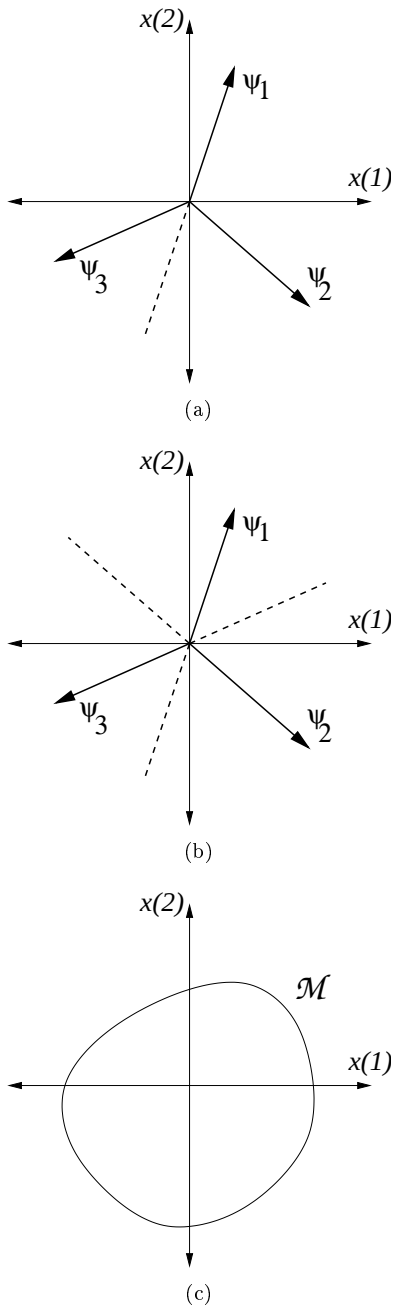


Figure 4.1: Simple models for signals in $\mathbb{R}^2$. (a) The linear space spanned by one element of the dictionary Ψ. The bold vectors denote the elements of the dictionary, while the dashed line (plus the corresponding dictionary element) denotes the subspace spanned by that dictionary element. (b) The nonlinear set of 1-sparse signals that can be built using Ψ. (c) A manifold $\mathcal{M}$.

For example, we may construct low-frequency signals using combinations of only the lowest frequency si-

nusoids from the Fourier dictionary. Similar subsets may be chosen from the wavelet dictionary; in particular, one may choose only elements that span a particular scaling space V_j. As we have mentioned previously, harmonic dictionaries such as sinusoids and wavelets are well-suited to representing smooth[2] signals. This can be seen in the decay of their transform coefficients. For example, we can relate the smoothness of a continuous 1-D function f to the decay of its Fourier coefficients $F(\omega)$; in particular, if $\int |F(\omega)| \left(1 + |\omega|^H\right) d\omega < \infty$, then $f \in \mathcal{C}^H$ [81]. In order to satisfy $\int |F(\omega)| \left(1 + |\omega|^H\right) d\omega < \infty$, a signal must have a sufficiently fast decay of the Fourier transform coefficients $|F(\omega)|$ as ω grows. Wavelet coefficients exhibit a similar decay for smooth signals: supposing $f \in \mathcal{C}^H$ and the wavelet basis function has at least H vanishing moments, then as the scale $j \to \infty$, the magnitudes of the wavelet coefficients decay as $2^{-j(H+1/2)}$ [81]. (Recall that $f \in \mathcal{C}^H$ implies f is well-approximated by a polynomial, and so due the vanishing moments this polynomial will have zero contribution to the wavelet coefficients.)

Indeed, these results suggest that the largest Fourier or wavelet coefficients of smooth signals tend to concentrate at the coarsest scales (lowest-frequencies). In Linear Approximation from Approximation (Section 5.1: Linear approximation) , we see that linear approximations formed from just the lowest frequency elements of the Fourier or wavelet dictionaries (i.e., the truncation of the Fourier or wavelet representation to only the lowest frequency terms) provide very accurate approximations to smooth signals. Put differently, smooth signals live near the subspace spanned by just the lowest frequency Fourier or wavelet basis functions.

4.2 Sparse (nonlinear) models

Sparse signal models can be viewed as a generalization of linear models. The notion of sparsity comes from the fact that, by the proper choice of dictionary Ψ, many real-world signals $x = \Psi\alpha$ have coefficient vectors α containing few large entries, but across different signals the locations (indices in α) of the large entries may change. We say a signal is strictly sparse (or "K-sparse") if all but K entries of α are zero.

Some examples of real-world signals for which sparse models have been proposed include neural spike trains (in time), music and other audio recordings (in time and frequency), natural images (in the wavelet or curvelet dictionaries [81], [49], [96], [78], [107], [59], [31], [19]), video sequences (in a 3-D wavelet dictionary [85], [95]), and sonar or radar pulses (in a chirplet dictionary [5]). In each of these cases, the relevant information in a sparse representation of a signal is encoded in both the **locations** (indices) of the significant coefficients and the **values** to which they are assigned. This type of uncertainty is an appropriate model for many natural signals with punctuated phenomena.

Sparsity is a **nonlinear** model. In particular, let Σ_K denote the set of all K-sparse signals for a given dictionary. It is easy to see that the set Σ_K is not closed under addition. (In fact, $\Sigma_K + \Sigma_K = \Sigma_{2K}$.) From a geometric perspective, the set of all K-sparse signals from the dictionary Ψ forms not a hyperplane but rather a union of K-dimensional hyperplanes, each spanned by K vectors of Ψ (see Figure 4.1(b)). For a dictionary Ψ with Z entries, there are $\binom{Z}{K}$ such hyperplanes. (The geometry of sparse signal collections has also been described in terms of orthosymmetric sets; see [58].)

Signals that are not strictly sparse but rather have a few "large" and many "small" coefficients are known as **compressible** signals. The notion of compressibility can be made more precise by considering the rate at which the **sorted** magnitudes of the coefficients α decay, and this decay rate can in turn be related to the ℓ_p norm of the coefficient vector α. Letting $\tilde{\alpha}$ denote a rearrangement of the vector α with the coefficients

[2] **Lipschitz smoothness** We say a continuous-time function of D variables has smoothness of order $H > 0$, where $H = r + \nu$, r is an integer, and $\nu \in (0, 1]$, if the following criteria are met [81], [49]:

- All iterated partial derivatives with respect to the D directions up to order r exist and are continuous.
- All such partial derivatives of order r satisfy a Lipschitz condition of order ν (also known as a Hölder condition).(A function $d \in \mathrm{Lip}(\nu)$ if $|d(t_1 + t_2) - d(t_1)| \leq C \parallel t_2 \parallel^{\nu}$ for all D-dimensional vectors t_1, t_2.)

We will sometimes consider the space of smooth functions whose partial derivatives up to order r are bounded by some constant Ω. With somewhat nonstandard notation, we denote the space of such bounded functions with bounded partial derivatives by $\mathcal{C}^H$, where this notation carries an implicit dependence on Ω. Observe that $r = \lceil H - 1 \rceil$, where $\lceil \cdot \rceil$ denotes rounding up. Also, when H is an integer $\mathcal{C}^H$ includes as a subset the space traditionally denoted by the notation "C^H" (the class of functions that have $H = r + 1$ continuous partial derivatives).

ordered in terms of decreasing magnitude, then the reordered coefficients satisfy [46]

$$\tilde{\alpha}_k \leq \| \alpha \|_{\ell_p} k^{-1/p}. \tag{4.3}$$

As we discuss in Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation), these decay rates play an important role in **nonlinear approximation**, where adaptive, K-sparse representations from the dictionary are used to approximate a signal.

We recall from Section 4.1 (Linear models) that for a smooth signal f, the largest Fourier and wavelet coefficients tend to cluster at coarse scales (low frequencies). Suppose, however, that the function f is piecewise smooth; i.e., it is $\mathcal{C}^H$ at every point $t \in \mathbb{R}$ except for one point t_0, at which it is discontinuous. Naturally, this phenomenon will be reflected in the transform coefficients. In the Fourier domain, this discontinuity will have a global effect, as the overall smoothness of the function f has been reduced dramatically from H to 0. Wavelet coefficients, however, depend only on local signal properties, and so the wavelet basis functions whose supports do not include t_0 will be unaffected by the discontinuity. Coefficients surrounding the singularity will decay only as $2^{-j/2}$, but there are relatively few such coefficients. Indeed, at each scale there are only $O(1)$ wavelets that include t_0 in their supports, but these locations are highly signal-dependent. (For modeling purposes, these significant coefficients will persist through scale down the parent-child tree structure.) After reordering by magnitude, the wavelet coefficients of piecewise smooth signals will have the same general decay rate as those of smooth signals. In Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation), we see that the quality of nonlinear approximations offered by wavelets for smooth 1-D signals is not hampered by the addition of a finite number of discontinuities.

4.3 Manifold models

Manifold models generalize the conciseness of sparsity-based signal models. In particular, in many situations where a signal is believed to have a concise description or "few degrees of freedom," the result is that the signal will live on or near a particular submanifold of the ambient signal space.

4.3.1 Parametric models

We begin with an abstract motivation for the manifold perspective. Consider a signal f (such as a natural image), and suppose that we can identify some single 1-D piece of information about that signal that could be variable; that is, other signals might rightly be called "similar" to f if they differ only in this piece of information. (For example, this 1-D parameter could denote the distance from some object in an image to the camera.) We let θ denote the variable parameter and write the signal as f_θ to denote its dependence on θ. In a sense, θ is a single "degree of freedom" driving the generation of the signal f_θ under this simple model. We let Θ denote the set of possible values of the parameter θ. If the mapping between θ and f_θ is well-behaved, then the collection of signals $\{f_\theta : \theta \in \Theta\}$ forms a 1-D path in the ambient signal space.

