第7章　特異値分解とその応用

7.1　特異値分解

対角化およびスペクトル分解を一般化したジョルダン分解は，正方行列にのみ定義された．非正方行列でさらに一般化された分解が**特異値分解 (singular value decomposition)** である．

【補題 7.1 】 $n \times m$ 行列 $A_{n\times m}$ $(m \geq n)$ の行ベクトルを直交系へ変換する適当な $n \times n$ 直交行列 $Q_{n\times n}$ が存在する．

【定理 7.1 特異値分解 **】** $n \times m$ 行列 A $(m \geq n)$ は，適当な $n \times n$ 直交行列 Q_1 および $m \times m$ 直交行列 Q_2 を用いて，

$$A = Q_1(\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)|0_{n\times(m-n)})Q_2^T$$

で分解できる．特に，$n \times m$ 行列

$$\Sigma = (\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)|0_{n\times(m-n)}) = \left(\begin{array}{cccc|ccc} \sigma_1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n & 0 & \ldots & 0 \end{array}\right)$$

の非負実数値 $\{\sigma_i \geq 0\}_{i=1}^n$ を行列 A の**特異値 (singular value)** という．

【系 7.1 対称行列の特異値分解 **】** 固有値 $\{\lambda_i\}_{i=1}^n$ を持つ対称行列 A の特異値は，$\{\sigma_i = |\lambda_i|\}_{i=1}^n$ となる．

【補題 7.1 の証明】対称行列 AA^T の対角化

$$A_{n\times m}A_{m\times n}^T = Q_{n\times n}\mathrm{diag}(\lambda_1, \ldots, \lambda_n)Q_{n\times n}^T$$

によって，

$$Q_{n\times n}^T A_{n\times m}(Q_{n\times n}^T A_{m\times n})^T = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

ここで，$Q^T A$ の行ベクトルを $\{p_i^T \in R^m\}_{i=1}^n$ とすると，

$$Q_{n\times n}^T A_{n\times m}\ (Q_{n\times n}^T A_{n\times m})^T = \begin{pmatrix} p_1^T \\ \hline \vdots \\ \hline p_n^T \end{pmatrix} (p_1|\ldots|p_n) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

から，$p_i^T p_j = \lambda_i \delta_{i,j}$．■

上記の証明において，定理 5.3 から半正定値行列 AA^T の固有値 $\{\lambda_i\}_{i=1}^n$ はすべて非負となるため，変換された行ベクトルのノルムは矛盾せずに求まる．以降，ノルムを $\| p_i \| = \sqrt{\lambda_i} = \sigma_i$ とする．

【定理 7.1 の証明】補題 7.1 における行ベクトルに $m-n$ 個の適当なベクトルを加えて得られる正規直交系を $\{q_i^T \in R^m\}_{i=1}^m$ とし，$Q_1 = Q_{n\times n}$ とすると，

$$Q_1^T A_{n\times m} = (\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)|0_{n\times(m-n)}) \begin{pmatrix} q_1^T \\ \hline \vdots \\ \hline q_m^T \end{pmatrix}.$$

ここで，

$$Q_2^T = \begin{pmatrix} q_1^T \\ \hline \vdots \\ \hline q_m^T \end{pmatrix}$$

とすると，$A = Q_1 \Sigma Q_2^T$ を得る．■

【例示 7.1　2×2 行列への適用】2×2 行列 $A = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$ を特異値分解してみる．まず，

$$AA^T = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} = Q_1 \Sigma \Sigma^T Q_1^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

したがって，

$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

から，特異値は $\{2, 1\}$ である．A の固有値は，純虚数 $\{\pm\sqrt{2}i\}$ より，固有値と特異値は異なることに注意を促す．同様に，

$$A^T A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} = Q_2 \Sigma^T \Sigma Q_2^T$$

より，

$$Q_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

したがって，特異値分解は，

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}.$$

【例示 7.2　2×3 行列への適用】2×3 行列 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ を特異値分解してみる．まず，

$$AA^T = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = Q_1 \Sigma \Sigma^T Q_1^T$$

となる．行列 AA^T は，$(AA^T)^T = AA^T$ を満足するため対称行列であり，実固有値を持ち，直交行列によって対角化される．この例では，既に対角行列となっており，

$$Q_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Sigma\Sigma^T = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

