Sparse Geodesic Paths

Mark A. Davenport and **Richard G. Baraniuk**
Department of Electrical and Computer Engineering
Rice University

Abstract

In this paper we propose a new distance metric for signals that admit a sparse representation in a known basis or dictionary. The metric is derived as the length of the sparse geodesic path between two points, by which we mean the shortest path between the points that is itself sparse. We show that the distance can be computed via a simple formula and that the entire geodesic path can be easily generated. The distance provides a natural similarity measure that can be exploited as a perceptually meaningful distance metric for natural images. Furthermore, the distance has applications in supervised, semi-supervised, and unsupervised learning settings.

1. Introduction

Distance metrics play a fundamental role in virtually all aspects of data processing. They are used to compare signals and to quantify how "similar" or "close" two signals are in such varied tasks as compression, filtering, clustering, anomaly detection, regression, and classification. The most common distance metric is the Euclidean, or ℓ_2, distance. This distance can be probabilistically motivated by assuming that a signal of interest has been corrupted with white Gaussian noise, in which case the ℓ_2 distance has a natural interpretation. There are a variety of common alternative distance metrics, including the ℓ_p norms for $p \neq 2$, but in the absence of any additional structure on the data, it is difficult to justify picking any of these over the Euclidean distance.

More recently, however, we have begun to appreciate the importance of incorporating data models in a variety of settings. Models are useful for distinguishing classes of interesting or probable signals from uninteresting or improbable signals. Models typically are effective because they enforce some low-dimensional structure exhibited by the data and thereby help us to avoid the curse of dimensionality. They also provide a mechanism for incorporating *a priori* knowledge about the data. In general, one can think of a data model simply as a set $\mathcal{X}$ near which the data is likely to lie. While this knowledge can be exploited in many possible ways, a natural approach is to reconsider our habitual use of the Euclidean distance and instead consider a distance more suitable for the space $\mathcal{X}$ in which our data lives. If we think of the Euclidean distance as the length of a straight line between two points, then we can see that unless $\mathcal{X}$ is a convex set, there will exist points where the straight line connecting them does not lie within $\mathcal{X}$. Hence, for non-convex $\mathcal{X}$ it is natural to replace the Euclidean distance with the *geodesic distance*. The geodesic distance for a set $\mathcal{X}$ is the length of the shortest path between a pair of points that never exits $\mathcal{X}$. A geodesic path is typically interpreted as a generalization of the notion of a "straight line" to an arbitrary space $\mathcal{X}$.

These ideas play a prominent role in a popular and powerful class of data models: low-dimensional *manifolds*. It has been well-established that if the data has a low-dimensional manifold structure, then exploiting this knowledge can lead to significant performance gains in a variety of data processing tasks (Donoho and Grimes 2005; Belkin and Niyogi 2004; Niyogi 2008). Nearly all methods for exploiting manifold structure ultimately rely on replacing the Euclidean distance with the geodesic distance. Unfortunately, while the manifold assumption is potentially powerful, in practice the manifold is typically not known, and so it is usually not possible to explicitly compute the geodesic distance. Given a sufficiently large data set, it is possible to estimate the geodesic distance between data points using various algorithms (J. Tenenbaum and Langford 2000; Roweis and Saul 2000; Donoho and Grimes 2003; Belkin and Niyogi 2003; Coifman and Maggioni 2006; Weinberger and Saul 2006). Unfortunately, these methods typically require a very large amount of data to obtain good estimates.

While in many applications the data exhibits a low-dimensional manifold structure, another large class of applications features data with low-dimensional *sparsity* structure. Sparse signals can be well-approximated by a linear combination of just a few elements from some basis or dictionary (DeVore 1998). Sparsity is a highly nonlinear model, since the choice of which elements are in the linear combination can change from signal to signal. This distinguishes sparsity models from more conventional models like subspaces and principal component analysis. In fact, it is easy to show that the set of all sparse signals consists of not one subspace but the union of a combinatorial number of subspaces. In contrast to a manifold model, a sparse model is completely characterized by the choice of basis/dictionary and the number of elements K in the linear combination.

Sparsity has been exploited heavily in fields such as *image processing* for tasks such as compression and denoising (Donoho 1995), since the multiscale wavelet transform (Mallat 1999) provides concise representations for many natural images. Sparsity also figures prominently in the nascent theory of *compressive sensing*, where sparse signals are stably recovered from just a few linear measurements via an optimization or greedy algorithm (Donoho 2006; Candès 2006), and in the study of the human visual system (Olshausen and Field 1996).

In this paper, we introduce the notion of the geodesic distance between two K-sparse signals, which we define as the shortest path between the two signals such that each point on the path is also K-sparse. Below we derive a simple formula for the distance and describe a simple algorithm for generating the geodesic path between any two sparse signals. After studying the distance's properties, we examine its use as a perceptually meaningful error metric for image analysis and discuss additional potential applications.

2. Sparse Geodesic Paths

2.1 Sparsity and unions of subspaces

A signal is K-*sparse* with respect to the basis or dictionary Ψ when we can represent x as $x = \Psi\alpha$ with $\|\alpha\|_0 \leq K$.[1] Denote the set of all sparse signals as $\Sigma_K(\Psi) = \{x \in \mathbb{R}^N : x = \Psi\alpha \text{ where } \|\alpha\|_0 \leq K\}$. One can see that $\Sigma_K(\Psi)$ can essentially be thought of as the union of all possible K-dimensional subspaces obtained by picking K elements of Ψ. Note that this definition is equally valid for the cases where Ψ represents an orthonormal *basis* and where Ψ is an over-complete *dictionary*. In the case where Ψ is overcomplete, the representation $x = \Psi\alpha$ is not unique, and thus even in the case where x has a K-sparse representation α, it may not be easy to find. However, this does not affect the geometry of $\Sigma_K(\Psi)$. For our purposes, the only significant difference between the two cases is that if Ψ is an orthonomral basis then if two signals x and y have representations with disjoint support, then we also have that x and y are orthogonal. This does not follow if Ψ is an overcomplete dictionary. Since this distinction will have relatively limited import, we will simply restrict our attention to the case where Ψ is the $N \times N$ identity matrix, and we will omit the dependence on Ψ from our notation. In Figure 1 (a) and (b) we illustrate Σ_K where $K = 2$ and $N = 3$ for the cases where $\Psi = I$ and where $\Psi \neq I$ is an orthonormal basis. Our goal is to obtain a simple formula for the geodesic distance, or the length of a geodesic path such as the one illustrated in Figure 1 (c).