More generally, when a signal has K degrees of freedom, we may model it as depending on some parameter θ that is chosen from a K-dimensional manifold Θ. (The parameter space Θ could be, for example, a subset of $\mathbb{R}^K$, or it could be a more general manifold such as SO(3).) We again let f_θ denote the signal corresponding to a particular choice of θ, and we let $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$. Assuming the mapping f is continuous and injective over Θ (and its inverse is continuous), then by virtue of the manifold structure of Θ, its image $\mathcal{F}$ will correspond to a K-dimensional manifold embedded in the ambient signal space (see Figure 4.1(c)).

These types of parametric models arise in a number of scenarios in signal processing. Examples include: signals of unknown translation, sinusoids of unknown frequency (across a continuum of possibilities), linear radar chirps described by a starting and ending time and frequency, tomographic or light field images with articulated camera positions, robotic systems with few physical degrees of freedom, dynamical systems with low-dimensional attractors [13], [15], and so on.

In general, parametric signals manifolds are **nonlinear** (by which we mean non-affine as well); this can again be seen by considering the sum of two signals $f_{\theta_0} + f_{\theta_1}$. In many interesting situations, signal manifolds are **non-differentiable** as well.

4.3.2 Nonparametric models

Manifolds have also been used to model signals for which there is no known parametric model. Examples include images of faces and handwritten digits [101], [9], which have been found empirically to cluster near low-dimensional manifolds. Intuitively, because of the configurations of human joints and muscles, it may be conceivable that there are relatively "few" degrees of freedom driving the appearance of a human face or the style of handwriting; however, this inclination is difficult or impossible to make precise. Nonetheless, certain applications in face and handwriting recognition have benefitted from algorithms designed to discover and exploit the nonlinear manifold-like structure of signal collections. Manifold Learning from Dimensionality Reduction (Section 7.1: Manifold learning) discusses such methods for learning parametrizations and other information from data living along manifolds.

Much more generally, one may consider, for example, the set of **all** natural images. Clearly, this set has small volume with respect to the ambient signal space — generating an image randomly pixel-by-pixel will almost certainly produce an unnatural noise-like image. Again, it is conceivable that, at least locally, this set may have a low-dimensional manifold-like structure: from a given image, one may be able to identify only a limited number of meaningful changes that could be performed while still preserving the natural look to the image. Arguably, most work in signal modeling could be interpreted in some way as a search for this overall structure.

Chapter 5

Approximation[1]

To this point, we have discussed signal representations and models as basic tools for signal processing. In the following modules, we discuss the actual application of these tools to tasks such as approximation and compression, and we continue to discuss the geometric implications.

5.1 Linear approximation

One common prototypical problem in signal processing is to find the best linear approximation to a signal x. By "best linear approximation," we mean the best approximation to x from among a class of signals comprising a linear (or affine) subspace. This situation may arise, for example, when we have a noisy observation of a signal believed to obey a linear model. If we choose an ℓ_2 error criterion, the solution to this optimization problem has a particularly strong geometric interpretation.

To be more concrete, suppose S is a K-dimensional linear subspace of $\mathbb{R}^N$. (The case of an affine subspace follows similarly.) If we seek

$$s^* := \underset{s \in S}{argmin} \|s - x\|_2, \tag{5.1}$$

standard linear algebra results state that the minimizer is given by

$$s^* = A^T A x, \tag{5.2}$$

where A is a $K \times N$ matrix whose rows form an orthonormal basis for S. Geometrically, one can easily see that this solution corresponds to an orthogonal projection of x onto the subspace S (see Figure 5.1(a)).

[1] This content is available online at <http://cnx.org/content/m18727/1.5/>.

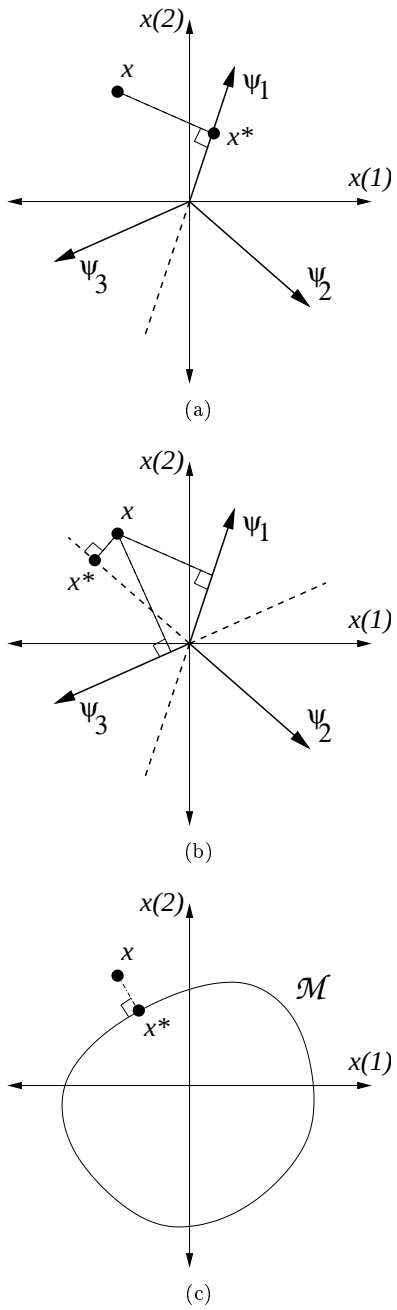
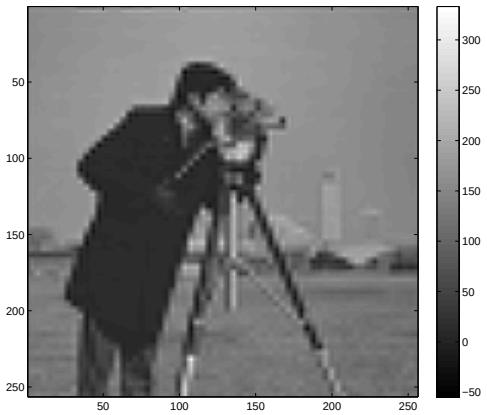


Figure 5.1: Approximating a signal $x \in \mathbb{R}^2$ with an ℓ_2 error criterion. (a) Linear approximation using one element of the dictionary Ψ corresponds to orthogonal projection of the signal onto the linear subspace. (b) Nonlinear approximation corresponds to orthogonal projection of the signal onto the nearest candidate subspace. In this case, we choose the best 1-sparse signal that can be built using Ψ. (c) Manifold-based approximation, finding the nearest point on $\mathcal{M}$.

The linear approximation problem arises frequently in settings involving signal dictionaries. In some settings, such as the case of an oversampled bandlimited signal, certain coefficients in the vector α may be assumed to be fixed at zero. In the case where the dictionary Ψ forms an orthonormal basis, the linear approximation estimate of the unknown coefficients has a particularly simple form: rows of the matrix A in (5.2) are obtained by selecting and transposing the columns of Ψ whose expansion coefficients are unknown, and consequently, the unknown coefficients can be estimated simply by taking the inner products of x against the appropriate columns of Ψ.

For example, in choosing a fixed subset of the Fourier or wavelet dictionaries, one may rightfully choose the lowest frequency (coarsest scale) basis functions for the set S because, as discussed in Linear Models from Low-Dimensional Signal Models (Section 4.1: Linear models) , the coefficients generally tend to decay at higher frequencies (finer scales). For smooth functions, this strategy is appropriate and effective; functions in Sobolev smoothness spaces are well-approximated using linear approximations from the Fourier or wavelet dictionaries [82]. For piecewise smooth functions, however, even the wavelet-domain linear approximation strategy would miss out on significant coefficients at fine scales. Since the locations of such coefficients are unknown a priority, it is impossible to propose a linear wavelet-domain approximation scheme that could simultaneously capture all piecewise smooth signals. As an example, Figure 5.2(a) shows the linear approximation of the Cameraman test image obtained by keeping only the lowest-frequency scaling and wavelet coefficients. No high-frequency information is available to clearly represent features such as edges.



(a)

(b)

Figure 5.2: Linear versus nonlinear approximation in the wavelet domain. (a) Linear approximation of the Cameraman test image obtained by keeping the $K = 4096$ lowest-frequency wavelet coefficients from the five-level wavelet decomposition. The MSE with respect to the original image is 353. (b) Nonlinear approximation of the Cameraman test image obtained by keeping the $K = 4096$ largest wavelet coefficients from the five-level wavelet decomposition. The MSE with respect to the original image is 72. Compared with linear approximation, more high frequency coefficients are included, which allows better representation of features such as edges.

5.2 Nonlinear approximation

A related question often arises in settings involving signal dictionaries. Rather than finding the best approximation to a signal f using a fixed collection of K elements from the dictionary Ψ, one may often seek the best K-term representation to f among all possible expansions that use K terms from the dictionary. Compared to linear approximation, this type of nonlinear approximation [45], [39] utilizes the ability of the dictionary to adapt: different elements may be important for representing different signals.

The K-term nonlinear approximation problem corresponds to the optimization

$$s^*_{K,p} := arg\,\min_{s \in \Sigma_K} \| s - f \|_p. \tag{5.3}$$

(For the sake of generality, we consider general L_p and ℓ_p norms in this section.) Due to the nonlinearity of the set Σ_K for a given dictionary, solving this problem can be difficult. Supposing Ψ is an orthonormal basis and $p = 2$, the solution to (5.3) is easily obtained by thresholding: one simply computes the coefficients α and keeps the K largest (setting the remaining coefficients to zero). The approximation error is then given simply by the coefficients that are discarded:

$$\| s^*_{K,2} - f \|_2 = \left(\sum_{k>K} \tilde{\alpha}_k^2 \right)^{1/2}. \tag{5.4}$$

When Ψ is a redundant dictionary, however, the situation is much more complicated. We mention more on this below (see also Figure 5.1(b)).

