日本音響学会誌 68 巻 11 号（2012），pp. 559–565

小特集—近年の音響信号処理における数理科学の進展—

非負値行列因子分解の音響信号処理への応用*

亀 岡 弘 和（東京大学/日本電信電話株式会社）**

43.10.Ln; 43.60.–c

1. は じ め に

実世界には，パワースペクトル，画素値，頻度など，非負値で表されるデータが多い。主成分分析や独立成分分析などの多変量解析では，所与のデータを複数の加法的な成分に分解することを目的とするが，これと同様に上述のような非負値のデータから構成成分を抽出することが役立つ場面が多い。例えば，複数の音源の音響信号が混在する多重音のパワースペクトルから個々の音源のパワースペクトルをうまく取り出すことができれば，雑音除去や音源分離などに役立てることができるし，顔画像データを目や鼻などの顔のパーツに該当する画像データにうまく分解することができれば，顔認証や顔画像合成などに役立てることができる。以上のように，非負値のデータを加法的な構成成分に分解することを目的とした多変量解析手法を非負値行列因子分解 (non-negative matrix factorization: NMF)[1] といい，様々な分野で近年注目を集めている。本稿では，NMF の定式化，基本的な性質，アルゴリズムの導出方法，音響信号処理の問題に焦点を当てた改良・拡張のアイディアについて解説する。

2. NMF とは

以後，データをベクトルで表記することとする。例えば，画像データであれば各ピクセルの画素値がベクトル要素となり，パワースペクトルであれば各周波数におけるパワー値がベクトル要素となる。今，N 個の非負値データベクトル $\boldsymbol{y}_1,\dots,\boldsymbol{y}_N \in \mathbb{R}^{\geq 0,K}$ が与えられたとしよう。これらを観測ベクトルと呼ぶ。ただし，$\mathbb{R}^{\geq 0,K}$ は K 次元の非負値ベクトル全体の集合を表す。NMF では，これら観測ベクトルが，いずれも M 個の基底ベクトルの適当な重みつき和によって表されたものと見なされ，すべての観測ベクトルを最も良く説明するような M 個の基底ベクトルおよび重み係数を推定することが目的である。よって，NMF では加法性が成り立つ量のみが対象となる。画素値やパワースペクトルはいずれも厳密には加法性が成り立つ量ではないが，NMF を適用する場面では近似的に加法性が成り立つと仮定されているということに注意が必要である。パワースペクトルの非加法性については 6 章 で詳しく述べる。以上の加法性に関する仮定の他に，基底ベクトルと重み係数がいずれも非負値であるという仮定が置かれる点も NMF における重要なポイントである。すなわち，NMF は，観測ベクトル $\boldsymbol{y}_n$ を基底ベクトル $\boldsymbol{h}_1,\dots,\boldsymbol{h}_M \in \mathbb{R}^{\geq 0,K}$ の非負結合 (結合係数 $u_{1,n},\dots,u_{M,n}$ が非負値の線形結合)

$$\boldsymbol{y}_n \simeq \sum_{m=1}^{M} \boldsymbol{h}_m u_{m,n} \quad (n=1,\dots,N) \qquad (1)$$

で近似する問題と見なせる。ここで，観測ベクトルを並べた行列を $\boldsymbol{Y} = [\boldsymbol{y}_1,\dots,\boldsymbol{y}_N] = (y_{k,n})_{K\times N}$，基底ベクトルを並べた行列を $\boldsymbol{H} = [\boldsymbol{h}_1,\dots,\boldsymbol{h}_M] = (h_{k,m})_{K\times M}$，結合係数 $u_{m,n}$ を m 行 n 列の要素とした行列を $\boldsymbol{U} = (u_{m,n})_{M\times N}$ とすると，式 (1) は $\boldsymbol{Y} \simeq \boldsymbol{HU}$ と同じ意味である。このように NMF は，観測ベクトルを並べた行列を二つの非負値行列の積に分解する問題と捉えることができる。これが，非負値行列因子分解と呼ばれる所以である。NMF のイメージを掴んでもらうために，[1]スペクトログラムを行列と見なして NMF を適用した例を図 1 に示す。

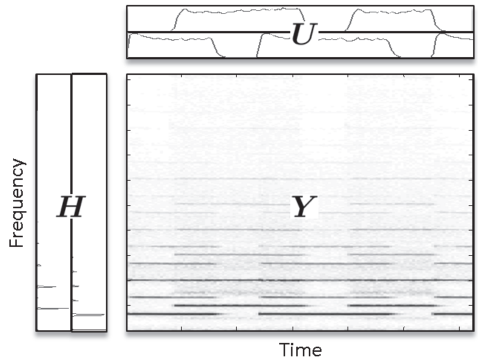
3. NMF の基本性質

NMF では通常，基底数 M は観測ベクトルの次元 K やデータの個数 N より小さく設定される。

[1]各時刻周辺における信号のスペクトルを時間順に並べて表示したものをスペクトログラムと呼ぶ。

* Non-negative matrix factorization with application to audio signal processing.

