MODULE 7.1
Singular Spectrum Analysis (SSA)
A PRELUDE

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

Problem area of interest:Alphabet soup

- Time Series(TS) Analysis(TSA) - univariate/scalar
  - multivariate/vector
- PCA/EoF (earlier modules)
- SSA is EoF applied to scalar TS
- Multi-channel SSA-(MSSA) : EoF applied to vector TS

Classical TSA

- TSA is a chapter in the vast and growing discipline of stochastic process in discrete time
- TSA had its beginnings in 1920's
- Developed by Yule, Walker, Wold, Wiener Kolmogorov, among others

Goals of TSA

- Given a TS, quantify the linear dependence structure using Auto Correlation Function(ACF) and Partial Auto Correlation Function(PACF)
- Identify a class of (empirical) discrete time stochastic dynamic models that could capture ACF and PACF
- Estimate the parameters of the chosen models

Goals of TSA - continued

- Compare the adequacy of the models using several measures - residual, Akaike information criterion etc.,
- Choose a very small number of "good" modules
- Develop algorithms for prediction and prediction error

Sources of TSA

- J.D.Hamilton(1995) Time Series Analysis, Princeton university press
- P.S.Brockwell and R.A.Davis(2013) Time Series Theory and Methods Springer
- W.A.Fuller(2009) Introduction to Statistical Time Series,John Wiley and sons

What is SSA?

- SSA is an alternate method to analyze time series data
- It starts by generating a data/trajectory matrix $x \in R^{m \times n}$ from the given scalar TS:$\{y_t/1 \leq t \leq N\}$
- SSA follows the foot steps of EoF analysis applied to x

What is SSA?

- The idea is to reconstruct the signal, trend, seasonal cycles and predictable parts and noise components using the spectral properties of the covariance matrix $\Sigma = \frac{1}{n}xx^T$
- Since the eigenvalues of the Gramian, $\frac{1}{n}xx^T$ are the singular values of x, the name Singular Spectrucm Analysis

Origins of SSA

- SSA was developed rather recently:
- J.M.Colebrook(1978)"Continous plakton records:Zooplakton and environment, North-East Atlantic and North Sea, 1948-1975",Oceanologica ACTA, vol 1, pp 9-23
- D.S.Broomhead and G.P.King(1986) "Extracting Qualitative Dynamics from Experimental Data" Physica, 20D, pp 217-236
- K.Fraedrich(1986) "Estimating the dimension of weather and climate attractor", Journal of Atmospheric Sciences, vol 43, pp 419-43

Literature on SSA

- J.B.Elsener and A.A.Tsonis(1996) Singular Spectrum Analysis, plenum press, New york
- N.Golyandina, V.Nekrutkin and A.Zhigljavsley(2001) Analysis of Time Series Structure: SSA and related techniques Chapman and Hall
- M.Ghil, et.al.(2002)"Advanced spectral methods for climate time series", Reviews of Geophysics, vol 40, 3-1 to 3-41 - Deals with SSA,MSSA and applications

From TS to Data/trajectory matrix: $x \in R^{m \times n}$

- Let $\{y_t | 1 \leq t \leq N\}$ be the given scalar TS data
- Pick an integer m : $1 < m \leq \frac{N}{2}$ and let $n = N - m + 1$
- m is called the window length and n is the number of continuous windows that can be formed from the given time series of length N

Columns of x

- The j^{th} column, x_{*j} of x is given by the m entries in the TS starting from location j where $1 \leq j \leq n$:

$$x_{*j} = (y_j, y_{j+1}, \ldots, y_{j+m-1})^T \tag{1}$$

- Then

$$x = [x_{*1}, x_{*2}, \ldots, x_{*n}] \in R^{m \times n} \tag{2}$$

matrix associated with the TS, y_t

Illustration

- Let N = 6 and m = 3. Then, $n = N - m + 1 = 4$
- $y_t = \{y_1, y_2, y_3, y_4, y_5, y_6\}$
- $x_{*j} = (y_j, y_{j+1}, \ldots, y_{j+m-1})^T$
- $x = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ y_2 & y_3 & y_4 & y_5 \\ y_3 & y_4 & y_5 & y_6 \end{bmatrix}$

Structure of x

- From the definition:

$$x = [x_{ij}] \quad and \quad x_{ij} = y_{i+j-1} \tag{3}$$

- The matrix x is such that elements along the anti diagonal for which $(i+j) = c$ for $2 \le c \le m+n$ are the same
- Such matrices are called Hankel matrices

Origin of trajectory matrices: Dynamical system

- Trajectory matrices were introduced to solve a class of inverse problems in dynamical system theory
- Let $f : R^n \to R^n$ be a smooth vector field and

$$\dot{x} = f(x) \tag{4}$$

be the given dynamical system(DS)

The direct problem

- Given the DS in (4) and an initial condition $x_0 \in R^n$, numerically compute the time series x(t) for $t \geq 0$
- The standard Runge - Kutta methods is often used for this purpose

The inverse problem

- Given only the time series of the i^{th} component, $x_i(t)$ of the solution x(t) of (4), infer the qualitative properties of the DS that generates x(t)
- These include fixed points and their stability, properties of altractors and their properties etc.,

Origin of the Trajectory matrix

- Trajectory matrix was introduced in the context of solving the above inverse problem: Derive the phase space characterization, from the given TS using the trajectory matrix defined above
- Packard et.al (1980)
- Ruelle (1980)
- Takens (1981)

Literature on DS and nonlinear Time Series

- There is a rich literature on the study of nonlinear time series in the context of chaotic dynamics
- References are given at the end of this module

Multi-channel SSA : MSSA

- Let $y_t \in R^L$, for some finite integer $L > 1$ be the given vector of time series for $1 \leq t \leq N$
- Define the integers m and n as above
- MSSA starts a trajectory matrix $x \in R^{m \times n}$ by stacking together the trajectory matrices $x(i) \in R^{m \times n}$ for each component y_i, t of the vector y(t)

