Mathematical Methods in Machine Learning

Wojciech Czaja

UMD, Spring 2016

Norbert Wiener Center
for Harmonic Analysis and Applications

Outline

Anisotropic Harmonic Analysis

- Harmonic analysis decomposes signals into simpler elements called *analyzing functions*.
- Classical HA methods include Fourier series and aforementioned wavelets. These have proven extremely influential and quite effective for many applications.
- However, they are fundamentally isotropic, meaning they decompose signals without considering how the signal varies directionally.
- Wavelets decompose an image signal with respect to translation and scale. Since the early 2000s, there have been several attempts to incorporate directionality into the wavelet construction.

Multiscale Directional Representations

- Early attempts to make wavelets more sensitive to directionality included appropriate filter design, anisotropic scaling, steerable filters, and similar techniques.
- **Directional wavelets:** J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. Ali introduced more complicated group actions for parametrization of 2-dimesnional wavelet transforms, including rotations or similitude group. These results were later generalized to construct wavelets on sphere and other manifolds.

J.-P. Antoine, D. Rosca, P. Vandergheynst,"Wavelet transform on manifolds: old and new approaches", ACHA, 2010, Vol. 28 (2),189–202.

- Subsequently **Radon transform** has been introduced in combination with wavelet transforms to replace the angular parametrization; This results in systems such as **ridgelets** (E. Candès and D. Donoho) or **Gabor ridge functions** (L. Grafakos and C. Sansing)
- **Contourlets:** M. Do and M. Vetterli constructed a discrete-domain multiresolution and multidirection expansion using non-separable filter banks, in much the same way that wavelets were derived from filter banks.

Multiscale Directional Representations

- **Curvelets:** E. Candès and D. L. Donoho introduced the curvelets as an efficient tool to extract directional information from images. Curvelets consist of translations and rotations of a sequence of basic functions depending on a parabolic scaling parameter. The curvelet transform is first developed in the continuous domain and then discretized for sampled data.
- **Wavelets with Composite Dilations:** K. Guo, D. Labate, W.-Q. Lim, B. Manning, G. Weiss, and E. Wilson studied affine systems built by using a composition of two sets of matrices as the dilation.
- **Shearlets:** D. Labate, K. Guo, G. Kutyniok, and G. Weiss introduced a special example of the Composite Dilation Wavelets.
- **Surfacelets** (Do, Lu), **bandlets** (Le Pennec, Mallat), **brushlets** (Meyer, Coifman), **wedgelets** (Donoho), **phaselets** (Gopinath), **complex wavelets** (Daubechies), **surflets** (Baraniuk), etc etc...

Anisotropic Harmonic Analysis

- These constructions incorporate directionality in a variety of ways.
- To summarize, some of the major constructions include:
  - Ridgelets.

E. Candès. *Ridgelets: theory and applications*. PhD thesis. (1998).

  - Curvelets.

D. Donoho and E. Candès. *Curvelets: A surprisingly effective nonadaptive representation for objects with edges*. Curve and Surface Fitting. (1999).

  - Contourlets.

M. Do and M. Vetterli. *Contourlets*. Beyond Wavelets. (2001).

  - Shearlets.

D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss. *Sparse multidimensional representation using shearlets*. Proc. SPIE 5914. (2005).

- Wavelets, ridgelets, curvelets, and shearlets are surprisingly related, as they all are special cases of the recently introduced $\alpha-$molecules.

P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer. $\alpha-$*molecules*. arXiv: 1407.4424. (2014).

Multiscale Directional Representations

- Many of the aforementioned representations were designed specifically for dealing with images, i.e., for the case of 2-dimensional Euclidean space.
- Multiscale directional representations can also be constructed analogously for higher dimensional spaces, as well as for some manifolds.
- A different approach is needed to deal with discrete structures, such as graphs, networks, or point clouds. R. Coifman and M. Maggioni proposed to use diffusion processes on such structures to introduce the notion of scale and certain directions.

R. R. Coifman and M. Maggioni, “Diffusion wavelets,” ACHA, 2006, Vol. 21(1), pp. 53–94.

Cartoon-like images

- A useful model for real images is the class of *cartoon-like images*, $\mathcal{E}^2(\mathbb{R}^2)$.
- Roughly, they are functions that are smooth away from a smooth curve of discontinuity.