** Hirokazu Kameoka (University of Tokyo / NTT)



図–1 NMF によるスペクトログラムの分解

例えば $M = K$ の場合，$\boldsymbol{H} = \boldsymbol{I}$ ($\boldsymbol{I}$ は単位行列) であるような分解表現 $\boldsymbol{Y} = \boldsymbol{HU}$ を得ることができるが，明らかにこの分解は意味をなさない。また，$M = N$ の場合，$\boldsymbol{U} = \boldsymbol{I}$ であるような分解表現 $\boldsymbol{Y} = \boldsymbol{HU}$ を得ることができるが，この分解からも意味を見出せない。$M < \min(K, N)$ のとき，NMF は観測行列 $\boldsymbol{Y}$ を低いランクの行列で近似しようとしていることに相当し，その場合に求まる基底行列と係数行列が重要な意味をもつ。

NMF では，観測データの中で共起する成分をひとまとめにしたものが基底ベクトルの推定結果になる傾向がある。例えば，どのデータも成分 a, b，c から成っているとき，a，b，c を基底にすれば全データを完全に表現することは可能である。ここでもし，成分 a や成分 b が生起するときいつも成分 c も揃って生起するようなら，成分 a と成分 c，および成分 b と成分 c をひとまとめにしたものを基底としても当該データを同様に表現可能である。この方が，少ない基底でデータを表現可能な分，より「節約的」である。このように，共起する成分をひとまとめにして基底と置いた方が，基底を節約でき，節約した分を他の成分にフィットするのに充てられるようになるわけである。先に述べたように，NMF では少ない基底で観測データをできるだけ良く表す必要があるため，上記のような基底が自然に得られる傾向にある。

NMF では，係数行列の非負性の制約により，基底同士を加算することしかできない (減算ができない) ため，係数行列の要素がスパース (疎) になる傾向がある。これはベクトルの合成を考えるとイメージしやすい。観測ベクトルを近似するには，観測ベクトルと近い方向を向いた基底ベクトル以外の係数はできるだけ小さくした方が良い。スパース性は統計的独立性と概念的に関連が深く，上述の共起する成分をひとまとめにしたものを基底にする働きは，なるべく各基底の係数が互いにコヒーレントにならないように基底を決定しようとする働きと見なすことができる。すなわち，NMF では，非負制約による副次的な効果として係数が互いに独立になるように基底が求まる傾向にある。

4. NMF アルゴリズム

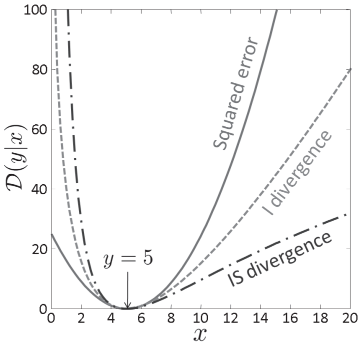
4.1 正値行列因子分解と NMF

観測行列を二つの非負値行列の積で表そうという NMF の基本概念自体は，Paatero らによって最初に提案されている [2]。Paatero らは誤差行列 $\boldsymbol{Y} - \boldsymbol{HU}$ の Frobenius ノルム (行列の要素の二乗和) で $\boldsymbol{HU}$ の $\boldsymbol{Y}$ からの乖離度 $D_{\mathrm{EU}}(\boldsymbol{H}, \boldsymbol{U})$ を定義し，さらに，$\boldsymbol{H}$ と $\boldsymbol{U}$ の各行列要素が非負であることを保証する目的で，各行列要素が 0 になろうとすると無限大のペナルティを課す対数障壁関数を定義し，これに適当な係数を掛けて乖離度 $D_{\mathrm{EU}}(\boldsymbol{H}, \boldsymbol{U})$ に足したものを目的関数とした最適化アルゴリズムを提案している。Paatero らの方法は，最終的に得られる $\boldsymbol{H}$ と $\boldsymbol{U}$ が上記の障壁関数の性質より正値行列に限られることから，positive matrix factorization (PMF) と呼ばれる。一方，上述のような障壁関数を用いずとも行列要素の非負性を保証しながら $\boldsymbol{Y} \simeq \boldsymbol{HU}$ となる $\boldsymbol{H}$ と $\boldsymbol{U}$ を効率的に得る反復アルゴリズムが Lee らによって考案され [1]，これをきっかけに NMF が広く知れ渡るようになった。

