

複素主成分解析の変動データへの応用

北野利一*・間瀬 肇

* 徳島大学工学部建設工学科

要 旨

主成分解析は、複雑なデータの特徴を抽出するために工学のあらゆる分野で広く用いられてきている統計処理の1手法である。しかし、変動データを対象に解析する場合、従来の研究では、固有関数が直交するという性質が強調されすぎており、得られた固有関数の性質が十分に活用されていない。本研究は、変動データの主成分解析法の特徴を整理するとともに、複雑なデータを縮約することにより得られる各軸の単変数の固有関数の有効な活用法について考察するものである。

キーワード：主成分解析(PCA)、経験的固有関数法、海底地形変動解析、Fourier級数

1. はじめに

主成分解析は、複雑なデータの特徴を抽出するために工学のあらゆる分野で広く用いられてきている統計処理の1手法で、別名を経験的固有関数法として知られる。特に地球物理学の分野では、ある物理量の空間および時間の多次元観測データを固有分解して、複雑なデータを数少ない変動関数の重ね合わせで表現し、その物理量の空間および時間変動の特徴を把握するのに利用されている(Wallance and Dickinson, 1972; Wallance, 1972; Barnett, 1983, 他多数)。また、海岸工学においても、湾の形状や海底地形の影響を受けた現地波浪観測データからのエッジ波の抽出(例えば、加藤ら, 1980)、複雑な深淺測量データからの海底地形変化の実態把握(Aubrey, 1979; Aubrey et al., 1980; 加藤・吉松, 1984 他)、スペクトルの発達特性の解析(Vincent and Resio, 1997)等の主成分解析の応用例がある。また、近年では、Liang and Seymour (1991)、Bosma and Dalrymple (1996) および横木ら(1998)は、現地データを用いて、複素値による主成分解析を行い、海底地形変動の進行波成分の抽出を検討している。また、橋本ら(1995)は、カルマンフィルターを用いて

気圧情報から波浪統計予測を行う際に、主成分解析から得られる気圧データの固有関数を予測式の説明変数に選ぶことにより、線形回帰の多重共線性をうまく回避している。

主成分解析に関する従来の研究では、変動データから固有関数に分解する手順やその可能性に主眼がおかれ、また固有関数が直交するという性質が強調されすぎているように思われる。それに比べ、橋本ら(1995)は固有関数の直交性を最大限に利用し、有効に活用している点で例外的である。しかし、観測データの特性の考察を主とする他の研究においては、極端に言えば、各軸の固有関数の解析結果を例示しているだけで、寄与率の高い第1成分の固有関数を再合成し、元データとの再現性を確認するといった表面的な考察で終わっている。複雑な変動データの縮約という観点からは、主成分解析が変数分離にもとづいた理論であると認識することが重要であると考ええる。

本研究では、変動データを対象とした主成分解析法の特徴を整理するとともに、複雑なデータを縮約することにより得られる各軸の単変数の固有関数の有効な活用法について考察する。

2. 主成分解析の理論的性質

本研究での理論構成として、主成分解析の骨子を、連続量の関数ベクトル変数と積分表現による内積演算を用いて簡潔に表現する。変動データを内積空間の基底関数により展開することにより、複雑な固有値・固有ベクトルの計算が直感的に理解できるようになり、理論的考察に適すると考える。これらの連続量による表現を用いることにより、変動データを対象にした主成分解析法の幾つかの理論的性質を調べる。

2.1 主成分解析理論の骨子

従来の主成分解析の定式においては、実際の数値データ処理を目的としているため、離散量のベクトルや行列が用いられる。つまり、解析対象となるデータは、離散化された時間や空間上の観測値で、行列で表現されるものであり、そのデータ行列の転置行列とデータ行列とを掛算して得られる相関行列の固有値・固有ベクトルを算出することにより、具体的に主成分解析が行われる。本研究での理論構成上考察するデータは、時間的にも空間的にも有限な区間 $(x, t) \in [0, L] \times [0, T]$ で定義される連続関数 $\eta(x, t)$ とし、次のように主成分解析の理論骨子を定式化する。

空間上での2地点 (x, y) の相関関数を

$$R(x, y) = {}^t \langle \eta(x, t), \eta(y, t) \rangle \quad (1)$$

と定義する。相関関数 R の固有値 λ 、固有ベクトル e は以下の条件を満たすものである。

$$\left. \begin{aligned} {}^x \langle R(x, y), e_p(x) \rangle &= \lambda_p e_p(y), \quad (1 \leq p \leq k) \\ {}^x \langle e_p(x), e_q(x) \rangle &= \delta_{pq} \end{aligned} \right\} \quad (2)$$

空間軸上での固有関数 $e(x)$ が決定できれば、その直交性を利用することにより、時間軸上での固有関数 $c(t)$ は

$$c_p(t) = {}^x \langle e_p(x), \eta(x, t) \rangle \quad (3)$$

により決定される。以上により、元関数 $\eta(x, t)$ は次式のような変数分離された k 個の固有関数を重ね合わせた表現を得る。

$$\eta(x, t) = \sum_{1 \leq p \leq k} e_p(x) c_p(t) \quad (4)$$

以上の元関数の分解は、2時刻での相関関数

$$R(t, s) = {}^x \langle \eta(x, t), \eta(x, s) \rangle \quad (5)$$

についての固有値問題を解くことによって、同様の結果を得ることができる。また、式(1), (2), (3) および (5) で用いられている内積は、次のように定義される。

$$\left. \begin{aligned} {}^t \langle f(t), g(t) \rangle &= T^{-1} \int_0^T \bar{f}(t) g(t) dt \\ {}^x \langle f(x), g(x) \rangle &= L^{-1} \int_0^L \bar{f}(x) g(x) dx \end{aligned} \right\} \quad (6)$$

ここで、関数 f が複素値をとる場合、 $\bar{f}$ は f の共役複素数である。式(6)は、つまるところ、時間軸および空間軸での平均操作である。このことから、主成分解析とは、「時間平均による各地点の相関行列の固有値問題（あるいは、空間平均による各時刻の相関行列の固有値問題）」である。そこで、単一軸の相関関数の考察に帰着させるために、2変数のいずれか一方の軸で「平均」することにより消去するという操作を必要とすることが重要となる。その結果、解析の対象となる関数を変数分離の形式で表現されるということが主成分解析の最大の特徴である。

ところで、主成分解析を成立させるためには、最低2元以上の変数をもつデータを対象としなければならない。なぜなら、上述の主成分解析の理論骨子において、例えば時間変数を固定すると、

$$\left. \begin{aligned} \eta(x) &= \sum_{1 \leq n \leq k} c_n {}^x \langle e_n(x), \eta(x) \rangle \\ {}^x \langle e_n(x), e_m(x) \rangle &= \delta_{nm} \end{aligned} \right\} \quad (7)$$

