

Bayesian Random Fourier Filters for Gaussian Noises

Shiyuan WANG^{1,2*}, Wanli WANG^{1,2,3}, Shukai DUAN^{1,2}, Lidan WANG^{1,2} & Chi K. TSE³

¹College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China;

²Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Chongqing 400715, China;

³Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China

Appendix A

Lemma 1. On the condition of appropriate dimensions for random variables, given a marginal Gaussian distribution for \mathbf{x} and a conditional Gaussian distribution for \mathbf{y} given \mathbf{x} in the following form

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{Q}_1), \quad (\text{A1})$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{H}\mathbf{x}, \mathbf{Q}_2), \quad (\text{A2})$$

the joint distribution (\mathbf{x}, \mathbf{y}) and the marginal distribution of \mathbf{y} are given by [1, 2]

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{H}\mathbf{m} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_1\mathbf{H}^T \\ \mathbf{H}\mathbf{Q}_1 & \mathbf{H}\mathbf{Q}_1\mathbf{H}^T + \mathbf{Q}_2 \end{bmatrix}\right), \quad (\text{A3})$$

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}; \mathbf{H}\mathbf{m}, \mathbf{H}\mathbf{Q}_1\mathbf{H}^T + \mathbf{Q}_2). \quad (\text{A4})$$

If $\mathbf{m} = \mathbf{0}$, the conditional distribution \mathbf{x} give \mathbf{y} are obtained by

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}; \mathbf{P}\mathbf{H}^T\mathbf{Q}_2^{-1}\mathbf{y}, \mathbf{P}), \quad (\text{A5})$$

where $\mathbf{P} = (\mathbf{Q}_1^{-1} + \mathbf{H}^T\mathbf{Q}_2^{-1}\mathbf{H})^{-1}$.

Appendix B

This Appendix shows the stability analysis of online BRFF.

Substituting (8) into (13) generates the following update form of $\hat{\boldsymbol{\theta}}_k$.

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k|k-1} + \mathbf{G}_k e_k. \quad (\text{B1})$$

Subtracting both sides of (B1) from $\boldsymbol{\theta}_k$ yields the relation between the estimated and predicted errors as

$$\tilde{\boldsymbol{\theta}}_k = \tilde{\boldsymbol{\theta}}_{k|k-1} - \mathbf{G}_k e_k. \quad (\text{B2})$$

The covariance matrix \mathbf{P}_k of the estimated error at time k given in (12) can be rewritten as

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{G}_k \mathbf{z}(\mathbf{x}_k)^T) \mathbf{P}_{k|k-1}. \quad (\text{B3})$$

According to the matrix inversion lemma in [3], i.e., $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$ with $\mathbf{A} = \mathbf{P}_{k|k-1}$, $\mathbf{B} = -\mathbf{P}_{k|k-1}\mathbf{z}(\mathbf{x}_k)$, $\mathbf{C} = (\mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_{k|k-1} \mathbf{z}(\mathbf{x}_k) + \delta_n^2)^{-1}$, $\mathbf{D} = \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_{k|k-1}$, after some manipulations we obtain

$$\mathbf{P}_k = (\mathbf{I} - \delta_n^{-2} \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \mathbf{z}(\mathbf{x}_k)^T) \mathbf{P}_{k|k-1}. \quad (\text{B4})$$

Combining (14), (B3), and (B4) into (B2), we have

$$\tilde{\boldsymbol{\theta}}_k = \tilde{\boldsymbol{\theta}}_{k|k-1} - \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \delta_n^{-2} e_k. \quad (\text{B5})$$

* Corresponding author (email: wsy@swu.edu.cn)

To perform stability analysis of online BRFF, we construct the Lyapunov function by $V_k = \tilde{\boldsymbol{\theta}}_k^T \mathbf{P}_k^{-1} \tilde{\boldsymbol{\theta}}_k$, which is a quadratic function associated with the estimated error $\tilde{\boldsymbol{\theta}}_k$ in the RFFS. Similarly, a quadratic function with the predicted error $\tilde{\boldsymbol{\theta}}_{k|k-1}$ is given by $V_k^- = \tilde{\boldsymbol{\theta}}_{k|k-1}^T \mathbf{P}_{k|k-1}^{-1} \tilde{\boldsymbol{\theta}}_{k|k-1}$.