4.2 行列の乖離度

NMF は，$\boldsymbol{HU}$ の $\boldsymbol{Y}$ からの乖離度を表す規準の定義に応じて異なる最適化問題に帰着する。Lee らは，前述の Frobenius ノルム規準に加え，I ダイバージェンス規準の NMF アルゴリズムを導出している [1]。当然ながら異なる乖離度規準を用いた場合では異なる解が最適解として得られる。このため，最適解が所望の解であるためには，背後にあるデータの生成プロセスに合った適切な規準を設定することが重要である。例えば，パワースペクトルを対象とした NMF では板倉斎藤擬距離が乖離度規準として用いられることがある [3] が，これは 6.2 節で述べる信号波形の生成プロセスに関する仮定に基づいている。

$y, x \in \mathbb{R}$ とすると，y の x からの二乗誤差，I ダイバージェンス，板倉斎藤擬距離はそれぞれ



図–2 $\mathcal{D}_{\mathrm{EU/KL/IS}}(y|x)$ を x の関数と見たときのグラフ

$$\mathcal{D}_{\mathrm{EU}}(y|x) = (y-x)^2 \tag{2}$$

$$\mathcal{D}_{\mathrm{KL}}(y|x) = y\log\frac{y}{x} - y + x \tag{3}$$

$$\mathcal{D}_{\mathrm{IS}}(y|x) = \frac{y}{x} - \log\frac{y}{x} - 1 \tag{4}$$

で与えられる。いずれも，$x=y$ のときにのみ 0 となり，x が y から離れるほど増大する関数である。図 2 にそれぞれを x の関数と見たときのグラフを示す。二乗誤差は y を中心に対称であるのに対し，I ダイバージェンスと板倉斎藤擬距離は非対称であり，x が y を下回る場合に，より過大なペナルティを課す関数であることが分かる。また，板倉斎藤擬距離は y と x の比のみで表される関数であるため，y と x のスケールに非依存である。これらを用いれば，$\boldsymbol{HU}$ の $\boldsymbol{Y}$ からの乖離度を，

$$D_{\cdot}(\boldsymbol{H},\boldsymbol{U}) = \sum_{k,n}\mathcal{D}_{\cdot}\Big(y_{k,n}\Big|\sum_m h_{k,m}u_{m,n}\Big)$$

で測ることができる ($\cdot$ は EU/KL/IS を表す)。

4.3 補助関数法

これらを最小化する非負値の $\boldsymbol{H}$ と $\boldsymbol{U}$ を求めるのが目的であるが，いずれも非負制約つき非線形最適化問題であり，解析的に解くことはできない。もし，目的関数が行列要素 $h_{k,m}, u_{m,n}$ ごとの関数の和に分離した形をしていれば，各行列要素ごとに非負値制約の中で解を探索することができて好都合であるが，残念ながらこれらの目的関数ではそうはなっていない。Lee らは，目的関数の上限となる補助関数を反復的に降下させることで目的関数を間接的に降下させていく，補助関数法と呼ぶ方法をベースに，補助関数として行列要素 $h_{k,m}$, $u_{m,n}$ ごとの関数の和に分離した形をとるものをうまく設計することで，当該制約付き非線形最適化問題の解を見通し良く探索することができると考えた。補助関数の定義と補助関数法の原理は以下のとおりである。

定義 1. $\theta = \{\theta_i\}_{1\le i\le I}$ をパラメータとする目的関数 $D(\theta)$ に対し，$D(\theta) = \min_{\bar\theta} G(\theta,\bar\theta)$ が成り立つとき，$G(\theta,\bar\theta)$ を $D(\theta)$ の**補助関数**と定義する。

定理 1 (補助関数法)**.** 補助関数 $G(\theta,\bar\theta)$ を，$\bar\theta$ に関して最小化するステップと，$\theta_1,\cdots,\theta_I$ に関して最小化するステップ

$$\bar\theta \leftarrow \underset{\bar\theta}{\mathrm{argmin}}\, G(\theta,\bar\theta) \tag{5}$$

$$\theta_i \leftarrow \underset{\theta_i}{\mathrm{argmin}}\, G(\theta,\bar\theta) \quad (i=1,\ldots,I) \tag{6}$$

を繰り返すと，目的関数 $D(\theta)$ の値は単調減少する。

本小特集に小野氏による補助関数法に関する解説があるので，詳細はそちらを参照して頂きたい。余談だが，著者は大学院生時代，不完全データに対する確率モデルのパラメータ推定手法である Expectation-Maximization アルゴリズムの拡張解釈を通し，Lee らとは独立に[2] 補助関数法を考案した。そして，当時の研究室スタッフであった小野氏とその性質や補助関数の設計方法について多くの議論を交わし，ともにこれまで補助関数法の音響信号処理における様々な最適化問題への応用に取り組んできた。小野氏の解説記事にはその取り組みの一部が紹介されているので，補助関数法の NMF 以外への適用例に興味のある読者は是非参照して頂きたい。

