The Discrete Shearlet Transform : A New Directional Transform and Compactly Supported Shearlet Frames

Wang-Q Lim

Abstract

It is now widely acknowledged that analyzing the intrinsic geometrical features of the underlying image is essential in many applications including image processing. In order to achieve this, several directional image representation schemes have been proposed. In this paper, we develop the discrete shearlet transform (DST) which provides efficient multiscale directional representation and show that the implementation of the transform is built in the discrete framework based on a multiresolution analysis (MRA). We assess the performance of the DST in image denoising and approximation applications. In image approximations, our approximation scheme using the DST outperforms the discrete wavelet transform (DWT) while the computational cost of our scheme is comparable to the DWT. Also, in image denoising, the DST compares favorably with other existing transforms in the literature.

Index Terms

Discrete shearlet transform, shearlets, wavelets, multiresolution analysis, image approximations, image denoising.

EDICS Category: SMR-REP, TEC-MRS and TEC-RST

I. Introduction

One of the most useful features of wavelets is their ability to efficiently approximate signals containing pointwise singularities. Consider a one-dimensional signal $s(t)$ which is smooth away from point

This work was supported by DFG Grant SPP-1324, KU 1446/13-1. W. Lim is with the Institue of Mathematics, University of Osnabrueck, Osnabrueck, 49080 Germany e-mail: wlim@mathematik.Uni-Osnabrueck.DE

discontinuities. If $s(t)$ is approximated using the best M-term wavelet expansion, then the rate of decay of the approximation error, as a function of M, is optimal. In particular, it is significantly better than the corresponding Fourier approximation error [1],[2].

However, wavelets fail to capture the geometric regularity along the singularities of surfaces, because of their isotropic support. To exploit the anisotropic regularity of a surface along edges, the basis must include elongated functions that are nearly parallel to the edges. Several image representations have been proposed to capture the geometric regularity of a given image [3],[4],[5],[6],[7],[8]. They include curvelets [7], contourlets [5] and bandelets [3]. In particular, the construction of curvelets is not built directly in the discrete domain and they do not provide a multiresolution representation of the geometry. In consequence, the implementation and the mathematical analysis are more involved and less efficient. Contourlets are bases constructed with elongated basis functions using the combination of a multiscale and a directional filter bank. However, contourlets have less clear directional features than curvelets, which leads to artifacts in denoising and compression. Bandelets are bases adapted to the function that is represented. Asymptotically, the resulting bandelets are regular functions with compact support, which is not the case of contourlets. However, in order to find basis optimally adapted to an image of size N, the bandelet transform searches for the optimal geometry. For an image of N pixels, the complexity of this best bandelet basis algorithm is $O(N^{3/2})$ which requires extensive computation [8].

Recently, a new representation scheme has been introduced in [9] and [10]. This is so called the *shearlet* representation which yields nearly optimal approximation properties [11]. Shearlets are frame elements used in this representation scheme. This new representation is based on a simple and rigorous mathematical framework which not only provides a more flexible theoretical tool for the geometric representation of multidimensional data, but is also more natural for implementation.As a result, the shearlet approach can be associated to a multiresolution analysis (MRA) and this leads to a unified treatment of both the continuous and discrete world [10]. However, all known constructions of shearlets so far are band-limited functions which have an unbounded support in space domain. In fact, in order to capture the local features of a given image efficiently, representation elements need to be compactly supported in the space domain. Furthermore, this property often leads to more convenient framework for practically relevant discrete implementation.

The paper is structured as follows. In Section II, we describe some basic properties of shearlets and explain an intuitive idea of how shearlets can provide efficient geometric representations. We construct compactly supported shearlets generated by separable functions in Section III. In this case, the separable generating functions are constructed using a MRA and this leads to the fast discrete shearlet transform

(DST) which we describe in Section IV. Applications of the DST in image approximations and denoising are discussed in Section V and VI. Concluding remarks are drawn in Section VII.

II. SHEARLETS

A family of vectors $\{\varphi_n\}_{n\in\Gamma}$ constitute a frame for a Hilbert space $\mathcal{H}$ if there exist two positive constants A, B such that for each $f \in \mathcal{H}$ we have

$$A\|f\|^2 \leq \sum_{n\in\Gamma} |\langle f, \varphi_n\rangle|^2 \leq B\|f\|^2.$$

In the event that $A = B$, the frame is said to be tight. Let us introduce some notations that we will use throughout this paper. The Fourier transform of f is defined by

$$\hat{f}(\omega) = \int_{\mathbb{R}^d} f(x)e^{-2\pi i x\cdot\omega}dx.$$

Also, for $t \in \mathbb{R}^d$ and $A \in GL_d(\mathbb{R})$, we define the following unitary operators.

$$T_t(f)(x) = f(x-t)$$

and

$$D_A(f)(x) = |A|^{-\frac{1}{2}} f(A^{-1}x).$$

Here, we denote $GL_d(\mathbb{R})$ the set of all $d \times d$ invertible matrices with real entries. Finally, for $q \in (\frac{1}{2}, 1]$ and $a > 1$, we define

$$A_0 = \begin{pmatrix} a^q & 0 \\ 0 & a^{\frac{1}{2}} \end{pmatrix} \quad \text{and} \quad B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{II.1}$$

and

$$A_1 = \begin{pmatrix} a^{\frac{1}{2}} & 0 \\ 0 & a^q \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{II.2}$$

We are now ready to define a shearlet frame as follows. Let $s_j = \lceil a^{j(q-1/2)} \rceil$ and $c \in \mathbb{R}^+$ be the sampling constant. For $\psi_0^1, \ldots, \psi_0^L, \psi_1^1, \ldots, \psi_1^L \in L^2(\mathbb{R}^2)$ and $\phi \in L^2(\mathbb{R}^2)$, we define

$$\Psi^0 = \{\psi^i_{jk\mathbf{m}} : j, k \in \mathbb{Z}, \mathbf{m} \in \mathbb{Z}^2, i = 1, \ldots, L\},$$

and

$$\begin{aligned} \Psi &= \{T_{cm}\phi : \mathbf{m} \in \mathbb{Z}^2\} \cup \{\psi^i_{jk\mathbf{m}} : j \geq 0, -s_j \leq k \leq s_j, \mathbf{m} \in \mathbb{Z}^2, i = 1, \ldots, L\} \\ &\cup \{\tilde{\psi}^i_{jk\mathbf{m}} : j \geq 0, -s_j \leq k \leq s_j, \mathbf{m} \in \mathbb{Z}^2, i = 1, \ldots, L\} \end{aligned}$$

where

$$\psi^i_{jk\mathbf{m}} = D_{A_0^{-j}B_0^{-k}} T_{c\mathbf{m}} \psi^i_0 \quad \text{and} \quad \tilde{\psi}^i_{jk\mathbf{m}} = D_{A_1^{-j}B_1^{-k}} T_{c\mathbf{m}} \psi^i_1. \tag{II.3}$$

If Ψ (or Ψ^0) is a frame for $L^2(\mathbb{R}^2)$ then we call the functions $\psi^i_{jk\mathbf{m}}$ and $\tilde{\psi}^i_{jk\mathbf{m}}$ in the system Ψ (or Ψ^0) *shearlets*.

Observe that shearlets in Ψ (or Ψ^0) are obtained by applying translations with the sampling constant c followed by applying anisotropic scaling matrices A^j_ℓ and shear matrices B^k_ℓ to the fixed generating functions ψ^i_0 and ψ^i_1. Those matrices A^j_ℓ and B^k_ℓ lead to windows which can be elongated along arbitrary directions and the geometric structures of singularities in images can be efficiently represented and analyzed using them. In fact, it can be shown that 2 dimensional piecewise smooth functions with C^2singularities can be approximated with nearly optimal approximation rate using shearlets. We refer to [11] for details. Furthermore, one can show that shearlets can completely analyze the singular structures of piecewise smooth images [12]. Those properties of shearlets are useful especially in image processing since singularities and irregular structures carry essential information in an underlying image. For example, discontinuities in the intensity of an image indicate the presence of edges. The following example shows

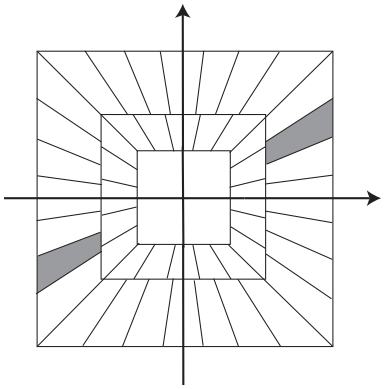

Fig. 1. The tiling of the frequency plane $\widehat{\mathbb{R}}^2$ induced by shearlets in Ψ. The tiling of horizontal conic region is illustrated in solid line and the tiling of vertical conic region is in dashed line.

how shearlets localize the frequency domain. Let us assume that $a = 2$ and $q = 1$ in (II.1) and (II.2). For any $\omega \in \widehat{\mathbb{R}}^2, \omega_1 \neq 0$ (or $\omega_2 \neq 0$), let ψ^1_0 and ψ^1_1 be given by

$$\begin{aligned} \hat{\psi}^1_0(\omega) &= \hat{\psi}_1(\omega_1)\hat{\psi}_2(\frac{\omega_2}{\omega_1}), \\ \hat{\psi}^1_1(\omega) &= \hat{\psi}_1(\omega_2)\hat{\psi}_2(\frac{\omega_1}{\omega_2}) \end{aligned}$$

where $\hat{\psi}_1, \hat{\psi}_2$ are infinitely smooth functions, $\mathbf{supp}(\hat{\psi}_1) \subset [-\frac{4}{5}, -\frac{1}{4}] \cup [\frac{1}{4}, \frac{4}{5}]$ and $\mathbf{supp}(\hat{\psi}_2) \subset [-1, 1]$. We assume that

$$\sum_{j \geq 0} |\hat{\psi}_1(2^{-j}\omega_1)|^2 = 1 \qquad \text{for} \quad |\omega_1| \geq \frac{1}{2}$$

and

$$\sum_{\ell=-1}^{1} |\hat{\psi}_2(\omega_1 + \ell)|^2 = 1 \qquad \text{for} \quad \omega_1 \in [-1, 1].$$

Then it can be easily shown that the band-limited functions $\hat{\psi}_0^1$ and $\hat{\psi}_1^1$ generate a tight frame Ψ for $L^2(\mathbb{R}^2)$ with an appropriate choice of ϕ and some modification of shearlet elements ψ_{jkm}^1 and $\tilde{\psi^1}_{jkm}$ whose frequency support intersects the line $\omega_1 = \omega_2$ or $\omega_1 = -\omega_2$ (see [11] for more details). Figure 1 shows the tiling of the frequency plane $\hat{\mathbb{R}^2}$ using the band-limited shearlet system Ψ that we described above. It was shown that this band-limited system Ψ can provide nearly optimal approximation for a piecewise smooth function f with C^2 smoothness except at points lying on C^2 curves [11]. However, the band-limited shearlet elements have an unbounded support in the space domain. Thus, they are not well localized in the space. Furthermore, it is not clear how to develop the discrete shearlet transform naturally derived from this construction so that it inherit the appropriate mathematical properties from the corresponding shearlet system on continuous domain.

The aim of this paper is basically to provide the novel construction of shearlets which overcomes the drawbacks of the band-limited shearlets and other directional representation systems. In addition to directionality and anisotropy, we identify a wish list for the new construction of shearlets :

1) Frame property : This will lead to a stable reconstruction of a given image.
2) Localization : The each of shearlet frame elements needs to be well localized in both the space and the frequency domain.
3) Efficient implementation : The discrete implementation needs to be naturally derived from the construction of shearlets so that it inherit the nice properties from the corresponding shearlet system.
4) Sparse approximation : The new construction scheme should provide sparse approximation comparable with the band-limited shearlets.

III. Construction of Shearlets

In this section, we will introduce some useful sufficient conditions to construct compactly supported shearlets associated with a MRA. Based on this construction, we will develop the DST in Section IV.

A. Some sufficient conditions for the construction of shearlets

Recall that $c > 0$ is the sampling constant in (II.3). We need the following result in [13] to derive the main result in this section.