5.2.1 Measuring approximation quality

One common measure for the quality of a dictionary Ψ in approximating a signal class is the fidelity of its K-term representations. Often one examines the asymptotic rate of decay of the K-term approximation error as K grows large. Defining

$$\sigma_K(f)_p := \| s^*_{K,p} - f \|_p, \tag{5.5}$$

for a given signal f we may consider the asymptotic decay of $\sigma_K(f)_p$ as $K \to \infty$. (We recall the dependence of (5.3) and hence (5.5) on the dictionary Ψ.) In many cases, the function $\sigma_K(f)_p$ will decay as K^{-r} for some r, and when Ψ represents a harmonic dictionary, faster decay rates tend to correspond to smoother functions. Indeed, one can show that when Ψ is an orthonormal basis, then $\sigma_K(f)_2$ will decay as K^{-r} if and only if $\tilde{\alpha}_k$ decays as $k^{-r+1/2}$ [47].

5.2.2 Nonlinear approximation of piecewise smooth functions

Let $f \in \mathcal{C}^H$ be a 1-D function. Supposing the wavelet dictionary has more than H vanishing moments, then f can be well approximated using its K largest coefficients (most of which are at coarse scales). As K grows large, the nonlinear approximation error will decay[2] as $\sigma_K(f)_2 \lesssim K^{-H}$.

Supposing that f is piecewise smooth, however, with a finite number of discontinuities, then (as discussed in Sparse (Nonlinear) Models from Low-Dimensional Signal Models (Section 4.2: Sparse (nonlinear) models)) f will have a limited number of significant wavelet coefficients at fine scales. Because of the concentration of these significant coefficients within each scale, the nonlinear approximation rate will remain $\sigma_K(f)_2 \lesssim K^{-H}$ as if there were no discontinuities present [82].

Unfortunately, this resilience of wavelets to discontinuities does not extend to higher dimensions. Suppose, for example, that f is a $\mathcal{C}^H$ smooth 2-D signal. Assuming the proper number of vanishing moments, a wavelet representation will achieve the optimal nonlinear approximation rate $\sigma_K(f)_2 \lesssim K^{-H/2}$ [37], [82]. As in the 1-D case, this approximation rate is maintained when a finite number of point discontinuities are introduced into f. However, when f contains 1-D discontinuities (edges separating the smooth regions), the approximation rate will fall to $\sigma_K(f)_2 \lesssim K^{-1/2}$ [82]. The problem actually arises due to the isotropic, dyadic supports of the wavelets; instead of $O(1)$ significant wavelets at each scale, there are now $O\left(2^j\right)$ wavelets overlapping the discontinuity. We revisit this important issue in Compression (Chapter 6).

Despite the limited approximation capabilities for images with edges, nonlinear approximation in the wavelet domain typically offers a superior approximation to an image compared to linear approximation in the wavelet domain. As an example, Figure 5.2(b) shows the nonlinear approximation of the Cameraman

[2] We use the notation $f(\alpha) \lesssim g(\alpha)$, or $f(\alpha) = O(g(\alpha))$, if there exists a constant C, possibly large but not dependent on the argument α, such that $f(\alpha) \leq C g(\alpha)$.

test image obtained by keeping the largest scaling and wavelet coefficients. In this case, a number of high-frequency coefficients are selected, which gives an improved ability to represent features such as edges. Better concise transforms, which capture the image information in even fewer coefficients, would offer further improvements in terms of nonlinear approximation quality.

5.2.3 Finding approximations

As mentioned above, in the case where Ψ is an orthonormal basis and $p = 2$, the solution to (5.3) is easily obtained by thresholding: one simply computes the coefficients α and keeps the K largest (setting the remaining coefficients to zero). Thresholding can also be shown to be optimal for arbitrary ℓ_p norms in the special case where Ψ is the canonical basis. While the optimality of thresholding does not generalize to arbitrary norms and bases, thresholding can be shown to be a near-optimal approximation strategy for wavelet bases with arbitrary L_p norms [47].

In the case where Ψ is a redundant dictionary, however, the expansion coefficients α are not unique, and the optimization problem (5.3) can be much more difficult to solve. Indeed, supposing even that an **exact** K-term representation exists for f in the dictionary Ψ, finding that K-term approximation is NP-hard in general, requiring a combinatorial enumeration of the $\binom{Z}{K}$ possible sparse subspaces [26]. This search can be recast as the optimization problem

$$\hat{\alpha} = argmin \| \alpha \|_0 \quad \text{s.t. } f = \Psi\alpha. \tag{5.6}$$

While solving (5.6) is prohibitively complex, a variety of algorithms have been proposed as alternatives. One approach convexifies the optimization problem by replacing the ℓ_0 fidelity criterion by an ℓ_1 criterion

$$\hat{\alpha} = argmin \| \alpha \|_1 \quad \text{s.t. } f = \Psi\alpha. \tag{5.7}$$

This problem, known as Basis Pursuit [34], is significantly more approachable and can be solved with traditional linear programming techniques whose computational complexities are polynomial in Z. The ℓ_1 criterion has the advantage of yielding a convex optimization problem while still encouraging sparse solutions due to the polytope geometry of the ℓ_1 unit ball (see for example [55] and [61]). Iterative greedy algorithms such as Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP) [82] have also been suggested to find sparse representations α for a signal f. Both MP and OMP iteratively select the columns from Ψ that are most correlated with f, then subtract the contribution of each column, leaving a residual. OMP includes an additional step at each iteration where the residual is orthogonalized against the previously selected columns.

5.3 Manifold approximation

We also consider the problem of finding the best manifold-based approximation to a signal (see Figure 5.1(c)). Suppose that $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ is a parametrized K-dimension manifold and that we are given a signal I that is believed to approximate f_θ for an unknown $\theta \in \Theta$. From I we wish to recover an estimate of θ. Again, we may formulate this parameter estimation problem as an optimization, writing the objective function (here we concentrate solely on the L_2 or ℓ_2 case)

$$D(\theta) = \|f_\theta - I\|_2^2 \tag{5.8}$$

and solving for

$$\theta^* = \underset{\theta \in \Theta}{argmin} D(\theta). \tag{5.9}$$

We suppose that the minimum is uniquely defined.

Standard nonlinear parameter estimation [8] tells us that, if D is differentiable, we can use Newton's method to iteratively refine a sequence of guesses $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \cdots$ to θ^* and rapidly convergence to the true value. Supposing that $\mathcal{F}$ is a **differentiable** manifold, we would let

$$J = [\partial D/\partial\theta_0 \;\; \partial D/\partial\theta_1 \;\; \cdots \;\; \partial D/\partial\theta_{K-1}]^T \tag{5.10}$$

be the gradient of D, and let H be the $K \times K$ Hessian, $H_{ij} = \frac{\partial^2 D}{\partial\theta_i \partial\theta_j}$. Assuming D is differentiable, Newton's method specifies the following update step:

$$\theta^{(k+1)} \leftarrow \theta^{(k)} + \left[H\left(\theta^{(k)}\right)\right]^{-1} J\left(\theta^{(k)}\right). \tag{5.11}$$

To relate this method to the structure of the manifold, we can actually express the gradient and Hessian in terms of signals, writing

$$D(\theta) = \| f_\theta - I \|_2^2 = \int (f_\theta - I)^2 \, dx = \int f_\theta^2 - 2 I f_\theta + I^2 \, dx. \tag{5.12}$$

Differentiating with respect to component θ_i, we obtain

$$\begin{array}{rcl} \frac{\partial D}{\partial\theta_i} = J_i &=& \frac{\partial}{\partial\theta_i}\left(\int f_\theta^2 - 2 I f_\theta + I^2 \, dx\right) \\ &=& \int \frac{\partial}{\partial\theta_i}\left(f_\theta^2\right) - 2I\frac{\partial}{\partial\theta_i} f_\theta \, dx \\ &=& \int 2 f_\theta \tau_\theta^i - 2 I \tau_\theta^i \, dx \\ &=& 2 < f_\theta - I, \tau_\theta^i >, \end{array} \tag{5.13}$$

where $\tau_\theta^i = \frac{\partial f_\theta}{\partial\theta_i}$ is a tangent signal. Continuing, we examine the Hessian,

$$\begin{array}{rcl} \frac{\partial^2 D}{\partial\theta_i\partial\theta_j} = H_{ij} &=& \frac{\partial}{\partial\theta_j}\left(\frac{\partial D}{\partial\theta_i}\right) \\ &=& \int \frac{\partial}{\partial\theta_j}\left(2 f_\theta \tau_\theta^i - 2 I \tau_\theta^i\right) \, dx \\ &=& \int 2\tau_\theta^i \tau_\theta^j + 2 f_\theta \tau_\theta^{ij} - 2 I \tau_\theta^{ij} \, dx \\ &=& 2 < \tau_\theta^i, \tau_\theta^j > + 2 < f_\theta - I, \tau_\theta^{ij} >, \end{array} \tag{5.14}$$

where $\tau_\theta^{ij} = \frac{\partial^2 f_\theta}{\partial\theta_i\partial\theta_j}$ denotes a second-derivative signal. Thus, we can interpret Newton's method geometrically as (essentially) a sequence of successive projections onto tangent spaces on the manifold.

Again, the above discussion assumes the manifold to be differentiable. Many interesting parametric signal manifolds are in fact nowhere differentiable — the tangent spaces demanded by Newton's method do not exist. However, in [105] we have identified a type of multiscale tangent structure to the manifold that permits a coarse-to-fine technique for parameter estimation.