2.2 Geodesic distance

In order to define the geodesic distance and the corresponding geodesic path for any space $\mathcal{X}$, we must first have some notion of the length of a path. We represent a path from a point x to a point y as a continuous function ϕ from the unit interval $[0, 1]$ to $\mathbb{R}^N$ such that $\phi(0) = x$ and $\phi(1) = y$. Next consider a partition $\bar{t}$ of the interval $[0, 1]$, by which we mean a sequence $\bar{t}_0, \bar{t}_1, \ldots, \bar{t}_M$ such that $0 = \bar{t}_0 < \bar{t}_1 < \ldots < \bar{t}_M = 1$. We let $T([0, 1])$ denote the set of all possible partitions of $[0, 1]$, where M is unbounded. A particular partition defines a sequence of $M + 1$ points on the path given by $\phi(\bar{t}_0), \phi(\bar{t}_1), \ldots, \phi(\bar{t}_M)$. Without placing any assumptions on the differentiability of ϕ, the length of the path ϕ is

$$L(\phi) = \sup_{\bar{t} \in T([0,1])} \sum_{m=0}^{M-1} \|\phi(\bar{t}_m) - \phi(\bar{t}_{m+1})\|_2. \quad (1)$$

In order to find the geodesic path, we must first restrict ourselves to paths that lie within our space of interest $\mathcal{X}$. Let $\Phi_{\mathcal{X}}(x, y)$ denote the set of all paths that satisfy $\phi(0) = x$, $\phi(1) = y$, and $\phi(t) \in \mathcal{X}$ for all $t \in [0, 1]$. The geodesic path from x to y in $\mathcal{X}$ is thus defined as

$$\gamma = \operatorname*{arg\,inf}_{\phi \in \Phi_{\mathcal{X}}(x,y)} L(\phi). \quad (2)$$

The geodesic distance is denoted by

$$d_{\mathcal{X}}(x, y) = L(\gamma). \quad (3)$$

Note that the γ may not be uniquely defined. For instance, a pair of antipodal points on a sphere can be connected by an infinite number of geodesic paths. However, the length of each path is the same, and thus $d_{\mathcal{X}}(x, y)$ is well-defined.

2.3 The sparse geodesic distance

We now consider the specific case where $\mathcal{X} = \Sigma_K$. In this case, for a particular path $\phi \in \Phi_{\Sigma_K}(x, y)$, each $t \in [0, 1]$ corresponds to a point $\phi(t) \in \mathbb{R}^N$ that satisfies $\|\phi(t)\|_0 \leq K$. We define $S_x = \text{supp}(x)$, i.e., S_x is the subset of at most K indices that correspond to the nonzeros of x. Similarly, $S_y = \text{supp}(y)$. We will assume throughout the following that $|S_x| = |S_y| = K$, since if $|S_x| \neq |S_y|$ we can always set $K = \max(|S_x|, |S_y|)$ and observe that both $x, y \in \Sigma_K$. In cases where we require that $|S_x| = |S_y|$, note that we can also always simply enlarge the smaller set by arbitrarily adding indices to equalize the size of the two sets.

We now begin by establishing the following elementary bounds on $d_{\Sigma_K}(x, y)$.

Proposition 1. *For any $x, y \in \Sigma_K$,*

$$\|x - y\|_2 \leq d_{\Sigma_K}(x, y) \leq \|x\|_2 + \|y\|_2. \quad (4)$$

Furthermore, if $S_x \cap S_y = \emptyset$, then[2]

$$\sqrt{\|x\|_2^2 + \|y\|_2^2} \leq d_{\Sigma_K}(x, y) \leq \|x\|_2 + \|y\|_2. \quad (5)$$

Proof. The upper bounds in (4) and (5) follow from the simple observation that there always exists a path $\phi \in \Phi_{\Sigma_K}(x, y)$ that consists of a straight line from x to the origin, and then a straight line from the origin to y, which results in a length of $\|x\|_2 + \|y\|_2$. The lower bounds are obtained by observing that $d_{\Sigma_K}(x, y)$ can be no less than the length of a straight line connecting x and y. We obtain (5) from (4) by simply enforcing the fact that if $S_x \cap S_y = \emptyset$, then x and y are orthogonal. □

[1] $\|\cdot\|_0$ denotes the ℓ_0 quasi-norm, which simply counts the number of nonzero entries of a vector.

[2] Note that the result in (5) holds only for the case where Ψ is an orthonormal basis, while the result in (4) in fact holds for the case where Ψ is an overcomplete dictionary.

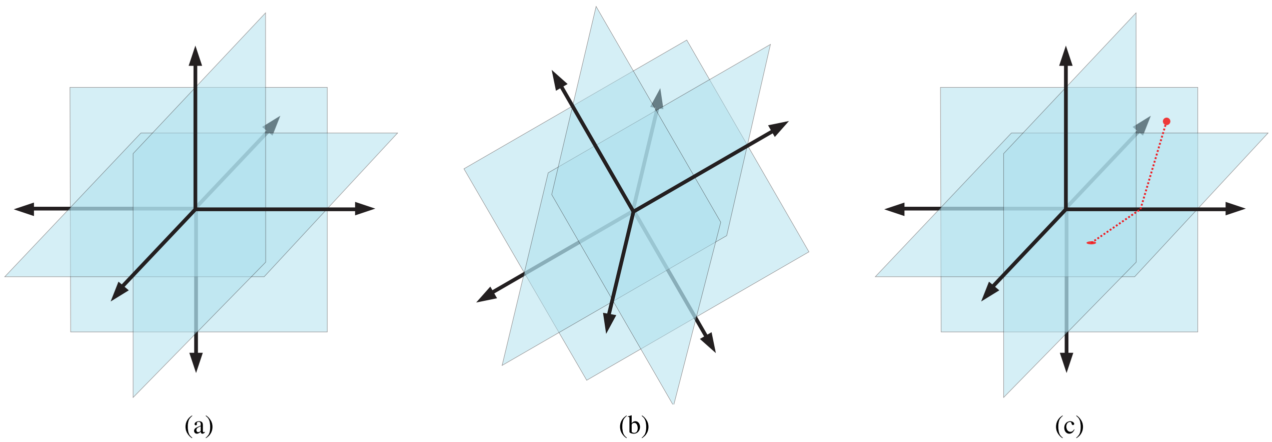


Figure 1: A depiction of Σ_2 embedded in $\mathbb{R}^3$. In (a) we have $\Sigma_2(I)$, i.e., the set of all 2-sparse signals in the canonical basis. In (b) we depict $\Sigma_2(\Psi)$, which is simply a rotated version of $\Sigma_2(I)$. In (c) we illustrate an example of a geodesic path in $\Sigma_2(I)$.

