Machine Learning for Seismic Signal Processing: Seismic Phase Classification on a Manifold

Juan Ramirez Jr.
Department of Electrical, Computer,
and Energy Engineering
University of Colorado
Boulder, Colorado 80309
Email: juan.ramirez@colorado.edu

François G. Meyer
Department of Electrical, Computer,
and Energy Engineering
University of Colorado
Boulder, Colorado 80309
Email: fmeyer@colorado.edu

Abstract—In this research, we consider the supervised learning problem of seismic phase classification. In seismology, knowledge of the seismic activity arrival time and phase leads to epicenter localization and surface velocity estimates useful in developing seismic early warning systems and detecting man-made seismic events. Formally, the activity arrival time refers to the moment at which a seismic wave is first detected and the seismic phase classifies the physics of the wave.

We propose a new perspective for the classification of seismic phases in three-channel seismic data collected within a network of regional recording stations. Our method extends current techniques and incorporates concepts from machine learning. Machine learning techniques attempt to leverage the concept of "learning" the patterns associated with different types of data characteristics. In this case, the data characteristics are the seismic phases. This concept makes sense because the characteristics of the phase types are dictated by the physics of wave propagation. Thus by "learning" a signature for each type of phase, we can apply classification algorithms to identify the phase of incoming data from a database of known phases observed over the recording network.

Our method first uses a multi-scale feature extraction technique for clustering seismic data on low-dimensional manifolds. We then apply kernel ridge regression on each feature manifold for phase classification. In addition, we have designed an information theoretic measure used to merge regression scores across the multi-scale feature manifolds. Our approach complements current methods in seismic phase classification and brings to light machine learning techniques not yet fully examined in the context of seismology.

We have applied our technique to a seismic data set from the Idaho, Montana, Wyoming, and Utah regions collected during 2005 and 2006. This data set contained compression wave and surface wave seismic phases. Through cross-validation, our method achieves a 74.6% average correct classification rate when compared to analyst classifications.

I. INTRODUCTION

A. Problem Statement

Seismic recording stations use sensors to collect non-stationary data describing motion observed at the station. Typically, three-channels of data are collected to represent the motion in the plane horizontal to the earth and the perpendicular direction (i.e. X, Y, and Z) at the sensing station. These signals consist of short bursts of energy that dissipate to ambient levels. The short bursts of energy result from seismic waves arriving at the station emanating from nearby sources. The analysis of these seismic recordings is at the core of technology used to mediate the effects of earthquakes on the human population. Two aspects critical to these technologies is the estimation of the seismic activity arrival time and the phase of a seismic wave. The arrival time is the moment where seismic activity is observed. On the other hand, a seismic phase is a label given to each instance of seismic activity that characterizes the wave as either (1) a compression wave (e.g. P-wave), shear wave (e.g. S-wave), or surface wave (e.g. L-wave / R-wave) and (2) any reflections through the earths inner structure along the waves path from the epicenter to the observation point.

This paper address the supervised learning of seismic phases from three-channel seismic recordings observed across a regional sensing network. We assume that seismic activity arrival times have been identified and are known to the algorithm. As a result of our data set, we demonstrate our algorithm on two classes of phases, the compressional wave -P and surface wave -L.

B. Related Work

Several methods exist to extract seismic event arrival time and phases from single station three-channel recordings. The most prominent method for arrival time estimation is the current-value-to-predicted-value [14]. The current-value is the short term average (STA) of the incoming signal and the predicted-value is the long term average (LTA). The ratio is expressed by STA/LTA. As data is collected, both STA and LTA are updated and a detection is declared when the ratio exceeds a predetermined threshold. A new method for arrival time estimation was presented by Taylor *et al.* (2010), where the seismic data is lifted into a high-dimensional space and graph-based methods are used to assess the presence of arrivals in the seismic data [12]. In essence, these methods attempt to locate the point within the seismic signal where the signal transitions from ambient levels to activity levels.

A common technique for phase classification is to examine the polarization of the three-channel data [5]. For example compression waves are typically linearly polarized signals

while surface waves are elliptically polarized. As a result of this distinguishing feature, most methods for the classification of seismic phase typically incorporate polarization estimates. These methods use three-channel seismic recordings to derive polarization measures from the eigen-spectra of the cross energy matrix. In essence, polarization analysis seeks to find the direction of maximum signal energy. Jackson *et al.* (1991) proposed to use the seismic trace data itself to estimate wave polarization [4]. Park *et al.* (1987), explored the concept of frequency dependent polarization by using the discrete Fourier transform of the seismic data to derive polarization measures. In their work, the authors found that P-phases exhibit different spectral characteristics in different frequency bands [8]. Anant and Dowla (1997), applied the discrete wavelet transform to the seismic data, resulting in a multi-scalar representation of the signal. They subsequently designed phase/arrival time locator functions for P- and S-phases in the time-frequency domain [1]. Similarly, Tilbuleac *et al.* (2003), showed an effective method for the detection of L- type waves using the wavelet transform [13]. These methods share the common thread of estimating polarization parameters in order to resolve the phase of the signal. Zhizhin *et al.* (2006), takes a different approach to source parameter estimation. Here, the authors take a supervised learning approach for source localization from three-channel seismic data. By designing features that allow data to cluster according to source location, they were able to estimate the location of new seismic sources occurring at a given recording station [15].