となり、相関関数による固有値計算は不要となる。従って、式(6)で定義された内積空間の基底関数は、三角関数である（付録1）ので、式(7)はFourier級数展開の定義に変更される。従って、固有関数はデータの特性に基づいたものでなくなる。この式(6)の内積の性質を利用して、解析対象の変動データをあらかじめFourier級数展開しておけば、特定の条件下では、相関関数および固有関数を陽的に表現することが可能になり、理論展開に適すると考えられる。

2.2 実数値解析と複素値解析の違い

本節では、最も基本的な波である定常波と進行波について、実数値で扱った場合と複素値で扱った場合の2種類での主成分解析の計算例を示し、その違いについて考察する。

実数関数の定常波 η_S^t と進行波 η_P^t は以下のように表される。

$$\eta_S^r(x, t) = \cos(kx) \cos(\omega t) \quad (8)$$

$$\eta_P^r(x, t) = \cos(\omega t - kx) \quad (9)$$

それぞれの波の時間平均による空間 2 地点の相関関数は、

$$R_S^r(x, y) = \frac{1}{2} \cos(kx) \cos(ky) \quad (10)$$

$$R_P^r(x, y) = \frac{1}{2} \{ \cos(kx) \cos(ky) + \sin(kx) \sin(ky) \} \quad (11)$$

となる。従って、定常波の相関関数 R_S^r は、

$$^x \langle R_S^r(x, y), \sqrt{2} \cos(kx) \rangle = \frac{1}{4} \times \sqrt{2} \cos(ky) \quad (12)$$

の関係を満たすので、主成分解析の結果は、

$$\begin{aligned} \{ \lambda, e(x), c(t) \} \\ = \left\{ \frac{1}{4}, \sqrt{2} \cos(kx), \frac{1}{\sqrt{2}} \cos(\omega t) \right\} \end{aligned} \quad (13)$$

となる。他方、進行波の相関関数 R_P^r は、

$$\begin{aligned} ^x \langle R_P^r(x, y), \sqrt{2} \cos(kx) \rangle &= \frac{1}{4} \times \sqrt{2} \cos(kx) \\ ^x \langle R_P^r(x, y), \sqrt{2} \sin(kx) \rangle &= \frac{1}{4} \times \sqrt{2} \sin(kx) \end{aligned} \quad (14)$$

および、

$$^x \langle \sqrt{2} \cos(kx), \sqrt{2} \sin(kx) \rangle = 0 \quad (15)$$

の関係を満たす。従って、主成分解析の結果は、

$$\begin{aligned} \{ \lambda, e(x), c(t) \} = \\ \left\{ \frac{1}{4}, \sqrt{2} \cos(kx), \frac{1}{\sqrt{2}} \cos(\omega t) \right\}, \\ \left\{ \frac{1}{4}, \sqrt{2} \sin(kx), \frac{1}{\sqrt{2}} \sin(\omega t) \right\} \end{aligned} \quad (16)$$

となり、進行波 η_P^r は 2 つの固有関数に分離され、1 つの固有関数で表現できないことがわかる。

複素関数の進行波 η_P^c と定常波 η_S^c は、以下のように表される。

$$\eta_P^c(x, t) = \exp i(\omega t - kx) \quad (17)$$

$$\eta_S^c(x, t) = \exp i(\omega t - kx) + \exp i(\omega t + kx) \quad (18)$$

それぞれの波の時間平均による空間 2 地点の相関関数は、

$$R_P^c(x, y) = \exp i k(x - y) \quad (19)$$

$$R_S^c(x, y) = (\exp(iky) + \exp(-iky)) \times (\exp(ikx) + \exp(-ikx)) \quad (20)$$

となる。進行波の相関関数 R_P^c は、

$$^x \langle R_P^c(x, y), \exp(-ikx) \rangle = \exp(-iky) \quad (21)$$

の関係を満たすので、主成分解析の結果は、

$$\begin{aligned} \{ \lambda, e(x), c(t) \} = \\ \{ 1, \exp(-ikx), \exp(i\omega t) \} \end{aligned} \quad (22)$$

となる。他方、定常波の相関関数 R_S^c は、

$$\begin{aligned} ^x \langle R_S^c(x, y), \frac{1}{\sqrt{2}} (\exp(ikx) + \exp(-ikx)) \rangle \\ = 2 \times \frac{1}{\sqrt{2}} (\exp(iky) + \exp(-iky)) \end{aligned} \quad (23)$$

の関係を満たし、主成分解析の結果は、

$$\begin{aligned} \{ \lambda, e(x), c(t) \} = \{ 2, \frac{1}{\sqrt{2}} (\exp(ikx) + \\ \exp(-ikx)), \sqrt{2} \exp(i\omega t) \} \end{aligned} \quad (24)$$

である。以上から、複素値での解析では、定常波も進行波も 1 つの固有関数で表現できるのに対し、実数値解析では、固有関数が定常波しか表現できないため、進行波は 2 つの定常波の和として表現されることがわかる。このことは、複素値と実数値で基底関数が異なり、べき乗数に純虚数をとる exp 関数では $\omega t - kx$ といった位相の移動を表現できる変数に書き換えることができるのに対し、sin および cos 関数では、1 組の空間関数と時間関数の積では実現できないことに起因する。

次に、固有関数の任意性について考察する。式 (10), (11), (19) および (20) のそれぞれの相関係数の固有関数導出においては、以下のような φ および γ の任意性が含まれる。

$$^x \langle R_S^r(x, y), \pm \sqrt{2} \cos(kx) \rangle = \pm \frac{1}{4} \times \sqrt{2} \cos(ky) \quad (25)$$

$$\begin{aligned} ^x \langle R_P^r(x, y), \pm \sqrt{2} \cos(kx + \varphi) \rangle = \\ \pm \frac{1}{4} \times \sqrt{2} \cos(kx + \varphi) \end{aligned} \quad (26)$$

$$^x \langle R_P^c(x, y), \exp i(\gamma - kx) \rangle = \exp i(\gamma - ky) \quad (27)$$

$$\begin{aligned} ^x \langle R_S^c(x, y), \frac{1}{\sqrt{2}} e^{i\gamma} (\exp(ikx) + \exp(-ikx)) \rangle \\ = 2 \times \frac{1}{\sqrt{2}} e^{i\gamma} (\exp(iky) + \exp(-iky)) \end{aligned} \quad (28)$$

