The Local Fourier Dictionary: A Natural Tool for Data Analysis

Naoki Saito

Department of Mathematics, University of California, Davis, CA 95616 USA

ABSTRACT

The *local Fourier dictionary* contains a large number of localized complex exponential functions. Representations of a function using the dictionary elements *locally* inherit many nice properties of the conventional Fourier representation, such as translation invariance and orientation selectivity. In this paper, after giving an intuitive review of its construction, we describe an algorithm to recover location-dependent shifts of local features in signals for matching and registration, and propose a *best local translation basis* selected from the local Fourier basis. Then we will report our preliminary results on the statistical analysis of natural scene images using the local Fourier dictionary, whose purpose is to examine the importance of *sparsity*, *statistical independence*, and *orientation selectivities* in representation and modeling of such images.

Keywords: Fourier transform, local Fourier transform, local cosine transform, best basis, translation invariance, sparse representation, statistical independence, statistics of natural scenes

1. INTRODUCTION

To analyze data, signals, and images measured by various sensors, the Fourier transform is undoubtedly one of the most popular tools. It decomposes the data into cosines and sines of different frequencies, and we can examine whether the data contains certain periodicity or not, or how the data are composed as a linear combinations of complex exponentials. Translation of a signal in the time domain is just a multiplication by the phase factor $\mathrm{e}^{-2\pi \mathrm{i} \tau \xi}$ in its Fourier representation. One can compute the amplitude, phase, and envelope of a signal by considering the analytic signal. For images, the Fourier transform has several additional advantages. The basis functions can capture *orientation* information unlike the DCT or wavelets. It also commutes with the rotation operations of an image. In spite of such convenient properties, the Fourier transform has a fundamental drawback: the basis functions are of completely global nature in time. Expansion coefficients of a signal are computed from all of the time samples because the basis functions oscillate globally over the entire support of the signal. For many signal and image analysis tasks, it is of critical importance to capture the local frequency contents and local orientation information. Splitting an image into pieces with the characteristic functions completely obscures the important frequency information due to the sharp edges of the characteristic functions. Therefore, we should use a smooth window function before performing local frequency analysis. The celebrated Balian-Low theorem[1] states that we cannot construct an orthonormal basis that optimizes localization of its basis functions jointly in the time and frequency domains if we restrict our basis functions in the following form of the windowed complex exponential functions:

$$g_{k,\ell}(t) \stackrel{\Delta}{=} g(t - k\Delta t)\mathrm{e}^{2\pi \mathrm{i} \ell \Delta \xi t},$$

where $g(\cdot)$ is a window function such as the Gaussian, Δt and $\Delta \xi$ are the basic sampling intervals along the time and the frequency axes, respectively. This is the form of Gabor elementary functions.

To circumvent these problems, the Wilson bases[2] and the local cosine/sine bases[3,4] have been developed. Both of these constructions, however, use sines and cosines rather than complex exponentials. For the applications requiring local phase and orientation information, the use of the complex exponentials is essential. Wickerhauser[5,6] developed the so-called *smooth orthogonal periodization* operator that smoothly restricts a function to an interval and periodizes it, which in turn, allows us to decompose it into localized complex exponentials with minimal edge effect. Thus, we can enjoy nice properties of the Fourier transform (e.g., translation invariance of the Fourier magnitudes and its commutativity with rotations) *locally* once we construct the local Fourier bases using these operators.

Despite its attractive properties, it seems to the author that the local Fourier dictionary (a hierarchical collection of localized complex exponential functions) has not gained popularity compared to the local cosine/sine dictionaries

Further author information: E-mail: saito@math.ucdavis.edu; WWW: http://math.ucdavis.edu/~saito

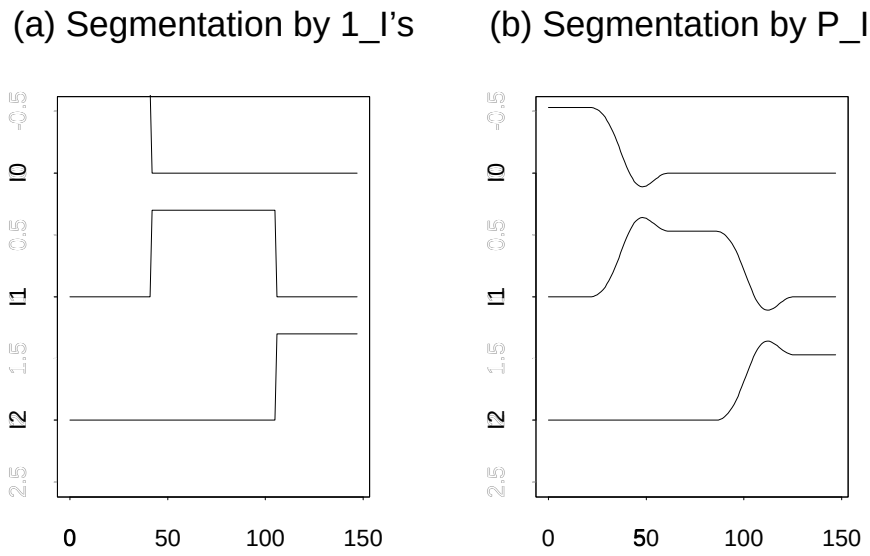


Figure 1. Splitting the constant function into three intervals, by the characteristic functions (a), and by the smooth orthogonal projections (b).

or the wavelet packet dictionaries. One of the purposes of this paper is to popularize the local Fourier dictionary. Therefore, in Section 2, we review the construction of the local Fourier bases and dictionary and describe the ideas behind their construction. The other purpose of this paper is to demonstrate its usefulness, in particular, for recovering location-dependent translations of features in signals (Section 3), and to examine the importance of the orientation information, sparseness of the image representations, and statistical independence of the coordinates for modeling or compressing natural scene images using this dictionary (Section 4).

2. CONSTRUCTION OF LOCAL FOURIER BASES

In this section, we review the construction of the local Fourier basis using the notation of Wickerhauser.[6] In order to understand the construction clearly, we need to contrast it with the local cosine basis. Therefore, we start with the smooth orthogonal projection, which is indispensable to construct the local cosine basis.

2.1. Smooth Orthogonal Projection

Suppose we are given a function $x(t) \in L^2(\mathbb{R})$. We would like to split this function smoothly into pieces each of which is supported on an interval I_k (with some overlap), where $\bigcup_{k\in\mathbb{Z}} I_k = \mathbb{R}$, and I_k's are disjoint. We define a sequence of subspaces Ω_k associated with the intervals I_k, and we demand orthogonality among Ω_k, i.e., $L^2(\mathbb{R}) = \bigoplus_{k\in\mathbb{Z}} \Omega_k$. (We refer the reader to Matviyenko[7] for biorthogonal cases that allow better time-frequency localization with additional procedures.) To achieve this goal, Coifman and Meyer[4] (see also Auscher, Weiss, and Wickerhauser[8]) introduced the following *smooth orthogonal projector* from $L^2(\mathbb{R})$ into $\Omega_k = P_{I_k} L^2(\mathbb{R})$:

$$P_{I_k} x(t) \triangleq U_{I_k}^* \mathbf{1}_{I_k} U_{I_k} x(t). \tag{1}$$

This operator consists of three operators, that is, U_{I_k} (*unitary folding operator*), $\mathbf{1}_{I_k}$ (restriction operator, i.e., $\mathbf{1}_{I_k} x(t) = x(t)$ if $t \in I_k$, $= 0$ otherwise), and $U_{I_k}^*$ (*unitary unfolding operator*, the adjoint of U_{I_k}). P_{I_k} is a smoother version of sharp segmentation $\mathbf{1}_{I_k}$. Figure 1 shows these two different segmentations of the constant function $x(t) \equiv 1$. Let $I_k = (\alpha_k, \alpha_{k+1})$. Then the unitary folding operator U_{I_k} for the interval I_k is defined as

$$U_{I_k} x(t) \triangleq U(r_k, \alpha_k, \epsilon_k) U(r_{k+1}, \alpha_{k+1}, \epsilon_{k+1}) x(t),$$