Chapter 6

Compression[1]

6.1 Transform coding

In Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation), we measured the quality of a dictionary in terms of its K-term approximations to signals drawn from some class. One reason that such approximations are desirable is that they provide concise descriptions of the signal that can be easily stored, processed, etc. There is even speculation and evidence that neurons in the human visual system may use sparse coding to represent a scene [89].

For data compression, conciseness is often exploited in a popular technique known as **transform coding**. Given a signal f (for which a concise description may not be readily apparent in its native domain), the idea is simply to use the dictionary Ψ to transform f to its coefficients α, which can then be efficiently and easily described. As discussed above, perhaps the simplest strategy for summarizing a sparse α is simply to threshold, keeping the K largest coefficients and discarding the rest. A simple encoder would then just encode the positions and quantized values of these K coefficients.

6.2 Metric entropy

Suppose f is a function and let $\hat{f}_R$ be an approximation to f encoded using R bits. To evaluate the quality of a coding strategy, it is common to consider the **asymptotic rate-distortion** (R-D) performance, which measures the decay rate of $\| f - \hat{f}_R \|_{L_p}$ as $R \to \infty$. The **metric entropy** [74] for a class $\mathcal{F}$ gives the best decay rate that can be achieved uniformly over all functions $f \in \mathcal{F}$. We note that this is a true measure for the complexity of a class and is tied to no particular dictionary or encoding strategy. The metric entropy also has a very geometric interpretation, as it relates to the smallest radius possible for a covering of 2^R balls over the set $\mathcal{F}$.

Metric entropies are known for certain signal classes. For example, the results of Clements [36] (extending those of Kolmogorov and Tihomirov [74]) regarding metric entropy give bounds on the optimal achievable asymptotic rate-distortion performance for D-dimensional $\mathcal{C}^H$-smooth functions f (see also [38]):

$$\| f - \hat{f}_R \|_{L_p} \lesssim \left(\frac{1}{R} \right)^{\frac{H}{D}}. \tag{6.1}$$

Rate-distortion performance measures the complexity of a representation and encoding strategy. In the case of transform coding, for example, R-D results account for the bits required to encode both the values of the significant coefficients **and** their locations. Nonetheless, in many cases transform coding is indeed an

[1] This content is available online at <http://cnx.org/content/m18729/1.3/>.

effective strategy for encoding signals that have sparse representations [40]. For example, in [38] Cohen et al. propose a wavelet-domain coder that uses a connected-tree structure to efficiently encode the positions of the significant coefficients and prove that this encoding strategy achieves the optimal rate

$$\| f - \hat{f}_R \|_{L_p} \lesssim \left(\frac{1}{R} \right)^{\frac{H}{D}}. \tag{6.2}$$

6.3 Compression of piecewise smooth images

In some cases, however, the sparsity of the wavelet transform may not reflect the true underlying structure of a signal. Examples are 2-D piecewise smooth signals with a smooth edge discontinuity separating the smooth regions. As we discussed in Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation), wavelets fail to sparsely represent these functions, and so the R-D performance for simple thresholding-based coders will suffer as well. In spite of all of the benefits of wavelet representations for signal processing (low computational complexity, tree structure, sparse approximations for smooth signals), this failure to efficiently represent edges is a significant drawback. In many images, edges carry some of the most prominent and important information [84], and so it is desirable to have a representation well-suited to compressing edges in images.

To address this concern, recent work in harmonic analysis has focused on developing representations that provide sparse decompositions for certain geometric image classes. Examples include curvelets [30], [20] and contourlets [52], slightly redundant tight frames consisting of anisotropic, "needle-like" atoms. In [77], bandelets are formed by warping an orthonormal wavelet basis to conform to the geometrical structure in the image. A nonlinear multiscale transform that adapts to discontinuities (and can represent a "clean" edge using very few coarse scale coefficients) is proposed in [3]. Each of these new representations has been shown to achieve near-optimal asymptotic approximation and R-D performance for piecewise smooth images consisting of C^H regions separated by discontinuities along C^H curves, with $H = 2$ ($H \geq 2$ for bandelets). Some have also found use in specialized compression applications such as identification photos [1].

In [33], we have presented a scheme that is based on the simple yet powerful observation that geometric features can be efficiently approximated using local, geometric atoms in the spatial domain, and that the projection of these geometric primitives onto wavelet subspaces can therefore approximate the corresponding wavelet coefficients. We prove that the resulting dictionary achieves the optimal nonlinear approximation rates for piecewise smooth signal classes. To account for the added complexity of this encoding strategy, we also consider R-D results and prove that this scheme comes within a logarithmic factor of the optimal performance rate. Unlike the techniques mentioned above, our method also generalizes to arbitrary orders of smoothness and arbitrary signal dimension.

Chapter 7

Dimensionality Reduction[1]

Recent years have seen a proliferation of novel techniques for what can loosely be termed "dimensionality reduction." Like the tasks of approximation and compression discussed above, these methods involve some aspect in which low-dimensional information is extracted about a signal or collection of signals in some high-dimensional ambient space. Unlike the tasks of approximation and compression, however, the goal of these methods is not always to maintain a faithful representation of each signal. Instead, the purpose may be to preserve some critical relationships among elements of a data set or to discover information about a manifold on which the data lives.

In this section, we review two general methods for dimensionality reduction. Section 7.1 (Manifold learning) begins with a brief overview of techniques for manifold learning. Section 7.2 (The Johnson-Lindenstrauss lemma) then discusses the Johnson-Lindenstrauss (JL) lemma, which concerns the isometric embedding of a cloud points as it is projected to a lower-dimensional space. Though at first glance the JL lemma does not pertain to any of the low-dimensional signal models we have previously discussed, we later see in Connections with dimensionality reduction (Section 8.6: Connections with dimensionality reduction) that the JL lemma plays a critical role in the core theory of CS, and we also employ the JL lemma in developing a theory for isometric embeddings of manifolds.

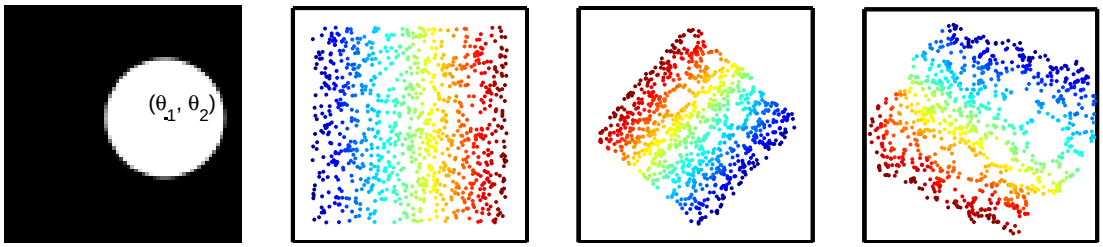
7.1 Manifold learning

Several techniques have been proposed for solving a problem known as **manifold learning** in which certain properties of a manifold are inferred from a discrete collection of points sampled from that manifold. A typical manifold learning setup is as follows: an algorithm is presented with a set of P points sampled from a K-dimensional submanifold of $\mathbb{R}^N$. The goal of the algorithm is to produce an mapping of these P points into some lower dimension $\mathbb{R}^M$ (ideally, $M = K$) while preserving some characteristic property of the manifold. Example algorithms include ISOMAP [98], Hessian Eigenmaps (HLLE) [60], and Maximum Variance Unfolding (MVU) [106], which attempt to learn isometric embeddings of the manifold (thus preserving pairwise geodesic distances in $\mathbb{R}^M$); Locally Linear Embedding (LLE) [93], which attempts to preserve local linear neighborhood structures among the embedded points; Local Tangent Space Alignment (LTSA) [108], which attempts to preserve local coordinates in each tangent space; and a method for charting a manifold [12] that attempts to preserve local neighborhood structures.

The internal mechanics of these algorithms differs depending on the objective criterion to be preserved, but as an example, the ISOMAP algorithm operates by first estimating the geodesic distance between each pair of points on the manifold (by approximating geodesic distance as the sum of Euclidean distances between pairs of the available sample points). After the $P \times P$ matrix of pairwise geodesic distances is constructed, a technique known as multidimensional scaling uses an eigendecomposition of the distance matrix to determine

[1] This content is available online at <http://cnx.org/content/m18732/1.5/>.

the proper M-dimensional embedding space. An example of using ISOMAP to learn a 2-dimensional manifold is shown in Figure 7.1.



(a) (b) (c) (d)

Figure 7.1: Manifold learning demonstration. (a) As input to the manifold learning algorithm, 1000 images of size 64×64 are created, where each image consists of a white disk translated to a random position (θ_1, θ_2). It follows that the images represent a sampling of 1000 points from a 2-dimensional submanifold of $\mathbb{R}^{4096}$. (b) Scatter plot of the true values for the (θ_1, θ_2) positions. For visibility in each plot, the color of each point indicates the true θ_1 value. (c) ISOMAP embedding learned from original data points in $\mathbb{R}^{4096}$. From the low-dimensional embedding coordinates we can infer the relative positions of the original high-dimensional images. (d) ISOMAP embedding learned from a random projection of the data set to $\mathbb{R}^M$, where $M = 15$.

These algorithms can be useful for learning the dimension and parametrizations of manifolds, for sorting data, for visualization and navigation through the data, and as preprocessing to make further analysis more tractable; common demonstrations include analysis of face images and classification of and handwritten digits. A related technique, the Whitney Reduction Network [14], [16], seeks a linear mapping to $\mathbb{R}^M$ that preserves ambient pairwise distances on the manifold and is particularly useful for processing the output of dynamical systems having low-dimensional attractors.