Theorem III.1. *Let $c > 0$ and for $s \in \widehat{\mathbb{R}}^2$, we define*

$$\Phi(s) = \operatorname*{ess\,sup}_{\omega \in \widehat{\mathbb{R}}^2} \sum_{i=1}^{L} \sum_{j,k \in \mathbb{Z}} |\hat{\psi}_0^i((B_0^T)^k A_0^j \omega) \hat{\psi}_0^i((B_0^T)^k A_0^j \omega + s)| \tag{III.4}$$

where $\psi_0^1, \ldots, \psi_0^L \in L^2(\mathbb{R}^2)$. If there exist $0 < D_1 < D_2 < \infty$ such that

$$D_1 \leq \sum_{i=1}^{L} \sum_{j,k \in \mathbb{Z}} |\hat{\psi}_0^i((B_0^T)^k A_0^j \omega)|^2 \leq D_2 \quad \text{for a.e } \omega \in \widehat{\mathbb{R}}^2 \tag{III.5}$$

and

$$\sum_{s \in \mathbb{Z}^2 \setminus \{0\}} \sqrt{\Phi\Big(\frac{s}{c}\Big)\Phi\Big(-\frac{s}{c}\Big)} = \epsilon_0 < D_1 \tag{III.6}$$

then Ψ^0 is a frame for $L^2(\mathbb{R}^2)$ with frame bounds C_1 and C_2 satisfying

$$\frac{1}{c^2}(D_1 - \epsilon_0) \leq C_1 \leq C_2 \leq \frac{1}{c^2}(D_2 + \epsilon_0).$$

By Theorem III.1, one can show the following result.

Proposition III.2. *Let $\alpha > 0$ and $\gamma > 2(\alpha + 2)$. If functions $\psi_0^1, \ldots, \psi_0^L$ satisfy*

$$|\hat{\psi}_0^i(\omega)| \leq C|\omega_1|^\alpha \frac{1}{(1 + |\omega_1|^2)^{\gamma/2}} \Big(\frac{|\omega_1|^2}{|\omega_1|^2 + |\omega_2|^2}\Big)^{\gamma/2} \tag{III.7}$$

for a.e $\omega \in \widehat{\mathbb{R}}^2$ and

$$D_1 \leq \sum_{i=1}^{L} \sum_{j,k \in \mathbb{Z}} |\hat{\psi}_0^i((B_0^T)^k A_0^j \omega)|^2 \quad \text{for a.e } \omega \in \widehat{\mathbb{R}}^2 \tag{III.8}$$

then there exists $c_0 > 0$ such that Ψ^0 is a frame for $L^2(\mathbb{R}^2)$ for any sampling constant $c \leq c_0$.

Proof : See *Appendix* A.

Proposition III.2 implies that if functions $\psi_0^1, \ldots, \psi_0^L$ have sufficient vanishing moments and fast decay in the frequency domain and III.8 is satisfied then there exists the sampling constant c_0 such that the functions $\psi_0^1, \ldots, \psi_0^L$ generate a frame Ψ^0 with any sampling constant $c \leq c_0$. This result can be easily extended to obtain similar sufficient conditions for Ψ to be a frame for $L^2(\mathbb{R}^2)$.

Corollary III.3. *Let α and γ be constants as in Proposition III.2. Assume that $\alpha' \geq \alpha + \gamma$ and $\gamma' \geq \alpha' - \alpha + \gamma$. For $i = 1, \ldots, L$, we define $\psi_0^i(x_1, x_2) = \gamma^i(x_1)\theta(x_2)$ such that*

$$|\hat{\gamma}^i(\omega_1)| \leq K_1 \frac{|\omega_1|^{\alpha'}}{(1 + |\omega_1|^2)^{\gamma'/2}}$$

and

$$|\hat{\theta}(\omega_2)| \leq K_2(1+|\omega_2|^2)^{-\gamma'/2}.$$

If

$$\operatorname*{ess\,inf}_{|\omega_2|\leq 1/2} |\hat{\theta}(\omega_2)|^2 \geq K_3 > 0 \tag{III.9}$$

and

$$\operatorname*{ess\,inf}_{\beta a^{-q}\leq|\omega_1|\leq\beta} \sum_{i=1}^{L} |\hat{\gamma}^i(\omega_1)|^2 \geq K_4 > 0 \quad \textit{for } 0<\beta\leq \min\left(1,\frac{a^q}{2}\right) \tag{III.10}$$

then there exists $c_0 > 0$ such that Ψ^0 is a frame for $L^2(\mathbb{R}^2)$ for all $c \leq c_0$.

Proof : First, it is easy to see that

$$\frac{|\omega_1|^{\alpha'}}{(1+|\omega_1|^2)^{\gamma'/2}(1+|\omega_2|^2)^{\gamma'/2}} \quad < \quad C\frac{|\omega_1|^{\alpha+\gamma}}{(1+|\omega_1|^2)^{\gamma/2}}\frac{1}{(|\omega_1|^2+|\omega_2|^2)^{\gamma/2}}.$$

This implies that $\psi_0^1,\ldots,\psi_0^L$ satisfy (III.7). Thus, it suffices to show (III.8). Let $\Lambda = \{(\omega_1,\omega_2)\in\widehat{\mathbb{R}}^2 : \beta a^{-q}\leq|\omega_1|\leq\beta,\ |\omega_2|\leq 1/2\}$. Direct computation gives us

$$\sum_{j,k\in\mathbb{Z}}\sum_{i=1}^{L}|\hat{\psi}_0^i((B^T)^kA^j\omega)|^2 \geq K_3^2K_4^2\sum_{j,k\in\mathbb{Z}}\chi_{A_0^j(B_0^T)^k\Lambda}(\omega)$$

for a.e $\omega\in\widehat{\mathbb{R}}^2$ and

$$\bigcup_{j\in\mathbb{Z}}\bigcup_{k\in\mathbb{Z}} A_0^j(B_0^T)^k\Lambda = \widehat{\mathbb{R}}^2.$$

We complete the proof. □

Corollary III.4. *For $i=1,\ldots,L$, let $\psi_0^i(\boldsymbol{x})=\gamma^i(x_1)\theta(x_2)$, $\psi_1^i(\boldsymbol{x})=\gamma^i(x_2)\theta(x_1)$ and $\phi(\boldsymbol{x})=\theta(x_1)\theta(x_2)$ where the functions γ^i and θ satisfy the assumptions in Corollary III.3. Then there exists $c_0>0$ such that ψ_ℓ^i and ϕ generate a shearlet frame Ψ of $L^2(\mathbb{R}^2)$ for $c\leq c_0$.*

Proof : Note that in this case, (III.4) and (III.5) become

$$\Phi(s) = \operatorname*{ess\,sup}_{\omega\in\widehat{\mathbb{R}}^2}\sum_{\ell=0}^{1}\sum_{i=1}^{L}\sum_{j=0}^{\infty}\sum_{k=-s_j}^{s_j}|\hat{\psi}_\ell^i((B_\ell^T)^kA_\ell^j\omega)\hat{\psi}_\ell^i((B_\ell^T)^kA_\ell^j\omega+s)| + |\hat{\phi}(\omega)\hat{\phi}(\omega+s)| \tag{III.11}$$

and

$$D_1 \leq \sum_{\ell=0}^{1}\sum_{i=1}^{L}\sum_{j=0}^{\infty}\sum_{k=-s_j}^{s_j}|\hat{\psi}_\ell^i((B_\ell^T)^kA_\ell^j\omega)|^2 + |\hat{\phi}(\omega)|^2 \leq D_2 \quad \text{for a.e } \omega\in\widehat{\mathbb{R}}^2 \tag{III.12}$$

where $s_j = \lceil a^{j(q-1/2)} \rceil$. (III.11) and (III.12) can be similarly estimated to obtain (III.6) and the upper bound D_2 in (III.12). We now estimate the lower bound in (III.12). Define

$$\Lambda_0 = \{\omega \in \widehat{\mathbb{R}}^2 \, : \, \beta a^{-q} \le |\omega_1| \le \beta, \, |\omega_2| \le 1/2\}$$

and

$$\Lambda_1 = \{\omega \in \widehat{\mathbb{R}}^2 \, : \, \beta a^{-q} \le |\omega_2| \le \beta, \, |\omega_1| \le 1/2\}.$$

Then we obtain

$$\sum_{\ell=0}^{1}\sum_{i=1}^{L}\sum_{j=0}^{\infty}\sum_{k=-s_j}^{s_j} |\hat{\psi}_\ell^i((B_\ell^T)^k A_\ell^j \omega)|^2 + |\hat{\phi}(\omega)|^2 \ge K \sum_{\ell=0}^{1}\sum_{j=0}^{\infty}\sum_{k=-s_j}^{s_j} \chi_{A_\ell^j (B_\ell^T)^k \Lambda_\ell}(\omega) + K' \chi_{[-1/2,1/2]^2}(\omega)$$

for a.e $\omega \in \widehat{\mathbb{R}}^2$ and

$$[-1/2,1/2]^2 \cup \Big(\bigcup_{\ell=0}^{1} \bigcup_{j=0}^{\infty} \bigcup_{k=-s_j}^{s_j} A_\ell^j (B_\ell^T)^k \Lambda_\ell \Big) = \widehat{\mathbb{R}}^2.$$

This completes the proof. □

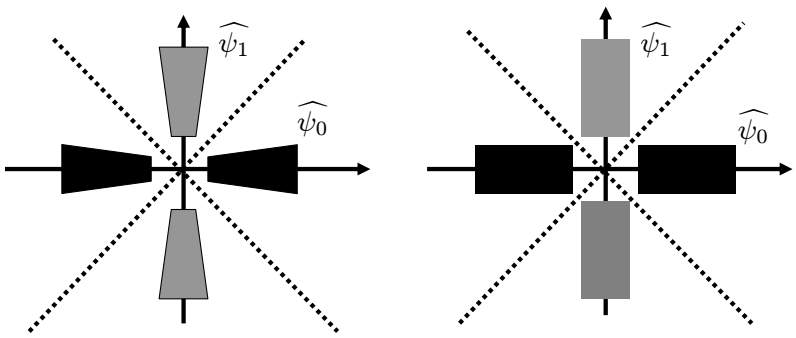

Fig. 2. Essential frequency supports of shearlets ψ_0^i and ψ_1^i in the frequency plane : Nonseparable shearlets (left) and separable shearlets (right). Both of separable and nonseparable windows can cover whole frequency plane by applying shear and scaling matrices.

Corollaries III.3 and III.4 show us that one can easily construct compactly supported shearlet frames generated by separable generating functions. Observe that horizontal and vertical rectangular windows can cover frequency plane by applying scaling and shear matrices (see Figure 2 (right)) and the resulting windows provide directional decomposition of the frequency plane. We should point out that compared to shearlets generated from the separable functions, shearlets generated from nonseparable functions whose essential frequency support looks as in Figure 2 (left) can more effectively cover the frequency plane $\widehat{\mathbb{R}^2}$. In fact, in this case there is basically no overlap between the essential supports of shearlets as we

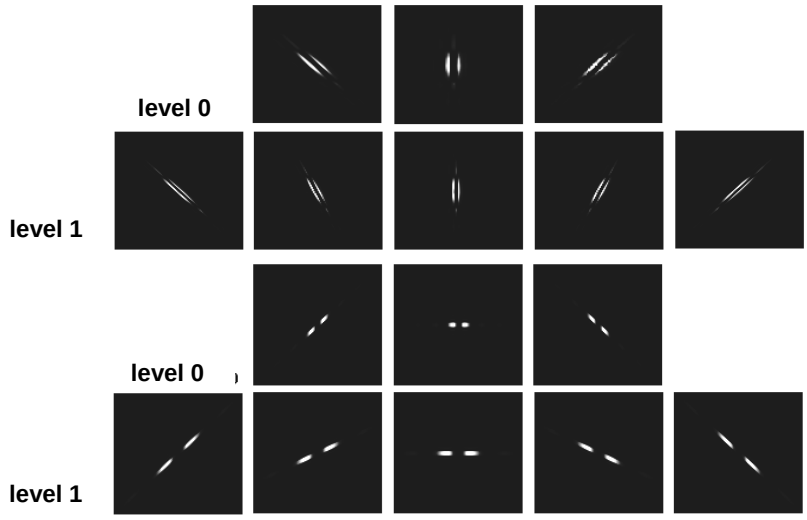

Fig. 3. Examples of shearlets in the space and frequency domain: Shearlets ψ^i_{jk0} associated with matrices A_0 and B_0 in (II.1) when $j = 0, 1$ (top rows). The magnitude of the Fourier transform of shearlets $|\hat{\psi}^i_{jk0}|$ associated with matrices A_0 and B_0 in (II.1) when $j = 0, 1$ (bottom rows).

have seen in Figure 1. This indicates nonseparable generating functions for shearlet system can provide the better frame bounds or even tight frame. However, the advantage of this separable construction is of course the simplicity of the construction and one can easily obtain shearlets with good properties such as compact support and regularity. On the other hand, this separibility leads to a fast implementation of computing the shearlet coefficients as we will show in the next sections. Figure 3 shows compactly supported shearlets associated with matrices A_0 and B_0 which provide directional windows in the space and frequency domain. Note that in this case, all of those are generated by separable functions of the form $\gamma^i(x_1)\theta(x_2)$. We now describe concrete examples of compactly supported shearlet frames using Corollaries (III.3) and (III.4).