We now turn our attention to how one might calculate $d_{\Sigma_K}(x, y)$. We begin by observing that any geodesic path must be well-behaved in a certain sense, so that we can restrict our attention to a reduced set of candidate paths. To do so we will need to establish some notation. Let $\phi_i^{-1}(0) = \{t : \phi_i(t) = 0\}$, and define $I_{\mathrm{z}}(\phi) = \{i : \phi_i^{-1}(0) \neq \emptyset\}$ and $I_{\mathrm{nz}}(\phi) = \{i : \phi_i^{-1}(0) = \emptyset\}$. Essentially $I_{\mathrm{z}}(\phi)$ is the set of indices for which $\phi_i(t) = 0$ for some value of t, and $I_{\mathrm{nz}}(\phi)$ is the set of indices for which $\phi_i(t) \neq 0$ for all t. Furthermore, for all $i \in I_{\mathrm{z}}(\phi)$, define

$$t_i = \inf_{t \in \phi_i^{-1}(0)} t \qquad \text{and} \qquad r_i = \sup_{t \in \phi_i^{-1}(0)} t. \tag{6}$$

Using this notation, we now demonstrate that any geodesic path is well-behaved in terms of its zero-crossings.

Lemma 1. *Suppose that $\phi \in \Phi_{\Sigma_K}(x, y)$. If ϕ is a geodesic path, then for all $i \in I_{\mathrm{z}}(\phi)$*

$$\phi_i(t) = 0 \tag{7}$$

for all $t \in [t_i, r_i]$.

Proof. The result follows from the assumption that ϕ is a geodesic path. To see this, recall that

$$\begin{aligned} L(\phi) &= \sup_{\bar{t} \in T([0,1])} \sum_{m=0}^{M-1} \|\phi(\bar{t}_m) - \phi(\bar{t}_{m+1})\|_2 \\ &= \sup_{\bar{t} \in T([0,1])} \sum_{m=0}^{M-1} \sqrt{\sum_{n=1}^{N} (\phi_n(\bar{t}_m) - \phi_n(\bar{t}_{m+1}))^2}. \end{aligned}$$

Suppose for the sake of a contradiction that $\phi_i^{-1}(0) \neq \emptyset$ and that there exists a $t' \in [t_i, r_i]$ such that $\phi_i(t') \neq 0$. Since the supremum is over all partitions $\bar{t}$, we are free to require that $t_i, t' \in \bar{t}$. For any such partition, there exists an m with $\bar{t}_m, \bar{t}_{m+1} \in [t_i, r_i]$ such that the term $(\phi_i(\bar{t}_m) - \phi_i(\bar{t}_{m+1}))^2 > 0$. However, if we modify ϕ by setting $\widetilde{\phi}_i(t) = 0$ for all $t \in [t_i, r_i]$ and $\widetilde{\phi}(t) = \phi(t)$ otherwise, then $\widetilde{\phi}_i$ will contribute zero to the terms in the sum where $\bar{t}_m, \bar{t}_{m+1} \in [t_i, r_i]$. Thus $L(\widetilde{\phi}) < L(\phi)$, and since the new path will still satisfy $\widetilde{\phi}(t) \in \Sigma_K$ for all $t \in [0, 1]$, this contradicts the assumption that ϕ is a geodesic path. □

From this we also obtain the following useful corollary.

Corollary 1. *Suppose that $\phi \in \Phi_{\Sigma_K}(x, y)$ and define $\alpha, \beta : [0, 1] \to \mathbb{R}^N$ as follows: For $i \in I_{\mathrm{z}}(\phi)$ set t_i and r_i according to (6), and define*

$$\alpha_i(t) = \begin{cases} \phi_i(t) & \text{for } t \in [0, t_i) \\ 0 & \text{for } t \in [t_i, 1] \end{cases} \tag{8}$$

and

$$\beta_i(t) = \begin{cases} 0 & \text{for } t \in [0, r_i] \\ \phi_i(t) & \text{for } t \in (r_i, 1]. \end{cases} \tag{9}$$

For $i \in I_{\mathrm{nz}}(\phi)$, define $\alpha_i(t) = \phi_i(t)$ and $\beta_i(t) = 0$. If ϕ is a geodesic path, then

$$\phi_i(t) = \alpha_i(t) + \beta_i(t) \tag{10}$$

and

$$\|\phi(t)\|_0 = \|\alpha(t)\|_0 + \|\beta(t)\|_0. \tag{11}$$

Proof. Suppose that $i \in I_{\mathrm{z}}(\phi)$. By definition, $t_i \leq r_i$, so that whenever $\alpha_i(t) \neq 0$, $\beta_i(t) = 0$, and whenever $\beta_i(t) \neq 0$, $\alpha_i(t) = 0$. Thus

$$\alpha_i(t) + \beta_i(t) = \begin{cases} \phi_i(t) & \text{for } t \in [0, t_i) \\ 0 & \text{for } t \in [t_i, r_i] \\ \phi_i(t) & \text{for } t \in (r_i, 1]. \end{cases}$$

From Lemma 1 we have that $\phi_i(t) = 0$ for all $t \in [t_i, r_i]$, so that this reduces to (10). For $i \in I_{\mathrm{nz}}(\phi)$, (10) holds trivially. Furthermore, this means that $\|\phi(t)\|_0 = \|\alpha(t) + \beta(t)\|_0$, but since $\alpha_i(t) \neq 0$ whenever $\beta_i(t) = 0$ and vice versa, this reduces to (11). □

We now introduce the concept of a "matching" between the nonzero coefficients of x and the nonzero coefficients of y, i.e., the elements of S_x and S_y. Matchings will provide a useful method for characterizing a path ϕ.

Definition 1. *A* matching *between S_x and S_y is sequence of K pairs of indices, $\mathcal{M} = \{(i_1, j_1), (i_2, j_2), \ldots, (i_K, j_K)\}$, such that each $i_k \in S_x$ is listed exactly once and similarly for each $j_k \in S_y$.*

Our goal is now to relate the task of finding the geodesic path to the task of identifying the optimal matching. Towards this end, we first note that a particular path $\phi \in \Phi_{\Sigma_K}(x, y)$ will only be "compatible" with certain matchings.

Definition 2. *We say that the path $\phi \in \Phi_{\Sigma_K}(x, y)$ is* compatible *with a matching $\mathcal{M}$ if for any $(i, j) \in \mathcal{M}$, we either have that $i, j \in I_{\mathrm{nz}}(\phi)$ and $i = j$ or that $i, j \in I_{\mathrm{z}}(\phi)$ and the t_i and r_j as defined in (6) satisfy $t_i \le r_j$.*

The key idea is that if we match a pair (i, j) such that $i \neq j$, then $\phi_i(t)$ must become zero before $\phi_j(t)$ becomes nonzero.