An illustration

- Let L = 3, N = 6, m = 3 and n = 4
- The given series:$y_t = (y_{1t}, y_{2t}, y_{3t})^T \in R^3 : 1 \leq t \leq N$
- Then

$$x(1) = \begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & y_{1,4} \\ y_{1,2} & y_{1,3} & y_{1,4} & y_{1,5} \\ y_{1,3} & y_{1,4} & y_{1,5} & y_{1,6} \end{bmatrix} \quad (5)$$

be the 3×4 trajectory matrix built out of the first component $\{y_{1,t} | 1 \leq t \leq 6\}$ of the given vector time series $\{y(t) | 1 \leq t \leq 6\}$

Illustration - continued

- Likewise build trajectory matrices x(2) and x(3) from the second and third components of $\{y_t \in R^3 | 1 \leq t \leq N\}$
- Then

$$x = \begin{matrix} & n \\ m & \begin{bmatrix} x(1) \\ \dots \\ x(2) \\ \dots \\ x(3) \end{bmatrix} \\ m & \\ m & \end{matrix} \quad (6)$$

is the trajectory matrix for the MSSA of $\{y_t\}$

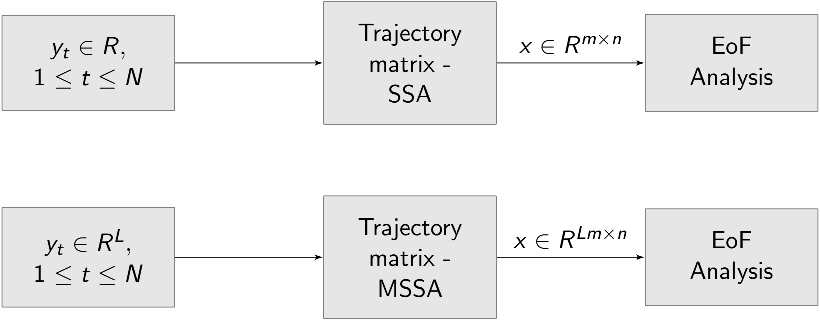
Illustration - continued

- Given $x \in R^{Lm \times n}$, compute the covariance matrix

$$\Sigma = \frac{1}{n} x x^T \in R^{Lm \times Lm} \tag{7}$$

- Σ contains the auto and cross covariances of the components of the vector $\{y_t\}$
- The singular values and vectors of x are closely related to the spectrum of Σ

Summary



- Except for the dimensionality, mathematical analysis of SSA and MSSA are quite similar

References: Use of trajectory matrix

- N.H.Packard, J.P.Crutchfield, J.D.Farmer and R.S.Shaw(1980)"Geometry from a time series", Physical Review Letter, A 45, 712-716
- D.Ruelle (1980)"Strange attractors",Mathematics Intelligence π,2,37-48
- F.Takens(1981)"Detecting strange attractors in turbulence" in: D.Rand and L.S.Young(Eds) Dynamical Systems and Turbulence, vol 898 of lecture notes in Mathematics, pp 366-381, Springer,Berlin

References

- H.Kanty and T.Schreiber(1997) Nonlinear Time series Analysis,Cambridge University Press
- H.Tong(1993) Non-linear Time series Analysis: A Dynamical Systems Approach, oxford university press, oxford

MODULE 7.2
SSA: a first look

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

Scalar time series

- Let y $= \{y_t | 1 \leq t \leq N\}$ denote the given scalar time series
- Examples: y_t is the global average temperature for year t
- y_t is the total number of deaths due to road accidents in the year t
- y_t is the Facebook stock price at the end of the day t

Trajectory / (lagged) data matrix $x \in R^{m \times n}$

- Let $1 < m < N/2$ and n = N-m+1. Then $m < n$
- Define the j^{th} column, x_{*j} of x:

$$x_{*j} = (y_j, y_{j+1}, \ldots, y_{j+m-1})^T \tag{1}$$

- The data matrix:

$$x = [x_{ij}] = \frac{1}{\sqrt{n}}[x_{*1}, x_{*2}, \ldots, x_{*n}] \in R^{m \times n} \tag{2}$$

where

$$x_{ij} = y_{i+j-1} \tag{3}$$

Example

- N $= 6$, m $=3$, n$=4$, $\sqrt{n} = 2$
- y $= \{y_1, y_2, y_3, y_4, y_5, y_6\}$
- x $= \frac{1}{2}\begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ y_2 & y_3 & y_4 & y_5 \\ y_3 & y_4 & y_5 & y_6 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}$
- Clearly the Hankel structure of x is evident:

$$x_{ij} = y_{i+j-1} \text{ for } i + j = c \tag{4}$$

where $2 \leq c \leq n + m$

A geometric view of x

- It is useful to consider the j^{th} column x_{*j} of x as coordinates of the j^{th} point in R^m
- Thus, the trajectory matrix, x describes the distribution of n points in R^m

Phase space trajectory induced by x

- The n points in R^m can be temporally ordered by connecting the point x_{*j} and $x_{*(j+1)}$ by a line segment for $1 \leq i \leq n$
- The resulting trajectory consisting of (n-1) piece-wise continuous line segments as $n \rightarrow \infty$ provides good amount of qualitative information on the system that generates the original time series

Sample second moment matrix: Version 1

- Let

$$\Sigma(1) = xx^T = \frac{1}{n}\begin{bmatrix} x_{*1} & x_{*2} & \ldots & x_{*n} \end{bmatrix}\begin{bmatrix} x_{*1}^T \\ x_{*2}^T \\ \vdots \\ x_{*n}^T \end{bmatrix} = \frac{1}{n}\sum_{k=1}^{n} x_{*k}x_{*k}^T \tag{5}$$

which is symmetric and is the average of the n outer product matrices

Example of Σ

- For the example with N = 6, m = 3 and n = 4:

$$\Sigma(1) = \frac{1}{4}\begin{bmatrix} \sum_{k=1}^{4} y_k^2 & \sum_{k=1}^{4} y_k y_{k+1} & \sum_{k=1}^{4} y_k y_{k+2} \\ \sum_{k=2}^{5} y_k y_{k-1} & \sum_{k=2}^{5} y_k^2 & \sum_{k=2}^{5} y_k y_{k+1} \\ \sum_{k=3}^{6} y_k y_{k-2} & \sum_{k=3}^{6} y_k y_{k-1} & \sum_{k=3}^{6} y_k^2 \end{bmatrix} \tag{6}$$

