非負値行列因子分解の最適化法について

香川高等専門学校 電気情報工学科
北村大地 (kitamura-d@t.kagawa-nct.ac.jp)

1. はじめに

非負値行列因子分解（nonnegative matrix factorization: NMF）[1] とは，1999 年に D. D. Lee と H. S. Seung によって提案された行列の分解アルゴリズムである．行列の分解は，線形方程式を解くための LU 分解や QR 分解，ベクトル空間の概念に基づく固有値分解（eigenvalue decomposition）や特異値分解（singular value decomposition: SVD）等が代表的であるが，NMF がこれらの行列分解手法と大きく異なるのは，その名の通り非負値行列を分解対象としている点である．非負値行列とは，「全ての要素が 0 以上（0 又は正値）」という制約を持つ一般行列である．

世の中で取り扱うべき 2 次元データ（即ち行列）は非負値行列であることが多い．例えば，5 人の顧客が 3 種類の商品の内，何をいくつ購入したかという購買データは，図 1(a) のように 3×5 の非負値行列で表される（-1 回購入した，というデータはあり得ない）し，8 ビット白黒画像は図 1(b) のように 0 から 255 の輝度を持つ非負値行列として表される．音響信号も例外ではなく，短時間フーリエ変換（short-time Fourier transform: STFT）を用いて時間信号（1 次元データ）を時間周波数信号（2 次元データ）に変換することで複素行列（全ての要素が複素数の行列）が得られ，その各要素の振幅値を取った振幅行列を定義すれば，図 1(c) のように非負値行列が現れる．

(a)

	顧客A	顧客B	顧客C	顧客D	顧客E
商品X	1	2	0	0	1
商品Y	1	3	1	2	2
商品Z	0	0	3	5	3

(b)

(c)

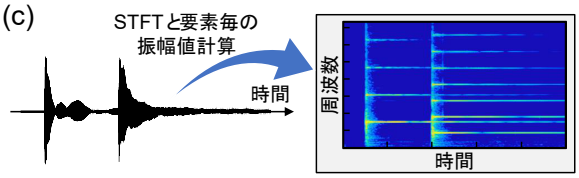

図 1　現実問題に現れる非負値行列：(a) 商品の購買データ，(b) 白黒画像，(c) 音響信号の時間周波数表現の振幅値．

このように，非負値行列は解くべき現実の問題の中に自然な形で現れることが多い．また，その非負値行列が「どのような構造を持っているか」を解析することは，工学的に極めて重要である．例えば，先の購買データの非負値行列（図 1(a)）を解析した結果，「商品 X を購入した顧客は商品 Y を購入しがちである」ということが分かれば，商品 X を購入した顧客に商品 Y を勧めることは効果的である．図 1(a) では，顧客 A，顧客 B，及び顧客 C の 3 人の購買データにこの傾向が確認できるが，例えば 1 万人の購買データを解析して 1000 人が同様の傾向を示していれば，この商品 X と商品 Y に関わる購買傾向はマーケティング戦略上重要である．このような「購買データ行列に隠れている特徴や傾向」は，統計的には「商品購入における相関関係」として解釈され，また代数的には「行列に含まれる潜在的な（行または列の）パターン」として解釈される．一般的に固有値分解や特異値分解で得られるベクトル（固有ベクトル又は特異ベクトル）は，まさに行列に含まれる潜在パターンを基底（basis）として網羅するものである．しかしながら，こと非負値行列においては，固有ベクトルや特異ベクトルを潜在パターンとして解釈しようとした場合，次のような問題が生じる．

今，実数行列 $\boldsymbol{R} \in \mathbb{R}^{I \times J}$ を考える．ここで，I と J はそれぞれ $\boldsymbol{R}$ の行数と列数である．2 次正方行列（$I = J = 2$）を例として，次の行列 $\boldsymbol{R}$ に固有値分解を適用する．

$$\boldsymbol{R} = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix} \tag{1.1}$$

固有値・固有ベクトルの定義 $\boldsymbol{R}\boldsymbol{e} = \lambda \boldsymbol{e}$（$\boldsymbol{e} \neq \boldsymbol{0}$）から固有方程式 $|\boldsymbol{R} - \lambda \boldsymbol{I}| = 0$ を導き，固有値 λ と固有ベクトル $\boldsymbol{e}$ をそれ

ぞれ計算すると，次式のようになる．

$$\text{固有値 } \lambda_1 = 3,\ \lambda_2 = 2 \qquad \text{固有ベクトル } \boldsymbol{e}_1 = C_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix},\ \boldsymbol{e}_2 = C_2 \begin{pmatrix} 1 \\ -2 \end{pmatrix} \tag{1.2}$$

ここで，$\boldsymbol{I}$ は単位行列，C_1 及び C_2 は任意定数である．この結果から，行列 $\boldsymbol{R}$ の固有値分解は次式のようになる．

$$\boldsymbol{R} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^{-1},\ \ \boldsymbol{E} = \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix},\ \ \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \tag{1.3}$$

但し，$C_1 = C_2 = 1$ とした．従って，行列 $\boldsymbol{R}$ に含まれる基底（潜在パターン）は $\boldsymbol{e}_1$ と $\boldsymbol{e}_2$ として得られる．では，行列 $\boldsymbol{R}$ が非負値行列（即ち $\boldsymbol{R} \in \mathbb{R}_{\geq 0}^{I \times J}$，ここで $\mathbb{R}_{\geq 0}$ は非負の実数の集合を表す）であった場合はどうか．次の非負値行列 $\boldsymbol{R}$ を例として考える．

$$\boldsymbol{R} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \tag{1.4}$$

固有値と固有ベクトルはそれぞれ次式で与えられる．

$$\text{固有値 } \lambda_1 = 5,\ \lambda_2 = 1 \qquad \text{固有ベクトル } \boldsymbol{e}_1 = C_1 \begin{pmatrix} 1 \\ 3 \end{pmatrix},\ \boldsymbol{e}_2 = C_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{1.5}$$

非負値行列 $\boldsymbol{R}$ の潜在パターンである固有ベクトルに着目すると，$\boldsymbol{e}_2$ は負値を含んでいることが分かる．これは，先の購買データの例で言えば，「商品 X を 1 個購入した顧客は商品 Y を -1 個購入しがちである」という傾向を，行列が含むパターンとして表していることとなり，途端に自然な解釈ができなくなる（「-1 個購入」を「1 個売却」と考えるのは自然かもしれないが，「売却」という概念を含まない購買データからそのような潜在パターンが得られてしまうのは理解し難い）．まして，行列 $\boldsymbol{R}$ が非正方行列であった場合固有値分解は適用できない．代わりに特異値分解は適用できるが，その結果得られる特異ベクトルは必ず線形独立（直交）である．2 次元座標における 2 本の非負直交基底は $\boldsymbol{e}_1 = C_1(1,0)^{\mathrm{T}}$ と $\boldsymbol{e}_2 = C_2(0,1)^{\mathrm{T}}$（$C_1, C_2 > 0$）しかなく，これは有意な潜在パターンとはいえない．ここで，$^{\mathrm{T}}$ はベクトル又は行列の転置を表す記号である．

以上の議論から，我々が期待する「非負値行列中の基底（潜在パターン）」は，非負値ベクトルであるものが望ましいことが分かる．このような目的を達成するアルゴリズムの代表例が NMF であり，その有用性から登場以降様々な分野で発展・拡張されてきた．非負値行列に対する非負の基底ベクトルの抽出は，固有値分解のように数学的に一意に求まるわけではなく，一般に NP 困難であることが証明されている [2]．しかしながら，基底ベクトルを何らかの非負の初期値で定めたうえで反復計算を繰り返すことで，局所最適解を推定するアルゴリズムが提案されており，そうして得られた非負基底ベクトルは，非負値行列の解析や処理に対してやはり有用であることが分かっている．

本稿では，NMF における最適化問題の定式化を示し，有用な解を得るための最適化アルゴリズムの導出方法をいくつか示す．NMF の最適化アルゴリズムで最も有名なものは，D. D. Lee と H. S. Seung が提案した majorization-minimization（MM）アルゴリズム [3, 4, 5][*1]に基づく乗算型更新式 [7] である．従って，本稿ではまず，2 章で最適化問題における基礎知識と MM アルゴリズムの詳細を解説する．次に，3 章で NMF の最適化手法について導出を示す．最後に，4 章でまとめを示す．

2. 最適化問題における基礎知識と MM アルゴリズム

本章では，NMF の最適化アルゴリズムを理解するうえで必要な数学的基礎知識として，まず最適化問題の定式化を行い，その一般的な解法である最急降下法と Newton 法について説明する．さらに，等式・不等式制約条件付き最適化問題の解法として有名な Lagrange の未定乗数法と Karush–Kuhn–Tucker（KKT）条件について説明し，最後に NMF の最適化において重要な MM アルゴリズムの詳細を述べる．

[*1] 日本語では「MM アルゴリズム」と同義で「補助関数法（auxiliary function technique）」と呼ぶことが多い．これは，文献 [7] の導出において上限関数（majorization function）を補助関数（auxiliary function）と呼んだことに起因している．しかしながら，近年は majorization-equalization（ME）アルゴリズム [6] という最適化手法も提案されていることに鑑み，MM アルゴリズムと ME アルゴリズムの両方を含む意味で補助関数法と呼ぶ傾向にある．

2.1　最適化問題

N 次元実数ベクトル $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_N)^{\mathrm{T}} \in \mathbb{R}^N$ を変数とし[*2]，実数のスカラー値を返すような目的関数 $\mathcal{J}(\boldsymbol{\theta})$ を考える．これを最小化[*3]する変数ベクトル $\boldsymbol{\theta}$ を求める問題は一般に最適化問題（optimization problem）と呼ばれる．本稿では，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ は不連続な点を含まない関数（連続関数）とする．最適化問題を定式化すると，次式のようになる．

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \tag{2.1}$$

ここで，$\min_x f(x)$ は $f(x)$ を x に関して最小化する問題を表す．また，式 (2.1) の解を $\boldsymbol{\theta}^*$ と表記すると，次式のようにも表現される．

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \tag{2.2}$$

ここで，$\arg\min_x f(x)$ は関数 $f(x)$ を最小化する x の集合（argument of the minimum）を返す作用素である．2 変数 $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ を持つ目的関数 $\mathcal{J}(\theta_1, \theta_2)$ の例として，図 2 のような関数を考える．この関数は様々な位置に極値（局所的な最大値や最小値）を持っている．例えば，$(\theta_1, \theta_2) = (3, -3)$ 付近や $(\theta_1, \theta_2) = (-4, 2)$ 付近等に最小値が存在する．このような点は停留点（stationary point）と呼ばれ，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の導関数が 0 になる．しかし，目的関数 $\mathcal{J}(\theta_1, \theta_2)$ の値が最も小さくなるのは（図 2 に見える範囲では）$(\theta_1, \theta_2) = (-2, 2)$ 付近の停留点であることが分かる．この点のように，目的関数値が最も小さくなる停留点を大域最小解（global minimum）と呼び，その他の局所的な停留点を局所最小解（local minimum）と呼ぶ．1 変数における大域最小解と局所最小解の例を図 3 に示す．

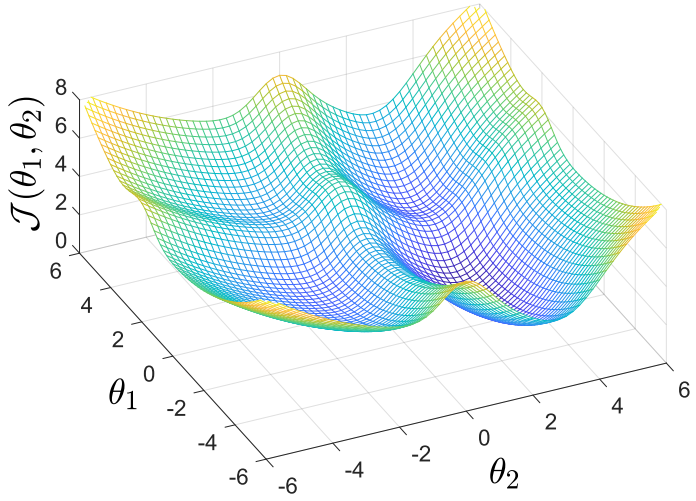

図 2　2 変数 $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ の目的関数 $\mathcal{J}(\theta_1, \theta_2)$ の例．

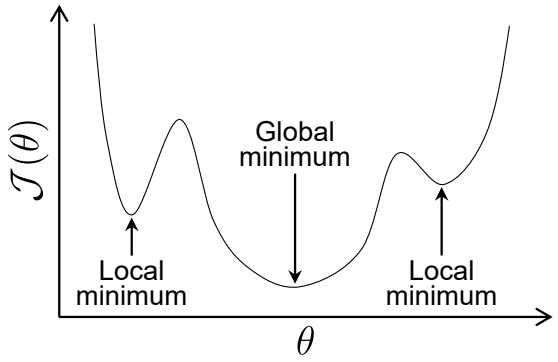

図 3　1 変数 θ の目的関数 $\mathcal{J}(\theta)$ の例．局所最小解と大域最小解の 2 種類が存在する．

2.2　最適性条件

最小化問題の解を求めるためには，解が満たすべき条件，即ち最適性条件（optimality condition）を考える必要がある．最適性条件は，次のように必要条件と十分条件の 2 つに分けることができる．

- 必要条件：点 $\boldsymbol{\theta}^*$ が局所最小解である　⇒　点 $\boldsymbol{\theta}^*$ は最適性の必要条件を満たす
- 十分条件：点 $\boldsymbol{\theta}^*$ が最適性の十分条件を満たす　⇒　点 $\boldsymbol{\theta}^*$ は局所最小解である

[*2] N 次元複素数ベクトルでも構わない．その場合は，実部と虚部を独立な 2 変数と考え，$2N$ 次元実数ベクトルを変数と考えることと等価である．
[*3] 最大化でも構わない．この違いは結局目的関数を -1 倍するか否かであるため，以後の議論に対して本質的ではない．

本来であれば，解の必要十分条件を導くことが望ましい．しかし，最適性の必要十分条件を持つような目的関数はその形が強く制限されるため，ここではより良い十分条件を求めることを考える．

十分条件は必要条件を満たさなければならないため，まずは必要条件を与える．今，$\boldsymbol{\theta}^*$ が局所最小解であるとする．この時，$\boldsymbol{\theta}^*$ から $\boldsymbol{d}$ 方向に長さ a だけ動いた点を $\boldsymbol{\theta}^* + a\boldsymbol{d}$ と定義する．ここで，$\boldsymbol{d} = (d_1, d_2, \cdots, d_N)^{\mathrm{T}}$ は長さ 1 の単位ベクトルであり，方向だけを定義するベクトルと考えてよい．$\boldsymbol{\theta}^*$ は局所最小解であるので，その近傍の関数値 $\mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d})$ は移動長 a が十分小さければ必ず増加するはずである．従って，

$$\begin{aligned} \mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) &\geq \mathcal{J}(\boldsymbol{\theta}^*) \\ \mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) - \mathcal{J}(\boldsymbol{\theta}^*) &\geq 0 \end{aligned} \tag{2.3}$$

が成立する．移動長 a に対する関数値の増分，即ち $\mathcal{J}(\boldsymbol{\theta}^*)$ 付近の $\boldsymbol{d}$ 方向への勾配を考えても同様のことがいえる．

$$\frac{\mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) - \mathcal{J}(\boldsymbol{\theta}^*)}{a} \geq 0 \tag{2.4}$$

a を正の領域から 0 に近づければ，合成関数の微分法を用いて次式が得られる．

$$\begin{aligned} \lim_{a \to +0} \frac{\mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) - \mathcal{J}(\boldsymbol{\theta}^*)}{a} &\geq 0 \\ \left(\frac{d\mathcal{J}(\boldsymbol{\theta}^*)}{d\boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{d} &\geq 0 \end{aligned} \tag{2.5}$$

ここで，スカラーを返す関数 $\mathcal{J}(\boldsymbol{\theta})$ に対するベクトル変数 $\boldsymbol{\theta}$ での微分 $d\mathcal{J}/d\boldsymbol{\theta}$ が登場するが，これは $\mathcal{J}$ の勾配を表し，$\boldsymbol{\nabla}$ 演算子[*4]を用いて $\boldsymbol{\nabla}\mathcal{J}$ と記述される[*5]．即ち，$\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{\theta})$ は各変数 θ_n による $\mathcal{J}(\boldsymbol{\theta})$ の偏微分を $n = 1, \cdots, N$ について並べたベクトルと等価である．従って，式 (2.5) の左辺は次式のように書き改められる．

$$\begin{aligned} \left(\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{\theta}^*) \right)^{\mathrm{T}} \boldsymbol{d} &= \left(\frac{\partial \mathcal{J}(\boldsymbol{\theta}^*)}{\partial \theta_1}, \frac{\partial \mathcal{J}(\boldsymbol{\theta}^*)}{\partial \theta_2}, \cdots, \frac{\partial \mathcal{J}(\boldsymbol{\theta}^*)}{\partial \theta_N} \right) \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{pmatrix} \\ &= \sum_{n=1}^{N} \frac{\partial \mathcal{J}(\boldsymbol{\theta}^*)}{\partial \theta_n} d_n \end{aligned} \tag{2.6}$$

以上より，局所最小解 $\boldsymbol{\theta}^*$ においては次式が成立する．

$$\sum_{n=1}^{N} \frac{\partial \mathcal{J}(\boldsymbol{\theta}^*)}{\partial \theta_n} d_n \geq 0 \tag{2.7}$$

式 (2.7) は，$\boldsymbol{d}$ がいかなる方向でも（あらゆる d_n に対しても）成立しなければならないため，結局最小解 $\boldsymbol{\theta}^*$ では，局所的か大域的に関わらず

$$\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{\theta}) = \frac{d\mathcal{J}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial \mathcal{J}(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \mathcal{J}(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \mathcal{J}(\boldsymbol{\theta})}{\partial \theta_N} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad \frac{\partial \mathcal{J}(\boldsymbol{\theta})}{\partial \theta_n} = 0 \ \ \forall n = 1, 2, \cdots, N \tag{2.8}$$

が満たされなければならない．式 (2.8) は目的関数の 1 階微分 $\boldsymbol{\nabla}\mathcal{J}$ に基づいていることから，1 次最適性条件（first-order optimality condition）と呼ばれる．なお，当たり前であるが，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ は局所最小解 $\boldsymbol{\theta}^*$ 近傍で 1 階微分可能である必要がある．

[*4] $\boldsymbol{\nabla}$ 演算子は各変数の微分演算子を並べた特殊なベクトルである．即ち，$\boldsymbol{\nabla} = (\partial/\partial\theta_1, \partial/\partial\theta_2, \cdots, \partial/\partial\theta_N)$ である．

[*5] 関数 $\mathcal{J}$ に対して $\boldsymbol{\nabla}$ 演算子を作用させた $\boldsymbol{\nabla}\mathcal{J}$ は勾配ベクトルと呼ばれ grad $\mathcal{J}$ とも記述される．勾配ベクトル $\boldsymbol{\nabla}\mathcal{J}$ は，ある点 $\boldsymbol{\theta}$ をいずれかの方向に動かしたとき，$\mathcal{J}$ の値が最も大きくなる方向を向いたベクトルであり，その大きさ $|\boldsymbol{\nabla}\mathcal{J}|$ は勾配がどの程度急かを表す．

1 次最適性条件（局所最小解の必要条件）

$\boldsymbol{\theta}^*$ を目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の最小化問題の局所又は大域最小解とする．また，$\mathcal{J}(\boldsymbol{\theta})$ は $\boldsymbol{\theta}^*$ の近傍で 1 階微分可能とする．このとき，

$$\nabla \mathcal{J}(\boldsymbol{\theta}^*) = \mathbf{0} \tag{2.9}$$

が成り立ち，これを 1 次最適性条件と呼ぶ．ここで，$\mathbf{0}$ は零ベクトルである．

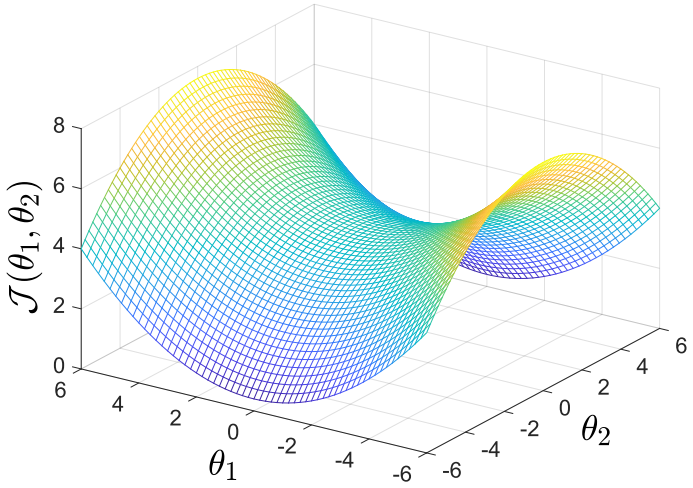

図 4　2 変数 $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ の目的関数 $\mathcal{J}(\theta_1, \theta_2)$ における鞍点の例．

以上の議論より最小解の必要条件が 1 つ与えられた．しかし，これは十分条件ではないことに注意しなければならない．例えば，図 4 における原点のように，θ_1 の最小値と θ_2 の最大値からなるような点も式 (2.9) を満たすが，これは局所最小解でも大域最小解でもない．このような点は鞍点（saddle point）と呼ばれ，極値を取らない停留点である．従って，式 (2.9) が局所最小解の十分条件ではないことが分かる．しかしながら，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が凸関数（convex function）[*6]であれば，凸関数の定義から変数の定義域全域で鞍点が存在しないことがいえるため，1 次最適性条件 (2.9) を満たす解は大域最小解となる．言い換えれば，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が凸関数ならば，1 次最適性条件 (2.9) は大域最小解の十分条件となる．

凸関数の最小解の十分条件

目的関数 $\mathcal{J}(\boldsymbol{\theta})$ を凸関数とする．このとき，1 次最適性条件を満たす点（停留点）は大域最小解である．

それでは，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が凸関数でない場合の解の十分条件はいかにして与えられるのだろうか．再び必要条件から考えてみる．非凸（non-convex）関数ならではの事項として，1 次最適性条件を満たす点（即ち停留点）が「極値を持つ点か鞍点か」ということ，また極値を持つ点であれば「極大値か極小値か」ということを考慮せねばならない．非凸関数の例である図 2 では，見える範囲に局所最小解が 4 点存在するが，そのほかに鞍点が 4 点存在し，さらに原点付近には極大値が 1 点存在する．これらは全て停留点であるため，1 次最適性条件を満たす．このように，非凸関数には最小解でない停留点が存在するため，目的関数の 1 階微分の情報だけを議論する 1 次最適性条件だけでなく，より高階の微分情報から最小解を判別する必要が生じる．

今，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が 2 階微分可能とする．$\mathcal{J}(\boldsymbol{\theta})$ に対して，局所最小解 $\boldsymbol{\theta}^*$ の周りで 2 次 Taylor 展開をすると，次式を得る．

$$\mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) = \mathcal{J}(\boldsymbol{\theta}^*) + a\left(\nabla \mathcal{J}(\boldsymbol{\theta}^*)\right)^{\mathrm{T}} \boldsymbol{d} + \frac{a^2}{2} \boldsymbol{d}^{\mathrm{T}} \left(\nabla^2 \mathcal{J}(\boldsymbol{\theta}^*)\right) \boldsymbol{d} + o(a) \tag{2.10}$$

[*6] 凸関数の定義は，関数上の任意の 2 点を結ぶ線分の内分点が常に関数値の上部にあるような関数である．あるいは，関数のエピグラフ（epigraph）が凸図形である，という定義も同値である．

ここで $o(a)$ は誤差項である．また，$\mathcal{J}$ の 2 階微分 $\boldsymbol{\nabla}^2\mathcal{J}$ が登場するが，これは次式に示すような行列となる[*7]．

$$\boldsymbol{\nabla}^2\mathcal{J} = \begin{pmatrix} \frac{\partial^2\mathcal{J}}{\partial{\theta_1}^2} & \frac{\partial^2\mathcal{J}}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2\mathcal{J}}{\partial\theta_1\partial\theta_N} \\ \frac{\partial^2\mathcal{J}}{\partial\theta_2\partial\theta_1} & \frac{\partial^2\mathcal{J}}{\partial{\theta_2}^2} & \cdots & \frac{\partial^2\mathcal{J}}{\partial\theta_2\partial\theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2\mathcal{J}}{\partial\theta_N\partial\theta_1} & \frac{\partial^2\mathcal{J}}{\partial\theta_N\partial\theta_2} & \cdots & \frac{\partial^2\mathcal{J}}{\partial{\theta_N}^2} \end{pmatrix} \tag{2.11}$$

