Harmonic Nonnegative Matrix Approximation for Multipitch Analysis of Musical Sounds *

Raczyński Stanisław A., Nobutaka Ono, Shigeki Sagayama
Graduate School of Information Science and Technology, The University of Tokyo

1 Introduction

Automatic music transcription of recorded music is usually a two stage process. The first stage is the event detection phase, where music events (note onsets, note offsets, pitch changes) are detected and identified. In the second stage, these events are transformed into a musical score. This paper focuses on the event detection stage, main part being multipitch analysis, which aims to uncover the fundamental frequencies of simultaneously played harmonic sounds. The procedure proposed in this paper is built upon a method from the family of Nonnegative Matrix Approximations (NNMA), which, under different names and in different varieties (e.g. NMF, NNSC, or SNMF2D), has recently received much attention, also from the music transcription community. 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 (described in [1]), 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 $\mathbf{X} \cong \mathbf{AS} = \widetilde{\mathbf{X}}$ (later refereed to as the basis matrix and the activity matrix, respectively). 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):

$$D_{KL}(\mathbf{P}, \mathbf{Q}) = \left| \mathbf{P} \odot \log \frac{\mathbf{P}}{\mathbf{Q}} - \mathbf{P} + \mathbf{Q} \right|, \quad (1)$$

where the logarithm, multiplication (denoted by $\odot$) and the division are calculated element-wise. Using the I-divergence leads to the Nonnegative Matrix Factorization (NMF), for which Lee and Seung [2] has proposed a very fast multiplicative update algorithm:

$$\mathbf{S} \leftarrow \mathbf{S} \odot \frac{\mathbf{A}^T \left(\frac{\mathbf{X}}{\mathbf{AS}} \right)}{\mathbf{A}^T \mathbf{1}}, \quad (2)$$

$$\mathbf{A} \leftarrow \mathbf{A} \odot \frac{\frac{\mathbf{X}}{\mathbf{AS}} \mathbf{S}^T}{\mathbf{1} \mathbf{S}^T}. \quad (3)$$

1.2 Penalized NNMA

By making the following assumption:

$$\nabla_{\mathbf{A}} \alpha(\mathbf{A}) \cong \nabla_{\mathbf{A}} \alpha(\mathbf{A}) \Big|_{\mathbf{A}=\mathbf{A}'}, \quad (4)$$

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 penalizations on both matrices. For description of derivation of these rules, see [3].

$$\mathbf{A} \leftarrow \mathbf{A} \odot \frac{\frac{\mathbf{X}}{\mathbf{AS}} \mathbf{S}^T}{\mathbf{1} \mathbf{S}^T + \nabla_{\mathbf{A}} \alpha(\mathbf{A})}, \quad (5)$$

$$\mathbf{S} \leftarrow \mathbf{S} \odot \frac{\mathbf{A}^T \frac{\mathbf{X}}{\mathbf{AS}}}{\mathbf{A}^T \mathbf{1} + \nabla_{\mathbf{S}} \beta(\mathbf{S})}. \quad (6)$$

2 Multipitch analysis procedure

The NNMA algorithm decomposes the data matrix $\mathbf{X}$, which contains contant-Q transforms of consecutive frames of musical signal to the basis matrix $\mathbf{A}$ and the activity matrix $\mathbf{S}$. The central frequencies of the constant-Q filters has been set to correspond to the frequencies of the notes of the fourty-eight-tone equal temperament scale (48-TET). Activity matrix is analyzed in the note detector described in section 2.3.

2.1 Matrix initialization in HNNMA

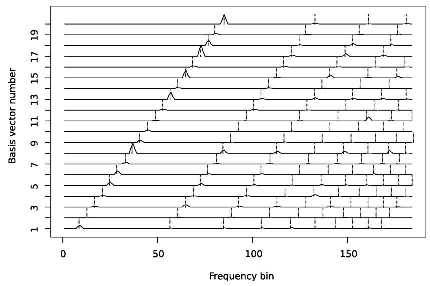
Because zero-valued elements of basis vectors will remain zero-valued throughout the learning process, we can initialize them to have zeros everywhere but at the positions of fundamentals of notes from a specific range of the 48-TET scale and their harmonics. That would guarantee that the basis vectors are sorted by their fundamental frequencies, and that corresponding rows in the activity matrix contain activities of consequent notes from that range, resulting in a harmonicaly-constrained NNMA. This would make analysis of the results of the algorithm straightforward – one would only have to analyze the note activities and find peaks corresponding to instances of these notes.

This technique is a good tradeoff between full basis estimation methods (such as NMF and other NNMA-based approaches) and methods that use pre-learned basis vectors. Figure 2.2 depicts an example basis matrix obtained using standard NNMA (NMF) method after fundamental frequency estimation and basis vector sorting and basis matrix obtained with HNNMA for the same data. The first matrix does not contain clear harmonic structure – many of the vectors have two (or more) dominant peaks, sometimes with highest peaks being the overtones instead of the fundamental, sometimes having the same fundamental frequency as different basis vectors.