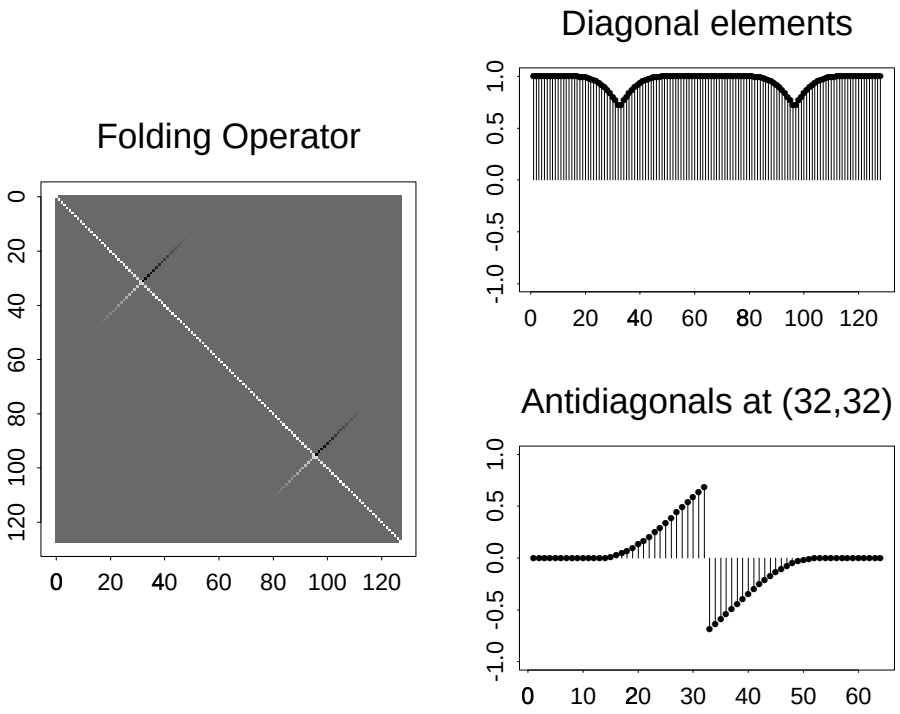


Figure 2. A typical folding operator U_I. In this case, $I = (31.5, 95.5)$, and $\epsilon = 16$ at both the left and right borders of the interval. The unfolding operator U_I^* is simply a transposition of U_I in this case.

where $U(r, \alpha, \epsilon)$ is a unitary folding operator associated with the *action region* $t \in (\alpha - \epsilon, \alpha + \epsilon)$ and is defined as

$$U(r,\alpha,\epsilon)x(t) \triangleq \begin{cases} r\left(\frac{t-\alpha}{\epsilon}\right) x(t) + r\left(\frac{\alpha-t}{\epsilon}\right) x(2\alpha - t) & \text{if } \alpha < t < \alpha + \epsilon, \\ r\left(\frac{\alpha-t}{\epsilon}\right) x(t) - r\left(\frac{t-\alpha}{\epsilon}\right) x(2\alpha - t) & \text{if } \alpha - \epsilon < t < \alpha, \\ x(t) & \text{otherwise.} \end{cases}$$

The function $r(t)$ above is called a *rising cutoff function*, which is a smooth version (e.g., $r \in C^d(\mathbb{R})$ with $d \in \mathbb{N}$) of the Heaviside step function satisfying the following condition:

$$|r(t)|^2 + |r(-t)|^2 = 1 \quad \text{for all } t \in \mathbb{R}, \quad \text{and} \quad r(t) = \begin{cases} 0 & \text{if } t \le -1, \\ 1 & \text{if } t \ge 1. \end{cases}$$

A typical example of $C^1(\mathbb{R})$ is the following iterated sine function:

$$r(t) = \begin{cases} 0, & \text{if } t \le -1, \\ \sin[\frac{\pi}{4}(1 + \sin\frac{\pi}{2}t)], & \text{if } |t| < 1, \\ 1 & \text{if } t \ge 1. \end{cases}$$

For various possible choices of the rising cutoff functions and their detailed properties, we refer the reader to the article of Matviyenko.[7] Figure 2 shows a typical folding operator for an interval in a matrix form. As one can see from the plots in Figure 2, the folding operator $U(r, \alpha, \epsilon)$ makes a signal *locally even* for the region $(\alpha, \alpha + \epsilon)$ and *locally odd* for the region $(\alpha - \epsilon, \alpha)$.

The adjoint operator of $U(r, \alpha, \epsilon)$ is called the unitary unfolding operator:

$$U^*(r,\alpha,\epsilon)x(t) \triangleq \begin{cases} r\left(\frac{t-\alpha}{\epsilon}\right) x(t) - r\left(\frac{\alpha-t}{\epsilon}\right) x(2\alpha - t) & \text{if } \alpha < t < \alpha + \epsilon, \\ r\left(\frac{\alpha-t}{\epsilon}\right) x(t) + r\left(\frac{t-\alpha}{\epsilon}\right) x(2\alpha - t) & \text{if } \alpha - \epsilon < t < \alpha, \\ x(t) & \text{otherwise.} \end{cases}$$

Both $U = U(r, \alpha, \epsilon)$ and $U^* = U^*(r, \alpha, \epsilon)$ are unitary isomorphisms of $L^2(\mathbb{R})$ since $U^*Ux(t) = UU^*x(t) = x(t)$ for $t \neq \alpha$, and $Ux(t) = U^*x(t) = x(t)$ if $|t - \alpha| \geq \epsilon$. For the operator $U_{I_k} = U(r_k, \alpha_k, \epsilon_k)U(r_{k+1}, \alpha_{k+1}, \epsilon_{k+1})$, we demand that the two action regions around α_k and α_{k+1} of I_k do not interfere, i.e., we must have $\alpha_k + \epsilon_k < \alpha_{k+1} - \epsilon_{k+1}$. The size of the action regions $2\epsilon_k$ around the boundary α_k and the rising cutoff function $r_k(t)$ can be either dependent or independent of I_k. A clever choice of ϵ_k and $r_k(t)$ dependent of I_k leads to better time-frequency localization schemes such as the multiple folding of Fang and Séré[9] and the time-frequency local cosines of Villemoes.[10]

Let $C_{m,k}(t) \triangleq \sqrt{2/|I_k|}\cos(\pi(m+\frac{1}{2})(t-\alpha_k)/|I_k|)$. Then $\{\mathbf{1}_{I_k}C_{m,k}\}_{m\in\{0\}\cup\mathbb{N}}$ form an orthonormal basis of $L^2(I_k)$. Coifman and Meyer[4] defined the *local cosine functions* $\phi_{m,k}(t) \triangleq U^*_{I_k}\mathbf{1}_{I_k}C_{m,k}$, and showed that the set $\{\phi_{m,k}\}_{m\in\{0\}\cup\mathbb{N}}$ forms an orthonormal basis of $\Omega_k = P_{I_k}L^2(\mathbb{R})$, thus the set $\{\phi_{m,k}\}_{m\in\{0\}\cup\mathbb{N},k\in\mathbb{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$. The fact that $\{\phi_{m,k}\}_{m\in\{0\}\cup\mathbb{N}}$ is an orthonormal basis of Ω_k can be proved by noticing that: 1) $\Omega_k = P_{I_k}L^2(\mathbb{R}) = U^*_{I_k}\mathbf{1}_{I_k}U_{I_k}L^2(\mathbb{R})$ is isomorphic to $U^*_{I_k}\mathbf{1}_{I_k}L^2(\mathbb{R}) = U^*_{I_k}L^2(I_k)$ because U_{I_k} is a unitary isomorphism on $L^2(\mathbb{R})$, and 2) $\{\mathbf{1}_{I_k}C_{m,k}\}_{m\in\{0\}\cup\mathbb{N}}$ is an orthonormal basis of $L^2(I_k)$.