B. Examples of compactly supported shearlets

Assume that $a = 4$ and $q = 1$ in (II.1). Then we have

$$A_0 = \begin{pmatrix} 4 & 0 \\ \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad B_0 = \begin{pmatrix} 1 & 1 \\ \\ 0 & 1 \end{pmatrix}.$$

Let us consider a box spline [1] of order m defined as follows.

$$\hat{\theta}_m(\omega_1) = \Big(\frac{\sin \pi\omega_1}{\pi\omega_1}\Big)^{m+1} e^{-i\epsilon\omega_1}$$

with $\epsilon = 1$ for m even and $\epsilon = 0$ for m odd. Observe that we have the following two scaling equation.

$$\hat{\theta}_m(2\omega_1) = m_0(\omega_1)\hat{\theta}_m(\omega_1)$$

and

$$m_0(\omega_1) = (\cos \pi\omega_1)^{m+1} e^{-i\epsilon\pi\omega_1}.$$

Let α' and γ' be positive real numbers as in Corollary III.3. We now define

$$\hat{\psi}_0^1(\omega) = (i)^\ell \Big(\sin \pi\omega_1\Big)^\ell \hat{\theta}_m(\omega_1)\hat{\theta}_m(\omega_2)$$

and

$$\hat{\psi}_0^2(\omega) = (i)^\ell \Big(\sin \frac{\pi\omega_1}{2}\Big)^\ell \hat{\theta}_m(\frac{\omega_1}{2})\hat{\theta}_m(\omega_2)$$

where $\ell \geq \alpha'$ and $m + 1 \geq \gamma'$. Then by Corollary III.3, ψ_0^1 and ψ_0^2 generate a frame Ψ^0 for $c \leq c_0$ with some $c_0 > 0$.

Define $\phi(x_1, x_2) = \theta_m(x_1)\theta_m(x_2)$,

$$\hat{\psi}_1^1(\omega) = (i)^\ell \Big(\sin \pi\omega_2\Big)^\ell \hat{\theta}_m(\omega_2)\hat{\theta}_m(\omega_1)$$

and

$$\hat{\psi}_1^2(\omega) = (i)^\ell \Big(\sin \frac{\pi\omega_2}{2}\Big)^\ell \hat{\theta}_m(\frac{\omega_2}{2})\hat{\theta}_m(\omega_1).$$

Then, by Corollary (III.4), the functions ϕ, ψ_ℓ^i for $\ell = 0, 1$ and $i = 1, 2$ generate a frame Ψ with $c \leq c_0$ for some $c_0 > 0$.

IV. Discrete Shearlet Transform

In the previous section, we constructed compactly supported shearlets associated with a MRA. In this section, we will show that this MRA leads to the fast DST which computes the shearlet coefficients $\langle f, \psi_{jk\mathbf{m}}^i \rangle$ and $\langle f, \tilde{\psi}_{jk\mathbf{m}}^i \rangle$. To be more specific, we let $a = 2$, $q = 1$ and the sampling constant $c = 1$ in (II.1) and (II.2). Also, to save notations, we slightly modify the notations we used in the previous section. First, for $j \geq 0$, define

$$A_{0,j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{\lfloor j/2 \rfloor} \end{pmatrix}, \; B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \; Q_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

and

$$A_{1,j} = \begin{pmatrix} 2^{\lfloor j/2 \rfloor} & 0 \\ 0 & 2^j \end{pmatrix}, \; B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ and } \; Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Finally, we let $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2, \mathbf{n} = (n_1, n_2), \mathbf{m} = (m_1, m_2), \mathbf{d} = (d_1, d_2) \in \mathbb{Z}^2$ and I_2 be a 2 by 2 identity matrix.

A. Forward Transform

Assume that $\theta^0 \in L^2(\mathbb{R})$ is a compactly supported function such that $\{\theta^0(\cdot - n_1) : n_1 \in \mathbb{Z}\}$ is an orthonormal sequence and

$$\theta^0(x_1) = \sum_{n_1 \in \mathbb{Z}} h^0(n_1)\sqrt{2}\theta^0(2x_1 - n_1). \tag{IV.13}$$

Also, define

$$\gamma(x_1) = \sum_{n_1 \in \mathbb{Z}} g(n_1)\sqrt{2}\theta^0(2x_1 - n_1) \tag{IV.14}$$

such that γ has sufficient vanishing moments and the pair of the filters h^0 and g is a pair of conjugate mirror filters. We also assume that γ and θ satisfy the conditions in Corollary III.3. For $j \in \mathbb{Z}^+ \cup \{0\}$, let

$$\theta^0_{j,n_1}(x_1) = 2^{j/2}\theta^0(2^j x_1 - n_1) \quad \text{and} \quad \gamma_{j,n_1}(x_1) = 2^{j/2}\gamma(2^j x_1 - n_1). \tag{IV.15}$$

Then $\{\theta^0_{j,n_1}\}_{n_1 \in \mathbb{Z}}$ and $\{\gamma_{j,n_1}\}_{n_1 \in \mathbb{Z}}$ are orthonormal bases of $V^0_j = \overline{\text{span}\{\theta^0_{j,n_1} : n_1 \in \mathbb{Z}\}}$ and $W_j = \overline{\text{span}\{\gamma_{j,n_1} : n_1 \in \mathbb{Z}\}}$ respectively. Also, we have $V^0_{j+1} = V^0_j \oplus W_j$ for each $j \in \mathbb{Z}$ and further assume subspaces V^0_j satisfy the MRA conditions. For $s \in [-1, 1]$, define

$$\psi^0_{js\mathbf{n}}(\mathbf{x}) = D_{B_0^{-s}A^{-1}_{0,j}} T_{\mathbf{n}} \psi^0_0$$

where $\psi^0_0(\mathbf{x}) = \gamma(x_1)\theta^0(x_2)$. In this case, we have

$$\psi^0_{j0\mathbf{n}}(\mathbf{x}) = \gamma_{j,n_1}(x_1)\theta^0_{\lfloor j/2 \rfloor, n_2}(x_2) \quad \text{and} \quad \psi^0_{js\mathbf{n}} = D_{B_0^{-s}}\psi^0_{j0\mathbf{n}}.$$

Similarly, we also define

$$\tilde{\psi}^0_{js\mathbf{n}}(\mathbf{x}) = D_{B_1^{-s}A^{-1}_{1,j}} T_{\mathbf{n}} \psi^0_1$$

where $\psi^0_1(\mathbf{x}) = \gamma(x_2)\theta^0(x_1)$. Let $s_{jk} = k/2^{\lceil j/2 \rceil}$ for $j \geq 0$ and $-2^{\lceil j/2 \rceil} \leq k \leq 2^{\lceil j/2 \rceil}$. Since $A_{\ell,j}B^{s_{jk}}_\ell = B^k_\ell A_{\ell,j}$ for $\ell = 0, 1$, we obtain shearlets

$$\psi^0_{js_{jk}\mathbf{n}} = D_{A^{-1}_{0,j}B_0^{-k}}\psi^0_0 \quad \text{and} \quad \tilde{\psi}^0_{js_{jk}\mathbf{n}} = D_{A^{-1}_{1,j}B_1^{-k}}\psi^0_1$$

which are the functions of the form in (II.3).

From now on, we will use the notations we defined above throughout this paper. We now explicitly derive discrete formulas to compute shearlet coefficients $\langle f, \psi^0_{js_{jk}\mathbf{n}}\rangle$ and $\langle f, \tilde{\psi}^0_{js_{jk}\mathbf{n}}\rangle$ for $j \geq 0, -2^{\lceil j/2 \rceil} \leq$

$k \le 2^{\lceil j/2 \rceil}$ and $f \in L^2(\mathbb{R}^2)$. We mainly consider $\langle f, \psi^0_{js_{jk}\mathbf{n}} \rangle$ and the rest of shearlet coefficients are similarly computed.

Assume that $f \in V^0_J \otimes V^0_J = \overline{\text{span}\{\theta^0_{J,n_1}(x_1)\theta^0_{J,n_2}(x_2) : (n_1,n_2) \in \mathbb{Z}^2\}}$. and $\phi^0_0(\mathbf{x}) = \theta^0(x_1)\theta^0(x_2)$ for fixed $J > 0$. Then we can write

$$f = \sum_{\mathbf{n} \in \mathbb{Z}^2} f_J(\mathbf{n}) D_{2^{-J}I_2} T_{\mathbf{n}} \phi^0_0 \tag{IV.16}$$

where $f_J(\mathbf{n}) = \langle f, D_{2^{-J}I_2} T_{\mathbf{n}} \phi^0_0 \rangle$. For $j > 0$, let g_j and h^0_j be the Fourier coefficients of

$$G_j(\omega_1) = \prod_{k=0}^{j-2} H^0(2^k\omega_1) G(2^{j-1}\omega_1) \quad \text{and} \quad H^0_j(\omega_1) = \prod_{k=0}^{j-1} H^0(2^k\omega_1)$$

respectively, where

$$H^0(\omega_1) = \sum_{n_1} h^0(n_1) e^{-2i\pi\omega_1} \quad \text{and} \quad G(\omega_1) = \sum_{n_1} g(n_1) e^{-2i\pi\omega_1}.$$

Also, we let $g_j(n_1) = h^0_j(n_1) = \delta(n_1)$ when $j = 0$. The following lemma is an immediate consequence of the cascade algorithm followed from (IV.13) and (IV.14).

Lemma IV.1. *Assume that θ^0 and $\gamma \in L^2(\mathbb{R})$ satisfy (IV.13) and (IV.14). For each $j \in \mathbb{Z}^+ \cup \{0\}$, define γ_{j,n_1} and θ^0_{j,n_2} as in (IV.15). Then we have*

$$\gamma_{j,n_1}(x_1) = \sum_{d_1} g_{J-j}(d_1 - 2^{J-j}n_1)\theta^0_{J,d1}(x_1). \tag{IV.17}$$

and

$$\theta^0_{j,n_2}(x_2) = \sum_{d_2} h^0_{J-j}(d_2 - 2^{J-j}n_2)\theta^0_{J,d2}(x_2) \tag{IV.18}$$

for $0 \le j < J$.

Proof : See *Appendix* B.