ただし、上式における定数 φ および γ は $[0, 2\pi)$ の範囲の任意の実数値である。ここで注目すべき点は、固有関数が決定された後も、各軸の関数に分離する際に、実数関数では、正負の任意性しか含まれない（進行波の実数解析での定数 φ の任意性の意味は異なる）のに対し、複素関数では、大きさ 1 の複素定数 $\exp(i\gamma)$ が含まれる点である。このことは、具体的な解析において固有関数を各軸の関数に分離

して考察する際に注意すべき性質である。

2.3 Hilbert 変換

前節では、実数値の波関数を暗黙の内に適切に複素化した関数を用いて、複素値解析の考察を行った。ここでは、実数値の一般的な波関数の複素化について考察する。

複素化には、以下に定義される Hilbert 変換 H を用いる。

$$H[f](z) = \text{p.v.} \int_{-\infty}^{\infty} \frac{f(y)dy}{\pi(z-y)} \quad (29)$$

ただし、p.v. は積分の主値を表す。Hilbert 変換は、実数軸上で定義された関数を複素平面上に解析接続するための虚数値を与える関数を与えるものであり、三角関数に対しては、位相を $\pi/2$ 進めた関数に変換され、次のようになる。

$$\left. \begin{aligned} H(\cos z) &\rightarrow \sin z \\ H(\sin z) &\rightarrow -\cos z \end{aligned} \right\} \quad (30)$$

対象とする波関数は、時間と空間の 2 変数の関数であり、いずれの変数に対して Hilbert 変換を適用すればよいのかについて検討する必要がある。以下では、実際に両変数に Hilbert 変換を適用して考察する。実数値関数 $\eta^r(x, t)$ を、以下のように空間と時間の 2 重 Fourier 級数に展開する。

$$\eta_0^r(x, t) = \sum_{m,n} \left\{ a_{mn}^{cc} \cos mkx \cos n\omega t + a_{mn}^{sc} \sin mkx \cos n\omega t + a_{mn}^{cs} \cos mkx \sin n\omega t + a_{mn}^{ss} \sin mkx \sin n\omega t \right\} \quad (31)$$

時間および空間軸に対する Hilbert 変換を H^t, H^x と表し、関数 η^r を複素化すると次のようになる。

$$\eta_0^r + iH^t[\eta_0^r] = \sum_{m,n} \left\{ c_{mn}^{(+)} \exp i(n\omega t + mkx) + c_{mn}^{(-)} \exp i(n\omega t - mkx) \right\} \quad (32)$$

および、

$$\eta_0^r + iH^x[\eta_0^r] = \sum_{m,n} \left\{ c_{mn}^{(+)} \exp i(mkx + n\omega t) + c_{mn}^{(-)} \exp i(mkx - n\omega t) \right\} \quad (33)$$

ここで、複素係数 $c_{mn}^{(+)}, c_{mn}^{(-)}$ は次のようになる。

$$\left. \begin{aligned} c_{mn}^{(+)} &= \frac{1}{2} \left\{ (a_{mn}^{cc} - a_{mn}^{ss}) - i(a_{mn}^{cs} + a_{mn}^{sc}) \right\} \\ c_{mn}^{(-)} &= \frac{1}{2} \left\{ (a_{mn}^{cc} + a_{mn}^{ss}) - i(a_{mn}^{cs} - a_{mn}^{sc}) \right\} \end{aligned} \right\} \quad (34)$$

時間軸に関して Hilbert 変換して得られた式(32)の複素値関数には、各成分波の周波数は $+n\omega$ のみであるのに対し、空間軸に関して Hilbert 変換した式(33)の複素値関数は $\pm n\omega$ であり、絶対値が等しく符号が異なる 2 種類の周波数が現れている。このことは、後者は未来から過去へも時間が流れることを許すことになり、物理的におかしくなる。従って、実数値関数を複素化する際には、間軸に関して Hilbert 変換することが妥当である。

2.4 数少ないモード関数で表現される条件

複雑な波関数の例を扱う前に、以下のような簡易な波関数でもう少し考察する。

$$\eta_1(x, t) = c_1 \exp i(\omega t - k_1 x) + c_2 \exp i(\omega t - k_2 x) \quad (35)$$

上式の波関数の相関関数は、

$$R_1(x, y) = \{ \bar{c}_1 \exp(ik_1 x) + \bar{c}_2 \exp(ik_2 x) \} \times \{ c_1 \exp(-ik_1 y) + c_2 \exp(-ik_2 y) \} \quad (36)$$

であり、以下の性質を有する。

$$\left. \begin{aligned} x \langle R_1(x, y), \exp(-ik_1 x) \rangle &= c_1 \bar{c}_1 \exp(-ik_1 y) + c_2 \bar{c}_1 \exp(-ik_2 y) \\ x \langle R_1(x, y), \exp(-ik_2 x) \rangle &= c_1 \bar{c}_2 \exp(-ik_1 y) + c_2 \bar{c}_2 \exp(-ik_2 y) \end{aligned} \right\} \quad (37)$$

従って、次の関係

$$\mu : 1 = c_1 \bar{c}_1 \mu + c_1 \bar{c}_2 : c_2 \bar{c}_1 \mu + c_2 \bar{c}_2 \quad (38)$$

を満足する μ を求めれば、空間軸の固有関数は、

$$e(x) = \frac{1}{\sqrt{1 + \mu \bar{\mu}}} \{ \mu \exp(-ik_1 x) + \exp(-ik_2 x) \} \quad (39)$$

と形式的に得られる。式(38)の関係は変形すると、

$$\bar{c}_1 c_2 \mu^2 - (c_1 \bar{c}_1 - c_2 \bar{c}_2) \mu - \bar{c}_2 c_1 = 0 \quad (40)$$

という 2 次方程式になり、その根は、

$$\mu = \mu_0, -\bar{\mu}_0^{-1} \quad (\mu_0 := c_1/c_2) \quad (41)$$

となる。しかし、根 $\mu = -\bar{\mu}_0^{-1}$ の方は論理的に、式(38)の関係に矛盾するので、次のような唯一 (2.2 で検討した大きさ 1 の不定定数による任意性は考慮

しない) の固有値・固有関数の組が得られる。

$$\begin{aligned} \{\lambda, e(x), c(t)\} = & \{c_1 \bar{c}_1 + c_2 \bar{c}_2, \\ & \sqrt{c_1 \bar{c}_1 + c_2 \bar{c}_2}^{-1} \{c_1 \exp(-ik_1 x) + c_2 \exp(-ik_2 x)\}, \\ & \sqrt{c_1 \bar{c}_1 + c_2 \bar{c}_2} \exp(iax)\} \end{aligned} \quad (42)$$