2.2. Smooth Orthogonal Periodization

Wickerhauser[5] found a modification to the smooth orthogonal projection P_{I_k}, and provided a way to smoothly restrict a function to an interval and periodize it, which in turn permits us to expand it into a periodic basis with minimal edge effect. This modification, the *smooth orthogonal periodization*, is defined as follows:

$$T_{I_k}x(t) \triangleq W^*_{I_k}\mathbf{1}_{I_k}U_{I_k}x(t). \tag{2}$$

In order to understand this operator, let us first define an *I-periodic extension* (or *I-periodization*) x_I of a function $x \in L^2(I)$ as

$$x_I(t) \triangleq \sum_{k\in\mathbb{Z}} x(t - k|I|).$$

Then, the operator T_{I_k} maps an $L^2_{loc}(\mathbb{R})$ function into an I_k-periodic extension of an $L^2(I_k)$ function. Moreover, T_{I_k} preserves the smoothness of the smooth functions: if $x \in C^d(\mathbb{R})$, then $T_{I_k}x$ has an I_k-periodic extension that also belongs to $C^d(\mathbb{R})$. The *periodized unfolding operator* $W^*_{I_k}$ in Equation (2) is defined as:

$$W^*_{I_k}x(t) = W^*(r_k, I_k, \epsilon_k)x(t) = \begin{cases} r_k(\frac{t-\alpha_k}{\epsilon_k})x(t) - r_k(\frac{\alpha_k-t}{\epsilon_k})x(\alpha_k + \alpha_{k+1} - t) & \text{if } \alpha_k < t < \alpha_k + \epsilon_k, \\ r_k(\frac{\alpha_{k+1}-t}{\epsilon_k})x(t) + r_k(\frac{t-\alpha_{k+1}}{\epsilon_k})x(\alpha_k + \alpha_{k+1} - t) & \text{if } \alpha_{k+1} - \epsilon_k < t < \alpha_{k+1}, \\ x(t) & \text{otherwise.} \end{cases}$$

This is the adjoint operator of the *periodized folding operator* W_{I_k}:

$$W_{I_k}x(t) = W(r_k, I_k, \epsilon_k)x(t) = \begin{cases} r_k(\frac{t-\alpha_k}{\epsilon_k})x(t) + r_k(\frac{\alpha_k-t}{\epsilon_k})x(\alpha_k + \alpha_{k+1} - t) & \text{if } \alpha_k < t < \alpha_k + \epsilon_k, \\ r_k(\frac{\alpha_{k+1}-t}{\epsilon_k})x(t) - r_k(\frac{t-\alpha_{k+1}}{\epsilon_k})x(\alpha_k + \alpha_{k+1} - t) & \text{if } \alpha_{k+1} - \epsilon_k < t < \alpha_{k+1} \\ x(t) & \text{otherwise.} \end{cases}$$

It is easy to show that both W_I and W^*_I are isomorphisms of $L^2(\mathbb{R})$ as well as $L^2(I)$. We note that the periodized folding and unfolding operators require the same action region size and the same rising cutoff function at the boundaries of the the intervals. We also note that the action regions and rising cutoff functions used in $W^*_{I_k}$ and U_{I_k} can be chosen differently. Figure 3 shows the periodized unfolding operator W^*_I for the same interval as Figure 2. Figure 4 contrasts T_I with P_I.

To see the difference between T_I and P_I more clearly, we applied these operators to a constant function and a linear function. The results are shown in Figure 5. What is the exact relationship between T_I and P_I? The answer is summarized in the following equations:

$$T^*_I T_I = P_I, \quad T_I T^*_I = \mathbf{1}_I. \tag{3}$$

These are easily derived from the definitions (1) and (2). For an I-periodic extension of a function supported on I, T_I simply restricts such extension to the interval I, i.e., does the role of $\mathbf{1}_I$, as shown in Figure 5. However, its adjoint $T^*_I = U^*_I\mathbf{1}_I W_I : L^2(I) \to U^*_I L^2(I) \simeq \Omega_k$, plays a key role for an I-periodic extension of a function as follows.

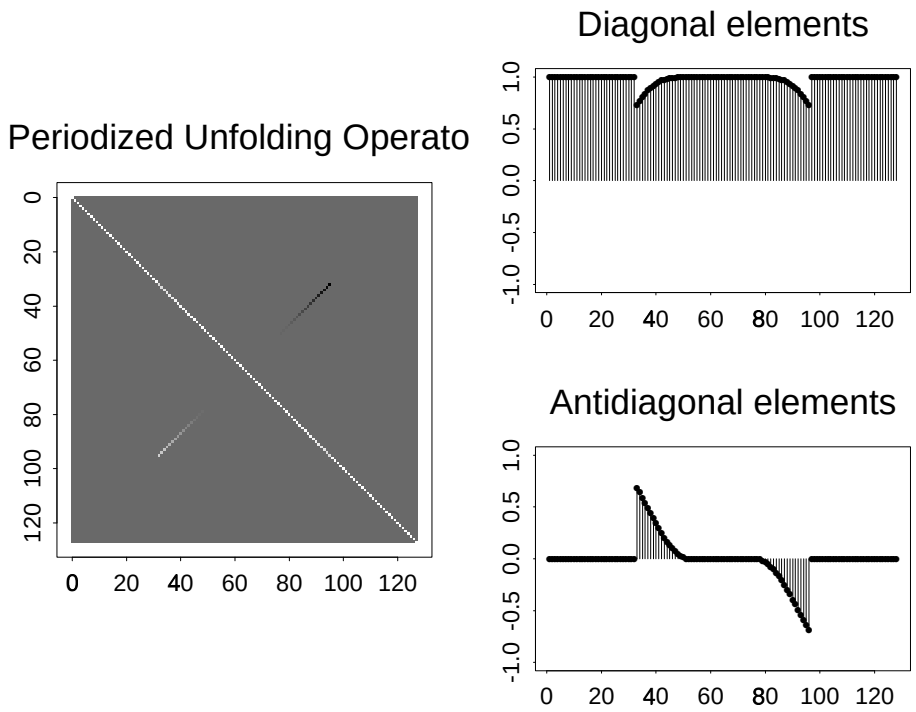


Figure 3. The periodized unfolding operator W_I^*. The size of the action region is the same as Figure 2. The periodized folding operator W_I is simply a transposition of W_I^* in this case.

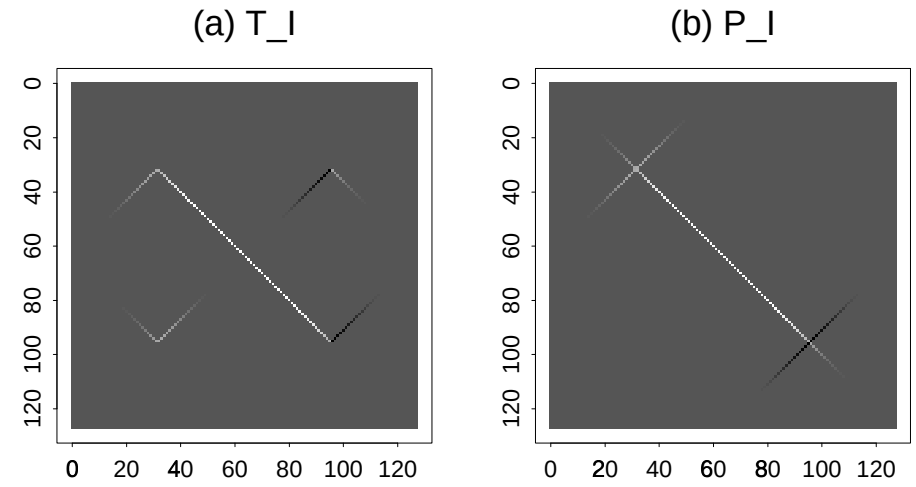


Figure 4. (a) The smooth orthogonal periodization operator T_I, and (b) the smooth orthogonal projector P_I.

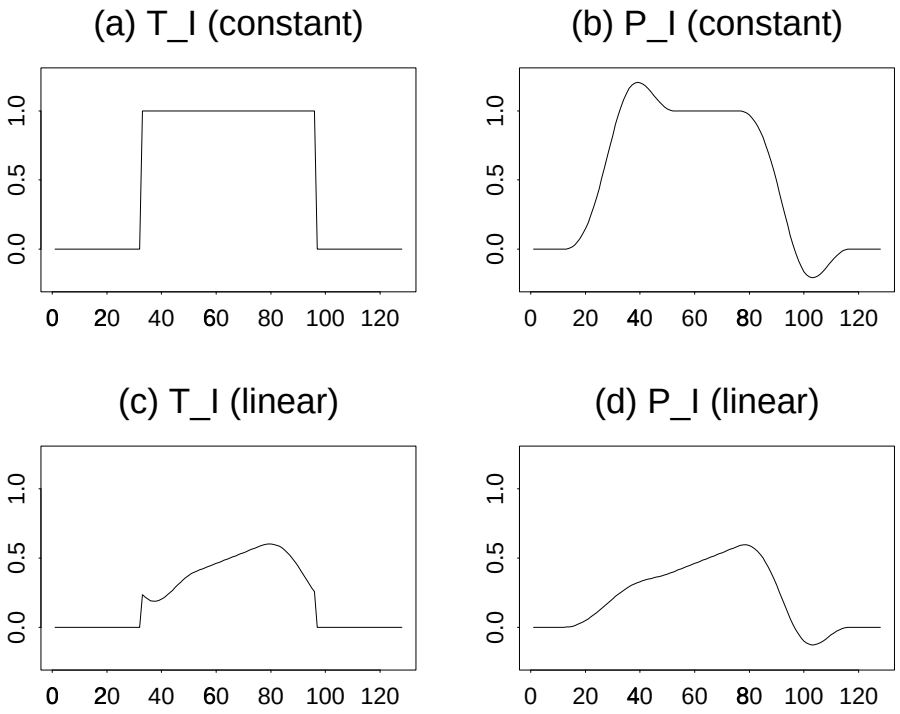


Figure 5. The operators in action. (a) T_I applied to the constant function 1. (b) P_I applied to the constant function. (c) T_I applied to a linear function. (d) P_I applied to the linear function. Note that if the original function is periodic, T_I does simple restriction.