General structure of Σ: Diagonal elements

- Using $x_{ij} = y_{i+j-1}$:

$$\Sigma_{ii}(1) = \frac{1}{n}\sum_{k=1}^{n} x_{ik}x_{ik} = \frac{1}{n}\sum_{k=1}^{n} y_{i+k-1}^2 = \frac{1}{n}\sum_{k=i}^{n+i-1} y_k^2 \qquad (7)$$

General structure of Σ: off-diagonal elements

- $$\begin{aligned}\Sigma_{ij}(1) &= \frac{1}{n}\sum_{k=1}^{n} x_{ik}x_{jk} \\ &= \frac{1}{n}\sum_{k=1}^{n} y_{i+k-1}y_{j+k-1} \\ &= \frac{1}{n}\sum_{k=i}^{n+i-1} y_k y_{k+|j-i|}\end{aligned} \tag{8}$$

- Let N be very large and m be small such that $n \approx N$
- In this case:

$$\Sigma_{ii}(2) = \frac{1}{N}\sum_{k=1}^{N} y_k^2 = c_0 \tag{9}$$

$$\Sigma_{ij}(2) = \frac{1}{N - |j-i|}\sum_{k=1}^{N-|j-i|} y_k y_{k+|j-i|} = c_{|j-i|} \tag{10}$$

Asymptotic approximation to Σ

- From (9) and (10)

$$\Sigma(2) \approx \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{m-1} \\ c_1 & c_0 & c_1 & \cdots & c_{m-2} \\ c_2 & c_1 & c_0 & \cdots & c_{m-3} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{m-1} & c_{m-2} & c_{m-3} & \cdots & c_0 \end{bmatrix} \qquad (11)$$

- The second moment matrix $\Sigma(2)$ in (11) in addition to being symmetric, also inherits the Toeplitz structure:

$$\Sigma(2) = [\Sigma_{ij}(2)] \text{ and } \Sigma_{ij}(2) = c_{|j-i|} \tag{12}$$

- That is, elements along the principal diagonal are the same as are those along the diagonals parallel to it

Properties of $\Sigma(1)$ and $\Sigma(2)$

- First choice (Broomhead and King (1986)): $\Sigma(1) \in \mathrm{R}^{m \times m}$ is computed using (7) and (8) which is symmetric
- Second choice (Vautard and Ghil (1989)): $\Sigma(2) \in \mathrm{R}^{m \times m}$ is computed using (9) and (10) which is symmetric and Toeplitz

Detrending

- The raw/original time series y_t may have non-stationary components such as trend(linear/non-linear)
- Estimate the trend using standard OLS method
- Detrend the given series by subtracting the trend component from the original series

Centering and normalization

- The centered version is obtained by subtracting the over all sample mean from each term of the series
- A centered series is also known as the anomaly series
- Normalized version is obtained by dividing each element by the overall sample standard deviation
- When there is a comparison of methods or different series, it is useful to work with normalized series

Comments

- Centering does not remove the trend
- If y_t is a centered second-order stationary series, then for large N

$$\Sigma(2) = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{m-1} \\ r_1 & r_0 & r_2 & \cdots & r_{m-2} \\ r_2 & r_1 & r_0 & \cdots & r_{m-3} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{m-1} & r_{m-2} & \cdots & \cdots & r_0 \end{bmatrix} \tag{13}$$

is the symmetric, Toeplitz auto covariance matrix

- If y_t is a normalized second-order stationary series, then for large N

$$\Sigma(2) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{m-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{m-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{m-1} & \rho_{m-2} & \rho_{m-3} & \dots & 1 \end{bmatrix} \tag{14}$$

is the symmetric, Toeplitz auto correlation matrix

Comments

- Since there is an intrinsic difference between $\Sigma(1)$ and $\Sigma(2)$ especially when N is small, great caution must be exercised in interpreting the results that are dependent on the properties of these matrices

Comments - white noise

- Another 500 pound Gorilla in the room is the property of the (stochastic) noise, ε_t component that induces randomness to y_t
- In many applications, this noise ε_t is modeled as a white noise: $E(\varepsilon_t) = 0 \quad E(\varepsilon_t\ \varepsilon_s) = 0$ for $t \neq s$
$= \sigma^2$ for $t = s$

Comments - Red noise

- In many geophysical applications, there is evidence that this corrupting noise ε_t is not a white noise
- Such a red noise is often modeled using a member of the ARMA(p, q) family
- Presence of red noise further complicates the analysis and the conclusions

Challenge of the red noise

- Allen and Smith (1994) provided a comprehensive discussion of the pitfalls in the analysis of time series using the SSA when the noise is not white
- As a further guide to the analysis, they proposed a Monte Carlo SSA that is useful in testing hypothesis relating to the properties of noise

- Let $\Sigma \in R^{m \times m}$ be the matrix computed using either of the two method using the raw, centered or normalized time series
- Assume that x is of full rank, that is

$$Rank(x) = min\{m, n\} = m \tag{15}$$

since $m < n$

- Then, $\Sigma(1)$ is SPD and $\Sigma(2)$ is SPD and Toeplitz

Eigen decomposition of Σ

- Let (λ_i, u_i), $1 \leq i \leq m$ be the eigenvalue vector pair for Σ, that is, $\Sigma u_i = u_i \lambda_i$ with

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m > 0 \tag{16}$$

- Setting $u = [u_1, u_2, \ldots, u_m] \in R^{m \times m}$
$\Lambda = Diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$
-

$$\Sigma u = u\Lambda, uu^T = u^T u = I_m \tag{17}$$

Principal patterns and components

- The eigenvector $u_i, 1 \leq i \leq m$ that constitute an orthogonal basis for R^m are the principal patterns
- The principal components are obtained by projecting the columns of x onto these principal pattern vectors

Principal component matrix: $A \in R^{m \times n}$

- The PC matrix A is given by

$$A = U^T x \in R^{m \times n} \tag{18}$$

- That is,

$$A = [A_{ij}] = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} \begin{bmatrix} x_{*1} & x_{*2} & \dots & x_{*n} \end{bmatrix}$$

-

$$A_{ij} = u_i^T x_{*j} (\text{inner product}) \tag{19}$$

Orthogonality of the rows of A

- $$AA^T = u^T xx^T u = u^T \Sigma u = \Lambda \tag{20}$$
- Each of the n columns of A give the coordinates of the n points with respect to new orthonormal coordinate system defined by the columns of u

