Detecting generalized synchronization between chaotic signals: a kernel-based approach

Hiromichi Suetani[1,3]**, Yukito Iba**[2] **and Kazuyuki Aihara**[1,3]

[1] Aihara Complexity Modelling Project, ERATO, JST, 3-23-5 Uehara, Shibuya-ku,
Tokyo 151-0064, Japan
[2] The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569,
Japan
[3] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku,
Tokyo 153-8505, Japan

E-mail: suetani@aihara.jst.go.jp

Abstract

A unified framework for analysing generalized synchronization in coupled
chaotic systems from data is proposed. The key of the proposed approach is the
use of the *kernel methods* recently developed in the field of machine learning.
Several successful applications are presented, which show the capability of
the kernel-based approach for detecting generalized synchronization, and
dynamical change of the coupling strength between two chaotic systems can
be captured by the proposed approach. It is also discussed how the kernel
parameter is suitably chosen from data.

PACS numbers: 05.45.Xt, 05.45.Tp, 02.50.Sk, 05.10.−a

1. Introduction

Synchronization of chaotic systems has been an active research area in recent years
[1]. Now the notion of synchronization of chaotic systems is extended far beyond
complete synchronization between identical systems [2–4], and various kinds of nonlinear
synchronization have been proposed [1]. A significant extension is generalized
synchronization (GS) [5], which is defined by a time-independent nonlinear functional relation
$y = \Psi(x)$ between the states x and y of two systems X and Y.

Experimental detection and characterization of GS from observed data is a challenging
problem, especially in biology, e.g., for study on nonlinear interdependence observed in
binding of different features in cognitive process [6] and epilepsies in the brain [7, 8].

In unidirectionally coupled systems, a way to detect GS is to make an identical copy Y' of
the response system Y driven by the common signal from the driver system X, then investigate

whether orbits of both Y and Y' coincide after transient [9, 10]. However, it is very difficult to prepare an identical copy of the response system, and this approach does not give any information on the structure of the synchronization manifold $\mathcal{M} = \{(x, y) : y = \Psi(x)\}$ in the state space.

On the other hand, several indices have been proposed [5, 7, 8, 11–13] to quantify the nonlinear dependence between X and Y based on a local relation between them. While the usefulness of these indices were shown in some examples [7, 8, 11], those approaches are still ad hoc, lacking systematic methods for unifying the local relation of different regions in the state space.

Recently, kernel methods have been attracting much attention in the machine learning community as powerful tools for analysing data with high nonlinearity [14–17]. In this paper, we propose a novel approach for analysing GS based on kernel canonical correlation analysis (kernel CCA) [18–21] which is a version of the kernel methods [22][4]. We will demonstrate that the kernel CCA provides a suitable index for measuring the nonlinear interdependence and gives global coordinates characterizing the synchronization manifold $\mathcal{M}$ of GS. Note that the latter has not been addressed in conventional studies on the analysis of GS. The proposed method does not require explicit knowledge on the underlying equations behind time series. Although we restrict our attention to the analysis of numerical experiments here, it is also applicable to experimental data.

The present paper is organized as follows. In section 2, we introduce a formulation of the kernel CCA. Sections 3 and 4 provide several successful applications of the kernel CCA for treating GS. In section 5, it is demonstrated that the nonstationary change of the coupling strength between two chaotic systems can be captured by the proposed approach. In section 6, we discuss an issue on the choice of kernel parameters. Our conclusion is given in section 7.

2. Kernel canonical correlation analysis

Canonical correlation analysis (CCA) is a multivariate analysis method to find a pair of vectors that maximize the correlation coefficient between projections of signals onto them [23]. CCA is useful for detecting a linear relation between a pair of multidimensional data sets, but cannot deal with the case like GS where there is a high nonlinear relation between two data sets. Extending the ability of CCA for analysing data with nonlinearity, kernelization of CCA has been proposed by several researchers [18–21]. An intuitive formulation of the kernel CCA is as follows.

For a pair of multivariate variables (x, y) with $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$, the kernel CCA seeks a pair of nonlinear scalar functions $f : \mathbb{R}^p \to \mathbb{R}$ and $g : \mathbb{R}^q \to \mathbb{R}$ such that an estimator of the correlation coefficient

$$\rho_{\mathcal{F}} = \frac{\mathsf{cov}(f(x), g(y))}{\sqrt{\mathsf{var}(f(x))}\sqrt{\mathsf{var}(g(y))}} \tag{1}$$

between $f(x)$ and $g(y)$ is maximized.

Methods based on expression (1) of the generalized correlation coefficient have a long history [24, 25]. However the use of the kernel approach gives a notable progress in the subject, with a remarkably simple and flexible way of treating (1).

---

[4] Preliminary results are reported in a conference proceedings [22] with a simpler example of detection of GS. Here, we give comprehensive treatments with examples of practical interests.

The essence of the kernel methods is an assumption that nonlinear functions f and g are well approximated by linear combinations of *kernels* on data points (x_n, y_n),

$$f(x) = \sum_{n=1}^{N} \alpha_n k(x_n, x), \qquad g(y) = \sum_{n=1}^{N} \beta_n k(y_n, y), \tag{2}$$

where $\{(x_n, y_n)\}_{n=1}^{N}$ denotes a training data set of (x, y).

Examples of kernels are $k(x, x') = e^{\frac{-\|x-x'\|^2}{2\sigma^2}}$ (Gaussian), $((x \cdot x') + \theta)^d)$(polynomial), and $\tanh(\eta(x \cdot x') + \theta)$(sigmoidal) [15–17]. A rich representation of nonlinear functions is allowed by suitably chosen weight coefficients $\{\alpha_n\}_{n=1}^{N}$, $\{\beta_n\}_{n=1}^{N}$ and kernel parameters. At first sight, the dependence of the expressions for f and g on the given data $\{(x_n, y_n)\}_{n=1}^{N}$ brings some ad hoc nature to the method. The assumption is, however, equivalent to the assumption that f and g are minimum norm functions in a reproducing kernel Hilbert space (RKHS). It is proved by invoking representer theorem in the theory of RKHS [15–17].

Substituting (2) into (1) and replacing covariance $\text{cov}(\cdot, \cdot)$ and variances $\text{var}(\cdot)$ with empirical averages over the data set $\{(x_n, y_n)\}_{n=1}^{N}$, the kernel CCA reduces to the maximization of

$$\rho_{\mathcal{F}} = \frac{{}^t\alpha K_X K_Y \beta}{\sqrt{{}^t\alpha K_X^2 \alpha}\sqrt{{}^t\beta K_Y^2 \beta}}, \tag{3}$$

where $\alpha = {}^t(\alpha_1, \alpha_2, \ldots, \alpha_N)$, $\beta = {}^t(\beta_1, \beta_2, \ldots, \beta_N)$ and K_X, K_Y are the *Gram matrices* $(K_X)_{i,j} = k(x_i, x_j)$ and $(K_Y)_{i,j} = k(y_i, y_j)$ defined with the given sample. The Gram matrix contains relevant topological information of data points in the state space. For simplicity, we tentatively assume that the averages of $\{f(x_n)\}_{n=1}^{N}$ and $\{g(y_n)\}_{n=1}^{N}$ are zero. Later we will show that subtraction of the averages from data can be done in an implicit way and does not cause any technical difficulty.

A naive generalization of CCA leads to the maximization of ${}^t\alpha K_X K_Y \beta$ subject to ${}^t\alpha K_X^2 \alpha = {}^t\beta K_Y^2 \beta = 1$. The corresponding Lagrangian is

$$\mathcal{L}_0 = {}^t\alpha K_X K_Y \beta - \frac{\rho_X}{2}\left({}^t\alpha K_X^2 \alpha - 1\right) - \frac{\rho_Y}{2}\left({}^t\beta K_Y^2 \beta - 1\right), \tag{4}$$

where ρ_X and ρ_Y are the Lagrange multipliers. Taking derivatives of $\mathcal{L}_0$ with respect to α and β, and $\rho_X = \rho_Y = \rho$ from the constraint, the kernel CCA is formulated as the following generalized eigenvalue problem:

$$\begin{pmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} K_X^2 & 0 \\ 0 & K_Y^2 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{5}$$

There is, however, a difficulty when the Gram matrices K_X and K_Y are invertible. By multiplying both sides of (5) by $\left(K_X^{-1} K_X^{-1}, K_Y^{-1} K_Y^{-1}\right)$ from the left-hand side, we have

$$\alpha = \frac{1}{\rho} K_X^{-1} K_Y \beta, \tag{6}$$

$$\beta = \frac{1}{\rho} K_Y^{-1} K_X \alpha. \tag{7}$$

Substituting (7) into (6), we obtain

$$I\alpha = \rho^2 \alpha, \tag{8}$$

which gives a trivial solution $\rho = \pm 1$. Such a situation is not uncommon, because the Gram matrix is invertible when a Gaussian kernel is used and data points are distinct from each other.

Thus the naive kernelization (5) of CCA does not provide useful information on the nonlinear dependence.

