THE LOCAL KARHUNEN-LOÈVE BASES

Ronald R. Coifman

Department of Mathematics
Yale University
New Haven, CT 06522 USA
email: coifman@math.yale.edu

Naoki Saito

Schlumberger-Doll Research
Old Quarry Road
Ridgefield, CT 06877 USA
email: saito@ridgefield.sdr.slb.com

ABSTRACT

We propose yet another dictionary of orthonormal bases (which has the same tree structure as the popular wavelet packet or local trigonometric dictionaries) adapted to a given ensemble of signals. These orthogonal waveforms are generated by a set of locally adapted versions of the Karhunen-Loève (KL) transform. The basis vectors in this dictionary represent *local features* in the time-frequency plane compared to the standard KL basis vectors. Because of the structure of the bases, the best basis selection algorithm of Coifman-Wickerhauser is readily applicable. Moreover, no a priori choice of conjugated quadrature filters or cosine/sine polarity is necessary; it is completely data driven. The computational cost to build this dictionary is comparable to or potentially less than that of the standard KL transform. As an application, we give an example of clustering geophysical acoustic waveforms.

1. INTRODUCTION

As a feature extraction tool for signal representation, compression, or clustering, the Karhunen-Loève (KL) expansion or the Principal Component Analysis (PCA) has been popular for more than 30 years [1]. The standard KL basis is an orthonormal basis which provides decorrelated coordinates for a given ensemble of signals or equivalently diagonalizes the covariance matrix of the ensemble. The KL basis is considered as the most efficient coordinate system for representing such an ensemble of signals under several criteria: it has the entropy-minimizing and ℓ^2-error-minimizing properties [1]. However, the KL basis has a few drawbacks. First, its computational cost is $O(n^3)$ (where n is a number of time samples in each signal) since it is a solution of the eigenvalue problem. Second, it is difficult to capture signal features localized in the time-frequency plane due to the global eigenvectors.

Over the last five years, many new tools for signal representation/compression have been proposed. These include the so-called *dictionaries of orthonormal bases* (see e.g., [2]). A dictionary of orthonormal bases consists of a redundant set of orthonormal bases and has a tree structure each of which node represents basis vectors spanning a subspace of specific time-frequency localization character. The wavelet packet dictionary and the local trigonometric dictionary are two popular examples of such a dictionary. This type of dictionary contains a large number of complete orthonormal bases. Therefore, this allows one to search the so-called *best basis* (from the dictionary) which is adapted to a particular input signal or to a specific task. This search is normally done by minimizing a certain information cost function (e.g., entropy) using the divide-and-conquer (or split-and-merge) algorithm. These new tools offer an ability to capture local features as well as computational efficiency. From the user's point of view, however, there is always a question: "Which pair of conjugated quadrature filters should I choose?" "Should I use local cosine dictionary or local sine dictionary?" Quite often, the answer is *"It depends on the characteristics of the input signals."*

In this paper, we propose yet another dictionary of orthonormal bases completely driven by a given ensemble of input signals. We call this a dictionary of *local Karhunen-Loève bases.* This dictionary is a collection of the KL basis vectors locally adapted in either the time or the frequency domain and has the same structure as the dictionaries of wavelet packets and local trigonometric bases. At least for building this dictionary, no specification is required from the users; i.e., *"The input signals speak for themselves."* Although the computational cost of constructing this dictionary is more expensive (slightly less than $O(\frac{4}{3}n^3)$) than the local trigonometric dictionary ($O(n[\log n]^2)$), it is comparable to the standard KL basis ($O(n^3)$), and can be much less than that as explained in the next section. Once the dictionary is built, selecting a basis from this dictionary is fast, i.e., $O(n)$, and expanding each signal into such a basis costs at most $O(n^2)$.

2. A DICTIONARY OF LOCAL KL BASES

In the following we describe the time domain version of the algorithm of building a dictionary of local KL bases and selecting a suitable basis from the dictionary for our purpose. The entire procedure can also be applied to the frequency domain representations of the signals, i.e., the time (sub)intervals below should be replaced by the frequency (sub)bands for that case. The algorithm is simple and straightforward:

SPLIT: split the whole time interval supporting the input signals into a redundant set of subintervals by the smooth orthogonal projection of Coifman and Meyer (see e.g., [2, Chap. 4]).

EIGEN: construct the standard KL basis on each subinterval.

Once such a dictionary of bases is computed and stored, then we can:

MERGE: select an optimal (in some sense) cover of the original time interval and the corresponding basis.

One can select a single basis which does a good job on the average for the whole input signals. One can also select a set of bases optimized for some subset of the input signals or even an optimized basis per input signal. It is at user's disposal once the dictionary is built.

