主成分分析

京都大学大学院工学研究科化学工学専攻
プロセスシステム工学研究室
加納　学

1997 年 1 月 第 1 版作成
2002 年 5 月 第 2 版作成

［注意事項］
自由に利用していただいて結構ですが，著作権は一切放棄していません．また，本資料の間違いなどによって生じた不利益などに対して，著者は一切責任を負いません．勿論，間違いの指摘やアドバイスは歓迎します.

目 次

1　主成分分析とは

身体検査で身長，体重という 2 種類の変数が 10 人について測定されており，このデータをプロットすると Fig.1 のようになったとする．この図から，身長が大きくなるほど体重も大きくなるという傾向が把握できる．この傾向を表現するためには，新たに図中 z_1 のような軸を考えればよいであろう．この z_1 軸は「体の大きさ」を表していると解釈できる．しかし,「体の大きさ」だけで各人の特徴をすべて表現できるわけではない．10 人の中には太っている人も痩せている人もいるのである．そこで，z_1 軸と直交する z_2 軸を考える．この z_2 軸は「肥満の程度」を表していると解釈できる．このように,「身長」や「体重」といった変数を独立に扱うのではなく,「体の大きさ」や「肥満の程度」といった総合的な指標を導入することによって，データに含まれる変数間の関係や特徴が容易に把握できるようになる．このような総合的指標を統計的に設定し，変数間の関係を把握するための手法が，**主成分分析** (PCA; Principal Component Analysis) と呼ばれるものである．主成分とは総合的指標のことであり，身体検査の例では,「体の大きさ」を表す z_1 が第 1 主成分,「肥満の程度」を表す z_2 が第 2 主成分と呼ばれる．

「体の大きさ」を表す z_1 によってデータの持つ特徴 (あるいは傾向) は概ね表現できているので，z_1 のみでデータを代表させることもできる．この場合，2 種類の変数によって表現されていたデータを 1 種類の変数で表現することになる．すなわち，データの持つ特徴を主成分を用いて表すことにより，情報の損失を最小限に抑えながら，2 次元のデータを 1 次元のデータに変換したわけである．このような観点から，主成分分析をデータの低次元化を行うための手法とみることもできる．

身体検査の結果を「体の大きさ」で代表させる場合,「肥満の程度」に関する情報は失われることになる．このようにデータの低次元化は情報の損失を伴うが，そのときの情報の損失量は各データ点から第 1 主成分 z_1 へおろした垂線の長さで表される．データの持つ特徴を最もよく表現するためには，情報損失量をできる限り小さくする必要がある．実際，主成分分析では，情報損失量を最小化するという条件の下で主成分が決定される．

情報損失量を最小にするということは，得られる情報量を最大にすることの裏返しである．これは，データのばらつきを最もよく表す方向，すなわち分散が最大となる方向に主成分を設定することによって実現できる．このことを Fig.1 を用いて示そう．いま，点 $A(x_{11}, x_{12})$ に着目すると，第 1 主成分 z_1 のみでデータを代表させる場合の情報損失量は点 A から直線 z_1 へ下ろした垂線の長さ $\overline{AB}$ となり，z_1 を用いて得られる情報量は $\overline{OB}$ で与えられる．ここで，点 O は z_1 軸の原点であり，データの重心である．この新しい情報量 $\overline{OB}$ は第 1 主成分得点と呼ばれる．情報損失量と新しい情報量 (主成分得点) との間には，明らかに

$$\overline{OA}^2 \quad = \quad \overline{OB}^2 + \overline{AB}^2 \tag{1}$$

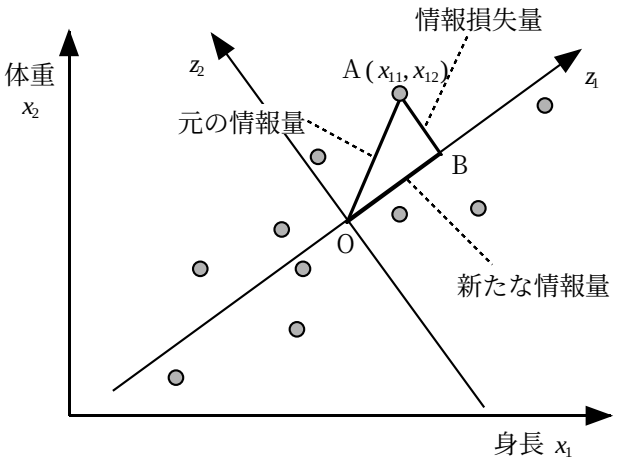

Fig. 1　身体検査の結果

という関係が成り立つ．$\overline{OA}$ を元の情報量と呼ぶことにすると，10 点のデータ全体について，

$$(\text{元の情報量の二乗和}) = (\text{新しい情報量の二乗和}) + (\text{情報損失量の二乗和})$$

という関係が成り立つ．ここで，元の情報量の二乗和はある定数であるから，情報損失量の二乗和を最小にすることと新しい情報量の二乗和を最大にすることとが等価であることがわかる．