To overcome this problem, we introduce small regularization terms in the denominator of the right-hand side of (3) as

$$
\rho_{\mathcal{F}} = \frac{{}^t\alpha K_X K_Y \beta}{\sqrt{{}^t\alpha K_X^2 \alpha + \kappa \|f\|_{\mathcal{F}}^2}\sqrt{{}^t\beta K_Y^2 \beta + \kappa \|g\|_{\mathcal{F}}^2}}, \tag{9}
$$

where $\|f\|_{\mathcal{F}}^2$ and $\|g\|_{\mathcal{F}}^2$ are the quadratic norms of f and g on the RKHS defined as

$$
\|f\|_{\mathcal{F}}^2 = \sum_{n'=1}^{N} \alpha_{n'} f(x_{n'}) = \sum_{n'=1}^{N}\sum_{n=1}^{N} \alpha_{n'}\alpha_n k(x_n, x_{n'}) = {}^t\alpha K_X \alpha, \tag{10}
$$

$$
\|g\|_{\mathcal{F}}^2 = \sum_{n'=1}^{N} \beta_{n'} g(y_{n'}) = \sum_{n'=1}^{N}\sum_{n=1}^{N} \beta_{n'}\beta_n k(y_n, y_{n'}) = {}^t\beta K_Y \beta, \tag{11}
$$

and κ is its parameter. Note that although the parameter $\kappa \neq 0$ is required for a nontrivial result, the precise value of κ is not important[5]. This insensitivity to the value of κ will be confirmed by a numerical experiment in the following section. As well as (5), the maximization of the numerator ${}^t\alpha K_X K_Y \beta$ in (9) subject to ${}^t\alpha K_X^2 \alpha + \kappa^t\alpha K_X \alpha = {}^t\beta K_Y^2 \beta + \kappa^t\beta K_Y \beta = 1$ reduces to the following generalized eigenvalue problem:

$$
\begin{pmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} K_X(K_X + \kappa I) & 0 \\ 0 & K_Y(K_Y + \kappa I) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{12}
$$

The first eigenvalue of (12) gives the maximal value $\rho_{\mathcal{F}}^{\max}$ of $\rho_{\mathcal{F}}$ in (9). $\rho_{\mathcal{F}}^{\max}$ is called the *canonical correlation coefficient*, and the variables $u = f(x)$ and $v = g(y)$ transformed by f and g are called the canonical variates of the kernel CCA.

So far, the averages of $\{f(x_n)\}_{n=1}^{N}$ and $\{g(y_n)\}_{n=1}^{N}$ are set to zero. In fact, it is not necessary to subtract the averages explicitly, because the subtraction of the means is equivalent to the replacement of the Gram matrices K with

$$
\tilde{K} = K - \frac{1}{N}(\mathbf{j}\,{}^t\mathbf{j})K - \frac{1}{N}K(\mathbf{j}\,{}^t\mathbf{j}) + \frac{1}{N^2}(\mathbf{j}\,{}^t\mathbf{j})K(\mathbf{j}\,{}^t\mathbf{j}), \tag{13}
$$

where $\mathbf{j}$ is an N-dimensional vector such that each component equals to the unity [15–17].

In the examples of this paper, the data $\{x_n\}_{n=1}^{N}$ and $\{y_n\}_{n=1}^{N}$ are normalized within a unit interval [0, 1] before applying the kernel CCA.

3. Detecting generalized synchronization by the kernel CCA

The kernel methods such as the kernel CCA have been mainly applied to problems of pattern recognition [17] and bioinformatics [27]. In this paper, we propose that the kernel CCA can also be a powerful tool for analysing nonlinear dynamics.

Voss *et al* [26] proposed a method for analysing dynamical systems by the maximization of the expression (1). It is based on the alternating conditional expectation (ACE) algorithm [25], where the transformations f and g in (1) are restricted to a linear combination of univariate functions such as $f(x_1, x_2, \ldots, x_p) = \sum_i \Phi_i(x_i)$, while the kernel CCA allows any nonlinear function $f(x_1, x_2, \ldots, x_p)$ of p variables in a RKHS. This difference makes the proposed

---

[5] Recently, statistical consistency of the kernel CCA, i.e., convergence of the estimators of f and g to the optimal ones is proved when f and g belong to a RKHS, under the condition that $\kappa/N \sim N^{-1/3+\delta}(0 < \delta < 1/3)$ with $N \to \infty$ [41].
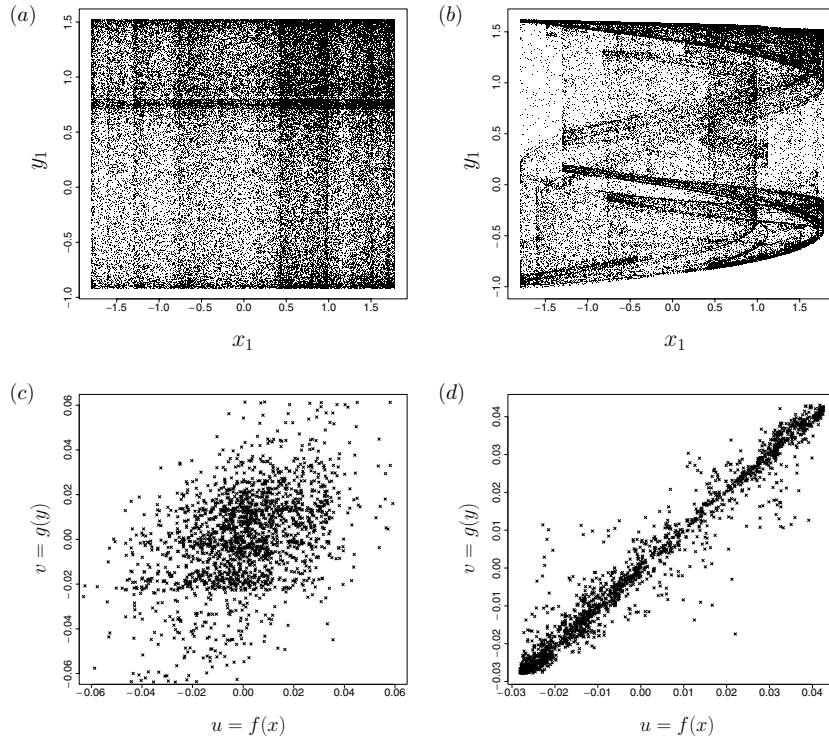
Figure 1. (*a*), (*b*) Projections of strange attractors of the coupled Hénon maps onto the (x_1, y_1) plane with $\gamma = 0.0$ in (*a*) and $\gamma = 0.25$ in (*b*). For both of (*a*) and (*b*), an orbit with length 5×10^4 is used for plotting. (*c*), (*d*) Scatter plots of the canonical variates of the kernel CCA with $\gamma = 0.0$ in (*c*) and $\gamma = 0.25$ in (*d*).

method much more powerful than the one based on ACE, especially in the cases where the nonlinear dependence between variables $x_1, x_2, \ldots$ is essential. Unlike ACE, the kernel CCA does not need any iterative procedure, and it is enough to solve a generalized eigenvalue problem by standard numerical algebra only once for each set of data and parameters.

In this section, we study two unidirectionally coupled Hénon maps [11] as an illustrative example and demonstrate the ability of the proposed approach.

3.1. Quantitative characterization of generalized synchronization

Two unidirectionally coupled Hénon maps are described by the following difference equations:

$$X : \begin{cases} x_1(t + 1) = 1.4 - x_1(t)^2 + 0.3x_2(t) \\ x_2(t + 1) = x_1(t), \end{cases} \tag{14}$$

$$Y : \begin{cases} y_1(t + 1) = 1.4 - \{\gamma x_1(t)y_1(t) + (1 - \gamma)y_1(t)^2\} + 0.1y_2(t) \\ y_2(t + 1) = y_1(t), \end{cases} \tag{15}$$

where $x(t) = (x_1(t), x_2(t))$ and $y(t) = (y_1(t), y_2(t))$ are state variables at time t and γ denotes the coupling strength between two systems X and Y.

Figures 1(*a*) and (*b*) show the projections of the strange attractors of (14) and (15) with $\gamma = 0.0$ and 0.25 onto the (x_1, y_1) plane, respectively. When the coupling strength γ is large,
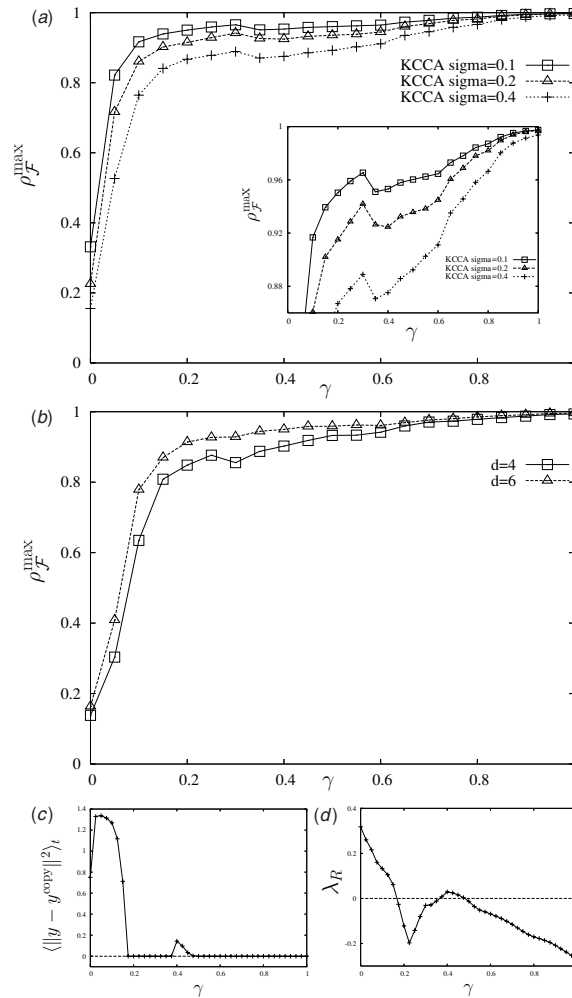
Figure 2. (*a*) The canonical correlation coefficient $\rho_{\mathcal{F}}^{\max}$ of the kernel CCA as a function of the coupling strength γ for three different values of σ with $\kappa = 0.1$ and $N = 2 \times 10^3$. Inset an enlargement of (*a*). (*b*) The index $\rho_{\mathcal{F}}^{\max}$ versus γ for the embedding dimension $d = 4$ and $d = 6$ with $\sigma = 1.0, \kappa = 0.1$, and $N = 2 \times 10^3$. The delay time l for embedding is set to 1. (*c*) The time average of the synchronization error between the response system and its identical copy as a function of γ. (*d*) The maximal Lyapunov exponent λ_R of the response system as a function of γ.

a complicated driver–response relation with high nonlinearity is formed between two different chaotic dynamics. It can be identified in figure 1(*b*) with a visual inspection, but it is not easy to represent it by naive statistical tools such as correlation coefficients.