Other algorithms have been proposed for characterizing manifolds from sampled data without constructing an explicit embedding in $\mathbb{R}^M$. The Geodesic Minimal Spanning Tree (GMST) [42] models the data as random samples from the manifold and estimates the corresponding entropy and dimensionality. Another technique [88] has been proposed for using random samples of a manifold to estimate its homology (via the Betti numbers, which essentially characterize its dimension, number of connected components, etc.). Persistence Barcodes [32] are a related technique that involves constructing a type of signature for a manifold (or simply a shape) that uses tangent complexes to detect and characterize local edges and corners.

Additional algorithms have been proposed for constructing meaningful functions on the point samples in $\mathbb{R}^N$. To solve a semi-supervised learning problem, a method called Laplacian Eigenmaps [10] has been proposed that involves forming an adjacency graph for the data in $\mathbb{R}^N$, computing eigenfunctions of the Laplacian operator on the graph (which form a basis for L_2 on the graph), and using these functions to train a classifier on the data. The resulting classifiers have been used for handwritten digit recognition, document classification, and phoneme classification. (The M smoothest eigenfunctions can also be used to embed the manifold in M, similar to the approaches described above.) A related method called Diffusion Wavelets [41] uses powers of the diffusion operator to model scale on the manifold, then constructs wavelets to capture local behavior at each scale. The result is a wavelet transform adapted not to geodesic distance but to diffusion distance, which measures (roughly) the number of paths connecting two points.

7.2 The Johnson-Lindenstrauss lemma

7.2.1 Fundamentals

As with the above techniques in manifold learning, the Johnson-Lindenstrauss (JL) lemma [70], [2], [43], [69] provides a method for dimensionality reduction of a set of data in $\mathbb{R}^N$. Unlike manifold-based methods, however, the JL lemma can be used for any arbitrary set Q of points in $\mathbb{R}^N$; the data set is not assumed to have any a priori structure.

Despite the apparent lack of structure in an arbitrary point cloud data set, the JL lemma suggests that there does exist a method for dimensionality reduction of that data set that can preserve key information while mapping the data to a lower-dimensional space $\mathbb{R}^M$. In particular, the original formulation of the JL lemma [70] states that there exists a Lipschitz mapping $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^M$ with $M = O\left(log\left(\#Q\right)\right)$ such that all pairwise distances between points in Q are approximately preserved. This fact is useful for solving problems such as **Approximate Nearest Neighbor** [69], in which one desires the nearest point in Q to some query point $y \in \mathbb{R}^N$ (but a solution not much further than the optimal point is also acceptable). Such problems can be solved significantly more quickly in $\mathbb{R}^M$ than in $\mathbb{R}^N$.

Recent reformulations of the JL lemma propose random linear operators that, with high probability, will ensure a near isometric embedding. These typically build on concentration of measure results such as the following.

Lemma 7.1:
[2], [43] Let $x \in \mathbb{R}^N$, fix $0 < \varepsilon < 1$, and let Φ be a matrix constructed in one of the following two manners:

1. Φ is a random $M \times N$ matrix with i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$ entries, where $\sigma^2 = 1/N$, or
2. Φ is random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$.

Then with probability exceeding

$$1 - 2exp\left(-\frac{M\left(\varepsilon^2/2 - \varepsilon^3/3\right)}{2}\right), \tag{7.1}$$

the following holds:

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}}. \tag{7.2}$$

The random orthoprojector referred to above is clearly related to the first case (simple matrix multiplication by a Gaussian Φ) but subtly different; one could think of constructing a random Gaussian Φ, then using Gram-Schmidt to orthonormalize the rows before multiplying x. We note also that simple rescaling of Φ can be used to eliminate the $\sqrt{\frac{M}{N}}$ in (7.2); however we prefer this formulation for later reference.

By using the union bound over all $\binom{\#Q}{2}$ pairs of distinct points in Q, Lemma "The Johnson-Lindenstrauss lemma" (Lemma 7.2, Johnson-Lindenstrauss, p. 33) can be used to prove a randomized version of the Johnson-Lindenstrauss lemma.

Lemma 7.2: Johnson-Lindenstrauss
Let Q be a finite collection of points in $\mathbb{R}^N$. Fix $0 < \varepsilon < 1$ and $\beta > 0$. Set

$$M \geq \left(\frac{4+2\beta}{\varepsilon^2/2 - \varepsilon^3/3}\right) ln\left(\#Q\right). \tag{7.3}$$

Let Φ be a matrix constructed in one of the following two manners:

1. Φ is a random $M \times N$ matrix with i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$ entries, where $\sigma^2 = 1/N$, or

2. Φ is random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$.

Then with probability exceeding $1 - (\#Q)^{-\beta}$, the following statement holds: for every $x, y \in Q$,

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x - \Phi y\|_2}{\|x-y\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}}. \tag{7.4}$$

Indeed, [2] establishes that both Lemma 7.1, p. 33 and Lemma 7.2, Johnson-Lindenstrauss, p. 33 also hold when the elements of Φ are chosen i.i.d. from a random Rademacher distribution ($\pm\sigma$ with equal probability 1/2) or from a similar ternary distribution ($\pm\sqrt{3}\sigma$ with equal probability 1/6; 0 with probability 2/3). These can further improve the computational benefits of the JL lemma.

7.2.2 Connections with compressed sensing

In the following module on Compressed Sensing we will discuss further topics in dimensionality reduction that relate to the JL lemma. In particular, as discussed in Connections with dimensionality reduction (Section 8.6: Connections with dimensionality reduction), the core mechanics of Compressed Sensing can be interpreted in terms of a stable embedding that arises for the family of K-sparse signals when observed with random measurements, and this stable embedding can be proved using the JL lemma. Furthermore, as discussed in Stable embeddings of manifolds (Section 8.7: Stable embeddings of manifolds), one can ensure a stable embedding of families of signals obeying manifold models under a sufficient number of random projections, with the theory again following from the JL lemma.

Chapter 8

Compressed Sensing[1]

A new theory known as Compressed Sensing (CS) has recently emerged that can also be categorized as a type of dimensionality reduction. Like manifold learning, CS is strongly model-based (relying on sparsity in particular). However, unlike many of the standard techniques in dimensionality reduction (such as manifold learning or the JL lemma), the goal of CS is to maintain a low-dimensional representation of a signal x from which a faithful approximation to x can be recovered. In a sense, this more closely resembles the traditional problem of data compression (see Compression (Chapter 6)). In CS, however, the encoder requires no a priori knowledge of the signal structure. Only the **decoder** uses the model (sparsity) to recover the signal. We justify such an approach again using geometric arguments.

8.1 Motivation

Consider a signal $x \in \mathbb{R}^N$, and suppose that the basis Ψ provides a K-sparse representation of x

$$x = \Psi\alpha, \tag{8.1}$$

with $\| \alpha \|_0 = K$. (In this section, we focus on exactly K-sparse signals, though many of the key ideas translate to compressible signals [28], [54]. In addition, we note that the CS concepts are also extendable to tight frames.)

As we discussed in Compression (Chapter 6), the standard procedure for compressing sparse signals, known as transform coding, is to (i) acquire the full N-sample signal x; (ii) compute the complete set of transform coefficients α; (iii) locate the K largest, significant coefficients and discard the (many) small coefficients; (iv) encode the **values and locations** of the largest coefficients.

This procedure has three inherent inefficiencies: First, for a high-dimensional signal, we must start with a large number of samples N. Second, the encoder must compute **all** N of the transform coefficients α, even though it will discard all but K of them. Third, the encoder must encode the locations of the large coefficients, which requires increasing the coding rate since the locations change with each signal.

8.2 Incoherent projections

This raises a simple question: For a given signal, is it possible to directly estimate the set of large $\alpha(n)$'s that will not be discarded? While this seems improbable, Candès, Romberg, and Tao [23], [28] and Donoho [54] have shown that a reduced set of projections can contain enough information to reconstruct sparse signals. An offshoot of this work, often referred to as **Compressed Sensing** (CS) [22], [28], [24], [25], [21], [54], [57], has emerged that builds on this principle.

[1] This content is available online at <http://cnx.org/content/m18733/1.5/>.

In CS, we do not measure or encode the K significant $\alpha(n)$ directly. Rather, we measure and encode $M < N$ projections $y(m) =< x, {\phi_m}^T >$ of the signal onto a **second set** of functions $\{\phi_m\}, m = 1, 2, ..., M$. In matrix notation, we measure

$$y = \Phi x, \tag{8.2}$$

where y is an $M \times 1$ column vector and the **measurement basis** matrix Φ is $M \times N$ with each row a basis vector ϕ_m. Since $M < N$, recovery of the signal x from the measurements y is ill-posed in general; however the additional assumption of signal **sparsity** makes recovery possible and practical.

The CS theory tells us that when certain conditions hold, namely that the functions $\{\phi_m\}$ cannot sparsely represent the elements of the basis $\{\psi_n\}$ (a condition known as **incoherence** of the two dictionaries [28], [23], [54], [99]) and the number of measurements M is large enough, then it is indeed possible to recover the set of large $\{\alpha(n)\}$ (and thus the signal x) from a similarly sized set of measurements y. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of a Fourier basis, or the Fourier basis and wavelets. Significantly, this incoherence also holds with high probability between an arbitrary fixed basis and a randomly generated one.