ここまで 2 変数の場合について述べてきたが，多変数の場合にも本質的な違いはない．すなわち，主成分分析とは，P 個の変数 $\{x_p\}(p = 1, 2, \cdots, P)$ の持つ情報を，情報の損失を最小限に抑えながら，$\{x_p\}$ の一次結合として与えられる互いに独立な $M(M \leq P)$ 個の**主成分** (総合的指標)$\{z_m\}$

$$z_m \quad = \quad \sum_{p=1}^{P} w_{pm} x_p \quad (m = 1, 2, \cdots, M) \tag{2}$$

を用いて表現する手法である．なお，z_m は第 m 主成分と呼ばれ，その結合係数 $\{w_{pm}\}(p = 1, 2, \cdots, P;\ m = 1, 2, \cdots, M)$ は以下の条件を満足するように決定される．

＜条件＞

第 1 主成分 z_1 の分散は $\{x_p\}(p = 1, 2, \cdots, P)$ のあらゆる 1 次式の持つ分散の中で最大であり，第 m 主成分 $\{z_m\}(m = 2, \cdots, M)$ の分散は $\{z_{m'}\}(m' = 1, 2, \cdots, m-1)$ の全てと無相関な 1 次式の持つ分散の中で最大である．ただし，

$$\sum_{p=1}^{P} w_{pm}^2 \quad = \quad 1 \quad (m = 1, 2, \cdots, M) \tag{3}$$

とする．

2　主成分の導出

2.1　準備

Sec.1 で述べた条件に従い，主成分の分散が最大となるように主成分を決定する方法について述べる．いま，P 個の変数について N 個のサンプルがある場合を考え，測定値を $\{x_{np}^*\}(n = 1, 2, \cdots, N; p = 1, 2, \cdots, P)$ とする．以下の議論を簡単にするために，各変数についてその平均値 $\{\bar{x}_p\}(p = 1, 2, \cdots, P)$ からの偏差 $\{x_{np}\}$ を導入する．すなわち，

$$x_{np} \quad = \quad x_{np}^* - \bar{x}_p \quad (n = 1, 2, \cdots, N; p = 1, 2, \cdots, P) \tag{4}$$

とする．このとき，測定データ全体は次の行列 $\boldsymbol{X}$ で与えられる．

$$\boldsymbol{X} \quad = \quad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1P} \\ x_{21} & x_{22} & \cdots & x_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NP} \end{pmatrix} \tag{5}$$

2.2　第 1 主成分の導出

第 1 主成分 z_1 は Eq.(2) で与えられるので，その結合係数を

$$\boldsymbol{w}_1 \quad = \quad \begin{pmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{P1} \end{pmatrix} \tag{6}$$

とすると，n 番目のサンプル

$$\boldsymbol{x}_n = \begin{pmatrix} x_{n1} & x_{n2} & \cdots & x_{nP} \end{pmatrix} \tag{7}$$

に対応する第 1 主成分 z_1 の値 t_{n1} は

$$\begin{aligned} t_{n1} &= \sum_{p=1}^{P} w_{p1} x_{np} \\ &= \boldsymbol{x}_n \boldsymbol{w}_1 \end{aligned} \tag{8}$$

となる．この第 1 主成分 z_1 の値 t_{n1} を**第 1 主成分得点**と呼ぶ．N 個のサンプルに対応する第 1 主成分得点を 1 つのベクトルにまとめ，

$$\boldsymbol{t}_1 = \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{N1} \end{pmatrix} \tag{9}$$

とおくと，

$$\boldsymbol{t}_1 = \boldsymbol{X}\boldsymbol{w}_1 \tag{10}$$

が成り立つ．第 1 主成分得点の平均値 $\bar{t}_1$ は

$$\begin{aligned} \bar{t}_1 &= \frac{1}{N}\sum_{n=1}^{N} t_{n1} \\ &= \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{w}_1 \\ &= \frac{1}{N}\sum_{n=1}^{N} \left(\sum_{p=1}^{P} w_{p1} x_{np} \right) \\ &= \frac{1}{N}\sum_{p=1}^{P} w_{p1} \left(\sum_{n=1}^{N} x_{np} \right) \\ &= 0 \end{aligned} \tag{11}$$

であるから，第 1 主成分 z_1 の分散 $\sigma_{z_1}^2$ は

$$\begin{aligned} \sigma_{z_1}^2 &= \frac{1}{N-1} \boldsymbol{t}_1^T \boldsymbol{t}_1 \\ &= \frac{1}{N-1} \left(\boldsymbol{X}\boldsymbol{w}_1 \right)^T \left(\boldsymbol{X}\boldsymbol{w}_1 \right) \\ &= \boldsymbol{w}_1^T \boldsymbol{V} \boldsymbol{w}_1 \\ &\geq 0 \end{aligned} \tag{12}$$

となる．なお，行列 $\boldsymbol{V}$ は共分散行列と呼ばれる非負定値行列であり，

$$\boldsymbol{V} = \frac{1}{N-1} \boldsymbol{X}^T \boldsymbol{X} \tag{13}$$

で与えられ，その (i, j) 要素 v_{ij} は

$$v_{ij} = \frac{1}{N-1}\sum_{n=1}^{N} x_{ni} x_{nj} \tag{14}$$

$$= \frac{1}{N-1}\sum_{n=1}^{N} \left(x_{ni}^* - \bar{x}_i \right) \left(x_{nj}^* - \bar{x}_j \right) \tag{15}$$