We apply the kernel CCA to the cases shown in figures 1(*a*) and (*b*). Here, we employ a Gaussian kernel with $\sigma = 0.1$ and κ is set to 0.3. We prepare an orbit with length $N = 2 \times 10^3$ as training data. As already remarked, each of the variables x_1, x_2, y_1 and y_2 is normalized within the unit interval [0, 1]. Figures 1(*c*) and (*d*) illustrate results of the kernel CCA. Figure 1(*d*) shows that GS is clearly identified as a cloud of points along the diagonal on the plane of the canonical variates u and v. On the other hand, when X and Y are independent, the correlation between the canonical variates is very weak as shown in figure 1(*c*). Figure 2(*a*)

shows the dependence of the canonical correlation coefficient $\rho_{\mathcal{F}}^{\max}$ on the coupling strength γ. The rapid increase of $\rho_{\mathcal{F}}^{\max}$ against γ in an interval $0 < \gamma < 0.17$ is indicated.

We will show that the kernel CCA can be used with time-delay embedding scheme for time series. In the time-delay embedding scheme, a pair of d-dimensional vectors $(x_1(t), x_1(t-l), \ldots, x_1(t-(d-1)l))$ and $(y_1(t), y_1(t-l), \ldots, y_1(t-(d-1)l))$, where l denotes the delay time, is used as a sample of training data. Figure 2(b) shows the graph of $\rho_{\mathcal{F}}^{\max}$ versus γ for two different values of the embedding dimension d. Shapes of the graphs shown in figure 2(b) does not change significantly compared to those shown in figure 2(a). This result indicates that the kernel CCA works for data obtained by using the time-delay embedding scheme.

3.2. Comparison with A conventional method

In [9, 10], an identification of GS based on the occurrence of the complete synchronization between the response system and its identical copy is proposed. For the coupled Hénon maps (14) and (15), we investigate the synchronization errors $\langle \| y - y^{\mathrm{copy}} \|^2 \rangle_t$ between the system Y and its identical copy Y' with variables $y^{\mathrm{copy}} = \left(y_1^{\mathrm{copy}}, y_2^{\mathrm{copy}} \right)$. The result is shown in figure 2(c). Here, the index $\langle \| y - y^{\mathrm{copy}} \|^2 \rangle_t$ between Y and Y' is measured by the average of 10^6 time steps and plotted as a function of γ. The index $\langle \| y - y^{\mathrm{copy}} \|^2 \rangle_t$ decreases with increasing γ and becomes zero at $\gamma \sim 0.17$. Our results shown in figures 2(a) and (b) are consistent with this result.

We also investigate the maximal Lyapunov exponent λ_R of the state $y(t) = y^{\mathrm{copy}}(t)$. The result is shown in figure 2(d). The index λ_R is determined from the first eigenvalue of the product of the Jacobian matrix of (15) with respect to the variables y, and an orbit with length 10^6 is used for its numerical evaluation. There is an interval where λ_R changes nonmonotonically against γ. Such nonmonotonic change is also observed in the graph of $\rho_{\mathcal{F}}^{\max}$ versus γ as shown in the inset of figure 2(a). These results tell us that $\rho_{\mathcal{F}}^{\max}$ defined by the kernel CCA can characterize subtle change as well as global tendency of GS.

3.3. Influence of noise and sample size

We consider how the performance of the kernel CCA is influenced by the introduction of the observational noise and the change of the size of training data. First, we consider the influence of noise. In numerical simulation, for each variable of (14) and (15) normalized within the interval $[0, 1]$, Gaussian random numbers with the mean zero and the standard deviation e are added as the observational noise. Results are shown in figure 3. For a low noise level ($e = 0.01$), the graph of $\rho_{\mathcal{F}}^{\max}$ versus γ is almost same as that for the noise free case. For higher noise levels ($e = 0.05$), there is a moderate decrease of $\rho_{\mathcal{F}}^{\max}$ in the whole interval of γ, however, the global tendency of the graph of $\rho_{\mathcal{F}}^{\max}$ versus γ is not lost. The proposed approach is fairly robust against noise except for extremely high noise level ($e = 0.3$).

Second, we study the influence of the size of training data. Figure 4(a) shows the average of $\rho_{\mathcal{F}}^{\max}$ over 20 realizations as a function of γ for several different values of the size N. With the exception of $\gamma = 0$, the graph of $\rho_{\mathcal{F}}^{\max}$ versus γ does not depend much on N. The result shows that the kernel CCA works even with relatively small size of training data. For the cases of $\gamma = 0$ and $\gamma = 0.25$, the graphs of $\rho_{\mathcal{F}}^{\max}$ versus N are shown in figure 4(b). Here, the vertical bars denote the corresponding standard deviation. The average of $\rho_{\mathcal{F}}^{\max}$ for $\gamma = 0$ increases monotonically with decreasing N whereas the one for $\gamma = 0.25$ does not almost change against N. The result with $\gamma = 0$ suggests that a proper choice of the kernel parameter for a given data is important for obtaining correct results. This issue is discussed in section 6.
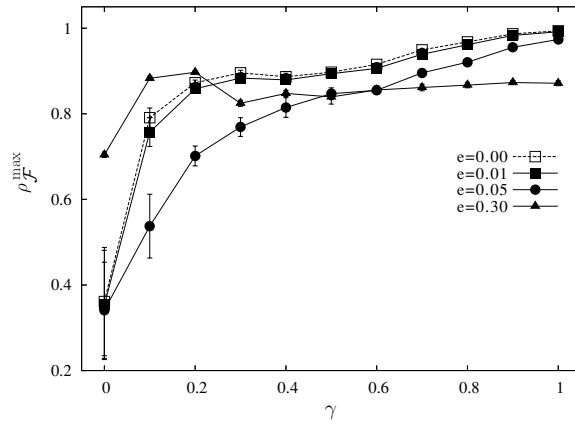
Figure 3. Influence of the observational noise on the index $\rho_{\mathcal{F}}^{\max}$ as a function of γ with $\sigma = 0.4$, $\kappa = 0.1$ and $N = 10^3$. The average and the standard deviation of $\rho_{\mathcal{F}}^{\max}$ over 20 realizations are plotted as symbols and vertical bars, respectively.
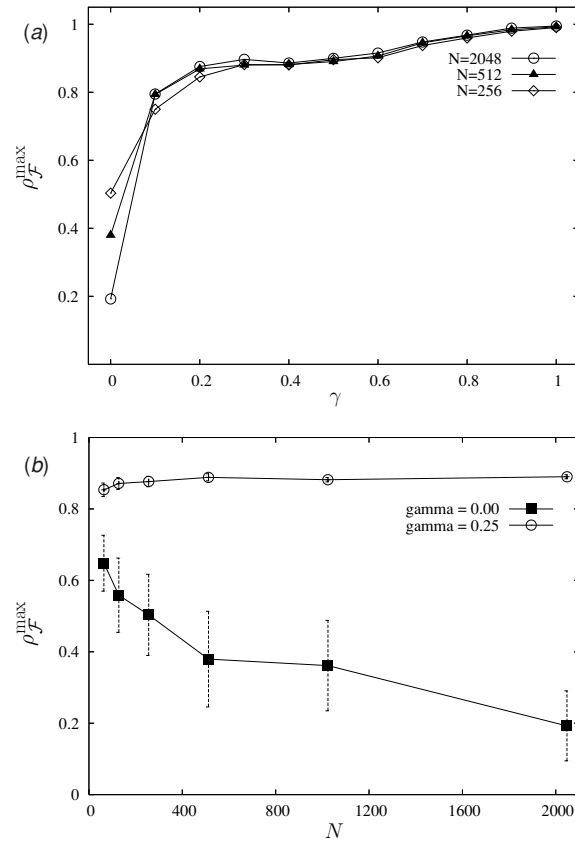

