MIREX 2007: MULTIPLE FUNDAMENTAL FREQUENCY ESTIMATION AND TRACKING
HARMONIC NONNEGATIVE MATRIX APPROXIMATION APPROACH

Stanisław A. Raczyński   Nobutaka Ono   Shigeki Sagayama
The University of Tokyo
Graduate School of Information Science and Engineering
E-mail: {raczynski,onono,sagayama}@hil.t.u-tokyo.ac.jp

ABSTRACT

This paper gives details to a Harmonic Nonnegative Matrix Approximation (HNMA) based multipitch detection algorithm submitted to the 2007 edition of the MIREX competition.

1 INTRODUCTION

This article describes an algorithm for multipitch analysis, which aims to uncover the fundamental frequencies of simultaneously played harmonic sounds. The proposed procedure is built upon a method from the family of Nonnegative Matrix Approximations (NNMA), which, under different names and in different varieties, has recently received much attention, also from the music transcription community (e.g. NMF [7], NNSC [1], or SNMF2D [5]). As it will be shown later in this paper, nature of musical signals can be exploited to increase the transcription potential of the NNMA algorithm.

1.1 Generalized Nonnegative Matrix Approximation

Generalized Nonnegative Matrix Approximation (GNMA, described in [2]), is a method for decomposition of a nonnegative (having only nonnegative elements) matrix $\mathbf{X}$ (later referred to as the data matrix) into a multiplication of two, also nonnegative, matrices $\mathbf{A}$ and $\mathbf{S}$ (later refereed to as the basis matrix and the activity matrix, respectively):

$$\mathbf{X} \cong \mathbf{AS} = \widetilde{\mathbf{X}}. \tag{1}$$

The Generalized NNMA solves this problem by minimizing a Bregman divergence between the data matrix $\mathbf{X}$ and its approximation $\widetilde{\mathbf{X}}$. A special case of Bregman divergence is the I-divergence (generalized Kullback-Leibler divergence):

$$D_{KL}(\mathbf{P},\mathbf{Q}) = \left|\mathbf{P} \odot \log\frac{\mathbf{P}}{\mathbf{Q}} - \mathbf{P} + \mathbf{Q}\right|, \tag{2}$$

where the logarithm and the division are calculated element-wise, and $\odot$ is an element-wise multiplication. Using an I-divergence leads to the Nonnegative Matrix Factorization (NMF), for which Lee and Seung [3] has proposed a very fast multiplicative algorithm:

$$\mathbf{S} \leftarrow \mathbf{S} \odot \frac{\mathbf{A}^T\left(\frac{\mathbf{X}}{\mathbf{AS}}\right)}{\mathbf{A}^T\mathbf{1}}, \tag{3}$$

$$\mathbf{A} \leftarrow \mathbf{A} \odot \frac{\frac{\mathbf{X}}{\mathbf{AS}}\mathbf{S}^T}{\mathbf{1}\mathbf{S}^T}. \tag{4}$$

1.2 Penalized NNMA

By making the following assumption:

$$\nabla_{\mathbf{A}}\alpha(\mathbf{A}) \cong \nabla_{\mathbf{A}}\alpha(\mathbf{A})\bigg|_{\mathbf{A}=\mathbf{A}'}, \tag{5}$$

which is asymptotically true, as difference between $\mathbf{A}$ in consequent iterations tends to $\mathbf{0}$, we can modify the multiplicative update rules to include a penalizations on both matrices:

$$\mathbf{A} \leftarrow \mathbf{A} \odot \frac{\frac{\mathbf{X}}{\mathbf{AS}}\mathbf{S}^T}{\mathbf{1}\mathbf{S}^T + \nabla_{\mathbf{A}}\alpha(\mathbf{A})}, \tag{6}$$

$$\mathbf{S} \leftarrow \mathbf{S} \odot \frac{\mathbf{A}^T\frac{\mathbf{X}}{\mathbf{AS}}}{\mathbf{A}^T\mathbf{1} + \nabla_{\mathbf{S}}\beta(\mathbf{S})}. \tag{7}$$

For description of derivation of these rules, see [4]. It must be noted that the new update rules may result in the matrices $\mathbf{A}$ and $\mathbf{S}$ becoming negative, so caution must be taken while constructing the objective function.