Proposition 2.1. *Let $x_I(t)$ be an I-periodic extension of $x \in L^2(I)$. Then,*

$$T_I^* x_I(t) = P_I x_I(t).$$

Proof. Lemma 4.7 of Wickerhauser[6] states that if W_I and U_I share the same action regions and the rising cut-off function, then for any $f \in L^2(\mathbb{R})$ we have

$$W_I \mathbf{1}_I f = \mathbf{1}_I U_I (\mathbf{1}_I f)_I,$$

where $(\mathbf{1}_I f)_I$ is an I-periodic extension of the restriction $\mathbf{1}_I f$. Therefore, if we start with the I-periodized function x_I instead of f above, we immediately have $W_I x_I = U_I x_I$. Using this fact, it is easy to derive

$$T_I^* x_I = U_I^* \mathbf{1}_I W_I x_I = U_I^* \mathbf{1}_I U_I x_I = P_I x_I.$$

□

Now for a general function $x \in L^2(\mathbb{R})$, using Equation (3), the smooth orthogonal projection is realized in *two steps*:

$$P_I x = T_I^* T_I x. \tag{4}$$

The first step $T_I x$ makes $x(t)$ smoothly localized and I-periodized, which allows us to expand $T_I x$ into a periodic basis. Then, the second step $T_I^*(T_I x)$ maps into $\Omega_I = P_I L^2(\mathbb{R})$. This is the significant difference from the local cosine functions, which is realized just by *one step*.

2.3. The local Fourier basis

Wickerhauser[5,6] constructed a smooth localized orthonormal basis using the smooth orthogonal periodization in the previous subsection. Let $\mathbb{R} = \bigcup_{k \in \mathbb{Z}} I_k$ where all the I_k's have disjoint action regions. Let $\{e_{m,k}(t) : m \in \mathbb{Z}\}$ be a

periodic orthonormal basis of $L^2(I_k) = \mathbf{1}_{I_k} L^2(\mathbb{R})$ with period $|I_k| = \alpha_{k+1} - \alpha_k$. A typical example is the complex exponentials, $e_{m,k}(t) = (1/\sqrt{|I_k|})e^{2\pi i m(t-\alpha_k)/|I_k|}$. Then, as we already indicated in Equation (4) and the remarks after that, we have the following theorem.

Theorem 2.2 (Wickerhauser[5,6]). *The set $\{T_{I_k}^* e_{m,k}\}_{m\in\mathbb{Z}}$ is an orthonormal basis of $\Omega_k = P_{I_k} L^2(\mathbb{R})$, and the set $\{T_{I_k}^* e_{m,k}\}_{(m,k)\in\mathbb{Z}^2}$ is an orthonormal basis of $L^2(\mathbb{R})$.*

From this theorem, it is easy to do both analysis and synthesis. We have the following expansion of $x(t)$:

$$\begin{aligned} x(t) &= \sum_k \sum_m < x, T_{I_k}^* e_{m,k} > T_{I_k}^* e_{m,k}(t) \\ &= \sum_k \sum_m < T_{I_k} x, e_{m,k} > T_{I_k}^* e_{m,k}(t). \end{aligned} \tag{5}$$

Hence, the expansion coefficients can be computed by $< T_{I_k} x, e_{m,k} >$, which are simply the expansion coefficients of the smoothly periodized function $T_{I_k} x(t)$ with respect to the periodic orthonormal basis $\{e_{m,k}\}$. For the discrete version using the complex exponentials as $e_{m,k}$, we can readily use the FFT algorithm. For synthesis, applying T_{I_k} on both sides of Equation (5) with Equation (3) leads to

$$T_{I_k} x(t) = \sum_m < T_{I_k} x, e_{m,k} > \mathbf{1}_{I_k} e_{m,k}(t).$$

This is just the Fourier synthesis using the coefficients $< T_{I_k} x, e_{m,k} >$ and the basis $\mathbf{1}_{I_k} e_{m,k}(t)$. For the discrete case, we can simply apply the inverse FFT to these coefficients. Once we get $T_{I_k} x(t)$, then applying $T_{I_k}^*$ gives us $T_{I_k}^* T_{I_k} x(t) = P_{I_k} x(t)$ by Equation (3). Because $L^2(\mathbb{R}) = \bigoplus_{k\in\mathbb{Z}} P_{I_k} L^2(\mathbb{R})$, we can simply sum them to get

$$x(t) = \sum_{k\in\mathbb{Z}} P_{I_k} x(t).$$

As long as the action regions of the intervals $\{I_k\}$ do not interfere with each other, each split $\{I_k\}_{k\in\mathbb{Z}}$ of the time axis $\mathbb{R}$ leads to an orthonormal basis. This naturally leads to the concept of a *dictionary of orthonormal* bases or a *time-frequency dictionary.*

2.4. The local Fourier dictionary

Recursively partitioning the time axis into a binary tree structured set of intervals, Coifman, Meyer, and Wickerhauser[4,11,6] created a notion of the *dictionary of orthonormal bases* or *time-frequency dictionary.* Using the same idea, we can easily construct the *local Fourier dictionary.* Below, we briefly describe its construction.

Let us define $I_{0,k} = [k, k+1)$. Then, of course, we have $\mathbb{R} = \bigcup_{k\in\mathbb{Z}} I_{0,k}$. We then recursively split the intervals at their midpoints. After jth recursion, each interval is of the form $I_{j,k} \triangleq [k/2^j, (k+1)/2^j)$, $k \in \mathbb{Z}$. Clearly, for each $j \in \mathbb{Z}$, we have $\mathbb{R} = \bigcup_{k\in\mathbb{Z}} I_{j,k}$ and $I_{j,k} = I_{j+1,2k} \cup I_{j+1,2k+1}$. For all practical purposes, we only consider signals with finite support in time below. Suppose all the signals are contained in the support $[a, b)$. Then, we can also assume that all the signals are periodic with period $b - a$. This can be done by taking $[a, b)$ wide enough so that all the signals are zero around the borders $t = a$ and $t = b$, or applying the smooth orthogonal periodization $T_{[a,b)}$ to the evenly reflected signals at the borders. Finally, we can map $[a, b)$ to $I_{0,0} = [0, 1)$ by a simple change of variable $(t - a)/(b - a)$. We can now assume that all the signals are in $L^2(I_{0,0})$ and periodic with period 1. We also want to stop the recursion at certain level $J \in \mathbb{N}$ in practice. Then, the dyadic intervals $\{I_{j,k}\}$, $j = 0, 1, \ldots, J$, $k = 0, 1, \ldots, 2^j - 1$ can be readily arranged as a binary tree with the root node $I_{0,0}$. For each interval $I_{j,k}$, we can associate a subspace $\Omega_{j,k} = P_{I_{j,k}} L^2(I_{0,0})$ with $\Omega_{0,0} = L^2(I_{0,0})$. This set of tree-structured subspaces (with the localized complex exponentials as orthonormal basis functions at each subspace) is called the *local Fourier dictionary.* This dictionary contains a huge number (more than $2^{2^{J-1}}$) of orthonormal bases since each cover of $I_{0,0}$ by a subset of $\{I_{j,k}\}$ corresponds to one orthonormal basis, as we mentioned in the previous subsection. For discrete and finite dimensional versions of this dictionary, we start with a set of discrete signals sampled on the regular grid in the interval $I_{0,0}$ with n time samples. Then, this version of the dictionary consists of a redundant number (e.g., $n \log n$) of the basis vectors with the specific characters in scale, position, and frequency. These basis vectors are organized as

a binary tree in a *hierarchical* manner ranging from very localized spikes to global oscillations on $I_{0,0}$ with different frequencies. Therefore, pattern analysis and interpretation tasks using this dictionary become more intuitive than using the standard basis or the discrete Fourier basis on the interval $I_{0,0}$. Decomposing or reconstructing a signal using this dictionary is fast, e.g., $O(n[\log n]^2)$, thanks to the celebrated FFT algorithm.