Theorem 1. *Suppose that $\phi \in \Phi_{\Sigma_K}(x, y)$. If ϕ is a geodesic path, then it is compatible with at least one matching $\mathcal{M}$.*

Proof. For any $i \in I_{\mathrm{nz}}(\phi)$, we add (i, i) to $\mathcal{M}$ and let $\ell = |I_{\mathrm{nz}}(\phi)|$. For the $K - \ell$ indices $i \in S_x \cap I_{\mathrm{z}}(\phi)$ and $j \in S_y \cap I_{\mathrm{z}}(\phi)$ let t_i and r_j be defined as in (6). We next sort the indices so that $i_1, i_2, \ldots, i_{K-\ell}$ and $j_1, j_2, \ldots, j_{K-\ell}$ satisfy $t_{i_1} \le t_{i_2} \le \cdots \le t_{i_{K-\ell}}$ and $r_{j_1} \le r_{j_2} \le \cdots \le r_{j_{K-\ell}}$. Our goal is to claim that by adding the (i_k, j_k) to $\mathcal{M}$ we have a valid matching, i.e., $t_{i_k} \le r_{j_k}$ for all k. Towards this end, observe that if ϕ is a geodesic path, then Corollary 1 implies that

$$\begin{aligned}\|\phi(t)\|_0 &= \|\alpha(t)\|_0 + \|\beta(t)\|_0 \\ &= \ell + |\{i_k : t < t_{i_k}\}| + |\{j_k : t > r_{j_k}\}|, \qquad (12)\end{aligned}$$

where the second equality follows from the facts that: *(i)* $\alpha_{i_k}(t) \neq 0$ if and only if $i \in I_{\mathrm{nz}}(\phi)$ or $i \in I_{\mathrm{z}}(\phi)$ and $t < t_{i_k}$, and *(ii)* $\beta_{j_k}(t) \neq 0$ if and only if $i \in I_{\mathrm{z}}(\phi)$ and $t > r_{j_k}$.

We now assume for the sake of a contradiction that $r_{j_m} < t_{i_m}$ for some m. Setting $t \in (r_{j_m}, t_{i_m})$, we wish to calculate $\|\phi(t)\|_0$. Since

$$t < t_{i_m} \le t_{i_{m+1}} \le \cdots \le t_{i_{K-\ell}}$$

we have that $\{i_k : t < t_{i_k}\}$ must contain (at least) $t_{i_m}, t_{i_{m+1}}, \ldots, t_{i_{K-\ell}}$. Thus, we have that

$$|\{i_k : t < t_{i_k}\}| \ge K - \ell - (m-1). \qquad (13)$$

Similarly, since

$$t > r_{j_m} \ge r_{j_{m-1}} \ge \cdots \ge r_{j_1}$$

we have that $\{j_k : t > r_{j_k}\}$ must contain (at least) $r_{j_1}, r_{j_2}, \ldots, r_{j_m}$. Thus, we also have that

$$|\{j_k : t > r_{j_k}\}| \ge m. \qquad (14)$$

Combining (13) and (14) and plugging these into (12) we obtain that for $t \in (r_{j_m}, t_{i_m})$,

$$\|\phi(t)\|_0 \ge \ell + (K - \ell - (m-1)) + m = K + 1.$$

This contradicts the assumption that $\phi \in \Phi_{\Sigma_K}(x, y)$, and hence we must in fact have that $r_{j_m} \ge t_{i_m}$. Since this holds for all m, we obtain that $\mathcal{M}$ is a valid matching. $\square$

Thus, given any potential geodesic path ϕ, there is at least one matching $\mathcal{M}$ compatible with ϕ. Now, supposing that a matching $\mathcal{M}$ is given, we will attempt to determine the shortest path $\phi^*_{\mathcal{M}}$ compatible with $\mathcal{M}$. This will allow us to find the geodesic path by finding the matching $\mathcal{M}$ that minimizes $L(\phi^*_{\mathcal{M}})$. As a first step in this direction, the following lemma provides a simpler characterization for any potential geodesic path ϕ that is compatible with a given matching $\mathcal{M}$.

Lemma 2. *Suppose that $\mathcal{M}$ is a matching of S_x and S_y and that $\phi \in \Phi_{\Sigma_K}(x, y)$ is a geodesic path compatible with $\mathcal{M}$. Define $\lambda : [0, 1] \to \mathbb{R}^K$ as*

$$\lambda_k(t) = |\alpha_{i_k}(t)| - |\beta_{j_k}(t)|, \qquad (15)$$

where $\alpha(t)$ and $\beta(t)$ are defined as in Corollary 1. Then $L(\phi) = L(\lambda)$, and for all k

$$\lambda_k(0) = |x_{i_k}| \qquad (16)$$

and

$$\lambda_k(1) = \begin{cases} |y_{j_k}| & \text{for } k : j_k \in I_{\mathrm{nz}}(\phi) \\ -|y_{j_k}| & \text{for } k : j_k \in I_{\mathrm{z}}(\phi). \end{cases} \qquad (17)$$

Proof. Since ϕ is compatible with $\mathcal{M}$, we have that $t_{i_k} \le r_{j_k}$. For k such that $i_k, j_k \in I_{\mathrm{z}}(\phi)$, using the same reasoning as in Corollary 1 we obtain

$$\lambda_k(t) = \begin{cases} |\alpha_{i_k}(t)| = |\phi_{i_k}(t)| & t \in [0, t_{i_k}) \\ 0 & t \in [t_{i_k}, r_{j_k}] \\ -|\beta_{j_k}(t)| = -|\phi_{j_k}(t)| & t \in (r_{j_k}, 1]. \end{cases} \qquad (18)$$

For k such that $i_k = j_k \in I_{\mathrm{nz}}(\phi)$, we have that

$$\lambda_k(t) = |\alpha_{i_k}(t)| = |\phi_{i_k}(t)|. \qquad (19)$$

Combining (18) and (19) we obtain (16) and (17).

Finally, recall that each term in the summation (1) for $L(\lambda)$ is of the form

$$\sqrt{\sum_{k=1}^{K} \left(\lambda_k(t_m) - \lambda_k(t_{m+1})\right)^2}. \qquad (20)$$

Since we ultimately take a supremum over all possible partitions $\bar{t}$, we are free to require that all partitions $\bar{t}$ contain t_{i_k}, r_{j_k} for all k such that $i_k, j_k \in I_z(\phi)$. This implies that for any k and for any pair $t_m, t_{m+1} \in \bar{t}$, $\mathrm{sgn}(\lambda_k(t_m)) = \mathrm{sgn}(\lambda_k(t_{m+1}))$, $\mathrm{sgn}(\phi_{i_k}(t_m)) = \mathrm{sgn}(\phi_{i_k}(t_{m+1}))$, and $\mathrm{sgn}(\phi_{j_k}(t_m)) = \mathrm{sgn}(\phi_{j_k}(t_{m+1}))$ for all k. Thus, each term in (20) reduces to either

$$\left(\phi_{i_k}(t_m) - \phi_{i_k}(t_{m+1})\right)^2$$

or

$$\left(\phi_{j_k}(t_m) - \phi_{j_k}(t_{m+1})\right)^2.$$

Thus, summing over all k and adding $N - K$ zero terms we obtain

$$\sqrt{\sum_{i=1}^{N} \left(\phi_i(t_m) - \phi_i(t_{m+1})\right)^2}.$$

Summing over all $t_m \in \bar{t}$ we obtain that $L(\lambda) = L(\phi)$. $\square$

Lemma 2 shows that any potential geodesic path ϕ that is compatible with a matching $\mathcal{M}$ can equivalently be thought of as a path λ in $\mathbb{R}^K$ between a pair of points, where the points are almost entirely defined by the choice of $\mathcal{M}$ — the starting point *is* uniquely determined by $\mathcal{M}$, and the ending point is determined up to an unknown sign for k such that $j_k \in I_{\mathrm{nz}}(\phi)$. The power in viewing ϕ in this manner is illustrated in the following theorem.

