OVERVIEW

Problem Statement
- Online square sparsifying transform learning & applications

Motivations
- Big data -> large training set -> batch learning expensive.
- Real-time applications = data arrives sequentially, and must be processed sequentially to limit latency.

Contributions
- We propose online and mini-batch transform learning methods that:
  - alternate between sparse coding and transform update
  - are amenable to big data, and real-time applications
  - converge to the set of stationary points of the cost
  - are cheap in computations, memory, and latency
  - are cheaper than online synthesis dictionary learning
  - encourage well-conditioning
- Proposed online schemes outperform batch methods in sparse representation & denoising, while being much faster.

BATCH LEARNING OF SPARSE MODELS

- Synthesis dictionary learning is typically non-convex and NP-hard, and algorithms [4] are computationally expensive.
- Batch Square Transform Learning [1, 2]

\[ \mathbf{W} \] is computed exactly & cheaply by thresholding \( \mathbf{W} \) by the largest magnitude elements. Least squares signal estimate: \( \mathbf{y} = \mathbf{W} \mathbf{x} \).

ONLINE SQUARE TRANSFORM LEARNING

- Online Learning & Sparse Coding: For \( t = 1, 2, 3, \ldots \), solve

\[ \min_{\mathbf{W} \in S} \frac{1}{2} \sum_{t} \| \mathbf{y}_t - \mathbf{W} \mathbf{x}_t \|^2_2 + \lambda \| \mathbf{W} \|_1 \]

- Mini-Batch Learning & Sparse Coding: For \( j = 1, 2, 3, \ldots \), solve

\[ \min_{\mathbf{W} \in S} \frac{1}{2} \sum_{t = jT}^{(j+1)T-1} \| \mathbf{y}_t - \mathbf{W} \mathbf{x}_t \|^2_2 + \lambda \| \mathbf{W} \|_1 \]

ONLINE LEARNING SCHEMATIC

CONVERGENCE ANALYSIS

- Prior work [3] showed the convergence of biconvex online synthesis learning.
- We make simpler assumptions here than in prior work [3]
- Our problems are not biconvex. We study the following costs:
  - The objective of the online transform update step is \( g_0(\mathbf{W}) = \frac{1}{2} \sum_{t} \| \mathbf{y}_t - \mathbf{W} \mathbf{x}_t \|^2_2 + \lambda \| \mathbf{W} \|_1 \)
  - The empirical objective function (for batch learning) is \( g_0(\mathbf{W}) = \frac{1}{2} \sum_{t} \| \mathbf{y}_t - \mathbf{W} \mathbf{x}_t \|^2_2 + \lambda \| \mathbf{W} \|_1 \)
  - The expected transform learning cost is \( g_0(\mathbf{W}) = \mathbb{E}_{\mathbf{x}}[\| \mathbf{y} - \mathbf{W} \mathbf{x} \|^2_2 + \lambda \| \mathbf{W} \|_1] \)

MINI-BATCH ALGORITHM FOR (P2)

- Sparse Coding: solve for \( x_t \) in (P1) with fixed \( W_t \): \( x_{t+1} \)
  - Cheap Solution: \( x_{t+1} = \arg \min_{x} \| \mathbf{W}_t \mathbf{x} - \mathbf{y}_t \|^2_2 + \lambda \| \mathbf{x} \|_1 \)
- Transform Update: solve for \( W_{t+1} \) in (P1) with fixed \( x_t \):
  - Minimize \( \frac{1}{2} \sum_{t} \| \mathbf{W}_{t+1} \mathbf{x}_t - \mathbf{y}_t \|^2_2 + \lambda \| \mathbf{W}_{t+1} \|_1 \)

MINI-BATCH DENOISING OF BIG IMAGES

ACKNOWLEDGEMENT

REFERENCES