Learning sparsifying filter banks

Luke Pfister & Yoram Bresler

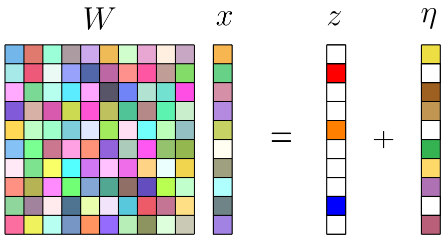
University of Illinois Urbana-Champaign

The Problem

$$y = Ax + e$$

- $x \in \mathbb{R}^{N^2}$: Desired image
- $y \in \mathbb{C}^{M}$: Measurements
- $A \in \mathbb{C}^{M \times N^2}$: System operator
- $e \in \mathbb{C}^{M}$: Noise
- Underdetermined or ill-conditioned $\Longrightarrow$ must incorporate prior information!
  - Assume signal has sparse representation

Transform Sparsity



- $W \in \mathbb{R}^{N_c \times K^2}$: Sparsifying transform
- $Wx \approx$ sparse
- W is left-invertible: $x \approx W^{\dagger} z$

Transform Sparsity

$$z^* = \min_z \frac{1}{2}\|Wx - z\|_2^2 + \nu\|z\|_0$$

$$z_i^* = \begin{cases} [Wx]_i & |[Wx]_i| \geq \sqrt{\nu} \\ 0 & \text{else} \end{cases}$$

$$\triangleq \mathcal{T}_\nu\left(Wx\right)$$

Analytic Sparse Models

- Designed for optimal properties on a mathematical class of signals
- Fast implementation
- Hard to design for high dimensional data

Adaptive Sparse Models

- Learn a good representation directly from data
  - Minimize objective function over training set
- Synthesis sparsity: Dictionary learning
- Transform sparsity: Transform learning

Transform Learning

$$\min_{W,Z} \frac{1}{2}\|WX - Z\|_F^2 + \nu\|Z\|_0 + J(W)$$

- X: Training data
- Z: Sparsified data
- J: Encourages well conditioned, left-invertible W

Transform Learning Algorithms

$$\min_{W,Z} \frac{1}{2}\|WX - Z\|_F^2 + \nu\|Z\|_0 + J(W)$$

- Algorithms to learn square, tall, separable, block structured ...
- Applications to image denoising, magnetic resonance imaging (MRI), computed tomography...
- Training set usually consists of patches of image or video data

Patches

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

Patches


1 2 3 4 5 6
7 8 9 10 11 12
13 14 15 16 17 18
19 20 21 22 23 24
25 26 27 28 29 30
31 32 33 34 35 36
1
2
7
8

x_1

Patches

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

1	2
2	3
7	8
8	9

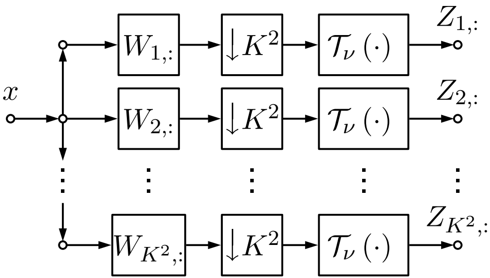
x_1 x_2

From Patches To Filter Banks

Non-overlapping Patches

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

Sparsifying Transforms as Filter Banks



Sparsifying Transforms as Filter Banks

- $\text{vec}(WX) = \mathcal{H}_W x$
- W is the **polyphase matrix** of a filter bank
- Defn: $\mathcal{H}_W$ is perfect reconstruction (PR) if $\mathcal{H}_W$ left invertible (LI).
- Properties of filter bank controlled by patch extraction and by W
  - Shape of patches $\rightarrow$ shape of filters
  - Rows of W $\rightarrow$ channels of filter bank
  - $\mathcal{H}_W$ is PR $\Leftrightarrow$ W is LI

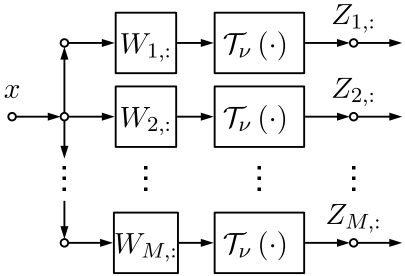
Sparsifying transforms as filter banks

- Take away: Existing transform learning algorithms learn perfect reconstruction downsampled filter banks!
  - ...but we usually use overlapping patches.
- Requiring W to be LI is stronger than requiring $\mathcal{H}_W$ to be PR!
- Two questions:
  1. Do we benefit by requiring $\mathcal{H}_W$ to be PR and relaxing the LI condition on W?
  2. Can we find an efficient algorithm to learn such a $\mathcal{H}_W$?