8.3 Methods for signal recovery

Although the problem of recovering x from y is ill-posed in general (because $x \in \mathbb{R}^N$, $y \in \mathbb{R}^M$, and $M < N$), it is indeed possible to recover **sparse** signals from CS measurements. Given the measurements $y = \Phi x$, there exist an infinite number of candidate signals in the shifted nullspace $\mathcal{N}(\Phi) + x$ that could generate the same measurements y (see Linear Models from Low-Dimensional Signal Models (Section 4.1: Linear models)). Recovery of the correct signal x can be accomplished by seeking a **sparse** solution among these candidates.

8.3.1 Recovery via combinatorial optimization

Supposing that x is exactly K-sparse in the dictionary Ψ, then recovery of x from y can be formulated as the ℓ_0 minimization

$$\hat{\alpha} = argmin \| \alpha \|_0 \quad \text{s.t. } y = \Phi\Psi\alpha. \tag{8.3}$$

Given some technical conditions on Φ and Ψ (see Theorem Section 8.3.1 (Recovery via combinatorial optimization)below), then with high probability this optimization problem returns the proper K-sparse solution α, from which the true x may be constructed. (Thanks to the incoherence between the two bases, if the original signal is sparse in the α coefficients, then no other set of sparse signal coefficients α' can yield the same projections y.) We note that the recovery program (8.3) can be interpreted as finding a K-term approximation to y from the columns of the dictionary $\Phi\Psi$, which we call the **holographic basis** because of the complex pattern in which it encodes the sparse signal coefficients [54].

In principle, remarkably few incoherent measurements are required to recover a K-sparse signal via ℓ_0 minimization. Clearly, more than K measurements must be taken to avoid ambiguity; the following theorem (which is proved in [7]) establishes that $K + 1$ random measurements will suffice. (Similar results were established by Venkataramani and Bresler [103].)

Theorem 8.1:
Let Ψ be an orthonormal basis for $\mathbb{R}^N$, and let $1 \leq K < N$. Then the following statements hold:

1. Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries with $M \geq 2K$. Then with probability one the following statement holds: all signals $x = \Psi\alpha$ having expansion coefficients $\alpha \in \mathbb{R}^N$ that satisfy $\| \alpha \|_0 = K$ can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0 optimization (8.3).
2. Let $x = \Psi\alpha$ such that $\| \alpha \|_0 = K$. Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries (notably, independent of x) with $M \geq K+1$. Then with probability one the

following statement holds: x can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0 optimization (8.3).

3. Let Φ be an $M \times N$ measurement matrix, where $M \leq K$. Then, aside from pathological cases (specified in the proof), no signal $x = \Psi\alpha$ with $\| \alpha \|_0 = K$ can be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$.

The second statement of the theorem differs from the first in the following respect: when $K < M < 2K$, there will necessarily exist K-sparse signals x that cannot be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$. However, these signals form a set of measure zero within the set of **all**K-sparse signals and can safely be avoided if Φ is randomly generated independently of x.

Unfortunately, as discussed in Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation), solving this ℓ_0 optimization problem is prohibitively complex. Yet another challenge is robustness; in the setting of Theorem "Recovery via ℓ 0 optimization" (Section 8.3.1: Recovery via combinatorial optimization), the recovery may be very poorly conditioned. In fact, **both** of these considerations (computational complexity and robustness) can be addressed, but at the expense of slightly more measurements.

8.3.2 Recovery via convex optimization

The practical revelation that supports the new CS theory is that it is not necessary to solve the ℓ_0-minimization problem to recover α. In fact, a much easier problem yields an equivalent solution (thanks again to the incoherency of the bases); we need only solve for the ℓ_1-sparsest coefficients α that agree with the measurements y [23], [22], [28], [24], [25], [21], [54], [57]

$$\hat{\alpha} = argmin \| \alpha \|_1 \quad \text{s.t. } y = \Phi\Psi\alpha. \tag{8.4}$$

As discussed in Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation), this optimization problem, also known as **Basis Pursuit** [35], is significantly more approachable and can be solved with traditional linear programming techniques whose computational complexities are polynomial in N.

There is no free lunch, however; according to the theory, more than $K + 1$ measurements are required in order to recover sparse signals via Basis Pursuit. Instead, one typically requires $M \geq cK$ measurements, where $c > 1$ is an **oversampling factor**. As an example, we quote a result asymptotic in N. For simplicity, we assume that the sparsity scales linearly with N; that is, $K = SN$, where we call S the **sparsity rate**.

Theorem 8.2:
[27], [56], [53] Set $K = SN$ with $0 < S \ll 1$. Then there exists an oversampling factor $c(S) = O(log(1/S))$, $c(S) > 1$, such that, for a K-sparse signal x in the basis Ψ, the following statements hold:

1. The probability of recovering x via Basis Pursuit from $(c(S) + \varepsilon) K$ random projections, $\varepsilon > 0$, converges to one as $N \to \infty$.
2. The probability of recovering x via Basis Pursuit from $(c(S) - \varepsilon) K$ random projections, $\varepsilon > 0$, converges to zero as $N \to \infty$.

In an illuminating series of recent papers, Donoho and Tanner [53], [56], [62] have characterized the oversampling factor $c(S)$ precisely (see also "The geometry of Compressed Sensing" (Section 8.5: The geometry of Compressed Sensing)). With appropriate oversampling, reconstruction via Basis Pursuit is also provably robust to measurement noise and quantization error [23].

We often use the abbreviated notation c to describe the oversampling factor required in various settings even though $c(S)$ depends on the sparsity K and signal length N.

A CS recovery example on the Cameraman test image is shown in Figure 8.1. In this case, with $M = 4K$ we achieve near-perfect recovery of the sparse measured image.

Figure 8.1: Compressive sensing reconstruction of the nonlinear approximation Cameraman image from Figure 5.2(b). Using $M = 16384$ random measurements of the K-term nonlinear approximation image (where $K = 4096$), we solve an ℓ_1-minimization problem to obtain the reconstruction shown above. The MSE with respect to the measured image is 0.08, so the reconstruction is virtually perfect.

8.3.3 Recovery via greedy pursuit

At the expense of slightly more measurements, iterative greedy algorithms such as Orthogonal Matching Pursuit (OMP) [99], Matching Pursuit (MP) [83], and Tree Matching Pursuit (TMP) [64], [76] have also been proposed to recover the signal x from the measurements y (see Nonlinear Approximation from Approximation (Section 5.2: Nonlinear approximation)). In CS applications, OMP requires $c \approx 2ln\,(N)$[99] to succeed with high probability. OMP is also guaranteed to converge within M iterations. We note that Tropp and Gilbert require the OMP algorithm to succeed in the first K iterations [99]; however, in our simulations, we allow the algorithm to run up to the maximum of M possible iterations. The choice of an appropriate practical stopping criterion (likely somewhere between K and M iterations) is a subject of current research in the CS community.

8.4 Impact and applications

CS appears to be promising for a number of applications in signal acquisition and compression. Instead of sampling a K-sparse signal N times, only cK incoherent measurements suffice, where K can be orders of magnitude less than N. Therefore, a sensor can transmit far fewer measurements to a receiver, which can reconstruct the signal and then process it in any manner. Moreover, the cK measurements need not be manipulated in any way before being transmitted, except possibly for some quantization. Finally, independent and identically distributed (i.i.d.) Gaussian or Bernoulli/Rademacher (random ± 1) vectors provide a useful **universal** basis that is incoherent with all others. Hence, when using a random basis, CS is universal in the sense that the sensor can apply the same measurement mechanism no matter what basis the signal is sparse in (and thus the coding algorithm is independent of the sparsity-inducing basis) [28], [54], [4].

These features of CS make it particularly intriguing for applications in remote sensing environments that might involve low-cost battery operated wireless sensors, which have limited computational and communication capabilities. Indeed, in many such environments one may be interested in sensing a **collection** of signals using a network of low-cost signals.

Other possible application areas of CS include imaging [97], medical imaging [23], [79], and RF environments (where high-bandwidth signals may contain low-dimensional structures such as radar chirps) [63]. As research continues into practical methods for signal recovery (see Section 8.3 (Methods for signal recovery)), additional work has focused on developing physical devices for acquiring random projections. Our group has developed, for example, a prototype digital CS camera based on a digital micromirror design [97]. Additional work suggests that standard components such as filters (with randomized impulse responses) could be useful in CS hardware devices [100].

8.5 The geometry of Compressed Sensing

It is important to note that the core theory of CS draws from a number of deep geometric arguments. For example, when viewed together, the CS encoding/decoding process can be interpreted as a linear projection $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^M$ followed by a nonlinear mapping $\Delta : \mathbb{R}^M \mapsto \mathbb{R}^N$. In a very general sense, one may naturally ask for a given class of signals $\mathcal{F} \in \mathbb{R}^N$ (such as the set of K-sparse signals or the set of signals with coefficients $\| \alpha \|_{\ell_p} \leq 1$), what encoder/decoder pair Φ, Δ will ensure the best reconstruction (minimax distortion) of all signals in $\mathcal{F}$. This best-case performance is proportional to what is known as the Gluskin n-width [71], [67] of $\mathcal{F}$ (in our setting $n = M$), which in turn has a geometric interpretation. Roughly speaking, the Gluskin n-width seeks the $(N-n)$-dimensional slice through $\mathcal{F}$ that yields signals of greatest energy. This n-width bounds the best-case performance of CS on classes of compressible signals, and one of the hallmarks of CS is that, given a sufficient number of measurements this optimal performance is achieved (to within a constant) [54], [48].