この行列はヘッセ行列（Hessian matrix）と呼ばれ，2 階偏微分を含む情報をまとめた行列である．目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が Schwarz の定理[*8]を満たす場合，$\partial^2\mathcal{J}/(\partial\theta_i\partial\theta_j) = \partial^2\mathcal{J}/(\partial\theta_j\partial\theta_i)$ が成立するため，ヘッセ行列 (2.11) は対称行列[*9]となる．式 (2.10) を変形すると，次式を得る．

$$\frac{\mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) - \mathcal{J}(\boldsymbol{\theta}^*)}{a^2} = \frac{1}{a}\left(\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{\theta}^*)\right)^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\left(\boldsymbol{\nabla}^2\mathcal{J}(\boldsymbol{\theta}^*)\right)\boldsymbol{d} + \frac{o(a)}{a^2} \tag{2.12}$$

ここで，$\boldsymbol{\theta}^*$ は最小解であり，1 次最適性条件を満たすことから $\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{\theta}^*) = \boldsymbol{0}$ となる．従って式 (2.12) の右辺第一項は 0 である．さらに，a を正の領域から 0 に近づければ，次式が得られる．

$$\lim_{a\to+0}\frac{\mathcal{J}(\boldsymbol{\theta}^* + a\boldsymbol{d}) - \mathcal{J}(\boldsymbol{\theta}^*)}{a^2} = \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\left(\boldsymbol{\nabla}^2\mathcal{J}(\boldsymbol{\theta}^*)\right)\boldsymbol{d} \tag{2.13}$$

以上の議論と式 (2.3) より，最小解 $\boldsymbol{\theta}^*$ に対して次の条件式が得られる．

$$\frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\left(\boldsymbol{\nabla}^2\mathcal{J}(\boldsymbol{\theta}^*)\right)\boldsymbol{d} \geq 0 \tag{2.14}$$

この条件式を解釈するために，下記で説明する行列の定値性（definiteness）を理解しておく必要がある．

ある対称行列 $\boldsymbol{S} \in \mathbb{R}^{N\times N}$ と変数ベクトル $\boldsymbol{x} \in \mathbb{R}^N$ に対して，次の項を考える．

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} &= \begin{pmatrix} x_1 & x_2 & \cdots & x_N \end{pmatrix}\begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1N} \\ s_{21} & s_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N1} & s_{N2} & \cdots & s_{NN} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \\ &= \sum_{n=1}^{N}\sum_{n'=1}^{N} s_{nn'}x_n x_{n'}\end{aligned} \tag{2.15}$$

ただし，$\boldsymbol{S}$ は対称行列なので $s_{nn'} = s_{n'n}$ である．式 (2.15) は行列 $\boldsymbol{S}$ とベクトル $\boldsymbol{x}$ に対する二次形式（quadratic form）と呼ばれ，すべての項の変数が 2 次の形をしている．式 (2.15) の値によって，行列 $\boldsymbol{S}$ の定値性が次のように与えられる．

- いかなる $\boldsymbol{x}$ に対しても $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} > 0$ が成立するとき，$\boldsymbol{S}$ は正定値行列（positive definite matrix）である
- いかなる $\boldsymbol{x}$ に対しても $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} \geq 0$ が成立するとき，$\boldsymbol{S}$ は半正定値行列（positive semi-definite matrix）である
- いかなる $\boldsymbol{x}$ に対しても $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} \leq 0$ が成立するとき，$\boldsymbol{S}$ は半負定値行列（negative semi-definite matrix）である
- いかなる $\boldsymbol{x}$ に対しても $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} < 0$ が成立するとき，$\boldsymbol{S}$ は負定値行列（negative definite matrix）である
- 上記のいずれにも該当しないとき，$\boldsymbol{S}$ は不定値行列（indefinite matrix）である

行列の定値性は固有値の符号と関連している．即ち，次のような性質がある．

- $\boldsymbol{S}$ の固有値がすべて正値のとき，$\boldsymbol{S}$ は正定値行列である
- $\boldsymbol{S}$ の固有値がすべて非負値のとき，$\boldsymbol{S}$ は半正定値行列である
- $\boldsymbol{S}$ の固有値がすべて非正値のとき，$\boldsymbol{S}$ は半負定値行列である
- $\boldsymbol{S}$ の固有値がすべて負値のとき，$\boldsymbol{S}$ は負定値行列である

[*7] 重要な事実として，ベクトルを変数としスカラーを返す関数 $\mathcal{J}$ の 1 階ベクトル微分 $\boldsymbol{\nabla}\mathcal{J}$ はベクトル，2 階ベクトル微分 $\boldsymbol{\nabla}^2\mathcal{J}$ は行列となる．

[*8] $\mathcal{J}$ の全ての変数に対する 1 階偏微分と 2 階偏微分が存在し，2 階偏微分が連続であるとき，2 階偏微分は変数の順番に依らない．Clairaut の定理とも呼ばれる．

[*9] 目的関数が複素ベクトルを変数としスカラーを返す複素関数である場合はエルミート行列になる．

- $\boldsymbol{S}$ の固有値が正負を含むとき，$\boldsymbol{S}$ は不定値行列である

このような行列の定値性は，スカラーの非負性を行列に対して拡張した概念である．実際に $N = 1$ のとき，行列（スカラー）S が正定値であることは $S > 0$，半正定値であることは $S \geq 0$ をそれぞれ意味している．

結局，目的関数 $\mathcal{J}$ の 2 階微分の情報を用いた最小解の必要条件式 (2.14) は，$\boldsymbol{d}$ がいかなる方向でも成立しなければならないため，ヘッセ行列 $\nabla^2\mathcal{J}(\boldsymbol{\theta}^*)$ が半正定値行列であることを示している．これを，2 次最適性条件（second-order optimality condition）と呼ぶ．

2 次最適性条件（局所最小解の必要条件）

$\boldsymbol{\theta}^*$ を目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の最小化問題の局所最小解とする．また，$\mathcal{J}(\boldsymbol{\theta})$ は $\boldsymbol{\theta}^*$ の近傍で 2 階微分可能とする．このとき，ヘッセ行列 $\nabla^2\mathcal{J}(\boldsymbol{\theta}^*)$ は半正定値行列となり，これを 2 次最適性条件と呼ぶ．

以上の内容をまとめると，非凸な目的関数の局所最小解の必要条件は次のように与えられる．

非凸関数の局所最小解の必要条件

$\boldsymbol{\theta}^*$ を目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の最小化問題の局所最小解とする．また，$\mathcal{J}(\boldsymbol{\theta})$ は非凸関数であり，$\boldsymbol{\theta}^*$ の近傍で 2 階微分可能とする．このとき，$\boldsymbol{\theta}^*$ の局所最小解の必要条件は，1 次最適性条件と 2 次最適性条件の両方を満たすことである．

続いて，非凸な目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の局所最小解の十分条件は次のように考えられる．ある点 $\boldsymbol{\theta}^*$ の近傍（一定の範囲内）で $\mathcal{J}(\boldsymbol{\theta})$ が狭義凸関数[*10]であり，$\boldsymbol{\theta}^*$ が凸関数の最小解の十分条件（1 次最適性条件）を満たしていれば，$\boldsymbol{\theta}^*$ が非凸な目的関数 $\mathcal{J}$ の局所的または大局的最小解であることが期待できる．ここで，$\boldsymbol{\theta}^*$ の近傍が凸関数ではなく狭義凸関数でならなければならない理由は，図 5 のような「近傍は凸関数だが局所最小解ではない場合」を除外するためである．狭義凸関数の定義として，$\boldsymbol{\theta}$ のヘッセ行列 $\nabla^2\mathcal{J}(\boldsymbol{\theta})$ はあらゆる点において正定値行列となる．従って，局所最小解 $\boldsymbol{\theta}^*$ においても，ヘッセ行列 $\nabla^2\mathcal{J}(\boldsymbol{\theta}^*)$ は正定値行列である．以上より，非凸関数の局所最小解の十分条件は，1 次最適性条件を満たす点の中で，ヘッセ行列が正定値行列となる点であることがいえる．

非凸関数の局所最小解の十分条件

非凸関数 $\mathcal{J}(\boldsymbol{\theta})$ において，点 $\boldsymbol{\theta}^*$ の関数値 $\mathcal{J}(\boldsymbol{\theta}^*)$ を考える．このとき，$\mathcal{J}(\boldsymbol{\theta})$ は $\boldsymbol{\theta}^*$ の近傍で 2 階微分可能とする．$\mathcal{J}(\boldsymbol{\theta}^*)$ が 1 次最適性条件を満たし，ヘッセ行列 $\nabla^2\mathcal{J}(\boldsymbol{\theta}^*)$ が正定値行列であるとき，$\boldsymbol{\theta}^*$ は局所最小解または大局最小解のいずれかである．

また，証明は割愛するが，上記の十分条件を一般化した性質として次のことが分かる．

- 一次最適性条件を満たす点 $\boldsymbol{\theta}$ に対してヘッセ行列が正定値行列のとき，関数値 $\mathcal{J}(\boldsymbol{\theta})$ は極小値である
- 一次最適性条件を満たす点 $\boldsymbol{\theta}$ に対してヘッセ行列が負定値行列のとき，関数値 $\mathcal{J}(\boldsymbol{\theta})$ は極大値である
- 一次最適性条件を満たす点 $\boldsymbol{\theta}$ に対してヘッセ行列が不定値行列のとき，関数値 $\mathcal{J}(\boldsymbol{\theta})$ は鞍点である

以上より，最適化問題において局所解を求めるための十分条件が与えられた．関数の勾配 $\nabla\mathcal{J}$ を用いた停留点という考え方と，ヘッセ行列 $\nabla^2\mathcal{J}$ の定値性による極値点・鞍点判別が，局所解を議論する上での基盤となる概念であることが分かる．次節では，これらの基盤概念に基づいて局所解を推定するアルゴリズムについて解説する．

2.3 微分可能な目的関数の一般的な解法

最適化問題を解くためには基本的に停留点を求める必要がある．2 次方程式における解の公式のように代数解[*11]として求められる場合は単純であるが，そうでない場合には，変数 $\boldsymbol{\theta}$ を何らかの値で初期化したうえで停留点に徐々に近づいていく最適化アルゴリズムが用いられる．代表的な最適化アルゴリズムに最急降下法（steepest descent）がある．これは，一階微分 $\nabla\mathcal{J}(\boldsymbol{\theta}) = d\mathcal{J}(\boldsymbol{\theta})/d\boldsymbol{\theta}$ を用いて関数の停留点に徐々に近づいていく手法である．いま，l 回反復時の変数を $\boldsymbol{\theta}^{(l)}$ と記述す

[*10] 凸関数でかつ最小点が唯一（1 点のみ）の関数を指す．
[*11] このような解は閉形式解（closed-form solution）と呼ばれる．

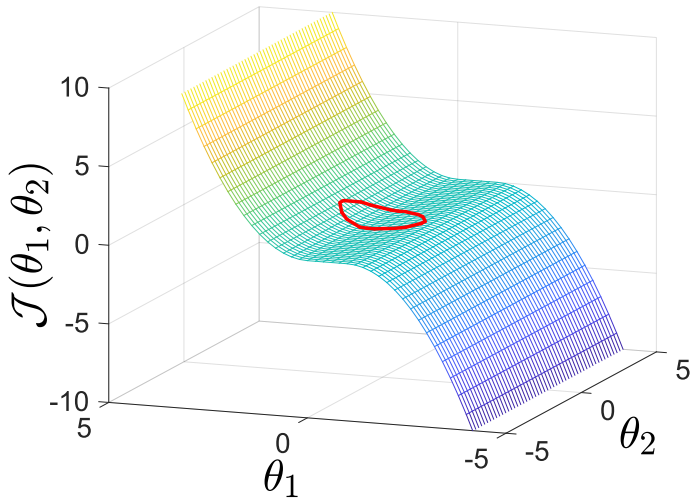

図 5　2 変数の目的関数 $\mathcal{J}(\theta_1, \theta_2)$ において，近傍では凸関数になるが局所最小解ではない例．赤線の範囲では $\mathcal{J}$ は凸関数であるが，近傍内での最小点 $\mathcal{J} = 0$ を複数点含むため狭義凸関数ではない．

るとき，最急降下法による変数の更新は次式で与えられる．

$$\begin{aligned} \boldsymbol{\theta}^{(l+1)} &= \boldsymbol{\theta}^{(l)} - \mu \boldsymbol{\nabla} \mathcal{J}(\boldsymbol{\theta}^{(l)}) \\ &= \boldsymbol{\theta}^{(l)} - \mu \begin{pmatrix} \frac{\partial \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_1} \\ \frac{\partial \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_N} \end{pmatrix} \end{aligned} \tag{2.16}$$

ここで，μ はステップサイズと呼ばれるパラメータであり，1 回の反復で $\boldsymbol{\theta}$ をどの程度移動させるかを調整する値である．最急降下法による変数更新の様子を図 6 に示す．この図から分かるように，最急降下法では適当に与えた初期値 $\boldsymbol{\theta}^{(0)}$ を起点として，目的関数の勾配を下る方向（$-\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{\theta})$）に変数 $\boldsymbol{\theta}$ を更新してゆく．但し，ステップサイズ μ が大きすぎると発散して解が求まらない場合があるため，アルゴリズムとしての変数の収束性は保証されない．また，停留点を求めるまでに多数の反復が必要（最適化が遅い）という点の他，先に述べた鞍点に捕まってしまう問題もある[*12]．

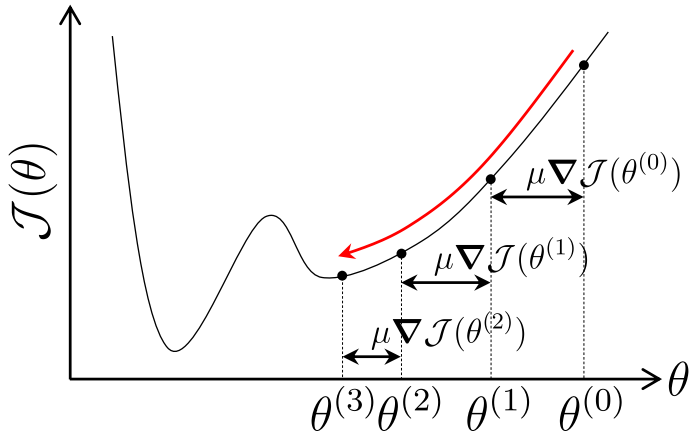

図 6　1 変数の目的関数 $\mathcal{J}(\theta)$ に最急降下法を適用したときの変数 θ の更新の様子．初期値 $\theta^{(0)}$ 近傍の局所最小解に向かって最適化が進む．

最急降下法よりも少ない反復回数で局所最小解に更新される最適化アルゴリズムとして Newton 法（Newton's method）

[*12] 但し，ステップサイズ μ を極限まで小さくすれば，最急降下法でも無限回の反復により鞍点に捕まる確率は 0 に収束することが証明されている [8].

がある．これは，1 階微分（勾配ベクトル）だけでなく，2 回微分（ヘッセ行列）の情報も用いた手法であり，変数の更新式は次式で与えられる．

$$\begin{aligned}\boldsymbol{\theta}^{(l+1)} &= \boldsymbol{\theta}^{(l)} - \mu \left(\boldsymbol{\nabla}^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})\right)^{-1} \boldsymbol{\nabla} \mathcal{J}(\boldsymbol{\theta}^{(l)}) \\ &= \boldsymbol{\theta}^{(l)} - \mu \begin{pmatrix} \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial {\theta_1}^2} & \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_1 \partial \theta_N} \\ \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial {\theta_2}^2} & \cdots & \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_2 \partial \theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_N \partial \theta_1} & \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_N \partial \theta_2} & \cdots & \frac{\partial^2 \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial {\theta_N}^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_1} \\ \frac{\partial \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \mathcal{J}(\boldsymbol{\theta}^{(l)})}{\partial \theta_N} \end{pmatrix} \end{aligned} \tag{2.17}$$

式 (2.17) と式 (2.16) を比較すると，最急降下法はヘッセ行列 $\boldsymbol{\nabla}^2 \mathcal{J}$ を単位行列 $\boldsymbol{I}$ とおいた場合に等しいことが分かる．Newton 法は多くの場合で最急降下法よりも高速であるが，毎回の反復においてヘッセ行列の逆行列を求めなければならず計算量の増加や数値安定性の問題が新たに生じる．また，ヘッセ行列が正則でない場合には何らかの正則化処理が必要となり，ヘッセ行列の対角成分に微小値を足す準 Newton 法や BFGS 法[*13]等を用いる必要がある．

2.4 制約条件付き最適化問題とその解法

最適化問題には，式 (2.1) において変数ベクトル $\boldsymbol{\theta}$ に対する何らかの制約が課せられる場合がある．例えば，変数の総和（変数ベクトル $\boldsymbol{\theta}$ の長さ）を制限する制約（$\sum_{n=1}^{N} \theta_n = 1$ 等）は等式の制約条件であり，変数の非負性を担保する制約（$\theta_n \geq 0 \ \ \forall n = 1, 2, \cdots, N$ 等）は不等式の制約条件である．このような制約条件が付く最適化問題を，明示的に制約条件付き最適化問題と呼ぶ．まずは，次式に示す等式制約条件付き最適化問題について取り上げる．

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \ \ \text{s.t. } f(\boldsymbol{\theta}) = 0 \tag{2.18}$$

ここで，等式制約条件を $f(\boldsymbol{\theta}) = 0$ という形で表現したが，これは一般的にあらゆる条件を含む．例えば，先の総和制約の例 $\sum_{n=1}^{N} \theta_n = 1$ を考えると，$f(\boldsymbol{\theta}) = \sum_{n=1}^{N} \theta_n - 1$ とおけば，これが 0 になる等式制約条件として式 (2.18) で表現される．以後，直感的理解を促すために 2 変数の目的関数で示す．即ち，式 (2.18) は次式となる．

$$\min_{\theta_1, \theta_2} \mathcal{J}(\theta_1, \theta_2) \ \ \text{s.t. } f(\theta_1, \theta_2) = 0 \tag{2.19}$$

ここで，目的関数 $\mathcal{J}(\theta_1, \theta_2)$ は 3 次元空間上の曲面であり，制約条件 $f(\theta_1, \theta_2) = 0$ は 2 次元平面上の曲線になる．例として，$\mathcal{J}(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2$ という目的関数と，$f(\theta_1, \theta_2) = 8\theta_1 - \theta_2^2 + 16$ という制約条件を考える．

$$\min_{\theta_1, \theta_2} \theta_1^2 + \theta_2^2 \ \ \text{s.t. } 8\theta_1 - \theta_2^2 + 16 = 0 \tag{2.20}$$

上式を図示したものが図 7 である．目的関数の曲面に対して，制約条件を満たす領域は $\theta_1 \theta_2$ 平面上の曲線として定義されている．この制約条件の曲線上で目的関数値が最も小さくなる点が等式制約条件付き最適化問題 (2.20) の解である．目的関数が $\mathcal{J}(\boldsymbol{\theta}) = \theta_1^2 + \theta_2^2 = |\boldsymbol{\theta}|^2$ であることに注目すれば，図 7 の赤い曲線上で変数ベクトル $\boldsymbol{\theta}$ の長さが最も短くなる（原点からの距離が最も近くなる）点が解となることが分かる．確かに，図 7 に示した解は，曲線上で変数ベクトル $\boldsymbol{\theta}$ が最小となる点と一致している．

では，図 7 の解はどのようにして求まるのだろうか．ここでは直感的な説明を後回しにし，等式制約条件付き最適化問題 (2.19) から次の関数を新たに考える．

$$\mathcal{L}(\theta_1, \theta_2, \lambda) = \mathcal{J}(\theta_1, \theta_2) - \lambda f(\theta_1, \theta_2) \tag{2.21}$$

ここで，λ は新たに追加したスカラーの変数である．式 $\mathcal{L}(\theta_1, \theta_2, \lambda)$ は Lagrange 関数（Lagrange function）と呼ばれ，元の最適化問題の目的関数 $\mathcal{J}(\theta_1, \theta_2)$，等式制約条件 $f(\theta_1, \theta_2)$，及び新しい変数 λ から構成されている．変数 λ は Lagrange 乗数（Lagrange multiplier）と呼ばれる．この Lagrange 関数を用いると，最適化問題の解は次の式を必ず満たす．

$$\frac{\partial \mathcal{L}}{\partial \theta_1} = \frac{\partial \mathcal{L}}{\partial \theta_2} = \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \tag{2.22}$$

[*13] BFGS 法の BFGS はそれぞれ Broyden，Fletcher，Goldfarb，Shanno の 4 名の頭文字に由来している．

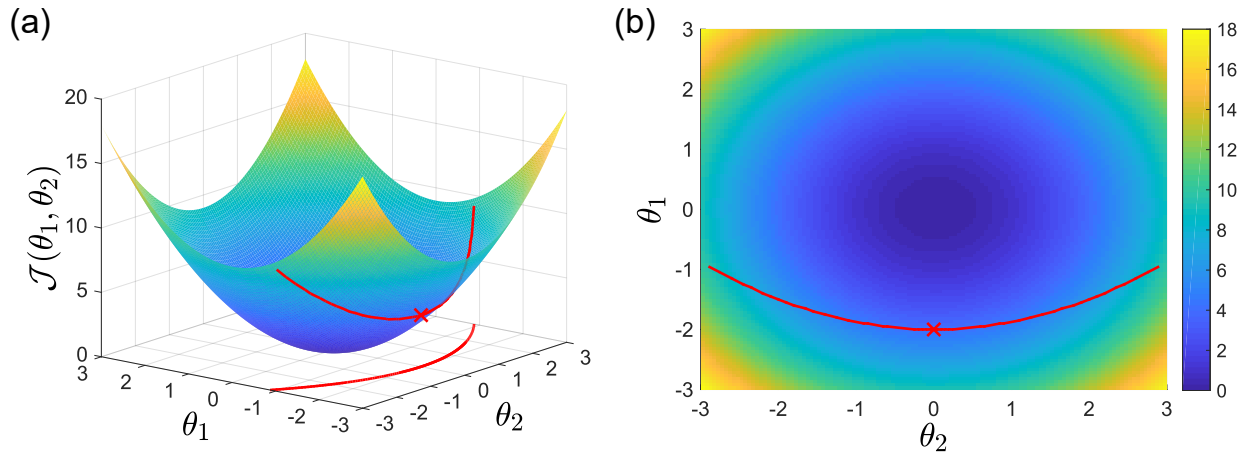

図 7　目的関数 $\mathcal{J}(\theta_1,\theta_2)=\theta_1^2+\theta_2^2$（曲面）と制約条件 $f(\theta_1,\theta_2)=8\theta_1-\theta_2^2+16$（$\theta_1\theta_2$ 平面上の赤い曲線）: (a) 3 次元図，(b) 2 次元図．制約条件を満たす変数範囲で目的関数値が最小となる点（この最適化問題の解）をクロスマークで示している．

つまり，Lagrange 関数を各変数（θ_1 及び θ_2）や Lagrange 乗数（λ）で偏微分した結果が全て 0 である，という式である．

例として，式 (2.20) の問題について計算してみると，Lagrange 関数は

$$\mathcal{L}(\theta_1,\theta_2,\lambda)=\theta_1^2+\theta_2^2-\lambda\left(8\theta_1-\theta_2^2+16\right) \tag{2.23}$$

であり，この偏微分と解の条件式 (2.22) は次のようになる．

$$\frac{\partial\mathcal{L}}{\partial\theta_1}=2\theta_1-8\lambda=0 \tag{2.24}$$

$$\frac{\partial\mathcal{L}}{\partial\theta_2}=2\theta_2+2\lambda\theta_2=0 \tag{2.25}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda}=-8\theta_1+\theta_2^2-16=0 \tag{2.26}$$

次に，式 (2.24)–(2.26) を満たす θ_1，θ_2，及び λ を求める．3 個の変数に対して独立な式が 3 個あるので，これを解くことができる．式 (2.24) 及び (2.25) より，次式が得られる．

$$\theta_1=4\lambda \tag{2.27}$$

$$\theta_2(2+2\lambda)=0 \quad\Leftrightarrow\quad \theta_2=0 \text{ or } \lambda=-1 \tag{2.28}$$

