EFFICIENT SPARSIFYING TRANSFORM LEARNING AND ITS APPLICATIONS
Saiprasad Ravishankar and Yoram Bresler
Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, USA

OVERVIEW
Problem Statement
- Learning square sparsifying transforms & applications
Contributions
- We propose algorithms for learning square transforms.
- Proposed alternating transform learning algorithms
  - have efficient closed-form solutions for subproblems
  - achieve global minimum of non-convex subproblems
  - achieve local minimum of overall problems
- are insensitive to initialization
- have low computational cost
- encourage well-conditioning
- Adapted transforms provide better image representations than analytical ones such as DCT, Wavelets.
- In denoising and compressed sensing, adaptive square transforms perform comparably or better than learnt overcomplete synthesis dictionaries, but are much faster.

LEARNING SPARSE MODELS
- Synthesis and analysis formulations are typically non-convex and NP-hard, and algorithms such as K-SVD [3] are computationally expensive.
- Square transform learning [2]
- Synthesis Model (SM): Given signal \( y \) and dictionary \( D \in \mathbb{R}^{m \times n} \), we have \( y = Dx + e \), with \( \|x\|_0 \ll K \), and \( e \) is a deviation term.
  - Synthesis sparse coding:
    \[
    \hat{x} = \text{argmin}_x \| y - Dx \|^2 \quad \text{s.t.} \quad \|x\|_0 \leq s
    \]
    - This is NP-hard.
  - Greedy and relaxation algorithms can be computationally expensive.
- Analysis Model (AM): Given signal \( y \) and analysis dictionary \( \Omega \in \mathbb{R}^{m \times n} \), \( \|dy\|_0 \ll m \).
- Noisy Analysis Model (NSAM) [1]: \( y = q + e \), with \( \Omega q \) sparse and \( e \) a small error term in the signal domain.
  - Analysis sparse coding:
    \[
    \hat{q} = \text{argmin}_q \| y - \Omega q \|^2 \quad \text{s.t.} \quad \|\Omega q\|_0 \leq s
    \]
    - This is NP-hard.
  - Approximate Algorithms - computationally expensive.

CONVERGENCE OF ALGORITHM
- By adding \( \|x_i\|_0 \leq s \) \( \forall i \) as a penalty in the cost of (P1) using a barrier function, (P1) becomes unconstrained.
- Denote the unconstrained objective by \( g(W, X) \).
- Let \( \beta_j(W) \) denote the magnitude of the \( j \)th largest element (magnitude-wise) of vector \( u \).
- Local Convergence: Let \((W^k, X^k)\) denote the iterate sequence generated by our alternating Algorithm with data \( Y \) and initial \((W^0, X^0)\). Then, \( g(W^k, X^k) \) is monotone decreasing, and converges to a finite value, say \( g^* \). Moreover, every accumulation point \((W^\star, X^\star)\) of the iterate sequence is a fixed point of the algorithm satisfying the following optimality condition.
  \[
  g(W + dW, X + dX) \geq g(W, X) \implies g^* = g(W^\star, X^\star) \quad (2)
  \]
- (2) holds for sufficiently small \( dW \) and \( dX \) in the union of the half space \( \{((W, X) : dX \leq 0 \} \) and the local region \( \{ \|x_i\|_0 \leq \max_j |\beta_j(W)\} : \|WY\|_0 > s \} \).
- tolerates arbitrarily large perturbations \( dX \).
- If \( \|WY\|_0 \leq s \) \( \forall i \), then \( dX \) can be arbitrary.
- Objective converges to local minimum and iters converge to equivalence class (in terms of cost) of local minimizers.
- Convergence irrespective of initialization.

APPLICATION: IMAGE DENOISING
- Goal: estimate image \( x \) from noisy image \( y = x + h \).
- \( R_j \) extracts 2D patch. \( R_j y = \) noiseless \( x_j \).
- Denoised image obtained by averaging \( x_j \)’s at their 2D locations.
- (P2) is solved by an efficient alternating scheme that uses closed-form updates, and \( x_j \) are found adaptively.

COMPUTED SENSING MRI
- Estimate image \( x \) from compressive Fourier measurements \( y \). 
- \( F_x \): undersampled Fourier encoding matrix.
- (P3) solved efficiently by alternating scheme (TMRI) that uses closed-form updates.

REFERENCE