Figure 4. (*a*) Dependence of the index $\rho_{\mathcal{F}}^{\max}$ on γ for three different values of the size N of training data with $\sigma = 0.4$, $\kappa = 0.1$. (*b*) Dependence of the index $\rho_{\mathcal{F}}^{\max}$ on N for $\gamma = 0$ and $\gamma = 0.25$ with $\sigma = 0.4$, $\kappa = 0.1$. The average of $\rho_{\mathcal{F}}^{\max}$ over 20 realizations are plotted as symbols in (*a*), and in addition to the average, the standard deviation is also plotted as vertical bars in (*b*).
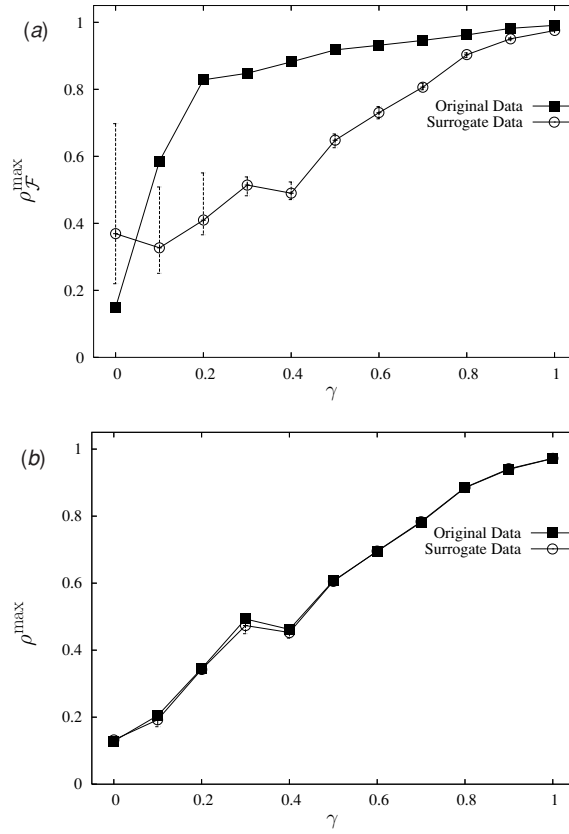
Figure 5. The canonical correlation coefficients $\rho_{\mathcal{F}}^{\max}$ of the kernel CCA (a) and $\rho^{\max}$ of the linear CCA (b) as functions of γ for the original data and the surrogate data. For the surrogate data, the average, and the maximal and minimal values of $\rho_{\mathcal{F}}^{\max}$ over 19 realizations are plotted as symbols, and both edges of vertical bars, respectively. Parameters are $d = 4, l = 1, \sigma = 1.0, \kappa = 0.1$ and $N = 10^3$ in (a) and $d = 4, l = 1$ and $N = 10^3$ in (b).

3.4. Assessing sensitivity of kernel CCA to nonlinear structure: surrogate data analysis

In order to investigate the ability of the kernel CCA to the nonlinear dependence between two systems (14) and (15), we use a method of surrogation [28]. Multivariate surrogate data are generated as follows: first, the Fourier transform of the time series is calculated for each of variables, then the *common* random numbers are added to the phase variables and finally the inverse Fourier transform is applied. The resulting multivariate time series have the same power spectra and cross-spectra as those of the original time series. By changing random numbers added to the phases, an arbitrary number of different time series which preserve the linear properties of the original is obtained. See papers [28–30] for technical details.

In numerical simulation, 19 realizations of the surrogate data for the time series $\{(x_1(t), y_1(t))\}$ of (14) and (15) are prepared by using the TISEAN package [28, 31]. We take $d = 4$ and $l = 1$ for time-delay embedding.

In figure 5(a), the index $\rho_{\mathcal{F}}^{\max}$ defined by the kernel CCA for the original data and that for the surrogate data as functions of γ are shown. For the surrogate data, the average over 19 realizations is plotted as a function of γ, and the corresponding maximal and minimal values
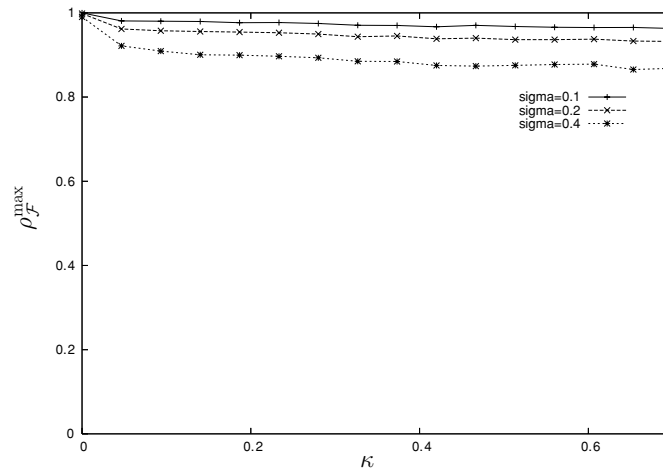
Figure 6. The index $\rho_{\mathcal{F}}^{\max}$ as a function of the regularization parameter κ for three different values of σ. $\gamma = 0.3$ and $N = 3 \times 10^2$.

are also shown as the both edges of vertical bars. For both of the original and the surrogate data, the index $\rho_{\mathcal{F}}^{\max}$ increases with increasing γ. Except for $\gamma = 0$ and $\gamma \sim 1$, however, the value of $\rho_{\mathcal{F}}^{\max}$ for the original data is significantly higher than that for the surrogate data. For larger values of γ, $\rho_{\mathcal{F}}^{\max}$ for the surrogate data increases monotonically, and $\rho_{\mathcal{F}}^{\max} \to 1$ as $\gamma \to 1$. This coincides with the fact that the attractor of the systems (14) and (15) is located around the plane $x_1 = y_1$ and $x_2 = y_2$, and the relation between two systems becomes almost linear one. The results suggest that the kernel CCA is sensitive to nonlinearity of the dependence between two systems.

We also investigate the performance of the linear CCA in the same way and the results are shown in figure 5(*b*). In this case, the difference between the maximal canonical correlation coefficients $\rho^{\max}$ of the linear CCA for the original data and that for the surrogate data is not significant for any value of γ. This indicates that the linear CCA can detect only the linear dependence between two systems.

3.5. The regularization parameter κ

As mentioned in the preceding section, the regularization terms are required for nontrivial results. Figure 6 shows the dependence of $\rho_{\mathcal{F}}^{\max}$ on the regularization parameter κ for three different values of σ. Although $\rho_{\mathcal{F}}^{\max} \to 1$ for too small κ, the value of $\rho_{\mathcal{F}}^{\max}$ decreases gradually and does not depend on the precise value of κ.

4. Other examples

In order to illustrate the capability of the kernel CCA in more complicated situations, we add the following three examples.

4.1. Coupled Rössler–Lorenz systems

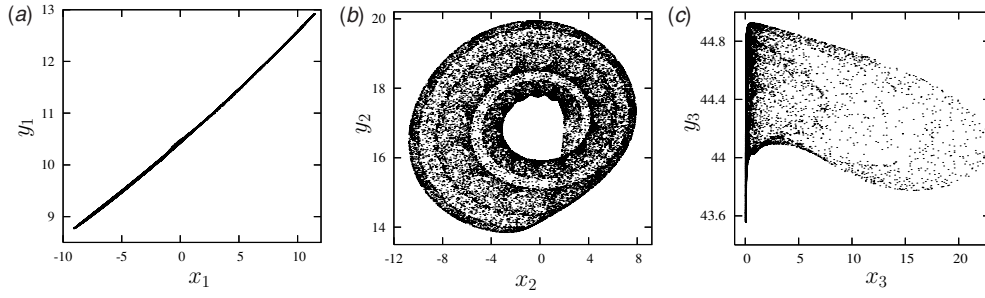First, we consider GS in a Lorenz system driven by a Rössler system [10]:

Figure 7. Projections of the strange attractor of the coupled Rössler and Lorenz systems with $\gamma = 10$ onto the planes of (x_1, y_1) in (a), (x_2, y_2) in (b) and (x_3, y_3) in (c).

$$X : \begin{cases} \dot{x}_1 = -(x_2 + x_3) \\ \dot{x}_2 = x_1 + 0.2x_2 \\ \dot{x}_3 = 0.2 + x_3(x_1 - 5.7), \end{cases} \tag{16}$$

$$Y : \begin{cases} \dot{y}_1 = 16(y_2 - y_1) - \gamma(y_1 - x_1) \\ \dot{y}_2 = -y_1 y_3 + 45.92 y_1 - y_2 \\ \dot{y}_3 = y_1 y_2 - 4y_3. \end{cases} \tag{17}$$

The coupling term is introduced in the first equation of (17), and γ is its strength.

We confirm that there is a sharp transition of GS at $\gamma \sim 4.8$ by investigating the long time average of the synchronization error between the system (17) and its identical copy driven by the common signal $x_1(t)$ of the system (16). There coexist two attractors in the state space after the transition of GS and we choose one of them. Figures 7(a)–(c) show the projections of the strange attractor of the systems (16) and (17) with $\gamma = 10$ onto the planes of (x_1, y_1), (x_2, y_2) and (x_3, y_3) respectively.

For each of the pairs shown in figures 7(a)–(c), the time series is embedded as points in a d-dimensional state space, and we apply the kernel CCA. We set the embedding dimension $d = 6$ and the delay time $l = 2.5$. The size of training data is $N = 10^3$. Results are shown in figure 8(a). All indices $\rho_{\mathcal{F}}^{\max}$ shown in figure 8(a) take large values for $\gamma \gtrsim 4.8$. The value $\gamma \sim 4.8$ agrees with the transition point of GS. In figure 8(b), results of the linear CCA applied to the same data are also shown. The indices $\rho^{\max}$ of the linear CCA for the cases of (x_1 versus y_1) and (x_2 versus y_2) take large values after the GS transition as well as the indices $\rho_{\mathcal{F}}^{\max}$ obtained by the kernel CCA. For the case of (x_3 versus y_3), however, $\rho^{\max}$ of the linear CCA is smaller than $\rho_{\mathcal{F}}^{\max}$ of the kernel CCA. This result suggests that the kernel CCA outperforms the linear CCA when the relation between two observed time series has high nonlinearity.