式 (2.28) において，$\lambda=-1$ の場合を考えると $\theta_1=-4$ かつ $\theta_2=0$ となる．しかし，この θ_1 と θ_2 の組み合わせは式 (2.26) を満たさないため，最小化問題の解ではない．一方，式 (2.28) において，$\theta_2=0$ の場合を考えると，式 (2.27) と $\theta_2=0$ を式 (2.26) に代入して次式が得られる．

$$-32\lambda-16=0$$

$$\lambda=-\frac{1}{2} \tag{2.29}$$

従って，式 (2.24)–(2.26) を全て満たす解は $\theta_1=-2$，$\theta_2=0$，及び $\lambda=-1/2$ として得られる．図 7 を見ると，確かに $(\theta_1,\theta_2)=(-2,0)$ の位置に解が存在していることが分かる．

さて，この Lagrange の未定乗数法がなぜ等式制約条件付き最適化問題の解を与えるのか，その理由について直感的な説明を示す．2 変数の最適化問題 (2.19) において，解の条件式 (2.22) について考える．まず，条件式 (2.22) 中の

$$\frac{\partial\mathcal{L}(\theta_1,\theta_2,\lambda)}{\partial\lambda}=0 \tag{2.30}$$

に着目すると，これは，制約条件の式 $f(\theta_1,\theta_2)$ を $-\lambda$ 倍した項（$-\lambda f(\theta_1,\theta_2)$）の λ による偏微分が 0 という条件であるため，等式制約条件 $f(\theta_1,\theta_2)=0$ を別の形で難しく表現したものにすぎない．事実として，式 (2.26) と最適化問題 (2.20) の

等式制約条件を比較すると，両者は -1 倍の違いを除いて一致している．次に，

$$\frac{\partial \mathcal{L}(\theta_1, \theta_2, \lambda)}{\partial \theta_1} = \frac{\partial \mathcal{L}(\theta_1, \theta_2, \lambda)}{\partial \theta_2} = 0 \tag{2.31}$$

に着目する．Lagrange 関数 $\mathcal{L}$ を展開して式 (2.31) を変形すると，次のようになる．

$$\frac{\partial \mathcal{L}(\theta_1, \theta_2, \lambda)}{\partial \theta_1} = \frac{\partial \mathcal{J}(\theta_1, \theta_2)}{\partial \theta_1} - \lambda \frac{\partial f(\theta_1, \theta_2)}{\partial \theta_1} = 0 \tag{2.32}$$

$$\frac{\partial \mathcal{L}(\theta_1, \theta_2, \lambda)}{\partial \theta_2} = \frac{\partial \mathcal{J}(\theta_1, \theta_2)}{\partial \theta_2} - \lambda \frac{\partial f(\theta_1, \theta_2)}{\partial \theta_2} = 0 \tag{2.33}$$

さらに λ の掛かる項を右辺に持っていけば

$$\frac{\partial \mathcal{J}(\theta_1, \theta_2)}{\partial \theta_1} = \lambda \frac{\partial f(\theta_1, \theta_2)}{\partial \theta_1} \tag{2.34}$$

$$\frac{\partial \mathcal{J}(\theta_1, \theta_2)}{\partial \theta_2} = \lambda \frac{\partial f(\theta_1, \theta_2)}{\partial \theta_2} \tag{2.35}$$

となる．これをベクトル形式でまとめて表現すると，

$$\begin{pmatrix} \frac{\partial \mathcal{J}(\theta_1, \theta_2)}{\partial \theta_1} \\ \frac{\partial \mathcal{J}(\theta_1, \theta_2)}{\partial \theta_2} \end{pmatrix} = \lambda \begin{pmatrix} \frac{\partial f(\theta_1, \theta_2)}{\partial \theta_1} \\ \frac{\partial f(\theta_1, \theta_2)}{\partial \theta_2} \end{pmatrix} \tag{2.36}$$

$$\boldsymbol{\nabla} \mathcal{J} = \lambda \boldsymbol{\nabla} f \tag{2.37}$$

という式に帰着する．式 (2.37) は，$\mathcal{J}$ の勾配ベクトル $\boldsymbol{\nabla}\mathcal{J}$ と制約条件式の勾配ベクトル $\boldsymbol{\nabla} f$ が平行（スカラー λ 倍）であることを表している．この勾配ベクトルの平行条件を最適化問題 (2.20) の例について 8 に示している．勾配ベクトル $\boldsymbol{\nabla}\mathcal{J}$ は目的関数 $\boldsymbol{\nabla}\mathcal{J}$ の等高線（$\mathcal{J}(\theta_1, \theta_2) = C$，ここで C は定数）に対して垂直なベクトルであり，また勾配ベクトル $\boldsymbol{\nabla} f$ も曲線 $f(\theta_1, \theta_2) = 0$ に対して垂直なベクトルである．これらが互いに平行になる点が最小化問題の解の候補であり，図 8 における $\theta_2 = 0$ の領域（直線上）に対応する．先の計算の例で，式 (2.24) 及び (2.25) より $\theta_2 = 0$ が先に導かれたのはこのためである．従って，$\boldsymbol{\nabla}\mathcal{J} = \lambda \boldsymbol{\nabla} f$（目的関数と制約条件の勾配ベクトルが平行になる点，即ち $\theta_2 = 0$ を満たす点）かつ $f(\theta_1, \theta_2) = 0$（制約条件の曲線上の点）は，結局 $\theta_2 = 0$ の直線と制約条件の曲線の交点を求める問題となり，変数 λ を媒介して θ_1 の解が得られる．なぜ式 (2.37) を満たす点が解の候補になるのか，という問については，次のように解釈できる．2 本の勾配ベクトル $\boldsymbol{\nabla}\mathcal{J}$ 及び $\boldsymbol{\nabla} f$ が互いに平行になる点は，即ち目的関数の等高線 $\mathcal{J}(\theta_1, \theta_2) = C$ と制約条件の曲線 $f(\theta_1, \theta_2) = 0$ が（交差ではなく）接する点である．もし目的関数の等高線と制約条件の曲線が交差しているならば，その点は「制約条件の曲線上で目的関数が最小となる点」ではないことが，図 8 より直感的に理解できよう[*14]．

以上より，2 変数の等式制約条件付き最小化問題は，確かに Lagrange 関数 $\mathcal{L}(\theta_1, \theta_2, \lambda)$ の偏微分条件式 (2.24)–(2.26) で求められることが分かった．一般の多変数の場合も，Lagrange 関数と偏微分条件式を多変数に自然に拡張することで等式制約条件付き最小化問題を解くことができる．さらに，満たすべき等式制約条件が $f_1(\boldsymbol{\theta}) = 0$，$f_2(\boldsymbol{\theta}) = 0$，$\cdots$，$f_M(\boldsymbol{\theta}) = 0$ として M 個ある場合でも，Lagrange 乗数とその項を自然に増やすことで，やはり最適化問題を解くことができる．

[*14] 但し，$\mathcal{J}$ の等高線が曲線 $f = 0$ に接することは $\mathcal{J}$ が極値をとる必要条件であるが，十分条件ではない．例えば，等高線が曲線 $f = 0$ の一方から接するように交わって反対側から離れていくなら，$\mathcal{J}$ はその点では極値をとらない．

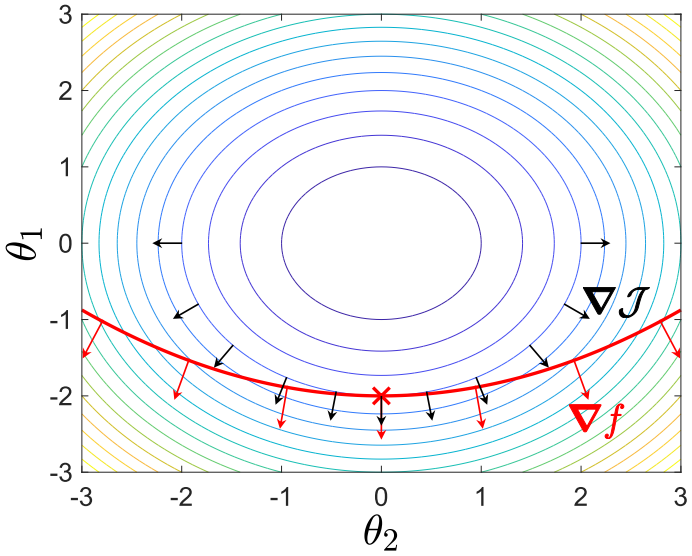

図 8　目的関数 $\mathcal{J}(\theta_1,\theta_2)=\theta_1^2+\theta_2^2$（曲面の等高線）と制約条件 $f(\theta_1,\theta_2)=8\theta_1-\theta_2^2+16$（赤い曲線）の 2 次元図．目的関数の等高線に対する勾配ベクトル $\nabla\mathcal{J}$ を黒い矢印，制約条件の曲線に対する勾配ベクトル ∇f を赤い矢印で示している．最適化問題の解であるクロスマークの位置では，$\nabla\mathcal{J}$ と ∇f が平行となる．

等式制約条件付き最適化問題の解の 1 次必要条件

N 次元ベクトル変数 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_N)^{\mathrm{T}}$ に対する目的関数 $\mathcal{J}(\boldsymbol{\theta})$ 及び等式制約条件 $f_1(\boldsymbol{\theta})=0$，$f_2(\boldsymbol{\theta})=0$，$\cdots$，$f_M(\boldsymbol{\theta})=0$ において，制約条件を全て満たすような $\mathcal{J}(\boldsymbol{\theta})$ の最小解 $\boldsymbol{\theta}^*$ を求める問題を考える．

$$\min_{\boldsymbol{\theta}}\mathcal{J}(\boldsymbol{\theta})\ \ \text{s.t. } f_m(\boldsymbol{\theta})=0\ \ \forall m=1,2,\cdots,M \tag{2.38}$$

このとき，最小解 $\boldsymbol{\theta}^*$ は，次の式を全て満たす．

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta}^*,\lambda_1,\lambda_2,\cdots,\lambda_M)}{\partial\theta_n}=0\ \ \forall n=1,2,\cdots,N \tag{2.39}$$

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta}^*,\lambda_1,\lambda_2,\cdots,\lambda_M)}{\partial\lambda_m}=0\ \ \forall m=1,2,\cdots,M \tag{2.40}$$

ここで，$\mathcal{L}(\boldsymbol{\theta},\lambda_1,\lambda_2,\cdots,\lambda_M)$ は次式で与えられる Lagrange 関数である．

$$\mathcal{L}(\boldsymbol{\theta},\lambda_1,\lambda_2,\cdots,\lambda_M)=\mathcal{J}(\boldsymbol{\theta})-\sum_{m=1}^{M}\lambda_m f_m(\boldsymbol{\theta}) \tag{2.41}$$

なお，式 (2.39) は $\nabla\mathcal{L}=\mathbf{0}$ と書くこともできる．このような等式制約条件付き最適化問題の解き方は Lagrange の未定乗数法（method of Lagrange multiplier）と呼ばれる．

以上が Lagrange の未定乗数法の本質であり，多変数となっても $\nabla\mathcal{J}=\lambda\nabla f$ かつ $f(\boldsymbol{\theta})=0$ を満たす解が等式制約条件付き最適化問題の解であるという性質は変わらない．また，制約条件が複数与えられた場合も同様に $\nabla\mathcal{J}=\sum_{m=1}^{M}\lambda_m\nabla f_m$ かつ全ての $m=1,2,\cdots,M$ について $f_m(\boldsymbol{\theta})=0$ を満たす点が解である．結局 Lagrange の未定乗数法とは，「等式制約条件付き最小化問題」を「制約条件無し停留点探索問題」に変換する手法と捉えることができる．注意すべき点は，Lagrange の未定乗数法は 1 次の必要条件（1 次最適性条件）しか用いていないことである．2.2 節で述べた通り，目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が凸関数ではない場合，1 次最適性条件で求まる解は極小値，極大値，鞍点の何れかであるため，どれに該当するのかをさらに調べる必要がある．具体的な方法としては，Lagrange 関数 $\mathcal{L}(\boldsymbol{\theta},\lambda_1,\lambda_2,\cdots,\lambda_M)$ の（未定乗数も含めた）ヘッセ行列[*15]の定値性により判定できる．

次に，制約条件が不等式として与えられた場合，即ち不等式制約条件付き最小化問題について考える．この場合も，等式

[*15] 縁付きヘッセ行列（bordered Hessian matrix）と呼ばれる．

制約条件の場合と同様に Lagrange の未定乗数法と同じ概念に基づいて解くことができる．一般に不等式制約条件付き最小化問題は

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \quad \text{s.t. } f(\boldsymbol{\theta}) \leq 0 \tag{2.42}$$

で与えられる．もし制約条件 $f(\boldsymbol{\theta}) \leq 0$ を満たす領域の内部に目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の局所最小値が存在する場合は，式 (2.42) の問題は制約条件の無い最小化問題（式 (2.1)）と等価である．これは即ち，Lagrange 関数 $\mathcal{L} = \mathcal{J}(\boldsymbol{\theta}) - \lambda f(\boldsymbol{\theta})$ において Lagrange 乗数が $\lambda = 0$ であることに対応している．逆に，制約条件を満たす領域内に目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の局所最小値が存在しない場合は，不等式制約条件付き最適化問題としての解は必ず制約条件を満たす領域の境界上に存在することがいえる．このような場合の例として，$\mathcal{J}(\theta_1, \theta_2) = -\theta_1 + \theta_2 + 5$ という目的関数と，$f(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2 - 1$ という制約条件を考える．

$$\min_{\theta_1, \theta_2} -\theta_1 + \theta_2 + 5 \quad \text{s.t. } \theta_1^2 + \theta_2^2 - 1 \leq 0 \tag{2.43}$$

この不等式制約条件は $\theta_1\theta_2$ 平面上の単位円の内部に相当する．図 9 は，式 (2.43) を図で示したものである．図から分かる通り，本来の目的関数 $\mathcal{J}(\theta_1, \theta_2)$ の最小値は $-\infty$ となる為，解は存在しない．しかしながら，不等式制約条件によって解の範囲が限定されることで，最小解が領域の境界に存在することになる．このように解が領域の境界上に存在する場合，不等式制約条件は結局等式 $f(\theta_1, \theta_2) = 0$ として満たされていることに相当するため，Lagrange の未定乗数法を適用した解と一致する．従って，$\boldsymbol{\nabla}\mathcal{J} = \lambda\boldsymbol{\nabla} f$ となる $\lambda < 0$ が存在することになる[*16]．また，そのときの解は領域の境界に存在するため，$f(\boldsymbol{\theta}) = 0$ を満たす．

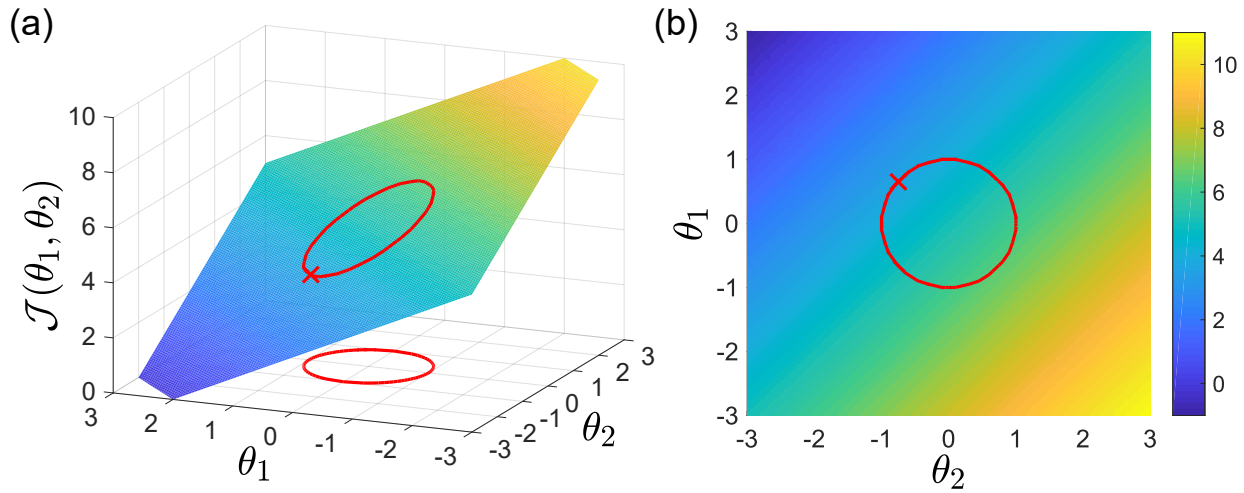

図 9　目的関数 $\mathcal{J}(\theta_1, \theta_2) = -\theta_1 + \theta_2 + 5$（平面）と不等式制約条件の境界 $f(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2 - 1$（$\theta_1\theta_2$ 平面上の赤い曲線）: (a) 3 次元図，(b) 2 次元図．不等式制約条件 $f(\theta_1, \theta_2) \leq 0$ は赤い曲線の内側に対応する．この不等式制約条件を満たす領域で目的関数値が最小となる点（この最適化問題の解）をクロスマークで示している．

以上をまとめると，次のようになる．

- 最小解が不等式制約条件の領域の内部にある場合，解は $\lambda = 0$ を満たす必要がある
- 最小解が不等式制約条件の領域の境界にある場合，解は $\lambda < 0$ かつ $f(\boldsymbol{\theta}) = 0$ を満たす必要がある

これらの条件より，λ か $f(\boldsymbol{\theta}) = 0$ の何れかが 0 なので $\lambda f(\boldsymbol{\theta}) = 0$ が得られ，不等式制約条件付き最小化問題の必要条件は

$$\lambda f(\boldsymbol{\theta}) = 0 \tag{2.44}$$
$$\lambda \leq 0 \tag{2.45}$$
$$f(\boldsymbol{\theta}) \leq 0 \tag{2.46}$$

[*16] 不等式制約条件付き最小化問題における Lagrange 乗数は，最大化問題では $\lambda > 0$，最小化問題では $\lambda < 0$ がそれぞれ存在することが保証されている．

となる．この必要条件は Karush–Kuhn–Tucker（KKT）条件と呼ばれる．従って，解は KKT 条件の下で，$L(\boldsymbol{\theta}, \lambda) = \mathcal{J}(\boldsymbol{\theta}) - \lambda f(\boldsymbol{\theta})$ に対する偏微分の条件式 (2.22) を満たす解を求めればよい．Lagrange の未定乗数法のときと同様に，複数の不等式制約条件の場合も，KKT 条件を自然な形で拡張できる．

不等式制約条件付き最小化問題の解の 1 次必要条件

N 次元ベクトル変数 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_N)^{\mathrm{T}}$ に対する目的関数 $\mathcal{J}(\boldsymbol{\theta})$ 及び不等式制約条件 $f_1(\boldsymbol{\theta}) \leq 0$，$f_2(\boldsymbol{\theta}) \leq 0$，$\cdots$，$f_M(\boldsymbol{\theta}) \leq 0$ において，制約条件を全て満たすような $\mathcal{J}(\boldsymbol{\theta})$ の最小解 $\boldsymbol{\theta}^*$ を求める問題を考える．

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \ \text{ s.t. } f_m(\boldsymbol{\theta}) \leq 0 \ \ \forall m = 1, 2, \cdots, M \tag{2.47}$$

このとき，最小解 $\boldsymbol{\theta}^*$ は，次の式を全て満たす．

$$\lambda_m f_m(\boldsymbol{\theta}^*) = 0 \ \ \forall m = 1, 2, \cdots, M \tag{2.48}$$
$$\lambda_m \leq 0 \ \ \forall m = 1, 2, \cdots, M \tag{2.49}$$
$$f_m(\boldsymbol{\theta}^*) \leq 0 \ \ \forall m = 1, 2, \cdots, M \tag{2.50}$$
$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}^*, \lambda_1, \lambda_2, \cdots, \lambda_M)}{\partial \theta_n} = 0 \ \ \forall n = 1, 2, \cdots, N \tag{2.51}$$

ここで，$\mathcal{L}(\boldsymbol{\theta}^*, \lambda_1, \lambda_2, \cdots, \lambda_M)$ は次式で与えられる Lagrange 関数である．

$$\mathcal{L}(\boldsymbol{\theta}, \lambda_1, \lambda_2, \cdots, \lambda_M) = \mathcal{J}(\boldsymbol{\theta}) - \sum_{m=1}^{M} \lambda_m f_m(\boldsymbol{\theta}) \tag{2.52}$$

また，等式制約条件と不等式制約条件が複数存在する最適化問題も，これまでの理論を応用することができる．

等式・不等式制約条件付き最小化問題の解の 1 次必要条件

N 次元ベクトル変数 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_N)^{\mathrm{T}}$ に対する目的関数 $\mathcal{J}(\boldsymbol{\theta})$ 及び等式制約条件 $f_1(\boldsymbol{\theta}) = 0$，$f_2(\boldsymbol{\theta}) = 0$，$\cdots$，$f_M(\boldsymbol{\theta}) = 0$ 及び不等式制約条件 $g_1(\boldsymbol{\theta}) \leq 0$，$g_2(\boldsymbol{\theta}) \leq 0$，$\cdots$，$g_L(\boldsymbol{\theta}) \leq 0$ において，制約条件を全て満たすような $\mathcal{J}(\boldsymbol{\theta})$ の最小解 $\boldsymbol{\theta}^*$ を求める問題を考える．

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \ \text{ s.t. } f_m(\boldsymbol{\theta}) = 0, g_l(\boldsymbol{\theta}) \leq 0 \ \ \forall m = 1, 2, \cdots, M, l = 1, 2, \cdots, L \tag{2.53}$$

このとき，最小解 $\boldsymbol{\theta}^*$ は，次の式を全て満たす．

$$\mu_l g_l(\boldsymbol{\theta}^*) = 0 \ \ \forall l = 1, 2, \cdots, L \tag{2.54}$$
$$\mu_l \leq 0 \ \ \forall l = 1, 2, \cdots, L \tag{2.55}$$
$$g_l(\boldsymbol{\theta}^*) \leq 0 \ \ \forall l = 1, 2, \cdots, L \tag{2.56}$$
$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}^*, \lambda_1, \lambda_2, \cdots, \lambda_M, \mu_1, \mu_2, \cdots, \mu_L)}{\partial \theta_n} = 0 \ \ \forall n = 1, 2, \cdots, N \tag{2.57}$$
$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}^*, \lambda_1, \lambda_2, \cdots, \lambda_M, \mu_1, \mu_2, \cdots, \mu_L)}{\partial \lambda_m} = 0 \ \ \forall m = 1, 2, \cdots, M \tag{2.58}$$

ここで，$\mathcal{L}(\boldsymbol{\theta}^*, \lambda_1, \lambda_2, \cdots, \lambda_M, \mu_1, \mu_2, \cdots, \mu_L)$ は次式で与えられる Lagrange 関数である．

$$\mathcal{L}(\boldsymbol{\theta}^*, \lambda_1, \lambda_2, \cdots, \lambda_M, \mu_1, \mu_2, \cdots, \mu_L) = \mathcal{J}(\boldsymbol{\theta}) - \sum_{m=1}^{M} \lambda_m f_m(\boldsymbol{\theta}) - \sum_{l=1}^{L} \mu_l g_l(\boldsymbol{\theta}) \tag{2.59}$$

2.5 MM アルゴリズム

これまでに述べた最適化アルゴリズムは全て目的関数 $\mathcal{J}(\boldsymbol{\theta})$ が $\boldsymbol{\theta}$ に関して微分可能であることを仮定している．もし目的関数が滑らかでない箇所を含むような，微分不可能な関数であった場合は，前述の最急降下法や Newton 法が適用できず，停留点を求めるのはより難しくなる．このような場合においても，凸最適化の分野では様々な最適化アルゴリズムが考案さ

れている [9]．本節ではとりわけ NMF に応用される MM アルゴリズムについて解説する．ここでは，先に簡単な例題を通して MM アルゴリズムの概要を理解した後に，MM アルゴリズムの一般的な概念について説明する．

今，次のような 1 変数の目的関数を考える．

$$\mathcal{J}(\theta) = \frac{1}{2}|\theta - 6| + |\theta - 1| + |\theta + 5| - 4 \tag{2.60}$$

図 10 は式 (2.60) を示した図である．式 (2.60) は絶対値関数を含むため，$\theta = -5,\ 1,\ 6$ の 3 点において微分不可能である．この関数が最小値となる変数 θ は図より $\theta = 1$ と確認できるが，以下では，この解を MM アルゴリズムにより求めてみる．

