Approximation of Piano Music with Exponential Damped Sinusoids

Martin Spiertz[1]

[1] *Institute for Communications Engineering, 52074 Aachen, Germany, email: spiertz@ient.rwth-aachen.de*

Introduction

A piano tone can be modeled as a short attack phase followed by a long decay of the string vibrations. This structure motivates an approximation of the decay by a sum of exponential damped sinusoids (EDS) for parameterization of a piano tone. A matching pursuit (MP) algorithm is used for extracting the EDS. The dictionary of the MP is adapted to the signal for reducing computational complexity. It can be shown, that in the case of piano music a large amount of signal energy can be represented by a small number of parameters. These parameters can be used for typical tasks in music analysis like music transcription or parameter based audio coding.

Basics

Matching Pursuit

A matching pursuit is an iterative algorithm for sparse signal decomposition by arbitrary basis functions (dictionary elements) [3]. It does not guarantee the sparsiest decomposition from a global point of view. Instead, it finds a nearly optimal decomposition by optimizing the energy reduction for each iteration. To start MP a residuum is initialized as the signal itself. At each iteration n the algorithm reduces the energy of the residuum r_{n+1} by subtracting the dictionary element d_{opt} multiplied with amplitude a_{opt}.

$$r_{n+1}(t) = r_n(t) - a_{opt} d_{opt} \tag{1}$$

The dictionary elements are typically designed as parameterized functions. In [3], a dictionary of gabor functions is used because of their optimal localization in time and frequency domain. The usage of EDS as dictionary elements is motivated by the asymmetric temporal structure of typical signals in audio processing as mentioned in [1]. Therefore in the following the MP with EDS as dictionary elements is discussed. EDS can be parameterized as follows[1]:

$$y(t) = ae^{-\frac{t-T}{\tau}} \cos(2\pi f(t-T) + \varphi)u(t) \tag{2}$$

An unit norm dictionary element can be defined by the dictionary parameters τ_k, f_k and φ_k as shown in equation 3.

$$d_k(t) = \frac{e^{-\frac{t}{\tau_k}} \cos(2\pi f_k t + \varphi_k)u(t)}{\sum_t (e^{-\frac{t}{\tau_k}} \cos(2\pi f_k t + \varphi_k)u(t))^2} \tag{3}$$

At each iteration n the metric defined in equation 4 is maximized by the optimal dictionary element d_{k_0} and the onset T_0. The amplitude a_0 as shown in equation 5 minimizes the energy of the residuum for this k_0 and T_0.

$$[k_0, T_0] = \underset{k,T}{\operatorname{argmax}} \, |\langle r_n, d_k(t-T)\rangle| \tag{4}$$

$$a_0 = \langle r_n, d_{k_0}(t-T_0)\rangle \tag{5}$$

Therefore the algorithm optimizes T_0 and a_0 for all d_k. After this, the amplitude a_0 and the corresponding parameters τ_0, f_0, and φ_0 of the dictionary element d_{k_0} are optimized by a least squares optimization for further minimization of the energy of the residuum. Here the Nelder-Mead simplex method is applied [2]. At least the residuum is updated as shown in equation 1.

It is possible to reduce the energy of the residuum below any threshold and therefore achieving an arbitrary accurate approximation of the signal just by applying a sufficient large number of iterations. For reducing evaluation time in our case, the stop condition is defined as an energy reduction from one iteration to the next lower than one percent of the energy of the signal itself:

$$\Delta E_n = \sum_t r_{n-1}^2(t) - \sum_t r_n^2(t) \tag{6}$$

$$\frac{\Delta E_n}{E_0} < \frac{1}{100} \tag{7}$$

The design of the dictionary for arbitrary signals is an open problem.

Constant Dictionary

In [1], a constant dictionary is used. Unlike the definition of d_k as shown in equation 3 the φ_k are not part of the set of parameters but are computed by a metric which finds a pair of complex conjugated EDS. In this case φ_k can be computed by the amplitude a_0, which is complex for this metric. Because a parameter set including φ_k as a third parameter leads to better signal decomposition by only few additional operations, the simulations for a constant dictionary use a parameter space with τ_k, f_k and φ_k. For reducing computational complexity, the τ_k, f_k and φ_k are quantized in an appropriate way. The dictionary is defined by all possible combinations of the quantized parameter values. Typically the quantization of parameter space leading to the sparsiest signal decomposition is not known a priori which leads to the following trade off:

Small Dictionaries reduce the computational complexity but lead to a suboptimal decomposition.

Large dictionaries usually enable a sparse decomposition of a wide range of signals at the cost of increasing computational complexity.