Example: Trajectory matrix

- Consider a series $\{1, 2, 3, 4, 3, 2, 1\}$ with N = 7. Set m = 3 and n = N - m + 1 = 5
- The trajectory matrix

$$x = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 & 3 & 4 & 3 \\ 2 & 3 & 4 & 3 & 2 \\ 3 & 4 & 3 & 2 & 1 \end{bmatrix} \in R^{3\times 5} \tag{21}$$

Example: Σ, u, Λ

- $\Sigma = \Sigma(1) = \begin{bmatrix} 7.8 & 7.6 & 6.2 \\ 7.6 & 8.4 & 7.6 \\ 6.2 & 7.6 & 7.8 \end{bmatrix}$
- $$\Lambda = Diag(0.093, 1.600, 22.307) = Diag(\lambda_3, \lambda_2, \lambda_1) \tag{22}$$
- $u = \begin{bmatrix} -0.4324 & -0.7071 & 0.5595 \\ 0.7912 & -0.0000 & 0.6115 \\ -0.4324 & 0.7071 & 0.5595 \end{bmatrix} = [u_3, u_2, u_1]$

Example:PC matrix A

- $A = \begin{bmatrix} -0.0658 & -0.0987 & 0.2551 & -0.0987 & -0.0658 \\ 0.6325 & 0.6325 & -0.0000 & -0.6325 & -0.6325 \\ 1.5478 & 2.3217 & 2.5952 & 2.3217 & 1.5478 \end{bmatrix}$
- Verify $AA^T = \Lambda$

- In the light of (17), from (18), it is immediate that

$$x = uA = uu^T x \tag{23}$$

- That is, the original matrix, x can be recovered from A using the operation uA

Reconstruction of the signal: $x(s) \in R^{m \times n}$

- In the light of ordering of λ's in (16), define an integer k such that $1 \leq k < m$ and

$$\sum_{i=1}^{k} \lambda_i \geq (1-\beta) \sum_{i=1}^{m} \lambda_i \tag{24}$$

- Define

$$I_m(m-k) = Diag(0, 0, \ldots, 0, 1, 1, \ldots, 1) \tag{25}$$

where 0's are k in number and 1's are m-k in number

Expression for $x(s)$

- Then
$$x(s) = u[I_m - I_m(m-k)]u^T x \tag{26}$$
$$= (\textstyle\sum_{i=1}^{k} u_i u_i^T)x$$
gives the signal component. The noise component is
$$x(n) = x - x(s) = u[I_m(m-k)]u^T x \tag{27}$$

Example - continued

- From (22): $\lambda_1 + \lambda_2 + \lambda_3 = 24$
- $\frac{\lambda_1}{24} = 0.929, \frac{\lambda_2}{24} = 0.0667, \frac{\lambda_3}{24} = 0.00388$
- Verify that $\lambda_1 + \lambda_2 = 23.907 > 0.99(\lambda_1 + \lambda_2 + \lambda_3) = 23.897$ where $\beta = 0.01$

$$\begin{aligned} x(s) &= (u_1 u_1^T + u_2 u_2^T)x \\ x(n) &= u_3 u_3^T x \end{aligned} \tag{28}$$

Example - continued

- Verify that (since n = 5)

$$\sqrt{5} \times (n) = \begin{bmatrix} 0.9364 & 1.9046 & 2.3467 & 3.9046 & 2.9364 \\ 2.1164 & 3.1747 & 3.5486 & 3.1747 & 2.1164 \\ 2.9364 & 3.9046 & 3.2467 & 1.9046 & 0.9364 \end{bmatrix} \tag{29}$$

is the signal part of the x recovered by the SSA analysis

Recovery of the signal part of the TS

- The signal recovered in (29) does not inherit the Hankel structure as the original data matrix

References

- D.S.Broomhead and G.P King (1986) "Extracting qualitative dynamics from experimental data", Physica, D 20, 217-236
- R. Vautard and M.Ghil(1989) "Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series", Physica D, 35, 395-424
- M.R.Allen and L.Smith (1996)"Monte Carlo SSA: Detecting irregular oscillations in the presence of coloured noise", Journal of climate, 93373-3404
- R.Vautard, P.Yiou and M.Ghil (1992), "Singular spectrum analysis: a tool kit for short, noisy and chaotic series", Physica D, 58, 95-126

MODULE 7.3
Singular spectrum Analysis(SSA): A second view

by

S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

What is SSA?

- SSA is a powerful tool for the analysis of time series(TS) using ideas from
  - multivariate statistics
  - geometry
  - dynamical system
  - signal processing

The goal of SSA

- To additively decompose a given TS as a sum of "independent" components that capture the
    - time varying trend
    - oscillatory component
    - noise

Two stages of SSA

1. Decomposition stage - two steps
   - 1.1 Embedding
   - 1.2 SVD Analysis
2. Reconstruction stage - two steps
   - 2.1 Grouping
   - 2.2 Diagonal averaging

Comments on the two approaches to SSA

SSA

Module 7.2

- Exploits the eigen decomposition of the covariance matrix of $x \in R^{m \times n}$
- Developed in Europe and USA
- Exploits the properties of the principal patterns

Module 7.3

- Exploits the SVD of $x \in R^{m \times n}$
- Developed by the Russian School
- Exploits the idea of grouping based on the notion of "separability"

References

- N. Golyandina, V. Nekrutkin and A. Zhigljavsky (2001) Analysis of Time series structure: SSA and Related Techniques, Chapman and Hall/ CRC 305 pages
- J. B. Elsner ana A. A. Tsonis(1996) Singular Spectrum Analysis: A new tool in Time Series Analysis, Plenum Press, NY, 164 pages

References

- H. Hassani (2007) "Singular Spectrum Analysis: Methodology and Comparison", Journal of Data Science, Vol 5, pp239 - 257

- Provides a nice summary of the key steps in the SSA based algorithm and a good comparison with other classical techniques from TSA.