Let us now describe the above three steps in detail.

Step SPLIT. This step applies the orthogonal projection operator using the smooth bell function to the signals to segment them smoothly into local pieces. See [2, Chap. 4] for details of this projection operator. By this projector, the signals supported on the original global interval I_0 can be decomposed into the local pieces supported on $\{I_k\}$. We are interested in the redundant set of subintervals $\{I_k\}$ and the associated subspaces and their orthonormal bases over $\{I_k\}$. Step MERGE selects an optimal non-redundant cover from $\{I_k\}$ (or equivalently an optimal complete orthonormal basis) for one's need. The selected basis vectors should capture the local features/phenomena of the signals. There are at least two efficient ways to perform the SPLIT and MERGE processes. One is to split the original interval into its left and right half intervals and repeat the process recursively. This process generates the popular binary-tree structured subspaces/subintervals. This binary tree contains more than $2^{2^{(J-1)}}$ complete orthonormal bases if we repeat the recursive splitting process J times. To obtain a best possible basis for one's need from this dictionary one can invoke the best-basis search algorithm of Coifman and Wickerhauser with a suitable cost function [2, Chap. 8]. This search process is rapid, i.e., $O(n)$.

The other method is due to X. Fang [3]; this splits the original interval I_0 into the many short subintervals of equal length, $I_1, I_2, \ldots, I_K$ (similar to the bottom level leaves of the binary tree representation). Then it examines whether it is worth merging the adjacent intervals from left to right. In other words, consider the first two subintervals I_1,I_2 and their parent $I_{1,2} = I_1 \cup I_2$. If it is worth uniting I_1 and I_2 under some criterion (which will be specified below), then use $I_{1,2}$. Next, examine whether I_3 should be merged with $I_{1,2}$. This process continues as long as it is worthwhile to merge the adjacent subintervals. Suppose it merged subintervals $I_1, I_2, \ldots, I_k$, but it decided not to merge the next subinterval I_{k+1}. Then, the merging process restarts by comparing I_{k+1}, I_{k+2} and $I_{k+1,k+2}$.

Step EIGEN. This process computes the KL basis on each subinterval. Thus, on each subinterval I_k, it requires $O(n_k^3)$ operations for diagonalizing the covariance matrix of the projected signals onto I_k, where $n_k \leq n$ is a length of each projected signal. Because this KL basis computation is required for each I_k, as a whole, it seems much more expensive than the one step global KL transform; however, our local KL transform with binary tree splits cost at most $O(\frac{4}{3}n^3)$ since each split reduces the dimensionality of the problem to half. In fact, if the levels (or depth) of the tree is J, then the computational cost is $O(\frac{4}{3}n^3(1-4^{-J-1}))$. We also need to store eigenvectors at each subinterval. The space required to store all of these eigenvectors is $2n^2(1-2^{-J-1})$ floating points. Expansion of an input signal into the bases costs $O(2n^2(1-2^{-J-1}))$ additions and multiplications. We note that the computational cost can be substantially smaller if we are only interested in the time intervals shorter than a certain length since this avoids the EIGEN step from the root node to the nodes at the corresponding level. Also, the splitting algorithm of Fang can be much more inexpensive than the binary tree version since it does not require the global KL basis computation in general.

Step MERGE. Once the dictionary is constructed, we want to select a best possible basis out of many possible bases from this dictionary for our needs. For the purpose of signal representation/compression or even signal clustering, the reduction of dimensionality is of critical importance. Main purpose here is to describe or represent a given ensemble of signals with a small number of basis functions each of which has some "physical meaning" or allows one to make "easy interpretation." Step MERGE is the key step for this purpose. It decides whether it is worth merging two adjacent subintervals or not in a recursive manner. For each subinterval, the Shannon entropy of the eigenvalues of the covariance (or autocorrelation) matrix measures the flatness of the variance (or energy) distribution over the KL coordinates under consideration; in other words, it measures inefficiency of the coordinates for compression. If we adopt this entropy measure as a cost functional to minimize, however, the original global interval is always selected because of the optimality property of the global KL basis [1]. Therefore, we must use other criteria to capture the local features of the signals. There are many possible criteria. Here, we only mention two simple ones. One is based on the $\ell^{1/2}$ norm of those eigenvalues. The other is based on the mean absolute deviations of the expansion coefficients. Let $\boldsymbol{\lambda}_\alpha = (\lambda_\alpha(1), \dots, \lambda_\alpha(n_\alpha))$, $n_\alpha < n$ be a set of eigenvalues (they are all nonnegative) of the covariance matrix of a particular subinterval I_α. In fact, these eigenvalues represent variances of the input signals under the KL coordinates over I_α. Then, we measure the inefficiency of the subinterval by the following nonlinear functional:

$$\mathcal{E}(\boldsymbol{\lambda}_\alpha; m) \triangleq \sum_{k=1}^{m} \left(\tilde{\lambda}_\alpha(k)\right)^p, \qquad (1)$$

where $m \leq n_\alpha$ is a specified number of the coordinates the user wants to use, say e.g., $m = 10$ for the signals of length $n = 256$, and $\tilde{\lambda}_\alpha(k)$s are decreasing rearrangement of the sequence $\boldsymbol{\lambda}_\alpha$, and $0 < p < 1$. With $p = \frac{1}{2}$, this cost functional measures the spread of the distribution of the standard deviations over the coordinates after ignoring the coordinates whose contribution is small.

The other criterion replaces the variances $\lambda_\alpha(k)$ by the mean absolute deviations of the ensemble of the input signals under the local KL coordinates. In this case, we set $\lambda_\alpha(k) = E|Y_\alpha(k)|$ where $Y_\alpha(k)$ is the kth component of the local KL coordinates of the projection of a random input signal $\boldsymbol{X}$ onto I_α, and E is the expectation operation. Then, for this cost functional, we use (1) with $p = 1$ instead of $p = \frac{1}{2}$.

Now let B_α be the standard KL basis computed over I_α, and be A_α be the best KL basis over I_α which may be B_α or direct sum of the shorter KL bases over some descendant nodes of I_α. Then, the following algorithm selects the so-called *Local Karhunen-Loève Basis* (LKLB):

- Set $A_\omega = B_\omega$ for every I_ω which is a leaf (a bottom level node) of the tree.
- For every I_α above the bottom level and its two children I_β, I_γ, do the following:

  If $\mathcal{E}(\boldsymbol{\lambda}_\alpha; m) < \mathcal{E}(\{\boldsymbol{\lambda}_\beta, \boldsymbol{\lambda}_\gamma\}; m)$,
  Then $A_\alpha = B_\alpha$.
  Else $A_\alpha = A_\beta \oplus A_\gamma$ and replace $\boldsymbol{\lambda}_\alpha$ by $\{\boldsymbol{\lambda}_\beta, \boldsymbol{\lambda}_\gamma\}$.

We note that (1) is not additive for $m \neq n$. The mechanism of this selection algorithm is essentially the same as the local regression basis selection proposed by the authors [4, Chap. 5], [5].

3. AN EXAMPLE

We apply the LKLB algorithm to the real geophysical dataset used in [4, Chap. 6], [6] which is shown in Figure 1. This dataset consists of 402 acoustic waveforms recorded in a borehole. Each waveform has 256 time samples with $\delta t = 10\mu$sec. The top 201 waveforms represent the ones propagated through sandstone layers ("sand waveforms") and the bottom 201 waveforms represent the ones through shale layers ("shale waveforms"). Figure 2 and 3 show the top 10 global KLB vectors and the top 10 LKLB vectors (with $m = 10$), respectively. The two different selection criteria discussed in the previous section yielded almost the same result; the only difference is the order of the last few basis functions. Figure 4 shows a cross-plot or projection onto the first and seventh LKLB vectors. The points indicated by $*$ and $\cdot$ correspond to sand and shale waveforms, respectively. A combination of the cross-plot and the plots of LKLB vectors allows one to make easier interpretation of waveform analysis and clustering.

4. DISCUSSION

The close relationship between the KL transform and discrete cosine transform (DCT) was pointed out in [7]. In fact, DCT is a limiting case of KLT if the signals under consideration obey the first order Markov process. Therefore, we conjecture that the local cosine transform (LCT) is a limiting case of the local KL transform proposed in this paper. Actual comparison between the LCT and LKLT is currently under investigation.

We also note that it is a straightforward exercise to extend the algorithm for images because of its construction. The image version of this algorithm may be useful for texture segmentation problems.