In most cases, the arrival time and phase are jointly estimated from the seismic signal. We propose to decouple the estimation of arrival time and phase by first applying an arrival time estimator and then a phase classifier. We also continue the use of machine learning techniques by merging the use of signal decomposition methods, polarization estimation, and supervised learning techniques to learn the phase of $K \times 3$ segments of seismic ground motion data.

II. Algorithm Overview

Our algorithm can be divided into two major stages (see Figure 1): (1) feature extraction and (2) supervised phase learning. During feature extraction, we seek to derive a multi-scale low-dimensional representation of each $K \times 3$ seismic ground motion data matrix. Each column of these matrices corresponds to a different sensing channel. Jackson *et al.* (1991), outlined a strategy to select $K \times 3$ windows of data from a three-channel seismic signal. The principles laid out are the following,

1) Each data window should only contain one activity arrival.
2) The window should attain a maximal signal-to-noise ratio.
3) The window length is sufficiently long to allow for the discrimination between ambient and activity signal levels.

These principles guide our window selection process. For each $K \times 3$ data matrix, we use an n-level stationary wavelet transform (where $n < \log_2(K)$) on each column of our data matrix to obtain a multi-scale representation. Using this multi-scale representation, we perform a singular value decomposition on the corresponding data matrices at each decomposition scale. By selecting the right principle singular vector (i.e. the right singular vector corresponding to the largest singular value) on each decomposition scale, we produce a multi-scale feature space where on each scale all $K \times 3$ data matrices are represented. The right principle singular vectors serves as our polarization signature. During this stage, we have transformed a single $K \times 3$ data matrix into a $n + 1$ set of vectors in $\mathbf{R}^3$, where each vector is embedded in three-dimensional multi-scale spherical manifold (see Figure 3).

In the supervised phase learning stage of our algorithm, we seek to use a database of labeled phases to classify seismic data for which the phase is unknown. We accomplish this by using kernel ridge regression on each spherical feature manifold where both the testing and training data reside. By merging the classification results from our multi-scale feature manifolds, we are able to produce a classification result for each unlabeled $K \times 3$ data matrix. Figure 1 is a block diagram of our algorithm.
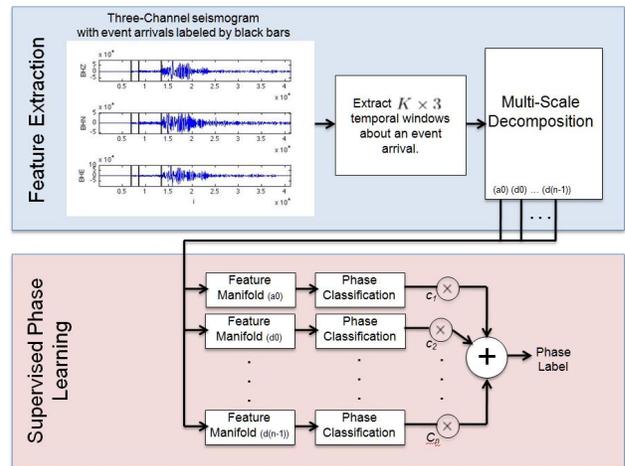

Fig. 1. Algorithm Flow

III. Theoretical Background

In this section, we present the two stages of our algorithm, feature extraction and supervised phase learning. Our goal is to extract signal features that allow for the discrimination between P-phases (rectilinearly polarized waves) and L-phases (elliptically polarized waves). Typical methods for seismic phase classification include estimating parameters to detect wave polarization states from three-channel data. This has been done by examining the singular value decomposition (SVD) of either the cross-energy matrix of the data, the data itself, or variations that include spectral decompositions of the data [[4], [8], [2]]. These parameters include the singular values and elements from the right principle singular vector obtained after SVD. In each case, these elements are used to

build models for phase classification. In our method, we use a stationary wavelet transform to estimate a wave polarization signature as a function of wavelet decomposition scale. We then build phase classification models on each decomposition scale by analyzing clusters of polarization signature vectors within the scale.