したがって，

$$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

から，特異値は $\{\sqrt{2}, 1\}$ である．同様に，行列 A^TA も対称行列であり，

$$A^TA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = Q_2\Sigma^T\Sigma Q_2^T$$

よりランクが 2 となる．これを対角化すると

$$A^TA = Q_2\Sigma^T\Sigma Q_2^T = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

したがって，

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

$m > n$ の場合，明らかに A^TA はフルランクなく，非正則となる．このため，零固有値に対する固有ベクトルの取り方が問題となるがグラム-シュミット直交化を利用すればよい．

7.2 不能線形方程式の最小 2 乗解

【定義 7.1 】 n 次元変数ベクトル x と $m \times n$ 係数行列 A による不能な線形方程式 $Ax \approx b$ において，2 乗誤差

$$\| b - Ax \|^2 = (b - Ax)^T(b - Ax)$$

を最小とする変数ベクトル x^* を不能線形方程式の**最小 2 乗解 (least squares solution)** という．

【定義 7.2 】 n 変数によるベクトル x と $m \times n$ 係数行列 A による不能な線形方程式 $Ax \approx b$ に関して，

$$A^TAx = A^Tb$$

を**正規方程式 (normal equation)** といい，正規方程式の解は最小 2 乗解を与える．

【定理 7.2 疑似逆行列 **】** A^TA が正則である場合，2 乗誤差

$$U(x) = \| Ax - b \|^2$$

を最小とする解は，

$$x^* = (A^TA)^{-1}A^Tb$$

であたえられる．特に, 行列 $A^+ = (A^TA)^{-1}A^T$ を，**疑似逆行列 (pseudinverse matrix)** という．

【証明】慣性法則によって，A の特異行列の積

$$\Sigma^T\Sigma = \left(\begin{array}{cccc|ccc} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n & 0 & \dots & 0 \end{array}\right) \left(\begin{array}{cccc} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ \hline 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{array}\right) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \tag{7.1}$$

のランクは，$A^TA = Q_2^T\Sigma^T\Sigma Q_2$ のランクと等しい．よって，特異行列はフルランクあり，全ての特異値は非零である．行列 A の特異値分解 $A = Q_1\Sigma Q_2^T$ を代入し，

$$U(x) = \| Q_1\Sigma Q_2^Tx - b \|^2 = \| Q_1(\Sigma Q_2^Tx - Q_1^Tb) \|^2 = \| \Sigma Q_2^Tx - Q_1^Tb \|^2 .$$

ここで，

$$\begin{array}{rcl} y & = & Q_2^Tx = (y_1, y_2, \dots, y_n)^T, \\ c & = & Q_1^Tb = (c_1, c_2, \dots, c_m)^T \end{array}$$

により直交変換すると，

$$\Sigma y - c = \left(\begin{array}{cccc} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ \hline 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{array}\right) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \left(\begin{array}{c} c_1 \\ c_2 \\ \vdots \\ c_n \\ \hline c_{n+1} \\ \vdots \\ c_m \end{array}\right).$$

よって，

$$U(y) = \sum_{i=1}^{n} |\sigma_i y_i - c_i|^2 + \sum_{i=n+1}^{m} |c_i|^2 \tag{7.2}$$

となり，$y_i = \frac{c_i}{\sigma_i} (i = 1, 2, \dots, n)$ で 2 乗誤差は最小となる．他方，(7.1) から，最小解を与える y^* は，

$$y^* = (\Sigma^T\Sigma)^{-1}\Sigma^Tc = (\Sigma^T\Sigma)^{-1}\Sigma^TQ_1^Tb$$

と表記でき，逆変換すると

$$x^* = Q_2(\Sigma^T\Sigma)^{-1}\Sigma^TQ_1^Tb.$$

ここで，擬似逆行列 [1]

$$\begin{array}{rcl} A^+ & = & (A^TA)^{-1}A^T = (Q_2\Sigma^T\Sigma Q_2^T)^{-1}Q_2\Sigma^TQ_1^T = (Q_2^T)^{-1}(Q_2\Sigma^T\Sigma)^{-1}Q_2\Sigma^TQ_1^T \\ & = & Q_2(\Sigma^T\Sigma)^{-1}(Q_2)^{-1}Q_2\Sigma^TQ_1^T = Q_2(\Sigma^T\Sigma)^{-1}\Sigma^TQ_1^T \end{array}$$