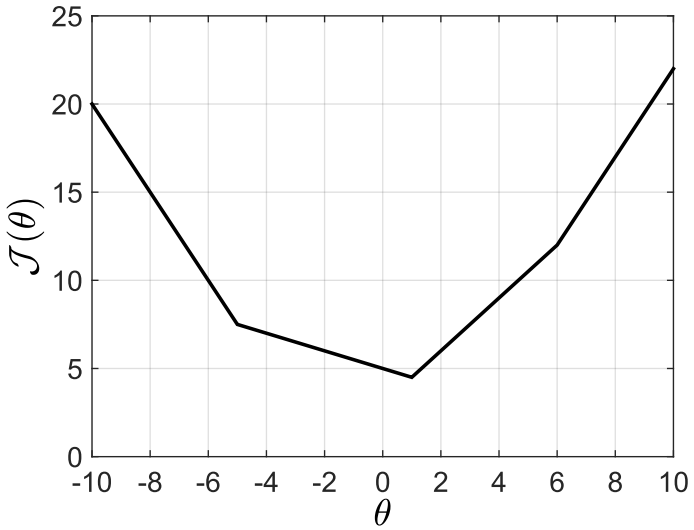

図 10　微分不可能な点を含む 1 変数の目的関数

式 (2.60) が微分不可能である要因は絶対値関数を含んでいることである．いま，仮に式 (2.60) 中の絶対値関数を全て微分可能な関数で置き換えた新しい関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ を定義する．ここで，新たに導入された変数 $\tilde{\theta}$ は「どのような関数で絶対値関数を置き換えるか」を左右するパラメータと理解すればよい．便宜上，この新たに導入した変数 $\tilde{\theta}$ を補助変数（auxiliary variable）と呼び，新たに得られた関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ を補助関数（auxiliary function）と呼ぶこととする．補助関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ は微分可能な関数であるため，$\partial \mathcal{J}^+ / \partial \theta = 0$ を直接計算でき，その最小点（停留点）を求めることができる．当然，本来の目的関数 $\mathcal{J}(\theta)$ を微分可能な補助関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ にすり替えてしまっているため，一見 $\mathcal{J}^+(\theta, \tilde{\theta})$ の最小点（$\partial \mathcal{J}^+ / \partial \theta = 0$ を満たす θ）には何も意味がないように思われる．しかし，補助関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ がもし次の性質を満たしていれば，これを用いて本来の微分不可能な関数 $\mathcal{J}(\theta)$ を間接的に最小化し解を導く手法が提案されており，これを MM アルゴリズムと呼ぶ．

- 定義域内の任意の θ 及び $\tilde{\theta}$ に対して $\mathcal{J}(\theta) \leq \mathcal{J}^+(\theta, \tilde{\theta})$ が常に成立する
- 定義域内の任意の θ に対して $\mathcal{J}(\theta) = \mathcal{J}^+(\theta, \tilde{\theta})$ を満たす $\tilde{\theta}$ が存在する

この 2 つの性質は次式で置き換えることもできる．

$$\mathcal{J}(\theta) = \min_{\tilde{\theta}} \mathcal{J}^+(\theta, \tilde{\theta}) \tag{2.61}$$

即ち，補助関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ は常に目的関数 $\mathcal{J}(\theta)$ の上側に存在し，最小点でのみ目的関数 $\mathcal{J}(\theta)$ と接している．改めて，式 (2.61) を満たす $\mathcal{J}^+(\theta, \tilde{\theta})$ を「補助関数」と定義する．なお，補助関数はその性質から上限関数（majorization function）と呼ばれることも多い．

微分可能な補助関数 $\mathcal{J}^+(\theta, \tilde{\theta})$ を使って目的関数 $\mathcal{J}(\theta)$ の最小解をどのように求めるかについて，式 (2.60) の例を用いて

説明する．今，絶対値関数 $\mathcal{Q}(\theta) = |\theta|$ に対する補助関数 $\mathcal{Q}^+(\theta, \tilde{\theta})$ を次のように与える．

$$\begin{aligned} \mathcal{Q}(\theta) &= |\theta| \\ &\leq \frac{\tilde{\theta}}{2}\theta^2 + \frac{1}{2\tilde{\theta}} \\ &\equiv \mathcal{Q}^+(\theta, \tilde{\theta}) \end{aligned} \tag{2.62}$$

絶対値関数 $\mathcal{Q}(\theta)$ と補助関数 $\mathcal{Q}^+(\theta, \tilde{\theta})$ を $\theta\tilde{\theta}$ 平面上に図示したものが図 11 である．図より，確かに $\mathcal{Q}^+(\theta, \tilde{\theta})$ は常に $\mathcal{Q}(\theta)$ の上側に存在し，ある $\theta = C$（C は定数）に対して補助関数 $\mathcal{Q}^+(\theta = C, \tilde{\theta})$ は最小点で $\mathcal{Q}(\theta)$ と接していることが分かる（図 11 における赤い曲線が接点の集合を示している）．従って，$\mathcal{Q}^+(\theta, \tilde{\theta})$ は式 (2.61) を満たしており，絶対値関数の補助関数と定義できる．また，不等式 (2.62) の等号が成立する条件（$\tilde{\theta}$ の値）について考えると，これは即ち図 11 における赤い曲線上の $\tilde{\theta}$ そのものである．このときの $\tilde{\theta}$ の値を求めると，次のようになる．

$$\begin{aligned} |\theta| &= \frac{\tilde{\theta}}{2}\theta^2 + \frac{1}{2\tilde{\theta}} \\ \theta^2\tilde{\theta}^2 - 2|\theta|\tilde{\theta} + 1 &= 0 \end{aligned}$$

従って，2 次方程式の解の公式より

$$\begin{aligned} \tilde{\theta} &= \frac{1}{2\theta^2}\left(2|\theta| \pm \sqrt{4|\theta|^2 - 4\theta^2}\right) \\ &= \frac{|\theta|}{\theta^2} \\ &= \frac{1}{|\theta|} \end{aligned} \tag{2.63}$$

が得られる．この式は反比例のグラフ $\tilde{\theta} = 1/\theta$ の $\theta < 0$ の領域（第 3 象限）を正（第 2 象限）に折り返したグラフである為，図 11 における赤い曲線と一致している．

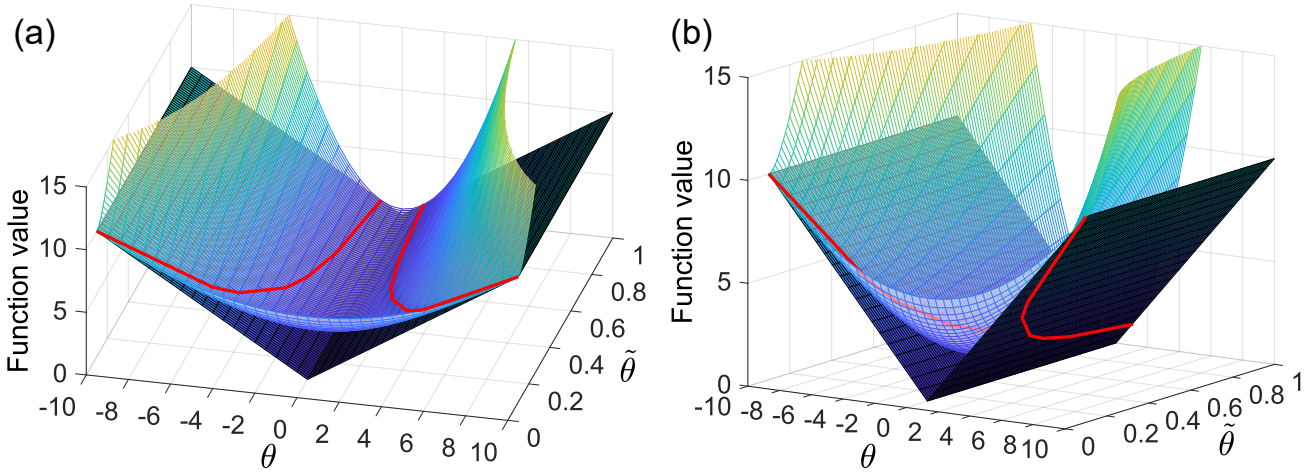

図 11　絶対値関数 $\mathcal{J}(\theta) = |\theta|$ を上から抑える補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}) = (\tilde{\theta}/2)\theta^2 + 1/(2\tilde{\theta})$ について 2 方向から描いたグラフ．目的関数と補助関数の接線は赤い曲線で示している．変数 θ をある値に固定したときの補助関数 $\mathcal{J}^+(\theta = C, \tilde{\theta})$（$C$ は定数）は常に最小点で目的関数 $\mathcal{J}(\theta)$ と接しており，$\mathcal{J}(\theta) = \min_{\tilde{\theta}} \mathcal{J}^+(\theta, \tilde{\theta})$ を満たしていることが分かる．また，補助変数 $\tilde{\theta}$ をある値に固定したときの補助関数 $\mathcal{J}^+(\theta, \tilde{\theta} = C)$ は，常に 2 点で目的関数 $\mathcal{J}(\theta)$ と接している．これらの接点は補助変数 $\tilde{\theta}$ の値が大きくなるほど原点に近づく．

絶対値関数に対する不等式 (2.62) を用いると，$\mathcal{Q}(\theta - C) = |\theta - C|$ に対しても $\mathcal{Q}^+(\theta - C, \tilde{\theta}) = (\tilde{\theta}/2)(\theta - C)^2 + 1/(2\tilde{\theta})$ という補助関数を作ることができるため，これを式 (2.60) に適用し，微分不可能な目的関数 $\mathcal{J}(\theta)$ の補助関数を次式のよう

に設計できる．

$$\begin{aligned}\mathcal{J}(\theta) &\leq \mathcal{J}^+(\theta,\tilde{\theta}_1,\tilde{\theta}_2,\tilde{\theta}_3)\\ &= \frac{1}{2}\left[\frac{\tilde{\theta}_1}{2}(\theta-6)^2+\frac{1}{2\tilde{\theta}_1}\right]+\frac{\tilde{\theta}_2}{2}(\theta-1)^2+\frac{1}{2\tilde{\theta}_2}+\frac{\tilde{\theta}_3}{2}(\theta+5)^2+\frac{1}{2\tilde{\theta}_3}-4\\ &= \frac{\tilde{\theta}_1}{4}(\theta-6)^2+\frac{\tilde{\theta}_2}{2}(\theta-1)^2+\frac{\tilde{\theta}_3}{2}(\theta+5)^2+\frac{1}{4\tilde{\theta}_1}+\frac{1}{2\tilde{\theta}_2}+\frac{1}{2\tilde{\theta}_3}-4\end{aligned} \tag{2.64}$$

ここで，3 つの絶対値関数に対して独立に不等式 (2.62) を適用しているため，3 種類の補助変数 $\tilde{\theta}_1$，$\tilde{\theta}_2$，及び $\tilde{\theta}_3$ を導入している．各不等式の等号成立条件は式 (2.63) より

$$\tilde{\theta}_1 = \frac{1}{|\theta-6|} \tag{2.65}$$

$$\tilde{\theta}_2 = \frac{1}{|\theta-1|} \tag{2.66}$$

$$\tilde{\theta}_3 = \frac{1}{|\theta+5|} \tag{2.67}$$

である．例えば，$\theta=-8$ で目的関数と接する補助関数 (2.64) を考える．ここで，$\theta=-8$ を θ の初期値という意味で $\theta^{(0)}$ と表す．この接点では等号成立条件が成立するため，$\theta=-8$ で目的関数と接する補助関数の補助変数は式 (2.65)–(2.67) に $\theta=-8$ を代入することで求めることができる．この補助変数は $\theta^{(0)}$ に依存して決まる定数であり，$\theta^{(0)}$ の点で目的関数と接するような補助関数の補助変数を新たに定めたという意味で，それぞれ $\tilde{\theta}_1^{(1)}$，$\tilde{\theta}_2^{(1)}$，及び $\tilde{\theta}_3^{(1)}$ と表す．従って，$\theta=-8$ で目的関数と接する補助関数は $\tilde{\theta}_1^{(1)}$，$\tilde{\theta}_2^{(1)}$，及び $\tilde{\theta}_3^{(1)}$ を式 (2.64) に代入した $\mathcal{J}^+(\theta,\theta_1^{(1)},\theta_2^{(1)},\theta_3^{(1)})$ として与えられる．これを図示したものが図 12(a) である．確かに $\theta=\theta^{(0)}$ で目的関数 $\mathcal{J}(\theta)$ と補助関数 $\mathcal{J}^+(\theta,\theta_1^{(1)},\theta_2^{(1)},\theta_3^{(1)})$ が接しており，それ以外の θ では常に補助関数が目的関数の上側に存在することがわかる．ここで，補助関数 $\mathcal{J}^+(\theta,\theta_1^{(1)},\theta_2^{(1)},\theta_3^{(1)})$ の最小点に着目すると，この点は $\partial\mathcal{J}^+/\partial\theta=0$ を満たしており，また補助関数が 2 次関数であることから，解析的に求められることが分かる．この補助関数の最小点を新たな（更新された）変数値という意味で $\theta^{(1)}$ と表すならば，補助関数の性質 (2.61) より，$\mathcal{J}(\theta^{(0)})\geq\mathcal{J}(\theta^{(1)})$ が成立する．即ち，変数 θ の更新前と更新後において目的関数値の単調非増加性が保証される．再び $\theta^{(1)}$ で目的関数と接する補助関数 $\mathcal{J}^+(\theta,\theta_1^{(2)},\theta_2^{(2)},\theta_3^{(2)})$ を式 (2.65)–(2.67) 及び式 (2.64) から設計し，同様の操作を繰り返すことで，図 12(b)–(f) のように目的関数の最小解に近づくことができる．変数の更新時にはやはり目的関数の単調非増加性 $\mathcal{J}(\theta^{(t)})\geq\mathcal{J}(\theta^{(t+1)})$（$t=1,2,\cdots$ は更新回数を表すインデックス）が保証されており，この反復更新アルゴリズムはやがて最小解で変数更新が止まる（目的関数値の列が収束する）ことが分かる[*17]．図 13 は初期値を $\theta^{(0)}=9$ として反復更新したときの様子である．$\theta^{(0)}=-8$ の図 12 の場合と比較して，早い段階で目的関数の最小解に近づき，以降の反復更新ではほとんど変化がなく収束している様子が分かる．

以上が MM アルゴリズムによる反復更新と最小解の推定法であるが，引き続き式 (2.60) を例にとり，数式的としての手続きをまとめる．初期値 $\theta^{(0)}$ で目的関数と接する補助関数が式 (2.64) によって $\mathcal{J}^+(\theta,\tilde{\theta}_1^{(1)},\tilde{\theta}_2^{(1)},\tilde{\theta}_3^{(1)})$ として与えられているため，次はこの関数の最小点 $\theta^{(1)}$ を求める必要がある．補助変数 $\tilde{\theta}_1^{(1)}$，$\tilde{\theta}_2^{(1)}$，及び $\tilde{\theta}_3^{(1)}$ は $\theta^{(0)}$ と等号成立条件 (2.65)–(2.67) から定数である点に注意して $\partial\mathcal{J}^+/\partial\theta=0$ を計算すると，次のように得られる．

$$\begin{aligned}\frac{\tilde{\theta}_1^{(1)}}{2}(\theta-6)+\tilde{\theta}_2^{(1)}(\theta-1)+\tilde{\theta}_3^{(1)}(\theta+5)&=0\\ \left(\frac{1}{2}\tilde{\theta}_1^{(1)}+\tilde{\theta}_2^{(1)}+\tilde{\theta}_3^{(1)}\right)\theta &= 3\tilde{\theta}_1^{(1)}+\tilde{\theta}_2^{(1)}-5\tilde{\theta}_3^{(1)}\\ \theta &= \frac{3\tilde{\theta}_1^{(1)}+\tilde{\theta}_2^{(1)}-5\tilde{\theta}_3^{(1)}}{\frac{1}{2}\tilde{\theta}_1^{(1)}+\tilde{\theta}_2^{(1)}+\tilde{\theta}_3^{(1)}}\end{aligned} \tag{2.68}$$

従って，式 (2.68) で求まる θ が更新後の変数 $\theta^{(1)}$ となる．MM アルゴリズムでは，式 (2.68) のように，補助関数の最小点が解析的に求められることが重要である．もし補助関数の最小点が解析的に求められない場合はラインサーチ法等を用いて

[*17] 但し，収束するためには目的関数 $\mathcal{J}(\theta)$ の値域が下に有界である（下限が存在する）必要がある．

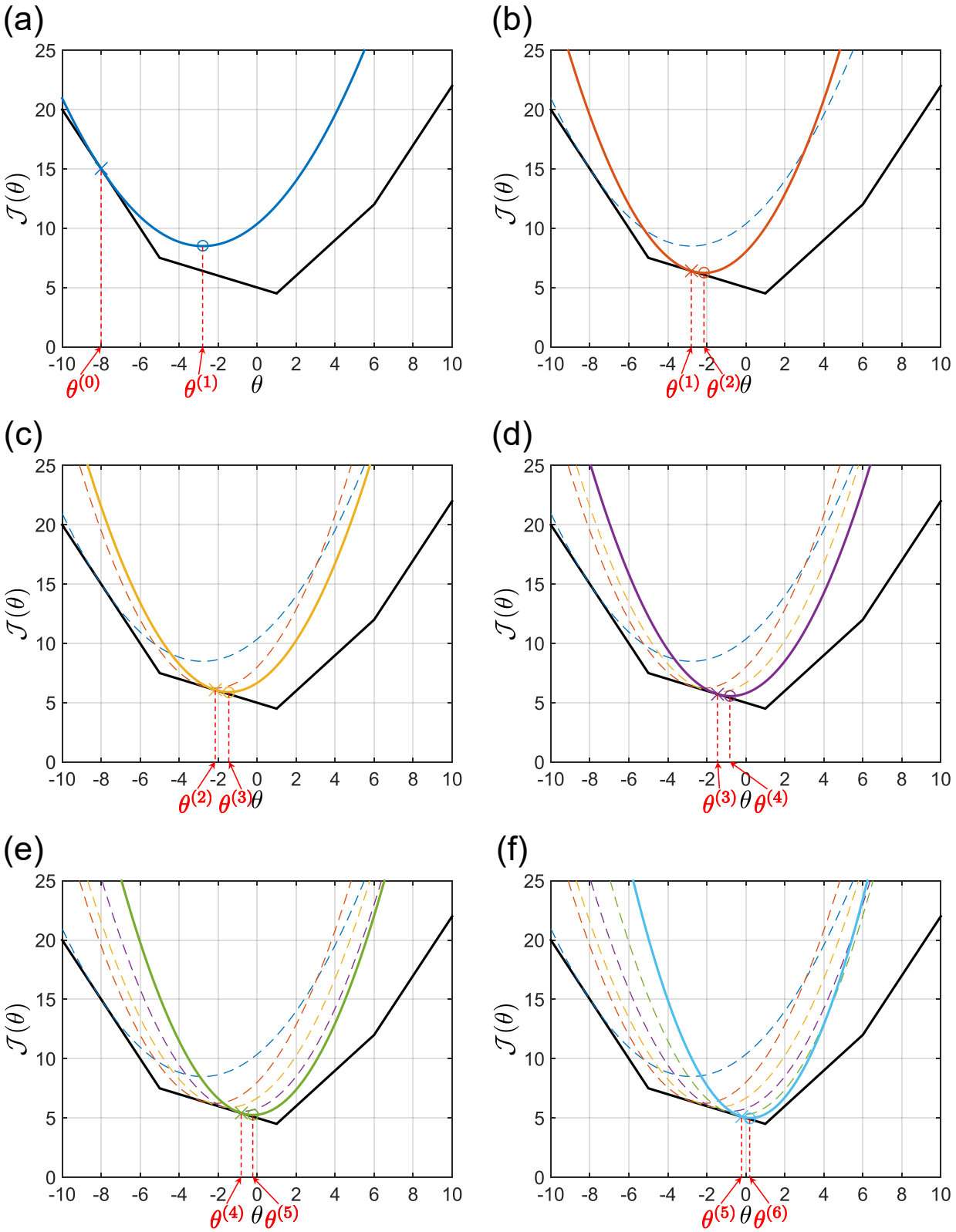

図 12　黒い線で表された目的関数 $\mathcal{J}(\theta)$ に対して初期値 $\theta^{(0)} = -8$ としたときの MM アルゴリズムによる変数更新の様子：(a) $\theta^{(0)}$ で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(1)}, \tilde{\theta}_2^{(1)}, \tilde{\theta}_3^{(1)})$，(b) $\theta^{(1)}$ で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(2)}, \tilde{\theta}_2^{(2)}, \tilde{\theta}_3^{(2)})$，(c) $\theta^{(2)}$ で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(3)}, \tilde{\theta}_2^{(3)}, \tilde{\theta}_3^{(3)})$，(d) $\theta^{(3)}$ で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(4)}, \tilde{\theta}_2^{(4)}, \tilde{\theta}_3^{(4)})$，(e) $\theta^{(4)}$ で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(5)}, \tilde{\theta}_2^{(5)}, \tilde{\theta}_3^{(5)})$，(f) $\theta^{(5)}$ で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(6)}, \tilde{\theta}_2^{(6)}, \tilde{\theta}_3^{(6)})$．2 つの関数の接点 $\mathcal{J}(\theta^{(t)}) = \mathcal{J}^+(\theta^{(t)}, \tilde{\theta}_1^{(t+1)}, \tilde{\theta}_2^{(t+1)}, \tilde{\theta}_3^{(t+1)})$ をクロスマークで示している．また，$\partial \mathcal{J}^+ / \partial \theta = 0$ を満たす補助関数の最小点での補助関数値 $\mathcal{J}^+(\theta^{(t+1)}, \tilde{\theta}_1^{(t+1)}, \tilde{\theta}_2^{(t+1)}, \tilde{\theta}_3^{(t+1)})$ をサークルで示している．

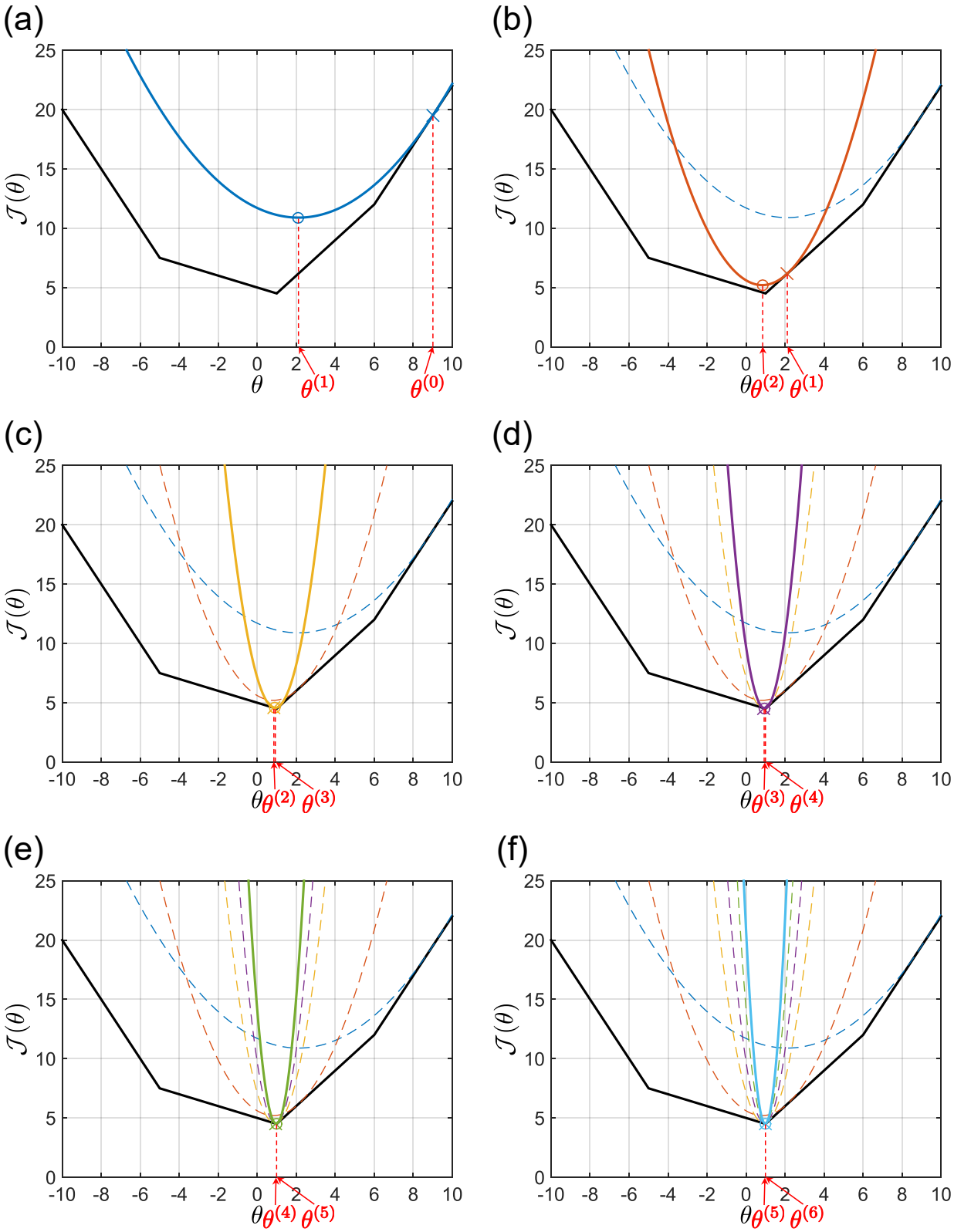

図 13　黒い線で表された目的関数 $\mathcal{J}(\theta)$ に対して初期値 $\theta^{(0)} = 9$ としたときの MM アルゴリズムによる変数更新の様子：(a) $\theta^{(0)}$ で接する補助関数 $\mathcal{J}^{+}(\theta, \tilde{\theta}_1^{(1)}, \tilde{\theta}_2^{(1)}, \tilde{\theta}_3^{(1)})$，(b) $\theta^{(1)}$ で接する補助関数 $\mathcal{J}^{+}(\theta, \tilde{\theta}_1^{(2)}, \tilde{\theta}_2^{(2)}, \tilde{\theta}_3^{(2)})$，(c) $\theta^{(2)}$ で接する補助関数 $\mathcal{J}^{+}(\theta, \tilde{\theta}_1^{(3)}, \tilde{\theta}_2^{(3)}, \tilde{\theta}_3^{(3)})$，(d) $\theta^{(3)}$ で接する補助関数 $\mathcal{J}^{+}(\theta, \tilde{\theta}_1^{(4)}, \tilde{\theta}_2^{(4)}, \tilde{\theta}_3^{(4)})$，(e) $\theta^{(4)}$ で接する補助関数 $\mathcal{J}^{+}(\theta, \tilde{\theta}_1^{(5)}, \tilde{\theta}_2^{(5)}, \tilde{\theta}_3^{(5)})$，(f) $\theta^{(5)}$ で接する補助関数 $\mathcal{J}^{+}(\theta, \tilde{\theta}_1^{(6)}, \tilde{\theta}_2^{(6)}, \tilde{\theta}_3^{(6)})$．目的関数の最小解付近では変数の更新が収束している様子が確認できる．

