FRAME BASED KERNEL METHODS FOR AUTOMATIC CLASSIFICATION IN HYPERSPECTRAL DATA

John J. Benedetto, Wojciech Czaja, Justin C. Flake, Matthew Hirn

Department of Mathematics
University of Maryland
College Park, MD 20742
jjb@math.umd.edu wojtek@math.umd.edu jcflake@math.umd.edu hirn@math.umd.edu

ABSTRACT

We propose a new kernel and frame based dimension reducing algorithm by exploiting the synergy between endmembers and kernel based classification schemes. This synergy becomes available by means of utilizing the notions of frames and frame representations.

1. INTRODUCTION

In many applications of machine learning, data mining, and image processing, high dimensional data is collected. Although the collected data is high dimensional, often it is the case that the data is intrinsically low dimensional. We think of this data as lying in a subspace or a manifold embedded in the larger space. Dimensionality reduction is the transformation of this data from a high dimensional space to a low dimensional space. Hopefully, the transformation creates a representation of the data that is low dimensional yet still preserves certain properties of the original data, e.g., [1], [2]. The goal of dimensionality reduction is to nd the minimum number of parameters (dimensions) necessary to account for the data. Dimensionality reduction techniques include: principal component analysis, locally linear embedding, neural networks, and genetic algorithms.

The novelty of our approach is based on the introduction of frame representations into the above DR scheme, instead of (orthonormal) bases. Our frames are data-dependent and are based on endmember demixing schemes like, e.g., [3], [4]. These frame representations yield multiple frame coefficient maps, which are then optimized with respect to the sparcity of the coefficients, and other measures. These concepts are explained in more detail below.

The authors gratefully acknowledge the support of the Department of Defence through contracts HM 1582-07-C-0008 and HM 1582-08-1-0009. Published in the IEEE 2009 International Geoscience & Remote Sensing Symposium (IGARSS 2009), scheduled for July 13-17, 2009 in Cape Town, South Africa.

2. ALGORITHM DESCRIPTION

Given a hyperspectral data set $X = \{x_i\}_{i=1}^N \subseteq \mathbb{R}^D$ consisting of N pixels in D dimensions, we propose the following algorithm for processing X: *1.) Landmarking, 2.) Kernel Application, 3.) Out of sample extension, 4.) Endmember selection, 5.) Frame coefficients.* Steps 1 and 3 enable the algorithm to run on large data sets. In step 2, we use kernel eigenmap methods to reduce the dimension of the data set X, thus creating a low dimensional data set $Y = \{y_i\}_{i=1}^N \subseteq \mathbb{R}^d$ that preserves the local geometry of X. Y consists of N d-dimensional data points, one for each element of X. We assume $d < D$. Step 4 selects endmembers for the lower dimensional data set Y. Unlike traditional endmember applications in which the the number of endmembers is fewer than the dimension of the data, we select more endmembers than the reduced dimension d. This creates a frame [5], Φ, for $\mathbb{R}^d$ by which we can represent the low dimensional data points Y. Frames provide overcomplete representations which gives flexibility in representing mixtures and pure elements. Step 5 computes the frame coefficients of the data points Y in terms of the endmembers Φ. There are infinitely many such frame representations - we highlight certain ones that are well suited for classification purposes.

2.1. Landmarking

Our first step is to reduce the complexity of the kernel eigenmap algorithm by selecting a subset of X on which to compute the kernel. We denote this subset as $Z = \{z_i\}_{i=1}^n \subseteq X$, where $n \ll N$. Our current results select the set Z uniformly at random from the set X. In the future we plan to investigate more systematic ways by which to sample X.

2.2. Kernel Application

Given $Z \subseteq X$, we construct a kernel for Z. Our results thus far focus on the locally linear embedding (LLE) kernel [2]. The general nature of our framework, though, al-

lows for the use of any kernel eigenmap method, including, e.g., Laplacian eigenmaps [1] or Isomap [6]. We diagonalize the resulting kernel K and select the d eigenvectors corresponding to the d smallest non-zero eigenvalues. Denote the j^{th} smallest non-zero eigenvector by v_j, and let the i^{th} entry of v_j be denoted by $v_j(i)$. The reduced dimension coordinates for the sampled points $z_i \in Z$ are then given by $y_i = (v_1(i), v_2(i), \ldots, v_d(i)) \in \mathbb{R}^d$, for all $i = 1, \ldots, n$.