以上の解析は、固有値問題を解かなくても、

$$\begin{aligned} \eta_1(x, t) = & \exp(iax) \\ & \times \{c_1 \exp(-ik_1 x) + c_2 \exp(-ik_2 x)\} \end{aligned} \quad (43)$$

と直ちに変数分離が可能である。

次に、次のような波数 $k_{1,2}$ も周波数 $\omega_{1,2}$ も異なる 2 成分からなる波

$$\begin{aligned} \eta_2(x, t) = & c_1 \exp i(\omega_1 t - k_1 x) + \\ & c_2 \exp i(\omega_2 t - k_2 x); \quad c_1 \neq c_2 \end{aligned} \quad (44)$$

についても、直ちに固有値・固有関数の組が得られる (実際、固有値問題を解いて確認してもよい)。

$$\begin{aligned} \{\lambda, e(x), c(t)\} = & \\ & \{c_1 \bar{c}_1, \exp(-ik_1 x), c_1 \exp(i\omega_1 t)\}, \\ & \{c_2 \bar{c}_2, \exp(-ik_2 x), c_2 \exp(i\omega_2 t)\} \end{aligned} \quad (45)$$

以上の結果から、波関数を Fourier 級数展開した時の全ての成分の周波数・波数が異なる場合、各成分波が固有関数となり、その固有値の比は波のエネルギーに比例したものになることがわかる。

さらに、次のような波関数については、どのような固有値・固有関数になるかを考える。

$$\begin{aligned} \eta_3(x, t) = & c_1 \exp i(\omega_1 t - k_1 x) + c_2 \exp i(\omega_1 t - k_2 x) \\ & + c_3 \exp i(\omega_2 t - k_1 x) + c_4 \exp i(\omega_2 t - k_2 x) \end{aligned} \quad (46)$$

式(39) で得られるのと同様な 2 次方程式を用いれば、固有関数は、形式的に以下ようになる。

$$\left. \begin{aligned} e_1(x) = & \sqrt{1 + \mu_0 \bar{\mu}_0}^{-1} \\ & \times \{\mu_0 \exp(-ik_1 x) + \exp(-ik_2 x)\} \\ e_2(x) = & \sqrt{1 + \mu_0 \bar{\mu}_0}^{-1} \\ & \times \{\exp(-ik_1 x) - \mu_0^{-1} \exp(-ik_2 x)\} \end{aligned} \right\} \quad (47)$$

ただし、 μ_0 は以下の関係を満たすものである。

$$\left. \begin{aligned} \mu_0 - \bar{\mu}_0^{-1} = & \\ & (c_1 c_1 + c_2 c_2 + c_3 c_3 + c_4 c_4) / (c_1 \bar{c}_2 + c_3 \bar{c}_4) \\ \mu_0 / \bar{\mu}_0 = & (\bar{c}_1 c_2 + \bar{c}_3 c_4) / (c_1 \bar{c}_2 + c_3 \bar{c}_4) \end{aligned} \right\} \quad (48)$$

また、固有値に関しては、その比が重要であり、

$$\begin{aligned} \frac{\lambda_1}{\lambda_1} = & (c_1 \bar{c}_2 + c_3 \bar{c}_4) \\ & \times \left(1 + \frac{1}{\mu_0 \bar{\mu}_0}\right) / \left(\frac{c_1 \bar{c}_2 + c_3 \bar{c}_4}{\bar{c}_1 c_2 + \bar{c}_3 c_4} - \bar{\mu}_0^{-1}\right) \end{aligned} \quad (49)$$

となる。ここで、固有値の比が無限大になる場合を考えて、

$$\frac{c_1 \bar{c}_2 + c_3 \bar{c}_4}{\bar{c}_1 c_2 + \bar{c}_3 c_4} - \bar{\mu}_0^{-1} = 0 \quad (50)$$

の時、式(48) から、

$$c_1 c_4 - c_2 c_3 = 0 \quad (51)$$

となる。また、

$$\bar{c}_1 c_2 + \bar{c}_3 c_4 = 0 \quad (52)$$

の場合、固有値の比が 1 となる。従って、各成分の振幅の比が式(51)あるいは式(52)を満たすような特殊な条件においてのみ、波関数 η_3 は 1 つの固有関数に縮退する。

以上の考察から、式(32) で示されるような一般的な変動関数 $\eta (= \eta'_0 + i H' [\eta'_0])$ の主成分解析により分離した結果は、形式的に、

$$\eta(x, t) = \sum_n \lambda_n c_n(t) e_n(x) \quad (53)$$

$$\left. \begin{aligned} c_n(t) = & \sum_p \alpha_p \exp(i\omega_p t) \\ e_n(x) = & \sum_q \{\beta_q \exp(ik_q x) + \chi_q \exp(-ik_q x)\} \end{aligned} \right\} \quad (54)$$

となる。微分方程式の変数分離法とは異なり、時間軸上の変動も単一周期でなく、固有な変動関数が得られるので、主成分解析は、変動のエネルギーが非定常に変化する波データを対象とするのに向いている。

また、主成分解析により分解される固有関数の個数の最大値は、波データを Fourier 級数展開した時の周波数・波数(正負を別と数える)の個数の大きい方の個数 M であり、さらに、周波数・波数の組み合わせやその振幅比の組み合わせが特殊な場合、 M より減少することがわかる。

2.5 固有関数が不定となる場合

本節では、2.2 で考察した固有関数の任意性とは別に、固有値の特殊性から生じる固有関数の不定性について考察する。

いま、ある波関数の相関関数 R_4 の 2 つの直交する固有関数 f, g が存在し、それぞれの固有値が等

しい時 ($\lambda_1 = \lambda_2 = \lambda$), 次式

$$\left. \begin{aligned} \tilde{f}(x) &= \alpha f(x) + \beta g(x) \\ \tilde{g}(x) &= \chi f(x) + \delta g(x) \end{aligned} \right\} \quad (55)$$

で表わせる関数 $\tilde{f}$ および $\tilde{g}$ もまた, 相関関数 R_4 の固有関数の組である. ただし, $\alpha, \beta, \chi, \delta$ は,

$$\left. \begin{aligned} \bar{\alpha}\delta + \bar{\beta}\chi &= 0 \\ |\alpha|^2 + |\beta|^2 &= |\chi|^2 + |\delta|^2 = 1 \end{aligned} \right\} \quad (56)$$

をみたす複素定数である. なぜなら,

$$\begin{aligned} {}^x \langle R(x, y), \tilde{f}(x) \rangle &= \\ \alpha {}^x \langle R(x, y), f(x) \rangle + \beta {}^x \langle R(x, y), g(x) \rangle &= \\ = \lambda \tilde{f}(x) \end{aligned} \quad (57)$$

であり, 同様に,

$${}^x \langle R(x, y), \tilde{g}(x) \rangle = \lambda \tilde{g}(x) \quad (58)$$

となる. なお,

$$\begin{aligned} {}^x \langle \tilde{f}(x), \tilde{f}(x) \rangle &= \alpha \bar{\alpha} {}^x \langle f(x), f(x) \rangle + \\ (\alpha \bar{\beta} + \beta \bar{\alpha}) {}^x \langle f(x), g(x) \rangle + \\ \beta \bar{\beta} {}^x \langle g(x), g(x) \rangle &= 1 \end{aligned} \quad (59)$$

および,

$${}^x \langle \tilde{g}(x), \tilde{g}(x) \rangle = 1 \quad (60)$$

$${}^x \langle \tilde{f}(x), \tilde{g}(x) \rangle = 0 \quad (61)$$

である. 式(56)を満たす $\alpha, \beta, \chi, \delta$ の組は無数に存在するので, 固有関数は不定となる.