4.2. Neural spike trains modulated by chaotic inputs

Second, we analyse the following FitzHugh–Nagumo (FHN) neuron model modulated by chaotic dynamics of the Rössler model:

$$X : \begin{cases} \dot{x}_1 = -\tau(x_2 + x_3) \\ \dot{x}_2 = \tau(x_1 + 0.36x_2) \\ \dot{x}_3 = \tau(0.4 + x_3(x_1 - 4.5)), \end{cases} \tag{18}$$

$$Y : \begin{cases} \dot{y}_1 = \{-y_1(y_1 - 0.5)(y_1 - 1) - y_2 + S(t)\}/0.02 \\ \dot{y}_2 = y_1 - y_2 - 0.15, \end{cases} \tag{19}$$
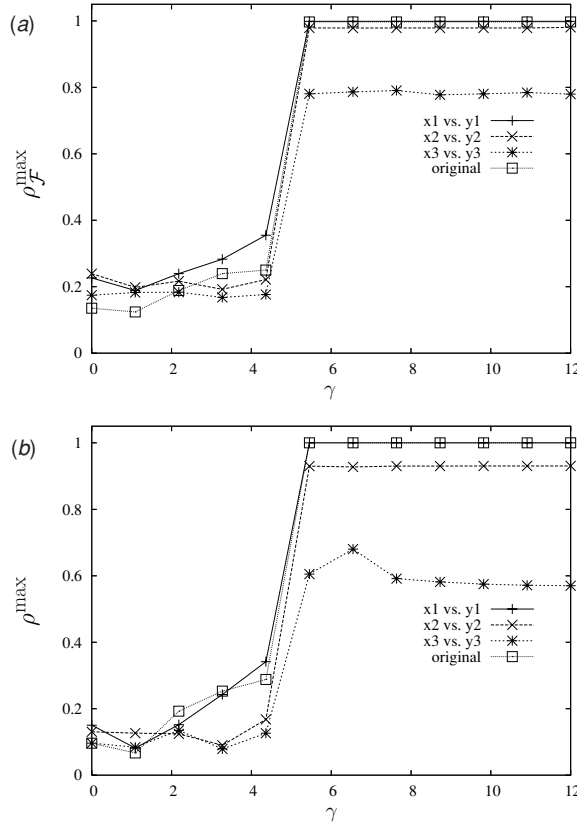
Figure 8. (*a*) The canonical correlation coefficient $\rho_{\mathcal{F}}^{\max}$ of the kernel CCA as a function of the coupling strength γ for different pairs of variables with $N = 10^3$, $d = 6$, $l = 2.5$, $\sigma = 1.5$ and $\kappa = 0.1$. (*b*) The canonical correlation coefficient $\rho^{\max}$ of the linear CCA as a function of the coupling strength γ for different pairs of variables with $N = 10^3$, $d = 6$, and $l = 2.5$. The results for the case where the original state variables (x_1, x_2, x_3) and (y_1, y_2, y_3) are used as training data are also shown.

where the term

$$S(t) = 0.23 + 0.0075 x_1(t) \tag{20}$$

defines a chaotic inputs to the neuron and τ is a parameter that controls the dominant time scale of the Rössler dynamics. The system composed of (18) and (19) has been investigated from the viewpoint of the problem whether the information on the input signal can be decoded from the output interspike intervals (ISIs) generated by a neuron or not [32–34]. Here, we focus on the relation between the input chaotic stimulus and the output ISIs from the viewpoint of GS between two oscillators with different dynamics.

We define the *i*th ISI as $s_i = t_i - t_{i-1}$, where t_i is onset time of the *i*th spike defined as the time when the variable y_1 makes upward crossing over some fixed threshold y_θ. The value of y_θ is set to 0.7 here. We also define the chaotic stimulus associated with the *i*th ISI as $r_i \equiv x_1(t_i)$, which is the value of x_1 in (18) at t_i. By using the delay embedding scheme, we transform the time series $\{r_i\}$ and $\{s_i\}$ into the state points $\{(r_i, r_{i-1}, \ldots, r_{i-d_X+1})\}$ and $\{(s_i, s_{i-1}, \ldots, s_{i-d_Y+1})\}$ in d_X- and d_Y-dimensional state spaces, respectively. We set $d_X = d_Y = 3$ here and apply the kernel CCA to these data sets.
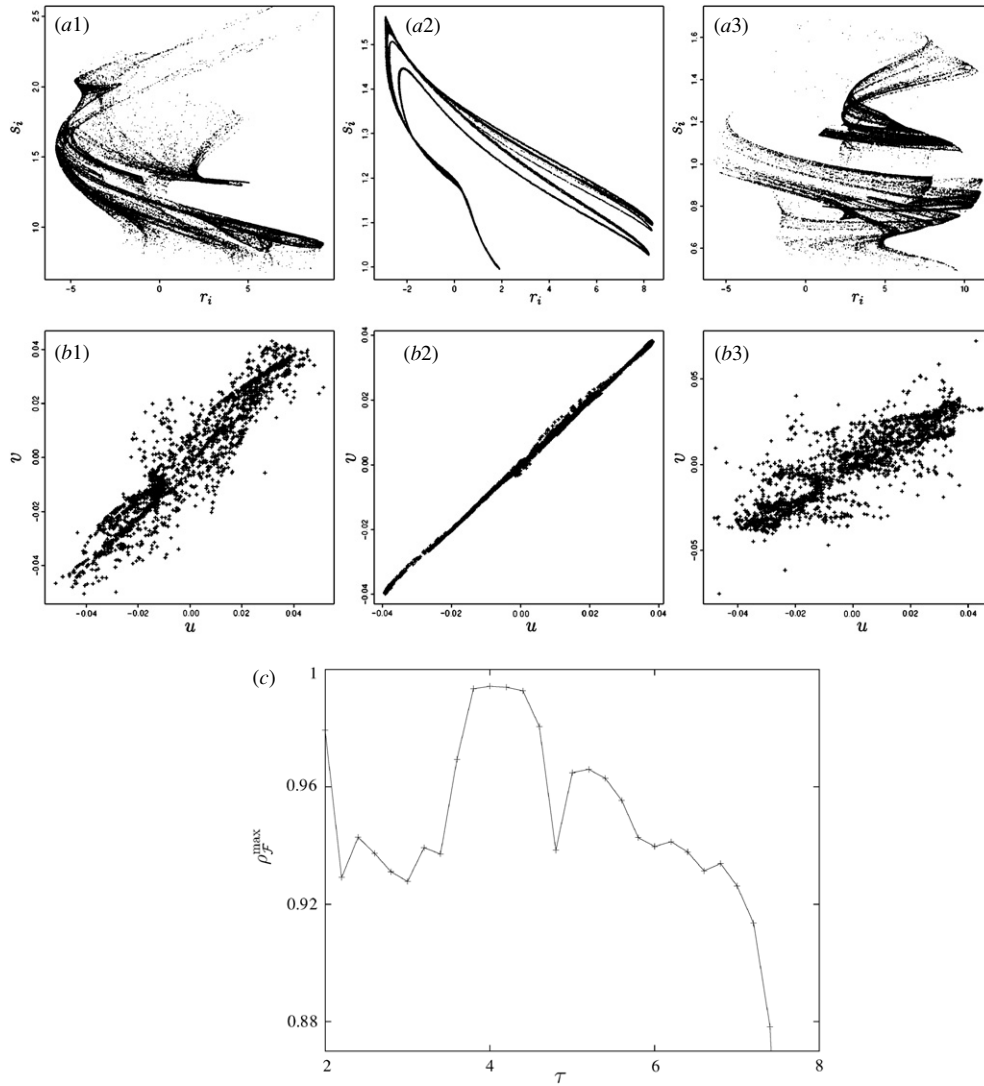
Figure 9. ($a1$)–($a3$) Scatter plots of r_i and s_i of (18) and (19) with $\tau = 2.7$ in ($a1$), $\tau = 4.0$ in ($a2$), and $\tau = 7.5$ in ($a3$). ($b1$)–($b3$) Scatter plots of canonical variates u and v of the kernel CCA with $\tau = 2.7$ in ($b1$), $\tau = 4.0$ in ($b2$), and $\tau = 7.5$ in ($b3$). $\sigma = 0.5$, $\kappa = 0.1$, and $N = 2 \times 10^3$. (c) The canonical correlation coefficient $\rho_{\mathcal{F}}^{\max}$ of the kernel CCA as a function of τ with $\sigma = 0.5$, $\kappa = 0.1$, and $N = 2 \times 10^3$.

In figures 9($a1$)–($a3$), the significant nonlinear dependences between the input chaos and the output ISI are observed in scatter plots of (r_i, s_i). Figures 9($b1$)–($b3$) also show scatter plots of the canonical variates of the kernel CCA associated with figures 9($a1$)–($a3$). It is easy to see that the correlation between canonical variates u and v shown in figures 9($b1$)–($b3$) corresponds to the complexity of the nonlinear dependence between r_i and x_i shown in figures 9($a1$)–($a3$).

The relation between r_i and s_i changes according to the value of the control parameter τ. In figure 9(c), the canonical correlation coefficient $\rho_{\mathcal{F}}^{\max}$ is plotted as a function of τ, which visualizes the change of the input–output relation between two systems of (18) and (19). The

index $\rho_{\mathcal{F}}^{\max}$ changes nonmonotonically with the increase of τ, and there is a regime around $\tau \sim 4$ where the value of $\rho_{\mathcal{F}}^{\max}$ is large. In addition to GS, chaotic phase synchronization (CPS) [35] occurs between two systems in this regime [34]. The increase of $\rho_{\mathcal{F}}^{\max}$ in this regime can be attributed to the occurrence of CPS.