Theorem 2. *Suppose that $\mathcal{M}$ is a matching of S_x and S_y and let $K_{\mathrm{nz}} = \{k : i_k = j_k$ and $\mathrm{sgn}(x_{i_k}) = \mathrm{sgn}(y_{j_k})\}$ and $K_{\mathrm{z}} = \{1, 2, \ldots, K\} \setminus K_{\mathrm{nz}}$. Suppose that $\mathcal{M}$ satisfies: (i) for any $i_k \in S_x \cap S_y$ such that $i_k = j_m$ with $k \neq m$, $|x_{i_k}| \cdot |y_{j_m}| \leq |x_{i_m}| \cdot |y_{j_k}|$, and (ii) for any $j_k \in S_x \cap S_y$ such that $j_k = i_m$ with $k \neq m$, $|x_{i_m}| \cdot |y_{j_k}| \leq |x_{i_k}| \cdot |y_{j_m}|$. Then the shortest path that is compatible with $\mathcal{M}$, ϕ^*, satisfies*

$$L^2(\phi^*) = \sum_{k \in K_{\mathrm{z}}} \left(|x_{i_k}| + |y_{j_k}|\right)^2 + \sum_{k \in K_{\mathrm{nz}}} \left(|x_{i_k}| - |y_{j_k}|\right)^2. \tag{21}$$

Proof. From Lemma 2, we have that any path that is compatible with the matching $\mathcal{M}$ defines a path λ such that $\lambda_k(0) = |x_{i_k}|$ and $\lambda_k(1) = \pm|y_{j_k}|$. Thus, any path ϕ compatible with $\mathcal{M}$ induces a path λ with starting point independent of the choice of ϕ. The ending point is allowed an unknown sign for $\lambda_k(1)$ when $k \in K_{\mathrm{nz}}$, but for all $k \in K_{\mathrm{z}}$ we again have that $\lambda_k(1)$ is uniquely determined independently of ϕ. We will now show that there exists a ϕ compatible with $\mathcal{M}$ such that the λ induced by this ϕ consists of a straight line which has the shortest length among all admissible choices of signs for $\lambda_k(1)$ with $k \in K_{\mathrm{nz}}$. In this case, there clearly cannot exist any shorter paths that admit the same matching.

Towards this end, we define $q_k = |x_{i_k}| + |y_{j_k}|$ and $p_k = |x_{i_k}|/q_k$. Next, consider the path defined by $\phi^* = \alpha^* + \beta^*$ where

$$\alpha^*_{i_k}(t) = \begin{cases} x_{i_k} - \mathrm{sgn}(x_{i_k}) q_k t & \text{for } t \in [0, p_k) \\ 0 & \text{for } t \in [p_k, 1] \end{cases}$$

for $k \in K_{\mathrm{z}}$,

$$\alpha^*_{i_k}(t) = x_{i_k} - (x_{i_k} - y_{j_k})t \quad \text{for } t \in [0, 1]$$

for $k \in K_{\mathrm{nz}}$, and $\alpha^*_i(t) = 0$ otherwise, and

$$\beta^*_{j_k}(t) = \begin{cases} 0 & \text{for } t \in [0, p_k] \\ y_{j_k} - \mathrm{sgn}(y_{j_k}) q_k (1-t) & \text{for } t \in (p_k, 1] \end{cases}$$

for $k \in K_{\mathrm{z}}$ and $\beta^*_j(t) = 0$ otherwise.

We now need to verify that ϕ^* is compatible with the matching $\mathcal{M}$. For $\phi^* = \alpha^* + \beta^*$, we clearly have that for any $(i_k, j_k) \in \mathcal{M}$ such that $k \in K_{\mathrm{nz}}$, we have that $i_k = j_k$ and $i_k, j_k \in I_{\mathrm{nz}}(\phi)$. Thus, to verify that ϕ^* is compatible with $\mathcal{M}$ we need only check that $t_{i_k} \leq r_{j_k}$ for $k \in K_{\mathrm{z}}$. Towards this end, we first must calculate $\phi^*_i(t)$. There are essentially five cases which must be considered.

Case 1: $i_k = j_k$ and $i_k, j_k \in S_x \cap S_y$. In this case

$$\phi^*_{i_k}(t) = \begin{cases} \alpha^*_{i_k}(t) & \text{for } t \in [0, p_k] \\ \beta^*_{i_k}(t) & \text{for } t \in (p_k, 1] \end{cases}$$

so that $(\phi^*_{i_k})^{-1}(0) = \{p_k\}$. Thus in this case we have that $t_{i_k} = r_{j_k} = p_k$, so that $t_{i_k} \leq r_{j_k}$.

Case 2: $i_k \neq j_k$, $i_k \in S_x \setminus S_y$, $j_k \in S_y \setminus S_x$. In this case we have that $\phi^*_{i_k}(t) = \alpha^*_{i_k}(t)$ and $\phi^*_{j_k}(t) = \beta^*_{j_k}(t)$. Thus, we again have that $t_{i_k} = r_{j_k} = p_k$ and hence $t_{i_k} \leq r_{j_k}$.

Case 3: $i_k \neq j_k$, $i_k \in S_x \cap S_y$, and $j_k \in S_y \setminus S_x$. In this case, since $i_k \in S_x \cap S_y$ but $i_k \neq j_k$, there must exist a j_m such that $m \neq k$ and $i_k = j_m$. From assumption *(i)* above we can show that $p_k \leq p_m$, and thus

$$\phi^*_{i_k}(t) = \begin{cases} \alpha^*_{i_k}(t) & \text{for } t \in [0, p_k] \\ 0 & \text{for } t \in (p_k, p_m) \\ \beta^*_{j_m}(t) & \text{for } t \in [p_m, 1]. \end{cases}$$

Thus, $t_{i_k} = p_k$. Since $j_k \in S_y \setminus S_x$, we again have that $\phi^*_{j_k}(t) = \beta^*_{j_k}(t)$ so that $r_{j_k} = p_k$. Hence, $t_{i_k} \leq r_{j_k}$.