Original Data

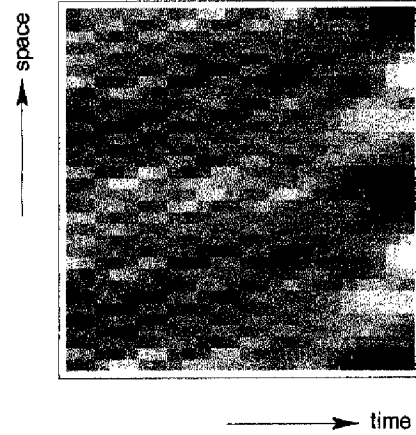


図-1 解析対象の具体的数値データの表示

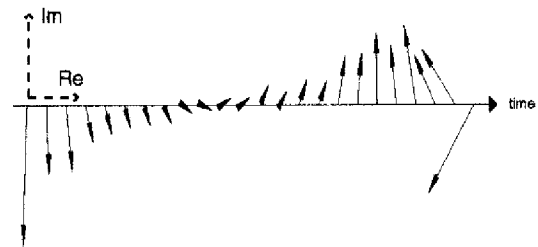


図-2 動径を用いた複素固有関数の図示

3. 複素値主成分解析の具体例

3.1 具体例その1

2. で考察したとおり, 主成分解析においては, 1) 時間および空間の各軸の固有関数に着目して考察する場合, 任意性が含まれること (2.2), 2) 時間軸上の変動も単一周期でなく, 固有な変動関数が得られること (2.4) 等の難点があるため, 具体的な数値データを解析する際には, 各軸の固有関数を合

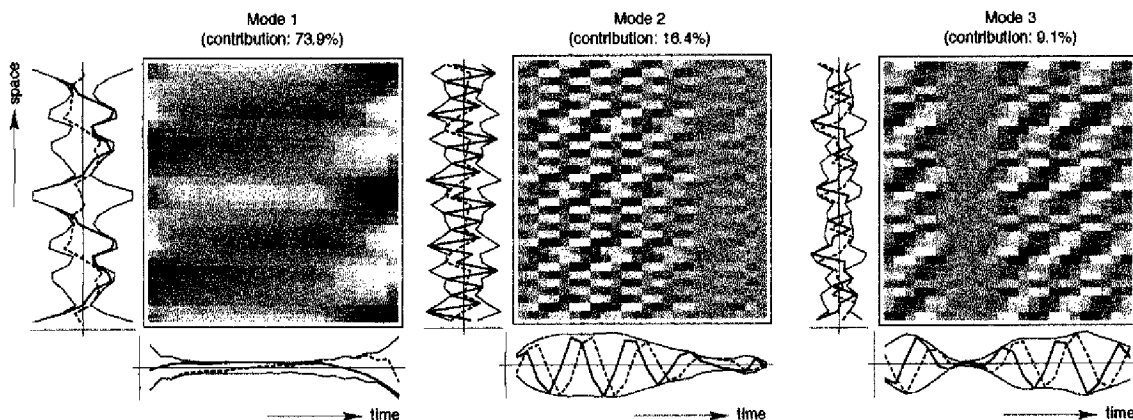


図-3 主成分解析による固有関数分解

成してデータを再構築しなければならない。相対的に固有値の大きい固有関数から再構築されたデータは、解析前の元データとほぼ同じようなものとなり、解析の目的を見失う危険性がある。極端に言えば、変動データに主成分解析を適用した従来の研究では、寄与率から元データの再現性を確認しているにすぎない。本章では、進行波成分も変数分離できる形式で表現できる利点から、複素主成分解析を具体的な数値変動データに適用し、得られる各軸の固有関数を考察する際の若干の工夫を提示する。さらに、前章で説明しきれていない主成分解析の特性についても考察する。

図-1 に示す 2 次元数値データを対象した場合、Hilbert 変換により複素化を行い、主成分解析により、寄与率が、0.74、0.16、0.09 の 3 成分の固有関数が得られた。各軸の固有関数は複素値であるので、従来は、図-2 に示すような各時刻、各地点を原点とする複素平面に連続的に図示している (Liang and Seymour, 1991; Liang et al, 1992; Bosma and Dalrymple, 1996 等) が、格子間隔が非常に狭いと煩雑になったり、その煩雑さを避けるために間引いて表示することは適切でない。そこで、実部と虚部のそれぞれの数値列 η_{Re} と η_{Im} を用いて

$$A = \sqrt{\eta_{Re}^2 + \eta_{Im}^2} \quad (62)$$

で定義される包絡線 A の時系列も併せて表示する。包絡線の時系列を表示するのは、大きさ 1 の複素数の位相の任意性(2.2) を回避するためである。図-3 に、各軸の固有関数とそれらを再構成した 2 次元データを示す。時間の周波数は単符号であり、各成分の時間の固有関数にのみ着目して考察する際に