Notation

- Let $\{y_t | 1 \leq t \leq N\}$ be the given scalar TS, where N is large.
- Let $1 < m < N/2$ and n = N - m + 1
- Define a lagged (Column) vector for $1 \leq j \leq n$:

$$x_{*j} = (y_j, y_{j+1}, \ldots, y_{i+m-1})^T \in R^m \tag{1}$$

Step 1.1 Embedding

- The given series is split into n lagged column vectors x_{*j}, $1 \leq j \leq n$
- Define the trajectory matrix: $x \in R^{m*n}$

$$x = [x_{*1}, x_{*2}, \ldots, x_{*n}] \in R^{m \times n} \quad (2)$$

Example

- Let N = 7, m = 3, n =5
- y = $\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$
- $x = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 \\ y_2 & y_3 & y_4 & y_5 & y_6 \\ y_3 & y_4 & y_5 & y_6 & y_7 \end{bmatrix} \in \quad R^{3\times 5}$
- Verify: $x_{ij} = y_{i+j-1}$ and matrix x inherits the Hankel structure - elements across the anti-diagonals (i + j = k, a constant) are the same

Step 1.2 : SVD of x $\in \mathbb{R}^{m \times n}$

- Since $1 < m < N/2$ and n = N - m +1, we have $m < n$
- Let x be full rank matrix: Rank(x) = min{m, n} = m
- Then (xx^T) is SPD and

$$(xx^T)u = u\Lambda \tag{3}$$

be the eigen decomposition of the smaller Gramian xx^T

- u $= [u_1, u_2, u_3, \ldots, u_m] \quad \in \quad \mathrm{R}^{m\times m}$

$$uu^T = u^T u = I_m \tag{4a}$$

- Let $\Lambda = Diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$:

$$\lambda_1 \geq \lambda_2 \geq \lambda_m \geq \lambda_m \tag{4b}$$

- Define

$$v_i = \frac{1}{\sqrt{\lambda_i}} x^T u_i \tag{5}$$

SVD of x

- v $= \{v_1, v_2, v_3, \ldots, v_m\} \in \quad \mathrm{R}^{n\times m}$ and

$$v^T v = I_m \tag{6}$$

$$(x^T x)v = v\Lambda \tag{7}$$

- A dual relation:

$$x^T u_i = v_i \lambda_i^{1/2} \quad and \quad x v_i = u_i \lambda_i^{1/2} \tag{8}$$

SVD of x

- From (5):

$$x^T u = v\Lambda^{1/2} \quad or \quad x^T = v\Lambda^{1/2} u^T \tag{9}$$

- From (8): SVD of x:

$$x = u\Lambda^{1/2} v^T \tag{10}$$

x as a sum of rank-1 matrices

- From (10):

$$x = \sum_{i=1}^{m} \lambda_i u_i v_i^T \tag{11}$$

- (λ_i, u_i, v_i) - i^{th} eigen triple of x, $1 \leq i \leq m$
- These m eigen triples are the basic building blocks for the reconstruction phase used in identifying the trend, oscillatory and noise components.

Energy in x

- Recall:

$$\begin{aligned} ||\lambda_i u_i v_i^T||_F^2 &= \lambda_i^2 \\ ||\lambda_i u_i v_i^T + \lambda_j u_j v_j^T||_F^2 &= \lambda_i^2 + \lambda_j^2 \end{aligned} \tag{12}$$

- From (11): The total energy in x:

$$||x||_F^2 = \sum_{i=1}^{m} \lambda_i^2 \tag{13}$$

Inherent optimality of SVD

- Let, for $1 \le r \le m$,

$$x(r) = \sum_{i=1}^{r} \lambda_i u_i v_i^T \in R^{m \times n} \tag{14}$$

- From (12):

$$||x(r)||_F^2 = \sum_{i=1}^{r} \lambda_i^2 \tag{15}$$

- Hence

$$||x - x(r)||^2 = \sum_{i=r+1}^{m} \lambda_i^2 \tag{16}$$

- In view of the ordering of the λ's in (4b), it is immediate that $x(r)$ is the best rank r approximation of x

Step 2.1: Grouping of indices

- Let $[m] = \{1, 2, 3, \ldots, m\}$
- Let for some $1 \leq p \leq m$, let $s_p \subset [m]$
- Grouping: Let $\{s_1, s_2, \ldots, s_p\}$ be partition of [m]

$$s_i \cap s_j = \emptyset, \quad \bigcup_{j=1}^{p} s_j = [m]$$

Resulting decomposition of x

- Consider the i^{th} group s_i of $|s_i|$ indices
- Define a matrix $x(i) \in R^{m\times n}$ of Rank(x(i)) $= |s_i|$ as

$$x(i) = \Sigma_{j \in s_i} \lambda_j u_j v_j^T \tag{17}$$

- Clearly,

$$x = \sum_{i=1}^{p} x(i) \tag{18}$$

where p is the number of sets in the partition of [m]

Step 2.2 - Diagonal averaging

- The matrix x(i) in (17) need not be a Hankel matrix
- Hankelization of x(i) relates to creating a Hankel matrix, $\bar{x}$(i) $= \mathcal{H}$x(i) for the x(i)
- This is the done by replacing each anti-diagonal in x(i) by the average of the elements in that diagonal in x(i)

Example of Hankelization

- Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \tag{19}$$

$$Z = \begin{bmatrix} z_1 & z_2 & z_3 & z_4 \\ z_2 & z_3 & z_4 & z_5 \\ z_3 & z_4 & z_5 & z_6 \end{bmatrix} = \mathcal{H}A \tag{20}$$

- Then

$$
\begin{aligned}
z_1 &= a_{11} \\
z_2 &= \frac{1}{2}(a_{12} + a_{21}) \\
z_3 &= \frac{1}{3}(a_{13} + a_{22} + a_{31}) \\
z_4 &= \frac{1}{3}(a_{14} + a_{23} + a_{32}) \\
z_5 &= \frac{1}{2}(a_{24} + a_{42}) \\
z_6 &= a_{34}
\end{aligned} \tag{21}
$$

Algorithm for Hankelization

- Let A $= [a_{ij}] \in R^{m \times n}$ matrix
- Let B $= \mathcal{H}A = [B_{ij}] \in R^{m \times n}$

$$B_{ij} = \begin{cases} \frac{1}{s-1} \sum_{p=1}^{s-1} a_{p,s-p} & \text{for } 2 \leq s \leq m-1 \\ \frac{1}{m} \sum_{p=1}^{m} a_{p,s-p} & \text{for } m \leq s \leq n+1 \\ \frac{1}{m+n-s+1} \sum_{p=s-n}^{L} a_{p,s-p} & \text{for } n+2 \leq s \leq n+m \end{cases} \tag{22}$$