Case 4: $i_k \neq j_k$, $i_k \in S_x \setminus S_y$, and $j_k \in S_x \cap S_y$. This case is symmetric to Case 3. Since $i_k \in S_x \setminus S_y$, we again have that $\phi^*_{i_k}(t) = \alpha^*_{i_k}(t)$ so that $t_{i_k} = p_k$. Since $j_k \in S_x \cap S_y$ but $i_k \neq j_k$, there must exist an i_m such that $m \neq k$ and $i_m = j_k$. From assumption *(ii)* above we can show that $p_m \leq p_k$, and thus

$$\phi^*_{j_k}(t) = \begin{cases} \alpha^*_{i_m}(t) & \text{for } t \in [0, p_m] \\ 0 & \text{for } t \in (p_m, p_k) \\ \beta^*_{j_k}(t) & \text{for } t \in [p_k, 1]. \end{cases}$$

Thus, $r_{j_k} = p_k$. Hence, $t_{i_k} \leq r_{j_k}$.

Case 5: $i_k \neq j_k$ and $i_k, j_k \in S_x \cap S_y$. Here, we have that $i_k = j_m$ with $m \neq k$ and $j_k = i_n$ with $n \neq k$. From assumption *(i)* above we can again show that $p_k \leq p_m$, and thus

$$\phi^*_{i_k}(t) = \begin{cases} \alpha^*_{i_k}(t) & \text{for } t \in [0, p_k] \\ 0 & \text{for } t \in (p_k, p_m) \\ \beta^*_{j_m}(t) & \text{for } t \in [p_m, 1]. \end{cases}$$

Similarly, from assumption *(ii)* above we can show that $p_m \leq p_k$, and thus

$$\phi^*_{j_k}(t) = \begin{cases} \alpha^*_{i_m}(t) & \text{for } t \in [0, p_m] \\ 0 & \text{for } t \in (p_m, p_k) \\ \beta^*_{j_k}(t) & \text{for } t \in [p_k, 1]. \end{cases}$$

Thus, $r_{j_k} = p_k$, and $t_{i_k} \leq r_{j_k}$.

Finally, one can easily check that if we form λ^* using (15) then we obtain

$$\begin{aligned} \lambda^*_k(t) &= |\alpha^*_{i_k}(t)| - |\beta^*_{i_k}(t)| \\ &= \begin{cases} |x_{i_k}| - (|x_{i_k}| - |y_{j_k}|)t & \text{for } k \in K_{\mathrm{nz}} \\ |x_{i_k}| - (|x_{i_k}| + |y_{j_k}|)t & \text{for } k \in K_{\mathrm{z}}. \end{cases} \end{aligned}$$

Thus, λ^* forms a straight line between $\lambda^*(0)$ and $\lambda^*(1)$ with length given by (21). Furthermore, switching the sign on any $\lambda^*_k(1)$ for any $k \in K_{\mathrm{nz}}$ can only increase this length. Thus, ϕ^* is the shortest path compatible with $\mathcal{M}$, as desired. □

Note that the assumptions in Theorem 2 are trivially satisfied for *any* matching when $S_x \cap S_y = \emptyset$. Thus, in this case we have reduced the problem of finding the geodesic path to the problem of identifying the matching between S_x and S_y that minimizes (21). This may not seem like much of an improvement at first, since there are $K!$ possible matchings, so an exhaustive search for the optimal matching will be computationally prohibitive for any moderately large value of K. However, we now demonstrate that such a search is unnecessary, and that the optimal matching can be obtained merely by sorting the nonzeros of x and y appropriately.

Theorem 3. *Suppose that $S_x \cap S_y = \emptyset$. Let $i_1, i_2, \dots, i_K$ denote the ordering of the indices of S_x that satisfies $|x_{i_1}| \leq |x_{i_2}| \leq \cdots \leq |x_{i_K}|$. Similarly, suppose that $j_1, j_2, \dots, j_K$ is the ordering of the indices of S_y that satisfies $|y_{j_1}| \geq |y_{j_2}| \geq \cdots \geq |y_{j_K}|$. This matching will result in the minimal length from (21), and hence the sparse geodesic distance is given by*

$$d^2_{\Sigma_K}(x, y) = \sum_{k=1}^{K} \left(|x_{i_k}| + |y_{j_k}|\right)^2 .$$

Proof. This can be verified by simply considering the impact of switching the order of any pair of indices and observing that it will necessarily increase $L(\phi^*)$. Specifically, recall that

$$\begin{aligned} L^2(\phi^*) &= \sum_{k=1}^{K} (|x_{i_k}| + |y_{j_k}|)^2 \\ &= \sum_{k=1}^{K} |x_{i_k}|^2 + |y_{j_k}|^2 + 2|x_{i_k}| \cdot |y_{j_k}| \\ &= \|x\|_2^2 + \|y\|_2^2 + 2\sum_{k=1}^{K} |x_{i_k}| \cdot |y_{j_k}| \end{aligned}$$

Observe that only the cross-product terms actually depend on our choice of $\mathcal{M}$. Thus, suppose that we switch the order of an arbitrary pair of indices. Without loss of generality, we assume that we switch i_1 with i_2 to obtain $\widetilde{\phi}$. Then cancelling out identical terms we obtain

$$\begin{aligned} L(\widetilde{\phi}) - L(\phi^*) = 2\,(&|x_{i_1}| \cdot |y_{j_2}| + |x_{i_2}| \cdot |y_{j_1}| \\ &-|x_{i_1}| \cdot |y_{j_1}| - |x_{i_2}| \cdot |y_{j_2}|) . \end{aligned} \tag{22}$$

Our goal is to show that if $|x_{i_1}| \leq |x_{i_2}|$ and $|y_{j_1}| \geq |y_{j_2}|$, then $L(\widetilde{\phi}) - L(\phi^*) \geq 0$, meaning that any other matching would necessarily lead to a longer path and hence ϕ^* is in fact a geodesic path. Upon simplification, the inequality that (22) is greater than zero is equivalent to

$$|x_{i_2}|(|y_{j_1}| - |y_{j_2}|) \geq |x_{i_1}|(|y_{j_1}| - |y_{j_2}|).$$

But since $|y_{j_1}| \geq |y_{j_2}|$, this is equivalent to

$$|x_{i_2}| \geq |x_{i_1}|,$$

which is precisely what we were assuming. Thus, the matching which swaps any pair of indices will result in a longer path as desired, and the provided matching has the minimum length. □

One can easily use the same technique to verify the following corollary.