A. Multi-scale Polarization Signature Feature Extraction

The goal of multi-scale analysis is to allow for the examination of signals at different frequency ranges or scales. Kumar and Foufola-Georgiu (1997), provide a detailed discussion of the use of the wavelet transform for geophysical applications [6]. However, in a similar manner proposed by Saragiotis *et al.* (1999), we use a stationary wavelet decomposition of the seismic signal to identify changes in the waveform in different frequency bands [10]. In addition, we use a symmetric wavelet to avoid introducing phase shift in our analysis [7]. The stationary wavelet decomposition is a redundant transform: we obtain $K \log_2 K$ coefficients, where K is the size of the original signal. Fortunately, there exists a fast algorithm to compute the stationary wavelet decomposition: the "à trou" algorithm [11]. Figure 2 shows a seismic signal, its corresponding spectogram, and stationary wavelet transform coefficients.
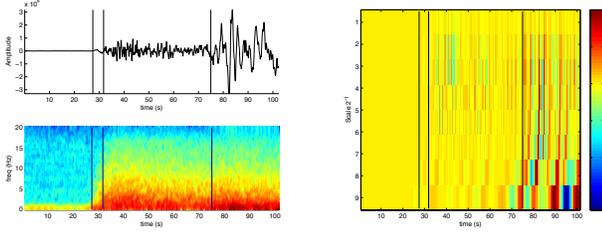

Fig. 2. Left: Z channel of a seismic signal (top) and spectrogram (bottom). Three seismic arrivals are marked by vertical bars. Note that the second and third arrivals hardly elicit any changes in the spectrogram. Right: stationary wavelet transform of the waveform shown in left. The magnitude is color coded and displayed as a function of time, from fine scale (top) to coarse scale (bottom).

Let Φ be a $K \times 3$ seismic data matrix, as described above. Applying an n-level stationary wavelet transform to each column of the data matrix Φ, we obtain the multi-scale representation,

$$\{\Phi^{a_0}, \Phi^{d_0}, \dots, \Phi^{d_{n-1}}\} \qquad (1)$$

where the superscripts $a_0, d_0, \cdots, d_{n-1}$ represent the different levels of wavelet decomposition applied to the seismic data matrix Φ. Let $j \in \{a_0, d_0, \cdots, d_{n-1}\}$ be an element of the set of decomposition indices. Each Φ^j is a matrix where each column is the j-th level decomposition in the stationary wavelet transform of the corresponding column of Φ.

The singular value decomposition of Φ^j is given by,

$$\Phi^j = U \Sigma V^T \qquad (2)$$

where U is an orthonormal basis of the row space of Φ^j, V is an orthonormal basis of the column space of Φ^j, and Σ is a diagonal matrix of the singular values. Let Γ^j be the right principle singular vector of Φ^j, which is known as the polarization vector in the seismic literature. The seismic data matrix Φ is thus represented on multiple scales of decomposition by the polarization vector $\Gamma^j \in \mathbf{R}^3$, where $j \in \{a_0, d_0, \cdots, d_{n-1}\}$. Each vector corresponds to a point on the sphere S^2, since $||\Gamma^j||_2 = 1$.

B. Spherical Feature Manifold

Applying the multi-scale polarization signature feature extraction method to each $K \times 3$ matrix Φ collected across a given seismic recording, we obtain spherical feature manifolds across multiple frequency scales. Figure 3 is an example of feature manifolds and wavelet decomposition coefficients across multiple scales for all data matrices Φ. This data was collected from 10 events recorded at various regional recording stations within the Rocky Mountain Seismic Sensing Network.
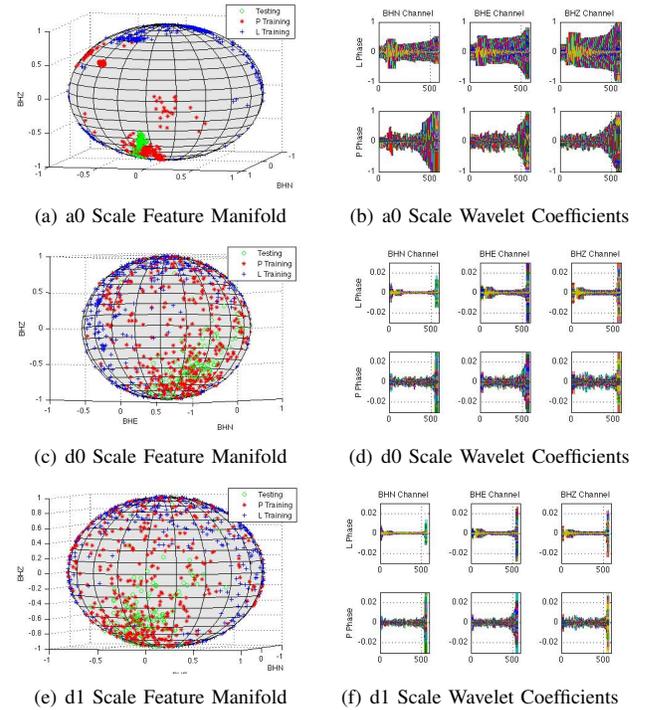

(a) a0 Scale Feature Manifold

(b) a0 Scale Wavelet Coefficients

(c) d0 Scale Feature Manifold

(d) d0 Scale Wavelet Coefficients

(e) d1 Scale Feature Manifold

(f) d1 Scale Wavelet Coefficients

Fig. 3. Feature manifolds and corresponding wavelet coefficient traces. Each $K \times 3$ seismic data matrix Φ is represented by a single point in the feature manifold across each wavelet decomposition scale. On each sphere, one realization of testing set and training set for classification is presented. We observe that across each decomposition scale, the unlabeled test data clusters near its true label class.