4.3. Bidirectionally coupled systems

The notion of GS is not restricted to unidirectionally coupled systems. In [36, 37], the occurrence of GS for bidirectionally coupled systems is also discussed. As an example of GS in bidirectionally coupled systems, we consider the following coupled Lorenz–Rössler systems [36]:

$$X : \begin{cases} \dot{x}_1 = 10(x_2 - x_1) + \gamma(y_1 - x_1) \\ \dot{x}_2 = 35x_1 - x_2 - x_1x_3 \\ \dot{x}_3 = -(8/3)x_3 + x_1x_2, \end{cases} \tag{21}$$

$$Y : \begin{cases} \dot{y}_1 = 5.5y_2 - y_3 + \gamma(x_1 - y_1) \\ \dot{y}_2 = 5.5y_1 + 0.165y_2 \\ \dot{y}_3 = 0.2 + y_3(y_1 - 10). \end{cases} \tag{22}$$

A mutual interaction between a Lorenz system (21) and a Rössler system (22) is introduced as diffusion terms in the first equations of both (21) and (22), and γ is its strength. In [36], Zheng *et al* try to define GS in bidirectionally coupled systems by considering identical copies X' and Y' of the systems X and Y. Using the approach in [36], two transitions of GS are found in the systems (21) and (22). At $\gamma \sim 1.2$, the Rössler system Y is entrained by the Lorenz system X in the sense that orbits of the systems Y and Y' completely coincide with each other by receiving the common signal of the system X. With the further increase of γ, the system X is also entrained by the system Y at $\gamma \sim 12.3$ which means that the complete synchronization between X and X' also occurs.

We apply the kernel CCA to the systems (21) and (22) and results are shown in figures 10 and 11. Figure 10(*a*1), (*b*1) and (*c*1) show the projections of attractors of the systems (21) and (22) with $\gamma = 0.4, 4.0$ and 13.0 onto the (x_1, y_1) planes, respectively. Figures 10(*a*2), (*b*2) and (*c*2) also show the corresponding results of the kernel CCA. Here, we use a Gaussian kernel with $\sigma = 1.0$. The state variables $\{(x_1(t_i), x_2(t_i), x_3(t_i)), (y_1(t_i), y_2(t_i), y_3(t_i))\}$ where $t_i = i\Delta t, i = 1, \ldots, N, \Delta t = 5.0, N = 2 \times 10^3$ are used as the training data set. It is observed that the state of GS is clearly identified as the high linear correlation between canonical variates u and v of the kernel CCA. Application of the kernel CCA to the bidirectionally coupled systems is straightforward while the approach of [36] requires rather subtle procedures. Figure 11 shows the index $\rho_{\mathcal{F}}^{\max}$ as a function of the coupling strength γ. There is a rapid increase of $\rho_{\mathcal{F}}^{\max}$ against γ in an interval $0 < \gamma < 1.2$. This result is consistent with the first transition of GS defined in [36]. The second transition of GS at $\gamma \sim 12.3$ defined in [36] is not seen in the graph of $\rho_{\mathcal{F}}^{\max}$ versus γ. A reason is that the value of $\rho_{\mathcal{F}}^{\max}$ already becomes nearly 1 at $\gamma \sim 5$. Another reason is that by definition, the proposed approach based on the kernel CCA is insensitive to the directionality of synchronization. It will be interesting to study modifications of the approach which can deal with the directionality of synchronization.

5. Nonstationary change of coupling strength

So far, we have focused on the dependence of the first eigenvalue $\rho_{\mathcal{F}}^{\max}$ on the coupling strength γ. We turn our attention to the eigenvector ${}^t(\alpha, \beta)$ in (12) and investigate changes of the structure of the dynamical system with the kernel CCA.
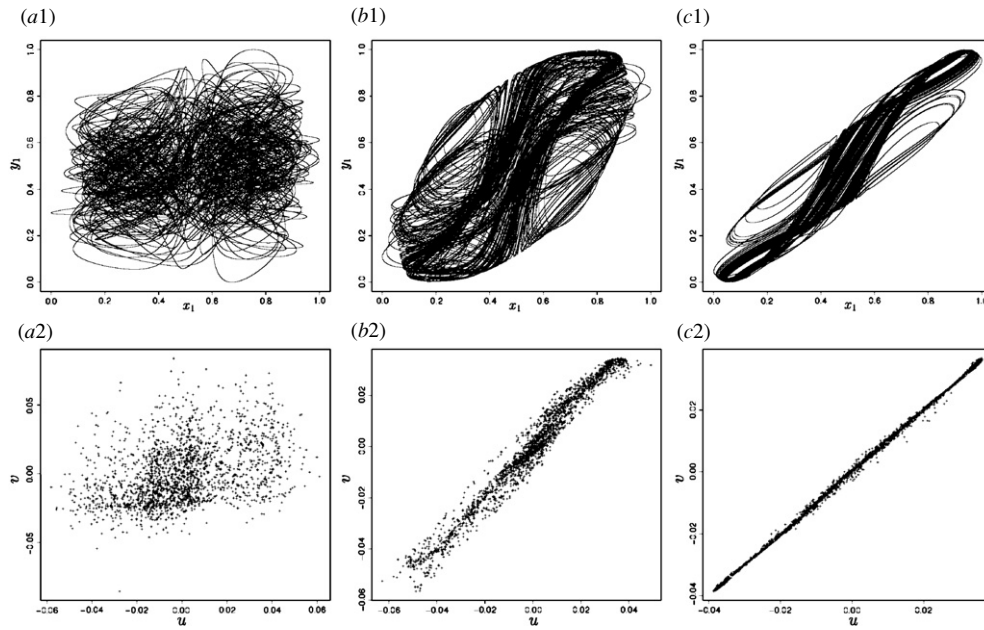
Figure 10. (a1), (b1), (c1) Projections of attractors of the bidirectionally coupled Lorenz and Rössler systems onto the plane of (x_1, y_1) with $\gamma = 0.4$ in (a1), $\gamma = 4.0$ in (b1) $\gamma = 13.0$ in (c1). (a2), (b2), (c2) Scatter plots of the canonical covariates of the kernel CCA with $\gamma = 0.4$ in (a1), $\gamma = 4.0$ in (b1) $\gamma = 13.0$ in (c1). Parameters are $\sigma = 1.0$, $\kappa = 0.1$, and $N = 2 \times 10^3$.
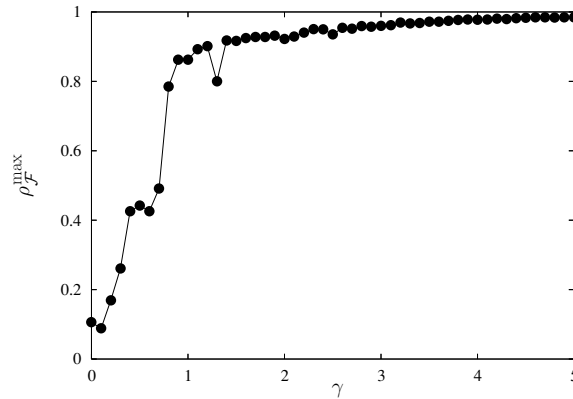

Figure 11. The index $\rho_{\mathcal{F}}^{\max}$ of the kernel CCA as a function of the coupling strength γ with $\sigma = 1.0$, $\kappa = 0.1$, and $N = 2 \times 10^3$.