Finally, we remark that the same splitting process can be used for signal classification problems. For classification, we replace the KL basis computation by the linear discriminant analysis of Fisher, and the cost function by the classification error or some functional measuring the "distances" among classes. We are currently investigating this *local linear discriminant analysis* (LLDA) and comparing it with the local discrimi-

Figure 1

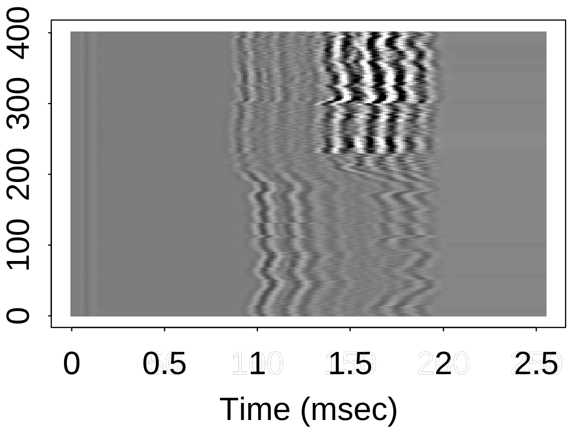


Figure 2

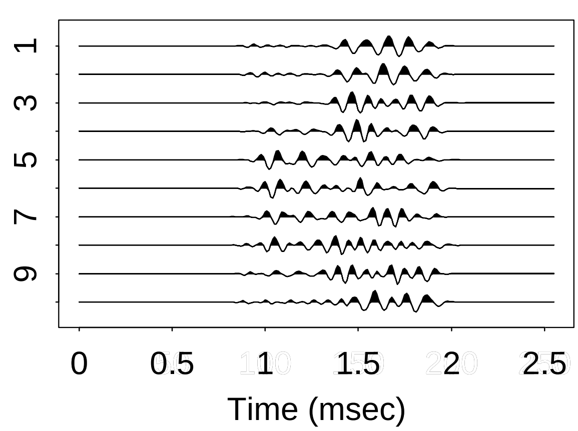


Figure 3

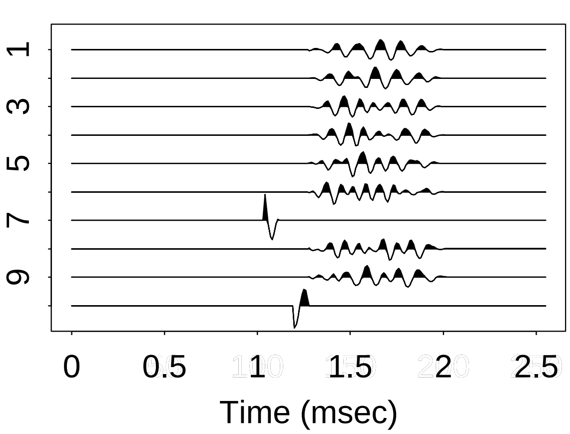
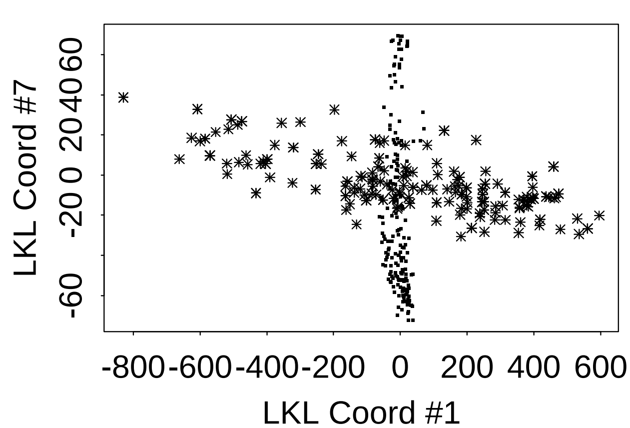


Figure 4



nant bases (LDB) of Saito and Coifman [5], [4, Chap. 4], [8].

5. REFERENCES

[1] S. Watanabe, "Karhunen-Loève expansion and factor analysis: theoretical remarks and applications," in *Trans. 4th Prague Conf. Inform. Theory, Statist. Decision Functions, Random Processes*, (Prague), pp. 635–660, Publishing House of the Czechoslovak Academy of Sciences, 1965.

[2] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*. Wellesley, MA: A K Peters, Ltd., 1994. with diskette.

[3] X. Fang, "Personal communication.".

[4] N. Saito, *Local Feature Extraction and Its Applications Using a Library of Bases*. PhD thesis, Dept. of Mathematics, Yale University, New Haven, CT 06520 USA, Dec. 1994.

[5] R. R. Coifman and N. Saito, "Constructions of local orthonormal bases for classification and regression," *Comptes Rendus Acad. Sci. Paris, Série I*, vol. 319, pp. 191–196, Jul. 1994.

[6] N. Saito and R. R. Coifman, "Geological information extraction from acoustic well-logging waveforms using time-frequency atoms," *Geophysics*, 1995. submitted.

[7] R. J. Clarke, "Relation between the Karhunen Loève and cosine transforms," *IEE Proceedings, Part F*, vol. 128, pp. 359–360, Nov. 1981.

[8] N. Saito and R. R. Coifman, "Local discriminant bases and their applications," *J. Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 337–358, 1995. Invited paper.