By embedding each data matrix Φ into multi-scale low-dimensional manifolds, we observe clustering among phase classes (see Figure 3-(a,c,e)). Characterizing each data matrix in these spaces allows for phase classification. We refer each point on the feature manifold as the polarization signature for a given data matrix Φ. The vectors Γ^j represent the direction of maximal energy in each decomposition scale.

C. Supervised Phase Learning

Using the multi-scale feature manifold representation of seismic data observed in the sensing network, we are able to apply a regression-based supervised learning algorithm for phase classification. In particular, we apply at each scale a kernel ridge regression on the non-linear feature manifold for classification. We first explore phase classification on a single feature manifold and then develop an information theoretic measure to merge the single scale classification results into a final phase classification result.

1) Single Scale Classification: In the feature manifold setting, we are interested in learning a function from binary phase data. The regression function $f_j : S_j^2 \longrightarrow [-1, 1]$ is a map from the feature manifold to a subset of real numbers. This function is derived from phase training data and evaluated on seismic data samples with an unknown phase. The function f_j, evaluated at a test point Γ^j, is given by,

$$f_j(\Gamma^j) = \beta_0^j + \sum_{i=1}^{N} \beta_i^j \kappa(\Gamma^j, \Gamma_i^j) \tag{3}$$

where $\kappa(\cdot, \cdot)$ is the kernel function, $\{\beta_1, \beta_2, \cdots, \beta_N\}$ are the model coefficients, and the index i runs over the number of training points, Γ_i^j. The kernel function is given by,

$$\kappa(\Gamma^j, \Gamma_i^j) = \exp\left(-\frac{d(\Gamma^j, \Gamma_i^j)}{\sigma(\pi - d(\Gamma^j, \Gamma_i^j))}\right) \tag{4}$$

where $d(\Gamma^j, \Gamma_i^j)$ is the geodesic distance between points Γ^j and Γ_i^j is on the sphere S^2. On S^2, the geodesic distance d is given by,

$$d(\Gamma^j, \Gamma_i^j) = \arccos((\Gamma^j)^T \Gamma_i^j) \tag{5}$$

Measuring the distance between two points on the sphere is the critical factor in implementing kernel ridge regression on the nonlinear feature manifold. In addition, this choice of kernel allows us to weight the importance of each training point on the sphere with respect to the test point. For training points where $d(\Gamma^j, \Gamma_i^j)$ is small, the argument of the kernel function, $\frac{d(\Gamma^j, \Gamma_i^j)}{(\pi - d(\Gamma^j, \Gamma_i^j))}$ will be near 0, resulting in the maximum kernel value of 1. As the distance between any two points approaches the maximum spherical distance of π radians, the kernel argument tends towards infinity and the kernel value approaches 0. In this way, the nonlinearity in the kernel argument serves to ensure the distance on the sphere is normalized appropriately.

In order to derive the function f_j, the model coefficients β_i^j must be determined from the training data. Let the training set on the j-th decomposition scale be given by,

$$\{y_1^j, \Gamma_1^j, y_2^j, \Gamma_2^j, \cdots, y_N^j, \Gamma_N^j\} \tag{6}$$

where y_i^j is a numeric phase class label: +1 for P-phase data or -1 for L-phase data and Γ_i^j are the corresponding polarization signature vectors. The model coefficients $\{\beta_i^j\}_{i=0}^{N}$ are found by solving the optimization problem,

$$\min_{\beta^j} \|y - \beta^j X^j\|_2^2 + \lambda \|\beta^j\|_2^2 \tag{7}$$

where $y = [y_1, y_2, \cdots, y_N]^t$, $\beta^j = [\beta_1^j, \beta_2^j, \cdots, \beta_N^j]$, and X is given by

$$X^j = \begin{pmatrix} \kappa(\Gamma_1^j, \Gamma_1^j) - \varphi_1^j & \cdots & \kappa(\Gamma_N^j, \Gamma_1^j) - \varphi_N^j \\ \vdots & \ddots & \vdots \\ \kappa(\Gamma_1^j, \Gamma_N^j) - \varphi_1^j & \cdots & \kappa(\Gamma_N^j, \Gamma_N^j) - \varphi_N^j \end{pmatrix} \tag{8}$$

where $\varphi_m^j = \frac{1}{N} \sum_{i=1}^{N} \kappa(\Gamma_m^j, \Gamma_i^j)$. The unique solution to the above optimization problem is given by $\beta^{j*} = (X^T X - \lambda I)^{-1} X^T y$ [3]. Using this procedure for each wavelet decomposition scale, we learn a function from the distributions of training data and assign a classification score to each testing point on each scale.

