Super-resolution of Musical Signals Using Approximate Matching Pursuit
Brennan P. Keegan¹, Steven K. Tjoa², and K. J. Ray Liu¹
¹Signals and Information Group, Dept. of Electrical and Computer Eng., Univ. of Maryland – College Park, MD, USA
²Imagine Research, San Francisco, CA, USA

Introduction

Super-resolution is well-studied for video and images but not for musical signals.

One idea [Smaragdis, et al., 2009]: For each low-resolution (LR) input spectrum, find its coefficients with respect to a low-resolution basis. Using the same coefficients, reconstruct a high-resolution (HR) output from a high-resolution basis.

But what if you use a very large, overcomplete dictionary of real-world musical atoms?

We propose a super-resolution method using Approximate Matching Pursuit.

Approximate Matching Pursuit

See [Tjoa and Liu, ISMIR 2011].

Basic Idea: Inside matching pursuit, match using an approximate nearest neighbor (ANN) method, not exhaustive linear search.

1. Input: spectrum \( x \in \mathbb{R}^M \); dictionary \( D \in \mathbb{R}^{M \times K} \).
2. Output: coefficients \( s \in \mathbb{R}^K \).
3. Initialize: \( s \leftarrow 0 \); active set: \( S \leftarrow \emptyset \); residual \( r \leftarrow x \); threshold \( \epsilon > 0 \).
4. While \( ||r|| > \epsilon \):
   a. Find any \( k \) such that dictionary atom \( a_k \) and residual \( r \) are “close enough”.
   b. Add atom to active set: \( S \leftarrow S \cup k \)
   c. Solve for \( \{s_j\}_{j \in S} \): \( \min_{\{s_j\}_{j \in S}} ||x - \sum_{j \in S} a_j s_j||_2 \)
   d. Update residual: \( r \leftarrow x - Ds \)
5. Return \( s \).

Evaluation: Start with ground-truth HR spectrogram, \( Y_0 \). Evaluate the reconstruction error, \( F = ||Y_0 - \hat{Y}||_F \).

Dictionary: Piano spectra from the U. Iowa Dataset.

Example

First five measures of Mozart’s “Rondo Alla Turca.” Spectrograms \( X \) (top) and \( D\hat{S} \) (bottom).

Proposed Super-resolution Method

1. Given a HR dictionary \( D \), truncate \( D \) in frequency to obtain the LR dictionary, \( D_\ell \).
2. Given LR input spectrogram \( X \) and LR dictionary \( D_\ell \), use AMP to obtain the coefficient matrix \( \hat{S} \).
3. Construct the HR output spectrogram, \( \hat{Y} = D\hat{S} \).

Experiments

1. \( F \) vs. number of LR frequency bins, \( M_{\text{low}} \) (synthetic input)
2. \( F \) vs. number of LR frequency bins, \( M_{\text{low}} \) (real musical input)
3. \( F \) vs. dictionary size, \( K \) (real musical input)

Conclusions

Super-resolution is achieved with a dictionary of musical spectra.

Proposed method works well for very LR inputs.

Larger dictionaries generally result in higher accuracy.

Future Work

• Test on heterogeneous musical inputs with acoustic overlap.
• Larger, more representative dictionaries.
• Use AMP for time-domain decomposition.
• Image and video processing.