求めることもできるが，そのような数値解析を行うのであれば，初めから目的関数に数値解析を適用すればよいため，MM アルゴリズムを使う意義はなくなる．従って，補助関数の設計において，微分可能かつ 2 次関数のように最小点が解析的に求められる関数を用いることが重要になる．

補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(1)}, \tilde{\theta}_2^{(1)}, \tilde{\theta}_3^{(1)})$ の最小点 $\theta^{(1)}$ を求めた後は，再びその点で接する補助関数 $\mathcal{J}^+(\theta, \tilde{\theta}_1^{(2)}, \tilde{\theta}_2^{(2)}, \tilde{\theta}_3^{(2)})$ を設計する．この補助関数の設計には，新しい補助変数 $\tilde{\theta}_1^{(2)}$，$\tilde{\theta}_2^{(2)}$，及び $\tilde{\theta}_3^{(2)}$ が必要であるため，式 (2.65)–(2.67) の等号成立条件よりこれらを求めればよい．

以上の手続きをまとめる．今，等号成立条件 (2.65)–(2.67) による補助変数の更新を，更新回数 t に対する一般形として

$$\tilde{\theta}_1^{(t+1)} = \frac{1}{|\theta^{(t)} - 6|} \tag{2.69}$$

$$\tilde{\theta}_2^{(t+1)} = \frac{1}{|\theta^{(t)} - 1|} \tag{2.70}$$

$$\tilde{\theta}_3^{(t+1)} = \frac{1}{|\theta^{(t)} + 5|} \tag{2.71}$$

と表し，さらに補助関数の最小点を与える式 (2.68) を，更新回数 t に対する一般形として

$$\theta^{(t+1)} = \frac{3\tilde{\theta}_1^{(t+1)} + \tilde{\theta}_2^{(t+1)} - 5\tilde{\theta}_3^{(t+1)}}{\frac{1}{2}\tilde{\theta}_1^{(t+1)} + \tilde{\theta}_2^{(t+1)} + \tilde{\theta}_3^{(t+1)}} \tag{2.72}$$

と表現しなおすとき，目的関数 (2.60) の MM アルゴリズムによる最小化手順は次のようになる．

1. 適当な初期値 $\theta^{(0)}$ をとる
2. 以下の操作を $t = 0, 1, 2, \cdots$ について繰り返す
   （a）式 (2.69)–(2.71) から補助変数 $\tilde{\theta}_1^{(t+1)}$，$\tilde{\theta}_2^{(t+1)}$，及び $\tilde{\theta}_3^{(t+1)}$ を求める
   （b）式 (2.72) から変数 $\theta^{(t+1)}$ を求める

上記アルゴリズムの 2.(a) は，点 $\theta^{(t)}$ で接する補助関数を求めることに対応しており，また 2.(b) は求めた補助関数の最小点を求めることに対応している．なお，補助関数の性質より

$$\mathcal{J}(\theta^{(t)}) = \mathcal{J}^+(\theta^{(t)}, \tilde{\theta}^{(t+1)}) \geq \mathcal{J}^+(\theta^{(t+1)}, \tilde{\theta}^{(t+1)}) \geq \mathcal{J}(\theta^{(t+1)}) \tag{2.73}$$

が成り立つため，このアルゴリズムの単調非増加性と収束性は保証される．さらに，等号成立条件と補助関数の最小点を求める式がいずれも解析的に与えられていることから，上記のアルゴリズムの 2.(a) 及び 2.(b) を 1 つにまとめて次式で表すこともできる．

$$\theta^{(t+1)} = \frac{\frac{3}{|\theta^{(t)}-6|} + \frac{1}{|\theta^{(t)}-1|} - \frac{5}{|\theta^{(t)}+5|}}{\frac{1}{2|\theta^{(t)}-6|} + \frac{1}{|\theta^{(t)}-1|} + \frac{1}{|\theta^{(t)}+5|}} \tag{2.74}$$

このようにして得られる式を θ の反復更新則（iterative update rule）と呼ぶ．

より一般的な話として，上記の最適化手法を多変量な関数にも適用できるように拡張した場合についてまとめる．

MM アルゴリズム

N 次元ベクトル変数 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_N)^{\mathrm{T}} \in \Theta$ に対する目的関数 $\mathcal{J}(\boldsymbol{\theta}) : \Theta \to \mathbb{R}$ において，$\mathcal{J}(\boldsymbol{\theta})$ の最小解を求める問題を考える．

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \tag{2.75}$$

MM アルゴリズムでは，任意の $\boldsymbol{\theta} \in \Theta$ 及び $\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}$ に対して次の式を満たす関数を補助関数として定義する．

$$\mathcal{J}(\boldsymbol{\theta}) = \min_{\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}} \mathcal{J}^{+}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \tag{2.76}$$

この補助関数に基づいて，次の手順によって目的関数 $\mathcal{J}(\boldsymbol{\theta})$ の最小解を求めることができる．

1. 適当な初期値 $\boldsymbol{\theta}^{(0)}$ をとる
2. 以下の操作を $t = 0, 1, 2, \cdots$ について繰り返す
   （a）次式を満たす $\tilde{\boldsymbol{\theta}}^{(t+1)}$ を求める

$$\tilde{\boldsymbol{\theta}}^{(t+1)} = \arg\min_{\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}} \mathcal{J}^{+}(\boldsymbol{\theta}^{(t)}, \tilde{\boldsymbol{\theta}}) \tag{2.77}$$

   （b）次式を満たす $\boldsymbol{\theta}^{(t+1)}$ を求める

$$\boldsymbol{\theta}^{(t+1)} = \arg\min_{\boldsymbol{\theta} \in \Theta} \mathcal{J}^{+}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}^{(t+1)}) \tag{2.78}$$

従って MM アルゴリズムは，設計した補助関数 $\mathcal{J}^{+}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ に対する式 (2.77) 及び (2.78) の交互反復最小化と捉えることができる．

最後に，補助関数の設計法について考える．MM アルゴリズムにおいて最も困難な部分は，如何にして式 (2.76) を満たす補助関数を見つけるかという問題である．さらに，前述の通り，補助関数の最小点が解析的に求められることも重要である．この問題に関しては一般的な解決法は現在提案されておらず，問題依存（目的関数依存）でかつ場当たり的，あるいは発見的な手法に頼るほかない．しかしながら，過去の文献で補助関数法の設計に利用されてきた不等式として，以下が挙げられる．

絶対値関数を 2 次関数で抑える不等式

$\theta \in \mathbb{R}$ に対して

$$|\theta| \leq \frac{\tilde{\theta}}{2}\theta^2 + \frac{1}{2\tilde{\theta}} \tag{2.79}$$

が成立する．なお，等号成立条件は $\tilde{\theta} = 1/|\theta|$ である．

Jensen の不等式

$\tilde{\theta}_1, \tilde{\theta}_2, \cdots, \tilde{\theta}_N \in \mathbb{R}_{\geq 0}$ を，$\sum_{n=1}^{N} \tilde{\theta}_n = 1$ を満たす正の実数の列としたとき，$\theta_1, \theta_2, \cdots, \theta_N \in \mathbb{R}$ に対して

$$f\left(\sum_{n=1}^{N} \tilde{\theta}_n \theta_n\right) \leq \sum_{n=1}^{N} \tilde{\theta}_n f\left(\theta_n\right) \tag{2.80}$$

が成立する．ここで，関数 $f : \mathbb{R} \to \mathbb{R}$ は凸関数である．なお，等号成立条件は $\theta_1 = \theta_2 = \cdots = \theta_N$ で与えられる．

凹関数の接線不等式

$\theta \in \mathbb{R}$ に対して

$$f(\theta) \leq f'(\tilde{\theta})\left(\theta - \tilde{\theta}\right) + f(\tilde{\theta}) \tag{2.81}$$

が成立する．ここで，関数 $f : \mathbb{R} \to \mathbb{R}$ は凹関数である．なお，等号成立条件は $\tilde{\theta} = \theta$ で与えられる．

特に NMF の反復更新則の導出では，2 次関数（$(\sum_{n=1}^{N}\theta_n)^2$ 等）への Jensen の不等式の適用や，対数関数（$\log\sum_{n=1}^{N}\theta_n$ 等）への接線不等式の適用が頻繁に用いられる．

3. NMF の最適化法

本章では，なぜ NMF がパターン抽出において一躍有名となったか，その理由である NMF の本質的な意味について先に説明する．次に，NMF における最適化問題を定式化し，その解を推定するための反復更新則の導出法についてまとめる．なお，反復更新則の導出については，2 章で取り扱った不等式制約条件付き最適化問題の解法や MM アルゴリズムを用いている．

3.1 NMF の本質的意味

NMF は観測された非負値行列 $\boldsymbol{X}\in\mathbb{R}_{\geq 0}^{I\times J}$ を次式で近似するアルゴリズムである．

$$\boldsymbol{X}\approx\boldsymbol{W}\boldsymbol{H} \tag{3.1}$$

ここで，$\boldsymbol{W}\in\mathbb{R}_{\geq 0}^{I\times K}$ 及び $\boldsymbol{H}\in\mathbb{R}_{\geq 0}^{K\times J}$ は NMF で求めるべき変数となる非負値行列である．また，K は $\boldsymbol{W}$ の列数又は $\boldsymbol{H}$ の行数であり，事前に決めておく定数である．式 (3.1) より，NMF は観測非負値行列 $\boldsymbol{X}$ を，別の二つの非負値行列 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ の行列積に近似的に分解していることが分かる．$\boldsymbol{X}$，$\boldsymbol{W}$，及び $\boldsymbol{H}$ の各要素をそれぞれ $x_{ij}\geq 0$，$w_{ik}\geq 0$，及び $h_{kj}\geq 0$ と表現すると，行列積である式 (3.1) は，要素毎に考えることで次式のようにも表現できる．

$$x_{ij}\approx\sum_{k=1}^{K}w_{ik}h_{kj}\quad\forall i=1,2,\cdots,I,\ j=1,2,\cdots,J,\ k=1,2,\cdots,K \tag{3.2}$$

さらに，ベクトルによる表現も定義しておく．$\boldsymbol{X}$ 及び $\boldsymbol{W}$ の各列ベクトルをそれぞれ $\boldsymbol{x}_j\in\mathbb{R}_{\geq 0}^{I}$ 及び $\boldsymbol{w}_k\in\mathbb{R}_{\geq 0}^{I}$ と定義し，$\boldsymbol{H}$ の行ベクトルを $\boldsymbol{h}_k^{\mathrm{T}}\in\mathbb{R}_{\geq 0}^{J}$ と定義すると，$\boldsymbol{X}$，$\boldsymbol{W}$，及び $\boldsymbol{H}$ は次のように表現できる．

$$\begin{aligned}\boldsymbol{X}&=\begin{pmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_J\end{pmatrix}\\&=\begin{pmatrix}x_{11} & x_{12} & \cdots & x_{1J}\\ x_{21} & x_{22} & \cdots & x_{2J}\\ \vdots & \vdots & \ddots & \vdots\\ x_{I1} & x_{I2} & \cdots & x_{IJ}\end{pmatrix}\end{aligned} \tag{3.3}$$

$$\begin{aligned}\boldsymbol{W}&=\begin{pmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_K\end{pmatrix}\\&=\begin{pmatrix}w_{11} & w_{12} & \cdots & w_{1K}\\ w_{21} & w_{22} & \cdots & w_{2K}\\ \vdots & \vdots & \ddots & \vdots\\ w_{I1} & w_{I2} & \cdots & w_{IK}\end{pmatrix}\end{aligned} \tag{3.4}$$

$$\begin{aligned}\boldsymbol{H}&=\begin{pmatrix}\boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots & \boldsymbol{h}_K\end{pmatrix}^{\mathrm{T}}\\&=\begin{pmatrix}\boldsymbol{h}_1^{\mathrm{T}}\\ \boldsymbol{h}_2^{\mathrm{T}}\\ \vdots\\ \boldsymbol{h}_K^{\mathrm{T}}\end{pmatrix}\\&=\begin{pmatrix}h_{11} & h_{12} & \cdots & h_{1J}\\ h_{21} & h_{22} & \cdots & h_{2J}\\ \vdots & \vdots & \ddots & \vdots\\ h_{K1} & h_{K2} & \cdots & h_{KJ}\end{pmatrix}\end{aligned} \tag{3.5}$$

このベクトル表記を用いると，式 (3.1) は次式のようにも表現できる．

$$\boldsymbol{X}=\sum_{k=1}^{K}\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}} \tag{3.6}$$

ここで，式 (3.6) の右辺の $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ はサイズが $I \times 1$ と $1 \times J$ の行列積と考えるため，結果はサイズが $I \times J$ の行列となり，$\boldsymbol{X}$ のサイズと一致する．

例えば観測非負値行列 $\boldsymbol{X}$ が

$$\boldsymbol{X} = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \tag{3.7}$$

で与えられる場合に，$K = 2$ という条件では

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{H} = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.8}$$

という非負値行列 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を求めることができれば，NMF の目的は達成されたことになる．但し，このような分解は $\boldsymbol{H}$ が単位行列 $\boldsymbol{I}$ であることから自明であり，とくに恩恵は見当たらない．また，この分解は一意ではなく，例えば $\boldsymbol{W} = \boldsymbol{I}$ かつ $\boldsymbol{H} = \boldsymbol{X}$ としても成立する．

一方，$K = 1$ という条件では，次のような行列分解ができる．

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{H} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 3 \end{pmatrix} \tag{3.9}$$

式 (3.9) はサイズが 2×1 の行列と 1×2 行列の行列積なので，その結果はサイズ 2×2 の行列となり，計算してみると $\boldsymbol{X}$ と一致することが分かる．式 (3.9) のような分解ができる理由は，観測非負値行列 $\boldsymbol{X}$ のランクが $\mathrm{rank}(\boldsymbol{X}) = 1$ であることにほかならない．つまり，$\boldsymbol{X}$ に含まれる 2 本の列（または行）ベクトルは互いに線形従属（linearly dependent）であり，同じ部分空間上に属している．今，$\boldsymbol{X} = (\boldsymbol{x}_1\ \boldsymbol{x}_2)$ と表現すると，確かに式 (3.7) の例では $\boldsymbol{x}_2 = 3\boldsymbol{x}_1$ が成り立っており，これらが互いに同一直線上に存在するため，線形独立（linearly independent）なベクトルが 1 本しかないことが分かる．即ち，行列 $\boldsymbol{X}$ の基底は 1 本である[*18]．

サイズが $I \times J$ の一般行列は最大で $\min(I, J)$ 本の線形独立なベクトル（基底）を持ちうる．言い換えれば，最大で $\min(I, J)$ のランクを持ちうる．実際に最大ランクを持つような行列は，一般にフルランク行列（full-rank matrix）と呼ばれる．サイズが 2×2 である $\boldsymbol{X}$ の取りうる最大ランクは 2 である．ところが，前述の通り，$\boldsymbol{X}$ が含む列（または行）ベクトルが互いに線形従属の関係にあれば，$\boldsymbol{X}$ のランクは最大ランクに達しないという状況になる．このような行列はランク落ち行列（rank-deficient matrix）と呼ばれる．また，例えばサイズが 100×200 という行列のランクが 10 であった場合，これは最大ランクが 100 であるにも関わらず，ほとんどの行（または列）ベクトルが線形従属という状況である．最大ランクに対して実際のランクがどの程度まで低い値になっているときに該当するかという明確な基準はないが，このような行列を一般に低ランク行列（low-rank matrix）と呼ぶ．式 (3.7) の例では，$\boldsymbol{X}$ のランクは 1 であり，$\boldsymbol{X}$ の基底である $\boldsymbol{x}_1$ が分解行列 $\boldsymbol{W}$ に，また基底の係数が $\boldsymbol{H}$ に現れている．

ここまでの例（式 (3.8) 及び (3.9)）は，$\mathrm{rank}(\boldsymbol{X}) \leq K$ となるように K を定めたときの NMF による行列分解であった．式 (3.9) のように $\boldsymbol{X}$ のランクと K が一致する状況を完備（complete）と呼び，また式 (3.8) のように $\boldsymbol{X}$ のランクを K が上回る状況を過完備（over-complete）と呼ぶ．完備または過完備な状況においては，NMF は $\boldsymbol{X}$ を $\boldsymbol{W}\boldsymbol{H}$ によって完全に再構成できる（式 (3.1) において等号を成立させることができる）．しかし，$\mathrm{rank}(\boldsymbol{X}) > K$ となる例では，$\boldsymbol{W}\boldsymbol{H}$ では完全に $\boldsymbol{X}$ を再構成することはできず，低ランク近似（low-rank approximation）分解となる．NMF の本質は，この低ランク近似にある．

NMF による低ランク近似分解の例を示す．観測非負値行列 $\boldsymbol{X}$ が

$$\boldsymbol{X} = \begin{pmatrix} 1 & 3 & 4 \\ 2 & 6 & 9 \\ 3 & 9 & 10 \end{pmatrix} \tag{3.10}$$

で与えられる場合を考える．$\boldsymbol{X}$ の 1 列目 $\boldsymbol{x}_1$ と 2 列目 $\boldsymbol{x}_2$ は $\boldsymbol{x}_2 = 3\boldsymbol{x}_1$ であるため，線形従属である．3 列目 $\boldsymbol{x}_3$ は $\boldsymbol{x}_1$ や $\boldsymbol{x}_2$ を用いて表せないため，$\boldsymbol{x}_1$ と $\boldsymbol{x}_3$，あるいは $\boldsymbol{x}_2$ と $\boldsymbol{x}_3$ は線形独立である．以上より，行列 $\boldsymbol{X}$ に含まれる線形独立なベクト

[*18] 本来，行列の基底とは線形独立なベクトルを成す集合の概念である．例えば，式 (3.7) の基底は $\{C\boldsymbol{x}_1 | C \in \mathbb{R}_{\neq 0}\}$ という集合であるため，1 つの集合を「1 本」と表現するのは誤りであるが，ここでは分かりやすさの為，基底を成す集合を「本」という単位で表現している．

ルの本数は 2 となり，$\mathrm{rank}(\boldsymbol{X}) = 2$ であることが分かる．この行列に対して，$K = 1$ という条件で NMF による分解を考えたとき，例えば以下のような結果があり得る．

$$\begin{aligned}\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} &= \begin{pmatrix}1\\2\\3\end{pmatrix}\begin{pmatrix}1 & 3 & 4\end{pmatrix}\\ &= \begin{pmatrix}1 & 3 & 4\\2 & 6 & 8\\3 & 9 & 12\end{pmatrix}\end{aligned} \tag{3.11}$$

式 (3.10) と式 (3.11) を比較すると，次式に示す誤差が生じていることが分かる．

$$\begin{aligned}\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_{\mathrm{Fr}}^2 &= \left\| \begin{pmatrix}1 & 3 & 4\\2 & 6 & 9\\3 & 9 & 10\end{pmatrix} - \begin{pmatrix}1 & 3 & 4\\2 & 6 & 8\\3 & 9 & 12\end{pmatrix} \right\|_{\mathrm{Fr}}^2\\ &= \left\| \begin{pmatrix}0 & 0 & 0\\0 & 0 & 1\\0 & 0 & -2\end{pmatrix} \right\|_{\mathrm{Fr}}^2\\ &= 5\end{aligned} \tag{3.12}$$

ここで，$\|\cdot\|_{\mathrm{Fr}}^2$ は Frobenius ノルムであり，実数一般行列を $\boldsymbol{A} \in \mathbb{R}^{I\times J}$，その要素を a_{ij} と定義したとき，次式で与えられる．

$$\|\boldsymbol{A}\|_{\mathrm{Fr}} = \sqrt{\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})} \tag{3.13}$$

$$= \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}|a_{ij}|^2} \tag{3.14}$$

また，$\mathrm{tr}(\cdot)$ は正方行列の跡（trace）であり，実数正方行列を $\boldsymbol{Q} \in \mathbb{R}^{I\times I}$，その要素を q_{ij} と定義したとき，次式で与えられる．

$$\mathrm{tr}(\boldsymbol{Q}) = \sum_{i=1}^{I} q_{ii} \tag{3.15}$$

式 (3.12) で求めた値は二乗 Euclid 距離に基づく誤差であり，$\boldsymbol{W}\boldsymbol{H}$ が $\boldsymbol{X}$ をどの程度近似できたかを表す指標といえる．この $\mathrm{rank}(\boldsymbol{X}) > K$ という状況の NMF では，$\boldsymbol{X}$ に含まれる基底の一部が $\boldsymbol{W}$ に現れ，その基底の係数が $\boldsymbol{H}$ に現れることになる．特に，式 (3.12) のように低ランク近似誤差が小さくなるように $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を求めることができたならば，そのときに $\boldsymbol{W}$ に含まれる K 本の基底は，$\boldsymbol{X}$ の大部分を表現できる重要なベクトルとなるはずである．これは，1 章で述べた「行列が含む潜在パターン」そのものである．従って NMF は，低ランク近似によって観測非負値行列中に含まれる少数（K 個）の潜在パターンを抽出するアルゴリズムと捉えることができる．

一例として，図 1(a) の購買データを観測非負値行列 $\boldsymbol{X}$ と考え，$\boldsymbol{X}$ から 2 本の基底を抽出するために $K = 2$ とおいて NMF で分解した結果を示す．観測非負値行列 $\boldsymbol{X}$ は

$$\boldsymbol{X} = \begin{pmatrix}1 & 2 & 0 & 0 & 1\\1 & 3 & 1 & 2 & 2\\0 & 0 & 3 & 5 & 3\end{pmatrix} \tag{3.16}$$

と与えられており，これを調べてみると線形独立なベクトルは 3 本あるため，$\mathrm{rank}(\boldsymbol{X}) = 3$ である．すなわち，$\boldsymbol{X}$ はフルランク行列である．NMF による分解例が

$$\begin{aligned}\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} &= \begin{pmatrix}1 & 0\\1 & 1\\0 & 3\end{pmatrix}\begin{pmatrix}1 & 2 & 0 & 0 & 1\\0 & 0 & 1 & 2 & 1\end{pmatrix}\\ &= \begin{pmatrix}1 & 2 & 0 & 0 & 1\\1 & 2 & 1 & 2 & 2\\0 & 0 & 3 & 6 & 3\end{pmatrix}\end{aligned} \tag{3.17}$$