2) Information Theoretic Confidence Measure: In order to quantify the reliability of the regression score $f_j(\Gamma^j)$ computed at scale j for test point Γ_j, we have designed a measure of the local phase signature consistency of training points near Γ^j. Our measure is the local information entropy in the distribution of training points located in a ball centered at Γ^j. A large local entropy suggest that the regression score may be less accurate, due to a nearly even distribution of training points from each phase class. On the other hand, a small local entropy suggests a phase class consistency among training points and possibly a more reliable regression score.

Let $B_\theta(\Gamma^j)$ be the ball of radius θ centered at Γ^j and let $Z_{\Gamma^j} \in \{-1, 1\}$ be a binary random variable defined over $B_\theta(\Gamma^j)$. To compute the local entropy, $H(Z_{\Gamma^j})$, we let the realization $Z_{\Gamma^j} = 1$ represent a P-phase signature training point and the realization $Z_{\Gamma^j} = -1$ represent an L-phase signature training point. Both $\Pr(Z_{\Gamma^j} = 1)$ and $\Pr(Z_{\Gamma^j} = -1)$ are found by taking the relative frequency approach within $B_\theta(\Gamma^j)$. The local entropy $H(Z_{\Gamma^j})$ is given by,

$$H(Z_{\Gamma^j}) = - \sum_{x \in \{-1, 1\}} \Pr(Z_{\Gamma^j} = x) \log(Pr(Z_{\Gamma^j} = x) \tag{9}$$

$H(Z_{\Gamma^j})$ gives a measure of the phase signature randomness found in the distribution of training points near Γ^j. We notice that when $\Pr(Z_{\Gamma^j} = 1) = \Pr(Z_{\Gamma^j} = -1) = \frac{1}{2}$, $H(Z_{\Gamma^j})$ takes the maximum of 1. In this case, the number of training points within $B_\theta(\Gamma^j)$ is equally distributed between both classes. This is not a favorable distribution, since the boundary between the classes is poorly defined. On the other hand, when $H(Z_{\Gamma^j})$ is near zero, the distribution of training points in $B_\theta(\Gamma^j)$ is heavily weighted towards only one of the two phase classes. In this case, we can trust the training data, as they fall within the boundary of one class. This observation makes it possible to design a regression confidence weight to associate to each regression score. Figure 4 shows the binary entropy with corresponding confidence measure and examples of training points with different local entropy. It is worthwhile to note that in some cases, a testing point Γ^j could be surrounded by training points from the opposite class, in which case the entropy would prove very misleading. As a result, one must consider all signal scales in order to resolve outlier issues.
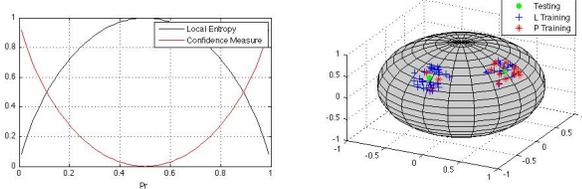
(a) Local Entropy and Confidence Measure (b) Local Training Point Distirbution

Fig. 4. (a) Shows both the binary entropy and confidence measure. Low entropy suggests high confidence while high entropy suggests low confidence. (b) Shows an example of two test points Γ^j. The point on the left has a low local entropy and therefore, a higher confidence in correct classification. The test point on the right has a high local entropy and thus a low confidence in correct classification.

Fig. 5. Map of Rocky Mountain Seismic Sensing Network.

3) Multi-scale Classification – How to Merge Scales: For each $K \times 3$ test point Φ, we have produced an $n+1$ multi-scale feature representation Γ^j where $j \in \{a_0, d_0, \cdots, d_{n-1}\}$ is the index of decomposition scale. On each scale j, we have found a regression score $f_j(\Gamma^j)$ and confidence measure $(1 - H(Z_{\Gamma^j}))$. The cumulative confidence weighted regression score is given by,

$$f(\Phi) = \sum_{j \in \{a_0, d_0, \cdots, d_{n-1}\}} (1 - H(\Gamma^j)) f_j(\Gamma^j) \qquad (10)$$

and is used for phase classification. In our case, Φ is classified as a P phase if $f(\Phi) > 0$ and a L phase if $f(\Phi) < 0$.

IV. EXPERIMENTAL APPROACH & RESULTS

We have applied our method to a data set from the Idaho, Montana, Wyoming, and Utah regions collected during 2005 and 2006. Arrival times and phases of all events were pre-determined by an expert analyst. This data set consists of 10 seismic sources measured from 12 three-channel seismic recording stations. Figure 4 shows a graphical representation of the recording stations and several seismic events. In this section, we outline our procedure for selecting data matrices Φ, describe our method for model parameter selection, and present cross-validation classification results. In addition, we compare a hypothesis testing method for phase classification presented in Jackson *et al.* (1991) [4].

