Intelligent sparse representations for speech
or how I learned to stop worrying and love the noise

Aditya Acharya, Karthik A Sankararaman, Manasij Venkatesh
ENEE662: Convex Optimization Group:13

Goals

Introduction to Speech Processing

Convex Programming Formulation

minimize \[ \sum_{i=1}^{m} \sqrt{\sum_{k \in \text{block } i} x_k^2} \]
subject to \[ \| Dx - y \|_2 < \epsilon \]

- The sparsity \( \ell_2,1 \) norm minimization problem is relaxed to an \( \ell_2,1 \) norm minimization problem.

Other Approaches

- **Lasso:** minimize \[ \frac{1}{2} \| Dx - y \|_2^2 + \alpha \| x \|_1 \]
- **Group Lasso:** minimize \[ \frac{1}{2} \| Dx - y \|_2^2 + \alpha \sum_{i=1}^{m} \sqrt{\sum_{k \in \text{block } i} x_k^2} \]
- We use ADMM for the above two problems. Sparsity improved by retaining the two most significant blocks.
- **Greedy Algorithms:** Greedily choosing the best two blocks By the way of Orthogonal Matching Pursuit.

Questions

- Can we have an intelligent representation in a different domain that exploits certain properties of speech and allows for sparser representation?
- Can such a representation allow us to denoise these signals and identify phonemes present in the signal?
- If \( y \) is a noisy signal, does imposing sparsity constraints while estimation result in denoising?

Preprocessing

- **Discrete Cosine Transform** is performed on 20ms samples of a phoneme to obtain a spectrogram.
- **PCA** on the spectrograms of all the phonemes gives the dictionary \( D \).
- **Objective:** To find a sparse vector \( x \) such that at most two phonemes are active in it and \( Dx = y \sim y \), the measured sample.

Experimental Details

Results

Performance of various methods for Estimation using Sparse representation

Performance of various methods for Denoising

Image Credits: Section 1-Noisy Signal: Ivan Selesnick