である．また，明らかに，$v_{ij} = v_{ji}$ すなわち $\boldsymbol{V} = \boldsymbol{V}^T$ が成り立つ．

前節の条件より，第 1 主成分 z_1 は Eq.(3) の下でその分散 $\sigma_{z_1}^2$ が最大となるように決定されなければならない．この最適化問題は Lagrange 乗数法を用いて容易に解くことができる．すなわち，Lagrange 乗数 λ を導入して

$$J_1 = \boldsymbol{w}_1^T \boldsymbol{V} \boldsymbol{w}_1 - \lambda \left(\boldsymbol{w}_1^T \boldsymbol{w}_1 - 1 \right) \tag{16}$$

とおき，J_1 を最大にするような結合係数 $\boldsymbol{w}_1$ を求めればよい．そこで，J_1 を $\boldsymbol{w}_1$ で偏微分して $\boldsymbol{0}$ とおくと，

$$\begin{aligned}
\frac{\partial J_1}{\partial \boldsymbol{w}_1} &= \begin{pmatrix} \dfrac{\partial J_1}{\partial w_{11}} \\ \dfrac{\partial J_1}{\partial w_{21}} \\ \vdots \\ \dfrac{\partial J_1}{\partial w_{P1}} \end{pmatrix} \\
&= \begin{pmatrix} 2\displaystyle\sum_{p=1}^{P} v_{1p} w_{p1} \\ 2\displaystyle\sum_{p=1}^{P} v_{2p} w_{p1} \\ \vdots \\ 2\displaystyle\sum_{p=1}^{P} v_{Pp} w_{p1} \end{pmatrix} - 2\lambda \begin{pmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{P1} \end{pmatrix} \\
&= 2\boldsymbol{V}\boldsymbol{w}_1 - 2\lambda \boldsymbol{w}_1 \\
&= \boldsymbol{0}
\end{aligned} \tag{17}$$

となることから

$$\left(\boldsymbol{V} - \lambda \boldsymbol{I} \right) \boldsymbol{w}_1 = \boldsymbol{0} \tag{18}$$

という条件式が得られる．これは固有値問題に他ならず，Lagrange 乗数 λ が満たすべき条件は固有方程式

$$\det \left| \boldsymbol{V} - \lambda \boldsymbol{I} \right| = 0 \tag{19}$$

を用いて表わすこともできる．以上より，Lagrange 乗数 λ および第 1 主成分 z_1 の結合係数 $\boldsymbol{w}_1$ はそれぞれ共分散行列 $\boldsymbol{V}$ の固有値および固有ベクトルとして与えられることがわかる．

さて，共分散行列 $\boldsymbol{V}$ は P 次元正方行列であるため，その固有値 λ は P 個存在するが，その中のいずれの固有値が第 1 主成分 z_1 の分散 $\sigma_{z_1}^2$ を最大にする結合係数 (固有ベクトル)$\boldsymbol{w}_1$ に対応しているのであろうか．この問題について考えてみよう．第 1 主成分 z_1 の分散 $\sigma_{z_1}^2$ は Eq.(12) で与えられ，その結合係数 $\boldsymbol{w}_1$ は条件 Eq.(18) を満足しなければならない．そこで，Eq.(12) に Eq.(18) を代入し，$\boldsymbol{w}_1^T \boldsymbol{w}_1 = 1$ に注意すると，

$$\begin{aligned}
\sigma_{z_1}^2 &= \boldsymbol{w}_1^T \boldsymbol{V} \boldsymbol{w}_1 \\
&= \boldsymbol{w}_1^T \lambda \boldsymbol{w}_1 \\
&= \lambda
\end{aligned} \tag{20}$$

を得る．これより，第 1 主成分 z_1 の分散 $\sigma_{z_1}^2$ は共分散行列 $\boldsymbol{V}$ の固有値 λ に等しいことがわかる．従って，最大にすべき第 1 主成分 z_1 の分散 $\sigma_{z_1}^2$ は共分散行列 $\boldsymbol{V}$ の最大固有値に等しくなり，その結合係数 $\boldsymbol{w}_1$ は最大固有値に対応する固有ベクトルとして求めることができる．

2.3 第 m 主成分の導出

第 2 主成分以下の結合係数 $\{\boldsymbol{w}_m\}(m=2,3,\cdots,M)$ も第 1 主成分と同様な手順で求めることができる．ただし，第 m 主成分 $z_m(m=2,\cdots,M)$ の分散 $\sigma^2_{z_m}$ が主成分 $\{z_{m'}\}(m'=1,2,\cdots,m-1)$ の全てと無相関な 1 次式の持つ分散の中で最大となるように，結合係数 $\boldsymbol{w}_m$ を決定しなければならない．当然ながら，Eq.(3) も満たされなければならない．

ここでは，数学的帰納法を用いて，第 m 主成分の結合係数 $\boldsymbol{w}_m$ を導出する方法を示す．いま，第 $m-1$ 主成分まで求められており，それらの結合係数 $\{\boldsymbol{w}_i\}(i=1,2,\cdots,m-1)$ は条件

$$
\begin{aligned}
(\boldsymbol{V}-\lambda_i\boldsymbol{I})\,\boldsymbol{w}_i &= \boldsymbol{0} && (21)\\
\boldsymbol{w}_i^T\boldsymbol{w}_j &= \begin{cases} 1 & (i=j) \\ 0 & (i\neq j) \end{cases} && (22)
\end{aligned}
$$