How can we select a good basis out of such many possible bases? First of all, we need to know what purpose such a basis is used for. In other words, we need to define a numerical criterion to evaluate the effectiveness of a basis for one's purpose. Then, we can use the bottom-up procedure to efficiently search a good basis tailored to the specific application from a huge number of possible bases by optimizing that criterion. This divide-and-conquer (or split-and-merge) algorithm is called the *best-basis* algorithm. Therefore, this dictionary provides us with a flexible, hierarchical, and computationally efficient set of feature extractors at our disposal.

We note that the most straightforward extension of the above constructs to higher dimensions can be achieved easily by the appropriate tensor products, which will be used in Section 4. The local Fourier dictionary for images is particularly attractive because it contains the basis vectors with *oblique* oscillations, which the usual wavelet packets and local cosine/sine dictionaries cannot have.

We also note that F. Meyer and R. R. Coifman[12] recently developed *brushlets*—basis functions efficient for capturing oriented textured patterns. They also use Wickerhauser's smooth orthogonal periodization and its extension to the biorthogonal case using the bell functions of Matviyenko[7] to partition the two-dimensional frequency domain instead of the space domain. Therefore, the brushlet dictionary is a dual of the local Fourier dictionary.

In the following sections, we will describe (only a few) potential applications of the local Fourier dictionary.

3. THE BEST LOCAL TRANSLATION BASIS

Suppose we are given two signals. Suppose we want to make a correspondence (or find a match) between these two signals. If all the features in the first signal are shifted by the same amount and there is no other distortion, i.e., if the difference between these two signals is simply a global shift, then the usual Fourier transform solves our problem: simply dividing the Fourier transform of the first signal by that of the second gives us a phase factor $\mathrm{e}^{-2\pi i b\xi}$ where b is the amount of the global shift. This strategy does not work if individual features move differently. We need to consider the *local shifts* (or *time-dependent shifts*) to deal with this situation, which is not at all uncommon in practical situations. For example, geophysical acoustic waveforms clearly have such local shifts since the compressional wave (P wave) and the shear wave (S wave) propagate the media with different speeds depending on the elastic properties of the media. Vertical seismic profile (VSP)[13] is another example.

In order to explain how we can recover the local shifts and what is the good basis to do this job, let us first consider a simple case of the global shift. Let $x(t)$ and $y(t) = x(t-b) + \varepsilon(t)$ be the two signals to be matched, where $\varepsilon(t)$ is i.i.d. Gaussian noise. Let $\widehat{x}(\xi)$ and $\widehat{y}(\xi)$ be their Fourier representations. As we already mentioned, the global shift τ in the time domain is simply the multiplication of $\mathrm{e}^{-2\pi\mathrm{i}\tau\xi}$ in the Fourier domain, we can invoke the following optimization to get the estimate of b, the true global shift.

$$b^{\bigstar} = \arg\min_{\tau\in\mathbb{R}} \|\mathrm{e}^{-2\pi\mathrm{i}\tau\xi}\widehat{x}(\xi) - \widehat{y}(\xi)\|_2. \tag{6}$$

Now, let us consider the discrete version of the above equation. Let $\boldsymbol{x} = (x_0, x_1, \ldots, x_{n-1})^T$ be a vector representing time samples of the signal $x(t)$ on the uniform sampling grid with enough number of samples so that no aliasing happens. Similarly, let $\boldsymbol{y} \in \mathbb{R}^n$ be a vector of time samples of $y(t)$. Let $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$ be their discrete Fourier transforms (DFT). The global time-shift operator is now represented by the following matrix:

$$\begin{aligned} S_\tau &= \mathrm{diag}(\omega_n^{-0\cdot\tau}, \omega_n^{-1\cdot\tau}, \ldots, \omega_n^{-(n-1)\cdot\tau}) \\ &= \mathrm{diag}(\omega_n^{-0\cdot\tau}, \omega_n^{-1\cdot\tau}, \ldots, \omega_n^{-n/2\cdot\tau}, \omega_n^{(n/2-1)\cdot\tau}, \ldots, \omega_n^{1\cdot\tau}), \end{aligned}$$

where $\omega_n = \mathrm{e}^{2\pi\mathrm{i}/n}$. Hence Equation (6) can be rewritten as

$$b^{\bigstar} = \arg \min_{-n/2\le\tau<n/2} \|S_\tau\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{y}}\|_2. \tag{7}$$

Because

$$\|S_\tau\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{y}}\|_2^2 = (S_\tau\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{y}})^*(S_\tau\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{y}}) = \|\widehat{\boldsymbol{x}}\|_2^2 + \|\widehat{\boldsymbol{y}}\|_2^2 - 2\mathrm{Re} < S_\tau\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}} >,$$

$\|\widehat{\boldsymbol{x}}\|_2 = \|\boldsymbol{x}\|_2$, and $\|\widehat{\boldsymbol{y}}\|_2 = \|\boldsymbol{y}\|_2$ by the Plancherel Theorem, the optimization in (7) is equivalent to

$$b^\star = \arg \max_{-n/2 \le \tau < n/2} \mathrm{Re} < S_\tau \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}} > .$$

We can use this optimization procedure locally at each subspace $\Omega_{j,k}$ of the local Fourier dictionary to recover the local shifts that best matches the two signals. Let us define a functional to be maximized at $\Omega_{j,k}$:

$$\mathcal{C}_{j,k}(\tau; \boldsymbol{x}, \boldsymbol{y}) \triangleq \mathrm{Re} < S_\tau^j \widehat{\boldsymbol{x}}_{j,k}, \widehat{\boldsymbol{y}}_{j,k} >, \tag{8}$$

where S_τ^j is a shift operator at level j subspaces (it does not depend on k if we use the homogeneous dyadic intervals $\{I_{j,k}\}$), $\widehat{\boldsymbol{x}}_{j,k}$ and $\widehat{\boldsymbol{y}}_{j,k}$ are the local Fourier coefficients of $\boldsymbol{x}$ and $\boldsymbol{y}$ at $\Omega_{j,k}$, respectively.

We now propose the following algorithm for recovering the local shifts.

Step 0: Decompose given two signals $\boldsymbol{x}$ and $\boldsymbol{y}$ into the local Fourier dictionary of depth $J \le \log_2 n$ (i.e., with levels $0, 1, \ldots, J$).

Step 1: At each subspace of the tree, $\Omega_{j,k}$, compute $\mathcal{C}^\star_{j,k} = \max \mathcal{C}_{j,k}(\tau_{j,k}; \boldsymbol{x}, \boldsymbol{y})$.

Step 2: Invoke the best basis algorithm to select the best subspaces $\Omega^\star_{j,k}$. This can be done by setting $\Omega^\star_{J,k} = \Omega_{J,k}$ at the bottom level, and then pruning the tree of the subspaces recursively from $j = J-1$ to $j = 0$ via

$$\Omega^\star_{j,k} = \begin{cases} \Omega_{j,k} & \text{if } \mathcal{C}^\star_{j,k} \ge \mathcal{C}^\star_{j+1,2k} + \mathcal{C}^\star_{j+1,2k+1}, \\ \Omega^\star_{j+1,2k} \oplus \Omega^\star_{j+1,2k+1} & \text{otherwise.} \end{cases}$$

We call the basis thus selected (i.e., the basis corresponding to the space $\Omega^\star_{0,0}$) the *best local translation basis* (BLTB). We also record the best local shifts corresponding to the best subspaces in the BLTB $\Omega^\star_{0,0}$, which can be used to undo the local shifts for matching these two signals.

Figure 6 demonstrates the use of BLTB using a synthetic example. Here, we first generated a signal of length 1024 consisting of a Gaussian-windowed cosine function in the first half of the time axis and a Gaussian-windowed sine function in the second half. Then we generated a second signal by shifting the feature in the first half by 100 units to the right and the one in the second half by 50 units to the left. Finally, the white Gaussian noise with $N(0,1)$ was added to both of them. The first signal is considered as a template signal and we want to locally shift this signal to match the second signal. For maximization of $\mathcal{C}_{j,k}$, we used Brent's method.[14] The local shifts recovered by the BLTB method were 98.1386 and -50.3255. The matched signals after undoing local shifts are displayed in Figure 6(b).