Definition

Cartoon-like image functions Let $f \in L^2(\mathbb{R}^2)$ be a function with support contained in the closed unit square $[0,1]^2$ and such that f can be written as

$$f = f_0 + \mathbb{1}_B f_1,$$

for some $B \subset [0,1]^2$ with a closed C^2 boundary. If $f_0 \in C^2(\mathbb{R}^2)$ and $f_1 \in C^2(\mathbb{R}^2)$, then we say that $f \in \mathcal{E}^2(\mathbb{R}^2)$.

Norbert Wiener Center for Harmonic Analysis and Applications

Representations of $\mathcal{E}^2(\mathbb{R}^2)$

- Let $f \in \mathcal{E}^2(\mathbb{R}^2)$ and let f_N be its best N-term approximation with respect to a set of analyzing functions. The optimal asymptotic decay rate of $||f - f_N||_2^2$ is $O(N^{-2})$, $N \to \infty$, achieved adaptively.

D.L. Donoho, Sparse components of images and optimal atomic decomposition, Constr. Approx. 17 (2001), 353–382.

Definition

Let $\{\psi_i : i \in I\} \subset L^2(\mathbb{R}^2)$ be a normalized frame for $L^2(\mathbb{R}^2)$. Then, we say that $\{\psi_i : i \in I\}$ provides optimally sparse approximation for $\mathcal{E}^2(\mathbb{R}^2)$ if the best N-term nonlinear approximation error in $L^2(\mathbb{R}^2)$:

$$||f - f_N||_2^2 = \left\| f - \sum_{i \in I_N} \langle f, \psi_i \rangle \psi_i \right\|_2^2,$$

where $\langle f, \psi_i \rangle$, $i \in I_N$, are the N largest coefficients in magnitude, satisfies

$$||f - f_N||_2^2 \leq C N^{-2},$$

as $N \to \infty$.

Representations of $\mathcal{E}^2(\mathbb{R}^2)$

- Up to a log factor, curvelets, contourlets, and shearlets satisfy this optimal decay rate (ridgelets are only optimal for linear boundaries). Hence, these analyzing functions are *essentially optimally sparse* for cartoon-like images. Wavelets can only achieve $O(N^{-1})$. Fourier series are even worse with $O(N^{-1/2})$.
- We focus on shearlets since they have multiple, efficient numerical implementations.
- Shearlets' optimality is nearly ideal for sufficiently chosen shear let frames:

$$||f - f_N||_2^2 \leq CN^{-2}\log^3(N), \quad \text{as} \quad N \to \infty.$$

G. Kutyniok, W.-Q. Lim, Compactly supported shearlets are optimally sparse. J. Approx. Theory 163 (2011), 1564–1589.

Shearlets

Continuous shearlets in $\mathbb{R}^2$ depend on three parameters: the scaling parameter $a > 0$, the shear parameter $s \in \mathbb{R}$, and the translation parameter $t \in \mathbb{R}^2$, and they are defined as follows:
We define the *parabolic scaling matrices*

$$A_a = \begin{pmatrix} a & 0 \\ 0 & a^{1/2} \end{pmatrix}, \qquad a > 0$$

and the *shearing matrices*

$$S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \qquad s \in \mathbb{R}.$$

Also, let D_M be the dilation operator defined by

$$D_M\psi = |\det M|^{-1/2}\psi(M^{-1}\cdot), \qquad M \in GL_2(\mathbb{R})$$

and T_t the translation operator defined by

$$T_t\psi = \psi(\cdot - t), \qquad t \in \mathbb{R}^2.$$

Shearlets

Definition

Let $\psi \in L^2(\mathbb{R}^2)$. The *Continuous Shearlet Transform* of $f \in L^2(\mathbb{R}^2)$ is

$$f \mapsto \mathcal{SH}_\psi f(a,s,t) = \langle f, T_t D_{A_a} D_{S_s} \psi \rangle, a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2.$$

- Parabolic scaling allows for directional sensitivity.
- Shearing allows us to change this direction.
- By carefully choosing ψ and discretizing the parameter space, we can decompose $f \in L^2(\mathbb{R}^2)$ into a Parseval frame.

Shearlets

- It's generally assumed that $\hat{\psi}$ splits as $\hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\hat{\psi}_2(\xi_2/\xi_1)$.
- The basic shearlet ψ is only used in a *horizontal cone*, while the reflection of ψ across the line $\xi_2 = \xi_1$ is used in a *vertical cone*. A scaling function ϕ is used for the low-pass region. This construction is known as *cone-adapted* shearlets.