4.4 $D_{\mathrm{EU/KL}}(\boldsymbol{H},\boldsymbol{U})$ を規準とした NMF

上述の補助関数法の原理に基づき，まず Lee らによる $D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U})$ を規準とした NMF アルゴリズム [1] の導出の概要を示す。$D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U})$ を展開し，$\boldsymbol{H}$ と $\boldsymbol{U}$ に依らない項を省略すると，

$$D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U}) \overset{H,U}{=} \sum_{k,n}(-2y_{k,n}x_{k,n} + x_{k,n}^2) \tag{7}$$

と書ける。ただし，

$$x_{k,n} = \sum_m h_{k,m}u_{m,n} \tag{8}$$

[2] 実は「補助関数法」という呼称は当時の指導教員と小野氏と共に考えたものである。画像処理分野では Majorization-Minimization アルゴリズムという呼び方が一般的なようである。なお，著者が NMF に興味をもったのは，NMF で補助関数法と等価な最適化手法が用いられていることを後に知ったのがきっかけである。

である。また，$\overset{z}{=}$ は z に関係する項のみに関する等号を表す。$D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U})$ において $x_{k,n}^2$ が，$\boldsymbol{H}$ と $\boldsymbol{U}$ の行列要素 $h_{k,1},\ldots,h_{k,M}$, $u_{1,n},\ldots,u_{M,n}$ を含んだ非線形関数項である。ここで，この項に対し，行列要素 $h_{k,m}$, $u_{m,n}$ ごとの関数の和に分離した形をした上限関数を設けたい。上述のような上限関数は，2 次関数が凸関数であることを利用し，以下の Jensen の不等式を用いることで設計可能である。Jensen の不等式とは，任意の凸関数 g，I 個の実数 $x_1,\ldots,x_I$，$\sum_i \lambda_i = 1$ を満たす正の重み係数 $\lambda_1,\ldots,\lambda_1$ のもとで成り立つ

$$g\Big(\sum_i \lambda_i z_i\Big) \le \sum_i \lambda_i g(z_i) \tag{9}$$

(等号成立は $z_1=\cdots=z_I$) のような不等式のことをいう。これを用いると，$x_{k,n}^2$ に対し，

$$x_{k,n}^2 \le \sum_m \lambda_{k,m,n}\left(\frac{h_{k,m}u_{m,n}}{\lambda_{k,m,n}}\right)^2 \tag{10}$$

のような不等式を立てることができる。ただし，$\lambda_{k,m,n}>0$, $\sum_m \lambda_{k,m,n}=1$ であり，等号は

$$\lambda_{k,m,n} = \frac{h_{k,m}u_{m,n}}{x_{k,n}} \tag{11}$$

のとき成立する。以上より，$D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U})$ における $x_{k,n}^2$ を式 (10) の右辺に置き換えることにより得られる関数 $G_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U},\boldsymbol{\lambda})$ は $D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U})$ の補助関数の要件を満たす。ただし，$\boldsymbol{\lambda} = \{\lambda_{k,m,n}\}_{K\times M\times N}$ とする。補助関数が得られれば，

$$\boldsymbol{\lambda} \leftarrow \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\, G_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U},\boldsymbol{\lambda}) \tag{12}$$

$$\boldsymbol{H} \leftarrow \underset{\boldsymbol{H}}{\operatorname{argmin}}\, G_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U},\boldsymbol{\lambda}) \tag{13}$$

$$\boldsymbol{U} \leftarrow \underset{\boldsymbol{U}}{\operatorname{argmin}}\, G_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U},\boldsymbol{\lambda}) \tag{14}$$

を反復的に行えば $D_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U})$ を単調に降下させていくことができる。式 (12) については既に述べたとおり式 (11) で与えられる。式 (13) と式 (14) は $h_{k,m}\ge 0$, $u_{m,n}\ge 0$ の制約のもとで解かれなければならない点に注意が必要だが，有り難いことに $G_{\mathrm{EU}}(\boldsymbol{H},\boldsymbol{U},\boldsymbol{\lambda})$ は行列要素ごとの関数の和に分かれた形をしているので，これらはいずれも行列要素ごとの一変数関数の非負制約つき最小化問題に帰着し，$\partial G_{\mathrm{EU}}/\partial h_{k,m}=0$, $\partial G_{\mathrm{EU}}/\partial u_{m,n}=0$ を解くことで更新式を容易に得ることができる。

$D_{\mathrm{KL}}(\boldsymbol{H},\boldsymbol{U})$ を規準とした NMF アルゴリズムも以上と同様にして導くことができる。負の対数関数が凸関数であることに注意すると，$-\log x_{k,n}$ の項に対し Jensen の不等式

$$-\log x_{k,n} \le -\sum_m \lambda_{k,m,n}\log\left(\frac{h_{k,m}u_{m,n}}{\lambda_{k,m,n}}\right)$$