Additionally, one may view the ℓ_0/ℓ_1 equivalence problem geometrically. In particular, given the measurements $y = \Phi x$, we have an $(N-M)$-dimensional hyperplane $\mathcal{H}_y = \{x^{'} \in \mathbb{R}^N : y = \Phi x^{'}\} = \mathcal{N}(\Phi) + x$ of feasible signals that could account for the measurements y. Supposing the original signal x is K-sparse, the ℓ_1 recovery program will recover the correct solution x if and only if $\| x^{'} \|_1 > \| x \|_1$ for every other signal $x^{'} \in \mathcal{H}_y$ on the hyperplane. This happens only if the hyperplane $\mathcal{H}_y$ (which passes through x) does not "cut into" the ℓ_1-ball of radius $\| x \|_1$. This ℓ_1-ball is a polytope, on which x belongs to a $(K-1)$-dimensional "face." If Φ is a random matrix with i.i.d. Gaussian entries, then the hyperplane $\mathcal{H}_y$ will have random orientation. To answer the question of how M must relate to K in order to ensure reliable recovery, it helps to observe that a randomly generated hyperplane $\mathcal{H}$ will have greater chance to slice into the ℓ_1 ball as $\dim(\mathcal{H}) = N - M$ grows (or as M shrinks) or as the dimension $K-1$ of the face on which x lives grows. Such geometric arguments have been made precise by Donoho and Tanner [53], [56], [62] and used to establish a series of sharp bounds on CS recovery.

8.6 Connections with dimensionality reduction

We have also identified [4] a fundamental connection between the CS and the JL lemma. In order to make this connection, we considered the **Restricted Isometry Property** (RIP), which has been identified as a key property of the CS projection operator Φ to ensure stable signal recovery. We say Φ has RIP of order K if for every K-sparse signal x,

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}}. \tag{8.5}$$

A random $M \times N$ matrix with i.i.d. Gaussian entries can be shown to have this property with high probability if $M = O(Klog(N/K))$.

While the JL lemma concerns pairwise distances within a finite cloud of points, the RIP concerns isometric embedding of an **infinite** number of points (comprising a union of K-dimensional subspaces in $\mathbb{R}^N$). However, the RIP can in fact be derived by constructing an effective **sampling** of K-sparse signals in $\mathbb{R}^N$, using the JL lemma to ensure isometric embeddings for each of these points, and then arguing that the RIP must hold true for **all** K-sparse signals. (See [4] for the full details.)

8.7 Stable embeddings of manifolds

Finally, we have also shown that the JL lemma can also lead to extensions of CS to other concise signal models. In particular, while conventional CS theory concerns sparse signal models, it is also possible to consider manifold-based signal models. Just as random projections can preserve the low- dimensional geometry (the union of hyperplanes) that corresponds to a sparse signal family, random projections can also guarantee a stable embedding of a low-dimensional signal manifold. We have the following result, which states that an RIP-like property holds for families of manifold-modeled signals.

Theorem 8.3:
Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having condition number $\frac{1}{\tau}$, volume V, and geodesic covering regularity R. Fix $0 < \varepsilon < 1\text{and}0 < \rho < 1$. Let Φ be a random M × N orthoprojector with

$$M = O\left(\frac{K\log\left(NVR\tau^{-1}\varepsilon^{-1}\right)\log\left(\frac{1}{\rho}\right)}{\varepsilon^2}\right) \tag{8.6}$$

If $M \leq N$, then with probability at least $1-\rho$ the following statement holds: For every pair of points x_1, $x_2 \in \mathcal{M}$,

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}} \tag{8.7}$$

The proof of this theorem appears in [6] and again involves the JL lemma. Due to the limited complexity of a manifold model, it is possible to adequately characterize the geometry using a sufficiently fine sampling of points drawn from the manifold and its tangent spaces. In essence, manifolds with higher volume or with greater curvature have more complexity and require a more dense covering for application of the JL lemma; this leads to an increased number of measurements. The theorem also indicates that the requisite number of measurements depends on the geodesic covering regularity of the manifold, a minor technical concept which is also discussed in [6].

This theorem establishes that, like the class of K-sparse signals, a collection of signals described by a K-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$ can have a stable embedding in an M-dimensional measurement space. Moreover, the requisite number of random measurements M is once again linearly proportional to the information level (or number of degrees of freedom) K in the concise model. This has a number of possible implications for manifold-based signal processing. Manifold-modeled signals can be recovered from compressive measurements (using a customized recovery algorithm adapted to the manifold model, in contrast with sparsity-based recovery algorithms) [44], [104]; unknown parameters in parametric models can be estimated from compressive measurements; multi-class estimation/classification problems can be addressed [44] by considering multiple manifold models; and manifold learning algorithms may be efficiently executed by applying them simply to the projection of a manifold-modeled data set to a low-dimensional measurement space [17]. (As an example, Figure 7.1(d) shows the result of applying the ISOMAP algorithm on a random projection of a data set from $\mathbb{R}^{4096}$ down to $\mathbb{R}^{15}$; the underlying parameterization of the manifold is extracted with little sacrifice in accuracy.) In all of this it is not necessary to adapt the sensing protocol to the model; the only change from sparsity-based CS would be the methods for processing or decoding the measurements. In the future, more sophisticated concise models will likely lead to further improvements in signal understanding from compressive measurements.

Bibliography

[1] *Let it Wave*. www.letitwave.fr.

[2] D. Achlioptas. Database-friendly random projections. In *Proc. Symp. Principles of Database Systems*, 2001.

[3] F. Arandiga, A. Cohen, M. Doblas, R. Donat, and B. Matei. Sparse representations of images by edge adapted nonlinear multiscale transforms. In *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Barcelona, Spain, Sept. 2003.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. The johnson-lindenstrauss lemma meets compressed sensing. 2006. Preprint.

[5] R. G. Baraniuk and D. L. Jones. Shear madness:new orthogonal bases and frames using chirp functions. *IEEE Trans. Signal Proc.*, 41(12):3543–3549, 1993.

[6] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 2008. To Appear.

[7] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. 2005. Preprint.

[8] D. M. Bates and D. G. Watts. *Nonlinear Regression Analysis and Its Applications*. John Wiley and Sons, New York, 1988.

[9] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), June 2003.

[10] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), June 2003.

[11] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, revised 2nd edition, 2003.

[12] M. Brand. Charting a manifold. In *Proc. Neural Inform. Processing Systems - NIPS*, 2002.

[13] D. S. Broomhead and M. Kirby. A new approach for dimensionality reduction: Theory and algorithms. *SIAM J. of Applied Mathematics*, 60(6), 2000.

[14] D. S. Broomhead and M. Kirby. A new approach for dimensionality reduction: Theory and algorithms. *SIAM J. of Applied Mathematics*, 60(6), 2000.

[15] D. S. Broomhead and M. J. Kirby. The whitney reduction network: A method for computing autoassociative graphs. *Neural Computation*, 13:2595–2616, 2001.

[16] D. S. Broomhead and M. J. Kirby. The whitney reduction network: A method for computing autoassociative graphs. *Neural Computation*, 13:2595–2616, 2001.

[17] M.B. Wakin C. Hegde and R.G. Baraniuk. Random projections for manifold learning. In *In Proc. Neural Information Processing Systems (NIPS)*, December 2007.

[18] E. Cand[U+FFFD]nd D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise singularities. *Comm. on Pure and Applied Math.*, 57:2198211;266, 2004.

[19] E. Cand[U+FFFD]nd D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise singularities. *Comm. on Pure and Applied Math.*, 57:219–266, 2004.

[20] E. Cand[U+FFFD]nd D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise singularities. *Comm. on Pure and Applied Math.*, 57:219–266, 2004.

[21] E. Cand[U+FFFD]nd J. Romberg. Practical signal recovery from random projections. 2005. Preprint.

[22] E. Cand[U+FFFD]nd J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. of Comp. Math.*, 2006. To appear.

[23] E. Cand[U+FFFD] J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2), February 2006.

[24] E. Cand[U+FFFD] J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 2006. To appear.

[25] E. Cand[U+FFFD]nd T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12), December 2005.

[26] E. Cand[U+FFFD]nd T. Tao. Error correction via linear programming. *Found. of Comp. Math.*, 2005. Preprint.

[27] E. Cand[U+FFFD]nd T. Tao. Error correction via linear programming. *Found. of Comp. Math.*, 2005. Preprint.

[28] E. Cand[U+FFFD]nd T. Tao. Near optimal signal recovery from random projections and universal encoding strategies. *IEEE Trans. Inform. Theory*, 2006. To appear.

[29] E. J. Cand[U+FFFD]nd D. L. Donoho. Curvelets 8212; a suprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L. L. Schumaker, editors, *Curve and Surface Fitting*. Vanderbilt University Press, 1999.

[30] E. J. Cand[U+FFFD]nd D. L. Donoho. Curvelets 8212; a suprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L. L. Schumaker, editors, *Curve and Surface Fitting*. Vanderbilt University Press, 1999.

[31] E. J. Cand[U+FFFD]nd D. L. Donoho. Curvelets: A suprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L. L. Schumaker, editors, *Curve and Surface Fitting*. Vanderbilt University Press, 1999.

[32] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. *Int. J. of Shape Modeling*. To appear.

[33] V. Chandrasekaran, M. B. Wakin, D. Baron, and R. Baraniuk. Representation and compression of multi-dimensional piecewise functions using surflets. to appear in {\em IEEE Trans. Inf. Theory}, 2008.

[34] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Sci. Comp.*, 20(1):33–61, 1998.

[35] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Sci. Comp.*, 20(1):33–61, 1998.

[36] G. F. Clements. Entropies of several sets of real valued functions. *Pacific J. Math.*, 13:1085–1095, 1963.

[37] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11:192–226, 2001.

[38] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11:192–226, 2001.

[39] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard. On the importance of combining wavelet-based nonlinear approximation with coding strategies. *IEEE Trans. Inform. Theory*, 48(7):1895–1921, July 2002.