であった場合，このときの低ランク近似誤差は

$$\begin{aligned}\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\|_{\mathrm{Fr}}^2 &= \left\| \begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 1 & 3 & 1 & 2 & 2 \\ 0 & 0 & 3 & 5 & 3 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 1 & 2 & 1 & 2 & 2 \\ 0 & 0 & 3 & 6 & 3 \end{pmatrix} \right\|_{\mathrm{Fr}}^2 \\ &= \left\| \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix} \right\|_{\mathrm{Fr}}^2 \\ &= 2 \end{aligned} \tag{3.18}$$

となる．分解して得られた $\boldsymbol{W}$ に含まれる基底ベクトルは $\boldsymbol{w}_1 = (1\ 1\ 0)^{\mathrm{T}}$ 及び $\boldsymbol{w}_2 = (0\ 1\ 3)^{\mathrm{T}}$ である．これは前者が「商品 X を 1 個購入した顧客は商品 Y を 1 個購入するという相関関係」を示しており，後者が「商品 Y を 1 個購入した顧客は商品 Z を 3 個購入するという相関関係」を示している．また，係数ベクトル $\boldsymbol{h}_1 = (1\ 2\ 0\ 0\ 1)^{\mathrm{T}}$ 及び $\boldsymbol{h}_2 = (0\ 0\ 1\ 2\ 1)^{\mathrm{T}}$ は，顧客 A から顧客 E のそれぞれがどの潜在パターンに当てはまるかを数値的に表している．とくに，顧客 E に関しては，潜在パターン $\boldsymbol{w}_1$ 及び $\boldsymbol{w}_2$ の両方の特性を持つ購入の仕方をしているが，そのような「潜在パターンの共起現象」も係数ベクトルが表現している．このような $\boldsymbol{X}$ 中の潜在パターンは，各商品の今後のマーケティング戦略に大いに役立てることができる．

また別の例として，図 1(c) の音響データ（振幅時間周波数行列）を観測非負値行列 $\boldsymbol{X}$ と考え，$\boldsymbol{X}$ から 2 本の基底を抽出するために $K = 2$ とおいて NMF で分解した結果を図 14 示す．この音響信号は，前半で倍音成分を含む音が生じており，

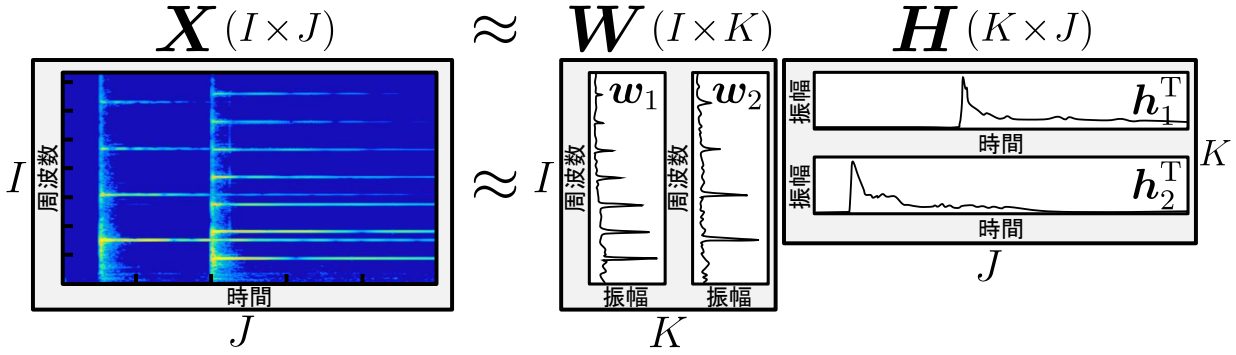

図 14　音響信号の振幅時間周波数行列を NMF で分解した結果例.

その音が鳴りやまぬ内に別の音高の音が重なるように生じている．NMF による分解では，それぞれの音の振幅スペクトルが基底ベクトル $\boldsymbol{w}_1$ 及び $\boldsymbol{w}_2$ として現れており，これは $\boldsymbol{X}$ 中の音色が潜在パターンとして抽出されていることを示している．また，また係数ベクトル $\boldsymbol{h}_1$ 及び $\boldsymbol{h}_2$ には，各音色がどの時間にどの程度の強度で生じたかを表しており，$\boldsymbol{H}$ は楽譜のような情報を示していることになる．

以上のように，NMF の本質は「低ランク近似分解による潜在パターン及びその係数の抽出」である．これまで $\boldsymbol{w}_k$ 及び $\boldsymbol{h}_k$ をそれぞれ基底ベクトル及び係数ベクトルと呼んだが，これに倣って行列 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ をそれぞれ基底行列（basis matrix）及び係数行列（coefficient matrix）と呼ぶ[*19]．注意すべき点は，上記に示した式 (3.17) 等の分解結果は単なる一例であり，NMF 分解で得られる $\boldsymbol{W}$ と $\boldsymbol{H}$ には原理的に次の 2 つの任意性が存在する．

- 順序の任意性（permutation ambiguity）
  $\boldsymbol{W}$ に含まれる K 本の基底ベクトル $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots \boldsymbol{w}_K$ の順序を入れ替えても，$\boldsymbol{H}$ に含まれる係数ベクトル $\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots \boldsymbol{h}_K$ を同様の順序に入れ替えれば，$\boldsymbol{W}\boldsymbol{H}$ は一致する．

$$\text{例：}\ \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{h}_1^{\mathrm{T}} \\ \boldsymbol{h}_2^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{w}_2 & \boldsymbol{w}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{h}_2^{\mathrm{T}} \\ \boldsymbol{h}_1^{\mathrm{T}} \end{pmatrix}$$

- 大きさの任意性（scale ambiguity）
  $\boldsymbol{W}$ に含まれる K 本の基底ベクトル $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots \boldsymbol{w}_K$ のそれぞれに対して $C_1, C_2, \cdots, C_K$ というスカラーを乗じて

[*19] 係数行列 $\boldsymbol{H}$ はアクティベーション行列（activation matrix）と呼ばれることも多い.

も，$\boldsymbol{H}$ に含まれる係数ベクトル $\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots \boldsymbol{h}_K$ のそれぞれに対して $C_1^{-1}, C_2^{-1}, \cdots, C_K^{-1}$ を乗じれば，$\boldsymbol{WH}$ は一致する．

$$\text{例：}\ \begin{pmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{h}_1^{\mathrm{T}} \\ \boldsymbol{h}_2^{\mathrm{T}}\end{pmatrix} = \begin{pmatrix}C_1\boldsymbol{w}_1 & C_2\boldsymbol{w}_2\end{pmatrix}\begin{pmatrix}C_1^{-1}\boldsymbol{h}_1^{\mathrm{T}} \\ C_2^{-1}\boldsymbol{h}_2^{\mathrm{T}}\end{pmatrix}$$

また，$\boldsymbol{W}$ 及び $\boldsymbol{H}$ の推定問題は解析的に解けるわけではなく，次節で説明する最適化問題を解く必要がある．さらに，定義する最適化問題によっては目的関数が凸関数にならないため，局所解の存在により必ずしも常に同じ分解結果が得られるわけではなくなる．

3.2 NMF の定式化と類似度関数

NMF で求めるべき変数は基底行列 $\boldsymbol{W}$ 及び係数行列 $\boldsymbol{H}$ の 2 つである．これは $IK + KJ$ 個の非負値要素 $w_{ik}, h_{kj}\ \forall i = 1, 2, \cdots, I, j = 1, 2, \cdots, J, k = 1, 2, \cdots, K$ を全て求めることに対応する．NMF の目的は $\boldsymbol{X}$ を $\boldsymbol{WH}$ で近似することであることから，式 (3.12) のような近似誤差を定義し，これを最小化する $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を求める最適化問題として定式化できる．ここでは，式 (3.12) に示す二乗 Euclid 距離だけでなく，一般的な類似度関数を用いて，NMF を次式で定式化する．

$$\min_{\boldsymbol{W},\boldsymbol{H}} \mathcal{D}(\boldsymbol{X}|\boldsymbol{WH})\ \ \text{s.t. } w_{ik}, h_{kj} \geq 0\ \forall i = 1, 2, \cdots, I, j = 1, 2, \cdots, J, k = 1, 2, \cdots, K \tag{3.19}$$

ここで，同サイズの行列 $\boldsymbol{A} \in \mathbb{R}^{I\times J}$ 及び $\boldsymbol{B} \in \mathbb{R}^{I\times J}$ を定義したとき，$\mathcal{D}(\boldsymbol{A}|\boldsymbol{B})$ は $\boldsymbol{A}$ と $\boldsymbol{B}$ の類似度を測る関数である．

用いる類似度関数応じて基底行列 $\boldsymbol{W}$ 及び係数行列 $\boldsymbol{H}$ の推定結果は変化する．特に，以下の 3 種類の類似度関数を用いる NMF が有名である．

- 二乗 Euclid 距離（squared Euclidean distance）

$$\begin{aligned}\mathcal{D}_{\mathrm{Eu}}(\boldsymbol{A}|\boldsymbol{B}) &= \|\boldsymbol{A} - \boldsymbol{B}\|_{\mathrm{Fr}}^2 \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}(a_{ij} - b_{ij})^2\end{aligned} \tag{3.20}$$

- 一般化 Kullback–Leibler ダイバージェンス（generalized Kullback–Leibler divergence）

$$\mathcal{D}_{\mathrm{KL}}(\boldsymbol{A}|\boldsymbol{B}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\left[a_{ij}\log\frac{a_{ij}}{b_{ij}} - (a_{ij} - b_{ij})\right] \tag{3.21}$$

- Itakura–Saito ダイバージェンス（Itakura–Saito divergence）

$$\mathcal{D}_{\mathrm{IS}}(\boldsymbol{A}|\boldsymbol{B}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{a_{ij}}{b_{ij}} - \log\frac{a_{ij}}{b_{ij}} - 1\right) \tag{3.22}$$

ここで，a_{ij} 及び b_{ij} はそれぞれ $\boldsymbol{A}$ 及び $\boldsymbol{B}$ の要素を表す．スカラーに対する二乗 Euclid 距離は，次に示す距離の公理を満たすため距離（distance）と呼ぶ．

1. 非負性：$\mathcal{D}(a|b) \geq 0\ \ \forall a, b \in \mathbb{R}$
2. 同一性：$\mathcal{D}(a|b) = 0 \Leftrightarrow a = b$
3. 対称性：$\mathcal{D}(a|b) = \mathcal{D}(b|a)\ \ \forall a, b \in \mathbb{R}$
4. 三角不等式：$\mathcal{D}(a|b) + \mathcal{D}(b|c) \geq \mathcal{D}(a|c)\ \ \forall a, b, c \in \mathbb{R}$

一方，スカラーに対する一般化 Kullback–Leibler ダイバージェンス及び Itakura–Saito ダイバージェンスは，いずれも上記の距離の公理の内対称性と三角不等式を満たさないため，擬距離あるいはダイバージェンス（divergence）と呼ぶ．図 15 にこの 3 種類の類似度関数を示す．ダイバージェンスは対称性を満たさないため，左右非対称な類似度関数となっている．一般化 Kullback–Leibler ダイバージェンスや Itakura–Saito ダイバージェンスの解釈としては，$a = 5$ という観測値に対して，推定値が $b > 5$ の場合には誤差として許容しやすいが，逆に推定値が $b < 5$ となる場合には大きなペナルティが課せられ

るような特性と捉えることができる．特に音響信号では，人間の聴覚特性が振幅スペクトルのピーク値に敏感であることから，NMF の基底ベクトルが確実にピークを表現するように，一般化 Kullback–Leibler ダイバージェンスや Itakura–Saito ダイバージェンスが頻繁に用いられる．

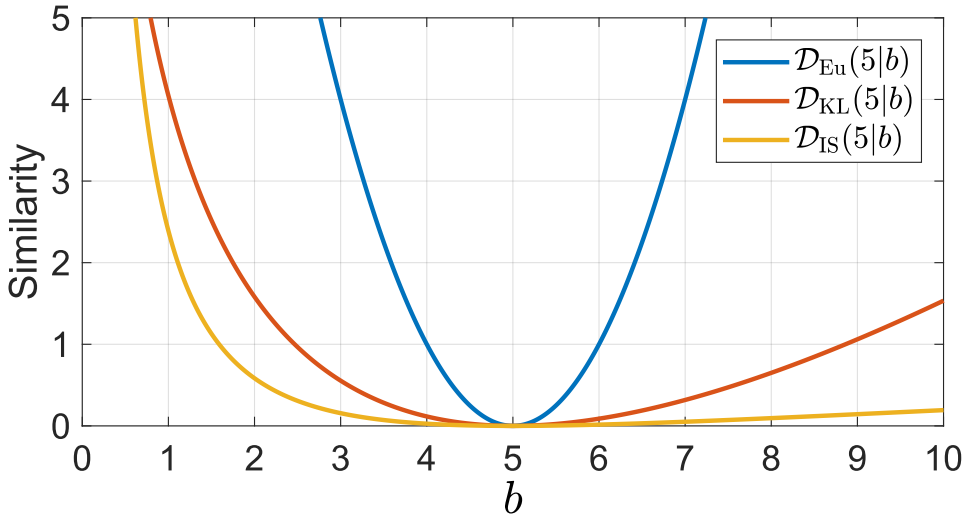

図 15　二乗 Euclid 距離，一般化 Kullback–Leibler ダイバージェンス，及び Itakura–Saito ダイバージェンスの概形．

また，上記の 3 種類の類似度関数を統一的に一般化して表現できる類似度関数として，次式で与えられる β ダイバージェンスがある [10][*20]．

$$\mathcal{D}_\beta(\boldsymbol{A}|\boldsymbol{B}) = \sum_{i=1}^{I}\sum_{j=1}^{J} d_\beta(a_{ij}|b_{ij}) \tag{3.23}$$

$$d_\beta(a_{ij}|b_{ij}) = \begin{cases} \frac{a_{ij}^\beta}{\beta(\beta-1)} + \frac{b_{ij}^\beta}{\beta} - \frac{a_{ij}b_{ij}^{\beta-1}}{\beta-1} & (\beta \in \mathbb{R}_{\neq 0,1}) \\ a_{ij}\log\frac{a_{ij}}{b_{ij}} - (a_{ij}-b_{ij}) & (\beta = 1) \\ \frac{a_{ij}}{b_{ij}} - \log\frac{a_{ij}}{b_{ij}} - 1 & (\beta = 0) \end{cases} \tag{3.24}$$

上式を見ると，$\beta = 2$，$\beta = 1$，及び $\beta = 0$ のときにそれぞれ $(1/2)\mathcal{D}_{\mathrm{Eu}}(\boldsymbol{A}|\boldsymbol{B})$，$\mathcal{D}_{\mathrm{KL}}(\boldsymbol{A}|\boldsymbol{B})$，及び $\mathcal{D}_{\mathrm{IS}}(\boldsymbol{A}|\boldsymbol{B})$ に一致することが分かる．

3.3　NMF の反復更新則の導出

以下では，二乗 Euclid 距離，一般化 Kullback–Leibler ダイバージェンス，及び Itakura–Saito ダイバージェンスを類似度関数に用いた時の 3 種類の NMF について，それらの最適化問題の解法を与える反復更新則の導出を示す．なお，途中で用いる行列演算における諸性質や行列微分等については文献 [11] を参照されたい．

3.3.1　二乗 Euclid 距離に基づく NMF の場合

二乗 Euclid 距離に基づく NMF（以下，Eu-NMF と表記する）の最適化問題は式 (3.19) 及び (3.20) より，次式のようになる．

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_{\mathrm{Fr}}^2 \ \text{ s.t. } w_{ik}, h_{kj} \geq 0 \ \forall i = 1, 2, \cdots, I, j = 1, 2, \cdots, J, k = 1, 2, \cdots, K \tag{3.25}$$

[*20] 但し，文献 [10] で示されている β ダイバージェンスの定義は，ここで示す式 (3.24) と β の値が 1 だけずれている．NMF で用いられる β ダイバージェンスは式 (3.24) がよく用いられる．

また，式 (3.25) は次式のようにスカラー形式でも表現できる．

$$\min_{\boldsymbol{W},\boldsymbol{H}} \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij}-\sum_{k=1}^{K} w_{ik}h_{kj}\right)^2 \quad \text{s.t. } w_{ik},h_{kj}\geq 0\ \forall i=1,2,\cdots,I, j=1,2,\cdots,J, k=1,2,\cdots,K \tag{3.26}$$

まず，行列形式で表された最適化問題 (3.25) に対して，不等式制約条件付き最適化問題の解法（KKT 条件）を適用して反復更新則を導出する．式 (3.25) の目的関数を，式 (3.13) を用いて次のように変形する．

$$\begin{aligned}
\mathcal{J}(\boldsymbol{W},\boldsymbol{H}) &= \|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\|_{\mathrm{Fr}}^2 \\
&= \mathrm{tr}\left[(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H})(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H})^{\mathrm{T}}\right] \\
&= \mathrm{tr}\left[(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H})(\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}})\right] \\
&= \mathrm{tr}\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}-\boldsymbol{W}\boldsymbol{H}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\right] \\
&= \mathrm{tr}\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}-(\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}})^{\mathrm{T}}+\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\right] \\
&= \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})-2\mathrm{tr}(\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}})
\end{aligned} \tag{3.27}$$

ここで，5 行目から 6 行目の変形には跡の性質 $\mathrm{tr}(\boldsymbol{Q}_1\boldsymbol{Q}_2)=\mathrm{tr}(\boldsymbol{Q}_2)+\mathrm{tr}(\boldsymbol{Q_2})$ 及び $\mathrm{tr}(\boldsymbol{Q})=\mathrm{tr}(\boldsymbol{Q}^{\mathrm{T}})$ を用いている（$\boldsymbol{Q}_1$，$\boldsymbol{Q}_2$，及び $\boldsymbol{Q}$ は正方行列である）．不等式制約条件 $w_{ik},h_{kj}\geq 0$ と目的関数 (3.27) より，Lagrange 関数は次のように構成できる．

$$\mathcal{L}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Gamma_W},\boldsymbol{\Gamma_H})=\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})-2\mathrm{tr}(\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}})-\mathrm{tr}(\boldsymbol{\Gamma}^{(\boldsymbol{W})}\boldsymbol{W}^{\mathrm{T}})-\mathrm{tr}(\boldsymbol{\Gamma}^{(\boldsymbol{H})}\boldsymbol{H}^{\mathrm{T}}) \tag{3.28}$$

ここで，$\boldsymbol{\Gamma}^{(\boldsymbol{W})}\in \mathrm{R}^{I\times K}$ 及び $\boldsymbol{\Gamma}^{(\boldsymbol{H})}\in \mathrm{R}^{I\times K}$ はそれぞれ不等式条件 $w_{ik}\geq 0$ 及び $h_{kj}\geq 0$ を表すための未定係数行列であり，KKT 条件より以下を満たす[*21]．

$$\gamma_{ik}^{(\boldsymbol{W})}\geq 0 \tag{3.29}$$

$$\gamma_{kj}^{(\boldsymbol{H})}\geq 0 \tag{3.30}$$

$$\gamma_{ik}^{(\boldsymbol{W})}w_{ik}=0\ \ \forall i=1,2,\cdots,I, k=1,2,\cdots,K \tag{3.31}$$

$$\gamma_{kj}^{(\boldsymbol{H})}h_{kj}=0\ \ \forall j=1,2,\cdots,J, k=1,2,\cdots,K \tag{3.32}$$

また，$\gamma_{ik}^{(\boldsymbol{W})}$ 及び $\gamma_{kj}^{(\boldsymbol{H})}$ はそれぞれ $\boldsymbol{\Gamma}^{(\boldsymbol{W})}$ 及び $\boldsymbol{\Gamma}^{(\boldsymbol{H})}$ の要素である．特に，式 (3.31) 及び (3.32) は KKT 条件における相補性条件（complementary slackness）と呼ばれる．この Lagrange 関数の停留点を求める為に，式 (3.28) を行列 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ で偏微分することを考える．行列での偏微分には次の性質 [11] を用いる．

$$\frac{\partial \mathrm{tr}\left[F(\boldsymbol{Y})\right]}{\partial \boldsymbol{Y}}=f(\boldsymbol{Y})^{\mathrm{T}} \tag{3.33}$$

ここで，$\boldsymbol{Y}$ は実数一般行列，$F(\boldsymbol{Y})$ は $\boldsymbol{Y}$ の要素のそれぞれに対して微分可能な関数，$f(\cdot)$ は $F(\cdot)$ をスカラーで微分したときに得られる導関数である．式 (3.33) を用いると，$F(\cdot)$ が変数に対して 1 次である場合の微分は次のように求めることができる．

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}\right]=\boldsymbol{I} \tag{3.34}$$

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}\boldsymbol{A}\right]=\boldsymbol{A}^{\mathrm{T}} \tag{3.35}$$

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}\right]=\boldsymbol{A}^{\mathrm{T}} \tag{3.36}$$

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\right]=\boldsymbol{A} \tag{3.37}$$

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}^{\mathrm{T}}\right]=\boldsymbol{A} \tag{3.38}$$

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}\boldsymbol{B}\right]=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}} \tag{3.39}$$

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{B}\right]=\boldsymbol{B}\boldsymbol{A} \tag{3.40}$$

[*21] 式 (3.29) 及び (3.30) において未定乗数の符号が式 (2.50) と逆になっているのは，不等式条件の不等号の向きが逆であることに起因する．

ここで，$\boldsymbol{A}$ 及び $\boldsymbol{B}$ は適当なサイズの定数行列である．また，$F(\cdot)$ が変数に対して 2 次である場合の微分は次のようになる．

$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}^2\right] = 2\boldsymbol{Y}^{\mathrm{T}} \tag{3.41}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right] = 2\boldsymbol{Y} \tag{3.42}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\right] = 2\boldsymbol{Y} \tag{3.43}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}^2\boldsymbol{A}\right] = (\boldsymbol{Y}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{Y})^{\mathrm{T}} \tag{3.44}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Y}\right] = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y} \tag{3.45}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\right] = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y} \tag{3.46}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\right] = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y} \tag{3.47}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}^{\mathrm{T}}\right] = \boldsymbol{Y}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Y}\boldsymbol{A} \tag{3.48}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right] = \boldsymbol{Y}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Y}\boldsymbol{A} \tag{3.49}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{A}\right] = \boldsymbol{Y}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Y}\boldsymbol{A} \tag{3.50}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}\boldsymbol{B}\boldsymbol{Y}\right] = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \tag{3.51}$$
$$\frac{1}{\partial \boldsymbol{Y}}\partial \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{B}\right] = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y} + \boldsymbol{B}\boldsymbol{A}\boldsymbol{Y} \tag{3.52}$$

これらの性質を用いて Lagrange 関数 (3.28) の偏微分を求める．まず，$\boldsymbol{W}$ による偏微分は，式 (3.38) 及び (3.48) を用いると

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}} &= -2\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{W}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{\mathrm{T}} + \boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{\Gamma}^{(\boldsymbol{W})} \\ &= -2\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} + 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{\Gamma}^{(\boldsymbol{W})}\end{aligned} \tag{3.53}$$

が得られる．従って，Lagrange 関数 (3.28) の $\boldsymbol{W}$ に関する停留点は次式で与えられる．

$$-2\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} + 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{\Gamma}^{(\boldsymbol{W})} = \boldsymbol{0} \tag{3.54}$$

式 (3.54) の ik 成分にのみ着目すると，次のような変形ができる．

$$\begin{aligned}\left[-2\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} + 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{\Gamma}^{(\boldsymbol{W})}\right]_{ik} &= 0 \\ \left[-2\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} + 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{\Gamma}^{(\boldsymbol{W})}\right]_{ik} w_{ik} &= 0\end{aligned} \tag{3.55}$$

ここで，$[\cdot]_{ij}$ は行列の ij 成分を表す．相補性条件 (3.31) より $\gamma_{ik}^{(\boldsymbol{W})} w_{ik} = 0$ であるので，更に変形して以下を得る．

$$\left[-2\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} + 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right]_{ik} w_{ik} = 0 \tag{3.56}$$
$$\left[-\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\right]_{ik} w_{ik} + \left[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right]_{ik} w_{ik} = 0 \tag{3.57}$$
$$\left[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right]_{ik} w_{ik} = \left[\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\right]_{ik} w_{ik} \tag{3.58}$$

