LEARNING DICTIONARY FOR EFFICIENT SIGNAL COMPRESSION

Afshin Abdi, Ali Payani, and Faramarz Fekri

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA
{abdi,payani,fekri}@ece.gatech.edu

Abstract

- Dictionary method for compression is considered.
- Instead of sparse or energy-compact representation (PCA), a data driven basis for compression purpose should work better.
- Given a class $\mathcal{D}$ of dictionaries, and data samples, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N]$,
  - *Learn* an appropriate dictionary, $\boldsymbol{D} \in \mathcal{D}$
  - Obtain a *representation* $\boldsymbol{w}_i$ for each $\boldsymbol{x}_i$
  - *Encode* each coefficient separately

  Such that
  - $\boldsymbol{x}_i \approx \boldsymbol{D}\boldsymbol{w}_i$
  - Transmission rate, $R(\boldsymbol{w}_i)$, is minimized

$$\min_{\boldsymbol{D},\boldsymbol{w}} \; R(\boldsymbol{w}) \quad s.t. \;\; \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{w}\|_2^2 \le \epsilon$$

or

$$\min_{\boldsymbol{D},\boldsymbol{w}} \; \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{w}\|_2^2 + \lambda R(\boldsymbol{w})$$

Compression Rate

Dictionary

- Cost of sending dictionary is fixed, if signal is stationary, and becomes negligible as number of samples increases.

Quantizing coefficients

- Each coefficient quantized separately.
- Using Lloyd-Max to design quantizers: complex but not much gain in performance
- Using uniform quantizer:

  Asymptotically efficient, Quantization error is uniformly bounded,
  For small quantization step-size, δ;

$$R\left(w^{(q)}\right) \approx H_d(w) - \log_2 \delta.$$

For our real seismic data and SNR range (20-40 dB), uniform quantizer performs similar to Lloyd-Max algorithm.

Encoding coefficients

- Assign probability distribution; Gaussian Mixture Model (GMM)
- Code length $\approx -\log_2 P(\boldsymbol{w})$

Outline of Dictionary Learning Algorithm

- Initialize Dictionary, Coefficients and Probability Distributions
- Repeat until convergence:
  - Update dictionary to minimize error: $\min_{\boldsymbol{D}} \sum_i \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{w}_i\|_2^2$
  - Given GMM parameters and $\boldsymbol{D}$, for each data sample $\boldsymbol{x}_i$, find the coefficients: $\min_{\boldsymbol{w}_i} \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{w}_i\|_2^2 - \lambda \log_2 P(\boldsymbol{w}_i)$
  - Update GMM parameters to fit $\boldsymbol{W} = [\boldsymbol{w}_1, \dots, \boldsymbol{w}_N]$ and reduce bit rate.

Updating Dictionary

- Orthonormal Dictionaries:

  $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{V}^T$ where $\boldsymbol{X}\boldsymbol{W}^T = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ is SVD decomposition.
- Unions of orthonormal dictionaries, $\boldsymbol{D} = [\boldsymbol{D}_1, \boldsymbol{D}_2, \dots, \boldsymbol{D}_L]$:

  Applying the above result on each $\boldsymbol{D}_l$, for $\boldsymbol{X}_l = \boldsymbol{X} - \sum_{k \ne l} \boldsymbol{D}_k \boldsymbol{W}_k$.
- Unit norm atoms: $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_m]$
  - Use K-SVD algorithm
  - $\boldsymbol{E}_l = \boldsymbol{X} - \sum_{k \ne l} \boldsymbol{d}_k \boldsymbol{w}_{k,:}$ ($\boldsymbol{w}_{k,:}$ is the k^{th} row of $\boldsymbol{W}$)

    $\boldsymbol{v} = \boldsymbol{E}_l \, \boldsymbol{w}_{l,:}^T$

    $\boldsymbol{d}_l = \boldsymbol{v} / \|\boldsymbol{v}\|$

Computing Coefficients

- $p(\boldsymbol{w}) = \sum_s \pi(\boldsymbol{s}) P(\boldsymbol{w}|\boldsymbol{s})$, $\boldsymbol{w}|\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$
- Using MAP estimator for source index in the GMM