が立てられる。これにより行列要素 $h_{k,m}$, $u_{m,n}$ ごとの関数の和に分離した形の補助関数を設けることができ，導出は以下同様である。

4.5 $D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U})$ を規準とした NMF

次に，著者らによる，板倉斎藤擬距離 $D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U})$ を規準とした NMF アルゴリズムの導出 [4] を以下に示す。$\boldsymbol{H}$, $\boldsymbol{U}$ に依らない項を省略すると，$D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U})$ は

$$D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U}) \overset{H,U}{=} \sum_{k,n}\left(\frac{y_{k,n}}{x_{k,n}} + \log x_{k,n}\right) \tag{15}$$

と書ける。先と同様の方針で補助関数法により当該非負制約つき最適化問題を解決するためには，$D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U})$ に対し，行列要素 $h_{k,m}$, $u_{m,n}$ ごとの関数の和に分離した形をした上限関数を設計できるかどうかがやはり鍵となる。まず，逆数関数は正領域において凸関数であるため，$1/x_{k,n}$ の項に関しては，Jensen の不等式

$$\frac{1}{x_{k,n}} \le \sum_m \lambda_{k,m,n}\left(1\middle/\frac{h_{k,m}u_{m,n}}{\lambda_{k,m,n}}\right) \tag{16}$$

が成り立つ。ただし，$\lambda_{k,m,n}$ は先と同様，$\lambda_{k,m,n}>0$, $\sum_m \lambda_{k,m,n}=1$ を満たす。次に，$\log x_{k,n}$ の項に関しては，正の対数関数は凹関数なので Jensen の不等式では上限関数が作れない。ここで，任意の微分可能な凹関数 g に対し，

$$g(x) \le g(\alpha) + (x-\alpha)g'(\alpha) \tag{17}$$

(等号成立は $x=\alpha$) が成り立つことを利用すると，

$$\log x_{k,n} \le \log\alpha_{k,n} + \frac{1}{\alpha_{k,n}}(x_{k,n}-\alpha_{k,n}) \tag{18}$$

のような不等式が立てられ，行列要素の 1 次式で表された上限関数を設計できる。以上の二つの不等式の等号成立条件は $\lambda_{k,m,n}$ と $\alpha_{k,n}$ がそれぞれ

$$\lambda_{k,m,n} = \frac{h_{k,m}u_{m,n}}{x_{k,n}},\quad \alpha_{k,n} = x_{k,n} \tag{19}$$

のときである。以上より，$D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U})$ における $1/x_{k,n}$ の項を式 (16) の右辺に，$\log x_{k,n}$ の項を式 (18) の右辺に置き換えることにより得られる関数

$G_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U},\boldsymbol{\lambda},\boldsymbol{\alpha})$ は $D_{\mathrm{IS}}(\boldsymbol{H},\boldsymbol{U})$ の補助関数の要件を満たす。あとは定理 1 に従って各更新式を求めれば板倉斎藤擬距離規準の NMF アルゴリズムが導かれる。

二乗誤差，I ダイバージェンス，板倉斎藤擬距離は，β ダイバージェンスと呼ぶ規準で統一的に記述することができ，これを規準とした NMF アルゴリズムが提案されている [5]。導出のポイントは前述の著者らのアイディア [4] と同様であり，凸関数の項に対しては式 (9) を，凹関数の項に対しては式 (17) の不等式を用いて補助関数を設計することができ，閉形式の更新式を求めることができる。

5. 観測行列の生成プロセス

二乗距離，I ダイバージェンス，板倉斎藤距離を規準とした NMF は，観測行列の要素 $y_{k,n}$ が $x_{k,n}$ を平均とした正規分布，Poisson 分布，指数分布

$$y_{k,n} \sim \mathcal{N}(y_{k,n}; x_{k,n}, \sigma^2) \tag{20}$$

$$y_{k,n} \sim \mathrm{Poisson}(y_{k,n}; x_{k,n}) \tag{21}$$

$$y_{k,n} \sim \mathrm{Exponential}(y_{k,n}; x_{k,n}) \tag{22}$$

に従って独立に生成されたと仮定した場合の $\boldsymbol{H}$, $\boldsymbol{U}$ の最尤推定問題と各々等価である。ただし，

$$\mathcal{N}(z;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(z-\mu)^2/2\sigma^2} \tag{23}$$

$$\mathrm{Poisson}(z;\mu) = \mu^z e^{-\mu}/z! \quad (z \geq 0) \tag{24}$$

$$\mathrm{Exponential}(z;\mu) = \frac{1}{\mu} e^{-z/\mu} \quad (z \geq 0) \tag{25}$$