[1] $u(t)$: unit step

Adaptive Dictionary

To avoid the trade off mentioned earlier for constant dictionaries, the dictionary is adapted to signal properties. First the spectrogram R_n of the residuum r_n is computed by a short-time Fourier transformation. Frequencies of interest are evaluated as peaks in the cumulated energy function $E(f)$.

$$E(f) = \sum_t R_n(t, f) \tag{8}$$

For each peak in $E(f)$ at frequency f_k, a dictionary element d_k is estimated. The remaining parameters τ_k and φ_k are estimated by the signal $x(t) = R_n(t, f_k)$, which is a frequency-filtered and subsampled version of the underlying EDS.

$$\tau_k = \frac{\Delta t}{\log(a_{t_0}/a_{t_1})} \tag{9}$$

$$a_k = \frac{a_{max}}{|\sum_t w(t)e^{-t/\tau_k}\cos(2\pi f_k t)e^{-j2\pi f_k t}|} \tag{10}$$

$$\varphi_k = \underset{\varphi}{\operatorname{argmax}} \sum_t \left(a_k e^{-\frac{t}{\tau_k}}\cos(2\pi f_k t + \varphi) - r_n(t)\right)^2 \tag{11}$$

Where Δt denotes the hopsize and $w(t)$ the window function of the short-time Fourier transformation, a_{t_0} and a_{t_1} two neighbouring values of $|x(t)|$, and a_{max} the maximum amplitude of $|x(t)|$. For sparsier signal decomposition the dictionary adaption can be done several times during the MP.

Results

MP applied to single piano tones

As a first test set, the piano samples of the University of Iowa[2] are used. In Figure 1, the energy of the residuum normalized by the energy of the signal itself is plotted over the number of iterations. First, the MP with adaptive dictionary decomposes the signals of the test set. Then the constant dictionary is generated by quantization of this parameter space. Two cases are shown for the constant and the adaptive dictionary. First, the piano tone with lowest energy of residuum is shown. Second, the mean value of the energy of the residuum over all piano tones for each iteration is plotted. It can be seen that the MP with adaptive dictionary performs better than the MP with constant dictionary. Therefore only the performance of the adaptive dictionary is investigated further.

MP applied to complete songs

The performance of the adaptive dictionary is now evaluated on real piano music. The songs are segmented by an appropriate note based segmentation algorithm, e.g. [4]. After this, the MP algorithm is applied to each segment. For *Beethoven, Für Elise* the energy of the residuum can be reduced to 12.08% of the original signal energy. For this energy reduction 10.91 iterations per

[2] http://theremin.music.uiowa.edu/MIS.html

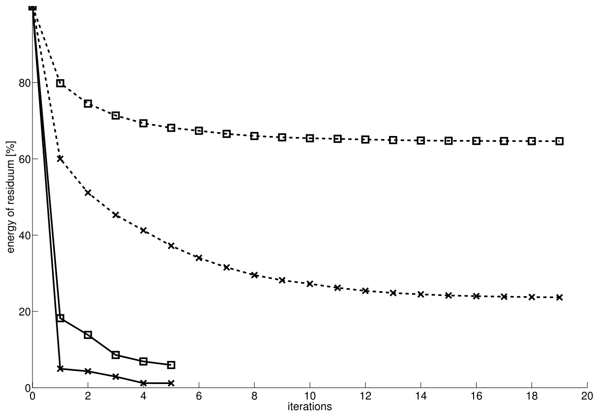

Figure 1: Comparison between constant and adaptive dictionary on single piano tones. The energy of the residuum normalized by the energy of the signal over the number of iterations is plotted. Best cases are plotted with solid lines, mean cases with dashed lines, squared marker are used for constant dictionary crossed marker for adaptive dictionary.

segment or equivalent 187.62 parameters per second are needed. Similar results are obtained for

- *Bach, Well-Tempered Piano, Part 1: Prelude A flat major*
- *Chopin, Nocturne E flat major Op.9 No.2*
- *Mozart, Sonata for Piano, B flat major, KV 281, First Movement*

Conclusions and Outlook

It can be shown that, for piano music, a compact parameterization of a large amount of signal energy with EDS is possible. Furthermore the adaptive dictionary performs better than the constant dictionary.
Future work will consider the usage of the EDS parameters as a feature vector for typical tasks in audio analysis, such as automatic music transcription, onset detection or parameter based audio coding.

References

[1] M.M. Goodwin and M. Vetterli. Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Transactions on Signal Processing*, 47(7):1890–1902, 1999.

[2] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder-Mead simplex algorithm in low dimensions. *SIAM Journal on Optimization*, 9:112–147, 1998.

[3] S.G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[4] J. Wellhausen. Note-accurate audio segmentation based on MPEG-7. In *Storage and Retrieval Methods and Applications for Multimedia 2004*, volume 5307 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 101–107, December 2003.