Hankelize the sum in (18)

- Recall that
  - $\mathcal{H}(\mathrm{A}+\mathrm{B}) = \mathcal{H}(\mathrm{A}) + \mathcal{H}(\mathrm{B})$
  - $\mathcal{H}(\mathrm{A}) = \mathrm{A}$ if A is a Hankel matrix.
- Operating both sides of (18) by $\mathcal{H}$:

$$x = \sum_{i=1}^{p} \mathcal{H}(x(i)) = \sum_{i=1}^{p} \bar{x}(i) \tag{23}$$

Decomposition of y_t

- Recall: Unique relation between time series and Hankel trajectory matrix
- Let $\{y_t | 1 \leq t \leq N\}$ be the TS for x and $\{\bar{y}_t(i) | 1 \leq t \leq N\}$ be the TS for (i)
- Then (23) becomes:

$$y_t = \sum_{i=1}^{p} \bar{y}_t(i) \text{ for each } 1 \leq t \leq N \qquad (24)$$

Comments: choice of L

- The results of the first stage of decomposition critically depends on the window length m
- If we already know that the given TS has periodic components with period T - say using spectral analysis, then L is proportional to this period.
- In any case L must be large but less than N/2, half the length of the series

Comments: challenge of grouping

- Of the four steps involved, embedding, SVD and Hankelization are quite algorithmic and can be easily implemented
- Grouping is the most demanding part of this approach to SSA
- Algorithm for optimal grouping is still evasive, extra / supplementary information about the series could be used as a guide to grouping

Guide to grouping - Scree plot

- Identifying breaks in the scree plot - plot of the eigenvalues Vs its rank, could help identify signals from noise
- Presence of white noise corresponds to a constant lower ceiling in the scree plot
- It is known that a harmonic component produces a pair of very close singular values

Guide to grouping - Periodogram

- Compute and plot the periodogram for the TS and identify the frequency with spikes in the spectrum
- We can then search for the eigen triple whose frequencies coincide with those identified by the spikes

Role of separability in grouping

- To partially automate the grouping operation, a notion of "separability" based on a "weighted correlation" is introduced
- Two series are separated from each other if their weighted correlation is low

Choice of Weights

- Let f_t and g_t be the two time series with $1 \leq t \leq N$
- Let $1 \leq m \leq N/2$ and n = N - m + 1
- Let

$$w_k = min\{k, m, N - k + 1\} \text{ for } 1 \leq k \leq N \tag{25}$$

Example

- Let N = 6, m = 3, n =4
- Then: w_k = min{k, 3, 7-k}
- Clearly,

k	1	2	3	4	5	6
w_k	1	2	3	3	2	1

Weighted inner product of two TS

- Let $\{f_t\}$ and $\{g_t\}$ be the two series for $1 \leq t \leq N$
- The weighted inner product between $\{f_t\}$ and $\{g_t\}$ is

$$< f, g >_w = \sum_{k=1}^{N} w_k f_k g_k \tag{26}$$

Norm of a given TS: $||f||_w$

- $$||f||_w^2 = < f, f >_w = \sum_{k=1}^{N} w_k f_k^2 \qquad (27)$$

Weighted correlation between f_t and g_t

- $$\rho_w(f,g) = \frac{< f,g >_w}{||f||_w ||g||_w} \tag{28}$$

  is the weighted correlation
- The idea is: if $\rho_w(f,g)$ is small, then the series $\{f_t\}$ and $\{g_t\}$ are almost w-orthogonal and do not share common information

An application

- Given the m-eigen triples $(\lambda_i^{1/2}, u_i, v_i)$ first compute

$$x(i) = \lambda_i^{1/2} u_i v_i^T \quad 1 \leq i \leq m \tag{29}$$

- Let $\bar{x}(i)$ be the Hankelized x(i) and $\{\bar{y}_t(i) | 1 \leq t \leq N\}$ be the corresponding TS

Compute their W-Correlations

- Given $\bar{x}_t(i)$ for $1 \leq i \leq m$, compute

$$R = [R_{ij}] \in R^{m \times m}$$

where

$$R_{ij} = \rho_w(\bar{x}_t(i), \bar{x}_t(j)) \tag{30}$$

- By examining the off-diagonal entries of this symmetric matrix, we may isolate the correlated pairs from others which in turn provide useful information for grouping

MODULE 17.1
Canonical Correlation Analysis (CCA)
Basic Theory

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

PCA

- Given a random field $x \in R^m$(indexed by points in a spatial domain $S \subseteq R^k, k = 1, 2, 3$), the goal of PCA is to decompose x as a linear combination of intrinsic spatial patterns which are related to the eigenvectors of cov(x) where the variance of the random coefficients in the linear combination are directly related to the corresponding eigenvalues of cov(x)
- If the eigenvalues of cov(x) are distinct and decreasing, this additive decomposition is often used as basis for "removing the chaff(noise) from the corn(signal)"

- Given two random fields $x_1 \in R^{m_1}$ and $x_2 \in R^{m_2}$(defined over two spatial domains S_1 and S_2) the goal of CCA is to express each field as a linear combination of its own spatial patterns in such a way that the corresponding spatial patterns exhibit maximum correlation
- An example : x_1 average SST over equatorial pacific during a month and x_2 could be the average rain fall across the USA in that same month

Second-order properties of two random fields

- Let $x_1 \in R^{m_1}$ and $x_2 \in R^{m_2}$ be two random fields of interest
-

$$E(x) = \mu_1 \quad and \quad E(y) = \mu_2 \tag{1}$$

$$cov(x) = E[(x - \mu_1)(x - \mu_1)^T] = \Sigma_{11} \in R^{m_1 \times m_1} \tag{2}$$

$$cov(y) = E[(y - \mu_2)(y - \mu_2)^T] = \Sigma_{22} \in R^{m_2 \times m_2} \tag{3}$$

- Assume that Σ_{11} and Σ_{22} are SPD

Covariance between x_1 and x_2

- $$cov(x_1, x_2) = E[(x_1 - \mu_1)(x_2 - \mu_2)^T] \\ = \Sigma_{12} \in R^{m_1 \times m_2} \quad (4)$$