Corollary 2. *Suppose that $|S_x \cap S_y| = \ell$ and let $i_1 = j_1, i_2 = j_2, \dots, i_\ell = j_\ell \in S_x \cap S_y$ be an enumeration of the elements of this set. Let $i_{\ell+1}, i_{\ell+2}, \dots, i_K$ denote the ordering of the indices of $S_x \setminus S_y$ that satisfies $|x_{i_{\ell+1}}| \leq |x_{i_{\ell+2}}| \leq \cdots \leq |x_{i_K}|$. Similarly, suppose that $j_{\ell+1}, j_{\ell+2}, \dots, j_K$ is the ordering of the indices of S_y that satisfies $|y_{j_{\ell+1}}| \geq |y_{j_{\ell+2}}| \geq \cdots \geq |y_{j_K}|$. Let $K_{\mathrm{nz}} = \{k : i_k = j_k$ and $\mathrm{sgn}(x_{i_k}) = \mathrm{sgn}(y_{j_k})$ and $K_{\mathrm{z}} = \{1, 2, \dots, K\} \setminus K_{\mathrm{nz}}$. This matching will result in a path of length*

$$\sqrt{\sum_{k \in K_{\mathrm{z}}} (|x_{i_k}| + |y_{j_k}|)^2 + \sum_{k \in K_{\mathrm{nz}}} (|x_{i_k}| - |y_{j_k}|)^2}$$

and thus we obtain that the sparse geodesic distance is bounded by

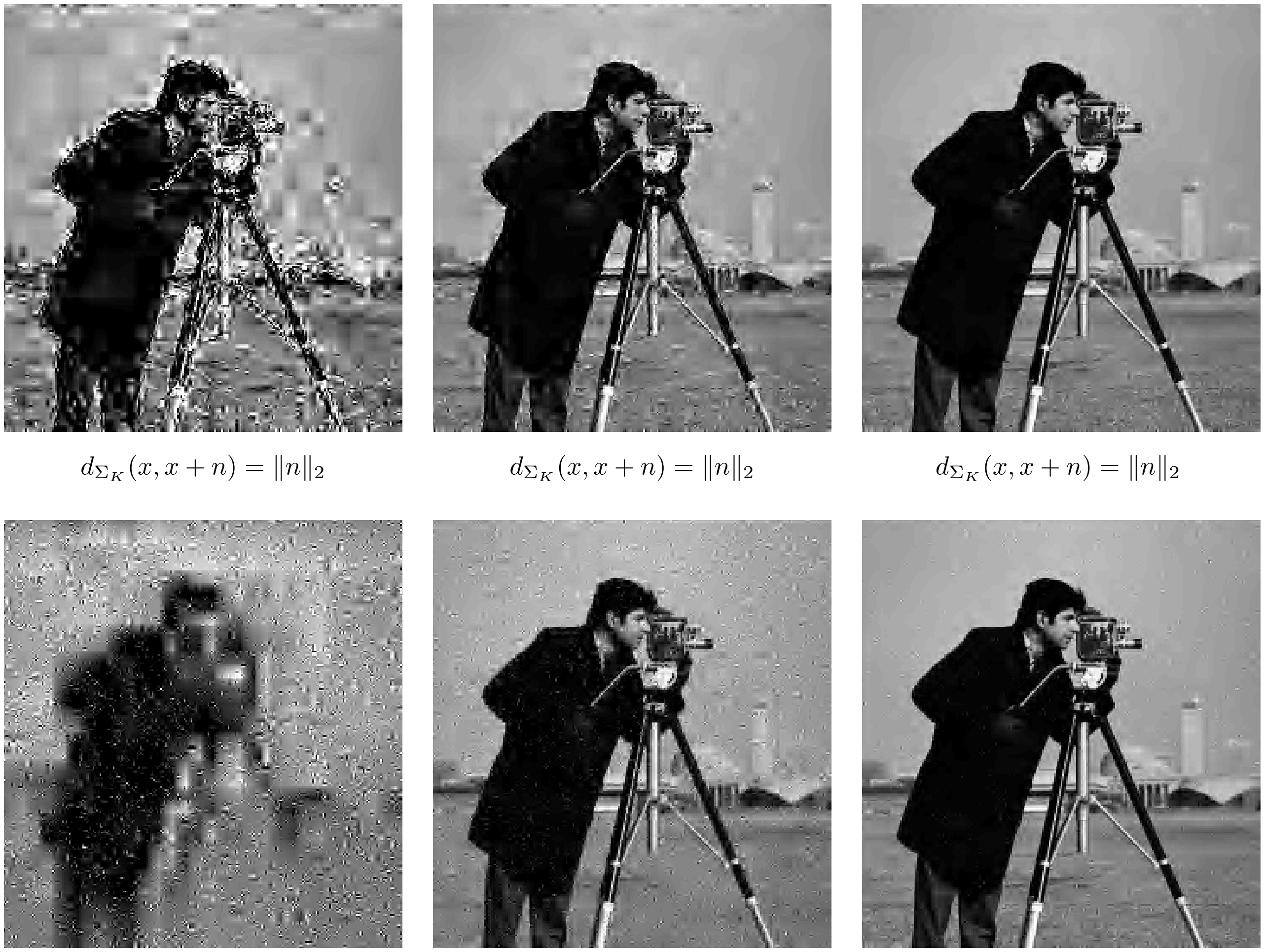
$$d^2_{\Sigma_K}(x, y) \leq \sum_{k \in K_{\mathrm{z}}} (|x_{i_k}| + |y_{j_k}|)^2 + \sum_{k \in K_{\mathrm{nz}}} (|x_{i_k}| - |y_{j_k}|)^2 .$$

3. Example

In order to help interpret the geodesic distance, in Figure 2 we present the results of adding a fixed amount of noise to a sparse image in two different ways. To form a sparse image, we computed the wavelet transform of the "camerman" test image and then zeroed out all but the K largest wavelet coefficients. In Figure 2, the top row of images shows the effect of adding noise only on the support of the nonzero wavelet coefficients. This results in a geodesic distance equal to the ℓ_2 distance. The bottom row of images shows the effect of adding noise that sets some nonzero wavelet coefficients to zero and forces previously zero-valued coefficients to become nonzero. This results in a geodesic distance that is larger than the ℓ_2 distance. Clearly, the visual impact of these two different kinds of noise is very different, and this difference is quantified much better by the sparse geodesic distance than the ℓ_2 distance.

