Fast Estimation of Optimal Sparseness of Music Signals

La Cour-Harbo, Anders

Published in:
Proceedings of SPRRA 2006

Publication date:
2006

Document Version
Early version, also known as pre-print

Citation for published version (APA):
la Cour-Harbo, A. (2006). Fast Estimation of Optimal Sparseness of Music Signals. In Proceedings of SPRRA 2006 ACTA Press.

FAST ESTIMATION OF OPTIMAL SPARSENESS OF MUSIC SIGNALS

Anders la Cour-Harbo

Department of Control Engineering
Aalborg University, Denmark
alc@control.aau.dk

ABSTRACT

We want to use a variety of sparseness measured applied to 'the minimal ℓ_1 norm representation' of a music signal in an over-complete dictionary as features for automatic classification of music. Unfortunately, the process of computing the optimal ℓ_1 norm representation is rather slow, and we therefore investigate the use of matching pursuit, alternating projection, and Moore-Penrose inverse for estimating the result of applying two different sparseness measures to 'the minimal ℓ_1 norm representation' without actually computing this representation.

1. INTRODUCTION

Automatic processing of music signals is a key component in applications such as recognition, classification, thumb-nailing, watermarking, and transcription of music. This processing forms the basis for making high level decisions such as what type or category the music belongs to, which 10 seconds of the music is most descriptive, which instruments or notes are present in the music, and so on. The common factor in all these applications is the need for feature extraction, which basically means some sort conversion from the music signal to a feature vector.

1.1. Sparseness as Feature

We believe that some of these features might come from the use of sparseness measures applied to a representation of the music signal in an over-complete dictionary. The basic idea of over-complete dictionaries is to represent the music signal by M (or fewer) elements chosen from a dictionary with N atoms, where M is the length of the signal and $N \gg M$. This allows for many different valid representations, so while this idea is an extension of the traditional complete linear transform, like the Fourier transform, and therefore in theory is at least as good, it is often difficult (and extremely measure dependent) to determine which representation is desirable. In this presentation we opt for sparseness as being a good measure of 'desirable' as a sparse representation in some sense captures the features of the music signal (provided that the dictionary is reasonably chosen).

1.2. Finding the Sparsest Representation

The perhaps most obvious choice for the sparseness measure is the ℓ_0 norm since this measures the number of non-zero entries. Unfortunately, finding this particular representation is in general NP hard, and thus not feasible for even moderately sized problems. A good alternative is the ℓ_1 norm, and we therefore seek the solution to the standard over-complete dictionary problem

$$\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the dictionary (atoms as columns), $\mathbf{b} \in \mathbb{R}^M$ the signal to be represented, and $\operatorname{rank}(\mathbf{A}) = M \leq N$. The ratio N/M is the redundancy factor.

This can be solved in a variety of ways, for instance by linear programming [1], quadratic programming [2], minimum fuel neural networks [3, 4], and FOCUSS [5]. Sub-optimal solutions can be obtained for instance by the pseudo inverse (also called Moore Penrose inverse, method of frames [6]; they solve the problem for minimal ℓ_2 norm), various types of matching pursuit [7, 8], and best orthogonal basis (like cosine packets [9], wavelet packets [10] etc.).

For a discussion of ℓ_0 versus ℓ_1 and the use of ℓ_1 as sparseness measure, see for instance [1, 11, 12].

2. METHODS

The fundamental question in this presentation is: How well is it possible to estimated the sparseness (measured in two different ways) of the solution to (1). The solution to (1) can be found for instance by interior point linear programming (abbreviated IP in the following) or minimum fuel neural network, but unfortunately all known methods for solving the problem are computational quite expensive. Since we believe that the degree of sparseness of a music signal can be an important parameter in classification of music we want to find an alternative way of estimating the sparseness without actually calculating the ℓ_1 optimal solution.

2.1. Dictionary for Representation of Music

To allow for a high degree of sparseness in the representation the dictionary should contain atoms that are well suited for representing music. Obviously, some frequency localized atoms should be included. Also, in our experience, time-frequency localized atoms are useful for capturing certain parts of music. Wavelets is chosen as the mean to accomplish this. Consequently, the dictionary consists of a wavelet packet dictionary and a local cosine dictionary each with 5 levels. The wavelet is a Symlet 4 (8 filter taps and 4 vanishing moments). See [13] for a discussion on the choice of wavelet for music signal representation.

This work is supported by the Danish Technical Science Foundation (STVF) grant no. 56-00-0143 (WAVES)