が導かれる．■

ここで，注意に値するのは誤差 (7.2) の末尾項 $\sum_{i=n+1}^{d} |c_i|^2$ である．一般に，変数の数 n と方程式の数 d が一致すると誤差 (7.2) はなくなる．この場合，疑似逆行列は，

$$A^+ = (A^TA)^{-1}A^T = A^{-1}(A^T)^{-1}A^T = A^{-1}$$

から逆行列となり，線形方程式 $Ax = b$ は完全に解けることになる．

[1] 同種の逆行列に関しては，より詳細な分類と名称が存在している．ここでは，この名称と定義を用いる．詳細は，正則化問題に関する文献等を参照してほしい．

7.3 情報処理への展開

7.3.1 最小 2 乗法と線形予測

ある観測量 $x \in R^n$ とそれに応じて変動する別な観測量 $y \in R^1$ があるとする．x と y の関数関係が線形

$$y = a^T x, \quad a \in R^n$$

であることを仮定し，過去の値から現在の値を推定する手法を**線形予測 (linear prediction)** という．例えば，m 回の観測結果 $\{y_i,\ x_i\}_{i=1}^m$ に対して推定性能を 2 乗誤差

$$E = \sum_{i=1}^{m} \| y_i - a^T x_i \|^2$$

で評価するとき，これを最小化する係数値 a^* が予測に適当と考えられる．この場合，

$$E = \| \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} x_1^T \\ \hline x_2^T \\ \hline \vdots \\ \hline x_m^T \end{pmatrix} a \|^2$$

から，正規方程式

$$(x_1|x_2|\dots|x_m) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = (x_1|x_2|\dots|x_m) \begin{pmatrix} x_1^T \\ \hline x_2^T \\ \hline \vdots \\ \hline x_m^T \end{pmatrix} a$$

を導き，最小 2 乗解

$$a^* = ((x_1|x_2|\dots|x_m) \begin{pmatrix} x_1^T \\ \hline x_2^T \\ \hline \vdots \\ \hline x_m^T \end{pmatrix})^{-1} (x_1|x_2|\dots|x_m) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

が得られる．

信号処理の音声合成あるいは音声信号の情報圧縮においては，線形予測が有効に用いられている．例えば，離散化された音声デ− タを，音声信号の最小単位である子音あるいは母音に分割して，それぞれに関して線形予測が適用される．厳密にいえば，実データには観測に伴う誤差や対象となる系に固有の外乱等が含まれている．この意味で，観測データに基づく最小 2 乗解が最良の解となるとはいい難い．この問題は，統計や統計数理において研究されている．

7.3.2 不良設定問題と正則化

未知変数よりも等式の数が多い不能線形方程式の解法を考えてきた．逆に，等式の数よりも未知変数が多い不定方程式の解法はどうであろうか．この問題は，**不良設定問題**（ill-posed problem）と言われており，**正則化**（regularization）という枠組でアプローチがなされている．例えば，有界な 1 次元格子 $\{i = 1, 2, \dots, 4M\}$ 上で $\{j = 1, 4, 8, \dots, 4M\}$ だけは高度データ $\{h_1, h_4, h_8, \dots, h_{4M}\}$ が既知であるとする．この情報から，全ての格子点上の高度 $\{x_i\ :\ i = 1, 2, \dots, 4M\}$ を推定する問題を考えてみる．誤差関数

$$U = |x_1 - h_1|^2 + |x_4 - h_4|^2 + \dots + |x_{4M} - h_{4M}|^2$$

を行列形式

$$\mathrm{diag}(1,0,0,1,\ldots,0,1)(x_1,x_2,\ldots,x_{4M})^T=(h_1,0,0,h_4,\ldots,0,h_{4M})^T$$

で記述すると，当然であるが作用素 $\mathrm{diag}(1,0,0,1,\ldots,0,1)$ は非正則となり，逆行列を持たない．そこで，両端点以外は，“隣接する左右の高度の平均値に近い値を持つ” という連続性を仮定し，誤差関数

$$U+\lambda\sum_{i=2}^{i=4M-1}|2x_i-x_{i-1}-x_{i+1}|^2 \quad (\lambda>0)$$

の最小解に注目する．ここで，連立方程式の作用素は，

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ -\lambda & 2\lambda & -\lambda & 0 & 0 & \ldots & 0 \\ 0 & -\lambda & 2\lambda & -\lambda & 0 & \ldots & 0 \\ 0 & 0 & -\lambda & 1+2\lambda & -\lambda & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