式 (3.58) は Eu-NMF の $\boldsymbol{W}$ に関する 1 次最適性条件である．即ち，$\boldsymbol{W}$ の最小解では，全ての i 及び k に関して $\left[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right]_{ik} = \left[\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\right]_{ik}$ または $w_{ik} = 0$ のいずれかが満たされる．

同様に，Lagrange 関数の係数行列 $\boldsymbol{H}$ に対する停留点を求める為に，式 (3.28) を行列 $\boldsymbol{H}$ で偏微分する．

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{H}} &= -2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H} + \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H} - \boldsymbol{\Gamma}^{(\boldsymbol{H})} \\ &= -2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X} + 2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H} - \boldsymbol{\Gamma}^{(\boldsymbol{H})}\end{aligned} \tag{3.59}$$

従って，Lagrange 関数 (3.28) の $\boldsymbol{H}$ に関する停留点は次式で与えられる．

$$-2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}+2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}-\boldsymbol{\Gamma}^{(\boldsymbol{H})}=\boldsymbol{0} \tag{3.60}$$

式 (3.60) の kj 成分にのみ着目すると，次のような変形ができる．

$$\begin{aligned}\left[-2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}+2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}-\boldsymbol{\Gamma}^{(\boldsymbol{H})}\right]_{kj}&=0\\ \left[-2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}+2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}-\boldsymbol{\Gamma}^{(\boldsymbol{H})}\right]_{kj}h_{kj}&=0\end{aligned} \tag{3.61}$$

相補性条件 (3.32) より $\gamma_{kj}^{(\boldsymbol{H})}h_{kj}=0$ であるので，更に変形して以下を得る．

$$\left[-2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}+2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}\right]_{kj}h_{kj}=0 \tag{3.62}$$

$$\left[-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\right]_{kj}h_{kj}+\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}\right]_{kj}h_{kj}=0 \tag{3.63}$$

$$\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}\right]_{kj}h_{kj}=\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\right]_{kj}h_{kj} \tag{3.64}$$

式 (3.64) は Eu-NMF の $\boldsymbol{H}$ に関する 1 次最適性条件である．即ち，$\boldsymbol{H}$ の最小解では，全ての k 及び j に関して $\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}\right]_{kj}=\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\right]_{kj}$ または $h_{kj}=0$ のいずれかが満たされる．

以上より，Eu-NMF の 1 次最適性条件が示された．

Eu-NMF の 1 次最適性条件

非負行列 $\boldsymbol{X}\in\mathbb{R}_{\geq 0}^{I\times J}$，$\boldsymbol{W}\in\mathbb{R}_{\geq 0}^{I\times K}$，及び $\boldsymbol{H}\in\mathbb{R}_{\geq 0}^{K\times J}$ において，Eu-NMF は次式で定式化される．

$$\min_{\boldsymbol{W},\boldsymbol{H}}\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\|_{\mathrm{Fr}}^{2}\ \text{ s.t. } w_{ik},h_{kj}\geq 0\ \forall i=1,2,\cdots,I,j=1,2,\cdots,J,k=1,2,\cdots,K \tag{3.65}$$

ここで，$[\boldsymbol{W}]_{ik}=w_{ik}$ 及び $[\boldsymbol{H}]_{kj}=h_{kj}$ である．この最小化問題の 1 次最適性条件は

$$\left[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right]_{ik}w_{ik}=\left[\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}\right]_{ik}w_{ik}\ \ \forall i=1,2,\cdots,I,j=1,2,\cdots,J \tag{3.66}$$

$$\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}\right]_{kj}h_{kj}=\left[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\right]_{kj}h_{kj}\ \ \forall i=1,2,\cdots,I,j=1,2,\cdots,J \tag{3.67}$$

である．

また，式 (3.57) 及び (3.63) から，Eu-NMF の最急降下法に基づく反復更新則を構築できる．

Eu-NMF の最急降下法に基づく反復更新則

Eu-NMF の最適化変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ は，次式を繰り返すことで最適化できる．

1. 適当な非負の初期値 $[\boldsymbol{W}^{(0)}]_{ik} = w_{ik}^{(0)}$ 及び $[\boldsymbol{H}^{(0)}]_{kj} = h_{kj}^{(0)}$ を全ての $i = 1, 2, \cdots, I$，$j = 1, 2, \cdots, J$，及び $k = 1, 2, \cdots, K$ についてとる
2. 以下の操作を $t = 0, 1, 2, \cdots$ について繰り返す
   （a）次式から $w_{ik}^{(t+1)}$ を求める

$$w_{ik}^{(t+1)} = w_{ik}^{(t)} + \mu_w \left[\sum_{j=1}^{J} x_{ij} h_{kj} - \sum_{j=1}^{J} \left(\sum_{k=1}^{K} w_{ik} h_{kj} \right) h_{kj} \right] w_{ik}^{(t)} \quad \forall i = 1, 2, \cdots, I, k = 1, 2, \cdots, K \tag{3.68}$$

   （b）次式から $h_{kj}^{(t+1)}$ を求める

$$h_{kj}^{(t+1)} = h_{kj}^{(t)} + \mu_h \left[\sum_{i=1}^{I} w_{ik} x_{ij} - \sum_{i=1}^{I} w_{ik} \left(\sum_{k=1}^{K} w_{ik} h_{kj} \right) \right] h_{kj}^{(t)} \quad \forall k = 1, 2, \cdots, K, j = 1, 2, \cdots, J \tag{3.69}$$

ここで，μ_w 及び μ_h はステップサイズパラメータである．なお，式 (3.68) 及び (3.69) は，次のように行列形式で表現することもできる．

$$\boldsymbol{W}^{(t+1)} = \boldsymbol{W}^{(t)} + \mu_w \left(\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} \right) \otimes \boldsymbol{W}^{(t)} \tag{3.70}$$

$$\boldsymbol{H}^{(t+1)} = \boldsymbol{H}^{(t)} + \mu_h \left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H} \right) \otimes \boldsymbol{H}^{(t)} \tag{3.71}$$

ここで，行列に対する演算子 $\otimes$ は要素毎の積（Hadamard 積）を表す．

反復更新則 (3.70) 及び (3.71) は，不等式制約条件付き最適化問題 (3.25) を，Lagrange 関数及び KKT 条件を用いて解くことで得られる Eu-NMF の最適化式である．一方で，MM アルゴリズムを用いて最急降下法よりも効率的な反復更新則を導出できることを，以下に示す [7]．まず，式 (3.26) の目的関数を次のように変形する．

$$\begin{aligned} \mathcal{J}(\boldsymbol{W}, \boldsymbol{H}) &= \sum_{i=1}^{I} \sum_{j=1}^{J} \left(x_{ij} - \sum_{k=1}^{K} w_{ik} h_{kj} \right)^2 \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} \left[x_{ij}^2 - 2 x_{ij} \sum_{k=1}^{K} w_{ik} h_{kj} + \left(\sum_{k=1}^{K} w_{ik} h_{kj} \right)^2 \right] \end{aligned} \tag{3.72}$$

式 (3.72) の目的関数の w_{ik} や h_{kj} に関する偏微分を計算する場合，第 3 項の k に関する総和を含む 2 次関数の展開を考えると煩雑である．そこで，次に示す Jensen の不等式 (2.80) を適用し，式 (3.72) を上から抑える補助関数を設計することを考える．

$$\begin{aligned} \left(\sum_{k=1}^{K} w_{ik} h_{kj} \right)^2 &= \left(\sum_{k=1}^{K} \delta_{ijk} \frac{w_{ik} h_{kj}}{\delta_{ijk}} \right)^2 \\ &\leq \sum_{k=1}^{K} \delta_{ijk} \left(\frac{w_{ik} h_{kj}}{\delta_{ijk}} \right)^2 \\ &= \sum_{k=1}^{K} \frac{w_{ik}^2 h_{kj}^2}{\delta_{ijk}} \end{aligned} \tag{3.73}$$

ここで，$\delta_{ijk} > 0$ は補助変数であり，$\sum_{k=1}^{K} \delta_{ijk} = 1$ を満たす．式 (3.73) は，Jensen の不等式 (2.80) において，凸関数 $f : \mathbb{R} \to \mathbb{R}$ を $f(\cdot) = (\cdot)^2$ として適用した不等式である．この不等式を式 (3.72) の第 3 項に適用することで，補助関数を次

式のように定義できる．

$$\begin{aligned}\mathcal{J}(\boldsymbol{W},\boldsymbol{H}) &\leq \mathcal{J}^{+}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta}) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij}^2 - 2x_{ij}\sum_{k=1}^{K} w_{ik}h_{kj} + \sum_{k=1}^{K}\frac{w_{ik}^2 h_{kj}^2}{\delta_{ijk}}\right)\end{aligned} \tag{3.74}$$

ここで，$\boldsymbol{\Delta} \in \mathbb{R}_{>0}^{I\times J\times K}$ は補助変数 δ_{ijk} を要素として含む 3 階のテンソルである．

次に，MM アルゴリズムの 1 つ目の操作 (2.77) として，設計した補助関数 $\mathcal{J}^{+}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta})$ を補助変数 $\boldsymbol{\Delta}$ について最小化する．いま，補助変数には等式制約条件 $\sum_{k=1}^{K}\delta_{ijk} = 1$ が課せられているため，次の等式制約条件付き最適化問題を考える必要がある．

$$\min_{\boldsymbol{\Delta}} \mathcal{J}^{+}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta}) \ \text{ s.t. } \sum_{k=1}^{K}\delta_{ijk} = 1 \ \ \forall i = 1,2,\cdots,I, j = 1,2,\cdots,J \tag{3.75}$$

最適化問題 (3.75) の Lagrange 関数は

$$\mathcal{L}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta},\lambda) = \mathcal{J}^{+} - \lambda\left(\sum_{k=1}^{K}\delta_{ijk} - 1\right) \tag{3.76}$$

で与えられるため，δ_{ijk} 及び λ の偏微分をそれぞれ 0 とおくと

$$\frac{\partial\mathcal{L}}{\partial\delta_{ijk}} = -\frac{w_{ik}^2 h_{kj}^2}{\delta_{ijk}^2} - \lambda = 0 \tag{3.77}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda} = -\sum_{k=1}^{K}\delta_{ijk} + 1 = 0 \tag{3.78}$$

が得られる．式 (3.78) は等式制約条件 $\sum_{k=1}^{K}\delta_{ijk} = 1$ そのものである．また，式 (3.77)，$\delta_{ijk} > 0$，及び $w_{ik}, h_{kj} \geq 0$ より

$$\begin{aligned}\delta_{ijk} &= \sqrt{-\frac{w_{ik}^2 h_{kj}^2}{\lambda}} \\ &= \frac{w_{ik}h_{kj}}{\sqrt{-\lambda}}\end{aligned} \tag{3.79}$$

が得られる．式 (3.79) の両辺に対して k について総和をとると

$$\sum_{k=1}^{K}\delta_{ijk} = \sum_{k=1}^{K}\frac{w_{ik}h_{kj}}{\sqrt{-\lambda}} \tag{3.80}$$

となり，等式制約条件 $\sum_{k=1}^{K}\delta_{ijk} = 1$ より

$$\sum_{k=1}^{K}\frac{w_{ik}h_{kj}}{\sqrt{-\lambda}} = 1 \tag{3.81}$$

が得られる．未定乗数 λ は k に依らない定数であるため，結局

$$\sqrt{-\lambda} = \sum_{k=1}^{K} w_{ik}h_{kj} \tag{3.82}$$

となり，これを式 (3.79) に代入することで，補助変数の最小解が次式として得られる．

$$\delta_{ijk} = \frac{w_{ik}h_{kj}}{\sum_{k'=1}^{K} w_{ik'}h_{k'j}} \tag{3.83}$$

ここで，式 (3.83) の左辺及び分子の k と分母の k は互いに無関係であるため，これらを区別するために分母のインデクス k を k' と記述している点に注意する．従って，式 (3.83) が補助関数 (3.75) の補助変数に関する最小解を与える．反復回数 t を考慮して記述すると，次のようになる．

$$\delta_{ijk}^{(t+1)} = \frac{w_{ik}^{(t)} h_{kj}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} \tag{3.84}$$

なお，最小解の式 (3.83) は不等式 (3.73) の等号成立条件に他ならない．Jensen の不等式における等号成立条件は，式 (2.80) における $\theta_1 = \theta_2 = \cdots = \theta_N$ で与えられることから，Lagrange 関数 (3.76) を定義して偏微分を計算しなくとも，直ちに

$$\frac{w_{i1}h_{1j}}{\delta_{ij1}} = \frac{w_{i2}h_{2j}}{\delta_{ij2}} = \cdots = \frac{w_{iK}h_{Kj}}{\delta_{ijK}} = \text{const.} \tag{3.85}$$

という条件式が導かれ，式 (3.79) の右辺の $\sqrt{-\lambda}$ を式 (3.85) の右辺の定数 const. とした式が得られることから，補助変数の最小解 (3.83) をより簡単に導くことができる．

続いて，MM アルゴリズムの 2 つ目の操作 (2.78) として，設計した補助関数 $\mathcal{J}^+(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\Delta})$ を本来の変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ について最小化する．まず，補助関数 (3.74) を w_{ik} で偏微分して 0 とおくと，次式が得られる[*22]．

$$\frac{\partial \mathcal{J}^+}{\partial w_{ik}} = \sum_{j=1}^{J} \left(-2x_{ij}h_{kj} + 2\frac{w_{ik}h_{kj}^2}{\delta_{ijk}} \right) = 0 \tag{3.86}$$

式 (3.86) を整理すると次式のようになる[*23]．

$$\begin{aligned} -\sum_{j=1}^{J} x_{ij}h_{kj} + \sum_{j=1}^{J} \frac{w_{ik}h_{kj}^2}{\delta_{ijk}} &= 0 \\ w_{ik}\sum_{j=1}^{J} \frac{h_{kj}^2}{\delta_{ijk}} &= \sum_{j=1}^{J} x_{ij}h_{kj} \\ w_{ik} &= \frac{\sum_{j=1}^{J} x_{ij}h_{kj}}{\sum_{j=1}^{J} \frac{h_{kj}^2}{\delta_{ijk}}} \end{aligned} \tag{3.87}$$

式 (3.87) が補助関数 (3.74) を最小化する w_{ik} である．反復回数 t を考慮して記述すると，次のようになる．

$$w_{ik}^{(t+1)} = \frac{\sum_{j=1}^{J} x_{ij}h_{kj}^{(t)}}{\sum_{j=1}^{J} \frac{{h_{kj}^{(t)}}^2}{\delta_{ijk}^{(t+1)}}} \tag{3.88}$$

同様に，補助関数 (3.74) を h_{kj} で偏微分して 0 とおくと，h_{kj} の最小解も次のように得られる．

$$\begin{aligned} \frac{\partial \mathcal{J}^+}{\partial h_{kj}} = \sum_{i=1}^{I} \left(-2x_{ij}w_{ik} + 2\frac{w_{ik}^2 h_{kj}}{\delta_{ijk}} \right) &= 0 \\ -\sum_{i=1}^{I} x_{ij}w_{ik} + \sum_{i=1}^{I} \frac{w_{ik}^2 h_{kj}}{\delta_{ijk}} &= 0 \\ h_{kj}\sum_{i=1}^{I} \frac{w_{ik}^2}{\delta_{ijk}} &= \sum_{i=1}^{I} x_{ij}w_{ik} \\ h_{kj} &= \frac{\sum_{i=1}^{I} x_{ij}w_{ik}}{\sum_{i=1}^{I} \frac{w_{ik}^2}{\delta_{ijk}}} \end{aligned} \tag{3.89}$$

[*22] 今，w_{ik} という 1 つの要素のみでの偏微分を考えているため，例えば w_{ik+1} のような ik 以外の要素は全て定数として扱われることに注意する．式 (3.86) において $\sum_i$ や $\sum_k$ の総和記号が消えるのはこの為である．

[*23] 今は補助関数を w_{ik} で偏微分しているため，偏微分を 0 とおいた式中の $\boldsymbol{w}_{ik}$ は新しい（更新後の）変数であり，偏微分定数である h_{kj} や δ_{ijk} は古い（更新前の）変数である点に注意する．従って，ここでは新しい変数 w_{ik} について解いた式が w_{ik} の反復更新則となる．

反復回数 t を考慮して記述すると，次のようになる．

$$h_{kj}^{(t+1)} = \frac{\sum_{i=1}^{I} x_{ij} w_{ik}^{(t)}}{\sum_{i=1}^{I} \frac{w_{ik}^{(t)\,2}}{\delta_{ijk}^{(t+1)}}} \tag{3.90}$$

従って，式 (3.84) で補助変数を更新し，式 (3.88) 及び (3.90) で本来の変数を更新する．この 2 つの手順を繰り返すことで，目的関数 (3.72) の値を最小化する $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を求めることができる．ここで，補助変数の反復更新則 (3.84) が解析的に記述できることから，式 (3.84) を式 (3.88) 及び (3.90) にそれぞれ代入することで，次式のような統合された反復更新則に書き換えることができる．

$$\begin{aligned} w_{ik}^{(t+1)} &= \frac{\sum_{j=1}^{J} x_{ij} h_{kj}^{(t)}}{\sum_{j=1}^{J} h_{kj}^{(t)\,2} \frac{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}{w_{ik}^{(t)} h_{kj}^{(t)}}} \\ &= w_{ik}^{(t)} \frac{\sum_{j=1}^{J} x_{ij} h_{kj}^{(t)}}{\sum_{j=1}^{J} h_{kj}^{(t)} \sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} \end{aligned} \tag{3.91}$$

$$\begin{aligned} h_{kj}^{(t+1)} &= \frac{\sum_{i=1}^{I} x_{ij} w_{ik}^{(t)}}{\sum_{i=1}^{I} w_{ik}^{(t)\,2} \frac{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}{w_{ik}^{(t)} h_{kj}^{(t)}}} \\ &= h_{kj}^{(t)} \frac{\sum_{i=1}^{I} x_{ij} w_{ik}^{(t)}}{\sum_{i=1}^{I} w_{ik}^{(t)} \sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} \end{aligned} \tag{3.92}$$

以上より，Eu-NMF の MM アルゴリズムに基づく反復更新則を構築できる．

Eu-NMF の MM アルゴリズムに基づく反復更新則

Eu-NMF の最適化変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ は，次式を繰り返すことで最適化できる．

1. 適当な非負の初期値 $[\boldsymbol{W}^{(0)}]_{ik} = w_{ik}^{(0)}$ 及び $[\boldsymbol{H}^{(0)}]_{kj} = h_{kj}^{(0)}$ を全ての $i = 1, 2, \cdots, I$，$j = 1, 2, \cdots, J$，及び $k = 1, 2, \cdots, K$ についてとる
2. 以下の操作を $t = 0, 1, 2, \cdots$ について繰り返す
   （a）次式から $w_{ik}^{(t+1)}$ を求める

$$w_{ik}^{(t+1)} = w_{ik}^{(t)} \frac{\sum_{j=1}^{J} x_{ij} h_{kj}^{(t)}}{\sum_{j=1}^{J} h_{kj}^{(t)} \sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} \quad \forall i = 1, 2, \cdots, I, k = 1, 2, \cdots, K \tag{3.93}$$

   （b）次式から $h_{kj}^{(t+1)}$ を求める

$$h_{kj}^{(t+1)} = h_{kj}^{(t)} \frac{\sum_{i=1}^{I} x_{ij} w_{ik}^{(t)}}{\sum_{i=1}^{I} w_{ik}^{(t)} \sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} \quad \forall k = 1, 2, \cdots, K, j = 1, 2, \cdots, J \tag{3.94}$$

なお，式 (3.93) 及び (3.94) は，次のように行列形式で表現することもできる．

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \frac{\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}} \tag{3.95}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}} \tag{3.96}$$

ここで，見やすさのために反復回数 t の標記は省略し，変数更新を表す演算子 $\leftarrow$ を用いている．また，行列に対する演算子 $\otimes$ は要素毎の積（Hadamard 積）を表し，行列の分数は要素毎の商を表す．一般に $K \ll \min(I, J)$ であることから，式 (3.95) 及び (3.96) は次のように計算順序を考慮した方が，計算量が小さくなる．

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \frac{\boldsymbol{X}\boldsymbol{H}^{\mathrm{T}}}{\boldsymbol{W}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})} \tag{3.97}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}}{(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})\boldsymbol{H}} \tag{3.98}$$

MM アルゴリズムによる反復更新則 (3.95) 及び (3.96) は，KKT 条件を用いた手法とは異なり，変数の非負性の制約条件 $w_{ik} \geq 0$ 及び $h_{kj} \geq 0$ を陽に考慮することなく導出されているため，反復更新における非負性の保証に関して疑問が残る．しかしながら，式 (3.95) 及び (3.96) が過去の変数値に対して何らかの値を乗算する更新式，即ち乗算型反復更新則（multiplicative update rule）になっていることを考えると，$w_{ik} \geq 0$ 及び $h_{kj} \geq 0$ を満たす初期値さえ与えられていれば，乗じられる分数項は常に非負であるため，反復更新の途中で w_{ik} 及び h_{kj} が負の値になってしまうことはなく，非負性は保証される．さらに，式 (3.70) 及び (3.71) の再急降下法に基づく反復更新則と比較すると，MM アルゴリズムの反復更新則はステップサイズパラメータを定める必要がなく，毎回の反復で目的関数値が増加することはない単調非増加性（monotonic non-increasing property）という望ましい性質を持っている．

3.3.2 KL-NMF の場合

一般化 Kullback–Leibler ダイバージェンスに基づく NMF（以下，KL-NMF と表記する）の最適化問題は式 (3.19) 及び (3.21) より，次式のようになる．

$$\begin{aligned} &\min_{\boldsymbol{W},\boldsymbol{H}} \sum_{i=1}^{I} \sum_{j=1}^{J} \left[x_{ij} \log \frac{x_{ij}}{\sum_{k=1}^{K} w_{ik} h_{kj}} - \left(x_{ij} - \sum_{k=1}^{K} w_{ik} h_{kj} \right) \right] \\ &\text{s.t. } w_{ik}, h_{kj} \geq 0 \ \forall i = 1, 2, \cdots, I, j = 1, 2, \cdots, J, k = 1, 2, \cdots, K \end{aligned} \tag{3.99}$$

この目的関数を変形すると

$$\begin{aligned}\mathcal{J}(\boldsymbol{W},\boldsymbol{H}) &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left[x_{ij}\log\frac{x_{ij}}{\sum_{k=1}^{K}w_{ik}h_{kj}} - \left(x_{ij}-\sum_{k=1}^{K}w_{ik}h_{kj}\right)\right] \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij}\log x_{ij} - x_{ij}\log\sum_{k=1}^{K}w_{ik}h_{kj} - x_{ij} + \sum_{k=1}^{K}w_{ik}h_{kj}\right)\end{aligned} \tag{3.100}$$

となる．Eu-NMF の場合と同様に w_{ik} や h_{kj} で偏微分を考えた時，式 (3.100) 中で計算が煩雑となる項は第 2 項の k に関する総和を含む負対数関数である．負対数関数は凸関数であることから，この項についても次式の Jensen の不等式 (2.80) を用いることができる．

$$\begin{aligned}-\log\sum_{k=1}^{K}w_{ik}h_{kj} &= -\log\sum_{k=1}^{K}\delta_{ijk}\frac{w_{ik}h_{kj}}{\delta_{ijk}} \\ &\leq -\sum_{k=1}^{K}\delta_{ijk}\log\frac{w_{ik}h_{kj}}{\delta_{ijk}}\end{aligned} \tag{3.101}$$

式 (3.73) と同様に，補助変数 $\delta_{ijk} > 0$ は $\sum_{k=1}^{K}\delta_{ijk} = 1$ を満たす．式 (3.101) は，Jensen の不等式 (2.80) において，凸関数 $f : \mathbb{R} \to \mathbb{R}$ を $f(\cdot) = -\log(\cdot)$ として適用した不等式である．この不等式を式 (3.100) の第 2 項に適用することで，補助関数を次式のように定義できる．

$$\begin{aligned}\mathcal{J}(\boldsymbol{W},\boldsymbol{H}) &\leq \mathcal{J}^{+}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta}) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(x_{ij}\log x_{ij} - x_{ij}\sum_{k=1}^{K}\delta_{ijk}\log\frac{w_{ik}h_{kj}}{\delta_{ijk}} - x_{ij} + \sum_{k=1}^{K}w_{ik}h_{kj}\right)\end{aligned} \tag{3.102}$$

