UNDERCOMPLETE BASIS FUNCTIONS FOR CREATIVE SIGNAL RECONSTRUCTION

Spencer Salazar

Center for Computer Research in Music and Acoustics
Stanford University
Stanford, California
spencer@ccrma.stanford.edu

ABSTRACT

Sparse coding and compressed sensing techniques have found a variety of uses in efficient signal encoding and compression. However these efforts typically seek to reconstruct a given signal with high-fidelity, and in the ideal case, with little to no perceptual variation from the original signal. In this project, we extended techniques for optimal L2+L1 norm sparse coding to retain characteristics of both the encoding basis set and the original input signals, for use as an effect in creative music applications. In this way, source signals can seem to be "recreated" or "remixed" by a different set of musical primitives, e.g. a symphony orchestra recreated by a 1980s video game console.

1. CONCEPT

Sparse coding and compressed sensing schemes encode a signal by projecting it on to a non-orthogonal basis set. Such bases are often supersets of a mathematically *complete* basis, such as the Fourier basis. In cases such as these, the basis set is known as *overcomplete*. Determining the representation under a given basis set is easily represented as a problem in various data fitting algorithms, such as various least-squares optimization algorithms or linear programming. Compressed sensing, being concerned with optimal sparsity, typically uses data fitting algorithms that optimize both the L2 and L1 norms. Optimizing the L1 norm favors solutions that have fewer non-zero components, allowing for greater compressability via traditional methods.

We define an *undercomplete* basis as one that mathematically is neither complete nor orthogonal. Projection of a signal on to an undercomplete basis will necessarily result in loss of data. However, reconstruction from the undercomplete encoding will heavily reflect musical characteristics of the underlying basis set. Depending on how close to complete the basis set is, and parameters of the data fitting algorithm used, the reconstruction will also retain salient musical characteristics of the source signal, thus combining characteristics of both the source and the basis set. While undercomplete reconstruction is unsatisfactory for general signal encoding and compression, it presents a number of possible artistic uses. For example, one might project/reconstruct a speech signal onto a basis set of piano notes, resulting in a "talking piano", or project/reconstruct a symphony orchestra onto a set of sawtooth waves, recreating the symphonic work a la an inexpensive vintage digital synthesizer.

In the case where we wish to preserve musical characteristics that are quantifiable (e.g. overall spectrum, spectral envelope, spectral peaks), our data fitting algorithm is sufficiently flexible that these characteristics can be optimized for. In this project,

we developed algorithms to fit both time-domain and frequency-domain representations of the signal. Based on our subjective listening of test signals, fitting the frequency-domain representation of the signal to the frequency-domain representation of the basis set produced a more musically recognizable reconstruction of the source signal. Attempting to best-fit other perceptually relevant characteristics of the signal, such as MFCCs or spectral peaks, presents an interesting avenue for further research.

2. TECHNIQUE

Initially we create a basis set to use for signal encoding and reconstruction. The manner in which these basis sets varies (see Section 2.1), but the result set is a collection of time-domain audio signals. Next we window the input signal according to a uniform block size. (The block size chosen depends largely on the length, or average length, of the basis set vectors.) If necessary, the features we wish to optimally match are extracted. For this project, only DFT magnitude spectrum and time domain signal were matched. The data fitting algorithm is executed given the basis set, the original signal, and some number of tuning parameters.

While we experimented with a variety of algorithms (see Section 2.2), each fundamentally is solving the same problem:

$$\min_x : \|Ax - y\|_2^2 + \tau\|x\|_1 \qquad (1)$$

where A is a matrix of basis vectors, x is the encoded signal according to that basis set, y is the windowed source signal, τ is a weighting parameter indicating desirability of sparseness of x, $\|v\|_2$ denotes the L2 (Euclidean) norm of v, and $\|v\|_1$ denotes the L1 norm of v. In the case of fitting the magnitude spectrum of the signal, A is a matrix of the magnitude spectrum of each basis vector, and y is the magnitude spectrum of the windowed source signal.

2.1. Basis Set

We have developed two different types of basis sets: uniform oscillator basis sets and non-uniform sampled basis sets. Uniform oscillator basis evenly distributes a number of standard oscillators (sawtooth, square wave, etc.) over a typically frequency range. Non-uniform sampled

2.2. Data Fitting Algorithms

3. REFERENCES

[1] S. K. Mitra and J. F. Kaiser, Eds., *Handbook for Digital Signal Processing*, J. Wiley & Sons, New York, NY, USA, 1993.

[2] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, second edition, 1991.

[3] X. Serra, *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, G. D. Poli, A. Picialli, S. T. Pope, and C. Roads Eds. Swets & Zeitlinger, Lisse, Switzerland, 1996.

[4] James A. Moorer, "Audio in the new millennium," *J. Audio Eng. Soc.*, vol. 48, no. 5, pp. 490–498, May 2000.

[5] A. Nackaerts, B. De Moor, and R. Lauwereins, "Parameter estimation for dual-polarization plucked string models," in *Proc. Intl. Computer Music Conf.*, Havana, Cuba, Sept. 17-23, 2001, pp. 203–206.

[6] D. Arfib, "Different ways to write digital audio effects programs," in *Proc. Digital Audio Effects (DAFx-98)*, Barcelona, Spain, Nov. 19-21, 1998, pp. 188–91.

[7] A. Askenfelt, "Automatic notation of played music (status report)," Tech. Rep., STL-QPSR, Vol. 1, pp. 1–11, 1976.

[8] E. B. Egozy, "Deriving musical control features from a real-time timbre analysis of the clarinet," M.S. thesis, Massachusetts Institute of Technology, 1995.

[9] P. Dutilleux, *Vers la machine à sculpter le son, modification en temps-réel des caractéristiques fréquentielles et temporelles des sons*, Ph.D. thesis, University of Aix-Marseille II, 1991.

[10] K. Fitz and L. Haken, "Current Research in Real-time Sound Morphing," Available at http://www.cerlsoundgroup.org/RealTimeMorph/, accessed March 08, 2006.