Department of Atmospheric Sciences
National Taiwan University

Chaos and Predictability/ HW4

Kai-Chih Tseng

2025 Fall

Question 1 : Fingerprint method

The **fingerprint method** is a statistical technique used to detect a *specific physical signal pattern* embedded in data. The core idea is to compare the covariance structure of the **experiment (or observational) data** with that of a **control run representing internal variability**.

We define the fingerprint matrix

$$S = C_{\text{exp}}\, C_{\text{ctrl}}^{-1},$$

where C_{exp} and C_{ctrl} denote the covariance matrices of the experiment and control datasets, respectively. An eigen-decomposition of S yields spatial patterns, and the eigenvector associated with the largest eigenvalue is called the **fingerprint**.

In other words, the fingerprint method identifies the spatial direction in which the signal is maximally amplified, while the contamination by internal variability is minimized.

Mathematical Formulation

Let $x(t) \in \mathbb{R}^n$ be an n-dimensional time series (e.g., climate fields expressed in an EOF basis). Assume the experiment data consist of internal variability plus a deterministic signal:

$$x_{\text{exp}}(t) = x_{\text{ctrl}}(t) + s(t)\, p, \tag{1}$$

where

- $p \in \mathbb{R}^n$ is the *spatial signal pattern*, and
- $s(t)$ is a scalar amplitude describing the temporal evolution of the signal.

We define the covariance matrices

$$C_{\text{ctrl}} = \mathbb{E}[\, x_{\text{ctrl}} x_{\text{ctrl}}^T \,], \qquad C_{\text{exp}} = \mathbb{E}[\, x_{\text{exp}} x_{\text{exp}}^T \,].$$

Substituting the signal—noise decomposition into C_{exp}, we obtain

$$C_{\text{exp}} = C_{\text{ctrl}} + \mathbb{E}[\, s^2(t) \,] pp^T.$$

Thus the difference between experimental and control covariances can be formulated as:

$$C_{\text{exp}} = C_{\text{ctrl}} + \alpha\, pp^T, \qquad \alpha = \mathbb{E}[\, s^2(t) \,].$$

The fingerprint matrix is defined as

$$S = C_{\text{exp}} C_{\text{ctrl}}^{-1} = I + \alpha\, pp^T C_{\text{ctrl}}^{-1}.$$

We perform an eigen-decomposition

$$Sv = \lambda v.$$

where the eigenvector associated with the largest eigenvalue rotates toward the direction of the signal pattern p. Therefore,

$$\text{largest eigenvector of } S \;=\; \text{fingerprint}.$$

In the hangouts we conducted over the past few weeks, the quantities appearing in Eq. (1) correspond to the following components in our implementation:

- $x_{\text{exp}}(t)$: the noisy experimental data, implemented as `noise[:,:,:] + signal[:,:,:]`
- $s(t)$: the imposed linear trend, given by `linspace(1,4,100)`
- p: the spatial signal pattern, represented by the 2nd-order Legendre polynomial
- $x_{\text{ctrl}}(t)$: the internally generated Gaussian noise, constructed as $\sum_n P_n(x), W_n(t)$

Problem

Based on this framework, please address the following tasks:

1. Propose your own version of Eq. (1) by choosing a different form of $s(t)$, p, or $x_{\text{ctrl}}(t)$, and briefly explain the reasoning behind your design.
2. Apply the fingerprint method to detect the spatial structure of your chosen signal.

Quention2 : Paper reading

1. In (Sen Zhao,2019), They proposed a stochastic dynamical model for Indian Ocean Dipole (IOD) :

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\lambda(t)T(t) + \alpha(t)T_{ENSO}(t) + \sigma_0\xi(t) \tag{2}$$

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = -m\xi(t) + w(t) \tag{3}$$

We note that this system is driven by an external forcing term $T_{\text{ENSO}}(t)$. Recall that in HW1, Questions 3.3 and 4, we analyzed how external forcing influences the characteristic time scale of a dynamical system.

Please discuss the similarities and differences between the present system and the HW1 system in terms of their characteristic time scales.

Also, we note that the red noise process/OU process here (eq.(3)) is the system we mentioned in Question2 of HW3, with $\Phi = -m\xi$. Based on your knowledge, please discuss how the behavior of T in eq.(2) and ξ in eq.(3) is affect by each term of the system.

