Mathematical Representation of Big Visual Data: A Data-driven Perspective

Ji Hui

National University of Singapore

May 2, 2017

Wavelet operator and block model

Recall a single-level synthesis operator of wavelet system

$$W^* = [W^*_{a_0}, \dots, W^*_{a_L}] \quad \text{with} \quad W^*_{a_\ell} c = a_\ell \circledast (c \uparrow_p).$$

The system X associated with W^* can be expressed as

$$X := \{X_n\}_{n\in\mathbb{Z}}, \quad \text{where} \quad X_n = [a_1(\cdot - np), \dots, a_L(\cdot - np)].$$

Suppose that the filter bank is finitely supported such that

$$\text{supp}(a_\ell) \subseteq \Omega = [0 : T-1], \quad \ell = 0, \dots, L-1.$$

Then, define the matrix form $D \in \mathbb{R}^{T\times L}$ of the filter bank by

$$D = (a_0[\Omega], a_1[\Omega], \dots, a_{L-1}[\Omega]).$$

Consider analyzing $f \in \ell^2(\mathbb{Z})$ by analysis operator W

$$c = Wf$$

which can be expressed in terms of blocks of c with step size p:

$$c[0 : L, \Omega_n] = Df[\Omega_n], \quad \forall \quad \Omega_n = [np : np + T - 1].$$

KSVD and frames

For a sequence $\ell^2(\mathbb{Z})$, partitioning it to finitely supported blocks (with possible overlap):

$$\{f_n\}_{n\in\mathbb{Z}} \subset \mathbb{R}^T.$$

The K-SVD method by Aaron et al. is to learn a dictionary $D \in \mathbb{R}^{T\times M}$ which sparsifies $\{f_k\}$, via solving

$$\min_{\{\|D_\ell=1\|\}_{\ell=1}^L,\{c_n\}_n} \sum_n \frac{1}{2}\|f_k - Dc_n\|_2^2 + \lambda\|c_k\|_0.$$

Suppose that (Remark: Condition (i) is not guaranteed in model)

(i) $\overline{\mathrm{span}\{D_1,\ldots,D_M\}} = \mathbb{R}^T$

(ii) $A\|f\|_2^2 \le \sum_k \|f_k\|_2^2 \le B\|f\|_2^2.$

Then the system $\{D_k,n\}_{n,k}$ forms a frame for $\ell^2(\mathbb{Z})$, where $D_{k,n}$ denotes the translated D_k w.r.t. f_n.

Data-driven tight frame

Data-driven frames

- No closed-form Linear expansion
- Completeness of D in $\mathbb{R}^T$ is difficult to be guaranteed.

Data-driven tight frames

- fast linear expansion: $f = \sum_k \langle f, x_k \rangle x_k$
- UEP on D for generating tight frames w/ multi-scale.

A general data-driven tight frame model

$$\min_{D,\{c_n\}_n} \sum_n \frac{1}{2} \|f_k - Dc_n\|_2^2 + \lambda \|c_k\|_0.$$

where the filter bank D satisfies the UEP.

The wavelet system generated by D forms a wavelet tight frames as long as $\{f_k\}_k$ is a uniform (overlapped) partition of f.

A simple and efficient construction

Consider a un-decimal discrete wavelet tight frame, i.e.,

$$X := \{X_n\}_{n\in\mathbb{Z}}, \quad \text{where} \quad X_n = [a_0[\cdot - n], \ldots, a_{L-1}[\cdot - n]].$$

The UEP for such a system is simplified to

$$\sum_{\ell=0}^{L-1} \sum_n a_\ell[n+k] a_\ell[n] = \delta_k.$$

Consider a finitely supported filter bank $\{a_0, \ldots, a_{L-1}\}$ with $\mathrm{supp}(a_\ell) \subset [0 : T-1]$. Define a dictionary $D \in \mathbb{R}^{T\times L}$

$$D_\ell = a_{\ell-1}[0 : T-1], \quad \ell = 1, \ldots, L.$$

Theorem (Local and global)

The filter bank $\{a_0, \ldots, a_{L-1}\}$ generates an un-decimal wavelet tight frame for $\ell^2(\mathbb{Z})$, provided that $\{\frac{1}{\sqrt{T}} D_\ell\}$ forms a tight frame for $\mathbb{R}^T$, i.e. $DD^* = T^{-1}I$.