を満たしているとして，第 m 主成分の結合係数 $\boldsymbol{w}_m$ を Lagrange 乗数法を用いて求める．すなわち，

$$
J_m = \boldsymbol{w}_m^T\boldsymbol{V}\boldsymbol{w}_m - \lambda_m\left(\boldsymbol{w}_m^T\boldsymbol{w}_m-1\right) - \sum_{i=1}^{m-1}\mu_i\boldsymbol{w}_m^T\boldsymbol{w}_i \tag{23}
$$

とおいて，J_m を最大にするような $\boldsymbol{w}_m$ を求めることにする．ここで，右辺最終項が第 m 主成分 $z_m(m=2,\cdots,M)$ と主成分 $\{z_{m'}\}(m'=1,2,\cdots,m-1)$ の全てが無相関となる条件を表している．J_m を $\boldsymbol{w}_m$ で偏微分して $\boldsymbol{0}$ とおくと，

$$
\begin{aligned}
\frac{\partial J_m}{\partial \boldsymbol{w}_m} &= 2\boldsymbol{V}\boldsymbol{w}_m - 2\lambda_m\boldsymbol{w}_m - \sum_{i=1}^{m-1}\mu_i\boldsymbol{w}_i \\
&= \boldsymbol{0} && (24)
\end{aligned}
$$

となる．$\boldsymbol{w}_j^T(j=1,2,\cdots,m-1)$ を左から掛けて，Eq.(22) を用いると，

$$
\boldsymbol{w}_j^T\boldsymbol{V}\boldsymbol{w}_m - \mu_j = 0 \quad (j=1,2,\cdots,m-1) \tag{25}
$$

を得る．ここで，左辺第 1 項は，$\boldsymbol{V}=\boldsymbol{V}^T$ および Eq.(21) より

$$
\begin{aligned}
\boldsymbol{w}_j^T\boldsymbol{V}\boldsymbol{w}_m &= \boldsymbol{w}_m^T\boldsymbol{V}\boldsymbol{w}_j \\
&= \boldsymbol{w}_m^T\lambda_j\boldsymbol{w}_j \\
&= 0 \quad (j=1,2,\cdots,m-1) && (26)
\end{aligned}
$$

となることから，Eq.(25) は

$$
\mu_j = 0 \quad (j=1,2,\cdots,m-1) \tag{27}
$$

となる．この結果を Eq.(24) に代入すると，最終的に

$$
(\boldsymbol{V}-\lambda_m\boldsymbol{I})\,\boldsymbol{w}_m = \boldsymbol{0} \tag{28}
$$

を得る．これは Eq.(21) と同一の式であり，第 m 主成分 z_m の分散 $\sigma^2_{z_m}$ もまた共分散行列 $\boldsymbol{V}$ の固有値に等しいことがわかる．ただし，大きい方から $m-1$ 個の固有値とそれに対応する固有ベクトルはすでに第 $m-1$ 番目までの主成分を表すのに使われているため，第 m 主成分 z_m の分散 $\sigma^2_{z_m}$ は共分散行列 $\boldsymbol{V}$ の m 番目に大きい固有値に等しくなり，結合係数 $\boldsymbol{w}_m$ はその固有値に対応する固有ベクトルとして求めることができる．

2.4 データの標準化

ここまでは，各変数の測定値をそのまま用いて，その共分散行列 $\boldsymbol{V}$ に基づいて主成分を決定する方法について述べた．しかし，各変数が異なる単位で測定されている場合には，単位の取り方によって異なる主成分が得られることになる．また，たとえ単位が同じであったとしても，大きく分散の異なる変数に対してそのまま主成分分析を適用すれば，その結果は分散の大きな変数の影響を強く受けることになり，変数間の関係を正しく把握することはできないであろう．従って，全ての変数を何らかの方法を用いて標準化する必要がある．

最も簡単で広く利用されている方法は，各変数を平均1，分散1となるように標準化する方法である．具体的には，P 個の変数について N 個のサンプルがある場合，その測定値 $x^*_{np}(n=1,2,\cdots,N; p=1,2,\cdots,P)$ を用いる代わりに，

$$\tilde{x}_{np} = \frac{x^*_{np} - \bar{x}_p}{\sigma_{x_p}} \tag{29}$$

を用いる．ここで，$\bar{x}_p, \sigma_{x_p}$ はそれぞれ p 番目の変数 x_p の平均値および標準偏差である．この場合，標準化された変数間の共分散は相関係数に等しくなるので，標準化されたデータ行列 $\tilde{\boldsymbol{X}}$ の相関行列を

$$\boldsymbol{R} = \frac{1}{N-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}} \tag{30}$$

とおくと，第 m 主成分 z_m の分散 $\sigma^2_{z_m}$ は相関行列 $\boldsymbol{R}$ の m 番目に大きな固有値に等しくなり，第 m 主成分 z_m の結合係数 $\boldsymbol{w}_m$ はその固有値に対応する固有ベクトルとして求めることができる．