A. Selecting Seismic Data Matrices Φ

Each data matrix Φ is a $K \times 3$ matrix, where each column is data collected form a different sensing channel over the same time period. The data matrix is obtained by a sliding window technique, where each data matrix is shifted in time by $\tau = 3$ samples around an activity arrival. Each matrix Φ is chosen to only contain one activity arrival. The window length parameter K is chosen to be of the form 2^m so to produce a sufficient number of wavelet decomposition levels. In our case, we have chosen $K = 512$.
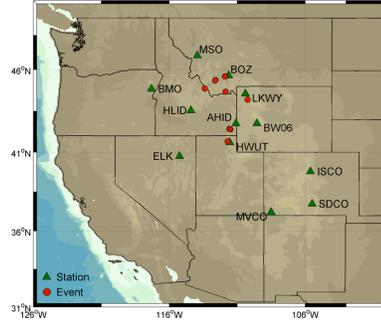
B. Parameter Selection

The model parameters used in kernel ridge regression are the regularization parameter, λ, and the kernel width parameter, σ. In machine learning, both of these parameters are typically determined by cross-validation for obtaining the best classification results. In our case, we estimate σ from the data. For each testing point Γ^j, we search for a radius θ_{Γ^j} of a ball centered at Γ^j that contains at least M training points. Depending on the density of the points near Γ^j, the search may be over exhaustive. In this case, the search is stopped upon initialization if the ball of radius θ_0 contains more than M training points. The parameter σ is given by,

$$\sigma_{\Gamma^j} = -\frac{\theta_{\Gamma^j}}{\ln(A)(\pi - \theta_{\Gamma^j})} \qquad (11)$$

where θ_{Γ^j} is the radius of a ball centered at Γ^j found by the above procedure and A is the kernel value at θ_{Γ^j}. This allows the kernel width parameter to be determined by the data. On the other hand, the regularization parameter λ is found by traditional cross-validation methods. However, λ can be interpreted as the ellasticity of the boundary between classification classes.

C. Supervised Learning Classification Results

Our experiments consist of two binary classification scenarios: (1) Pn-phase vs. Lg-phase and (2) Pg-phase vs. Lg-phase. In each scenario we are testing our classifier on compressional and surface type seismic waves. The first letter in the phase label represents the type of wave observed, either compressional, surface, shear, etc. Subsequent letters represent the path or reflections the wave experiences as it travels from the source to the observation point. Both Pn- and Pg-phase types represent primary compressional waves where "n" represents a wave that travels along the boundary between the crust and mantle and "g" represents a wave that only travels through the crust. The Lg-phase type is a surface wave that only travels through the crust.

Using the 10 seismic sources present in our data set, we test our model using a cross-validation method. Each testing instance is performed on a set of sources realized across the sensing network such that each source is only measured at a

TABLE I
PERCENTAGE CORRECT CLASSIFICATION: PN-PHASE VS. LG-PHASE

Testing Realization	Entropy Weighted		Unity Weighted	
	%Lg	%Pg	%Lg	%Pg
#1; (15, 0.01, 0.01)	76.12	85.60	69.90	81.10
#2; (15, 0.01, 0.01)	57.50	88.20	45.70	81.70
#3; (30, 0.001, 0.01)	70.00	84.50	55.60	79.60
#4; (30, 0.001, 0.01)	70.60	82.30	54.90	82.70
#5; (15, 0.01, 0.01)	76.10	55.00	64.3	46.70
Mean	**70.06**	**79.12**	**58.08**	**74.36**

Model Parameters: $\lambda = .001$, and triple above is $(M, \theta_0, \theta_{step})$

TABLE II
PERCENTAGE CORRECT CLASSIFICATION: PG-PHASE VS. LG-PHASE

Testing Realization	Entropy Weighted		Unity Weighted	
	%Lg	%Pg	%Lg	%Pg
#1; (15, 0.01, 0.01)	75.50	90.40	68.00	84.50
#2; (15, 0.001, 0.01)	70.10	78.90	59.30	75.20
#3; (30, 0.001, 0.01)	65.30	80.50	52.10	83.00
#4; (15, 0.001, 0.01)	71.00	69.5	62.5	69.6
#5; (15, 0.001, 0.01)	73.8	71.4	67.7	71.8
Mean	**71.14**	**78.14**	**61.92**	**76.82**

Model Parameters: $\lambda = .0001$ and triple above is $(M, \theta_0, \theta_{step})$

single station. Table I and Table II show classification results for five testing realizations of data for the two testing scenarios. In addition, we compare the results for two regression models, entropy weighted and unity weighted. The weighted regression model is given by,

$$f(\Phi) = \sum_{j \in \{a_0, d_0, \cdots, d_{n-1}\}} \kappa_j f_j(\Gamma^j) \qquad (12)$$

where $\kappa_j = (1 - H(\Gamma^j))$ is in the entropy weighted case and $\kappa_j = 1$ in the unity weighted case. The difference among these models is that in one case, we use our notion of regression confidence to weight the regression scores instead of simply summing regression scores across the decomposition scales. Table I and II show simulation results for these two models over several testing instances.