As an illustration, let us consider the problem of extracting nonstationary changes of the coupling strength γ from time series data generated by the coupled Hénon maps. An example of such changes is shown in figure 12(b). When the value of γ varies from γ_0, an orbit leaves from the synchronization manifold $\mathcal{M}$ with γ_0 and is attracted again to it when the value of γ is restored to γ_0. As shown in figure 12(a), the time series of the original variables (x_1, y_1) does not tell us whether the orbit is lying on $\mathcal{M}$ or not at a given time t. By using the kernel CCA, a deviation of the orbit from $\mathcal{M}$ will be detected as a large value of $|f(x) - g(y)|$, where $f(\cdot)$ and $g(\cdot)$ are nonlinear transformations determined by the first eigenvector $^t(\alpha, \beta)$.
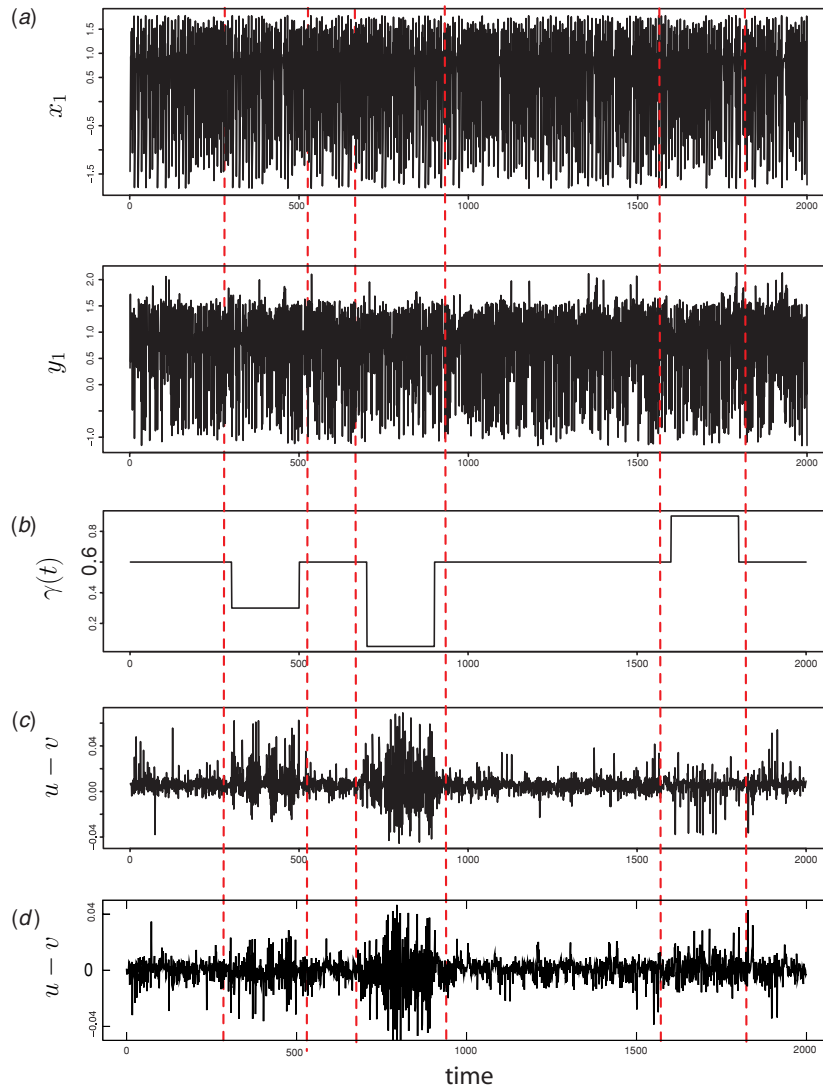
Figure 12. (*a*) Time series of the variables x_1 and y_1 of the coupled Hénon maps when the value of γ changes temporally as shown in (*b*). (*c*), (*d*) Time series of the difference of the canonical variates of the kernel CCA with $\sigma = 0.1$ and $\kappa = 0.1$.

Numerical experiments are performed with $\gamma_0 = 0.6$ in two different conditions, and results are shown in figures 12(*c*) and (*d*), respectively. Figure 12(*c*) shows the case where $f(\cdot)$ and $g(\cdot)$ are estimated from an orbit with $\gamma = \gamma_0$, which is prepared separately from the orbit to be analysed. In figure 12(*c*), time intervals where the values of γ are different from γ_0 are clearly detected as successive bursts in the time series of $|f(x) - g(y)|$.

In figure 12(*d*), we show the case where the orbit to be analysed is also used for estimating $f(\cdot)$ and $g(\cdot)$. We see that even when the estimation of nonlinear transformations is affected by 'noise' from the points not lying on $\mathcal{M}$, successive bursts are still observed in the time series of $|f(x) - g(y)|$ except for those in a time interval $1600 < t < 1800$.
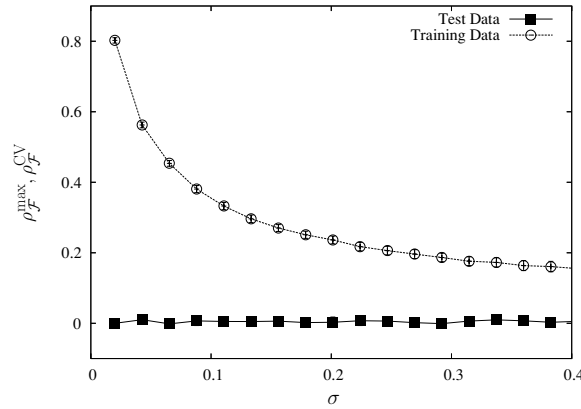
Figure 13. Results of cross-validation for two uncoupled Hénon maps ($\gamma = 0.0$). The dotted line with open circles shows $\rho_{\mathcal{F}}^{\max}$ versus σ and the solid line with filled squares shows $\rho_{\mathcal{F}}^{\mathrm{CV}}$ versus σ. We prepare 20 data sets with size $N = 10^3$ for training the kernel CCA, and a test data with size $\tilde{N} = 2 \times 10^3$ for cross-validation. The average over 20 realizations are plotted as symbols. The standard deviation is also plotted as vertical bars, but the ones with small values of errors are hidden by symbols. κ is set to 0.01.

6. Choice of kernel parameters

In the preceding section, we set the value of the kernel parameter such as the width σ of a Gaussian kernel in an ad hoc manner. If σ is too small, the nonlinear transformations $f(\cdot)$ and $g(\cdot)$ in (2) cannot interpolate between data points of the training sample. In contrast, if σ is too large, (2) cannot represent a highly nonlinear structure such as the synchronization manifold $\mathcal{M}$ of GS. Thus a proper choice of the kernel parameter is crucial for obtaining the good performance of the method. In this section, we discuss how the kernel parameter can suitably be chosen from a given data set.

A naive way of choosing σ is to find the value of σ that maximizes $\rho_{\mathcal{F}}^{\max}$. The dotted line with open circles in figure 13 shows the graph of an index $\rho_{\mathcal{F}}^{\max}$ as a function of σ when two Hénon maps (14) and (15) are uncoupled ($\gamma = 0$). Here, an orbit of (14) and (15) with length $N = 10^3$ is used as a set of the training data, and the average and the standard deviation over 20 realizations are plotted as symbols and vertical bars, respectively. The index $\rho_{\mathcal{F}}^{\max}$ increases monotonically with the decrease of σ, and $\rho_{\mathcal{F}}^{\max} \sim 1$ is attained in the limit of $\sigma \to 0$, while there is no interaction between the two systems X and Y. This monotonic tendency of $\rho_{\mathcal{F}}^{\max}$ does not determine an optimal value of the kernel parameter σ.

As a way to overcome this difficulty, the following procedure is proposed. First, we set aside the data $\{(\tilde{x}_n, \tilde{y}_n)\}_{n=1}^{\tilde{N}}$ for assessing the performance of the kernel CCA (we call this 'test' data) separately from the data used for training the kernel CCA. Then, we estimate the nonlinear transformations $f(\cdot)$ and $g(\cdot)$ from the training data and calculate the correlation coefficient $\rho_{\mathcal{F}}^{\mathrm{CV}}$ between the variables $\tilde{u}_n = f(\tilde{x}_n)$ and $\tilde{v}_n = g(\tilde{y}_n)$, $n = 1, 2, \dots, \tilde{N}$ defined as

$$\rho_{\mathcal{F}}^{\mathrm{CV}} = \frac{\langle (\tilde{u}_n - \langle \tilde{u}_n \rangle) \cdot (\tilde{v}_n - \langle \tilde{v}_n \rangle) \rangle}{\sqrt{\langle (\tilde{u}_n - \langle \tilde{u}_n \rangle)^2 \rangle} \cdot \sqrt{\langle (\tilde{v}_n - \langle \tilde{v}_n \rangle)^2 \rangle}}, \tag{23}$$

where $\langle \cdots \rangle$ is the empirical average over $\tilde{N}$ samples. This strategy for assessing the performance of the estimated model with new data is regarded as a version of *cross-validation* (CV) [38–40].
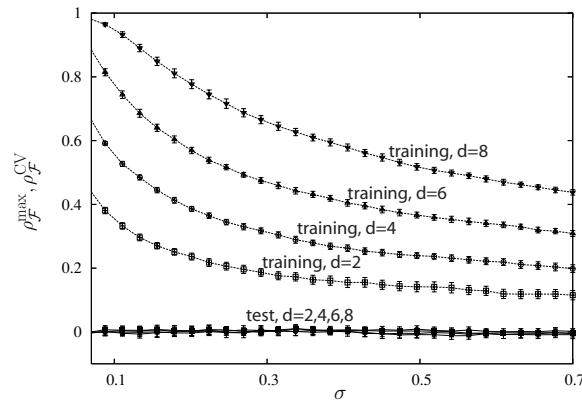
Figure 14. Results of cross-validation with $\gamma = 0$ for four different values of the embedding dimension d. We take the delay time $l = 1$. The condition for numerical experiments is the same as figure 13.
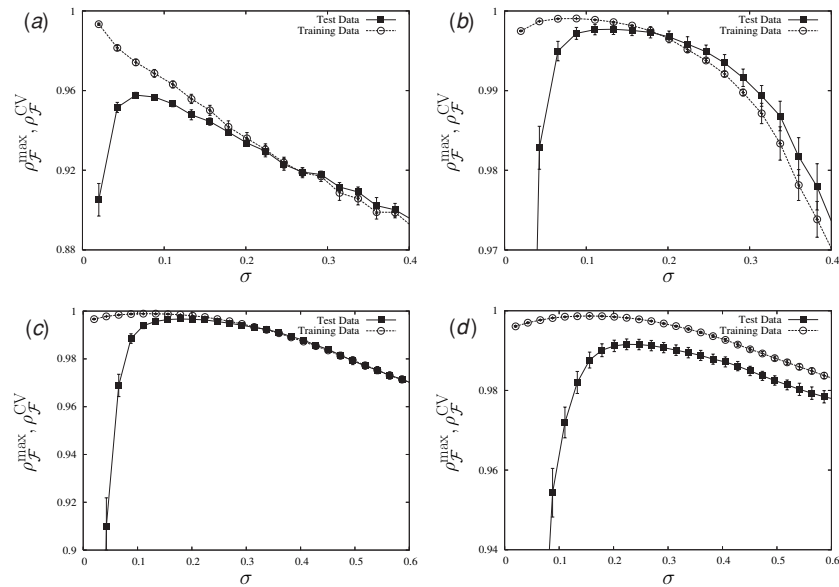

Figure 15. Results of cross-validation for two coupled Hénon maps with $\gamma = 0.25$. The embedding dimension $d = 2$ in (*a*), $d = 4$ in (*b*), $d = 6$ in (*c*) and $d = 8$ in (*d*), and the delay time $l = 1$. The condition for numerical experiments is the same as figure 13.