Variational model for dictionary learning

The sparsity-based model for data-driven tight frame

$$\min_{D,\{c_k\}} \sum_k \|f_k - Dc_k\|_2^2 + \mu\|c_k\|_0, \quad \text{s.t.} \quad DD^* = T^{-1}I_L,$$

which is equivalent to the following real-valued dictionary learning model ($D \leftarrow \sqrt{T}D$):

$$\min_{D,C} \|Y - DC\|_F^2 + \lambda\|C\|_0, \quad \text{s.t.} \quad DD^\top = I,$$

where $Y = \sqrt{T}[\dots, f_{-1}, f_0, f_1, \dots]$ and $C = [\dots, c_{-1}, c_0, c_1, \dots]$,

- The optimization problem is a challenging non-convex problem.
- When D is a over-complete tight frame, the subproblem of calculating the sparse code C under D is a NP-hard problem.

Fast methods for orthogonal dictionary learning

A further simplification by consider a square matrix $D \in \mathbb{R}^{T \times T}$.

$$D^* D = DD^* = I \Longrightarrow \min_{D^* D = I, \{c_k\}} \sum_k \|C - D^* Y\|_2^2 + \lambda |C\|_0,$$

Alternating iteration scheme: for $k = 0, 1, 2, \ldots$,

$$\begin{cases} \text{P1}: & C_{k+1} := \text{argmin}_C \|C - D^* Y\|_F^2 + \lambda \|C\|_0 \\ \text{P2}: & D_{k+1} := \text{argmin}_{D^* D = I} \|Y - DC\|_F^2 \end{cases}$$

Each step in the iteration has closed-form solution

$$\begin{cases} \text{P1}: & C_{k+1} := \Gamma_{\sqrt{\lambda}}(D^* Y) \\ \text{P2}: & D_{k+1} := UV^*, \end{cases}$$

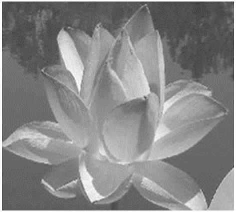
where Γ_μ denote the hard-thresholding operator:

$$\Gamma_\mu(x) = x \text{ if } |x| > \mu \text{ and } 0 \text{ otherwise},$$

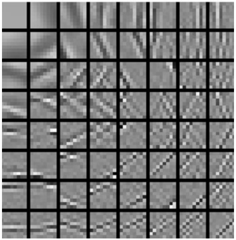
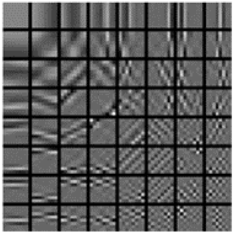
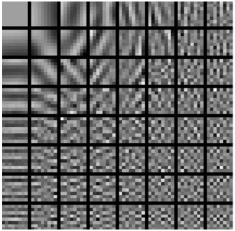
and (U, V) denotes the orthogonal matrices of SVD of YC_k^* such that $YC_k^* = U\Sigma V^*$.

Demonstration of data-driven filter bank

Images


Data-driven filter banks

Reference

1. K. Kreutz-Delgado, J. Murray, B. Rao, K. Engan, T. Lee, T. Sejnowski, Dictionary learning algorithms for sparse representation, *Neural Computation* 15 (2) (2003)
2. J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, *Supervised dictionary learning*, *NIPS*, 2009
3. M. Aharon, M. Elad, and A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.*, 54 (11), 2006.
4. M. Elad, M. Ahron, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Processing*, 54 (12) (2006)
5. J. Cai, H. Ji, Z. Shen and G. Ye, Data-driven tight frame construction and image denoising, *Applied and Computational Harmonic Analysis*, 37(1), 2014
6. Y. Quan, H. Ji and Z. Shen, Data-driven multi-scale non-local wavelet frame construction and image recovery, *Journal of Scientific Computing*, 63(2), 2015
7. C. Bao, H. Ji and Z. Shen, Convergence analysis for iterative data-driven tight frame construction scheme, *Applied and Computational Harmonic Analysis*, 2015