上記の標準化を施すことによって，単位に依存しない分析を行うことが可能になる．しかし，実際の測定データは大きさや性質の異なる様々な誤差を含むため，誤差の影響を強く受けている変数を用いた分析は誤った結果を導く恐れがある．そこで，誤差の影響の程度に応じて各変数に異なる重みを付けるという方法が考えられる．重みの付け方としてはいろいろ考えられるが，誤差の影響を全く受けていない変数は全て分散が 1 となるように，誤差のみによって変動している変数についてはその分散が 0 となるようにするのが直感的にも妥当であろう．このような重みの付け方として，測定値 $x^*_{np}(n=1,2,\cdots,N; p=1,2,\cdots,P)$ の代わりに，

$$\tilde{x}'_{np} = \frac{x^*_{np} - \bar{x}_p}{\sigma_{x_p}} \frac{\sigma_{x_p} - \sigma_{x_{pe}}}{\sigma_{x_p}} \tag{31}$$

を用いる方法がある．ここで，$\sigma_{x_{pe}}$ は p 番目の変数 x_p に含まれる誤差の標準偏差である．現実には，誤差の標準偏差を正確に把握することは困難であるが，何らかの方法でその推定値が与えられている場合には，ここで述べたような重み付けが有効となる．

3 寄与率と因子負荷量

3.1 寄与率

主成分分析とは少数の総合的指標 (主成分) を用いて変数間の関係や特徴を把握するための統計的手法である．従って，各主成分が元のデータに含まれる特徴をどの程度表現しているのか，あるいは何個の主成分を採用すれば元のデータに含まれる特徴を十分に表現できるのかを知ることが必要になる．このための指標として，寄与率および累積寄与率がある．

合計 P 個の変数の分散の和は，その共分散行列を $\boldsymbol{V}$ とすると，$\boldsymbol{V}$ の (p,p) 要素 v_{pp} が変数 x_p の分散に等しくなるので，

$$\begin{aligned}\sum_{p=1}^{P}\sigma^2_{x_p} &= \sum_{p=1}^{P} v_{pp} \\ &= tr(\boldsymbol{V})\end{aligned} \tag{32}$$

で与えられる．一方，第 m 主成分の分散 $\sigma_{z_m}^2$ は共分散行列 $\boldsymbol{V}$ の m 番目に大きな固有値 λ_m に等しいので，合計 P 個の主成分の分散の和は

$$\begin{aligned} \sum_{p=1}^{P} \sigma_{z_p}^2 &= \sum_{p=1}^{P} \lambda_p \\ &= tr(\boldsymbol{V}) \end{aligned} \tag{33}$$

で与えられる．すなわち，変数の分散の総和と主成分の分散の総和とは等しくなる．そこで，第 m 主成分が元のデータに含まれる特徴をどの程度表現しているかを示す指標として，第 m 主成分の分散が分散の総和に占める割合

$$C_m = \frac{\lambda_m}{\sum_{p=1}^{P} \lambda_p} = \frac{\lambda_m}{tr(\boldsymbol{V})} \tag{34}$$

を利用することができる．C_m は**寄与率** (proportion) と呼ばれ，通常%を用いて表される．また，第 m 主成分までの分散の和が分散の総和に占める割合

$$P_m = \sum_{i=1}^{m} C_i = \frac{\sum_{i=1}^{m} \lambda_i}{tr(\boldsymbol{V})} \tag{35}$$

は**累積寄与率** (accumulated proportion, cumulative proportion) と呼ばれる．なお，各変数を予め分散 1 に標準化してある場合には，共分散行列 $\boldsymbol{V}$ は相関行列 $\boldsymbol{R}$ に等しくなるため，相関行列を用いて寄与率および累積寄与率を求めることもできる．

上述の累積寄与率を用いて採用する主成分数を決定する場合，通常，累積寄与率が 80%程度となるように主成分数が決められる．しかし，80%という数値に深い意味はなく，常にこの基準が妥当であるとは限らない．従って，累積寄与率を参考にしながら，主成分分析の目的や適用対象に応じて採用する主成分数を適切に決定する必要がある．

各変数の分散を予め 1 に標準化する場合，すなわち相関行列に基づいて主成分分析を行う場合には，元の変数が持つ分散よりも大きな分散を持つ主成分のみを採用するという考え方に従い，大きさが 1 以上の固有値に対応する主成分を全て採用するという基準が利用されることもある．この場合，採用するべき主成分の個数は大きさが 1 以上の固有値の数として自動的に与えられることになる．

3.2 因子負荷量

主成分分析の結果を解釈するということは，主成分 (総合的指標) の持つ意味を解釈するということである．主成分は各変数の線形結合として与えられるので，その解釈のためには，主成分と各変数との相関を把握することによって主成分に強く影響を及ぼす変数を特定することが有効である．このための指標が**因子負荷量** (factor loading) であり，主成分と変数との相関係数として定義される．すなわち，第 m 主成分 z_m と p 番目の変数 x_p との間の因子負荷量は

$$r_{z_m x_p} = \frac{\sigma_{z_m x_p}^2}{\sigma_{z_m} \sigma_{x_p}} \tag{36}$$

で与えられる．ここで，$\sigma_{z_m}, \sigma_{x_p}$ はそれぞれ z_m, x_p の標準偏差であり，$\sigma^2_{z_m x_p}$ は共分散を表す．いま，サンプル数が N であるとすると，z_m, x_p の要素はそれぞれ

$$z_m \quad : \quad \begin{pmatrix} t_{1m} \\ t_{2m} \\ \vdots \\ t_{Nm} \end{pmatrix} = \boldsymbol{X}\boldsymbol{w}_m \tag{37}$$

$$x_p \quad : \quad \begin{pmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{Np} \end{pmatrix} = \boldsymbol{X}\boldsymbol{c}^{(p)} \tag{38}$$