First we check that spurious detection of synchronization can be avoided by the procedure based on CV. The solid line with filled squares in figure 13 shows the dependence of an index $\rho_{\mathcal{F}}^{\mathrm{CV}}$ on σ with $\gamma = 0$. When there is no interaction between two systems, the value of $\rho_{\mathcal{F}}^{\mathrm{CV}}$ is nearly equal to 0 for any σ. Corresponding results using the time-delay embedding are shown in figure 14. In figure 14, the dependencies of $\rho_{\mathcal{F}}^{\max}$ and $\rho_{\mathcal{F}}^{\mathrm{CV}}$ on σ with $\gamma = 0$ are plotted for several different values of the embedding dimension d. Again, the value of $\rho_{\mathcal{F}}^{\mathrm{CV}}$ is almost equal to 0 for any σ and d, whereas the value of $\rho_{\mathcal{F}}^{\max}$ as a function of σ increases monotonically

with the increase of d. This result suggests that the procedure based on CV effectively erases spurious detection even when the data are embedded in a high-dimensional state space

Next we show that the cross validation procedure is useful for choosing optimal values of σ. Figure 15 shows the values of $\rho_{\mathcal{F}}^{\max}$ and $\rho_{\mathcal{F}}^{CV}$ as functions of σ for the coupled Hénon maps (14) and (15) with $\gamma = 0.25$. The condition for the numerical experiment is the same as $\gamma = 0$. For all of graphs, the value of $\rho_{\mathcal{F}}^{CV}$ with the dotted line takes its maximum at a nonzero value of σ, whereas $\rho_{\mathcal{F}}^{\max}$ increases monotonically with the decrease of σ. This result indicates that we can choose this value as an optimal width of a Gaussian kernel.

7. Conclusion

In conclusion, we have proposed a new approach for analysing GS in a unified framework of a kernel method. We have tested the proposed approach by applying it to several examples exhibiting GS, and demonstrated that the canonical correlation coefficient of the kernel CCA is a suitable index for the characterization of GS. In addition, it has been shown that nonstationary changes of the coupling are detected from the time series by the difference between canonical variates of the kernel CCA. It has been also discussed how the parameter of the kernel function can suitably be chosen from data by the procedure of cross-validation. Our experiments show that a method based on CV gives promising results in optimizing the parameter σ. The cross-validation procedure is also useful to circumvent spurious detection of GS by overfitting.

The approach based on the kernel CCA provides not only an index for measuring the nonlinear interdependence. It also provides global nonlinear coordinates, and these coordinates allow a representation of the interaction between the dynamical systems under investigation. Note that the linear CCA also provides global coordinates, but it cannot discriminate between linear and nonlinear relations. In this respect, it goes beyond the conventional methods of analysing GS. Our attempts open a new possibility of the kernel methods for analysing complex dynamics observed in nonlinear systems.

Acknowledgments

References

[1] Pikovsky A, Rosenblum M and Kurths J 2002 *Synchronization, A Universal Concept in Nonlinear Sciences* (Cambridge: Cambridge University Press)
[2] Fujisaka H and Yamada T 1983 Stability theory of synchronized motion in coupled-oscillator system *Prog. Theor. Phys.* **69** 32–47
[3] Pikovsky A S 1984 On the interaction of strange attractors *Z. Phys.* B **55** 149–54
[4] Pecora L M and Carroll T L 1990 Synchronization in chaotic systems *Phys. Rev. Lett.* **64** 821–4
[5] Rulkov N F, Sushchik M M, Tsimring L S and Abarbanel H D I 1995 Generalized synchronization of chaos in directionally coupled chaotic systems *Phys. Rev.* E **51** 980–94
[6] Farmer S F 1998 Rhythmicity, synchronization and binding in human and primate motor systems *J. Physiol.* **509** 3–14
[7] Le Van Quyen M, Martinerie J, Adam C and Varela F J 1999 Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy *Physica* D **127** 250–66
[8] Arnhold J, Grassberger P, Lehnertz K and Elger C E 1999 A robust method for detecting interdependences: application to intracranially recorded EEG *Physica* D **134** 419–30

[9] Kocarev L and Parlitz U 1996 Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems *Phys. Rev. Lett.* **76** 1816–9

[10] Abarbanel H D I, Rulkov N F and Sushchik M M 1996 Generalized synchronization of chaos: the auxiliary system approach *Phys. Rev.* E **53** 4528–35

[11] Schiff S J, So P, Chang T, Burke R E and Sauer T 1996 Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble *Phys. Rev.* E **54** 6708–24

[12] Quian Quiroga R, Arnhold J and Grassberger P 2000 Learning driver-response relationships from synchronization patterns *Phys. Rev.* E **61** 5142–8

[13] Stam C J and Van Dijk B W 2002 Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets *Physica* D **163** 236–51

[14] Schölkopf B, Smola A J and Müller K-R 1998 Nonlinear component analysis as a kernel eigenvalue problem *Neural Comput.* **10** 1299–319

[15] Müller K-R, Mika S, Rätsch G, Tsuda K and Schölkopf B 2001 An introduction to kernel-based learning algorithms *IEEE Trans. Neural Netw.* **12** 181–201

[16] Schölkopf B and Smola A J 2002 *Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond* (Cambridge, MA: MIT Press)

[17] Shawe-Taylor J and Cristianini N 2004 *Kernel Methods for Pattern Recognition* (Cambridge: Cambridge University Press)

[18] Lai P L and Fyfe C 2000 Kernel and nonlinear canonical correlation analysis *Int. J. Neural Syst.* **10** 365–77

[19] Akaho S 2001 A kernel method for canonical correlation analysis *Proc. International Meeting of the Psychometric Society (IMPS2001)* pp 123–8

[20] Melzer T, Reiter M and Bischof H 2001 Nonlinear feature extraction using generalized canonical correlation analysis *Proc. International Conference on Artficial Neural Networks (ICANN)* (London: Springer) pp 353–60

[21] Bach F R and Jordan M I 2002 Kernel independent component analysis *J. Mach. Learn. Res.* **3** 1–48

[22] Suetani H, Iba Y and Aihara K 2006 *Prog. Theor. Phys. Suppl.* **161** 340–3

[23] Hotelling H 1936 Relation between two sets of variates *Biometrika* **28** 321–77

[24] Buja A 1990 Remarks on functional canonical variates, alternating least squares methods and ACE *Ann. Stat.* **18** 1032–69

[25] Breiman L and Friedman J H 1985 Estimating optimal transformations for multiple regression and correlation *J. Am. Stat. Assoc.* **80** 580–98

[26] Voss H and Kurths J 1997 Reconstruction of non-linear time delay models from data by the use of optimal transformations *Phys. Lett.* A **234** 336–44

[27] Schölkopf B, Tsuda K and Vert J-P (ed) 2004 *Kernel Methods in Computational Biology* (Cambridge, MA: MIT Press)

[28] Schreiber T and Schmitz A 2000 Surrogate time series *Physica* D **142** 346–82

[29] Prichard D and Theiler J 1994 Generating surrogate data for time series with several simultaneously measured variables *Phys. Rev. Lett.* **73** 951–4

[30] Schreiber T and Schmitz A 1996 Improved surrogate data for nonlinearity tests *Phys. Rev. Lett.* **77** 635–8

[31] Hegger R, Kantz H and Schreiber T 1999 Practical implementation of nonlinear time series methods: the TISEAN package *Chaos* **9** 413–35

[32] Sauer T 1994 Reconstruction of dynamical systems from interspike intervals *Phys. Rev. Lett.* **72** 3811–4

[33] Racicot D M and Longtin A 1997 Interspike interval attractors from chaotically driven neuron models *Physica* D **104** 184–204

[34] Han S K, Kim W S and Kook H 2002 Synchronization and decoding interspike intervals *Int. J. Bifurcation Chaos* **12** 983–99

[35] Rosenblum M G, Pikovsky A S and Kurths J 1996 Phase synchronization of chaotic oscillators *Phys. Rev. Lett.* **76** 1804–7

[36] Zheng Z, Wang X and Cross M C 2002 Transitions from partial to complete generalized synchronizations in bidirectionally coupled chaotic oscillators *Phys. Rev.* E **65** 056211

[37] Osipov G V, Hu B, Zhou C, Ivanchenko M V and Kurths J 2003 Three types of transitions to phase synchronization in coupled chaotic oscillators *Phys. Rev. Lett.* **91** 024101

[38] Stone M 1974 Cross-validatory choice and assessment of statistical prediction *J. R. Stat. Soc.* B **36** 111–47

[39] Silverman B W 1985 Some aspects of the spline smoothing approach to non-parametric regression curve fitting *J. R. Stat. Soc.* B **47** 1–52

[40] Wahba G 1990 *Spline Models for Observational Data* (Philadelphia: SIAM)

[41] Fukumizu K, Bach F R and Gretton A 2005 Consistency of kernel canonical correlation analysis *ISM Research Memorandum* no 942