2 PROCEDURE OVERVIEW

2.1 HNMA

Harmonic NMA (HNMA) imposes a harmonic constraint on the basis matrix in the penalized version of NMA. This constraint is enforced by initializing the basis matrix with zeros everywhere but at bins corresponding to frequencies of consecutive notes from an equal temperament musical scale, and the multiples of those frequencies (harmonic tones). Values initialized to zeros will not change due to the multiplicative nature of the update algorithm, forcing it to approximate the input data using only harmonically structured basis vectors.

In this implementation, a different set of penalty functions than in [4]. Only the following penalty function is used:

$$\alpha(\mathbf{A}) = \left| (\mathbf{A} - \mathbf{S}_D \mathbf{A} \mathbf{S}_R) \odot (\mathbf{A} - \mathbf{S}_D \mathbf{A} \mathbf{S}_R) \right|, \quad (8)$$

where $\mathbf{S}_D$ and $\mathbf{S}_R$ are the down-, and right-shifting matrices, respectively (identity matrices shifted down and right). In other words, it is a Froebenius norm of the difference between the basis matrix and its version shifted one vector right and one semitone down. This results in all basis vectors being similar to each other and in practice prevents their harmonic elements to grow above the fundamental one.

It can be shown that

$$\nabla_{\mathbf{A}} \alpha(\mathbf{A}) = 2\mathbf{A} - 2\mathbf{S}_D^T \mathbf{A} \mathbf{S}_R^T - 2\mathbf{S}_D \mathbf{A} \mathbf{S}_R + 2\mathbf{S}_D^T (\mathbf{S}_D \mathbf{A} \mathbf{S}_R) \mathbf{S}_R^T, \quad (9)$$

which is used with equation 6 to obtain the learning algorithm.

As the input of the HNMA algorithm, an amplitude spectrogram generated using the Constant-Q Transform was chosen. Such a representation has many advantages over the standard amplitude or power spectrogram – the number of frequency bins is smaller (which results in smaller amount of memory used and the HNMA algorithm running faster), the frequency scale is *de facto* logarithmic (so we can use the similarity penalty), and the frequency bins correspond to an equal temperament scale (in this algorithm a 36-Tone Equal Temperament scale), which also makes many tasks easier.

2.2 Postprocessing

Results of the HNMA algorithm are used to detect multiple pitches in each frame (see fig. 1 for details). Window of the median algorithm is set to be 9 frames wide (equivalent to 90 ms), and the threshold is fixed to a value of one tenth of the largest value in the smoothed coefficient matrix;

3 REFERENCES

[1] Abdallah, S.A. and Plumbley, M.D. “Unsupervised analysis of polyphonic music by sparse coding,” *IEEE Trans. on Neural Networks*, vol. 17, no. 1, pp. 179–196, 2006.

[2] Dhillon, I.S. and Sra, S. “Generalized Nonnegative Matrix Approximations with Bregman Divergences,” *Proc. Neural Information Processing Systems,* Vancouver, USA 2005.

[3] Lee, D.D. and Seung, H.S. “Learning the parts of objects by nonnegative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[4] Raczyński, S.A., Ono, N., Sagayama, S. “Multipitch analysis with Harmonic Nonnegative Matrix Approximation,” to be published in *Proc. 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007.

[5] Schmidt, M.N. and Mørup M. “Sparse Non-negative Matrix Factor 2-D Deconvolution for Automatic Transcription of Polyphonic Music,” *Proc. 6th International Symposium on Independent Component Analysis and Blind Signal Separation*, Charleston, USA, 2006.

[6] Sha, F. and Saul, L.K. “Real-Time Pitch Determination of One or More Voices by Nonnegative Matrix Factorization,” *Advances in Neural Information Processing Systems*, vol. 17, 2005.

[7] Smaragdis, P. and Brown. J.C. “Non-Negative Matrix Factorization for Polyphonic Music Transcription,” *Proc. 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, 2003.

