

1 Single Scale Adaptive Wavelet Tight Frames

As a first step of the plan, we plan to construct wavelet tight frames in a manner that is adapted to the given data. The resulting model will be a building block of the multi-layer structure developed in later sections. We begin with a brief review of existing construction of wavelet tight frames.

1.1 Wavelet Tight Frames

In this subsection, we give a very brief introduction to wavelet tight frames, for a detailed introduction to this subject, the readers may refer to [1].

Let \mathcal{H} be a Hilbert space, we are mainly concerned with the case when $\mathcal{H} = L_p(\mathbb{R}^d)$. a system $X \subset \mathcal{H}$ is called a tight frames is

$$\|f\|_2^2 = \sum_{x \in X} |\langle f, x \rangle|^2, \quad \text{for any } f \in \mathcal{H}$$

There are two operators that are associated with a tight frame, one is the analysis operator defined as

$$W : f \in \mathcal{H} \rightarrow \{\langle x, f \rangle\}_{x \in X} \in l_2(\mathbb{N})$$

and the other is the synthesis operator W^T , which is the adjoint operator of the analysis operator defined as

$$W^T : \{a_n\} \in l_2(\mathbb{N}) \rightarrow \sum_{x \in X} a_n x \in \mathcal{H}.$$

The system X is called a tight frame if and only if $W^T W = I$, where $I : \mathcal{H} \rightarrow \mathcal{H}$ is the identity operator. In other words, given a tight frame X , we have the following canonical expansion:

$$f = \sum_{x \in X} \langle x, f \rangle x, \quad \text{for any } f \in \mathcal{H}$$

The sequence $Wf := \{\langle f, x \rangle\}_{x \in X}$ are called the canonical tight frame coefficients. Thus, tight frames are often viewed as generalizations of orthonormal basis. In fact, a tight frame X is an orthonormal basis for \mathcal{H} if and only if $\|x\| = 1, \forall x \in X$.

In signal processing applications, one widely used class of tight frames is the wavelet tight frames. The construction starts with a finite set of generators $\Psi := \{\psi^1, \dots, \psi^m\}$. Then consider the affine system defined by the shifts and dilations of the generators:

$$X(\Psi) = \{M^{j/2} \psi^l(M^j \cdot -k), 1 \leq l \leq m, j, k \in \mathbb{Z}, M \in \mathbb{Z}^+\}.$$

The affine system $X(\Psi)$ is called a wavelet tight frame if it is a tight frame satisfying

$$f = \sum_{x \in X(\Psi)} \langle f, x \rangle x, \forall f \in \mathcal{H}.$$

Wavelet tight frames used in practice are usually constructed from multi-resolution analysis(MRA). This is because a MRA structure is crucial for fast decomposition and reconstruction algorithms. The MRA construction usually starts with a compactly supported scaling function ϕ with a refinement mask a_0 satisfying

$$\hat{\phi}(M \cdot) = \hat{a}_0 \hat{\phi}.$$

where $\hat{\phi}$ is the Fourier transform of ϕ , and \hat{a}_0 is the discrete Fourier series defined as $\hat{a}_0(\omega) := \sum_{k \in \mathbb{Z}} a_0(k) e^{-ik\omega}$ and $\hat{a}_0(0) = 1$. After obtaining the scaling function, the next step is to find an appropriate set of filters $\{a_1, \dots, a_m\}$ and define the set of functions called framelets $\Psi = \{\psi_1, \dots, \psi_m\}$ by

$$\hat{\psi}_i(M \cdot) = \hat{a}_i \hat{\phi}, i = 1, \dots, m$$

such that the affine system $X(\Psi)$ forms a wavelet tight frame. It is natural to ask when does such a system form a wavelet tight frame. Sufficient and necessary conditions are given by the so called Unitary Extension Principle(UEP). There are different versions of UEP principles, we are only concerned with

one version that is associated with discrete wavelet tight frames, which we will state after a description of the decomposition and reconstruction operations. For a survey of UEP, the reader may refer to [1].