- $$cov(x_2, x_1) = E[(x_2 - \mu_2)(x_1 - \mu_1)^T] \\ = \Sigma_{21} = \Sigma_{12}^T \in R^{m_2 \times m_1} \quad (5)$$

Normalize x_1 and x_2

- Let $D_1 \in R^{m_1 \times m_2}$ = Diagonal matrix with the diagonal elements of Σ_{11}

  $D_2 \in R^{m_1 \times m_2}$ = Diagonal matrix with the diagonal elements of Σ_{22}
- Normalize x_1 and x_2:
- Then

$$\hat{x}_1 = D_1^{-1/2}(x_1 - \mu_1)$$
$$\hat{x}_2 = D_2^{-1/2}(x_2 - \mu_2) \quad (6)$$

are the centered and normalized versions of x_1 and x_2

- $$\begin{aligned} cor(x_1, x_2) &= cov(\hat{x}_1, \hat{x}_2) = E[\hat{x}_1(\hat{x}_2)^T] \\ &= D_1^{-1/2}\Sigma_{12}D_2^{-1/2} \in R^{m_1 \times m_2} \end{aligned} \tag{7}$$

Spatial patterns for random fields x_1 and x_2

- Let

$$F = [f_1, f_2, \ldots, f_{m_1}] \in R^{m_1 \times m_1} \quad (8)$$

be a matrix whose linearly independent column vectors span R^{m_1} and constitute m_1 distinct spatial patterns for x_1

- Similarly, let

$$G = [g_1, g_2, \ldots, g_{m_2}] \in R^{m_2 \times m_2} \quad (9)$$

be that for x_2

- By resolving x_1 along each of the spatial patterns f_i, it follows that

$$x_1 = \sum_{i=1}^{m_1} (x_1^T f_i) f_i = \sum_{i=1}^{m_1} (\alpha_i f_i) \tag{10}$$

where the coefficients $\alpha_i = (x_1^T f_i)$ of the linear combination in (10) are random variables since x_1 is random

Expansion - continued

- Similarly, resolving x_2 along each of the spatial patterns g_j, it follows that

$$x_2 = \sum_{j=1}^{m_2} (x_2^T g_j) g_j = \sum_{j=1}^{m_2} (\beta_j g_j) \tag{11}$$

- Here again, the coefficients $\beta_j = (x_2^T g_j)$ inherit their randomness from that of x_2

Statement of the problem

- The goal is to find the spatial patterns $\{f_i\}$ and $\{g_j\}$ such that

$$cor(\alpha_1, \beta_1) \geq cor(\alpha_2, \beta_2) \geq \cdots \geq cor(\alpha_k, \beta_k) > 0 \qquad (12)$$

and

$$cor(\alpha_i, \beta_i) \textit{ is the maximum for } (f_i, g_i) \qquad (13)$$

for each $i = 1, 2, \ldots k \leq min\{m_1, m_2\}$

Expression for typical correlation

- Let $f \in R^{m_1}$ and $g \in R^{m_2}$ be two typical spatial patterns for x_1 and x_2
- Define:

$$\alpha = x_1^T f \quad and \quad \beta = x_2^T g \tag{14}$$

- Then

$$\rho = cor(\alpha, \beta) = \frac{cov(\alpha, \beta)}{[var(\alpha)var(\beta)]^{1/2}} \tag{15}$$

Expression for cov(α, β)

- $$cov(\alpha, \beta) = E[(x_1 - \mu_1)^T f (x_2 - \mu_2)^T g]$$
  $$= f^T cov(x_1, x_2) g$$
  $$= f^T \Sigma_{12} g = f^T \Sigma_{21}^T g \tag{16}$$

Expression for $var(\alpha)$ and $var(\beta)$

- $$Var(\alpha) = E[(x_1 - \mu)^T f (x_1 - \mu_1)^T f]$$
$$= f^T E[(x_1 - \mu)(x_1 - \mu_1)^T] f$$
$$= f^T \Sigma_{11} f > 0 \tag{17}$$
- Similarly:
$$Var(\beta) = g^T \Sigma_{22} g > 0 \tag{18}$$

Typical correlation

- Substituting (16), (17) and (18) in (15):

$$\rho = cor(\alpha, \beta) = \frac{f^T \Sigma_{12} g}{(f^T \Sigma_{11} f)^{1/2} (g^T \Sigma_{22} g)^{1/2}} \tag{19}$$

- Goal is to find the pattern pair (f,g) that maximizes the right hand side of (19)

Scale invariance of ρ

- Let a,b be two positive real constants
- It can be verified

$$\rho = cov(\alpha, \beta) = cov(a\alpha, b\beta) \tag{20}$$

- That is, ρ is invariant under the scaling of the spatial patterns f and g

Normalized spatial patterns

- With out loss of generality, assume that the patterns f and g are normalized:

$$f^T\Sigma_{11}f = 1 \quad and \quad g^T\Sigma_{22}g = 1 \tag{21}$$

- Define

$$\bar{f} = \Sigma_{11}^{1/2}f \quad and \quad \bar{g} = \Sigma_{22}^{1/2}g \tag{22}$$

- Then

$$\bar{f}^T\bar{f} = 1 = \bar{g}^T\bar{g} \tag{23}$$

New expression for ρ

- Substituting (21), (22) and (23) in (19), we get a bilinear form ρ given by:

$$\rho = \bar{f}^T(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2})\bar{g}$$

$$= \bar{f}^T A\bar{g} \tag{24}$$

where

$$A = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} \in R^{m_1 \times m_2} \tag{25}$$

- The problem is to find $\bar{f} \in R^{m_1}$ and $\bar{g} \in R^{m_2}$ that maximizes ρ in (24) under the constraints (23)

Lagrangian approach

- Consider the lagragian

$$L(\bar{f}, \bar{g}, a, b) = \bar{f}^T A\bar{g} + a(1 - \bar{f}^T f) + b(1 - \bar{g}^T g) \qquad (26)$$

where a and b are the (scalar) unknown Lagrangian multipliers

Conditions for the maximum

- $$\nabla_{\bar{f}} L = A\bar{g} - 2a\bar{f} = 0 \quad (27)$$
  $$\nabla_{\bar{g}} L = A^T \bar{f} - 2b\bar{g} = 0 \quad (28)$$
  $$\nabla_a L = 1 - \bar{f}^T \bar{f} = 0 \quad (29)$$
  $$\nabla_b L = 1 - \bar{g}^T \bar{g} = 0 \quad (30)$$