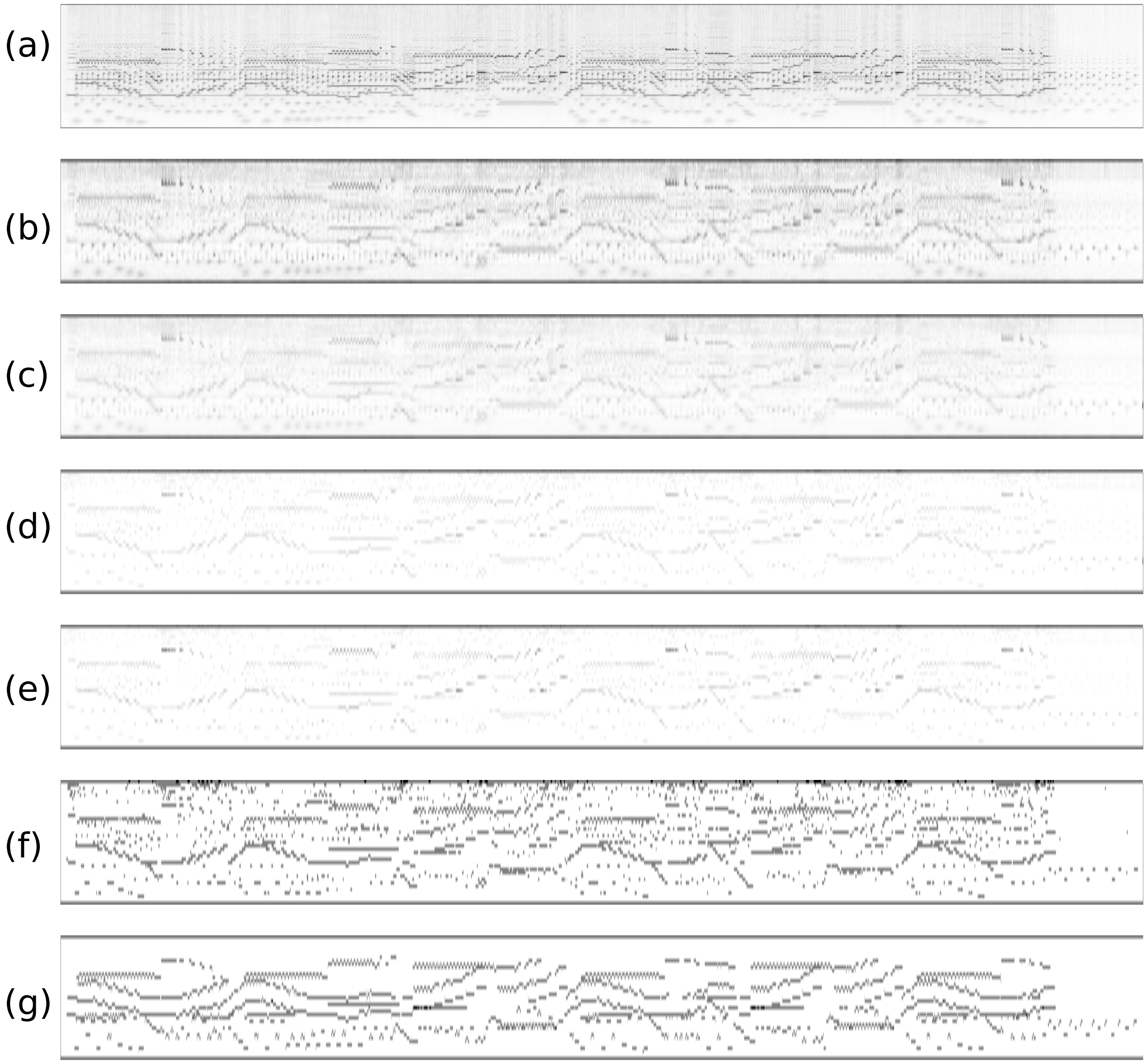

Figure 1. Stages of processing input data in the described algorithm. First a Constant-Q Transform is calculated (a), HNMA is performed to get the note coefficients (b), coefficients are smoothed (row-wise) with median filter (c), for each frame 5 largest peaks are found (d), results are again smoothed with a median filter (e) and finally thresholded (f). Subfigure (g) shows the groundtruth data for reference.