Given a filter $a \in l_2(\mathbb{Z})$, the discrete decomposition and reconstruction transform are defined in the following way. Define the one dimensional down-sampling and up-sampling operator:

$$\begin{aligned} [v \downarrow M](n) &:= v(Mn), \quad n \in \mathbb{Z} \\ [v \uparrow](Mn) &:= v(n), \quad n \in \mathbb{Z} \end{aligned}$$

where M a positive integer denoting the down-sampling or up-sampling factor. Down-sampling and up-sampling operators in higher dimensions are carried out by performing one dimensional operators along each dimension.

Define the linear convolution operator $S_a : l_2(\mathbb{Z}) \rightarrow l_2(\mathbb{Z})$ by

$$[S_a(v)](n) := [a * v](n) = \sum_{k \in \mathbb{Z}} (a(-\cdot) * v \downarrow)M, \forall v \in l_2(\mathbb{Z})$$

For a set of filters $\{a_i\}_{i=1}^m \subset l_2(\mathbb{Z})$, we define its analysis operator W by

$$W = [S_{a_1(-\cdot)}, S_{a_2(-\cdot)}, \dots, S_{a_m(-\cdot)}]^T.$$

Its synthesis operator is defined as the transpose of W :

$$W^T = [S_{a_1(\cdot)}, S_{a_2(\cdot)}, \dots, S_{a_m(\cdot)}].$$

Now we may state the UEP for this situation.

Proposition 1 (cite). *Let a_1, \dots, a_m be finitely supported filters, the following are equivalent:*

1. $W_a^T W_a = I$

2. for all $\omega \in [0, 1)^d \cup M^{-1}\mathbb{Z}^d$,

$$\sum_{i=1}^m \hat{a}_i \overline{\hat{a}_i(\xi + 2\pi\omega)} = \delta(\omega);$$

3. for all $k, \gamma \in \mathbb{Z}^d$,

$$\sum_{i=1}^m \sum_{n \in \mathbb{Z}^d} \overline{a_i(k + Mn + \gamma)} a_i(Mn + \gamma) = M^{-d} \delta(k).$$

In particular, if the data are real numbers and no down-sampling is performed, then $W^T W = I$ is equivalent to

$$\sum_{i=1}^m \sum_{n \in \mathbb{Z}^d} a_i(k + n) a_i(n) = \delta_k, \forall k \in \mathbb{Z}^d. \quad (1)$$

The linear B-spline wavelet tight frame used in many image restoration tasks is constructed via the UEP. Its associated tree filters are :

$$a_1 = \frac{1}{4}(1, 2, 1)^T; \quad a_2 = \frac{\sqrt{2}}{4}(1, 0, -1)^T; \quad a_3 = \frac{1}{4}(-1, 2, -1)^T.$$

Once the 1D filter $\{a_i\}_{i=1}^m$ for generating a tight frame for $l_2(\mathbb{Z})$ is constructed, the traditional way of generating higher dimensional tight frames is to use tensor products of 1D filters. But in this paper, we are going to construct 2D filters directly.

1.2 Adaptive Construction

We use shift-invariant systems because we accept the premise that at a proper scale, the statistical properties of image patches are translational invariant. Let W_a be the matrix form of the analysis operator generated by the filters $\{a_i\}_{i=1}^m$, and let W_a^T be the matrix form of the synthesis operator. Define \mathcal{C} to be the set of filters that satisfy the full UEP condition, that is $\mathcal{C} = \{\{a_i\}_{i=1}^m : W_a^T W_a = I\}$. Sometimes in image processing tasks, we can tolerance a little bit of loss of information, so we also consider the filters that approximately satisfy the full UEP condition within error δ , $\mathcal{C}_\delta = \{\{a_i\}_{i=1}^m : \|W_a^T W_a - I\|_* \leq \delta\}$, where $\|\cdot\|_*$ means the operator norm.