Figure 7 shows the results applied to the real geophysical acoustic waveforms propagated through the sandstone layers located at different depths. One can observe that most of the local phases are matched after undoing the estimated local shifts to the template signal.

4. STATISTICS OF NATURAL SCENES VIA LOCAL FOURIER DICTIONARY

In this section, we report the preliminary results of our experiments on the analysis of statistics of natural scene images, which is a joint work with Brons Larson at UCD. The more complete report[15] is in preparation.

Statistical analysis of natural scene images has recently drawn considerable attention particularly in the field of neuroscience such as Field,[16] Olshausen & Field,[17,18] Bell & Sejnowski,[19] van Hateren and van der Schaaf,[20] to name a few. Their main motivation is to understand the receptive field properties of simple cells in the mammalian primary visual cortex by analyzing the statistics of natural scenes. Field[16] suggested that neurons with line and edge selectivities in primary visual cortex may provide *sparse* representation of natural scenes. This may imply that mammals exploit the sparsity for image representations in their brain. Many neuroscientists have tried to answer the following reverse proposition, which is also very interesting: Immersed in the natural environment, whether the receptive fields of simple cells of mammals autonomously form edge or line detectors. If one can demonstrate this, it may be a convincing argument about why mammals have edge detectors and why mammals are exploiting sparse and efficient representations of natural scenes. This proposal has generated many interesting algorithms and numerical

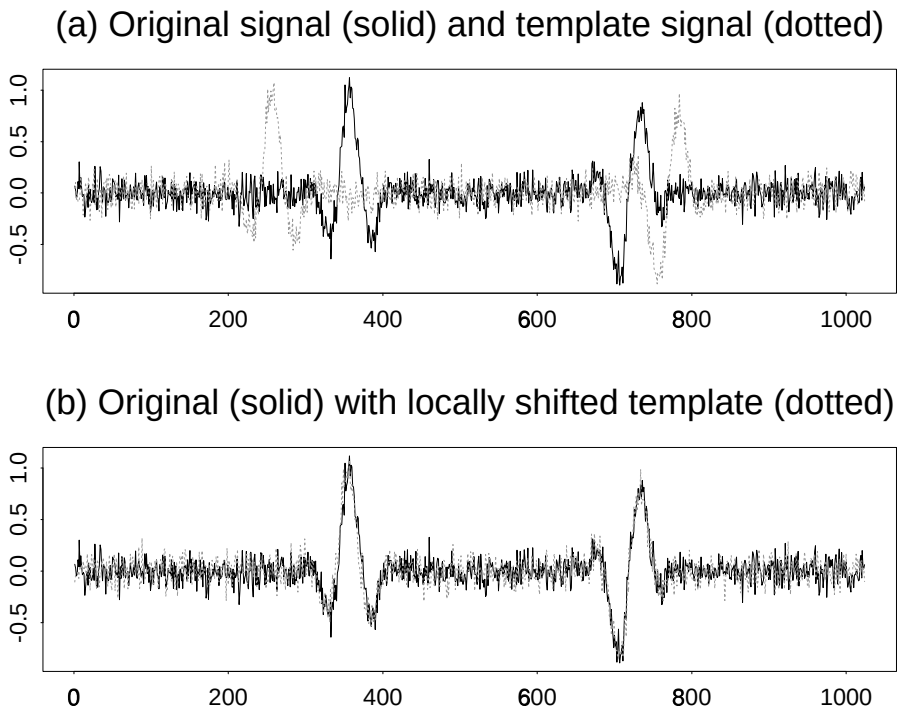


Figure 6. Before and after the match obtained by the BLTB for the synthetic signals.

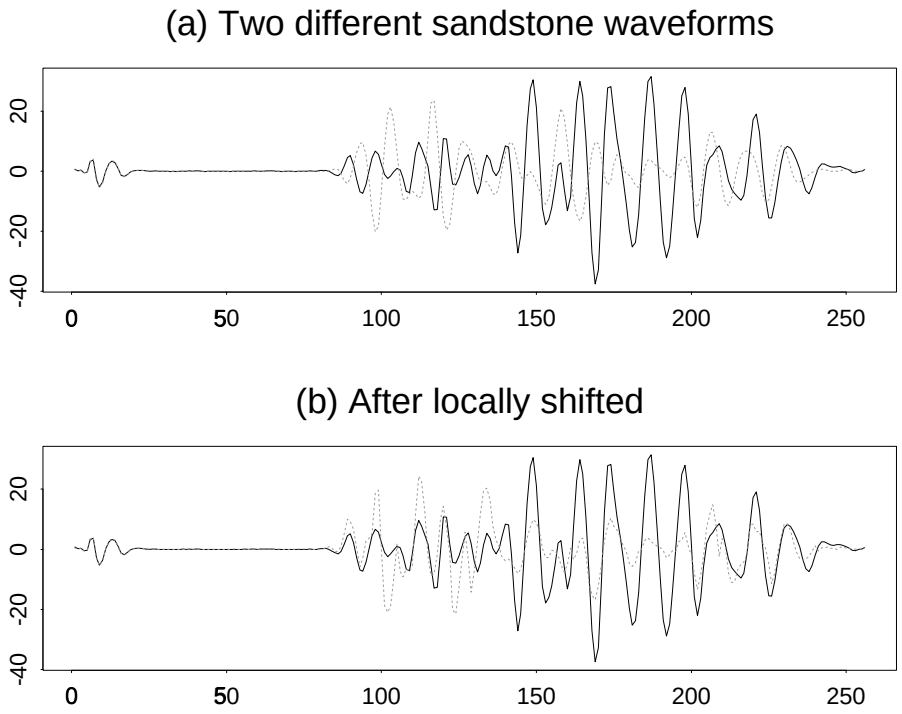


Figure 7. Before and after the match obtained by the BLTB. These are geophysical acoustic waveforms propagated through two different sandstone layers.

experiments.[17–20] All of these approaches essentially try to find the basis functions from some overcomplete set of bases by optimizing either *sparsity* or *statistical independence* (among the expansion coefficients) using the neural networks or learning algorithms. In fact, they found that the estimated basis functions all resemble Gabor functions or oriented DOG filters. On one hand, their algorithms are truly self-organizing since they can build such basis functions completely from scratch. On the other hand, their computational cost prevents us from conducting experiments on large image patches. Because of the high computational cost, most of their experiments used small image patches (e.g., 12×12 or 16×16 pixels) extracted from a set of natural scene images.

Mathematically, these approaches can be formulated as follows. Let X be a random vector representing an image patch extracted from a collection of natural scene images. Let $\mathbf{\Phi}$ be a matrix whose columns are basis vectors to be determined. Then, their approaches can be written as

$$\min_{\mathbf{\Phi} \in \mathcal{D}} E_X \left\{ \|X - \mathbf{\Phi} Y\|_2^2 + \lambda S(Y) \right\}, \tag{9}$$

where E_X is the expectation operator with respect to X, $Y = Y(X)$ is a vector of the expansion coefficients of an input X relative to the basis $\mathbf{\Phi}$, and $\lambda > 0$ is a weight controlling the relative importance of the fidelity term and the term $S(Y)$ that is the focal point of various discussions, i.e., a measure of the effectiveness of the basis $\mathbf{\Phi}$. Olshausen and Field[17,18] used the sparsity constraint

$$S(Y) = \sum_i \log(1 + Y_i^2),$$

whereas Bell and Sejnowski[19] used the 'infomax' constraint

$$S(Y) = -H(g(Y)),$$

where H is the Shannon entropy, and $g(\cdot)$ is a nonlinear sigmoidal function. (We use the minimization scheme here so, we put the minus sign in front of H, i.e., minimization of *negentropy*.)

Donoho[21] recently suggested that the concept of the Sparse Component Analysis (SCA) and conducted a detailed mathematical analysis, where he argued that sparsity and overcompleteness may be more plausible on biological grounds and are more important for practical data compression purposes than statistical independence.

In the mean time, from a completely different motivation (stochastic modeling of a class of similar images), the author developed an algorithm of computing the *least statistically-dependent basis* (LSDB) selected from time-frequency dictionaries.[22,23] This can be viewed as a dictionary version of the independent component analysis (ICA).[24,25]

This series of works has motivated us to study the following questions:

1. To form edge or line detectors, sparsity is more important than statistical independence?
2. What is the effect of the sizes of the image patches used?
3. What is the effect of orthonormality?
4. What is the effect of overcompleteness?
5. What is the effect of orientation selectivities of basis functions?