From the above tables, it is clear that by using a entropy type confidence weight to asses the confidence in regression we are able to achieve stronger results than simply summing regression scores across decomposition scale. In addition, experimental results not shown here suggest that relying on a subset of the regression scores does not produce a stronger classification result. In essence, it is necessary to include each regression score from each wavelet decomposition scale to successfully classify a wave as a compressional or surface phase. Our experiments show a total average correct classification of seismic data matrices Φ of 74.61%.

D. Hypothesis Testing Classification Results [4].

Jackson *et al.* (1991) describes an *F*-test method for phase classification. In this work, they use time domain descriptions of the seismic data window Φ to extract polarization estimates. By performing a SVD of the matrix Φ they obtain,

$$\Phi^j = U\Sigma V^T \qquad (13)$$

and from the singular values construct models for the residual energy for both linearly and elliptically polarized waves. These models are given by $\sigma_1^2 + \sigma_2^2 + \frac{\sigma_2^2 + \sigma_3^2}{2}$ for linearly polarized waves and $3(\sigma_3^2)$ for elliptically polarized waves. The values $\sigma_1, \sigma_2,$ and σ_3 are the singular values of Φ. They proposed using the singular values to estimate the probability that the signal contained in Φ is rectlinearly polarized as opposed to elliptically polarized.

In their method, they tested the null hypothesis (H_0) that *the wave captured in the window Φ is linearly polarized* against the alternative hypothesis (H_1) *the wave captured in the window Φ is elliptically polarized*. By assigning F to be the ratio of the energies,

$$F = \frac{\sigma_1^2 + \sigma_2^2 + \frac{\sigma_2^2 + \sigma_3^2}{2}}{3(\sigma_3^2)} \qquad (14)$$

and using

$$\Psi_{\frac{N}{N+NF}}\left(\frac{N}{2}, \frac{N}{2}\right) = \frac{1}{\beta\left(\frac{N}{2}, \frac{N}{2}\right)} \int_0^{\frac{N}{N+NF}} t^{\frac{N}{2}-1}(1-t)^{\frac{N}{2}-1} dt \qquad (15)$$

to evaluate to significance of this hypothesis. The function Ψ is the incomplete beta function where N is a parameter that counts the number of independent samples in the signal window Φ and β is the standard beta function. The number of independent samples is determined by counting the number of peaks present in the Fourier transform of the seismic signal. Table III shows results for running this algorithm on data from testing realization #1. Due to space limitations, one test realization is shown.

Table III shows, *F*-test classification accuracy compared against the analysts classification. For a null hypothesis acceptance threshold of $\eta = 0.30$, we see that for Lg- test data the null hypothesis is accepted 32.1% resulting in a correct classification on 67.9% of the test data. Similarly at the same threshold on Pg-phase and Pn-phase data the null hypothesis is accepted for only 1.3% and 8.9% of test cases, respectively. Thus, yielding poor classification on Pn- and Pg-phase test data. As the threshold decreases, we see that classification performance on Lg- test data decreases while performance increases on Pn- and Pg- data.

V. DISCUSSION

In this work, we have merged two fields of research, seismology and machine learning. We have demonstrated a supervised learning algorithm for P- and L- phase classification in $K \times 3$ seismic data. In this section, we explore the consequences of this research from the viewpoint of

TABLE III
HYPOTHESIS TEST RESULTS: TEST REALIZATION #1

η	$\Pr(Lg\|H_0) > \eta$	$\Pr(Pg\|H_0) > \eta$	$\Pr(Pn\|H_0) > \eta$
0.50	0.000	0.000	0.000
0.40	0.120	0.013	0.089
0.30	0.321	0.115	0.229
0.20	0.604	0.288	0.464
0.10	0.770	0.471	0.836
0.05	0.772	0.589	0.949

η: null hypothesis acceptance

our experimental results, dimensionality reduction, real-time implementation, and generalizations.

Our experimental results show that despite the fact that we train our classifier using seismic data observed across a variety of regional recording stations, we are still able to distinguish between P- and L-phases. These results suggest the local topographical structure surrounding the recording station does not impact the local structure of the feature manifolds. In addition, P- and L-phases cluster near each other regardless of which station in the network measured them, respectively. The strength of our method is a direct result of data clustering exhibited in the feature manifolds across each decomposition scale. Comparing to a classification method presented in [4], we see that our technique leads to stronger classification results.