2.3. Out of Sample Extension

Given the n low dimensional coordinates $\{y_i\}_{i=1}^n$ corresponding to the sampled set $Z = \{z_i\}_{i=1}^n \subseteq X$, we wish to extend these new coordinates to all of X via an out of sample extension [7]. To do so we extend the definition of k-nearest neighbors to include reference points $x \in X$ that are out of sample:

$$\mathcal{N}_k(x) = \{\tilde{z} \in Z \mid \tilde{z} \text{ is one of the } k \text{ closest points to } x\}.$$

Notice that while the reference point may now come from X, the neighbors are still selected only from the sampled points Z. In the case of LLE, if $x \in X$ is out of sample, then we define the new low dimensional coordinates of x as:

$$y_x = \sum_{i=1}^{n} y_i w(x, z_i),$$

where $w(x, z_i)$ is the weight of z_i in the LLE reconstruction of x using the k-nearest neighbors selected only from the sampled values Z. After a suitable re-indexing of the low dimensional coordinates, we are left with a set $Y = \{y_i\}_{i=1}^N \subseteq \mathbb{R}^d$, where y_i is the new low dimensional representation of the original high dimensional data point $x_i \in X \subseteq \mathbb{R}^D$.

2.4. Endmember Selection

The fourth step in our algorithm is to select endmembers for the low dimensional space $Y \subseteq \mathbb{R}^d$. Traditional applications of endmember algorithms are run on the original high dimensional data set $X \subseteq \mathbb{R}^D$, and if s denotes the number of endmembers, then $s < D$. Since we are finding endmembers for the space Y, we propose finding $s > d$ endmembers, thus creating a frame $\Phi = \{\varphi_i\}_{i=1}^s$ for Y. Frames arise naturally in dimension in reduction, and are in fact a generalization of orthonormal bases. There are many endmember selection algorithms available, e.g., N-FINDR [8], ORASIS [9], and Pixel Purity Index [10]; see also [11] and [4]. The results of this paper employ the Support Vector Data Description (SVDD), see, e.g., [3] algorithm for selecting endmembers. The core idea of SVDD is to obtain a minimal spherical shaped boundary around the data set, which in turn gives a description of the data in terms of a set of support vectors.

2.5. Frame Coefficients

Given a frame $\Phi = \{\varphi_i\}_{i=1}^s$ for Y, we shall find a set of coefficients $C = \{c_{i,j}\}_{i,j=1}^{N,s}$ that represents Y in terms of Φ:

$$y_i = \sum_{j=1}^{s} c_{i,j}\varphi_j \quad \text{for all} \quad i = 1, \ldots, N.$$

We propose two separate ways to find C. The first is based on the frame operator $S : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, which is:

$$Sy = \sum_{i=1}^{s} \langle y, \varphi_i \rangle \varphi_i \quad \text{for all} \quad y \in \mathbb{R}^d.$$

For any frame Φ, the frame operator S is invertible, and in fact gives the following representation:

$$y = \sum_{i=1}^{s} \langle y, S^{-1}\varphi_i \rangle \varphi_i \quad \text{for all} \quad y \in \mathbb{R}^d.$$

Thus, we can define the coefficient set $C = \{c_{i,j}\}_{i,j=1}^{N,s}$ as:

$$c_{i,j} = \langle y_i, S^{-1}\varphi_j \rangle \quad \text{for all} \quad i = 1, \ldots, N, \;\; j = 1, \ldots, s.$$

The coefficients $\{c_{i,j}\}$ are called the *canonical coefficients* and they minimize the ℓ^2 energy of the coefficient set C. An alternative to the canonical coefficient set is to find sparse coefficient representations. Such coefficients are found by minimizing the ℓ^p energy of the coefficients, where $0 < p \leq 1$:

$$c_{i,\cdot} = \arg\min_{\tilde{c}} \|\tilde{c}\|_{\ell^p} \quad \text{subject to} \quad y_i = \sum_{j=1}^{s} \tilde{c}_j \varphi_j.$$

3. RESULTS

The results of this section are based on the Smith Island PROBE2 HSI data published by C. M. Bachmann et al. in [12], [13], [14], [15].