Our plan is to examine these questions using the same natural scene images used by Olshausen and Field[17,18] and using various basis dictionaries such as local Fourier, brushlets,[12] edgetlets,[26] local cosine, wavelet packets, which we will report in the near future.[15] In this paper, we will show preliminary results using the local Fourier dictionary. This dictionary is particularly interesting since it provides oriented feature extractors, which the wavelet packets and local cosine/sine dictionaries cannot provide, and such oriented pattern detectors are prevalent in human vision systems.[27]

Now we describe our experiments in this paper more precisely. Our setting is the following.

1. Set $\mathcal{D}$ = 2D Local Fourier Dictionary.

2. Each $\mathbf{\Phi}$ examined in the optimization is a complete orthonormal basis selected from $\mathcal{D}$ by the joint best basis (JBB) algorithm[11,28] (with different optimization criteria from the original proposals). This selection makes the first term of the optimization (9) always 0.

3. Set $S(Y) = \|Y\|_1$ (ℓ^1 norm) for the sparsity constraint, and $S(Y) = \sum_i H(Y_i)$ (sum of the differential entropy of the individual coordinates) for the independence constraint.

4. Image patch size ranges from 16×16 to 128×128.

The reason why $S(Y) = \|Y\|_1$ measures sparsity of the sequence of the coefficients Y is the following.[21] Let $\|Y\|_p = (\sum_i |Y_i|^p)^{1/p}$ be the ℓ^p norm. Then, for p, $0 < p \leq 1$, this is a robust version of the true sparsity measure $\|Y\|_0 = \#\{i : Y_i \neq 0\}$ since $\lim_{p \downarrow 0} \|Y\|_p^p = \|Y\|_0$. In this report, we only examine $p = 1$.

The reason why $S(Y) = \sum_i H(Y_i)$ measures the statistical dependence among the coordinates Y_i is the following.[22,23] The mutual information of Y is defined as

$$I(Y) = I(Y_1, \ldots, Y_n) = \int f_Y(y_1, \ldots, y_n) \log \frac{f_Y(y_1, \ldots, y_n)}{\prod_{i=1}^n f_{Y_i}(y_i)} \, dy_1 \cdots dy_n = -H(Y) + \sum_{i=1}^n H(Y_i),$$

where f_Y is a joint probability density function (pdf) of Y, and f_{Y_i} is a marginal pdf by integrating the joint pdf with respect to all but Y_i. The mutual information $I(Y)$ measures statistical dependence among the coordinates Y_i: the more dependent Y_i's are, the larger $I(Y)$ gets, $I(Y)$ is always nonnegative, and $I(Y) = 0$ if and only if Y_i's are statistically independent. Now, by simple computation, as long as the matrix $\mathbf{\Phi}$ is selected from $\mathrm{SL}(n, \mathbb{C})$, which includes both orthonormal and appropriately normalized biorthogonal bases, the change of the coordinates by $\mathbf{\Phi}$ preserves the joint entropy $H(Y)$, that is, $H(Y) = H(X)$ for any such $\mathbf{\Phi}$. Therefore, minimizing $I(Y)$ is equivalent to minimizing $\sum_i^n H(Y_i)$ as long as $\mathbf{\Phi} \in \mathrm{SL}(n, \mathbb{C})$. In practice, we need to use a certain density estimation method such as average shifted histograms for two-dimensional data since the local Fourier dictionary generates a table of complex-valued coefficients. The computational cost of our approach regardless of using the sparsity constraint or the independence constraint is less expensive than the neural network based optimization; indeed, it only costs $O(n[\log n]^2)$.

Some of the results of our experiments are shown in Figures 8 and 9. These are the most energetic 128 basis vectors (each basis vector is complex-valued) for each case.

We observe the following from these figures.

1. The sparsity constraint tends to split the image patches into a finer segments (i.e., local Fourier basis functions of size 8×8 pixels) than the independence constraint.

2. The independence constraint prefers the basis functions with larger support (32×32 pixels). The patches with 128×128 (not shown) and 64×64 lead to the basis functions whose supports are 32×32 pixels while the patches with 32×32 (not shown) and 16×16 were not split at all.

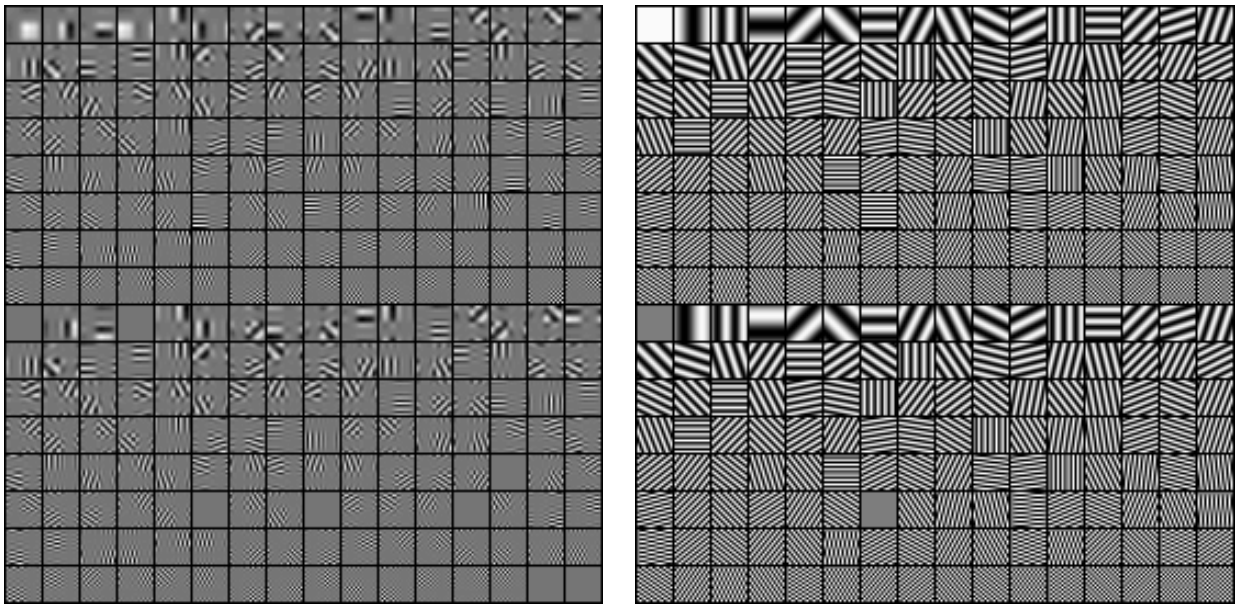
3. The basis functions with rather high frequency nature are included in the list since we impose to select the complete basis.

From this set of experiments, in order to form small scale edge or line detectors, the sparsity seems more important than the statistical independence as Donoho suggested, mainly because experiments on the larger image patches (e.g., 64×64, 128×128) consistently selected small scale basis vectors (8×8). On the other hand, we need more experiments to answer the above questions more definitely by examining various different dictionaries, and using robust entropy estimators for the statistical independence.