For a given class of images and a specific task at hand, there are infinitely many discrete wavelet tight frames to choose from. Although they all provide perfect reconstruction of the input signal, some of them may provide sparser representations than the rest. Therefore, we propose a model that provides the sparsest representation of the class of signals at hand. That is, the filters we are looking for is the minimizer of the following optimization program:

$$\begin{aligned} & \min_{a_1, \dots, a_m} \sum_{i=1}^m \Phi(v_i) \\ \text{subject to } & v_i = a_i(-) * x, \quad i = 1, \dots, m \\ & \{a_i\}_{i=1}^m \in \mathcal{C} \end{aligned} \quad (2)$$

where $\Phi(x; a_i)$ is a sparsity inducing function, it can be chosen as, for example, the l_1 norm or l_0 "norm" or the Huber loss. We will focus the l_1 norm in the numerical illustrations given later as we find it attractive in terms of quality. That is,

$$\begin{aligned} & \min_{a_1, \dots, a_m} \sum_{i=1}^m \|v_i\|_1 \\ \text{subject to } & v_i = a_i(-) * x, \quad i = 1, \dots, m \\ & \{a_i\}_{i=1}^m \in \mathcal{C} \end{aligned} \quad (3)$$

As this optimization program is not convex, a local minimum is we can hope at best. We have no guarantee of the global optimality of the solution. Surprisingly, the local minimum obtained by interior point method are of very good quality in numerical experiments.

In some image processing tasks, such as pattern recognition, a small deviation from the perfect reconstruction filters is allowed. In that case, we consider

$$\begin{aligned} & \min_{a_1, \dots, a_m} \sum_{i=1}^m \|v_i\|_1 \\ \text{subject to } & v_i = a_i(-) * x, \quad i = 1, \dots, m \\ & \{a_i\}_{i=1}^m \in \mathcal{C}_\delta \end{aligned} \quad (4)$$

In this case, we can get an approximate solution to this model by penalty method, in fact, we have the following proposition:

Proposition 2. *Let $\{a^*\}_{i=1}^m$ be the minimizer of (4), let $\{\hat{a}_i\}_{i=1}^m$ be the solution to the following program:*

$$\begin{aligned} & \min_{a_1, \dots, a_m} \sum_{i=1}^m \|v_i\|_1 + \eta \|y - \sum_j a_j(-) * a_j * y\|_2 \\ \text{subject to } & v_i = a_i(-) * x, \quad i = 1, \dots, m \end{aligned} \quad (5)$$

where y is Gaussian random variables of sufficient length. Then for appropriate choice of η , there exist a constant $c \ll 1$, such that

$$(1 - c) \leq \sup_i \frac{\|a_i^* - \hat{a}_i\|_\infty}{\|a_i^*\|} \leq (1 + c)$$

with high probability? or asymptotic results?

Optimization program (5) is relatively easier to solve, we call it the sampling version of model (4). In practice, η is chosen based on our tolerance of deviation from perfection reconstruction filters. By letting η goes to ∞ , we recover (4). As numerical algorithms is not the focus of this paper, we omit the details of implementation here. A brief description of numerical algorithms is included in the appendix, numerical illustrations are reproducible through the MATLAB code online at [\[\]](#). So far, the models are based on the premise that the signals can be written as a sparse linear combination of translational invariant wavelets. Even if the signal is really generated this way, inevitably there will be perturbations or noises added to the coefficients. Hence, it is helpful to consider a variant of model (4):

$$\begin{aligned} & \min_{a_1, \dots, a_m, v_1, \dots, v_m} \sum_{i=1}^m \|v_i\|_1 + \lambda \sum_{i=1}^m \|v_i - a_i(-\cdot) * x\|_2^2 \\ & \text{subject to } \{a_i\}_{i=1}^m \in \mathcal{C}_\delta \end{aligned} \quad (6)$$

The sampling version can be written correspondingly.