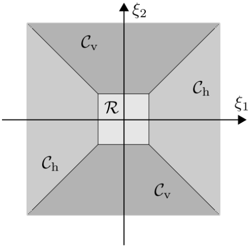

Figure : Frequency tiling for cone-adapted shearlets.

G. Kutyniok and D. Labate, eds. *Shearlets: Multiscale analysis for multivariate data.* Birkhäuser. (2012).

Shearlet Implementations

- Shearlets have several efficient numerical implementations in MATLAB that are freely available.
  - 2D Shearlet Toolbox (Easley, Labate, and Lim). [1]
  - Shearlab (Kutyniok, Shahram, Zhuang et al.). [2]
  - Fast Finite Shearlet Transform (Häuser and Steidl).[3]
- We used the last option (FFST) here, which is in many ways the most intuitive of the implementations.

[1] http://www.math.uh.edu/~dlabate/software.html

[2] http://www.shearlab.org/

[3] http://www.mathematik.uni-kl.de/imagepro/software/ffst/

Fast Finite Shearlet Transform (FFST)

- Consider an $M \times N$ image. Define $j_0 := \lfloor \log_2 \max\{M, N\} \rfloor$. We discretize the parameters as follows:

$$a_j := 2^{-2j} = \frac{1}{4^j}, \qquad j = 0, \ldots, j_0 - 1,$$

$$s_{j,k} := k2^{-j}, \qquad -2^j \leq k \leq 2^j,$$

$$t_m := \left(\frac{m_1}{M}, \frac{m_2}{N}\right), \quad m_1 = 0, \ldots, M-1, \; m_2 = 0, \ldots, N-1.$$

- Note that the shears vary from -1 to 1. To fill out the remaining directions, we also shear with respect to the y-axis.
- Shearlets whose supports overlap are “glued” together.
- The transform is computed through the $2D$ FFT and iFFT.

Norbert Wiener Center for Harmonic Analysis and Applications

Fast Finite Shearlet Transform

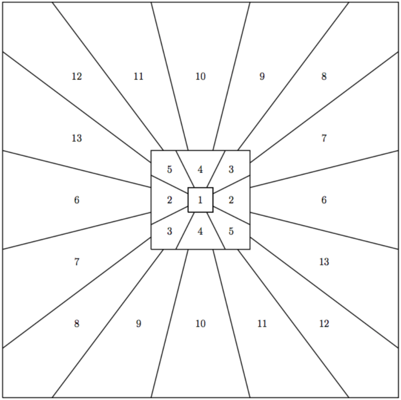

Figure : Frequency tiling for FFST.

S. Häuser and G. Steidl. *Fast finite shearlet transform: a tutorial*. arXiv:1202.1773. (2014).

Additional Picture for FFST

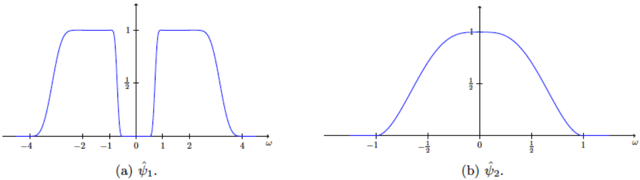

Figure : $\hat{\psi}_1$ and $\hat{\psi}_2$ for the FFST.

ibid.

Fast Finite Shearlet Transform

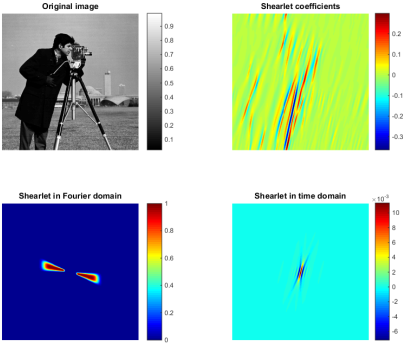

Figure : Demonstration of output from the FFST on the cameraman image. The shearlet coefficients are from scale 3 (out of 4) in the direction of slope 4.

ibid

Norbert Wiener Center for Harmonic Analysis and Applications

Summary and Conclusions

- We have covered some of the multiscale directional representations systems in use today.
- These systems are equipped with good approximation properties, exceeding in certain aspects what wavelet theory provides us with, and they have fast implementations.
- But we have to address the question of determining the specific set of directions/parameters needed for any given data of interest.