Optimal solution

- Conditions (29) and (30) follows from (23)
- Optimal $(\bar{f}, \bar{g})$ are obtained as the solution of (27) and (28) written as

$$A\bar{g} = 2a\bar{f} \tag{31}$$

$$A^T\bar{f} = 2b\bar{g} \tag{32}$$

A related eigenvalue problem

- Substituting (32) in (31) and vice versa:

$$AA^T\bar{f} = 4ab\bar{f} \tag{33}$$

$$A^TA\bar{g} = 4ab\bar{g} \tag{34}$$

- That is, setting $\lambda = 4ab$, it follows that $(\lambda, \bar{f})$ is an eigen pair of AA^T and $(\lambda, \bar{g})$ is an eigen pair of A^TA

A first look at the solution to the problem

- Indeed, the pair $(\bar{f}, \bar{g})$ of spatial patterns that maximizes ρ in (24) are given by the eigenvectors of the two Grammian matrices AA^T and $A^T A$ respectively where $A \in R^{m_1 \times m_2}$ is the matrix of the bilinear form in (25)
- Also, notice that the eigenvalues λ are related to the product of the Lagrangian multipliers a and b

Properties of A^TA and AA^T

- For concreteness, let $m_2 < m_1$ and A be of full rank
- $Rank(A) = Rank(A^T) = min\{m_1, m_2\} = m_2$
- $Rank(AA^T) = Rank(A^TA) = m_2$
- Thus, smaller Gramian A^TA is of full rank, and SPD but longer Gramian AA^T is rank deficient and symmetric positive semi-definite

- Let $(\lambda_i, \bar{g}_i)$, $1 \leq i \leq m_2$ be the eigen pairs of $A^T A$. That is, $(A^T A)\bar{g}_i = \bar{g}_i \lambda_i$, $\bar{g}_i \in R^{m_2}$ where

$$\lambda_1 > \lambda_2 > \lambda_3 \cdots > \lambda_{m_2} > 0 \tag{35}$$

- Let

$$\begin{aligned} \bar{G} &= [\bar{g_1}, \bar{g_2}, \ldots, \bar{g_{m_2}}] \in R^{m_2 \times m_2} \\ \Lambda &= Diag(\lambda_1, \lambda_2, \lambda_3 \ldots, \lambda_{m_2}) \end{aligned} \tag{36}$$

- Then :

$$\begin{aligned} (A^T A)\bar{G} &= \bar{G}\Lambda \\ \bar{G}^T \bar{G} = \bar{G}\bar{G}^T &= I_{m_2} \end{aligned} \tag{37}$$

Eigen structure of AA^T

- Define

$$\bar{f}_i = \frac{1}{\sqrt{\lambda_i}} A\bar{g}_i \in R^{m_1}, 1 \leq i \leq m_2 \tag{38}$$

- Verify:

$$(AA^T)\bar{f}_i = \lambda_i \bar{f}_i \tag{39}$$

- Set: $\bar{F} = [\bar{f}_1, \bar{f}_2, \ldots, \bar{f_{m_2}}] \in R^{m_1 \times m_2}$
- Then

$$\begin{aligned} (AA^T)\bar{F} &= \bar{F}\Lambda \\ \bar{F}^T\bar{F} &= I_{m_2} \end{aligned} \tag{40}$$

- Rewriting (38)

$$A\bar{g}_i = \bar{f}_i \sqrt{\lambda_i}, \quad 1 \leq i \leq m_2$$

- That is,

$$A\bar{G} = \bar{F}\Lambda^{1/2} \quad or \quad A = \bar{F}\Lambda^{1/2}\bar{G}^T \tag{41}$$

is called the singular value decomposition of A

- λ_i's are the eigenvalues of A^TA and AA^T and $\sqrt{\lambda_i}$ are the singular values of A

Back to correlation ρ

- From (24): For $1 \leq i \leq m_2$,

$$\rho_i = \bar{f}_i^{\,T} A \bar{g}_i \tag{42}$$

- Substitute for A using (41):

$$\rho_i = \bar{f}_i^{\,T} \bar{F} \Lambda \bar{G}^T g_i \tag{43}$$

Simplification of (43)

- $$\bar{f}_i^T F = [\bar{f}_i^T \bar{f}_1, \bar{f}_i^T \bar{f}_2, \ldots, \bar{f}_i^T \bar{f}_i, \ldots, \bar{f}_i^T \bar{f}_{m_2}]$$
$$= (0, 0, \ldots, 1, \ldots, 0) = e_i^T \in R^{m_2} \qquad (44)$$
the i^{th} unit vector in R^{m_2}

- $$\bar{G}^T g_i = \begin{bmatrix} \bar{g}_1^T g_i \\ \bar{g}_2^T g_i \\ \vdots \\ \bar{g}_i^T g_i \\ \vdots \\ \bar{g}_{m_2}^T g_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = e_i \qquad (45)$$

Maximum value of ρ_i along $(\bar{f}_i, \bar{g}_i)$

- Substituting (44) and (45) in (43):

$$\rho_i = e_i^T \Lambda^{1/2} e_i = \sqrt{\lambda_i} \tag{46}$$

  the i^{th} singular value of A

- Since $\lambda_i > 0$ and are ordered as in (35),it follows that ρ_i attains its i^{th} maximum value along the spatial pattern pair $(\bar{f}_i, \bar{g}_i)$

Singular value and Lagrangian multipliers

- From (33)-(34):

$$\lambda = 4ab \quad or \quad \sqrt{\lambda} = 2\sqrt{ab}$$

  that is singular values are proportional to the square root of the product of the two Lagrangian multipliers

- When $m_1 = m_2$ from (31)-(32), we get

$$0 = \bar{f}^T A\bar{g} - \bar{g}^T \bar{A}\bar{f} = 2\bar{f}^T \bar{g}(a - b) \tag{47}$$

  it follows that a $=$ b since $\bar{f}^T \bar{g} \neq 0$

References

- The books
  W.Hardle and L.Simar(2003) Applied Multivariate Statistical Analysis, Springer Verlag
- H.Von Storch and F.Zwiers(1999) Statistical Analysis in climate Research, Cambridge University Press, contain very good introduction to CCA.