The key feature of this variant, is that unlike the previous model, the wavelet coefficients $\{v_i\}$ is no longer linear dependent on x given the filters $\{a_i\}_{i=1}^m$ yet still can be computed efficiently. Indeed, given the learned filters $\{a_i\}_{i=1}^m$ and the new input signal x , the coefficients is obtained by

$$\min_{v_1, \dots, v_m} \sum_{i=1}^m \|v_i\|_1 + \lambda \sum_{i=1}^m \|v_i - a_i(-\cdot) * x\|_2^2,$$

the solution of which we can readily write explicitly:

$$v_i = \mathcal{T}_{1/2\lambda}(a_i(-\cdot) * x), \quad i = 1, \dots, m$$

where $\mathcal{T} : \mathbb{R} \mapsto \mathbb{R}$ is the soft-thresholding operator defined by

$$\mathcal{T}_a(x) = \begin{cases} (|x| - a)\text{sign}(x), & \text{if } |x| > a \\ 0, & \text{otherwise} \end{cases}.$$

When \mathcal{T} operates on a vector, it operates on each component of the vector.

A special case of this variant, when the support of a_i is of size $r \times r$ and $m = r^2$ and the filters are orthogonal to each other is proposed independently in [\[\]](#), and local solution is found by iteratively solving the $\{v_i\}_{i=1}^m$ and $\{a_i\}_{i=1}^m$.

To this end, we have introduced the construction of adaptive wavelet tight frames for one layer. When plugged into applications, we found the filters learned are of high quality, the results produced are comparable to those obtained by the dictionary learning paradigm, numerical illustrations are given in a later section, this is more or less expected. In addition, we observed some quite unexpected and intriguing phenomena that we would like to share with the readers. **quadratic term needs to be verified**

1.3 Some Intriguing Observations

A unique low frequency filter. The most commonly used wavelets and wavelet tight frames constructed using MRA has only one low frequency filter. As a result, it enables a fast multi-level decomposition and reconstruction algorithm, the architecture is illustrated in Figure. However, the UEP does not distinguish between low frequency and high frequency filters, the concept of a unique low frequency filter is more of mathematical considerations rather than practical need. Indeed, one need not be bounded by this restriction, and can use multiple low frequency filter. From a practitioner's point of view, as long as the filters provide sparse representations of the signal, we do not care much if there is one or multiple low frequency filters. Indeed, in proposing the previous models, we adopt this practitioner's point of view, hence no constraint on the number of low frequency filters is added in the models. One would expect the resulting filters would be of diverse frequencies.

Surprisingly, we found that, on many datasets, the filters learned contains exactly one low frequency filter. That is, among the m filters trained, one and only one of them sums up to 1, the rest each sum up to 0 or very close to 0, no intermediate values are presented. This phenomenon is stable regardless of the value of m and support size of the filters. We observed this phenomenon on many datasets, including the Yale-face dataset, caltech-101 dataset, the fingerprint dataset, and a large number of randomly chosen

natural photos from the Internet. However, this phenomenon is not entirely universal, we observed multiple low frequency filters on the dataset mnist. Apparently, this phenomenon has to do with the characteristics of the particular class of images and reflects the structure of the function space of all "natural images". A natural question is: for what class of images, can we observe such a phenomenon?

What this phenomenon means is that: the filters that give rise to the sparsest representation of image signals happen to have only one low frequency filter, which coincides with the traditional wavelets we have been using successfully in the past two decades. Some classes of images, exhibit sparse representations in these basis or tight frames, some don't. Notably, the majority the class of photos we conceptually take as "natural images" always exhibit such a phenomenon. (We haven't yet seen counter examples, but Conversely, exhibiting this phenomenon can be thought of as a characteristic of the so called "natural images").