である。このことは，以下により確かめられる。式 (20)〜(22) により立てられる $x_{k,n}$ の対数尤度 $L(x_{k,n}) = \log p(y_{k,n}|x_{k,n})$ はいずれも $x_{k,n} = y_{k,n}$ のときに最大となる。すなわち，$L(y_{k,n}) \geq L(x_{k,n})$ である。よって，対数尤度差 $L(y_{k,n}) - L(x_{k,n})$ は，$x_{k,n} = y_{k,n}$ のときにのみ 0 になる，$y_{k,n}$ と $x_{k,n}$ の近さを表す非負の尺度と見なせ，式 (20)〜(22) の場合の対数尤度差 $L(y_{k,n}) - L(x_{k,n})$ は，それぞれ式 (2)〜(4) と等しい。

6. スペクトログラムモデルとしての NMF

Smaragdis らは，音楽音響信号の自動採譜を目的とし，音楽音響信号の (振幅またはパワー) スペクトログラムを観測行列と見なして NMF を適用し，楽音ごとのスペクトログラムに分解する方法 [6] を提案している。瞬時瞬時の観測スペクトルが構成音のスペクトルの重みつき和によって表される (スペクトルは加法的である)，という仮定と，各構成音は周波数成分比が時不変でゲインのみが時間変化する，という仮定の下で，観測スペクトログラムから個々の構成音のスペクトル (周波数成分比) と各時刻におけるゲインを推定する問題は，NMF と同形の問題と見なせることが本アプローチのポイントである。例えば，音楽は平均律音階に該当する限られた種類の音高の音で構成されるため，各音高の音が時刻に依らずいつも同じ周波数成分比を有すると仮定できる場合，音楽音響信号のスペクトログラムに NMF を適用すると各基底ベクトルが一つの音階の音のスペクトルとなり，結合係数が各時刻におけるそれぞれの音階の音のゲインとなるように得られる。以上のアプローチの有効性は経験的に知られているが，上述の，核となる 2 つの仮定は実際には成り立たない。本章では，これらの仮定をより現実に即したものに改変した改良モデルを紹介する。

6.1 複素 NMF[7]

音の信号波形は加法的である。短時間 Fourier 変換やウェーブレット変換などの時間周波数分解は基本的には信号波形の線形変換であるため，時間周波数分解により得られる複素スペクトログラムも加法的である。しかし，複素スペクトログラムから振幅 (またはパワー) スペクトログラムへの変換は非線形であるため，(振幅およびパワー) スペクトログラムは実際には非加法的である。端的に言えば，波形 A と波形 B の和のスペクトルは波形 A のスペクトルと波形 B のスペクトルの和とは等しくなると限らない。従って，振幅 (またはパワー) スペクトログラムを加法的な成分に分解したところで，それぞれの成分が物理的に何に対応するかは定かではない。

NMF をスペクトログラムに適用するアプローチでは，周波数成分比さえ同じであれば波形 (ゲインや位相スペクトル) が異なっていたとしても同一音と見なそう，という考え方がベースとなっているが，著者らはこの考え方を基に複素スペクトログラムをモデル化し，観測複素スペクトログラムを加法的な成分に分解する「複素 NMF」と呼ぶ方法を提案している [7]。

6.2 板倉斎藤擬距離規準の NMF[3]

先述のとおりパワースペクトル自体は加法的ではないが，各構成音の信号が互いに統計的に独立のときそれぞれのパワースペクトルの期待値は加

法的となる。これは，NMF において置かれるスペクトルの加法性の仮定が，期待値の意味では正当化される場合があるということを示唆している。

多重音信号中の各構成音の信号が短時間区間毎に平均が 0 の (巡回) 定常 Gauss 過程に従って生成されたと仮定すると，各区間における離散 Fourier 変換の各周波数成分は平均が 0 の複素正規分布に従う。すなわち，区間 n における m 番目の構成音の周波数 k の成分 (構成音 m の複素スペクトログラム) を $s_{k,n}^{(m)}$ とすると，$s_{k,n}^{(m)}$ は $s_{k,n}^{(m)} \sim \mathcal{N}_{\mathbb{C}}(s_{k,n}^{(m)}; 0, \nu_{k,n}^{(m)})$ に従う。ただし，$\mathcal{N}_{\mathbb{C}}(z; \mu, \nu) = \frac{1}{\pi\nu} e^{-|z-\mu|^2/\nu}$ である。$\nu_{k,n}^{(m)}$ は構成音 m のパワースペクトログラムの期待値 $\mathbb{E}[|s_{k,n}^{(m)}|^2]$ を表すパラメータである。ここで，観測信号の複素スペクトログラム $y_{k,n}$ が $y_{k,n} = \sum_m s_{k,n}^{(m)}$ のように構成音の複素スペクトログラムの和で与えられ，$s_{k,n}^{(m)}$ と $s_{k,n}^{(m')}$ $(m \neq m')$ が互いに独立と仮定できるなら，$y_{k,n} \sim \mathcal{N}_{\mathbb{C}}(y_{k,n}; 0, \sum_m \nu_{k,n}^{(m)})$ に従う。よって，$x_{k,n} = \sum_m \nu_{k,n}^{(m)}$ と置くと，$y_{k,n}$ が与えられたもとでの $x_{k,n}$ の対数尤度は $L(x_{k,n}) = -\log \pi x_{k,n} - |y_{k,n}|^2/x_{k,n}$ となる。この対数尤度は $x_{k,n} = |y_{k,n}|^2$ のとき最大となるので $L(|y_{k,n}|^2) \geq L(x_{k,n})$ である。実は，対数尤度差 $L(|y_{k,n}|^2) - L(x_{k,n}) \geq 0$ は，$|y_{k,n}|^2$ と $x_{k,n}$ の板倉齋藤擬距離 $\mathcal{D}_{\mathrm{IS}}(|y_{k,n}|^2|x_{k,n})$ と等しい。従って，各構成音のパワースペクトログラムの期待値を表す $\nu_{k,n}^{(m)}$ に関し，周波数成分比が時不変であるような構造 $\nu_{k,n}^{(m)} = h_{k,m} u_{m,n}$ を仮定すれば，$\boldsymbol{H} = (h_{k,m})_{K \times M}$ と $\boldsymbol{U} = (u_{m,n})_{M \times N}$ の最尤推定問題は，観測パワースペクトログラム $|y_{k,n}|^2$ を要素にもつ行列 $\boldsymbol{Y}$ に対し板倉斎藤擬距離規準の NMF を行うことと等価となる [3]。