[40] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard. On the importance of combining wavelet-based nonlinear approximation with coding strategies. *IEEE Trans. Inform. Theory*, 48(7):1895–1921, July 2002.

[41] R. R. Coifman and M. Maggioni. Diffusion wavelets. *Appl. Comput. Harmon. Anal.*, 2005. To appear.

[42] J. A. Costa and A. O. Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. Signal Processing*, 52(8), August 2004.

[43] S. Dasgupta and A. Gupta. An elementary proof of the johnson-lindenstrauss lemma. Technical report TR-99-006, Berkeley, CA, 1999.

[44] M.A. Davenport, M.F. Duarte, M.B. Wakin, J.N. Laska, D. Takhar, K.F. Kelly, and R.G. Baraniuk. The smashed filter for compressive classification and target recognition. In *Proc. Computational Imaging V at SPIE Electronic Imaging*, January 2007.

[45] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[46] R. A. DeVore. Lecture notes on compressed sensing. *Rice University ELEC 631 Course Notes*, Spring 2006.

[47] R. A. DeVore. Lecture notes on compressed sensing. *Rice University ELEC 631 Course Notes*, Spring 2006.

[48] R. A. DeVore. Lecture notes on compressed sensing. *Rice University ELEC 631 Course Notes*, Spring 2006.

[49] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Inform. Theory*, 38(2):719–746, Mar. 1992.

[50] M. N. Do and M. Vetterli. Contourlets: A directional multiresolution image representation. In *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Rochester, New York, Oct. 2002.

[51] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Trans. Image Processing*, 2005. To appear.

[52] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Trans. Image Processing*, 2005. To appear.

[53] D. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. January 2005. Preprint.

[54] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4), April 2006.

[55] D. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. 2005. Preprint.

[56] D. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. 2005. Preprint.

[57] D. Donoho and Y. Tsaig. Extensions of compressed sensing. 2004. Preprint.

[58] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.*, 1(1):100–115, Dec. 1993.

[59] D. L. Donoho. Denoising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, May 1995.

[60] D. L. Donoho and C. E. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA*, 100(10):5591–5596, May 2003.

[61] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when then projection radically lowers dimension. Technical report 2006-11, Stanford University Department of Statistics, 2006.

[62] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when then projection radically lowers dimension. Technical report 2006-11, Stanford University Department of Statistics, 2006.

[63] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk. Sparse signal detection from incoherent projections. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 2006.

[64] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk. Fast reconstruction of piecewise smooth signals from random projections. In *Proc. SPARS05*, Rennes, France, Nov. 2005.

[65] F. C. A. Fernandes, R. L. C. van Spaendonck, and C. S. Burrus. A new framework for complex wavelet transforms. *IEEE Trans. Signal Processing*, July 2003.

[66] F. C. A. Fernandes, M. B. Wakin, and R. G. Baraniuk. Non-redundant, linear-phase, semi-orthogonal, directional complex wavelets. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, Quebec, Canada, May 2004.

[67] A. Garnaev and E. D. Gluskin. The widths of euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

[68] M. W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Springer, 1976.

[69] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimenstionality. In *Proc. Symp. Theory of Computing*, pages 604–613, 1998.

[70] W. B Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. In *Proc. Conf. Modern Analysis and Probability*, pages 189–206, 1984.

[71] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia*, (41):334–351, 1977.

[72] N. Kingsbury. Image processing with complex wavelets. *Phil. Trans. R. Soc. Lond. A*, 357, Sept. 1999.

[73] N. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comp. Harm. Anal.*, 10:234–253, 2001.

[74] A. N. Kolmogorov and V. M. Tihomirov. -entropy and -capacity of sets in functional spaces. *Amer. Math. Soc. Transl. (Ser. 2)*, 17:277–364, 1961.

[75] J. Kova269;evi263; and A. Chebira. Life beyond bases: The advent of frames. 2006. Preprint.

[76] C. La and M. N. Do. Signal reconstruction using sparse tree representation. In *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, August 2005. SPIE.

[77] E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Trans. Image Processing*, 14(4):423–438, April 2005.

[78] S. LoPresto, K. Ramchandran, and M. T. Orchard. Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework. In *Proc. Data Compression Conf.*, pages 221–230, Snowbird, Utah, March 1997.

[79] M. Lustig, D. L. Donoho, and J. M. Pauly. Rapid mr imaging with compressed sensing and randomly under-sampled 3dft trajectories. In *Proc. 14th Ann. Mtg. ISMRM*, May 2006.

[80] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, CA, USA, 1999.

[81] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, CA, USA, 1999.

[82] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, CA, USA, 1999.

[83] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, CA, USA, 1999.

[84] D. Marr. *Vision*. W. H. Freeman and Company, San Francisco, 1982.

[85] N. Mehrseresht and D. Taubman. An efficient content-adaptive motion compensated 3d-dwt with enhanced spatial and temporal scalability. 2004. Preprint.

[86] F. Morgan. *Riemannian Geometry: A Beginner's Guide*. A K Peters, 2nd edition, 1998.

[87] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with confidence from random samples. 2004. Preprint.

[88] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with confidence from random samples. 2004. Preprint.

[89] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Res.*, 37:311–3325, 1997.

[90] B. O'Neill. *Elementary Differential Geometry*. Harcourt Academic Press, 2nd edition, 1997.

[91] M. T. Orchard and H. Ates. Equiripple design of real and complex filter banks. Technical report, Rice University, 2003.

[92] I. Ur Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schroeder. Multiscale representations for manifold-valued data. 2004. Preprint.

[93] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[94] I. W. Selesnick. The design of approximate hilbert transform pairs of wavelet bases. *IEEE Trans. Signal Processing*, 50(5), May 2002.

[95] I. W. Selesnick and K. L. Li. Video denoising using 2d and 3d dual-tree complex wavelet transforms. In *Proc. SPIE Wavelet Applications Signal Image Processing X*, 2003.

[96] J. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Processing*, 41(12):3445–3462, Dec. 1993.

[97] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K. F. Kelly, and R. G. Baraniuk. A compressed sensing camera: New theory and an implementation using digital micromirrors. In *Proc. Computational Imaging IV at SPIE Electronic Imaging*, San Jose, January 2006. SPIE.

[98] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[99] J. Tropp and A. C. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. April 2005. Preprint.

[100] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 2006.

[101] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1), 1991.

[102] R. van Spaendonck, T. Blu, R. Baraniuk, and M. Vetterli. Orthogonal hilbert transform filter banks and wavelets. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2003.

[103] R. Venkataramani and Y. Bresler. Further results on spectrum blind sampling of 2d signals. In *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, volume 2, Chicago, Oct. 1998.

[104] M. B. Wakin. *The Geometry of Low-Dimensional Signal Models*. Ph. d. thesis, department of electrical and computer engineering, Rice University, Houston, Tx, August 2006.

[105] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk. High-resolution navigation on non-differentiable image manifolds. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*. IEEE, 2005.

[106] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *Int. J. Computer Vision 8211; Special Issue: Computer Vision and Pattern Recognition-CVPR 2004*, 70(1):77–90, 2006.

[107] Z. Xiong, K. Ramchandran, and M. T. Orchard. Space-frequency quantization for wavelet image coding. *IEEE Trans. Image Processing*, 6(5):677–693, 1997.

[108] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via tangent space alignment. *SIAM J. Scientific Comput.*, 26(1), 2004.

Attributions

Collection: *Concise Signal Models*
Edited by: Michael Wakin
URL: http://cnx.org/content/col10635/1.4/
License: http://creativecommons.org/licenses/by/2.0/

Module: "Introduction to Concise Signal Models"
By: Michael Wakin
URL: http://cnx.org/content/m18720/1.5/
Pages: 1-2
Copyright: Michael Wakin
License: http://creativecommons.org/licenses/by/2.0/

Module: "Signal Dictionaries and Representations"
By: Michael Wakin
URL: http://cnx.org/content/m18724/1.5/
Pages: 3-9
Copyright: Michael Wakin
License: http://creativecommons.org/licenses/by/2.0/

Module: "Manifolds"
By: Michael Wakin
URL: http://cnx.org/content/m18722/1.4/
Pages: 11-13
Copyright: Michael Wakin
License: http://creativecommons.org/licenses/by/2.0/

Module: "Low-Dimensional Signal Models"
By: Michael Wakin
URL: http://cnx.org/content/m18726/1.4/
Pages: 15-19
Copyright: Michael Wakin
License: http://creativecommons.org/licenses/by/2.0/

Module: "Approximation"
By: Michael Wakin
URL: http://cnx.org/content/m18727/1.5/
Pages: 21-27
Copyright: Michael Wakin
License: http://creativecommons.org/licenses/by/2.0/

Module: "Compression"
By: Michael Wakin
URL: http://cnx.org/content/m18729/1.3/
Pages: 29-30
Copyright: Michael Wakin
License: http://creativecommons.org/licenses/by/2.0/

Module: "Dimensionality Reduction"
By: Michael Wakin
URL: http://cnx.org/content/m18732/1.5/
Pages: 31-34

Module: "Compressed Sensing"
By: Michael Wakin
URL: http://cnx.org/content/m18733/1.5/
Pages: 35-40

Concise Signal Models
This collection reviews fundamental concepts underlying the use of concise models for signal processing. Topics are presented from a geometric perspective and include low-dimensional linear, sparse, and manifold-based signal models, approximation, compression, dimensionality reduction, and Compressed Sensing.

About Connexions
Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities.

Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai. Connexions is part of an exciting new information distribution system that allows for **Print on Demand Books**. Connexions has partnered with innovative on-demand publisher QOOP to accelerate the delivery of printed course materials and textbooks into classrooms worldwide at lower prices than traditional academic publishers.