で与えられる．ここで，$\boldsymbol{c}^{(p)}$ は第 p 列のみを取り出すためのベクトルであり，

$$\boldsymbol{c}^{(p)} \quad = \quad \begin{pmatrix} 0^{(1)} \\ \vdots \\ 1^{(p)} \\ \vdots \\ 0^{(P)} \end{pmatrix} \tag{39}$$

とする．ただし，$\boldsymbol{c}^{(p)}$ の各要素の右肩にある () 内の番号は行番号を表している．このとき，z_m と x_p の共分散は

$$\begin{aligned} \sigma^2_{z_m x_p} &= \frac{1}{N-1}(\boldsymbol{X}\boldsymbol{w}_m)^T \boldsymbol{X}\boldsymbol{c}^{(p)} \\ &= \frac{1}{N-1}\boldsymbol{w}_m^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{c}^{(p)} \\ &= (\boldsymbol{V}\boldsymbol{w}_m)^T \boldsymbol{c}^{(p)} \\ &= \lambda_m \boldsymbol{w}_m^T \boldsymbol{c}^{(p)} \\ &= \lambda_m w_{pm} \end{aligned} \tag{40}$$

となる．従って，因子負荷量は，$\sigma^2_{z_m} = \lambda_m$ より

$$\begin{aligned} r_{z_m x_p} &= \frac{\lambda_m w_{pm}}{\sqrt{\lambda_m}\sigma_{x_p}} \\ &= \frac{\sqrt{\lambda_m} w_{pm}}{\sigma_{x_p}} \end{aligned} \tag{41}$$

で与えられる．なお，各変数の分散が予め 1 に標準化されている場合には，因子負荷量は

$$r_{z_m x_p} \quad = \quad \sqrt{\lambda_m} w_{pm} \tag{42}$$

となる．

4 主成分分析の行列表現

4.1 主成分得点の計算

P 個の変数について N 個のサンプルを含む，適当に標準化されたデータ行列 $\boldsymbol{X}$ があり，その共分散行列の固有値が大きい方から順に $\lambda_i (i = 1, 2, \cdots, P)$ で与えられているとする．共分散行列は非負定値行列で

あるから，その固有値は必ずゼロ以上の値をとる．いま，

$$\lambda_i \begin{cases} > 0, & \text{for } i = 1, 2, \cdots, r \\ = 0, & \text{for } i = r+1, r+2, \cdots, P \end{cases} \tag{43}$$

であるとすると，$r+1$ 番目以降の固有値はいずれもゼロであるから，それらに対応する主成分の分散もゼロとなる．すなわち，$r+1$ 番目以降の主成分の方向にはデータは分散を持たない．これは，元々P 次元空間に含まれているサンプル $\{\boldsymbol{x}_n\}(n = 1, 2, \cdots, N)$ が実は高々r 次元部分空間に含まれることを意味する．従って，実際に得られる主成分の数はゼロでない固有値の数 (r 個) に等しくなり，ゼロでない固有値に対応する規格化された固有ベクトル $\{\boldsymbol{w}_i\}(i = 1, 2, \cdots, r)$ が r 次元部分空間の正規直交基底となる．

この場合，i 番目に大きな固有値に対応する固有ベクトル $\boldsymbol{w}_i$ が第 i 主成分の結合係数となるので，第 i 主成分 z_i は

$$z_i \quad = \quad \sum_{p=1}^{P} w_{pi} x_p \tag{44}$$

となり，第 i 主成分得点 $\boldsymbol{t}_i$ は

$$\boldsymbol{t}_i \quad = \quad \boldsymbol{X}\boldsymbol{w}_i \tag{45}$$

で与えられる．ここで，第 $M(M \leq P)$ 主成分までを考慮し，主成分得点からなる行列 (score matrix) を

$$\boldsymbol{T} \quad = \quad \begin{pmatrix} \boldsymbol{t}_1 & \boldsymbol{t}_2 & \cdots & \boldsymbol{t}_M \end{pmatrix} \tag{46}$$

とおき，さらに，主成分の結合係数からなる行列 (loading matrix) を

$$\boldsymbol{P} \quad = \quad \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_M \end{pmatrix} \tag{47}$$

とおくと，

$$\boldsymbol{T} \quad = \quad \boldsymbol{X}\boldsymbol{P} \tag{48}$$

を得る．

新しく得られたサンプルが既に得られているサンプルと比較してどのような特徴を持つかを知りたい場合，既存のサンプルから得られる主成分に基づいて新しいサンプルの主成分得点を求め，各サンプルの主成分得点を比較することが有効である．この際，主成分得点の計算は以下の手順で行われる．

1. 既存のサンプルの loading matrix $\boldsymbol{P}$ を求める．
2. 既存のサンプルの主成分得点を Eq.(48) を用いて計算する．
3. 新しく得られたサンプル $\boldsymbol{x}_{new}$ の主成分得点 $\boldsymbol{t}_{new}$ を次式を用いて計算する．

$$\boldsymbol{t}_{new} \quad = \quad \boldsymbol{x}_{new}\boldsymbol{P} \tag{49}$$