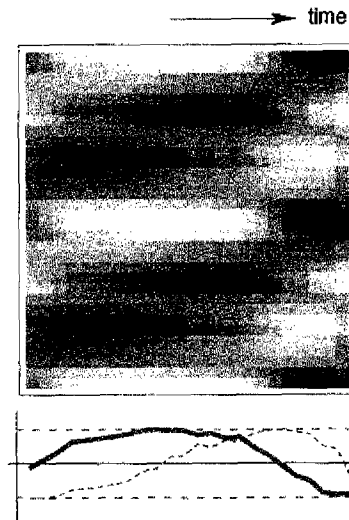


図-4 式(63)の時間関数より再構成された固有関数

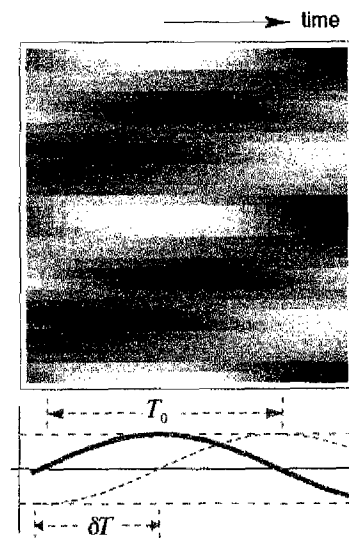


図-5 式(65)の時間関数より再構成された固有関数

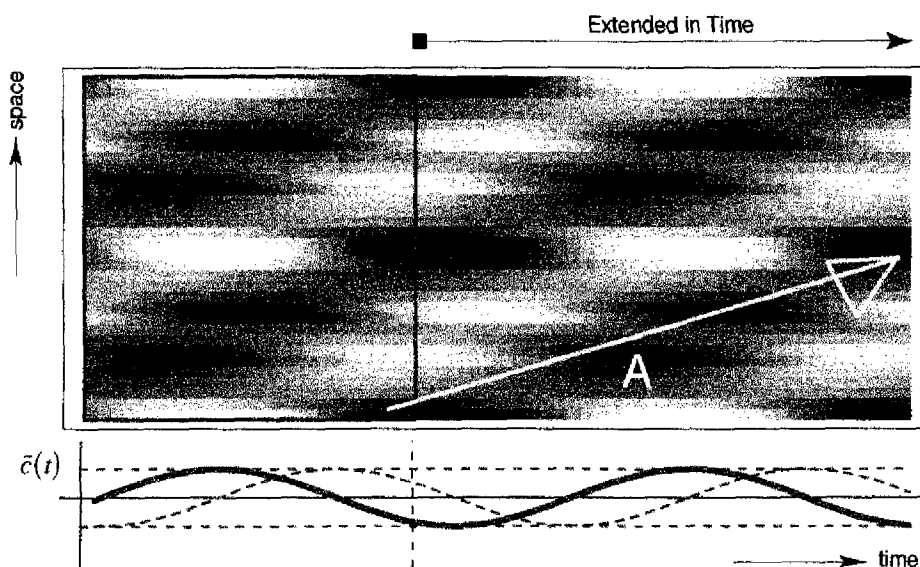


図-6 外挿を行った時間関数から作成される固有関数

は、実部や虚部の時系列そのものに意味はなく、むしろ、それらが包絡線の時系列に接する時刻の間隔が、変動の周期性しているという点で重要である。特に、漂砂による地形変動の観測データのような変動の卓越周期に対して観測期間が十分でない場合、実数値のみによる解析では包絡線の時系列を得ることができないため、変動の周期性が不明となる。この点からもデータを複素化して主成分解析を行う方が得られる情報は多い。また、空間の固有関数については、式(54)に見るとおり、空間波数は両符合が含まれるため、空間軸の固有関数を単純に見ただけでは十分に考察できない。波数の符合を分離することにより、空間軸の固有関数の特性を検討することも可能である(北野ら, 1999)。

さらに、各成分の固有関数に変数分離された形式になっているという観点から、空間の固有関数 $e^{(1)}$ の変動が時間の固有関数 $c^{(1)}$ の振動で伝播して行くと解釈される。振動関数のエネルギーが時間的に変化することに注意し、エネルギー変動を無視するために、まず、次式で表される振動関数での正規化を試みる。

$$\bar{c}^{(1)} = c(t) \| c(t) \| / | c(t) | \quad (63)$$

ただし、

$$\| f(t) \| := \int_0^T | c(t) | dt \quad (64)$$

であり、全時間におけるエネルギーの平均値を表すものである。図-4は、式(63)により正規化した時間関数 $\bar{c}^{(1)}$ と空間の固有関数により再構築した2次元データを図示したものである。対象とした変動データの時間軸のエネルギー変化が非常に大きいため、図-3では時間軸の中央部で不明瞭であった変動の伝播の様子が浮き上がってくるのがわかる。

次に、周期 T_0 と位相を表す時間差 δT を読み取ることにより、時間の固有関数 $\bar{c}^{(1)}$ を単一周期関数に置き換える。

$$\bar{c}^{(1)} = \| c(t) \| \exp \{ 2\pi i (t - \delta T) / T_0 \} \quad (65)$$

式(65)で表される $\bar{c}^{(1)}$ をもとに再構築された2次元データを図-5に図示する。図-4と比較して変動の伝播の様子が滑らかになっているが、関数 $\bar{c}^{(1)}$ を用いる最大の利点は、対象としたデータの変動の周期特性が未来や過去にわたって同様なものであると仮定のもとで、時間軸上での外挿を行なえる点にある。

