特異値分解による時系列データの主成分分析

千葉豪

平成 28 年 9 月 21 日

文献 [1] に述べられている、特異値分解を用いた時系列データの主成分分析について、以下、簡単にまとめる。

N_d 個の観測時系列データ（$x(t_i), i = 1, 2, \cdots, N_d$）が与えられたとき、系の動力学的性質は遅れ座標系で再構築された運動軌道の幾何学的性質を解析することで調べることができる。埋め込み空間次元を n とし、その空間内の運動軌道の各点を時系列データから以下のように与える。

$$\mathbf{x}_i^T = [x(t_i), x(t_{i+1}), \cdots, x(t_{i+n-1})] \tag{1}$$

上式を以下のように記述すれば、後の理解も容易となるであろう。

$$\begin{array}{lcllllllll}
\mathbf{x}_1^T & = & [x(t_1) & x(t_2) & \cdots & x(t_n)] & & & & \\
\mathbf{x}_2^T & = & & [x(t_2) & x(t_3) & \cdots & x(t_{n+1})] & & & \\
\mathbf{x}_3^T & = & & & [x(t_3) & x(t_4) & \cdots & x(t_{n+2})] & & \\
\vdots & & & & & & & & & \\
\mathbf{x}_N^T & = & & & & [x(t_N) & x(t_{N+1}) & \cdots & x(t_{N_d})] &
\end{array} \tag{2}$$

ここで、$N = N_d - (n-1)$ である。Takens の埋め込み定理によると、このように遅れ時間座標系で表された運動軌道は、時系列データを生み出した現実の位相空間上の運動軌道と位相同系な軌道となる。

ここで、$\mathbf{x}_i$ で構成される行列 $\mathbf{X}$ を

$$\mathbf{X} = \frac{1}{\sqrt{N}} (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) \tag{3}$$

と定義し、さらに共分散行列 $\mathbf{C}$ を次のように定義する。

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T = \frac{1}{N} (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k \mathbf{x}_k^T \tag{4}$$

なお、この共分散行列 $\mathbf{C}$ の (i,j) 要素 c_{ij} は

$$c_{ij} = \frac{1}{N} \sum_{k=1}^{N} x(t_{k+i-1}) x(t_{k+j-1}) \tag{5}$$

と書ける。すなわち c_{ij} は、「$\mathbf{x}_k$ の i 番目と j 番目の積を全ての k について和をとったもの」「運動軌道の i 番目の要素と j 番目の要素の共分散（相関）に対応するもの」と理解することが出来る。

行列 $\mathbf{C}$ の固有値を σ_i^2、対応する固有ベクトルを $\mathbf{v}_i$ と書くとすると、以下の方程式が成り立つ。

$$\mathbf{CV} = \mathbf{V\Sigma} \tag{6}$$

ここで、$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$、$\mathbf{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$ であり、$\sigma_1^2 > \sigma_2^2 > \cdots > \sigma_n^2$ とする。この固有ベクトル $\mathbf{v}_i$ は正規直交系を構成する。すなわち、

$$\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij} \tag{7}$$

である。ここで、δ_{ij} はクロネッカーのデルタを示す。

なお、σ_i は行列 $\mathbf{X}$ の特異値に、$\mathbf{v}_i$ は左特異ベクトルに対応する。すなわち、

$$\mathbf{X} = \mathbf{UDV}^T, \qquad \mathbf{D} = diag\left(\sigma_1, \sigma_2, \cdots, \sigma_n\right) \tag{8}$$

なので、

$$\mathbf{C} = \mathbf{XX}^T = \mathbf{UDV}^T\mathbf{VDU}^T = \mathbf{UD}^2\mathbf{U}^T \tag{9}$$

である。

式 (6) より

$$\mathbf{V}^T\mathbf{CV} = \mathbf{\Sigma} \tag{10}$$

なので、式 (4) を用いると

$$\mathbf{V}^T\left(\mathbf{XX}^T\right)\mathbf{V} = \mathbf{V}^T\mathbf{X}\cdot\mathbf{X}^T\mathbf{V} = \left(\mathbf{X}^T\mathbf{V}\right)^T\left(\mathbf{X}^T\mathbf{V}\right) = \mathbf{\Sigma} \tag{11}$$