From (IV.17) and (IV.18), we obtain

$$\psi^0_{j\mathbf{0}\mathbf{n}}(\mathbf{x}) = \sum_{\mathbf{d} \in \mathbb{Z}^2} g_{J-j}(d_1 - 2^{J-j}n_1) h^0_{J-\lfloor j/2 \rfloor}(d_2 - 2^{J-\lfloor j/2 \rfloor}n_2)\theta^0_{J,d1}(x_1)\theta^0_{J,d2}(x_2). \tag{IV.19}$$

On the other hand, we have

$$\langle f, \psi^0_{js_{jk}\mathbf{n}} \rangle = \langle D_{B_0^{s_{jk}}} f, \psi^0_{j0\mathbf{n}} \rangle \tag{IV.20}$$

We note that if the function $D_{B_0^{s_{jk}}} f \in V_J \otimes V_J$ so that it can be represented by a linear combination of integer translates of the scaling function ϕ^0_0 at Jth level then the shearlet coefficients in (IV.20) can be easily computed by applying the separable wavelet transform associated with the anisotropic scaling

matrix $A_{0,j}$. However, this is not the case unless $k = 0$. In MRA approach, it is essential to represent an underlying function by a linear combination of integer translates of scaling function at the given resolution level. We now show that this can be done by refining integer grid $2^{-J}\mathbb{Z}^2$ along the horizontal axis. Lemma IV.1 implies that

$$\theta^0_{J,n_1}(x_1) = \sum_{d_1\in\mathbb{Z}} h^0_{\lceil j/2\rceil}(d_1 - 2^{\lceil j/2\rceil}n_1)\theta^0_{J+\lceil j/2\rceil,d_1}(x_1).$$

Hence, we have

$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{Z}^2} ((f_J)_{\uparrow 2^{\lceil j/2\rceil}} *_1 h^0_{\lceil j/2\rceil})(\mathbf{n}) D_{2^{-J}Q_0^{-\lceil j/2\rceil}} T_{\mathbf{n}}\phi^0_0(\mathbf{x}) \tag{IV.21}$$

where $*_1$ and $\uparrow 2^j$ are 1D convolution and upsampling by 2^j along the horizontal axis and

$$D_{B_0^{s_{jk}}} f = \sum_{\mathbf{n}\in\mathbb{Z}^2} ((f_J)_{\uparrow 2^{\lceil j/2\rceil}} *_1 h^0_{\lceil j/2\rceil})(B_0^{-k}\mathbf{n}) D_{2^{-J}Q_0^{-\lceil j/2\rceil}} T_{\mathbf{n}}\phi^0_k \tag{IV.22}$$

where $\phi^0_k = D_{B_0^k}\phi^0_0$. (IV.22) implies that the function $D_{B_0^{s_{jk}}} f$ can be represented by a linear combination of integer translates of sheared scaling function ϕ^0_k on anisotropic grid $2^{-J-\lceil j/2\rceil}\mathbb{Z}^2 \times 2^{-J}\mathbb{Z}^2$ obtained by refinement along the horizontal axis. For nonnegative integers j_1 and j_2, define a linear transform $\mathcal{W}_{j_1,j_2} : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)$ by

$$\mathcal{W}^0_{j_1,j_2}(f)(n_1,n_2) = \sum_{\mathbf{m}\in\mathbb{Z}^2} g_{j_1}(m_1 - 2^{j_1}n_1) h^0_{j_2}(m_2 - 2^{j_2}n_2) f(m_1,m_2)$$

for $f \in \ell^2(\mathbb{Z}^2)$. This is simply the separable wavelet transform associated with a diagonal sampling matrix whose diagonal entries are 2^{j_1} and 2^{j_2}. We also let $\Phi_k(\mathbf{n}) = \langle \phi^0_k, T_n\phi^0_0\rangle$ and $\overline{h^0_j}(n_1) = h^0_j(-n_1)$. Then, by (IV.19),(IV.20) and (IV.22),

$$\langle f, \psi^0_{js_{j,k}\mathbf{n}}\rangle = \mathcal{W}^0_{J-j,J-\lfloor j/2\rfloor}\Big(\big([((f_J)_{\uparrow 2^{\lceil j/2\rceil}} *_1 h^0_{\lceil j/2\rceil})(B_0^{-k}\cdot) * \Phi_k] *_1 \overline{h^0}_{\lceil j/2\rceil}\big)_{\downarrow 2^{\lceil j/2\rceil}}\Big)(\mathbf{n}) \tag{IV.23}$$

where $*_1$, $\uparrow 2^j$ and $\downarrow 2^j$ are 1D convolution, upsampling and downsampling by 2^j along the horizontal axis, respectively (see Figure 4(a)). Notice that $\big([((f_J)_{\uparrow 2^{\lceil j/2\rceil}} *_1 h^0_{\lceil j/2\rceil})(B_0^{-k}\cdot) * \Phi_k] *_1 \overline{h^0}_{\lceil j/2\rceil}\big)_{\downarrow 2^{\lceil j/2\rceil}}$ computes interpolating sample values $f_J(B_0^{s_{jk}}\mathbf{n})$ from $f_J(\mathbf{n})$. Alternatively, one can approximate $f_J(B_0^{s_{jk}}\mathbf{n})$ by taking $f_J(\mathbf{m})$ where $\mathbf{m} \in \mathbb{Z}^2$ is an integer lattice point closest to $B_0^{s_{j,k}}\mathbf{n}$. This suggests the following formula :

$$\langle f, \psi^0_{js_{jk}\mathbf{n}}\rangle \approx \mathcal{W}^0_{J-j,J-\lfloor j/2\rfloor}(P_{s_{jk}}(f_J))(\mathbf{n}) \tag{IV.24}$$

where

$$P_s(f_J)(\mathbf{n}) = f_J\big(n_1 + \lfloor sn_2\rfloor, n_2\big)$$

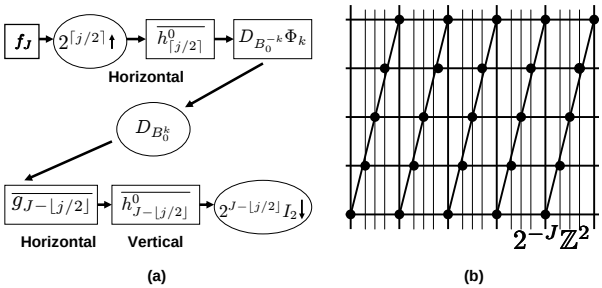

Fig. 4. (a) Two computation steps to compute the shearlet coefficients : Refinement along the horizontal axis (top) and resampling associated with shear matrix (middle) followed by the separable wavelet transform (bottom) (b) Refinement along the horizontal to obtain the coefficients of $D_{B_0^{k/2^{\lceil j/2 \rceil}}} f$ when $j = 4$ and $k = 1$.

for $s \in [-1, 1]$. Similarly, the shearlet coefficients $\langle f, \tilde{\psi}^0_{js_{j,k}\mathbf{n}} \rangle$ can be explicitly computed. Also, we have approximating formula as follows :

$$\langle f, \tilde{\psi}^0_{js_{j,k}\mathbf{n}} \rangle \approx \widetilde{\mathcal{W}^0}_{J-j,J-\lfloor j/2 \rfloor}(\tilde{P}_{s_{j,k}}(f_J))(\mathbf{n}) \tag{IV.25}$$

where

$$\widetilde{\mathcal{W}^0}_{j_1,j_2}(f)(n_1, n_2) = \sum_{\mathbf{m} \in \mathbb{Z}^2} g_{j_1}(m_2 - 2^{j_1} n_2) h^0_{j_2}(m_1 - 2^{j_2} n_1) f(m_1, m_2)$$

and

$$\tilde{P}_s(f_J)(\mathbf{n}) = f_J\big(n_1, n_2 + \lfloor s n_1 \rfloor\big).$$

Using (IV.24) and (IV.25), we now define the DST $S : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)$ which computes the shearlet coefficients of $f \in V^0_J \otimes V^0_J$. Define an index set

$$\Lambda_0 = \{(j, k) : j = 0, 1, \ldots, J-1, -2^{\lceil j/2 \rceil} \le k \le 2^{\lceil j/2 \rceil}\}$$

and

$$S(f) = \Big((\langle f, T_{\mathbf{n}} \phi^0_0 \rangle)_{\mathbf{n} \in \mathbb{Z}^2}, S_0(f), \tilde{S}_0(f) \Big)$$

where

$$\begin{cases} S_0(f) = \Big(\mathcal{W}^0_{J-j,J-\lfloor j/2 \rfloor}(P_{s_{jk}}(f_J)) \Big)_{(j,k) \in \Lambda_0} \\ \tilde{S}_0(f) = \Big(\widetilde{\mathcal{W}^0}_{J-j,J-\lfloor j/2 \rfloor}(\tilde{P}_{s_{jk}}(f_J)) \Big)_{(j,k) \in \Lambda_0} \end{cases} \tag{IV.26}$$

We should point out that directional filters associated with the discrete shear operations P_s and $\tilde{P}_s$ in

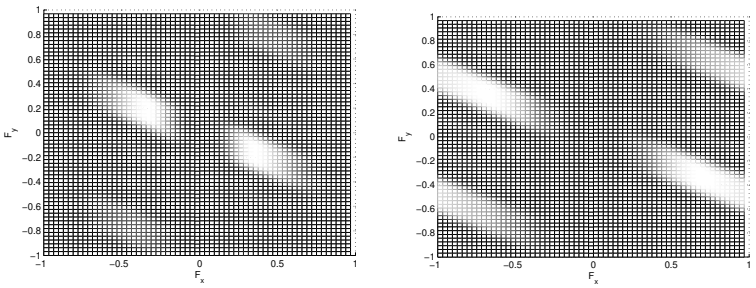

Fig. 5. (a): Directional filters with the grid refinement associated with $B_0^{k/2}$ for k = 1. (b): Aliased directional filters associted with $B_0^{k/2}$ for k = 1.

(IV.26) provide unsatisfactory frequency localization mainly due to the periodicity of 2D spectrum of discrete signals. Especially when the power of shear matrix B_0 (or B_1) is a non integer, the resulting shear matrix B_0^s (or B_1^s) does not preserve the structure of integer lattice $\mathbb{Z}^2$ and this makes the corresponding directional filter (in IV.26) aliased as we can see in Figure 5(b). On the other hand, we notice that the aliasing components introduced by the shear operations $D_{B_0^{s_{jk}}}$ with non integer powers $s_{jk} = k/2^{\lceil j/2 \rceil}$ can be removed by upsampling $2^{\lceil j/2 \rceil} \uparrow$ along the horizontal axis, followed by lowpass filtering as we proposed in (IV.23) (see Figure 4(a)). In fact, this refinement along the horizontal axis leads to shear matrices B_0^k (from $B_0^{s_{jk}}$) which preserve integer lattice and the aliasing components due to the shear operations $D_{B_0^{s_{jk}}}$ are removed. Figure 5(a) shows the frequency support of one of the resulting directional filters with this refinement.

In order to deal with images defined on finite domain, modification of our DST is necessary for shearlets near the boundary of the domain. The simplest way to deal with these boundary issues for linear transform involving convolution is to periodize an underlying image. Of course, this simple scheme can easily adapted to the DST and we used this for numerical tests for image approximations we will describe in later section. However, the periodization produces artificial line singularities due to shear operations as well as singularities on the boundary. For those artificial line singularities, one can apply a mirror extension along the horizontal or vertical axis depending on the choice of shear matrices B_0 and B_1 respectively.

Before we finish this section, we briefly mention about the redundancy of the DST. For simplicity, we only consider the case when the sampling constant $c = 1$ and this can be easily generalized. Let us

assume f_J is a $2^J \times 2^J$ discrete image. We notice that shear operations $P_{s_{jk}}$ and $\tilde{P}_{s_{jk}}$ do not change the size of the image and the anisotropic wavelet transforms $\mathcal{W}^0_{j_1,j_2}$ and $\widetilde{\mathcal{W}^0}_{j_1,j_2}$ produce $2^{J-j_1} \times 2^{J-j_2}$ and $2^{J-j_2} \times 2^{J-j_1}$ subimages respectively. Also, the transform S_0 is to apply $2^{\lceil j/2 \rceil+1}+1$ shear operations followed by anisotropic wavelet transform $\mathcal{W}^0_{J-j,J-\lfloor j/2 \rfloor}$ at each jth level. Thus, S_0 provides $2^{\lceil j/2 \rceil+1}+1$ subimages of size $2^j \times 2^{\lfloor j/2 \rfloor}$ at each jth level. This indicates that the total number of shearlet coefficients obtained from S_0 is

$$\#(S_0) = \sum_{j=0}^{(J-1)} (2^{\lceil j/2 \rceil+1}+1)(2^{j+\lfloor j/2 \rfloor})$$

and $\#(\tilde{S}_0) = \#(S_0)$. Finally, the size of DC coefficients $\langle f, T_{\mathbf{n}}\phi^0_0 \rangle$ is 1×1. Therefore, the total number of shearlet coefficients $\#(S)$ obtained from the DST S is

$$\begin{aligned} \#(S) &= 2\sum_{j=0}^{(J-1)} (2^{\lceil j/2 \rceil+1}+1)(2^{j+\lfloor j/2 \rfloor}) + 1 \\ &= 2\Big(2\sum_{j=0}^{J-1} 4^j + \sum_{j=0}^{J-1} (2\sqrt{2})^j\Big) + 1 \\ &= \frac{4}{3}(4^J-1) + \frac{2}{2\sqrt{2}-1}((2\sqrt{2})^J - 1) + 1. \end{aligned}$$

From this, redundancy ratio is given by

$$\frac{\#(S)}{2^{2J}} = \frac{4}{3}(1-4^{-J}) + \frac{2}{2\sqrt{2}-1}(2^{-J/2}-4^{-J}) + 4^{-J} \leq 2$$

for $J \geq 1$.