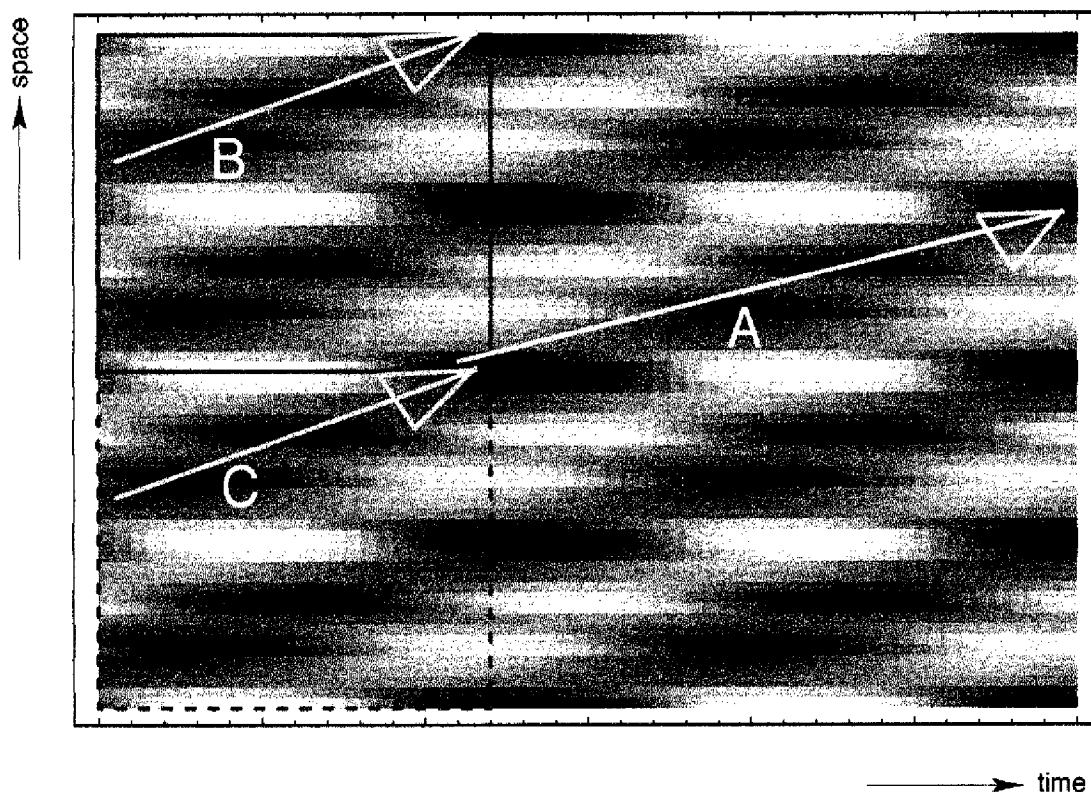


図-7 固有関数のリンク効果

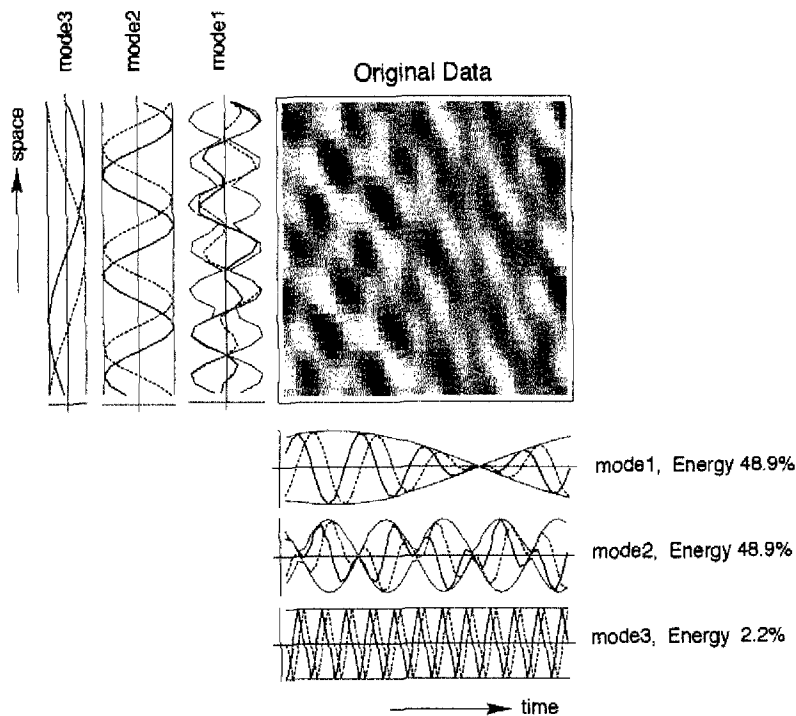


図-8 式(66)で表される具体的数値データの表示

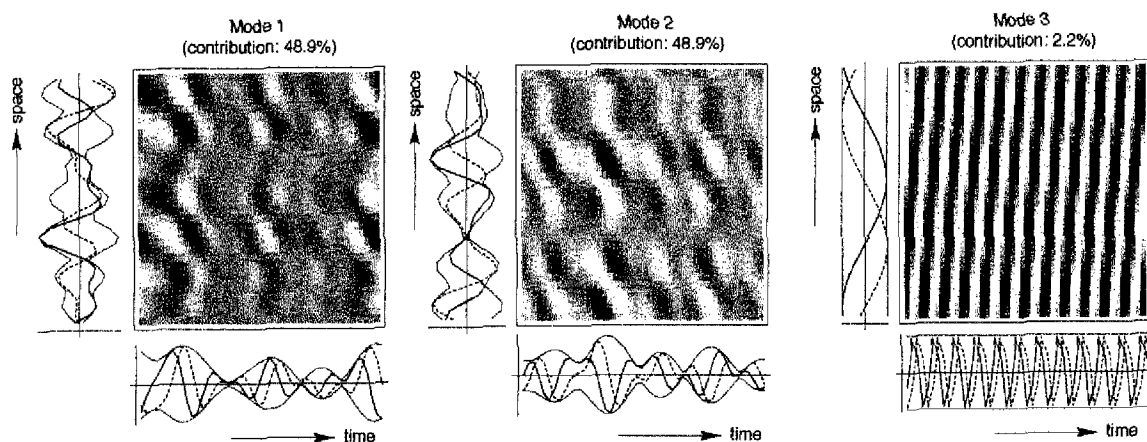


図-9 式(66)の数値データの主成分解析結果

未来の方向に $\tilde{e}^{(1)}$ を延長した時間関数から再構築される2次元データを図-6に示す。ただし、これらの考察で注意すべき点は、外挿されるエネルギー変動を無視していることと、境界付近の変動の伝播については入念な考察が必要となることである。つまり、後者については、図-6での時空間上の伝播経路Aは、元データの範囲には存在しない変動の特性が不定な領域から伝播したものと考えられ、空間軸方向に2つつなぎ合わせて作成した図-7を見るとおり、その伝播経路Aは、経路Bの複写した経路Cの延長上にあることが根拠になっているこ

とがわかる。つまり、式(6)での内積が有限区間での定義であることから、Fourier級数展開における境界近傍でのリンク効果(大崎, 1976)と同じような現象が主成分解析においても生じることが原因である。従って、データ処理の前にデータの範囲をよく吟味し、境界部の変動より中央部の変動の特性を重視することが重要である。

3.2 具体例その2

次の具体的数値データ例として、次式で表される2次元データについて考察する。

$$\eta(x,t) = \left\{ \exp(i\omega_1 t + \theta_1) + \exp(i\omega_2 t + \theta_2) \right\} \times \\ \left\{ \exp(ik_1 x + \varphi_1) + \exp(-ik_2 x + \varphi_2) \right\} + \\ \sqrt{2} \left\{ \exp(i\omega_3 t + \theta_3) + \exp(i\omega_4 t + \theta_4) \right\} \times \\ \exp(ik_3 x + \varphi_3) + 0.4242 \exp(i\omega_5 t + ik_4 x + \varphi_4) \quad (66)$$

上式を見るとおり、周波数・波数が全て異なるので、各項がそれぞれ固有関数に相当すると期待される。図-8に各項の固有関数とともに元データを図示する。また、著者らが作成した主成分解析のプログラムによる結果を図-9に示す。式(66)の第1および2項の変動のエネルギーは、いずれも2となり、解析対象となる2次元データに固有値が等しい成分が含まれる場合である。2.5で考察したとおり、固有関数が不定となるにもかかわらず、無数にある固有関数の組み合わせの1つ(固有ベクトル算定のアルゴリズムに依存)だけを選んでいることに注意する。従って、観測データの誤差を考慮すると、ほぼ同じような固有値が得られる場合には、以上で考察した固有関数の不定性に相当しないかについて検討しなければならない。