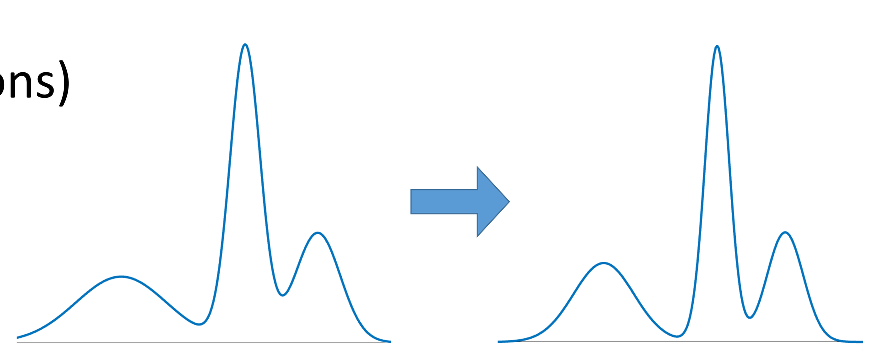
  Code length $\approx -\left(\log_2 \pi(\hat{\boldsymbol{s}}) + \log_2 P(\boldsymbol{w}|\hat{\boldsymbol{s}})\right)$
- Knowing $\hat{\boldsymbol{s}}$: $\boldsymbol{w}^* = \boldsymbol{\mu}_s + \left(\boldsymbol{D}^T\boldsymbol{D} + \frac{\lambda}{2} \boldsymbol{\Sigma}_s^{-1}\right)^{-1} \boldsymbol{D}^T(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\mu}_s)$
- How to estimate $\hat{\boldsymbol{s}}$:
  - Orthonormal Dictionary

$$s_i^* = \operatorname{argmin}_s \frac{(\mu_{i,s} - y_i)^2}{\lambda + 2\sigma_{i,s}^2} + \ln \frac{\sigma_{i,s}}{\pi_i(s)} \quad \text{where} \quad \boldsymbol{y} = \boldsymbol{D}^T\boldsymbol{x}$$

  - Unions of orthonormal dictionaries: treat each part ($\boldsymbol{w}_l$) separately
  - General dictionary: iteratively estimate $\hat{\boldsymbol{s}}$ and $\boldsymbol{w}$

Updating GMM parameters

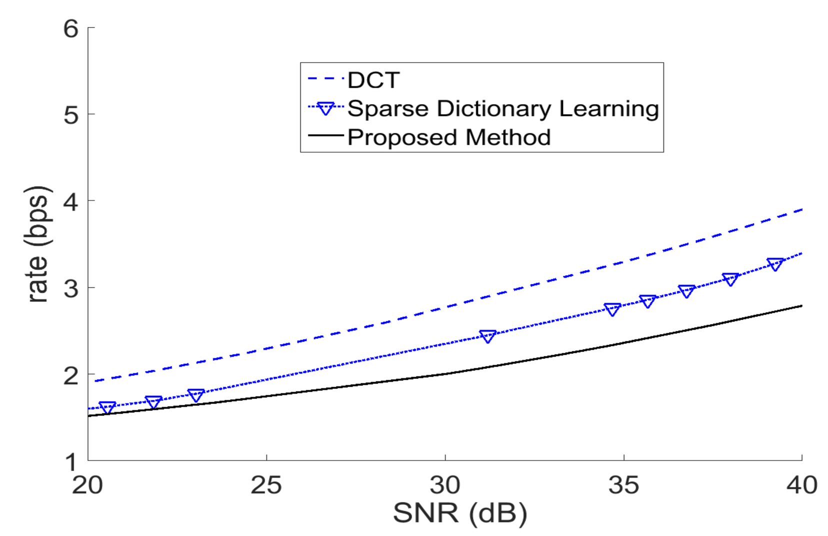
- Use EM algorithm (few iterations)
- Gradient Descend



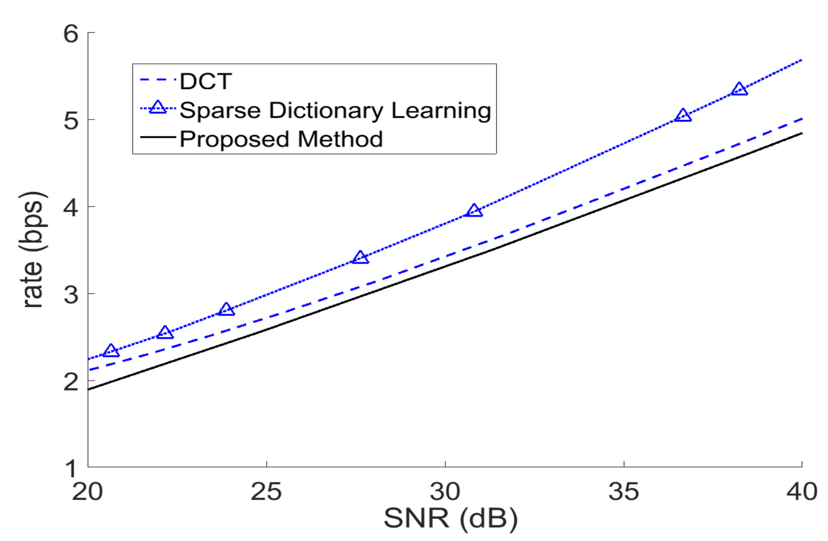
Simulation Results on Seismic Signals

- Verified the algorithm on two publicly available seismic databases (UTAH and USGS)
- Considered different number of sources in the GMM model, $K = 5$ gave best results
- Seismic traces are divided into segments of length 16 or 32.
- Compared with DCT and sparse dictionary learning

UTAH database, orthonormal dictionary of size 32×32



USGS database, orthonormal dictionary of size 32×32



USGS database, union of two orthonormal dictionaries of size 16×32