B. Inverse Transform

It is well known that an inverse frame operator with reasonable frame bounds can be effectively computed by applying iterative schemes such as conjugate gradient iterations [1]. Let S^* be the adjoint operator of the DST S. Then the pseudo inverse to the frame operator S is given by $(S^*S)^{-1}S^*$ and this can be efficiently computed by conjugate gradient iterations. As one can see in Figure 6(a), only few iterations provide good approximation to a given image f from the shearlet coefficients.

C. Extended Discrete Shearlet Transform

In the previous section, we derived a fast algorithm to compute the shearlet frame coefficients based on a standard MRA approach. However, it is often convenient to work with a tight frame rather than a general frame in many applications. In particular, one can not directly compute an inverse of frame

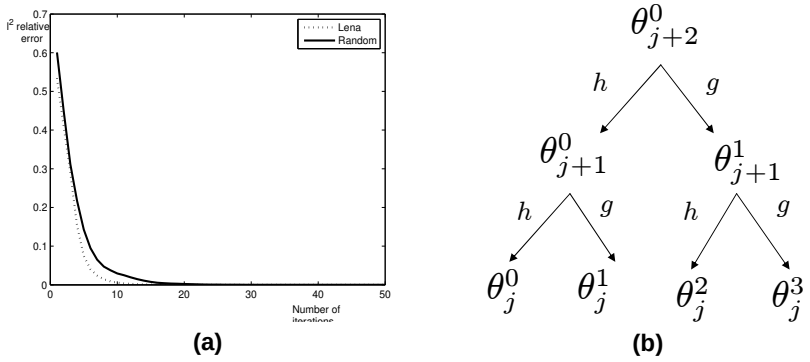

Fig. 6. (a) Convergence rate of approximating inverse DST for a random image whose entries are uniformly distributed and Lena : vertical axis : ℓ^2 relative error $\frac{\|f-\hat{f}\|_2}{\|f\|_2}$ and horizontal axis : the number of iterations in conjugate gradient iterations. Random image (solid line) and Lena (dashed line). (b) Tree structure of wavelet packet basis elements θ^p_j when $\ell_0 = 2$.

operator (if it is not tight) and needs to use an iterative scheme to approximate the inverse. In this section, we will extend the shearlet frame which we constructed in the previous section to obtain a tight frame composed of directional orthonormal bases. We also show that for this extended construction, each of those directional orthonormal bases provides anisotropic window functions elongated along the direction determined by the action of shear matrix. This extended construction will be basically the same as the construction of shearlet system we described in the previous section except that we add extra basis elements to the shearlet system to obtain an orthonormal basis for each shear matrix. Based on this construction, we will then develop decomposition algorithm which explicitly decomposes a given image into directional components using orthogonal transforms. This decomposition algorithm will be directly extended from the DST and we will use definitions in IV.A.

For fixed shear parameter $s \in [-1, 1]$, consider a shearlet system Ψ_s associated with the shear matrix B^s_0 as follows:

$$\Psi_s = \{\psi^0_{js\mathbf{n}} : j \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\}.$$

For each shear parameter s, we now construct an orthonormal basis Ψ'_s containing Ψ_s such that each basis element in Ψ'_s is an anisotropic window function elongated along the direction determined by the shear parameter s. We only consider the orthonormal bases Ψ'_s associated with the shear matrices B^s_0 for details and similar arguments can be applied for the systems associated with the shear matrices B^s_1.

We first need the following facts which are immediate consequences of the basic ideas of wavelet packet basis [14].

Lemma IV.2. *For $\theta^p \in L^2(\mathbb{R})$, let $\theta^p_{j,n_1}(x_1) = 2^{j/2}\theta^p(2^j x_1 - n_1)$. Define $V_j^p = \overline{span(\theta^p_{j,n_1})_{n_1\in\mathbb{Z}}}$ for each $j, p \in \mathbb{Z}^+ \cup \{0\}$. Assume that $(\theta^p_{0,n_1})_{n_1\in\mathbb{Z}}$ is an orthonormal basis of V_0^p for a fixed p and let*

$$\theta^{2p}_{j,0} = \sum h^0(n_1)\theta^p_{j+1,n_1} \quad \textit{and} \quad \theta^{2p+1}_{j,0} = \sum g(n_1)\theta^p_{j+1,n_1}. \tag{IV.27}$$

where h^0 and g are conjugate mirror filters. Then $(\theta^{2p}_{j,n_1})_{n_1\in\mathbb{Z}}$ and $(\theta^{2p+1}_{j,n_1})_{n_1\in\mathbb{Z}}$ are orthonormal bases of V_j^{2p} and V_j^{2p+1}. Furthermore, $V_{j+1}^p = V_j^{2p} \oplus V_j^{2p+1}$.

Proof : For the proof, we refer to [14].

Corollary IV.3. *Let j be a fixed nonnegative integer. Assume that $(\theta^0_{0,n_1})_{n_1\in\mathbb{Z}}$ forms an orthonormal basis of V_0^0 and subspaces $(V^0_{j'})_{j'\in\mathbb{Z}}$ satisfy the MRA conditions. For $\ell = 1, 2, \dots$ and $p = 0, 1, \dots$, let*

$$\theta^{2p}_{j+\ell-1,0} = \sum_{n_1\in\mathbb{Z}} h^0(n_1)\theta^p_{j+\ell,n_1} \quad \textit{and} \quad \theta^{2p+1}_{j+\ell-1,0} = \sum_{n_1\in\mathbb{Z}} g(n_1)\theta^p_{j+\ell,n_1} \tag{IV.28}$$

where h^0 and g are conjugate mirror filters. Then $\{\theta^p_{j,n_1} : n_1 \in \mathbb{Z}, p = 0, 1, 2, \dots\}$ forms an orthonormal basis of $L^2(\mathbb{R})$.

Proof : We may assume that $j = 0$ without loss of generality. The recursive formula (IV.28) provides a sequence of functions $(\theta^p_{0,0})_{p=0,\dots,2^{\ell_0}-1}$ when $\ell = \ell_0 > 0$ (see Figure 6(b)). By Lemma IV.2, $(\theta^p_{0,n_1})_{n_1\in\mathbb{Z}}$ forms an orthonormal basis of V_0^p for each $p = 0, 1, \dots, 2^{\ell_0} - 1$ and $\oplus_{p=0}^{2^{\ell_0}-1} V_0^p = V_{\ell_0}^0$. Recall that MRA implies that $L^2(\mathbb{R}) = \overline{\cup_{j=0}^{\infty} V_j^0}$ and $V_j^0 \subset V_{j+1}^0$. Thus our claim is obvious since ℓ_0 is arbitrary. □

Using (IV.28), define

$$\Psi'_0 = \{\psi^p_{j0\mathbf{n}} : j, p \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\} \cup \{\phi^p_{0\mathbf{n}} : p \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\}$$

where

$$\psi^p_{j0\mathbf{n}}(\mathbf{x}) = \gamma_{j,n_1}(x_1)\theta^p_{\lfloor j/2\rfloor,n_2}(x_2) \text{ and } \phi^p_{0\mathbf{n}}(\mathbf{x}) = \theta^0_{0,n_1}(x_1)\theta^p_{0,n_1}(x_2).$$

By Corollary IV.3, we notice that $\{\psi^p_{j0\mathbf{n}} : p \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\}$ and $\{\phi^p_{0\mathbf{n}} : p \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\}$ are orthonormal bases of $W_j \otimes L^2(\mathbb{R})$ and $V_0^0 \otimes L^2(\mathbb{R})$ respectively. Therefore, Ψ'_0 is an orthonormal basis of $L^2(\mathbb{R}^2)$ and $\Psi_0 \subset \Psi'_0$. For each shear parameter $s \in [-1, 1]$, we also define $\Psi'_s = D_{B_0^{-s}}\Psi'_0$. That is,

$$\Psi'_s = \{\psi^p_{js\mathbf{n}} : j, p \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\} \cup \{\phi^p_{s\mathbf{n}} : p \in \mathbb{Z}^+ \cup \{0\}, \mathbf{n} \in \mathbb{Z}^2\}$$

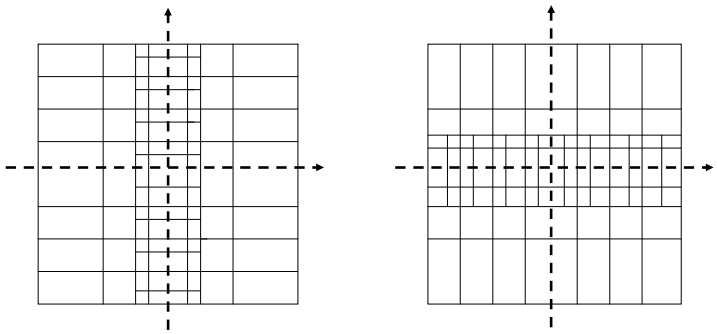

Fig. 7. Frequency tilings associated with bases Ψ'_s(left) and $\tilde{\Psi}'_s$(right) when $s = 0$.

where $\psi^p_{j\mathbf{sn}} = D_{B_0^{-s}}\psi^p_{j0\mathbf{n}}$ and $\phi^p_{\mathbf{sn}} = D_{B_0^{-s}}\phi^p_{0\mathbf{n}}$. It is clear that Ψ'_s is an orthonormal basis of $L^2(\mathbb{R}^2)$. Similarly, we define

$$\tilde{\psi}^p_{j0\mathbf{n}}(\mathbf{x}) = \gamma_{j,n_2}(x_2)\theta^p_{\lfloor j/2\rfloor,n_1}(x_1), \;\; \tilde{\phi}^p_{0\mathbf{n}} = \theta^0_{0,n_2}(x_2)\theta^p_{0,n_1}(x_1)$$

and

$$\tilde{\psi}^p_{j\mathbf{sn}} = D_{B_1^{-s}}\tilde{\psi}^p_{j0\mathbf{n}}, \;\; \tilde{\phi}^p_{s\mathrm{n}} = D_{B_1^{-s}}\tilde{\phi}^p_{0\mathbf{n}}.$$

Then

$$\tilde{\Psi}'_s = \{\tilde{\psi}^p_{j\mathbf{sn}} : j, p \in \mathbb{Z}^+ \cup \{0\},\, \mathbf{n} \in \mathbb{Z}^2\} \cup \{\tilde{\phi}^p_{\mathbf{sn}} : p \in \mathbb{Z}^+ \cup \{0\},\, \mathbf{n} \in \mathbb{Z}^2\}$$

is an orthonormal basis of $L^2(\mathbb{R}^2)$ for each $s \in [-1, 1]$. Figure 7 shows the frequency tilings obtained by bases Ψ'_s and $\tilde{\Psi}'_s$ when $s = 0$.