Our method for multi-scale feature extraction reduces the $K \times 3$ seismic ground motion matrices to multi-scale three-dimensional polarization signature vectors. This transformation can be viewed as a non-linear dimensionality reduction from the space of $K \times 3$ matrices to multi-scale feature manifolds embedded in $\mathbf{R}^3$. The dramatic reduction in dimensionality seems to preserve enough wave information useful for classification. In addition to using the right principle singular vector for phase classification, we have experimented with using the singular values and principle components obtained in the SVD stage of our algorithm. Experiments, not shown in this paper, suggest that these other components do not improve our classification results.

In order to extend our method to run in a real-time setting where the $K \times 3$ seismic data matrix is sliding across the event arrival, the singular value decomposition step would need to be updated in real-time. Rosenburg (2010), showed an algorithm which estimated polarization parameters in real-time by using a matrix update procedure for the singular value decomposition [9]. Adapting this algorithm for field use could improve early warning system lead time and allow for more efficient use of the analysts' resources.

The strength of our technique is in the generality of the construction and processing of data in the multi-scale feature manifold setting. Our algorithm relies heavily on access to a closed form expression for the geodesic distance on the manifold. In cases where the distance on the feature manifold is known or can be estimated, kernel methods can be lifted to the manifold for direct processing.

VI. CONCLUSION

We have developed a multi-scale manifold based supervised learning algorithm for P- and L- seismic phase classification. In particular, we have transformed $K \times 3$ seismic ground motion matrices into $n + 1$ three-dimensional polarization signature feature vectors residing on spherical manifolds. By lifting linear regression techniques to the non-linear feature manifolds, we are able to resolve fundamental wave characteristics for phase classification. In addition, we have designed an information theoretic confidence measure for multi-scale model boosting. We intend to extend our methods to examine alternate supervised learning techniques and fully examine its performance against standard methods in this field.

REFERENCES

[1] K. S. Anant and F. U. Dowla. Wavelet Transform Methods for Phase Identification in Three-Component Seismograms. *Bulletin of the Seismological Society of America*, 87(6):1598–1612, December 1997.

[2] L. D'Auria, F. Giudicepietro, M. Martini, M. Orazi, R. Peluso, and G. Scarpato. Polarization analysis in the discrete wavelet domain: An application to volcano seismology. *Bulletin of the Seismological Society of America*, 100(2):670, 2010.

[3] T. Hastie, R. Tibshirani, and Friedman J. *The Elements of Statistical Learning*. Springer Verlag, 2009.

[4] G. M. Jackson, I. M. Mason, and S. A. Greenhalgh. Principle Component Transforms of Triaxial Recordings by Singular Value Decomposition. *Geophysics*, 56(4):528–533, April 1991.

[5] E. R. Kanasewich. *Time Sequence Analysis In Geophysics*. The University of Alberta Press, 1985.

[6] P. Kumar and E. Foufoula-Georgiou. Wavelet Analysis for Geophysical Applications. *Reviews of Geophysics*, 35(4):385–412, November 1997.

[7] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[8] J. Park, F. L. Vernon III, and C. R. Lindburg. Frequency Dependent Polarization of High-Frequency Seismograms. *Journal of Geophysical Research*, 92(B12):12,664–12,674, November 1987.

[9] A. Rosenburg. Real-Time Ground-Motion Analysis: Distinguishing P and S Arrivals in a Noisy Environment. *Bulletin of the Seismological Society of America*, 100(3):1252–1262, June 2010.

[10] C.D. Saragiotis, L.J. Hadjileontiadis, and S.M. Panas. A higher-order statistics-based phase identification of three-component seismograms in a redundant wavelet transform domain. In *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, pages 396–399. IEEE, 1999.

[11] M.J. Shensa. The discrete wavelet transform: Wedding the à trous and Mallat algorithms. *IEEE Trans. on Signal Processing*, 40(10):2464–2482, 1992.

[12] K. M. Taylor, M. J. Procopio, C. J. Young, and F. G. Meyer. Estimation of arrival times from seismic waves: a manifold based approach. *Geophysical Journal International*, 185:435–452, 2011.

[13] I. M. Tibuleac, E. T. Herrin, J. M. Britton, R. Shumway, and A. C. Rosca. Automatic Secondary Seismic Phase Picking Using the Wavelet Transform. *Seismological Research Letters*, 74(6):884–892, November/December 2003.

[14] M. Withers, R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo. A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bull. seism. Soc. Am.*, 88(1):95–106, Feb. 1998.

[15] M. N. Zhizhin, D. Rouland, J. Bonnin, A. D. Gvishiani, and A. Burtsev. Rapid Estimation of Earthquake Source Parameters from Patter Analysis of Waveforms Recorded at a Single Three-Component Broadband Station. *Bulletin of the Seismological Society of America*, 96(6):2329–2347, December 2006.