Previous Work

- Connection between patch based analysis operators and convolution previously known
- Convolution often used as a computational tool
- (Cai, 2014) learned a PR filter bank by requiring $W^T W = I$

Maximally Overlapping Patches



- Maximally overlapping patches
- W is no longer the polyphase matrix for $\mathcal{H}_W$
- Generally difficult to check the PR condition

Maximally overlapping Patches

- Simplify: replace linear convolution with circular convolution
- Patch interpretation: 'wrap around' image boundary
- Third representation
  - C_j : 2D circulant matrix such that $C_j x = w_j \circledast x$
  - $C_W \triangleq [C_{w_1}^T, C_{w_2}^T, \ldots, C_{w_{N_C}}^T]^T$
  - $C_W x = \mathcal{H}_W x = \mathrm{vec}\,(WX)$
  - C_W is highly structured: block-circulant with circulant blocks

A few ingredients...

- $\Phi \in \mathbb{C}^{N^2 \times N^2}$: Normalized 2D discrete Fourier transform (DFT) matrix.
- $\bar{\Phi} \in \mathbb{C}^{N^2 \times K^2}$: Zero-padded DFT of $K \times K$ signal
- $\mathbf{1}_k \triangleq \underbrace{[1, 1, \ldots, 1]^T}_{k \text{ entries}}$
- $C_j = \Phi^H \operatorname{ddiag}\left(\bar{\Phi} W_{j,:}^T\right) \Phi$

The Key Property

- Each frequency must pass through at least one channel!

Diagonalization

$$C_W^T C_W = \Phi^H \operatorname{ddiag}\left(\left| \bar{\Phi} W^T \right|^2 \mathbf{1}_{N_c} \right) \Phi$$

Perfect Recovery Condition

$\mathcal{H}_W$ is PR $\Leftrightarrow$ each entry of $\left| \bar{\Phi} W^T \right|^2 \mathbf{1}_{N_c} > 0$

- Decouples the choice of N_c and K
- Especially attractive for high dimensional data

Learning a sparsifying filter bank

Learning Formulation

- Desiderata
  - Parameterize with few degrees of freedom
  - $\mathcal{H}_W x$ should be (approximately) sparse
  - $\mathcal{H}_W$ should be PR and well conditioned
  - No identically zero filters
  - No duplicated filters

Learning Formulation

- WX should be (approximately) sparse

$$F(W, Z, x) \triangleq \frac{1}{2}\|WX - Z\|_F^2 + \nu\|Z\|_0$$

Learning Formulation

- $\mathcal{H}_W$ should be PR and well conditioned
- No uniformly zero filters

$$J_1(W) = \left\| \left| \bar{\Phi} W^T \right|^2 \mathbf{1}_{N_c} - \mathbf{1}_{N^2} \right\|_2^2 - \beta \sum_{j=1}^{N_c} \log \left(\| W_{j,:} \|_2^2 \right)$$

Learning Formulation

- No duplicated filters

$$J_2(W) = \sum_{1 \le i < j \le N_c} -\log\left(1 - \left(\frac{\langle W_{i,:}, W_{j,:}\rangle}{\|W_{i,:}\|_2 \|W_{j,:}\|_2}\right)^2\right)$$

Learning Formulation

$$\min_{W,Z} \frac{1}{2}\|WX - Z\|_F^2 + \alpha J_1(W) + \gamma J_2(W) + \nu\|Z\|_0$$

Alternating minimization:

- $Z^{k+1} = \arg\min_Z \frac{1}{2}\|W^k X - Z\|_F^2 + \nu\|Z\|_0$
- $W^{k+1} = \arg\min_W \frac{1}{2}\|WX - Z^{k+1}\|_F^2 + \alpha J_1(W) + \gamma J_2(W)$

Alternating Minimization

$$Z^{k+1} = \arg\min_{Z} \frac{1}{2}\|WX - Z\|_F^2 + \nu\|Z\|_0$$

- Closed form: $Z^{k+1} = \mathcal{T}_\nu\left(WX\right)$

Alternating Minimization

$$W^{k+1} = \arg\min_{W} \frac{1}{2}\|WX - Z\|_F^2 + \alpha J_1(W) + \gamma J_2(W)$$

- Minimize using L-BFGS

Application to Magnetic Resonance Imaging

Imaging Model

- Imaging Model: Undersampled Fourier measurements

$$y = \Gamma\Phi x + e$$

- $x \in \mathbb{R}^{N^2}$: Input image
- $\Phi \in \mathbb{C}^{N^2 \times N^2}$: DFT matrix
- $\Gamma \in \mathbb{C}^{M \times N^2}$: Row selection matrix
- $e \in \mathbb{C}^{M}$: Zero mean Gaussian noise