6.3 可変基底 NMF[8–10]

NMF をスペクトログラムに適用するアプローチでは基本的に，分解したい各構成音の周波数成分比は時不変であるという仮定が置かれるが，音楽音響信号を楽音ごとに分解する問題を扱う上では必ずしもこの仮定は成り立たない。歌声や音声はもちろん，ピアノやヴァイオリンの各音符に対応する音響信号のスペクトルは時々刻々と変化する。このように同一音源の音が持続していたとしてもスペクトルが時間変化する場合，通常の NMF を適用しても，時刻ごとのスペクトルが別の音としてばらばらに分解されてしまうことが多い。そこで，観測スペクトログラムを個々の持続音ごとに分解することを目的として，基底スペクトルが時間変化するよう改良された NMF の拡張モデルが提案されている [8–10]。

6.4 その他の拡張モデル

NMF の楽音分離や自動採譜への応用においては，音楽スペクトログラムを各楽音のノートごとのスペクトログラムに分解することが目的となるが，その成功の鍵は，楽音がもつ構造，性質，傾向をいかに見つけ，NMF モデルにどう組み込めるか，にある。このような動機からさまざまな拡張モデルが提案されている。例えば，各基底スペクトルのゲインが時間方向に連続的となるよう制約が組み込まれたもの [11]，各基底スペクトルが調波構造をなすよう制約されたもの [12]，各基底スペクトルにソースフィルタモデルの拘束が組み込まれたもの [13, 14]，基底スペクトルが音色特徴量空間においていくつかのクラスタを形成するよう基底のクラスタリングと NMF によるスペクトログラム分解が同時に行われるもの [15]，などが検討されている。

NMF において基底数をいかに決定するかは重要課題の一つである。Cemgil らは，5 章で述べたような NMF の生成モデルとしての解釈を通して，NMF の基底数決定問題を周辺尤度 ($\boldsymbol{H}$ と $\boldsymbol{U}$ に関して周辺化したデータ行列の生成確率 $p(\boldsymbol{Y})$) に基づくモデル選択の問題として定式化している [16]。モデル選択により NMF の基底数を決定するためには，さまざまな基底数のもとで NMF を実行し，情報量規準や周辺尤度を算出して比較する手続きが必要となる。これに対し，パラメータとモデルの複雑度を同時に推論する枠組であるノンパラメトリックベイズアプローチにより NMF の基底数を決定する方法も提案されている [10, 17]。

7. 音響信号処理への応用

NMF は，自動採譜の他，音声強調・分離 [18]，帯域拡張 [19]，音楽信号からのボーカル成分抽出 [20]・ドラム音抽出 [21]，音声認識のための音素特徴量抽出 [22]，ホルマントトラッキング [23]，エコーキャンセラ [24] などの音響信号処理問題に応用展開されており，著者らも NMF のアルゴリズムをヒントにしたブラインド残響除去法を提案している [25]。また，NMF はモノラル信号の音源分離のための有効なアプローチとして認識されるよう

になって以来，そのモデル化の考え方は多チャネル信号の音源分離にも効果的であろうという期待から，NMF の多チャネル拡張に関する検討も近年進められている [26–29]。

8. ま と め

本稿では，NMF の基本性質，アルゴリズムの導出方法，生成モデルとしての解釈，音響信号処理への応用とそのための拡張モデルについて解説した。より理解を深めたい読者は他の解説記事 (例えば [30, 31]) も是非とも参照して頂きたい。

文 献

[1] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in Adv. NIPS, pp. 556–562 (2000).