5. DISCUSSION

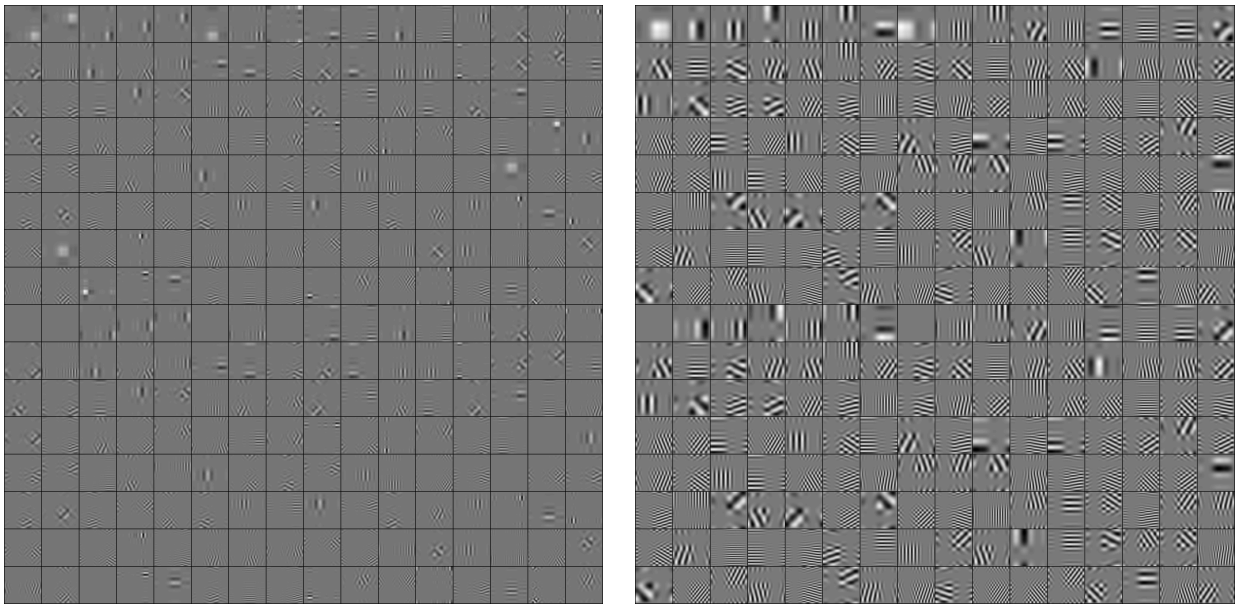
5.1. Locally analytic signals

Once we select a basis from the local Fourier dictionary by a certain criterion, it is straightforward to construct the *analytic signals* locally at each segment by computing the Hilbert transform. This may also be combined with the latest method to factor a signal into the well-behaved phase and amplitude components via the Blaschke products being developed by Coifman and Nahon.[29] Such methods are also important for robust estimation of local shifts in signals discussed in Section 3 since the local shifts of highly oscillatory features create many local maxima in Equation (8), which make optimization task much tougher.



(a) JBB–ℓ^1 norm minimization (b) LSDB

Figure 8. The top 128 basis vectors for 16×16 patches. The upper and lower halves of each matrix show the real and imaginary components of the basis vectors, respectively.



(a) JBB–ℓ^1 norm minimization (b) LSDB

Figure 9. The top 128 basis vectors for 64×64 patches.

5.2. Local deformation, signal matching, and data compression

The local shifts discussed in Section 3 is the simplest kind of general local deformation mechanisms including localized rotation and nonlinear warping. Recovering localized versions of shifts, rotations, and other deformations has at least two important implications. First, this recovery leads to *image registration* giving a clue to the *change* between the two images, which may in turn deepen our understanding of the objects under investigation. Second, this may lead to compact data representation. The idea here is to represent a sequence of signals or images by a template with the deformation parameters. The template itself can also be compressible by the time-frequency dictionary. Then, the whole dataset such as the geophysical acoustic waveforms used by the author[30] can be represented by a single template signal (the mean signal is one of the candidates for such a template) and all the deformation parameters that make the template signal look like the other signals. If the dataset consists of completely different signals, then this strategy will not work: too many bits are required to describe the deformation parameters, which will not help compression. But if the dataset consists of similar but slightly different signals and if these differences can be captured by the deformation parameters compactly, then the gain might be huge. This idea may be viewed as a nonlinear generalization of the Karhunen-Loève transform.

5.3. Further extension

Matviyenko[7] developed biorthogonal window functions for the local cosine/sine bases that have better time-frequency localization property while maintaining stability and efficiency in computations compared to the Gabor bases. Bennett[31] recently extended the 2D local cosine/sine and wavelet packet dictionaries by splitting an image into a *forest* of dyadic segments which includes a quadtree and allows *rectangular* partitioning of the input image. He also developed a fast best-basis search algorithm from such a forest of subspaces. These generalizations and extensions are yet to be investigated under the local Fourier context.

ACKNOWLEDGMENTS

The author would like to thank Brons Larson for conducting numerical experiments on Section 4, and Bruno Olshausen for stimulating discussions and for providing the digitized natural scene images. This research was supported in part by research contract from SRI International (via AFOSR F49520-98-C-0041).

REFERENCES

1. R. Balian, "Une principle d'incertitude fort en théorie du signal ou en mécanique quantique," *C. R. Acad. Sci. Paris, Serié 2* **292**, pp. 1357–1362, 1981.
2. I. Daubechies, S. Jaffard, and J. L. Journé, "A simple Wilson orthonormal basis with exponential decay," *SIAM J. Math. Anal.* **22**, pp. 554–572, 1991.
3. H. S. Malvar, "The LOT: transform coding without blocking effects," *IEEE Trans. Acoust., Speech, Signal Process.* **37**, pp. 553–559, 1989.
4. R. R. Coifman and Y. Meyer, "Remarques sur l'analyse de Fourier à fenêtre," *Comptes Rendus Acad. Sci. Paris, Série I* **312**, pp. 259–261, 1991.
5. M. V. Wickerhauser, "Smooth localized orthonormal bases," *C. R. Acad. Sci. Paris, Série I* **316**, pp. 423–427, 1993.
6. M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A K Peters, Ltd., Wellesley, MA, 1994. with diskette.
7. G. Matviyenko, "Optimized local trigonometric bases," *Appl. Comput. Harmonic Anal.* **3**, pp. 301–323, 1996.
8. P. Auscher, G. Weiss, and M. V. Wickerhauser, "Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets," in *Wavelets: A Tutorial in Theory and Applications*, C. K. Chui, ed., pp. 237–256, Academic Press, 1992.
9. X. Fang and E. Séré, "Adapted multiple folding local trigonometric transforms and wavelet packets," *Appl. Comput. Harmonic Anal.* **1**, pp. 169–179, 1994.
10. L. F. Villemoes, "Adapted bases of time-frequency local cosines," *Applied and Computational Harmonic Analysis* , 1999. submitted.
11. R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory* **38**, pp. 713–719, Mar. 1992.

12. F. G. Meyer and R. R. Coifman, "Brushlets: a tool for directional image analysis and image compression," *Appl. Comput. Harmonic Anal.* **4**, pp. 147–187, 1997.
13. W. M. Telford, L. P. Geldart, and R. E. Sheriff, *Applied Geophysics*, Cambridge Univ. Press, second ed., 1990.
14. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge Univ. Press, second ed., 1992.
15. N. Saito and B. Larson, "Analysis of natural images via local Fourier dictionary: Sparsity vs Independence," tech. rep., Dept. Math., University of California, Davis, 1999. in preparation.
16. D. J. Field, "What is the goal of sensory coding?," *Neural Computation* **6**, pp. 559–601, 1994.
17. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**, pp. 607–609, 1996.
18. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Research* , pp. 3311–3325, 1997.
19. A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research* **37**, pp. 3327–3338, 1997.
20. J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc. Royal Soc. London, Ser. B* **265**, pp. 359–366, 1998.
21. D. L. Donoho, "Sparse components of images and optimal atomic decompositions," tech. rep., Dept. Statistics, Stanford University, 1998.
22. N. Saito, "Least statistically-dependent basis and its application to image modeling," in *Wavelet Applications in Signal and Image Processing VI*, A. F. Laine, M. A. Unser, and A. Aldroubi, eds., vol. Proc. SPIE 3458, pp. 24–37, 1998.
23. N. Saito, "The least statistically-dependent basis and its applications," in *Proc. 32nd Asilomar Conference on Signals, Systems, and Computers*, pp. 732–736, IEEE, 1998.
24. P. Comon, "Independent component analysis, a new concept?," *Signal Processing* **36**, pp. 287–314, 1994.
25. A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation* **7**, pp. 1129–1159, 1995.
26. A. Averbuch, R. Coifman, D. Donoho, M. Israeli, and J. Waldén, "The pseudopolar FFT and its applications," tech. rep., Dept. Computer Science, Yale University, 1999.
27. D. H. Hubel, *Eye, Brain, and Vision*, Scientific American Library, 1995.
28. M. V. Wickerhauser, "Fast approximate factor analysis," in *Curves and Surfaces in Computer Vision and Graphics II*, pp. 23–32, Oct. 1991. Proc. SPIE 1610.
29. R. R. Coifman and M. Nahon, "Personal communication," 1999.
30. N. Saito and R. R. Coifman, "Extraction of geological information from acoustic well-logging waveforms using time-frequency wavelets," *Geophysics* **62**(6), pp. 1921–1930, 1997.
31. N. N. Bennett, "Fast algorithms for best anisotropic Walsh bases and relatives," *Applied and Computational Harmonic Analysis* , 1998. submitted.