Image Reconstruction

$$\min_{x,\mathcal{H}_W,z} \frac{1}{2}\|y-\Gamma\Phi x\|_2^2+\lambda\left(\frac{1}{2}\|\mathcal{H}_W x - z\|_2^2 + \nu\|z\|_0 + \alpha J_1(\mathcal{H}_W) + \gamma J_2(\mathcal{H}_W)\right)$$

- Data fidelity
- Transform learning
- Solve using alternating minimization

Image Update

$$\min_{x} \frac{1}{2}\|y-\Gamma\Phi x\|^2+\frac{\lambda}{2}\|\mathcal{H}_W x-z\|_2^2$$

Transform Update

$$\min_{W} \frac{1}{2}\|\mathcal{H}_W x-z\|_F^2+\alpha J_1(W)+\gamma J_2(W)$$

Sparse Code Update

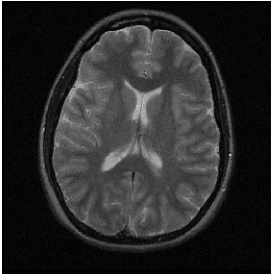
$$\min_{z} \frac{1}{2}\|\mathcal{H}_W x - z\|_2^2 + \nu\|z\|_0$$

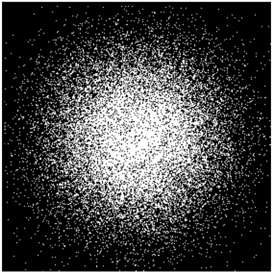
Experiments

- Synthetic MR data from magnitude image
- $\approx$ 5 fold undersampling
- Vary filter size $\&$ number of channels
- Compare against square patch-based transform learning:

$$\min_{W,x,Z} \frac{1}{2}\|y - \Gamma\Phi x\|_2^2 + \frac{\lambda}{2}\|WX - Z\| + \nu\|Z\|_0$$
$$+ \alpha\|W\|_F^2 - \beta \log\det W$$

- Solved using alternating minimization
- Initialized with DCT matrix

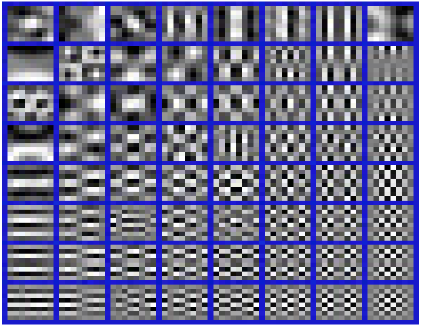


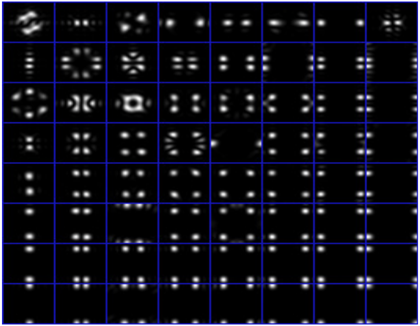


Reconstruction PSNR (dB)

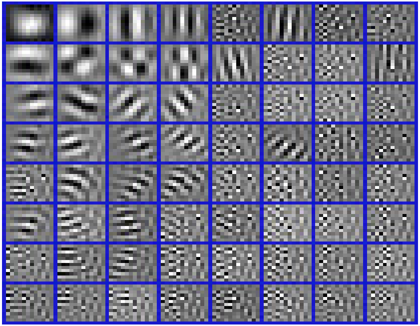
σ / PSNR In	Filter Bank			Patch Based
	$N_c = 64$, $K = 8$	$N_c = 128$, $K = 8$	$N_c = 64$, $K = 12$	64×64
0 / 29.6	35.2	**35.2**	35.1	34.6
$\frac{10}{255}$ / 28.8	32.6	**32.7**	32.6	32.5
$\frac{20}{255}$ / 26.9	**31.6**	31.6	31.2	31.3

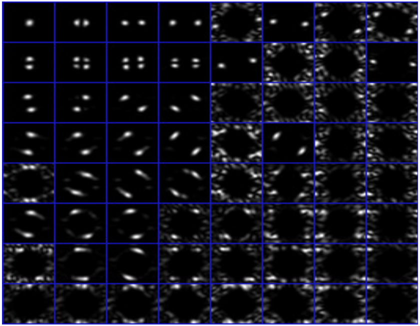
Learned filters 8×8





Learned filters 12×12





Conclusion

Conclusion

- New framework for learning filter bank sparsifying transforms
- Replace patch recovery conditions with image recovery
- Decouples number of channels from filter length
- Outperforms patch based transform for MR reconstruction

Conclusion

- Extended version on ArXiV soon!
- For more on transform learning
  - Sai Ravishankar's talk (Tuesday Afternoon)
  - `http://transformlearning.csl.illinois.edu`
  - `http://lukepfister.me`
- Supported in part by NSF CCF-1320953 and the Andrew T. Yang Fellowship.