Next, we derive a fast algorithm to compute the extended shearlet coefficients $\langle f, \psi_\mu\rangle$ and $\langle f, \tilde{\psi}_{\mu'}\rangle$ for $\psi_\mu \in \Psi'_\mu$ and $\tilde{\psi}_{\mu'} \in \tilde{\Psi}'_s$. For $p \in \mathbb{Z}^+ \cup \{0\}$, consider binary expansion of p so that $p = \sum_{k=0}^{j-1} \epsilon_k 2^k$ with $\epsilon_k \in \{0, 1\}$. Let h^p_j be the Fourier coefficients of $\prod_{k=0}^{j-1} H^{\epsilon_k}(2^k\omega_1)$ where

$$H^0(\omega_1) = \sum_{n_1\in\mathbb{Z}} h^0(n_1)e^{-2\pi\omega_1 n_1} \quad \text{and} \quad H^1(\omega_1) = \sum_{n_1\in\mathbb{Z}} g(n_1)e^{-2\pi\omega_1 n_1}.$$

For $p = 0, 1, \ldots, 2^{j_2} - 1$ and $i = 0, 1$, define linear maps $\mathcal{W}^{p,i}_{j_1,j_2} : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)$ and $\widetilde{\mathcal{W}}^{p,i}_{j_1,j_2} : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)$ as follows.

$$\begin{cases} \mathcal{W}^{p,i}_{j_1,j_2}(f)(\mathbf{n}) = \sum_{m_1}\sum_{m_2} f(m_1, m_2)g^i_{j_1}(m_1 - 2^{j_1}n_1)h^p_{j_2}(m_2 - 2^{j_2}n_2) \\ \widetilde{\mathcal{W}}^{p,i}_{j_1,j_2}(f)(\mathbf{n}) = \sum_{m_1}\sum_{m_2} f(m_1, m_2)g^i_{j_1}(m_2 - 2^{j_1}n_2)h^p_{j_2}(m_1 - 2^{j_2}n_1) \end{cases}$$

for $f \in \ell^2(\mathbb{Z}^2)$, where the filter coefficients $g^1_{j_1} = g_{j_1}$(highpass filter) and $g^0_{j_1} = h^0_{j_1}$(lowpass filter). We assume (IV.16) for $f \in V_J \otimes V_J$ with positive integer J. Using the same arguments as in the previous section, one can show that for $j = 0, 1, \ldots, (J-1)$ and $-2^{\lceil j/2 \rceil} \le k \le 2^{\lceil j/2 \rceil}$,

$$\langle f, \psi^p_{js_{jk}\mathbf{n}} \rangle = \mathcal{W}^{p,1}_{J-j,J-\lfloor j/2 \rfloor}\Big(\big([((f_J)_{\uparrow 2^{\lceil j/2 \rceil}} *_1 h^0_{\lceil j/2 \rceil})(B_0^{-k}\cdot) * \Phi_k] *_1 \overline{h^0}_{\lceil j/2 \rceil} \big)_{\downarrow 2^{\lceil j/2 \rceil}} \Big)(\mathbf{n})$$

and this indicates that

$$\langle f, \psi^p_{js_{jk}\mathbf{n}} \rangle \approx \mathcal{W}^{p,1}_{J-j,J-\lfloor j/2 \rfloor}(P_{s_{j,k}}(f_J))(\mathbf{n}).$$

Similarly, we have

$$\langle f, \tilde{\psi}^p_{js_{jk}\mathbf{n}} \rangle \approx \widetilde{\mathcal{W}}^{p,1}_{J-j,J-\lfloor j/2 \rfloor}(\tilde{P}_{s_{j,k}}(f_J))(\mathbf{n})$$

We are now ready to define the extended DST $\hat{S} : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)$. We first let

$$\mathcal{W}_{j_1,j_2} = (\mathcal{W}^{p,1}_{j_1,j_2})_{p=0,\ldots,2^{j_2}-1} \quad \text{and} \quad \widetilde{\mathcal{W}}_{j_1,j_2} = \widetilde{(\mathcal{W}}^{p,1}_{j_1,j_2})_{p=0,\ldots,2^{j_2}-1}.$$

For $M \in \mathbb{Z}^+$, let us define

$$\hat{S} = (\hat{S}_{-2^M}, \ldots, \hat{S}_{2^M}, \hat{\tilde{S}}_{-2^M}, \ldots, \hat{\tilde{S}}_{2^M})$$

where

$$\hat{S}_k = \Big((\mathcal{W}^{p,0}_{J,J}((P_{k/2^M})))^{2^J-1}_{p=0}, (\mathcal{W}_{J-j,J-\lfloor j/2 \rfloor}(P_{k/2^M}))^{J-1}_{j=0} \Big)$$

and

$$\hat{\tilde{S}}_k = \Big((\widetilde{\mathcal{W}}^{p,0}_{J,J}(\tilde{P}_{k/2^M}))^{2^J-1}_{p=0}, (\widetilde{\mathcal{W}}_{J-j,J-\lfloor j/2 \rfloor}(\tilde{P}_{k/2^M}))^{J-1}_{j=0} \Big)$$

for $-2^M \le k \le 2^M$. Basically, each subtransform $\hat{S}_k$ or $\hat{\tilde{S}}_k$ consists of the discrete shear transform $P_{k/2^M}$ or $\tilde{P}_{k/2^M}$ and the anisotropic discrete wavelet transform (ADWT) $\mathcal{W}_{j1,j2}$ (horizontal direction) or $\widetilde{\mathcal{W}}_{j1,j2}$ (vertical direction). Figure 8 summarizes our extended DST.

Finally, It is clear that each $\hat{S}_k$ and $\hat{\tilde{S}}_k$ are orthogonal transforms. Therefore, the extended DST S consists of those $2^{M+2}+2$ orthogonal transforms and $2^{M+2}+2$ is the number of directions. This implies that the inverse DST is simply the adjoint of S with normalization. For $f_J \in \ell^2(\mathbb{Z}^2)$, we have

$$\frac{1}{2^{M+2}+2} \hat{S}^*(\hat{S}(f_J)) = f_J. \tag{IV.29}$$

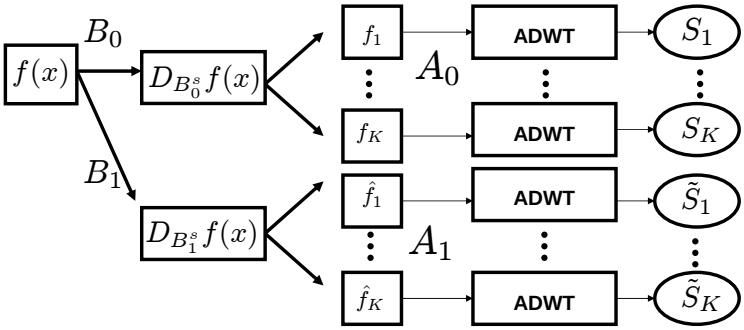

Fig. 8. Block diagram of the extended DST

V. Image Approximations Using DST

In this section, we introduce an adaptive image approximation scheme using the extended DST that we described in the previous section. Suppose that $\mathcal{B}_k$ are orthonormal bases associated with orthogonal transforms $\hat{S}_{k-2^M-1}$ for $k = 1, \ldots, 2^{M+1}+1$ and associated with orthogonal transforms $\hat{\tilde{S}}_{k-3(2^M)-2}$ for $k = 2^{M+1}+2, \ldots, 2^{M+2}+2$. The main idea of our adaptive approximation scheme is similar to the Matching Pursuit (MP) introduced by Mallat and Zhong [15]. The matching pursuit selects vectors one by one from a given basis dictionary, while optimizing the signal approximation at each step. On the other hand, our approximation scheme selects the best basis so that it provides the best approximation with P nonzero terms among all of the bases $\mathcal{B}_k$ at each step. Let

$$\mathcal{B} = \bigcup_{k=1}^{2^{M+2}+2} \mathcal{B}_k.$$

and this is a basis dictionary consisting of $2^{M+2}+2$ orthonormal bases. Observe that for $f \in \ell^2(\mathbb{Z}^2)$, each subtransform $\hat{S}_k$ or $\hat{\tilde{S}}_k$ of the DST $\hat{S}$ can be written as $(\langle f, \varphi_\gamma^k \rangle)_{\gamma \in \Gamma}$ where $\varphi_\gamma^k \in \mathcal{B}_k$.

The best P term approximation for given signal $f \in \ell^2(\mathbb{Z}^2)$ with respect to basis $\mathcal{B}_k$ is

$$Q_{k,P}(f) = \sum_{\gamma \in I_P^k} \langle f, \varphi_\gamma^k \rangle \varphi_\gamma^k \quad \text{where} \quad \varphi_\gamma^k \in \mathcal{B}_k.$$

Here, I_P^k is an index set of P largest coefficients $\langle f, \varphi_\gamma^k \rangle$. In this case, the residual is

$$R(f) = f - Q_{k,P}(f).$$

Since $R(f)$ is orthogonal to $Q_{k,P}$, we have

$$\|f\|^2 = \|R(f)\|^2 + \|Q_{k,P}(f)\|^2$$

and

$$\|Q_{k,P}(f)\|^2 = \sum_{\gamma \in I_P^k} |\langle f, \varphi_\gamma^k \rangle|^2$$

To minimize $\|R(f)\|$, we must choose the best orthogonal projection $Q_{k,P}$ such that $\|Q_{k,P}(f)\|$ is maximum among $\|Q_{\ell,P}(f)\|$ for $\ell = 1, \ldots, 2^{M+2}+2$. Let $R^0(f) = f$. Suppose that the mth order residue $R^m(f)$ is already computed for $m \geq 0$. The next iteration chooses the index $k_m \in \{1, \ldots, 2^{M+2}+2\}$ such that

$$k_m = \text{argsup}_k \sum_{\gamma \in I_P^k} |\langle R^m(f), \varphi_\gamma^k \rangle|^2$$

where $\varphi_\gamma^k \in \mathcal{B}_k$ for each $k = 1, \ldots, 2^{M+2}+2$ and

$$R^m(f) = Q_{k_m,P}(R^m(f)) + R^{m+1}(f) \tag{V.30}$$

The orthogonality of $R^{m+1}(f)$ and $Q_{k_m,P}(R^m(f))$ implies

$$\|R^m(f)\|^2 = \|Q_{k_m,P}(R^m(f))\|^2 + \|R^{m+1}(f)\|^2 \tag{V.31}$$

Summing (IV.30) from m between 0 and $K-1$ yields

$$f = \sum_{m=0}^{K-1} Q_{k_m,P}(R^m(f)) + R^K(f). \tag{V.32}$$

Similarly, summing (IV.31) from m between 0 and $K-1$ gives

$$\|f\|^2 = \sum_{m=0}^{K-1} \|Q_{k_m,P}(R^m(f))\|^2 + \|R^K(f)\|^2. \tag{V.33}$$

It is easy to show that $\|R^m\|$ converges exponentially to 0 when m tends to infinity.

Our adaptive approximation scheme using the DST is implemented with the following steps.

1. **Initialization :** Set $m = 0$ and compute the extended DST ; $\hat{S}(R^m(f))$

2. **Best match :** Find the best orthogonal projection $Q_{k_m,P}$ such that

$$k_m = \text{argsup}_k \|Q_{k,P}(R^m(f))\|$$

3. **Update :** Compute $m+1$ th order residual $R^{m+1}(f)$;

$$R^{m+1}(f) = R^m(f) - Q_{k_m,P}(R^m(f)).$$

4. **Stopping rule :** If

$$\|R^{m+1}(f)\|^2 = \|R^m(f)\|^2 - \|Q_{k_m,P}(R^m(f))\|^2 \leq \epsilon^2 \|f\|^2$$

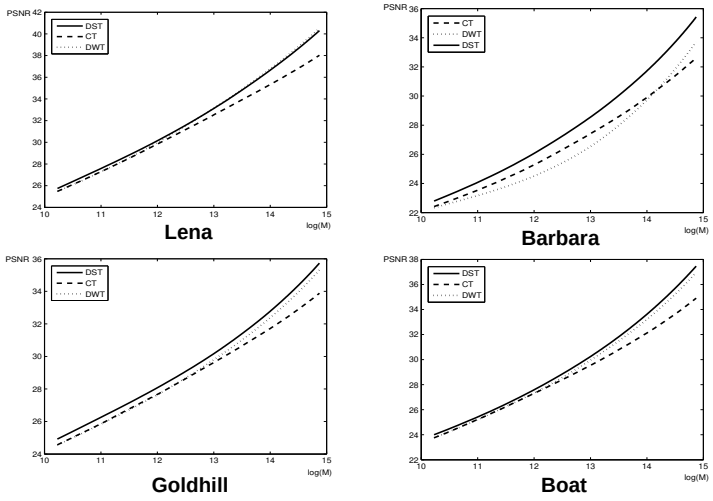

Fig. 9. Nonlinear approximation results on different test images: From top left, clockwise: 'Lena', 'Barbara', 'boat' and 'goldhill'. Horizontal axis : the number of nonzero coefficients in log scale and Vertical axis : PSNR

then stop. Otherwise m = m+1 and go to 1.