2. In (Ting Liu, 2021), they discussed the ENSO Predictability over the Past 137 Years. based on the concepts mentioned in course materials, they defined a new predictability measures under Gaussian assumption.

   Please discuss the new insights using relative entropy, in compare with actual prediction skill measures.

Question 3: Paper reading (Granger causality)

Consider 2 variables X, Y. Intuitutively, Granger causality is based on this argument:

$$\text{If we have an additional information X, does it improves the prediction of Y ?} \tag{4}$$

If so, in an appropriate statistical sense, then we call it "X Granger causes Y". For example, Let us say we want to predict if it will rain today (Y), then we consider 2 scenarios :

- Scenario A : we only look at yesterday's rain data. then we can make a decent guess
- Scenario B : we look at yesterday's rain data plus the humidity from yesterday (X).

If Scenario B gives you a significantly more accurate prediction than Scenario A, we say that Humidity Granger-causes Rain. It simply means humidity contains unique information about future rain that past rain data does not.

Mathematical formulation

A p-order vector autoregressive model, VAR(p), for an n-dimensional stationary stochastic process $\mathbf{U}$, sampled at discrete time indices t, is defined as

$$\mathbf{U}_t = \sum_{i}^{p} \mathbf{A}_i \mathbf{U}_{t-i} + \varepsilon_t \tag{5}$$

The intuition of this model is to suppose $\mathbf{U}_t$ is a function, in the form of linear combination, of past states $\mathbf{U}_{t-1}, \mathbf{U}_{t-2}, \cdots \mathbf{U}_{t-p}$. If p=1, then this model is exactly the red noise process, which is the discretized version of eq.(3). Based on this model, we consider 2 scenario :

- Scenario A : we only look at $\mathbf{Y}$ it self,

$$\mathbf{Y}_t = \sum_{i=1}^{p} \mathbf{A}'_{yy} \mathbf{Y}_{t-i} + \varepsilon'_{y,t} \tag{6}$$

- Scenario B : we look at past data of $\mathbf{Y}$ **plus** the past data of $\mathbf{X}$:

$$\mathbf{Y}_t = \sum_{i=1}^{p} \mathbf{A}_{yy} \mathbf{Y}_{t-i} + \sum_{i=1}^{p} \mathbf{A}_{yx} \mathbf{X}_{t-i} + \varepsilon_{y,t} \tag{7}$$

To quantify the improvement of prediction, we consider the statistical difference of 2 residual $\varepsilon_{y,t}$, $\varepsilon'_{y,t}$, calculate the log-likelihood ratio

$$\mathscr{F}_{\mathbf{X}\to\mathbf{Y}} = \log \frac{|\Sigma'_{yy}|}{|\Sigma_{yy}|} \tag{8}$$

where $\Sigma_{yy} = cov(\varepsilon_{y,t})$, $\Sigma_{yy} = cov(\varepsilon_{y,t})$

Assumption and Limitation

For a Granger Causality to be valid, our data should satisfy:

- Stationarity: The mean and variance of the data should not change over time. if, not, then we have to transform it to a stationary variable.
- Linearity: we assume the current state is linearly depends on the past states.
- Lag selection: the result might depends on the selection of p in eq.(5). An practical way to have an objective selection is using AIC (Akiake information criterion), which is ,roughly speaking, select such p that minimize prediction error defined by p and log-likelihood function.

Also, we note that "X Granger cause Y" doesn't mean "X cause Y" physically, it only depends on the chronological order of X and Y. For example, we know "A rooster crows before the sun rises every morning". If we run a Granger Causality test, we obtain "The rooster Granger cause the sunrise" because the crowing is a perfect predictor of the sun coming up. However, in reality, we know the rooster doesn't cause the sunrise.

In short, Granger causality detects temporal precedence (which comes first) and predictive power, not physical mechanisms.

Problems

In (Eviatar 2019), they considered $\mathscr{F}_{Atmos \to SST}$ and $\mathscr{F}_{SST \to Atmos}$, discussed the difference between these 2 amount under different setting. Please discuss what they found by using Granger causality method, and it's interpretation.