[2] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," Environmetrics, vol. 5, pp. 111–126 (1994).

[3] C. Févotte, N. Bertin and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," Neural Computation, vol. 21, no. 3, pp. 793–830 (2009).

[4] 亀岡弘和, 後藤真孝, 嵯峨山茂樹, "スペクトル制御エンベロープによる混合音中の周期および非周期成分の選択的イコライザ," 情処研報, 2006-MUS-66-13, pp. 77–84 (2006).

[5] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta-divergence," in Proc. MLSP, pp. 283–288 (2010).

[6] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for music transcription," in Proc. WASPAA, pp. 177–180 (2003).

[7] 亀岡弘和, 小野順貴, 柏野邦夫, 嵯峨山茂樹, "複素 NMF: 新しいスパース信号分解表現と基底系学習アルゴリズム," 音講論 (秋), 2-8-13, pp. 657–660 (2008).

[8] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in Proc. ICA 2004, pp. 494–499 (2004).

[9] A. Ozerov, C. Févotte and M. Charbit, "Factorial scaled hidden Markov model for polyphonic audio representation and source separation," in Proc. WASPAA 2009, pp. 121–124 (2009).

[10] 中野允裕, ルルー ジョナトン, 亀岡弘和, 中村友彦, 小野順貴, 嵯峨山茂樹, "スペクトログラムのベイジアンノンパラメトリックモデリングに基づく音楽信号の解析," 情処研報, 2011-MUS-91-6 (2011).

[11] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," IEEE Trans. ASLP, vol. 15, no. 3, pp. 1066–1074 (2007).

[12] S. A. Raczynski, N. Ono and S. Sagayama, "Multipitch analisys with harmonic nonnegative matrix approximation," in Proc. of ISMIR 2007, pp. 381–386 (2007).

[13] T. Virtanen and A. Klapuri, "Analysis of polyphonic audio using source-filter model and non-negative matrix factorization," in Adv. NIPS (2006).

[14] H. Kameoka and K. Kashino, "Composite autoregressive system for sparse source-filter representation of speech," in Proc. ISCAS2009, pp. 2477–2480 (2009).

[15] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino and S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," in Proc. ICASSP, pp. 5365–5368 (2012).

[16] A. T. Cemgil, "Bayesian inference for nonnegative matrix factorization models," Tech. Rep. CUED/F-INFENG/TR.609, University of Cambridge (2008).

[17] M. Hoffman, D. Blei and P. Cook, "Bayesian nonparametric matrix factorization for recorded music," in Proc. ICML, pp.439–446 (2010).

[18] P. Smaragdis, B. Raj and M.V. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in Proc. ICA 2007, pp. 414–421 (2007).

[19] P. Smaragdis and B. Raj, "Example-driven bandwidth expansion," in Proc. WASPAA, pp. 135–138 (2007).

[20] J.-L. Durrieu, G. Richard, B. David and C. Févotte, "Source/filter model for unsupervised main melody extraction from polyphonic audio signals," IEEE Trans. ASLP, vol. 18, no. 3, pp. 564–575 (2010).

[21] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," in Proc. EUSIPCO (2005).

[22] A. Hurmalainen, J. Gemmeke and T. Virtanen, "Non-negative matrix deconvolution in noise robust speech recognition," in Proc. ICASSP, pp. 4588–4591 (2011).

[23] J.-L. Durrieu, J.-P. Thiran, "Sparse non-negative decomposition of speech power spectra for formant tracking," in Proc. ICASSP, pp. 5260–5263 (2011).

[24] 戸上, 川口, "準ブラインド非負行列分解を用いたマルチチャンネル非線形エコーキャンセラ," 音講論 (秋), 3-Q-9, pp. 757–758 (2009).

[25] H. Kameoka, T. Nakatani and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," in Proc. ICASSP, pp. 45–48 (2009).

[26] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," IEEE Trans. ASLP, vol. 18, no. 3, pp. 550–563 (2010).

[27] Y. Kitano, H. Kameoka, Y. Izumi, N. Ono and S. Sagayama, "A sparse component model of source sinals and its application to blind source separation," in Proc. ICASSP, pp. 4122–4125 (2010).

[28] H. Sawada, H. Kameoka, S. Araki and N. Ueda, "New formulations and efficient algorithms for multichannel NMF," in Proc. WASPAA, pp. 153–156 (2011).

[29] H. Sawada, H. Kameoka, S. Araki and N. Ueda, "Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization," in Proc. ICASSP, pp. 261–264 (2012).

[30] 亀岡弘和, "非負値行列因子分解," 計測と制御, vol. 51, no. 9, pp. 835–844 (2012).

[31] 澤田宏, "非負値行列因子分解 NMF の基礎とデータ／信号解析への応用," 信学誌, vol. 95, no. 9 (2012).