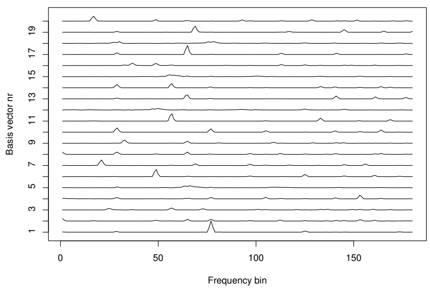
2.2 Additional penalties in HNNMA

HNNMA extends the regular NNMA to include additional penalties on the activity matrix $\mathbf{S}$. We

* 調波的非負行列近似による多重ピッチ解析、ラチンスキ・スタニスワブ、小野順貴、嵯峨山茂樹 (東大情報理工)



(a) Basis vectors after analysis with the HNNMA.



(b) Basis vectors after analysis with regular NMF

Fig. 1 Basis vectors after analysis of the *Ode to Joy*. Harmonic structure in (a) is clearly visible; vertical dotted lines indicate expected harmonic peaks positions.

Table 1 Piano pieces used for algorithm evaluation

Composer	**Title**	**Notes**	**Acc.**	**Corr.**
L. Beethoven	Symphony in D minor, Op. 125, No. 9 (last movement, Ode to joy)	101	96%	86%
F. Chopin	Nocturne in E# major, Op. 9, No. 2 (part)	328	87%	72%
F. Chopin	Nocturne in Bb minor, Op. 9, No. 1 (part)	358	70%	74%
J. S. Bach	Minuet No 4 in G	102	97%	100%

would like to find such an activity matrix that would be sparse, i.e. each row should contain only very few non-zero elements, contain mutually uncorrelated rows. The above can be reformulated, accordingly, in terms of a objective function β:

$$\begin{aligned}\beta(\mathbf{S}) &= -\mu_1 \left|\log(1+\mathbf{S}\odot\mathbf{S})\right| \\ &\quad +\mu_2\left(\left|\mathbf{S}^T\mathbf{S}\right| - \left|\mathbf{S}\odot\mathbf{S}\right|\right), \qquad (7)\end{aligned}$$

$$\nabla_{\mathbf{S}}\beta(\mathbf{S}) = -2\mu_1\mathbf{S}/\left(1+\mathbf{S}\odot\mathbf{S}\right)+2\mu_2\mathbf{S}\left(\mathbf{1}-\mathbf{I}\right). \quad (8)$$

2.3 Note detector

In our experiments a relatively simple note detector was used. After normalization of the note activities, the detector would mark all peaks above a fixed threshold. For each detected peak, local maxima that would have at least one local minimum below a certain threshold were marked as sub-peaks, and, consequently, as different notes.

3 Experimental results

All analyzed recordings were played on piano, as listed in Table 1, but experiments show equally good results for acoustic guitar and violin. The input data was first mixed down to a monaural signal and resampled to 11025 kHz. The constant-Q transform was calculated for frames shifted 12 ms. After learning the basis matrix contained very well structured vectors, each one having a stronger peak for the fundamental tone and weaker peaks for the harmonics (Figure 2.2). The results of note detection for few example pieces of music are presented in Table 1.

Fig. 2 Three bars from the middle of Chopin's Nocturne in E# major, Op. 9, No. 2

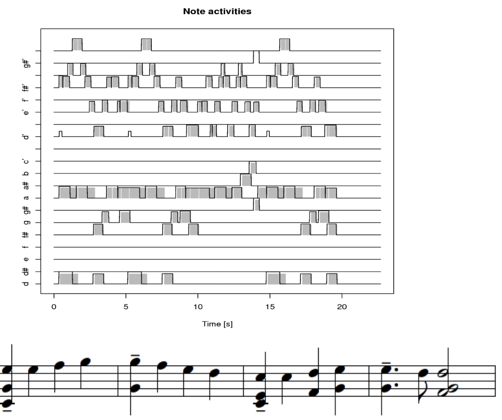


Fig. 3 Note detections (black lines) for *Ode to Joy* compared to the original notes (grey squares) and the first 4 bars of the song as a reference.

4 Conclusion

In this paper, we discussed the use of Harmonic Nonnegative Matrix Approximation for multipitch analysis of polyphonic music signals. By initializing the basis matrix with harmonic structure and using new penalties of sparsity and uncorrelation of rows of the activity matrix, this approach yielded higher note detection accuracy compared with previous extensions of the Nonnegative Matrix Approximation algorithm. The future work includes improving the post-processing of the HNNMA results by incorporating models of musical rhythm and harmonicity.

References

[1] I.S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with Bregman divergences. *Proc. Neural Information Processing Systems (NIPS) Conference*, 2005.

[2] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[3] N. Raczyński, S.A. Ono and S. Sagayama. Multipitch analysis with Harmonic Nonnegative Matrix Approximation. *Proc. 8th International Conference on Music Information Retrieval*, 2007.