In this case K is the number of iterations and P nonzero coefficients are added at each iteration. Therefore, KP coefficients are produced after K iterations in the MP step. Also we note that our extended DST requires $O((2^{M+2}+2)N)$ operations where N is the size of input image. Therefore, the total computation cost of this adaptive approximation scheme is $O((2^{M+2}+2)NK)$ and this could be highly expensive depending on the choice of M and K. However, our tests indicate that only a few iterations and directions can still provide good approximations. For numerical tests, we compare the performance of our adaptive approximation scheme using the extended DST to other non adaptive transforms such as the discrete biorthogonal CDF 9/7 wavelet transform (DWT)[16] and contourlet transform (CT)[5] in image compression (see Figures 9 and 10). We used 5 level decomposition and only two directions (vertical and horizontal) and applied one iteration in the MP step. In this case, the running time of our proposed scheme is competitive to the standard DWT while it outperforms other transforms. In fact, with these choices, our adaptive approximation MATLAB routine takes 0.65 seconds while the DWT MATLAB routine (available at www-stat.stanford.edu/wavelab/) takes 0.51 seconds (see Figure 11 (a)). Finally, we note that one can achieve significant improvement over other transforms by increasing the number of iterations but this will increase computation cost as well.

Fig. 10. Compression results of 'Barbara' image of size 512×512: The image is reconstructed from 5024 most significant coefficients. (a) Zoomed original image (b) Zoomed image reconstructed by the DWT (PSNR = 25.01) (c) Zoomed image reconstructed by the CT (PSNR = 25.87) (d) Zoomed image reconstructed by the DST (PSNR = 26.73).

(a)

Transform (Forward+Inverse)	Redundancy	Speed
DST	2 (5 level decomposition)	0.65 sec
DWT (9/7 CDF)	1 (4 level decomposition)	0.51 sec

(b)

Transform	Redundancy	Speed
DST 10	10 directions X mirror extensions (2) = 20	14.17 sec
DST 18	18 directions X 2 shifts X (2) = 72	40.93 sec
TIWT	64 shifts = 64	14.15 sec
NSCT	16 + 16 + 8 + 8 + 4 directions +1 = 53	1227.87 sec

Fig. 11. Running time of DST and DWT for image approximations and denoising on desktop with 2.53GHz, 1.98 GB of RAM. (a) Image approximations and (b) denoising

TABLE I

		DST10	DST18	NSCT	TIWT	DST10	DST18	NSCT	TIWT
σ	Noisy	Lena				Boat			
$\sigma = 10$	28.15dB	35.75dB	35.96dB	35.47dB	35.16dB	34.22dB	34.42dB	33.67dB	33.64dB
$\sigma = 20$	22.14dB	32.28dB	32.51dB	32.32dB	31.54dB	30.49dB	30.71dB	30.34dB	29.85dB
$\sigma = 30$	18.60dB	30.33dB	30.59dB	30.53dB	29.51dB	28.48dB	28.69dB	28.60dB	27.78dB
$\sigma = 40$	16.11dB	28.87dB	29.15dB	29.12dB	27.93dB	27.21dB	27.44dB	27.39dB	26.43dB
$\sigma = 50$	14.17dB	27.82dB	28.11dB	28.09dB	26.76dB	26.21dB	26.45dB	26.50dB	25.33dB
σ	Noisy	Peppers				Goldhill			
$\sigma = 10$	28.15dB	34.45dB	35.67dB	35.23dB	35.20dB	33.30dB	33.47dB	32.76dB	32.64dB
$\sigma = 20$	22.14dB	32.21dB	32.44dB	32.28dB	31.83dB	30.09dB	30.27dB	29.90dB	29.33dB
$\sigma = 30$	18.60dB	30.23dB	30.48dB	30.48dB	29.65dB	28.41dB	28.60dB	28.44dB	27.57dB
$\sigma = 40$	16.11dB	28.85dB	29.11dB	29.11dB	28.00dB	27.23dB	27.43dB	27.38dB	26.26dB
$\sigma = 50$	14.17dB	27.78dB	28.06dB	28.02dB	26.72dB	26.41dB	26.61dB	26.61dB	25.39dB

Fig. 12. Comparison of various transforms in denoising

VI. Image Denoising Using DST

In order to illustrate the potential of the DST, we study additive white Gaussian noise removal by means of hard thresholding estimator. We want to estimate signals that are considered to be realization of real random vector f. The noisy measurements are

$$g = f + W$$

where W is a zero-mean Gaussian white noise of variance σ^2.

For our numerical tests. We compared the performance of the extended DST to other transforms such as the Nonsubsampled Contourlet Transform (NSCT)[17] and Translation Invariant Wavelet Transform

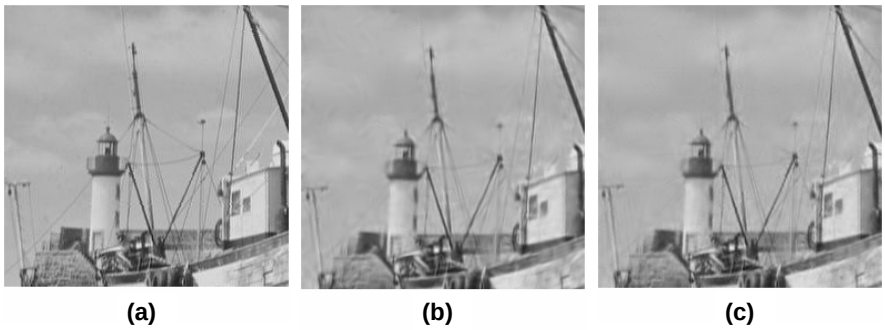

Fig. 13. Comparison between the DST and NSCT for denoising Gaussian noise with $\sigma = 15$. (a) Original Boat image. (b) Denoised with NSCT [17], PSNR = 31.66dB. (c) Denoised with the DST18, PSNR = 32.17dB

(TIWT)[18] in this denoising task. To highlight the performance of the extended DST relative to other transforms, we apply hard threshold on the coefficients of various transforms. We choose the threshold

$$T_j = K_j \sigma$$

where j is the index of decomposition level, K_j are constants experimentally determined so that they can produce overall optimal performance for each transform and σ is the variance of noise. We tested two types of the extended DST for which one uses 18 directions (DST 18) and the other one uses 10 directions (DST 10). Also, in order to avoid artificial discontinuities due to shear operations and periodization, we applied the DST to a mirror extended image. In this case, we used a mirror extension along the horizontal or vertical axis depending on the choice of shear matrices B_0 and B_1 respectively. Table I shows the PSNR results for various transforms and noise intensities. The results show the DST is consistently superior to the TIWT and NSCT in PSNR measure while the running time of our proposed transform is quite competitive to other transforms.

VII. Conclusion

We have constructed compactly supported shearlet system which can provide efficient directional image representation. We also have developed the two types of fast discrete implementations of shearlets called the DST and the extended DST. Especially the extended DST is simply applying shear transforms on the discrete domain followed by the anisotropic wavelet transforms. Thus, the extended DST is implemented by applying finitely many orthogonal transforms and the inverse transform is straightforward.

Application of our proposed transform in image approximations and denoising were studied. In image approximations, the results obtained with our adaptive image representation using the extended DST are

consistently superior to other transforms such as the DWT and CT both visually and with respect to PSNR. In denoising, we studied the performance of the extended DST coupled with a hard thresholding estimator. Our results indicate that the DST consistently outperforms other competing transforms such as the NSCT and TIWT in PSNR.

One of the main features of our DST is that it can exactly compute the shearlet coefficients based on a MRA (see IV.23). In fact, shear operations preserve the structure of integer lattice and this allows them to be faithfully adapted for discrete framework such that the DST is naturally derived from the construction of shearlet frame. Furthermore, it can be shown that our compactly supported shearlet frame generated by separable wavelets can provide nearly optimal approximation for a piecewise smooth function. In other words, one can achieve that

$$\|f - f_M\|^2 \leq CM^{-2}(\log M)^3 \qquad \text{as} \quad M \to \infty$$

where f is C^2 smooth function except at points lying on C^2 curves and f_M is shearlet nonlinear approximation obtained by taking the M largest shearlet coefficients. This approximation rate is same as one obtained with curvelets or band-limited shearelts and the proof of this claim will appear in elsewhere [21]. This indicates that our new construction scheme of shearlets provides sparse approximation as well as efficient implementation which explicitly computes the shearlet coefficients.

However, we should point out that there are basically two issues that need to be resolved concerning our proposed construction of shearlets. First of all, our proposed DST does not share tree structure based on filter banks like other multiscale transforms such as DWT and CT. Therefore, computation cost increases as we increase the number of shear operations. Secondly, the ratio of the frame bounds for our construction depends on the choice of the sampling constant c. Especially, this implies that one needs to increase the redundancy of shearlet frame in order to achieve more reasonable frame bounds.

Thus, for future study, we will focus on improving construction of shearlets such that the resulting construction can provide more structured discrete transform as well as less redundant shearlet frame.

Acknowledgements

The author would like to thank G. Kutyniok and P. Kittipoom for useful discussions. The author also would like to point out that Proposition III.2 is a special case of one of main theorems in [20].

APPENDIX A
PROOF OF PROPOSITION III.2

Without loss of generality, we may assume that $1 \leq |\omega_1| \leq a^q$ and $\omega_2 \in \widehat{\mathbb{R}}$. Also, we let $\omega = (\omega_1, \omega_2) \in \widehat{\mathbb{R}}^2$ and $s = (s_1, s_2) \in \widehat{\mathbb{R}}^2$. We first observe that

$$\operatorname*{ess\,sup}_{\omega_1 \in \widehat{\mathbb{R}}} \sum_{j \in \mathbb{Z}} \frac{(a^{qj}|\omega_1|)^\alpha}{(1 + |a^{qj}\omega_1|^2)^{\beta/2}} < \infty \quad \text{with } \beta > \alpha. \tag{A.34}$$

(For the proof, see [19].) Also, it is easy to see that

$$\sup_x \sum_{k \in \mathbb{Z}} \frac{1}{((x+k)^2 + 1)^{\rho/2}} < \infty \quad \text{if } \rho > 1. \tag{A.35}$$

Notice that we may assume that for $s_1 \in \widehat{\mathbb{R}}$, $a^{qj}\omega_1 \neq -s_1$ for all $j \in \mathbb{Z}$ since a set

$$\{(\omega_1, \omega_2) \in \widehat{\mathbb{R}}^2 \,:\, a^{qj}\omega_1 = -s_1 \text{ for some } j \in \mathbb{Z}\}$$

is a set of measure 0. Let us estimate $\Phi(s)$. Direct computation gives

$$\begin{aligned}
\sum_{j,k \in \mathbb{Z}} |\hat{\psi}_0^i((B_0^T)^k A_0^j \omega)\hat{\psi}_0^i((B_0^T)^k A_0^j \omega + s)| &\leq C_1 \sum_{j \in \mathbb{Z}} \frac{a^{q\alpha j}|\omega_1|^\alpha}{(1 + |a^{qj}\omega_1|^2)^{\gamma/2}} \frac{|a^{qj}\omega_1 + s_1|^\alpha}{(1 + |a^{qj}\omega_1 + s_1|^2)^{\gamma/2}} \\
&\times \sum_{k \in \mathbb{Z}} \Big(1 + |a^{j(1/2-q)}\frac{\omega_2}{\omega_1} + k|^2\Big)^{-\frac{\gamma}{2}} \\
&\times \Big(1 + \Big|\frac{a^{j/2}\omega_2 + k a^{qj}\omega_1 + s_2}{a^{qj}\omega_1 + s_1}\Big|^2\Big)^{-\gamma/2} \\
&= \sum_{|a^{qj}\omega_1| > 1/2\|s\|_\infty} \cdots + \sum_{|a^{qj}\omega_1| \leq 1/2\|s\|_\infty} \cdots \\
&= E_1 + E_2.
\end{aligned}$$

where $\|s\|_\infty = \max(|s_1|, |s_1|)$. By (A.34) and (A.35), we have

$$E_1 \leq C_1' \sum_{|a^{qj}\omega_1| > 1/2\|s\|_\infty} (a^{qj}|\omega_1|)^{\alpha - \gamma}$$

where $\alpha - \gamma < 0$ by assumption. In fact, one can show that

$$\sum_{|a^{qj}\omega_1| > 1/2\|s\|_\infty} (a^{qj}|\omega_1|)^{\alpha-\gamma} \leq C_2 \|s\|_\infty^{\alpha-\gamma} \tag{A.36}$$

(For the proof, see [19].)