となり

$$\begin{pmatrix} \left(\mathbf{X}^T\mathbf{v}_1\right)^T \\ \left(\mathbf{X}^T\mathbf{v}_2\right)^T \\ \vdots \\ \left(\mathbf{X}^T\mathbf{v}_n\right)^T \end{pmatrix} \left(\mathbf{X}^T\mathbf{v}_1, \mathbf{X}^T\mathbf{v}_2, \cdots, \mathbf{X}^T\mathbf{v}_n\right) = \mathbf{\Sigma} \tag{12}$$

すなわち

$$\left(\mathbf{X}^T\mathbf{v}_i\right)^T \cdot \left(\mathbf{X}^T\mathbf{v}_j\right) = \sigma_i^2\delta_{ij} \tag{13}$$

と書くことが出来る。ベクトル $\mathbf{X}^T\mathbf{v}_i$ は

$$\mathbf{X}^T\mathbf{v}_i = \frac{1}{\sqrt{N}}\begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \cdots \\ \mathbf{x}_n^T \end{pmatrix}\mathbf{v}_i = \frac{1}{\sqrt{N}}\begin{pmatrix} \mathbf{x}_1^T\mathbf{v}_i \\ \mathbf{x}_2^T\mathbf{v}_i \\ \cdots \\ \mathbf{x}_n^T\mathbf{v}_i \end{pmatrix} \tag{14}$$

と書けるが、これより、このベクトルの k 成分は $\mathbf{x}_k$ の $\mathbf{v}_i$ への写像成分（$\mathbf{x}_k$ の $\mathbf{v}_i$ 成分）に対応していること、また、式 (13) より、運動軌道の $\mathbf{v}_i$、$\mathbf{v}_j$ への写像は互いに直交することが分かる。埋め込み空間次元が 2 のときのイメージ図を**図 1** に示す。運動軌道点 $\mathbf{x}$ が二次元平面上の楕円を周回するとした場合、その軌道点の $\mathbf{v}_1$、$\mathbf{v}_2$ の写像の時系列データ（下図）は、お互いに直交することになる。

さて、埋め込み空間内の任意の運動軌道点 $\mathbf{x}_i$ は $\mathbf{v}_j$ を用いて以下のように記述できる。

$$\mathbf{x}_i = \sum_j \left(\mathbf{v}_j^T\mathbf{x}_i\right)\mathbf{v}_j \tag{15}$$