既存のサンプルの主成分得点に対する新しいサンプルの主成分得点の特徴 (大きい，平均値 0 に近いなど) を各主成分について検討することにより，新しいサンプルの特徴を把握することができる．例えば，Sec.1 で述べた身体検査の例では，第 1 主成分が「体の大きさ」を，第 2 主成分が「肥満の程度」を表しているので，新しいサンプルの第 1 主成分得点が大きく，第 2 主成分得点が小さければ，その人は体が大きいが痩せていることがわかる．この例は高々 2 変数の場合であるので，主成分による特徴抽出の効果はほとんど無いに等しい．しかし，プロセスデータのような多変数データを扱う場合には，生データから特徴を把握することは一般に困難であり，主成分分析による特徴抽出が威力を発揮するであろう．実際に，運転状態の監視や異常検出，さらには異常診断といった分野に主成分分析が利用されている．

4.2 データの再構築

第 1 主成分得点は P 次元空間内の各サンプル $\boldsymbol{x}_n (n = 1, 2, \cdots, N)$ を第 1 主成分を表す z_1 軸に射影して得られるベクトルの長さであるから，あるサンプル $\boldsymbol{x}_n$ を第 1 主成分のみで代表させる場合，その予測値 $\hat{\boldsymbol{x}}_n$ は

$$\begin{aligned} \hat{\boldsymbol{x}}_n &= t_{n1}\boldsymbol{w}_1^T \\ &= \boldsymbol{x}_n \boldsymbol{w}_1 \boldsymbol{w}_1^T \end{aligned} \tag{50}$$

で与えられる．ここで，t_{n1} はサンプル $\boldsymbol{x}_n$ の第 1 主成分得点，$\boldsymbol{w}_1$ は第 1 主成分の結合係数である．サンプル $\boldsymbol{x}_n$ を第 M 主成分までの全てを用いて表現する場合には，その予測値 $\hat{\boldsymbol{x}}_n$ は各主成分方向の成分の線形結合として

$$\hat{\boldsymbol{x}}_n = \sum_{m=1}^{M} t_{nm}\boldsymbol{w}_m^T \tag{51}$$

で与えられる．簡単な例として，$P = 3, M = 2$ の場合にデータが再構築される様子を Fig.2 に示しておく．Eq.(51) は loading matrix $\boldsymbol{P}$ を用いて

$$\begin{aligned} \hat{\boldsymbol{x}}_n &= \begin{pmatrix} t_{n1} & t_{n2} & \cdots & t_{nM} \end{pmatrix} \boldsymbol{P}^T \\ &= \boldsymbol{x}_n \boldsymbol{P}\boldsymbol{P}^T \end{aligned} \tag{52}$$

と表すこともできる．これより，N 個のサンプル全体については，

$$\hat{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{P}\boldsymbol{P}^T \tag{53}$$

が成り立つ．主成分数 M が共分散行列 (あるいは相関行列) のゼロでない固有値の数に等しい場合には，$\hat{\boldsymbol{X}}$ は元のデータ行列 $\boldsymbol{X}$ と正確に一致する．しかし，それ以外の場合には $\hat{\boldsymbol{X}}$ は $\boldsymbol{X}$ と一致しない．このとき，元データと再構築データの残差行列 $\boldsymbol{E}$ は

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{X} - \hat{\boldsymbol{X}} \\ &= \boldsymbol{X} - \boldsymbol{T}\boldsymbol{P}^T \\ &= \boldsymbol{X} - \boldsymbol{X}\boldsymbol{P}\boldsymbol{P}^T \\ &= \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^T) \end{aligned} \tag{54}$$

となる．

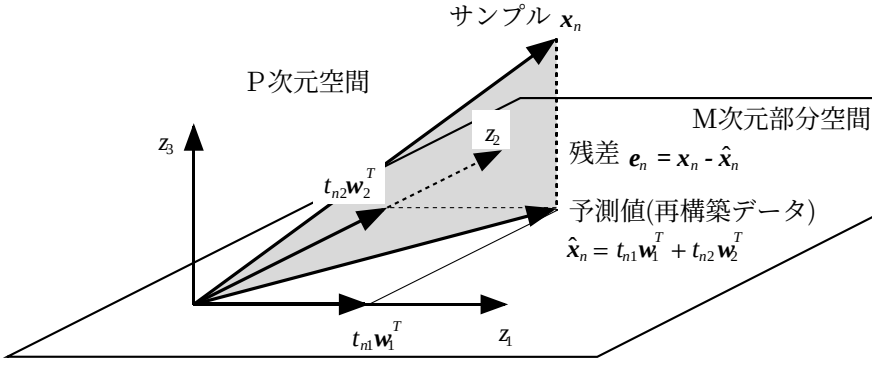

Fig. 2　データの再構築 ($P = 3, M = 2$ の場合)

4.3 主成分分析と特異値分解

ここまで，主成分が固有値や固有ベクトルと深く関係していることをみてきた．その関係とは，適当に標準化されたデータ行列が与えられたとき，その共分散行列の固有ベクトルが主成分の結合係数になり，固有値が主成分の分散に等しくなるというものであった．