4. 結語

主成分解析の理論を連続量による表現に書き改め、また、対象となる2次元変動データをFourier級数展開することにより、得られる固有関数が直感的にわかるように整理した。その結果、変動データを対象とした主成分解析法の特徴と問題点およびその活用法について幾つか点が明らかになった。以下にそれらをまとめておく。

- 1) 実数値による主成分解析の固有関数は、定常波しか含まれないが、複素値による場合は、定常波も進行波も含まれる。従って、変動データの主成分解析を行う際は、データを複素化した方が有効である。
- 2) 実数値データの複素化の際には、時間軸変動に関して Hilbert 変換を行うのが正しい。
- 3) 固有関数の個数は、解析対象となるデータを Fourier 展開した時の波数と周波数のいずれか大きい数以下であり、波数と周波数の組み合わせやその振幅比の組み合わせが特殊な場合、さらに減少する。
- 4) 固有関数が不定となる場合に注意する。
- 5) 複素データの主成分解析により得られる各軸の固有関数は、包絡線とともに考察するとその特

性が理解しやすい。

最後に、本研究でとりあげていない主成分解析法の特徴や問題点が他にまだ残されていないかを検討するとともに、理論構築において自由度が含まれる3元変数の変動データでの解析法、他の内積定義の採用、固有関数の算出過程での条件付加等、主成分解析法にはまだまだ発展の余地が残されている。今後これらの点を、検討する予定である。

謝辞：主成分解析法の今後の改良点に関して、運輸省港湾技術研究所の橋本典明博士との有意義な議論は、本研究をまとめるにあたって非常に参考になり、ここに感謝致します。

参考文献

- 大崎順彦(1976): 地震動のスペクトル解析入門, 東京, 鹿島出版会, 260p.
- 加藤一正・灘岡和夫・田中則男(1980): Empirical eigenfunction によるエッジ波の解析, 第27回海岸工学講演会論文集, pp.178-182.
- 加藤一正・吉松 晃(1984): 三次元の経験的固有関数による深淺図解析法, 港湾技術研究所報告, 第23巻, 第2号, pp.24-47.
- 北野利一・中野 晋・岡 彰紀・間瀬 肇(1999): 主成分解析による新たな地形変動分析法に関する理論的検討, 海岸工学論文集, 第46巻, (投稿中)。
- 橋本典明・永井紀彦・清水勝義・菅原一晃・久高将信・田中 聡(1995): 主成分分析とカルマンフィルタを用いた統計的波浪予測手法の適用性について, 海岸工学論文集, 第42巻, pp.336-340.
- 横木裕宗・三村信男・佐藤圭輔(1998): 主成分解析を用いた大洗港周辺の海浜地形解析, 海岸工学論文集, 第45巻, pp.571-575.
- Aubrey, D. G. (1979): Seasonal patterns of onshore/off-shore sediment movement, J.G.R., Vol.84, No.C10, pp.6347-6354.
- Aubrey, D. G., D. L. Inman and C. D. Winant (1980): The statistical prediction of beach changes in Southern California, J.G.R., Vol.85, No.C6, pp.6264-3276.
- Barnett, T. P. (1983): Interaction of the monsoon and Pacific trade wind systems at Interannual time scale. Part I: The equatorial zone, Mon. Wea. Rev. Vol.111, pp.756-773.
- Bosma, K. F. and R. A. Darlymple (1996): Beach profile analysis around Indian River Inlet, Delaware, USA,

- Proc. 25th Coastal Eng. Conf., ASCE, pp.401-405.
- Liang, G., T. E. White and R. J. Seymour (1992): Complex principal component analysis of seasonal variation in nearshore bathymetry, Proc. 23rd Coastal Eng. Conf., ASCE, pp.2242-2250.
- Liang, G. and R. J. Seymour (1991): Complex principal component analysis of wave-like sand motions, Proc. Coastal Sediments '91, ASCE, pp.2175-2186.
- Vincent, C. L. and D. T. Resio (1977): An eigenfunction parameterization of a time sequence of wave spectra, Coastal Eng., Vol.1, pp.185-205.
- Wallace, J. M. and R. E. Dickinson (1972): Empirical orthogonal Representation of time series in the frequency domain. Part I: Theoretical considerations, J. Appl. Meteor., Vol.11, pp.887-892.
- Wallace, J. M. (1972): Empirical orthogonal Representation of time series in the frequency domain. Part II: Application to the study of tropical wave disturbances, J. Appl. Meteor., Vol.11, pp.893-990.

付録1 三角関数の公式

$$\phi \langle f, g \rangle = (2\pi)^{-1} \int_0^{2\pi} f(\phi)g(\phi) d\phi \quad (A1)$$

で内積が定義された関数空間において,

$$\left. \begin{aligned} \phi \langle \sqrt{2} \cos(m\phi), \sqrt{2} \cos(n\phi) \rangle \\ \phi \langle \sqrt{2} \sin(m\phi), \sqrt{2} \sin(n\phi) \rangle \end{aligned} \right\} = \delta_{mn} \quad (A2)$$

$$\phi \langle \sqrt{2} \cos(m\phi), \sqrt{2} \sin(n\phi) \rangle = 0 \quad (A3)$$

および,

$$\phi \langle \exp(im\phi), \exp(in\phi) \rangle = \delta_{mn} \quad (A4)$$

が成立する。ただし、 δ_{mn} は Kronecker のデルタである。従って、この内積空間での正規基底関数は、

$$\sqrt{2} \cos(n\phi), \sqrt{2} \sin(n\phi); \quad n \geq 1 \quad (A5)$$

および,

$$\exp(in\phi); \quad n = \text{integer} \quad (A6)$$

である。

Theoretical Considerations for Applying Complex Principal Component Analysis to Fluctuated Data

Toshikazu KITANO* and Hajime MASE

* Dept. of Civil Engineering, The University of Tokushima

Synopsis

The Principal Component Analysis (PCA) has been widely used in all fields of engineering, and it is known as a tool of statistical analysis in order to extract the characteristics from experimental data. In the existing studies dealing with fluctuated data, the orthogonality of eigen functions by PCA has been emphasized and the other properties have not been utilized enough. This study proposes more practical use of the eigen function of single variable, by examining the own properties of PCA in analyzing wave-like fluctuating data.

Keywords: PCA (Principal Component Analysis), CPCA, Empirical Eigen Function, Beach Profile Analysis, Fourier Series