と正則化されるため，逆行列により妥当な解を得ることができる．無論，正則化項強度を決める λ の選択方法は，真の曲面の局所構造 (不連続性あるいは高次微分可能性) に依存して巧みに選択する必要がある．これに関しては，ベイズ統計を例とする統計数理の研究対象となっている．

章末問題

【問題 7.1 特異値分解 **】** 2 つの直交行列を求め，特異値分解せよ．(注意：特異値は正か零であることに気をつけよ.)

$$(1)\quad \begin{pmatrix} \frac{-1}{\sqrt{2}} & 1 & -1 \\ \frac{-1}{\sqrt{2}} & -1 & 1 \end{pmatrix} \qquad (2)\quad \begin{pmatrix} \frac{-1}{2} & 1 & \frac{-1}{2} & 1 \\ \frac{-1}{2} & -1 & \frac{-1}{2} & -1 \end{pmatrix}$$

【問題 7.2 正規方程式の導出**】** 解析学の知識から，$m\times n$ 係数行列 A，定数列ベクトル $a\in R^m$ と変数列ベクトル $x=(x_1,x_2,\ldots,x_n)^T$ で与えられる *2* 次曲面

$$U(x)=\| Ax-a \|^2$$

を極小とする値 $x=x^*$ は，連立方程式

$$\mathrm{grad}\ U(x)=\begin{pmatrix} \frac{\partial U(x)}{\partial x_1} \\ \vdots \\ \frac{\partial U(x)}{\partial x_n} \end{pmatrix}=0 \tag{7.3}$$

の解となることが知られている．(第 9 章参照）(7.3) に関して次の問に答えよ．尚，ベクトル関数

$$v\ :\ x\mapsto v(x)=(v_1(x),v_2(x),\ldots,v_n(x))^T$$

のヤコビ行列は，

$$Dv(x)=\begin{pmatrix} \frac{\partial v_1(x)}{\partial x_1} & \ldots & \frac{\partial v_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial v_n(x)}{\partial x_1} & \ldots & \frac{\partial v_n(x)}{\partial x_n} \end{pmatrix}$$

で表記する．

(1) B を $n \times n$ 定数行列および $b \in R^n$ を定数列ベクトルとするとき，grad $(b^T x)$ および grad $(x^T b)$ を求めよ．また，$D(Bx)$ を求めよ．

(2) grad $U(x)$ を求め，(7.3) が正規方程式 $A^T Ax - A^T a = 0$ となることを示せ．

(3) 極値が存在する条件，すなわち，$U(x)$ のヘッセ行列式が正であるならば，正規方程式 (7.3) が解ける（孤立解を持つ）ことを示せ．(ヒント：ヘッセ行列式は $\det(D \text{ grad } U(x))$ となることに注意せよ.)

【問題 7.3 不能方程式の最小 2 乗解 **】** 次の係数行列 A と定数ベクトル b で与えられる不能方程式の疑似逆行列 A^+ および最小 2 乗解を計算せよ．

$$(1) \quad A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 2 \\ 1 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad (2) \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

【問題 7.4 データフィッティング **】** 区間 I 上のある曲線 $y = f(x)$ は，異なる 6 つの点

$$\{(x_i, y_i) \ : \ i = 1, 2, \ldots, 6\}$$

を通るとする．この曲線 $y = f(x)$ を多項式 $ax^3 + bx^2 + cx + d$ で最小 2 乗の意味で近似するとき, 具体的な計算方法を述べよ．

【問題 7.5 正則化 **】** 本章で議論した有界な *1* 次元格子点上の正則化において，隣接する左右の高度の平均と等しくなるという誤差関数

$$\sum_{i=2}^{i=4M-1} |2x_i - x_{i-1} - x_{i+1}|^2$$

は，

$$\sum_{i=2}^{i=4M-1} |(x_i - x_{i-1}) - (x_{i+1} - x_i)|^2$$

と変形することで左右の勾配（差分）が等しくなる一次微係数の保存を意味する．同様にして，*2* 次微係数の保存を意味する誤差関数を求めよ．