まず，補助関数 $\mathcal{J}^{+}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta})$ に対する補助変数 δ_{ijk} の最小化について考える．Eu-NMF の場合と同様に，δ_{ijk} の最小解は不等式 (3.101) の等号成立条件から与えられるものである．不等式 (3.101) においても，補助変数が式 (3.85) を満たすときに等号が成立することから，同様の議論を経て式 (3.84) として補助変数 δ_{ijk} の反復更新則が得られることが分かる．

次に，補助関数 $\mathcal{J}^{+}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta})$ に対する本来の変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ の最小化について考える．まず，補助関数 (3.102) を w_{ik} で偏微分して 0 とおくと，次式が得られる．

$$\frac{\partial\mathcal{J}^{+}}{\partial w_{ik}} = \sum_{j=1}^{J}\left(-x_{ij}\delta_{ijk}\frac{\delta_{ijk}}{w_{ik}h_{kj}}\frac{h_{kj}}{\delta_{ijk}} + h_{kj}\right) = 0 \tag{3.103}$$

式 (3.103) を整理すると次式のようになる．

$$\begin{aligned}-\sum_{j=1}^{J}x_{ij}\frac{\delta_{ijk}}{w_{ik}} + \sum_{j=1}^{J}h_{kj} &= 0 \\ w_{ik}\sum_{j=1}^{J}h_{kj} &= \sum_{j=1}^{J}x_{ij}\delta_{ijk} \\ w_{ik} &= \frac{\sum_{j=1}^{J}x_{ij}\delta_{ijk}}{\sum_{j=1}^{J}h_{kj}}\end{aligned} \tag{3.104}$$

式 (3.104) が補助関数 (3.102) を最小化する w_{ik} である．反復回数 t を考慮して記述すると，次のようになる．

$$w_{ik}^{(t+1)} = \frac{\sum_{j=1}^{J}x_{ij}\delta_{ijk}^{(t+1)}}{\sum_{j=1}^{J}h_{kj}^{(t)}} \tag{3.105}$$

同様に，補助関数 (3.102) を h_{kj} で偏微分して 0 とおくと，h_{kj} の最小解も次のように得られる．

$$\begin{aligned}
\frac{\partial \mathcal{J}^{+}}{\partial h_{kj}} = \sum_{i=1}^{I} \left(-x_{ij}\delta_{ijk} \frac{\delta_{ijk}}{w_{ik}h_{kj}} \frac{w_{ik}}{\delta_{ijk}} + w_{ik} \right) = 0 \\
-\sum_{i=1}^{I} x_{ij} \frac{\delta_{ijk}}{h_{kj}} + \sum_{i=1}^{I} w_{ik} = 0 \\
h_{kj} \sum_{i=1}^{I} w_{ik} = \sum_{i=1}^{I} x_{ij}\delta_{ijk} \\
h_{kj} = \frac{\sum_{i=1}^{I} x_{ij}\delta_{ijk}}{\sum_{i=1}^{I} w_{ik}}
\end{aligned} \tag{3.106}$$

反復回数 t を考慮して記述すると，次のようになる．

$$h_{kj}^{(t+1)} = \frac{\sum_{i=1}^{I} x_{ij}\delta_{ijk}^{(t+1)}}{\sum_{i=1}^{I} w_{ik}^{(t)}} \tag{3.107}$$

従って，式 (3.84) で補助変数を更新し，式 (3.105) 及び (3.107) で本来の変数を更新する．この 2 つの手順を繰り返すことで，目的関数 (3.100) の値を最小化する $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を求めることができる．Eu-NMF の場合と同様に，補助変数の反復更新則 (3.84) を式 (3.105) 及び (3.107) にそれぞれ代入することで，次式のような統合された反復更新則に書き換えることができる．

$$\begin{aligned}
w_{ik}^{(t+1)} &= \frac{\sum_{j=1}^{J} x_{ij} \frac{w_{ik}^{(t)} h_{kj}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}{\sum_{j=1}^{J} h_{kj}^{(t)}} \\
&= w_{ik}^{(t)} \frac{\sum_{j=1}^{J} \frac{x_{ij}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} h_{kj}^{(t)}}{\sum_{j=1}^{J} h_{kj}^{(t)}}
\end{aligned} \tag{3.108}$$

$$\begin{aligned}
h_{kj}^{(t+1)} &= \frac{\sum_{i=1}^{I} x_{ij} \frac{w_{ik}^{(t)} h_{kj}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}{\sum_{i=1}^{I} w_{ik}^{(t)}} \\
&= h_{kj}^{(t)} \frac{\sum_{i=1}^{I} \frac{x_{ij}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} w_{ik}^{(t)}}{\sum_{j=1}^{J} w_{ik}^{(t)}}
\end{aligned} \tag{3.109}$$

以上より，KL-NMF の MM アルゴリズムに基づく反復更新則を構築できる．

KL-NMF の MM アルゴリズムに基づく反復更新則

KL-NMF の最適化変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ は，次式を繰り返すことで最適化できる．

1. 適当な非負の初期値 $[\boldsymbol{W}^{(0)}]_{ik} = w_{ik}^{(0)}$ 及び $[\boldsymbol{H}^{(0)}]_{kj} = h_{kj}^{(0)}$ を全ての $i = 1, 2, \cdots, I$，$j = 1, 2, \cdots, J$，及び $k = 1, 2, \cdots, K$ についてとる
2. 以下の操作を $t = 0, 1, 2, \cdots$ について繰り返す
   （a）次式から $w_{ik}^{(t+1)}$ を求める

$$w_{ik}^{(t+1)} = w_{ik}^{(t)} \frac{\sum_{j=1}^{J} \frac{x_{ij}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} h_{kj}^{(t)}}{\sum_{j=1}^{J} h_{kj}^{(t)}} \quad \forall i = 1, 2, \cdots, I, k = 1, 2, \cdots, K \tag{3.110}$$

   （b）次式から $h_{kj}^{(t+1)}$ を求める

$$h_{kj}^{(t+1)} = h_{kj}^{(t)} \frac{\sum_{i=1}^{I} \frac{x_{ij}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}} w_{ik}^{(t)}}{\sum_{j=1}^{J} w_{ik}^{(t)}} \quad \forall k = 1, 2, \cdots, K, j = 1, 2, \cdots, J \tag{3.111}$$

なお，式 (3.110) 及び (3.111) は，次のように行列形式で表現することもできる．

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \frac{\frac{\boldsymbol{X}}{\boldsymbol{W}\boldsymbol{H}}\boldsymbol{H}^{\mathrm{T}}}{\boldsymbol{1}\boldsymbol{H}^{\mathrm{T}}} \tag{3.112}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \frac{\boldsymbol{W}^{\mathrm{T}}\frac{\boldsymbol{X}}{\boldsymbol{W}\boldsymbol{H}}}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{1}} \tag{3.113}$$

ここで，見やすさのために反復回数 t の標記は省略し，変数更新を表す演算子 $\leftarrow$ を用いている．また，$\boldsymbol{1} \in \{1\}^{I \times J}$ は要素が全て 1 の行列である．さらに，行列に対する演算子 $\otimes$ は要素毎の積（Hadamard 積）を表し，行列の分数は要素毎の商を表す．

3.3.3 IS-NMF の場合

Itakura–Saito ダイバージェンスに基づく NMF（以下，IS-NMF と表記する）の最適化問題は式 (3.19) 及び (3.22) より，次式のようになる．

$$\begin{aligned} &\min_{\boldsymbol{W},\boldsymbol{H}} \sum_{i=1}^{I}\sum_{j=1}^{J} \left(\frac{x_{ij}}{\sum_{k=1}^{K} w_{ik}h_{kj}} - \log \frac{x_{ij}}{\sum_{k=1}^{K} w_{ik}h_{kj}} - 1 \right) \\ &\text{s.t. } w_{ik}, h_{kj} \geq 0 \ \forall i = 1, 2, \cdots, I, j = 1, 2, \cdots, J, k = 1, 2, \cdots, K \end{aligned} \tag{3.114}$$

この目的関数を変形すると

$$\begin{aligned} \mathcal{J}(\boldsymbol{W}, \boldsymbol{H}) &= \sum_{i=1}^{I}\sum_{j=1}^{J} \left(\frac{x_{ij}}{\sum_{k=1}^{K} w_{ik}h_{kj}} - \log \frac{x_{ij}}{\sum_{k=1}^{K} w_{ik}h_{kj}} - 1 \right) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J} \left(\frac{x_{ij}}{\sum_{k=1}^{K} w_{ik}h_{kj}} - \log x_{ij} + \log \sum_{k=1}^{K} w_{ik}h_{kj} - 1 \right) \end{aligned} \tag{3.115}$$

となる．Eu-NMF の場合と同様に w_{ik} や h_{kj} で偏微分を考えた時，式 (3.115) 中で計算が煩雑となる項は第 2 項及び第 3 項の k に関する総和を含む逆数関数及び対数関数である．逆数関数は凸関数であることから Jensen の不等式 (2.80) を適用で

きる．

$$\begin{aligned}\frac{1}{\sum_{k=1}^{K} w_{ik}h_{kj}} &= \frac{1}{\sum_{k=1}^{K} \delta_{ijk}\frac{w_{ik}h_{kj}}{\delta_{ijk}}} \\ &\leq \sum_{k=1}^{K} \delta_{ijk}\frac{\delta_{ijk}}{w_{ik}h_{kj}} \\ &= \sum_{k=1}^{K} \frac{\delta_{ijk}^2}{w_{ik}h_{kj}}\end{aligned} \tag{3.116}$$

式 (3.73) と同様に，補助変数 $\delta_{ijk} > 0$ は $\sum_{k=1}^{K} \delta_{ijk} = 1$ を満たす．式 (3.116) は，Jensen の不等式 (2.80) において，凸関数 $f : \mathbb{R} \to \mathbb{R}$ を $f(\cdot) = (\cdot)^{-1}$ として適用した不等式である．また，対数関数は凹関数であることから接線不等式 (2.81) を適用することができる．

$$\begin{aligned}\log \sum_{k=1}^{K} w_{ik}h_{kj} &\leq \frac{1}{\omega_{ij}} \left(\sum_{k=1}^{K} w_{ik}h_{kj} - \omega_{ij} \right) + \log \omega_{ij} \\ &= \frac{1}{\omega_{ij}} \sum_{k=1}^{K} w_{ik}h_{kj} - 1 + \log \omega_{ij}\end{aligned} \tag{3.117}$$

ここで，$\omega_{ij} > 0$ は補助変数である．不等式 (3.116) 及び (3.117) を式 (3.115) の第 2 項及び第 3 項にそれぞれ適用することで，補助関数を次式のように定義できる．

$$\begin{aligned}\mathcal{J}(\boldsymbol{W}, \boldsymbol{H}) &\leq \mathcal{J}^{+}(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\Delta}, \boldsymbol{\Omega}) \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} \left(x_{ij} \sum_{k=1}^{K} \frac{\delta_{ijk}^2}{w_{ik}h_{kj}} - \log x_{ij} + \frac{1}{\omega_{ij}} \sum_{k=1}^{K} w_{ik}h_{kj} - 1 + \log \omega_{ij} - 1 \right)\end{aligned} \tag{3.118}$$

ここで，$\boldsymbol{\Omega} \in \mathrm{R}_{>0}^{I \times J}$ は補助変数 ω_{ij} を要素として含む行列である．

まず，補助関数 $\mathcal{J}^{+}(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\Delta}, \boldsymbol{\Omega})$ に対する補助変数 δ_{ijk} 及び ω_{ij} の最小化について考える．Eu-NMF の場合と同様に，δ_{ijk} の最小解は不等式 (3.116) の等号成立条件から与えられるものである．不等式 (3.116) においても，補助変数が式 (3.85) を満たすときに等号が成立することから，同様の議論を経て式 (3.84) として補助変数 δ_{ijk} の反復更新則が得られることが分かる．また，ω_{ij} の最小解は，やはり不等式 (3.117) の等号成立条件から得られる．接線不等式 (2.81) の等号成立条件は $\tilde{\theta} = \theta$ であることから，補助変数の反復更新則は次式となる．

$$\omega_{ij} = \sum_{k=1}^{K} w_{ik}h_{kj} \tag{3.119}$$

次に，補助関数 $\mathcal{J}^{+}(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\Delta}, \boldsymbol{\Omega})$ に対する本来の変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ の最小化について考える．まず，補助関数 (3.118) を w_{ik} で偏微分して 0 とおくと，次式が得られる．

$$\frac{\partial \mathcal{J}^{+}}{\partial w_{ik}} = \sum_{j=1}^{J} \left(-x_{ij}\frac{\delta_{ijk}^2}{w_{ik}^2 h_{kj}} + \frac{h_{kj}}{\omega_{ij}} \right) = 0 \tag{3.120}$$

式 (3.120) を整理すると次式のようになる．

$$\begin{aligned}-\sum_{j=1}^{J} x_{ij}\frac{\delta_{ijk}^2}{w_{ik}^2 h_{kj}} + \sum_{j=1}^{J} \frac{h_{kj}}{\omega_{ij}} &= 0 \\ w_{ik}^2 \sum_{j=1}^{J} \frac{h_{kj}}{\omega_{ij}} &= \sum_{j=1}^{J} x_{ij}\frac{\delta_{ijk}^2}{h_{kj}} \\ w_{ik} &= \sqrt{\frac{\sum_{j=1}^{J} x_{ij}\frac{\delta_{ijk}^2}{h_{kj}}}{\sum_{j=1}^{J} \frac{h_{kj}}{\omega_{ij}}}}\end{aligned} \tag{3.121}$$

式 (3.121) が補助関数 (3.118) を最小化する w_{ik} である．反復回数 t を考慮して記述すると，次のようになる．

$$w_{ik}^{(t+1)} = \sqrt{\frac{\sum_{j=1}^{J} x_{ij} \frac{{\delta_{ijk}^{(t+1)}}^2}{h_{kj}^{(t)}}}{\sum_{j=1}^{J} \frac{h_{kj}^{(t)}}{\omega_{ij}{}^{(t+1)}}}} \tag{3.122}$$

同様に，補助関数 (3.118) を h_{kj} で偏微分して 0 とおくと，h_{kj} の最小解も次のように得られる．

$$\begin{aligned}
\frac{\partial \mathcal{J}^{+}}{\partial h_{kj}} &= \sum_{i=1}^{I} \left(-x_{ij} \frac{\delta_{ijk}^2}{w_{ik} h_{kj}^2} + \frac{w_{ik}}{\omega_{ij}} \right) = 0 \\
-\sum_{i=1}^{I} x_{ij} \frac{\delta_{ijk}^2}{w_{ik} h_{kj}^2} &+ \sum_{i=1}^{I} \frac{w_{ik}}{\omega_{ij}} = 0 \\
h_{kj}^2 \sum_{i=1}^{I} \frac{w_{ik}}{\omega_{ij}} &= \sum_{i=1}^{I} x_{ij} \frac{\delta_{ijk}^2}{w_{ik}} \\
h_{kj} &= \sqrt{\frac{\sum_{i=1}^{I} x_{ij} \frac{\delta_{ijk}^2}{w_{ik}}}{\sum_{i=1}^{I} \frac{w_{ik}}{\omega_{ij}}}}
\end{aligned} \tag{3.123}$$

反復回数 t を考慮して記述すると，次のようになる．

$$h_{kj}^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{I} x_{ij} \frac{{\delta_{ijk}^{(t+1)}}^2}{w_{ik}^{(t)}}}{\sum_{i=1}^{I} \frac{w_{ik}^{(t)}}{\omega_{ij}{}^{(t+1)}}}} \tag{3.124}$$

従って，式 (3.84) 及び (3.119) で補助変数を更新し，式 (3.122) 及び (3.124) で本来の変数を更新する．この 2 つの手順を繰り返すことで，目的関数 (3.115) の値を最小化する $\boldsymbol{W}$ 及び $\boldsymbol{H}$ を求めることができる．Eu-NMF 及び KL-NMF の場合と同様に，補助変数の反復更新則 (3.84) 及び (3.119) を式 (3.122) 及び (3.124) にそれぞれ代入することで，次式のような統合された反復更新則に書き換えることができる．

$$\begin{aligned}
w_{ik}^{(t+1)} &= \sqrt{\frac{\sum_{j=1}^{J} x_{ij} \frac{{w_{ik}^{(t)}}^2 {h_{kj}^{(t)}}^2 \Big/ \left(\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)} \right)^2}{h_{kj}^{(t)}}}{\sum_{j=1}^{J} \frac{h_{kj}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}} \\
&= w_{ik}^{(t)} \sqrt{\frac{\sum_{j=1}^{J} \frac{x_{ij}}{\left(\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)} \right)^2} h_{kj}^{(t)}}{\sum_{j=1}^{J} \frac{h_{kj}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}}
\end{aligned} \tag{3.125}$$

$$\begin{aligned}
h_{kj}^{(t+1)} &= \sqrt{\frac{\sum_{i=1}^{I} x_{ij} \frac{{w_{ik}^{(t)}}^2 {h_{kj}^{(t)}}^2 \Big/ \left(\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)} \right)^2}{w_{ik}^{(t)}}}{\sum_{i=1}^{I} \frac{w_{ik}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}} \\
&= h_{kj}^{(t)} \sqrt{\frac{\sum_{i=1}^{I} \frac{x_{ij}}{\left(\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)} \right)^2} w_{ik}^{(t)}}{\sum_{j=1}^{J} \frac{w_{ik}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}}
\end{aligned} \tag{3.126}$$

以上より，IS-NMF の MM アルゴリズムに基づく反復更新則を構築できる．

IS-NMF の MM アルゴリズムに基づく反復更新則

IS-NMF の最適化変数 $\boldsymbol{W}$ 及び $\boldsymbol{H}$ は，次式を繰り返すことで最適化できる．

1. 適当な非負の初期値 $[\boldsymbol{W}^{(0)}]_{ik} = w_{ik}^{(0)}$ 及び $[\boldsymbol{H}^{(0)}]_{kj} = h_{kj}^{(0)}$ を全ての $i = 1, 2, \cdots, I$，$j = 1, 2, \cdots, J$，及び $k = 1, 2, \cdots, K$ についてとる
2. 以下の操作を $t = 0, 1, 2, \cdots$ について繰り返す
   （a）次式から $w_{ik}^{(t+1)}$ を求める

$$w_{ik}^{(t+1)} = w_{ik}^{(t)} \sqrt{\frac{\sum_{j=1}^{J} \frac{x_{ij}}{\left(\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}\right)^2} h_{kj}^{(t)}}{\sum_{j=1}^{J} \frac{h_{kj}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}} \quad \forall i = 1, 2, \cdots, I, k = 1, 2, \cdots, K \tag{3.127}$$

   （b）次式から $h_{kj}^{(t+1)}$ を求める

$$h_{kj}^{(t+1)} = h_{kj}^{(t)} \sqrt{\frac{\sum_{i=1}^{I} \frac{x_{ij}}{\left(\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}\right)^2} w_{ik}^{(t)}}{\sum_{j=1}^{J} \frac{w_{ik}^{(t)}}{\sum_{k'=1}^{K} w_{ik'}^{(t)} h_{k'j}^{(t)}}}} \quad \forall k = 1, 2, \cdots, K, j = 1, 2, \cdots, J \tag{3.128}$$

なお，式 (3.127) 及び (3.128) は，次のように行列形式で表現することもできる．

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{\frac{\boldsymbol{X}}{(\boldsymbol{W}\boldsymbol{H})^{\cdot 2}} \boldsymbol{H}^{\mathrm{T}}}{\frac{\boldsymbol{1}}{\boldsymbol{W}\boldsymbol{H}} \boldsymbol{H}^{\mathrm{T}}} \right)^{\cdot \frac{1}{2}} \tag{3.129}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^{\mathrm{T}} \frac{\boldsymbol{X}}{(\boldsymbol{W}\boldsymbol{H})^{\cdot 2}}}{\boldsymbol{W}^{\mathrm{T}} \frac{\boldsymbol{1}}{\boldsymbol{W}\boldsymbol{H}}} \right)^{\cdot \frac{1}{2}} \tag{3.130}$$

ここで，見やすさのために反復回数 t の標記は省略し，変数更新を表す演算子 $\leftarrow$ を用いている．また，$\boldsymbol{1} \in \{1\}^{I \times J}$ は要素が全て 1 の行列である．さらに，行列に対する演算子 $\otimes$ は要素毎の積（Hadamard 積），行列の分数は要素毎の商，行列のドット付き指数乗は要素毎の指数乗を表す．

実際に NMF の MM アルゴリズムを計算する場合は，乗算型反復更新則であることから，w_{ik} や h_{kj} の初期値に 0 を与えてしまうと，その要素は全く更新されなくなってしまう．この現象を避けるために，初期値 $\boldsymbol{W}^{(0)}$ 及び $\boldsymbol{H}^{(0)}$ には正の乱数を与える等の手法がとられる．その他，より効率的な初期値を観測非負値行列 $\boldsymbol{X}$ から推定する手法も提案されている [12]．また，反復更新の途中で浮動小数点の 0 以上の最小値を下回り，0 となる要素が生じた場合には，その要素については反復更新を止めるという処置をすべきである．これは，例えば w_{ik} が $k = 1, 2, \cdots, K$ について全て 0 となってしまった場合，反復更新則の計算過程において零割りが生じるためである[*24]．このような数値的不安定性を解消するために，各変数の更新後に次式のようなフロアリング処理を施すことが有用である．

$$\boldsymbol{W} \leftarrow \max(\boldsymbol{W}, \varepsilon) \tag{3.131}$$

$$\boldsymbol{H} \leftarrow \max(\boldsymbol{H}, \varepsilon) \tag{3.132}$$

ここで，$\max(\cdot, \cdot)$ は各要素に関して何れか大きい方を返す演算であり，ε は計算機イプシロン（machine epsilon）を表す．

最後に，本章では Eu-NMF，KL-NMF，及び IS-NMF の 3 種類について反復更新則を導出したが，これらを統一的に表した β ダイバージェンスに基づく NMF の MM アルゴリズムによる反復更新則が導出されている [13]．その他にも，用途に応じて様々な類似度関数が目的関数に用いられており，本章と同様の手法で反復更新則が導出されている例も多い．

[*24] とはいえ，NMF の目的が低ランク近似であることを考えると，ある基底ベクトル $\boldsymbol{w}_k$ が零ベクトルとなることは考えにくい．

4. まとめ

本稿では，機械学習の分野で頻繁に用いられる NMF の本質的な説明，反復更新則の導出に必要な最適化の基礎理論，及び有名な 3 種類の NMF の反復更新則の導出を示した．NMF は，観測値の非負性を陽に考慮した教師無し学習と捉えることができ，多くの応用が挙げられる．また，その最適化理論として MM アルゴリズムに基づく手法が有名であり，2000 年代から理論体系が確立・発展している．本稿に示した NMF の導出の方法をよく理解すれば，応用として NMF を活用した様々な変形問題・拡張問題を考え，さらに実践することが期待される．なお，本稿の記述に誤りを見つけた場合は，その旨を著者に連絡をして頂ければ幸いである．

参考文献

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

[2] S. A. Vavasis, “On the complexity of nonnegative matrix factorization,” *SIAM J. Optimization*, vol. 20, no. 3, pp. 1364–1377, 2009.

[3] Y. Cao, P. P. B. Eggermont, and S. Terebey, “Cross Burg entropy maximization and its application to ringing suppression in image reconstruction,” *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 286–292, 1999.

[4] K. Lange, *MM Optimization Algorithms*, Society for Industrial & Applied Mathematics, U.S., 2016

[5] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. Signal Processing*, vol. 65, no. 3, 2017.

[6] C. Févotte and J. Idier, “Algorithms for nonnegative matrix factorization with the β-divergence,” *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[7] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization” *Proc. Neural Information Processing Systems*, pp. 556–562, 2000.

[8] S. Smale, “Differentiable dynamical systems,” *Bulletin of the American mathematical Society*, vol. 73, no. 6, pp. 747–817, 1967.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University press, 2004.

[10] S. Eguchi and Y. Kano, “Robustifying maximum likelihood estimation,” *The Institute of Statistics and Mathematics*, Technical Report, 2001.

[11] K. B. Petersen and M. S. Pedersen, “The matrix cookbook,” *Technical University of Denmark*, vol. 7, no. 15, 2008.

[12] D. Kitamura and N. Ono, “Efficient initialization for nonnegative matrix factorization based on nonnegative independent component analysis,” *Proc. International Workshop on Acoustic Signal Enhancement*, 2016.

[13] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, “Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with beta-divergence,” *Proc. International Workshop on Machine Learning for Signal Processing*, pp. 283–288, 2010.