The Smith P20011018 data set consists of 640976 pixels, broken into 22 distinct classes. The results of the Smith trial were obtained with the following settings:

- Kernel: LLE
- Number of neighbors: 50
- Number of samples: 40000 pixels
- Number of reduced dimensions: 21
- Endmember algorithm: SVDD
- Number of endmembers: 69
- Type of coefficients: canonical
- Classifier: vector angle

Sample coefficient maps and a pseudo-color image of Smith Island are given in figure 3.

Classification results are summarized in table 1. Here # denotes the number of samples, #c denotes the number of correctly identified samples, #fp denotes the number of false positives, and #fn denotes the number of false negatives.

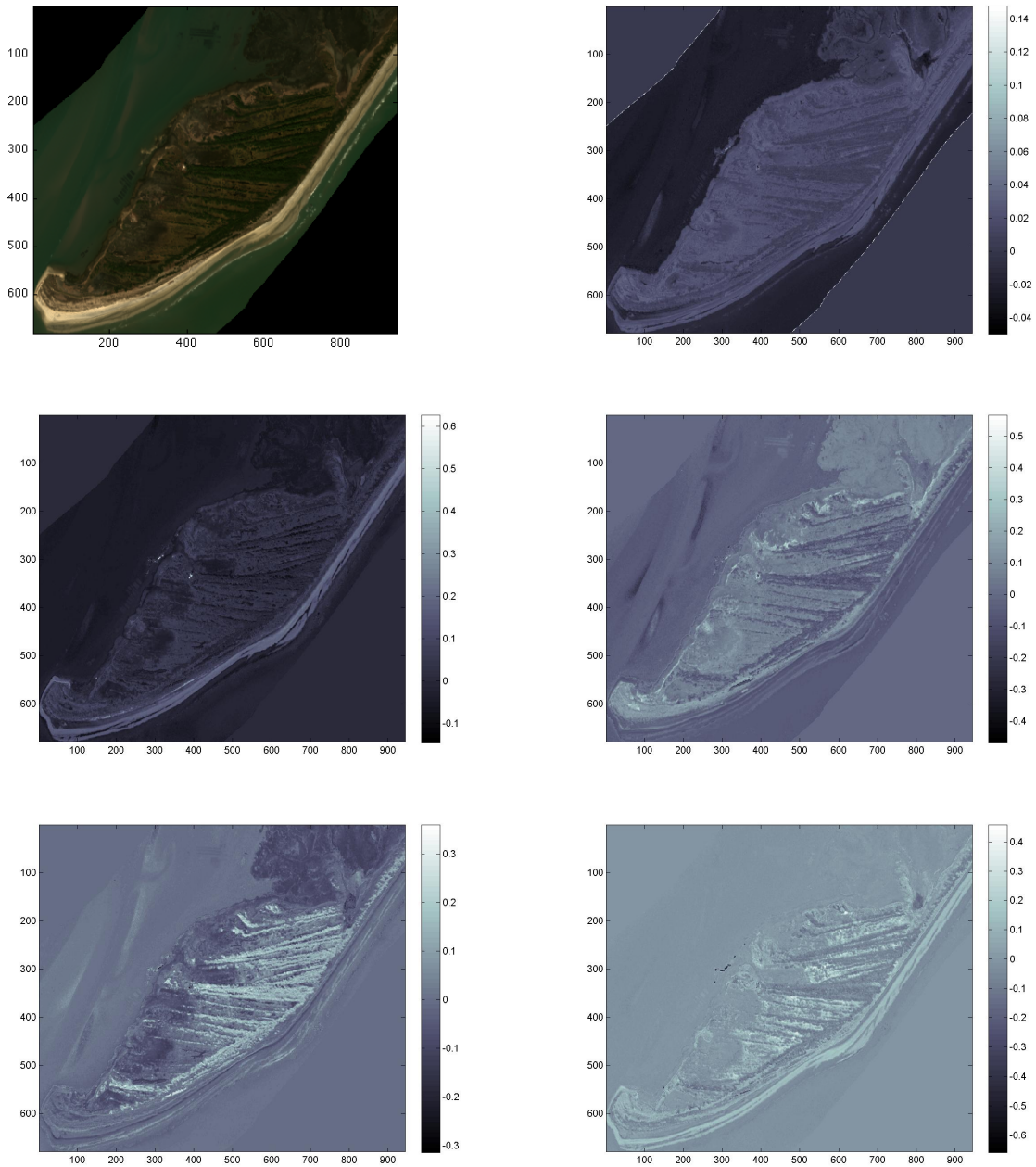

Fig. 1. Frame coefficient maps of Smith Island

Table 1. Smith trial ground truth results

	#	#c	%	#fp	#fn
phrag4	196	138	70%	68	58
scirpus	246	155	63%	55	91
juncus	184	116	63%	33	68
patens	66	57	86%	33	9
distichlis	97	90	93%	18	7
andropogon	57	38	67%	9	19
ammophila	32	25	78%	29	7
mud	70	63	90%	25	7
alterniora	200	182	91%	60	18
borrichia	90	84	93%	24	6
salicornia	76	58	76%	3	18
iva	58	49	84%	51	9
pine	166	134	81%	59	32
hardwood_mixed	328	193	59%	41	135
pond_water	105	69	66%	3	36
sand	159	157	99%	0	2
wrack	144	97	67%	11	47
myrica	167	132	79%	54	35
seaoats	18	13	72%	0	5
typha	44	18	41%	59	26
Water_Nshore	206	206	100%	0	0
SubmergedNets1	34	34	100%	0	0
Total	2743	2108	77%	635	635

Table 2. Smith competing ground truth results (total only)

	#	#c	%	#fp	#fn
Raw data	2743	1957	71%	786	786
SVDD	2743	2088	76%	655	655
LLE ($d = 19, \ldots, 25$)	2743	2059	75%	684	684

4. REFERENCES

[1] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, pp. 585–591, 2002.