Next, we estimate E_2. First assume that $\|s\|_\infty = s_1$. By (A.35), we obtain

$$E_2 \leq C_3 \sum_{|a^{qj}\omega_1| \leq 1/2\|s\|_\infty} \frac{a^{q\alpha j}|\omega_1|^\alpha}{(1 + |a^{qj}\omega_1|^2)^{\gamma/2}} \frac{|a^{qj}\omega_1 + s_1|^\alpha}{(1 + |a^{qj}\omega_1 + s_1|^2)^{\gamma/2}}$$

Note that $|a^{qj}\omega_1 + s_1| \geq |s_1| - |a^{qj}\omega_1| \geq 1/2|s_1|$ and $|a^{qj}\omega_1 + s_1| \leq 3/2|s_1|$. Thus, by (A.34),

$$E_2 \leq C_4 \sum_{j\in\mathbb{Z}} \frac{a^{q\alpha j}|\omega_1|^\alpha}{(1+|a^{qj}\omega_1|^2)^{\gamma/2}} |s_1|^{\alpha-\gamma} \leq C_5 \|s\|_\infty^{\alpha-\gamma}. \tag{A.37}$$

Finally, assume that $\|s\|_\infty = |s_2|$, then we have

$$\begin{aligned} E_2 &\leq C_5 \sum_{|a^{qj}\omega_1|\leq 1/2\|s\|_\infty} \frac{a^{q\alpha j}|\omega_1|^\alpha}{(1+|a^{qj}\omega_1|^2)^{\gamma/2}} \frac{|a^{qj}\omega_1 + s_1|^\alpha}{(1+|a^{qj}\omega_1+s_1|^2)^{\gamma/2}} \\ &\times \sum_{k\in\mathbb{Z}} \Big(1+|a^{j(1/2-q)}\frac{\omega_2}{\omega_1}+k|^2\Big)^{-\frac{\gamma}{4}} \Big(1+|a^{j(1/2-q)}\frac{\omega_2}{\omega_1}+k|^2\Big)^{-\frac{\gamma}{4}} \\ &\times \frac{|1+a^{-qj}\omega_1^{-1}s_1|^{\gamma/2}}{\Big(|1+a^{-qj}\omega_1^{-1}s_1|^2 + |k+a^{j(1/2-q)}\frac{\omega_2}{\omega_1} + \frac{s_2 a^{-qj}}{\omega_1}|^2\Big)^{\gamma/4}} \end{aligned}$$

Observe that

$$\frac{z^2}{(1+x^2)(z^2+(x+y)^2)} \leq C_6 \text{max}(z^2,1)\frac{1}{y^2} \quad y \neq 0.$$

Assume that $z^2 = |1+a^{-qj}\omega_1^{-1}s_1|^2 \geq 1$. By (A.34) and (A.35), we have

$$\begin{aligned} E_2 &\leq C_7 \sum_{|a^{qj}\omega_1|\leq 1/2\|s\|_\infty} \frac{a^{q\alpha j}|\omega_1|^\alpha}{(1+|a^{qj}\omega_1|^2)^{\gamma/2}} \frac{|a^{qj}\omega_1 + s_1|^\alpha}{(1+|a^{qj}\omega_1+s_1|^2)^{\gamma/2}} \\ &\times \frac{|a^{qj}\omega_1+s_1|^{\gamma/2}}{|s_2|^{\gamma/2}} \leq C_8 \|s\|_\infty^{\alpha-\gamma/2} \end{aligned} \tag{A.38}$$

If $z^2 = |1+a^{-qj}\omega_1^{-1}s_1|^2 \leq 1$ one can also similarly estimate E_2 to obtain (A.38). By (A.36),(A.37)and (A.38), we have

$$\sum_{s\in\mathbb{Z}^2\setminus\{0\}} \sqrt{\Phi\Big(\frac{s}{c}\Big)\Phi\Big(-\frac{s}{c}\Big)} \leq C_8 \Big(\sum_{m\in\mathbb{Z}^2\setminus 0} \|m\|_\infty^{\alpha-\gamma/2}\Big) c^{\alpha-\gamma/2}$$

and $\sum_{m\in\mathbb{Z}^2\setminus 0}\|m\|_\infty^{\alpha-\gamma/2} < \infty$ when $\gamma/2-\alpha > 2$. Therefore, for any $D_1 > 0$, there exists $c_0 > 0$ such that

$$D_1 - \sum_{s\in\mathbb{Z}^2\setminus\{0\}} \sqrt{\Phi\Big(\frac{s}{c}\Big)\Phi\Big(-\frac{s}{c}\Big)} > 0$$

for all $c \leq c_0$. This proves the condition (III.6) in Theorem III.1. Also, $\gamma - \alpha > 0$ and $\alpha > 0$. The inequality (A.35) implies that for $1 \leq |\omega_1| \leq a^q$ and $\omega_2 \in \mathbb{R}$, we have that

$$\begin{aligned} \sum_{i=1}^L \sum_{j,k\in\mathbb{Z}} |\hat{\psi}_0^i((B_0^T)^k A_0^j\omega)|^2 &\leq K \sum_{j\in\mathbb{Z}} \frac{|a^{qj}\omega_1|^{2\alpha}}{(1+|a^{qj}\omega_1|^2)^{-\gamma}} \sum_{k\in\mathbb{Z}} \Big(1+|a^{j(1/2-q)}\frac{\omega_2}{\omega_1}+k|^2\Big)^{-\gamma} \\ &\leq K \sum_{j\in\mathbb{Z}} |a^{qj}\omega_1|^{2\alpha}(1+|a^{qj}\omega_1|^2)^{-\gamma} \\ \leq K\Big(\sum_{j=0}^\infty |a^{q(-j+1)}|^{2\alpha} + \sum_{j=1}^\infty (1+|a^{qj}|^2)^{-(\gamma-\alpha)}\Big) &< \infty. \end{aligned}$$

This implies the upper bound condition in (III.5) in Theorem III.1. □

APPENDIX B
PROOF OF LEMMA IV.1

We only prove (IV.18) and (IV.17) is proved similarly. Taking the Fourier transform on (IV.13) gives

$$\hat{\theta}^0(\omega_1)\frac{1}{\sqrt{2}}H^0\Big(\frac{\omega_1}{2}\Big)\hat{\theta}^0\Big(\frac{\omega_1}{2}\Big).$$

Then direct computation gives

$$\begin{aligned}
\widehat{D_{2^{-j}}T_{2^{j-J}n_1}\theta^0}(\omega_1) &= 2^{-j/2}\hat{\theta}^0\Big(\frac{\omega_1}{2^j}\Big)e^{-i2\pi\frac{n_1}{2^J}\omega_1} \\
&= 2^{-\frac{j+1}{2}}H^0\Big(\frac{\omega_1}{2^{j+1}}\Big)\hat{\theta}^0\Big(\frac{\omega_1}{2^{j+1}}\Big)e^{-i2\pi\frac{n_1}{2^J}\omega_1} \\
&= 2^{-J/2}H^0\Big(\frac{\omega_1}{2^{j+1}}\Big)\cdots H^0\Big(\frac{\omega_1}{2^J}\Big)\hat{\theta}^0\Big(\frac{\omega_1}{2^J}\Big)e^{-i2\pi\frac{n_1}{2^J}\omega_1} \\
&= \prod_{k=0}^{J-j-1} H^0\Big(2^k\frac{\omega_1}{2^J}\Big)2^{-J/2}\hat{\theta}^0\Big(\frac{\omega_1}{2^J}\Big)e^{-i2\pi\frac{n_1}{2^J}\omega_1}
\end{aligned}$$

Thus we have

$$\widehat{D_{2^{-j}}T_{2^{j-J}n_1}\theta^0}(\omega_1) = H^0_{J-j}\Big(\frac{\omega_1}{2^J}\Big)2^{-J/2}\hat{\theta}^0\Big(\frac{\omega_1}{2^J}\Big)e^{-i2\pi\frac{n_1}{2^J}\omega_1} \tag{B.39}$$

Taking an inverse Fourier transform on (B.49) gives

$$D_{2^{-j}}T_{2^{j-J}n_1}\theta^0(x_1) = \sum_{m_1} h^0_{J-j}(m_1)D_{2^{-J}}T_{n_1+m_1}\theta^0(x_1).$$

This implies (IV.18). □

REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.

[2] D. Donoho, M. Vetterli, R. Devore and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inform. Theory*, 44, 2435-2476, 1998.

[3] E. Pennec and S. Mallat, "Sparse geometric image representation with bandelets," *IEEE Trans. Image Process.,* vol. 14, no. 4, pp 423-438, Apr. 2005.

[4] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli and P.L. Dragotti, " Directionlets: Anisotropic multidirectional representation with separable filtering," *IEEE Trans. Image Process.*, Vol. 15, Nr. 7, pp. 1916-1933, 2006.

[5] M. Do and M. Vetterli, " The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091-2106, Dec. 2005.

[6] I. W. Selesnick, R. G. Baraniuk, and N. Kingsbury, " The dual-tree complex wavelet transform - A coherent framework for multiscale signal and image processing," *IEEE Signal Processing Magazine*, 22(6):123-151, November 2005.

[7] E. Candes and D. Donoho, " New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Commun. Pure Appl. MAth*, vol. 57, no. 2, pp. 219-266, Feb. 2004.

[8] G. Peyre and S. Mallat, " Discrete bandelets with geometric orthogonal filters," *Proceedings of ICIP,* Sept. 2005.

[9] K. Guo, D. Labate, W. Lim, G. Weiss, and E. Wilson, " Wavelets with composite dilations and their MRA properties", *Appl. Comput. Harmon. Anal.*, 20 , pp. 231-249 (2006).

[10] D. Labate, W. Lim, G. Kutyniok and G. Weiss, and " Sparse multidimensional representation using shearlets", Wavelets XI (San Diego, CA, 2005), 254-262, SPIE Proc. 5914 , SPIE, Bellingham, WA, 2005.

[11] K. Guo and D. Labate, " Optimally sparse multidimensional representation using shearlets,", *SIAM J Math. Anal.*, 39 pp. 298-318, 2007.

[12] K. Guo, D. Labate and W. Lim, " Edge analysis and identification using the continuous shearlet transform", to appear in *Appl. Comput. Harmon. Anal*.

[13] G. Kutyniok and D. Labate, " The construction of regular and irregular shearlet frames", *J. Wavelet Theory Appl.*, 1, pp. 1-10, 2007.

[14] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, " Wavelet analysis and signal processing"', *Wavelets and their Applications*, pp.153-178, 1992.

[15] S. Mallat and S. Zhang, " Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, pp. 3397-3415, Dec. 1993.

[16] A. Cohen, I. Daubechies and J. Feauveau, " Biorthogonal bases of compactly supported wavelets," *Commun. on Pure and Appl. Math.*, 45:485-560, 1992.

[17] A. Cunha, J. Zhou, and M. Do, " The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089-3101, Oct. 2006.

[18] R. Coifman and D. Donoho, " Translation invariant de-noising," *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York: SPringer-Verlag, pp. 125-150, 1995.

[19] D. Yang and X. Zhou, "Irregular wavelet frames on $L^2(\mathbb{R}^n)$", *Science in China Ser. A Mathematics*, pp. 277-287, 2005.

[20] P. Kittipoom, G. Kutyniok and W. Lim, "Construction of Compactly Supported Shearlets", preprint.

[21] G. Kutyniok and W. Lim, "Compactly Supported Shearlets are Optimally Sparse", preprint.

PLACE
PHOTO
HERE