4. Discussion and Extensions

While sparsity is a general model of broad applicability, it is also a fairly weak model in that the values and locations of the nonzero coefficients of a sparse signal are arbitrary. In many applications, sparse signals possess a secondary structure on the nonzero coefficients. For example, wavelet transforms of piecewise smooth signals are near sparse but also tend to live on a connected tree structure in the wavelet domain (Baraniuk 1999; Cohen et al. 2001; Baraniuk et al. 2002). Such *structured sparsity* renders invalid some subspaces in the union of subspaces Σ_K (Blumensath and Davies 2007). Structured sparsity has been exploited in state-of-the-art image compression algorithms (Cohen et al. 2001) and in compressive sensing (Baraniuk et al. 2008; Eldar and Mishali 2009). We hope to extend the sparse geodesic distance to a structured sparse geodesic distance (that follows the shortest path such that each point along the path is a structured sparse signal). In order to do so, it will



$d_{\Sigma_K}(x, x+n) = \|n\|_2$ $\quad$ $d_{\Sigma_K}(x, x+n) = \|n\|_2$ $\quad$ $d_{\Sigma_K}(x, x+n) = \|n\|_2$

$d_{\Sigma_K}(x, x+n) > \|n\|_2$ $\quad$ $d_{\Sigma_K}(x, x+n) > \|n\|_2$ $\quad$ $d_{\Sigma_K}(x, x+n) > \|n\|_2$

(SNR = 10dB) (SNR = 20dB) (SNR = 30dB)

Figure 2: Images (x) corrupted with equal amounts of noise (n) but having different geodesic distances from the true sparse image. The images in the top row have the minimum geodesic distance possible for a given signal-to-noise ratio (SNR), i.e., the geodesic distance $d_{\Sigma_K}(x, x+n)$ is equal to the ℓ_2 distance $\|x - (x+n)\|_2 = \|n\|_2$. The images in the bottom row have greater geodesic distances from the true image as compared to the top row, i.e., $d_{\Sigma_K}(x, x+n) > \|n\|_2$ For a given SNR, it is thus clear that larger geodesic distances correspond to significant visual artifacts.

be necessary to reexamine the Theorem 2, since there are potentially even more matchings $\mathcal{M}$ for which the path ϕ^* will fail to be compatible with $\mathcal{M}$. This will occur since compatibility will now also require that ϕ^* exhibits a kind of structured sparsity — so even if ϕ^* is K-sparse, it might not lie within our set of interest. It seems unlikely that we will obtain an analytical formula for such a geodesic distance, but it may be possible to use dynamic programming techniques to calculate the distance efficiently.

Furthermore, while sparse signals are known to provide a good approximation to many classes of signals, in many cases it is more natural to consider related models such as ℓ_p or weak-ℓ_p balls for $p < 1$. Just as we have done in this work, one could define the geodesic distance for these sets as the shortest path ϕ from x to y such that $\|\phi(t)\|_{\ell_p} \leq R$ or $\|\phi(t)\|_{w\ell_p} \leq R$ for all t. Since for these examples the admissible set is again non-convex (and in some sense can be thought of as a relaxation of Σ_K), it is conceivable that the techniques developed in this work could be useful in analyzing these alternative notions of geodesic distance as well. This could provide a more rigorous framework for measuring geodesic distances between non-sparse signals compared to the obvious approach of thresholding x and y to obtain K-sparse signals and then measuring their geodesic distance.

Finally, as noted in the Introduction, in the case where data lies on a manifold, it has been shown to be extremely beneficial to exploit this structure — particularly in the context of semi-supervised classification (Belkin and Niyogi 2004; Niyogi 2008). In this setting we have a large amount of unlabeled training data and a small amount of labeled data from which we wish to learn a classifier. A potentially powerful method for tackling this problem is to use the unlabeled data to learn the geometric structure of the data set so that we are able to obtain a more meaningful measure of the distance between the labeled training data. In settings where we have prior knowledge or reason to believe that our data should be sparse with respect to a potentially completely unknown dictionary, we could use the unlabeled data to *learn* a good dictionary for the purpose of sparse representation using an algorithm such as the KSVD (Aharon, Elad, and Bruckstein 2006). Using this learned dictionary we could then calculate sparse geodesic distances between the labeled training data as input to a nearest-neighbor or SVM-type classifier. We plan to explore these and other ideas more extensively in a sequel.

Acknowledgements

This work was supported by grants NSF CCF-0431150 and CCF-0728867, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936 and N00014-08-1-1112, AFOSR FA9550-07-1-0301, ARO MURI W311NF-07-1-0185, and the TI Leadership Program.

References

Aharon, M.; Elad, M.; and Bruckstein, A. 2006. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing* 54(11):4311–4322.

Baraniuk, R.; DeVore, R.; Kyriazis, G.; and Yu, X. 2002. Near best tree approximation. *Advances in Comp. Math.* 16(4):357–373.

Baraniuk, R.; Cevher, V.; Duarte, M.; and Hegde, C. 2008. Model-based compressive sensing. Preprint.

Baraniuk, R. 1999. Optimal tree approximation with wavelets. In *Proc. SPIE Wavelet Applications in Signal and Image Processing VII*, volume 3813, 196–207.

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

Belkin, M., and Niyogi, P. 2004. Semi-supervised learning on Riemannian manifolds. *Machine Learning Journal* 56:209–239.

Blumensath, T., and Davies, M. E. 2007. Sampling theorems for signals from the union of linear subspaces. Preprint.

Candès, E. 2006. Compressive sampling. In *Proc. Int. Congress of Mathematics*, volume 3, 1433–1452.

Cohen, A.; Dahmen, W.; Daubechies, I.; and DeVore, R. 2001. Tree approximation and optimal encoding. *Appl. Comp. Harm. Anal.* 11(2):192–226.

Coifman, R. R., and Maggioni, M. 2006. Diffusion wavelets. *Appl. Comput. Harmon. Anal.* 21(1):53 – 94.

DeVore, R. 1998. Nonlinear approximation. *Acta Numerica* 7:51–150.

Donoho, D., and Grimes, C. 2003. Hessian Eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* 100(10):55915596.

Donoho, D., and Grimes, C. 2005. Image manifolds isometric to Euclidean space. *J. Math. Imaging and Computer Vision* 23(1):5–24.

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

Donoho, D. 2006. Compressed sensing. *IEEE Trans. Inform. Theory* 6(4):1289 – 1306.

Eldar, Y. C., and Mishali, M. 2009. Robust recovery of signals from a structured union of subspaces. To appear in IEEE Trans. Inform. Theory.

J. Tenenbaum, V. d. S., and Langford, J. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):23192323.

Mallat, S. 1999. *A wavelet tour of signal processing*. Academic Press.

Niyogi, P. 2008. Manifold regularization and semi-supervised learning: Some theoretical analyses. Technical Report TR-2008-01, Computer Science Dept., University of Chicago.

Olshausen, B., and Field, D. 1996. Emergence of simple-cell receptive field properties by learning a sparse representation. *Nature* 381:607–609.

Roweis, S., and Saul, L. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):23192323.

Weinberger, K., and Saul, L. 2006. Unsupervised learning of image manifolds by semidefinite programming. *Int. J. Computer Vision Special Issue: Computer Vision and Pattern Recognition* 70(1):7790.