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

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

[4] D. Gillis, J. Bowles, and M. Winter, "Using endmembers as a coordinate system in hyperspectral imagery," *Proceedings of SPIE*, vol. 4816, pp. 346–354, 2002.

[5] J. Kovacevic and A. Chebira, "Live beyond bases: The advent of frames," *IEEE Signal Processing Magazine*, 2007.

[6] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering," *Advances in Neural Information Processing Systems*, vol. 16, pp. 177–184, 2004.

[8] M. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE*, vol. 3753, pp. 266–275, 1999.

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

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

[11] C. Wu and C.-I Chang, "A new simplex growing algorithm for endmember extraction," *IEEE International Geoscience and Remote Sensing Symposium*, July 2005.

[12] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large scale hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. vol. 44 (10), pp. 2786–2803, 2006.

[13] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. vol. 43 (3), pp. 441–454, 2005.

[14] C. M. Bachmann, "Improving the performance of classifiers in high-dimensional remote sensing applications: An adaptive resampling strategy for error-prone exemplars (aresepe)," *IEEE Trans. on Geoscience and Remote Sensing*, vol. vol. 41 (9), pp. 2101–2112, 2003.

[15] C. M. Bachmann, T. F. Donato, G. M. Lamela, W. J. Rhea, M. H. Bettenhausen, R. A. Fusina, K. Du Bois, J. H. Porter, , and B. R. Truitt, "Automatic classification of land-cover on smith island, va using hymap imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. vol. 40 (10), pp. 2313–2330, 2002.