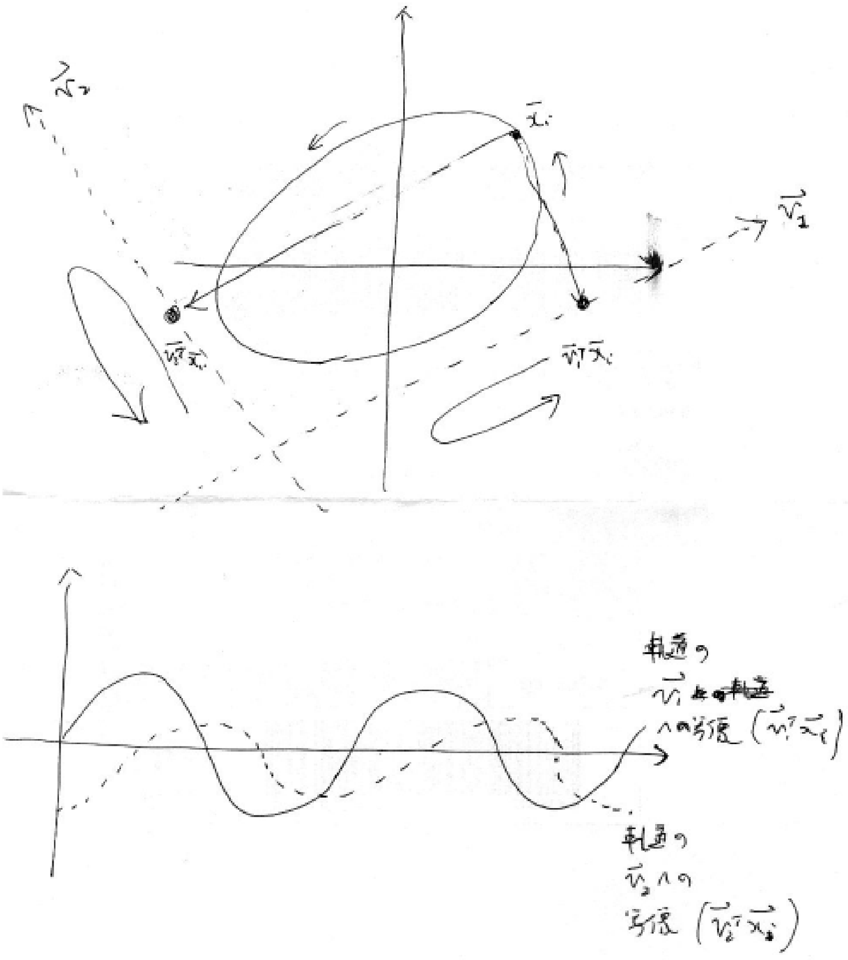

図 1: イメージ図

従って、ベクトル $\mathbf{x}_i$ の最初の要素である $x(t_i)$ は以下のように記述することができる。[1] [2]

$$x(t_i) = \sum_j \left(\mathbf{v}_j^T \mathbf{x}_i\right) (\mathbf{v}_j)_1 = \sum_j \varphi_{ji} (\mathbf{v}_j)_1 \tag{18}$$

以上のように、時系列データ $x(t_i)$ は $\varphi_{ji} (\mathbf{v}_j)_1$ のように成分分解することが出来るが、成分 i と

[1] $x(t_i)$ はここで示した以外の方法でも計算することが可能である。例えば、$i < n$ であれば

$$x(t_i) = \sum_j \left(\mathbf{v}_j^T \mathbf{x}_1\right) (\mathbf{v}_j)_i \tag{16}$$

としても求めることが出来るはずである。ここで、例えば式 (18) で $(\mathbf{v}_j)_1 = 0$ のときを考えてみよう。このとき、j 成分は時系列データに含まれないことになる。一方、式 (16) に基づいた場合には、$(\mathbf{v}_j)_1 = 0$ であっても j 成分の全てがゼロとなることは無い。このことは、時系列データを特異値分解によって成分毎に分ける方法には任意性があることを示している。

[2] なお、$i > N$ については

$$x(t_{N+k}) = \sum_j \left(\mathbf{v}_j^T \mathbf{x}_N\right) (\mathbf{v}_j)_{k+1} \tag{17}$$

として求めてやればよいであろう。

成分 j の内積は以下のように書ける。

$$\sum_{k=1}^{N}\varphi_{ik}\left(\mathbf{v}_i\right)_1\cdot\varphi_{jk}\left(\mathbf{v}_j\right)_1 = \left(\mathbf{v}_i\right)_1\left(\mathbf{v}_j\right)_1\sum_{k=1}^{N}\varphi_{ik}\varphi_{jk} = \left(\mathbf{v}_i\right)_1\left(\mathbf{v}_j\right)_1\sum_{k=1}^{N}\left(\mathbf{v}_i^T\mathbf{x}_k\right)\left(\mathbf{v}_j^T\mathbf{x}_k\right) \quad (19)$$

$$= \left(\mathbf{v}_i\right)_1\left(\mathbf{v}_j\right)_1\cdot N\sigma_i^2\delta_{ij} \quad (20)$$

従って、

$$\sum_{k=1}^{N}\left(\varphi_{ik}\left(\mathbf{v}_i\right)_1\right)^2 = \left(\mathbf{v}_i\right)_1^2\cdot N\sigma_i^2 \quad (21)$$

となり、時系列データ中に含まれる i 成分の強度は $\sigma_i^2\left(\mathbf{v}_i\right)_1^2$ に比例することが分かる。また、$i\neq j$ ならば、

$$\sum_{k=1}^{N}\varphi_{ik}\left(\mathbf{v}_i\right)_1\cdot\varphi_{jk}\left(\mathbf{v}_j\right)_1 = 0 \quad (22)$$

となり、異なる成分は直交関係にあることが分かる。

参考文献

[1] M. Tsuji, Y. Shimazu, 'Evaluation of Decay Ratio of BWRs using Singular Value Decomposition Method,' *J. Nucl. Sci. Technol.*, **42**, p. 169-182 (2005).