P 個の変数について N 個のサンプルからなるデータ行列 $\boldsymbol{X}$ が与えられている場合，その共分散行列の i 番目に大きな固有値を λ_i，それに対応する規格化された固有ベクトルを $\boldsymbol{w}_i$ とすると，次式が成り立つ．

$$\frac{1}{N-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}_i = \lambda_i\boldsymbol{w}_i \quad (i=1,2,\cdots,P) \tag{55}$$

ここで，$(N-1)\lambda_i$ を改めて λ_i とおくと，

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}_i = \lambda_i\boldsymbol{w}_i \quad (i=1,2,\cdots,P) \tag{56}$$

を得る．この場合，λ_i は行列 $\boldsymbol{X}^T\boldsymbol{X}$ の i 番目に大きな固有値，$\boldsymbol{w}_i$ はそれに対応する固有ベクトルとなる．ここで，loading matrix を

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_P \end{pmatrix} \tag{57}$$

とおくと，$\boldsymbol{P}$ が正規直交行列であることから，

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{P}\boldsymbol{P}^T \tag{58}$$

が成り立つ．さらに，

$$\sigma_i = \sqrt{\lambda_i} \quad (i=1,2,\cdots,P) \tag{59}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_P \end{pmatrix} \tag{60}$$

$$\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{X}\boldsymbol{w}_i \quad (i=1,2,\cdots,P) \tag{61}$$

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_P \end{pmatrix} \tag{62}$$

とおくと，

$$\boldsymbol{X}\boldsymbol{P} = \boldsymbol{U}\boldsymbol{\Sigma} \tag{63}$$

が成り立つので，これを Eq.(58) に代入すると，最終的に

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{P}^T \tag{64}$$

を得る．ここで，行列 $\boldsymbol{\Sigma}$ は対角行列であり，行列 $\boldsymbol{U}$ および $\boldsymbol{P}$ は正規直交行列であることから，Eq.(64) は，行列 $\boldsymbol{X}$ の特異値分解に他ならないことがわかる．なお，行列 $\boldsymbol{P}$ が正規直交行列であることは，その列が規格化された固有ベクトルであり，Eq.(22) を満たしていることから明らかである．一方，行列 $\boldsymbol{U}$ が正規直交行列であることは，行列 $\boldsymbol{U}^T\boldsymbol{U}$ の各要素を考えることにより容易に示すことができる．すなわち，行列 $\boldsymbol{U}^T\boldsymbol{U}$ の (i,j) 要素を $(\boldsymbol{U}^T\boldsymbol{U})_{ij}$ とすると，

$$\begin{aligned} (\boldsymbol{U}^T\boldsymbol{U})_{ij} &= \frac{1}{\sigma_i}(\boldsymbol{X}\boldsymbol{w}_i)^T \cdot \frac{1}{\sigma_j}\boldsymbol{X}\boldsymbol{w}_j \\ &= \frac{1}{\sigma_i\sigma_j}\boldsymbol{w}_i^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}_j \end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda_j}{\sigma_i \sigma_j} \boldsymbol{w}_i^T \boldsymbol{w}_j \\
&= \begin{cases} 1 & (i \neq j) \\ 0 & (i = j) \end{cases}
\end{aligned} \tag{65}
$$

となるので，行列 $\boldsymbol{U}$ は正規直交行列である．

行列 $\boldsymbol{X}$ の特異値分解である Eq.(64) は，ゼロでない固有値の数を r とすると，

$$
\boldsymbol{X} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{w}_i^T \tag{66}
$$

と表現することもできる．さらに，

$$
\boldsymbol{u}_i^T \boldsymbol{u}_j = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{67}
$$

$$
\boldsymbol{w}_i^T \boldsymbol{w}_j = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{68}
$$

であることから，

$$
\boldsymbol{X}^T \boldsymbol{X} = \sum_{i=1}^{r} \lambda_i \boldsymbol{w}_i \boldsymbol{w}_i^T \tag{69}
$$

$$
\boldsymbol{X} \boldsymbol{X}^T = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T \tag{70}
$$

となる．これらの式はそれぞれ非負定値行列 $\boldsymbol{X}^T \boldsymbol{X}$ および $\boldsymbol{X} \boldsymbol{X}^T$ のスペクトル分解と呼ばれるものであり，$\boldsymbol{X}^T \boldsymbol{X}$ と $\boldsymbol{X} \boldsymbol{X}^T$ の固有値は等しくなることを示している．これらのスペクトル分解を行列を用いて表現すると，

$$
\boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{P} \boldsymbol{\Sigma}^2 \boldsymbol{P}^T \tag{71}
$$

$$
\boldsymbol{X} \boldsymbol{X}^T = \boldsymbol{U} \boldsymbol{\Sigma}^2 \boldsymbol{U}^T \tag{72}
$$

となる．

以上より，主成分分析は特異値分解およびスペクトル分解と密接に関係していることがわかる．その関係を簡単にまとめると，

1. 特異値分解に基づく場合
   第 i 主成分の結合係数 $\boldsymbol{w}_i$ は，データ行列 $\boldsymbol{X}$ の i 番目に大きな特異値 σ_i に対応する右特異ベクトルとして与えられる．
2. スペクトル分解に基づく場合
   第 i 主成分の結合係数 $\boldsymbol{w}_i$ は，行列 $\boldsymbol{X}^T \boldsymbol{X}$ の i 番目に大きな固有値 λ_i に対応する固有ベクトルとして与えられる．

となる．