According to (18), we have

$$\begin{aligned} V_k^- &= \tilde{\boldsymbol{\theta}}_{k|k-1}^T \mathbf{P}_{k|k-1}^{-1} \tilde{\boldsymbol{\theta}}_{k|k-1} \\ &= \tilde{\boldsymbol{\theta}}_{k-1}^T \boldsymbol{\beta}_{k-1}^T \mathbf{P}_{k|k-1}^{-1} \boldsymbol{\beta}_{k-1} \tilde{\boldsymbol{\theta}}_{k-1}. \end{aligned} \quad (\text{B6})$$

Based on (7), (20), and $\delta_D^2 \geq 0$, the following inequality associated with $\mathbf{P}_{k|k-1}$ and \mathbf{P}_{k-1} is established.

$$\mathbf{P}_{k|k-1}^{-1} = (\mathbf{P}_{k-1} + \delta_D^2 \mathbf{I})^{-1} \leq \mathbf{P}_{k-1}^{-1}. \quad (\text{B7})$$

Equation (B7) holds from the fact that $\mathbf{M}^{-1} > (\mathbf{M} + \mathbf{H})^{-1}$ [4] on the condition of any $n \times n$ matrices $\mathbf{M}, \mathbf{H} > \mathbf{0}_{n \times n}$. The following inequality is therefore derived.

$$V_k^- - V_{k-1} \leq \tilde{\boldsymbol{\theta}}_{k-1}^T (\boldsymbol{\beta}_{k-1}^T \mathbf{P}_{k-1}^{-1} \boldsymbol{\beta}_{k-1} - \mathbf{P}_{k-1}^{-1}) \tilde{\boldsymbol{\theta}}_{k-1}. \quad (\text{B8})$$

Substituting (21) into (B8) yields $V_k^- - V_{k-1} \leq 0$.

Expanding V_k with (B5) and combining (B4), we have

$$\begin{aligned} V_k &= \tilde{\boldsymbol{\theta}}_k^T \mathbf{P}_k^{-1} \tilde{\boldsymbol{\theta}}_k \\ &= (\tilde{\boldsymbol{\theta}}_{k|k-1} - \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \delta_n^{-2} e_k)^T \mathbf{P}_k^{-1} (\tilde{\boldsymbol{\theta}}_{k|k-1} - \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \delta_n^{-2} e_k) \\ &= \tilde{\boldsymbol{\theta}}_{k|k-1}^T \mathbf{P}_k^{-1} \tilde{\boldsymbol{\theta}}_{k|k-1} - \tilde{\boldsymbol{\theta}}_{k|k-1}^T \mathbf{z}(\mathbf{x}_k) \delta_n^{-2} e_k - \mathbf{z}(\mathbf{x}_k)^T \tilde{\boldsymbol{\theta}}_{k|k-1} \delta_n^{-2} e_k \\ &\quad + \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \delta_n^{-4} e_k^2 \\ &= V_k^- + \tilde{\boldsymbol{\theta}}_{k|k-1}^T \delta_n^{-2} \mathbf{z}(\mathbf{x}_k) \mathbf{z}(\mathbf{x}_k)^T \tilde{\boldsymbol{\theta}}_{k|k-1} - \tilde{\boldsymbol{\theta}}_{k|k-1}^T \mathbf{z}(\mathbf{x}_k) \delta_n^{-2} e_k \\ &\quad - \mathbf{z}(\mathbf{x}_k)^T \tilde{\boldsymbol{\theta}}_{k|k-1} \delta_n^{-2} e_k + \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \delta_n^{-4} e_k^2. \end{aligned} \quad (\text{B9})$$

Substituting (19) into (B9) generates

$$V_k - V_k^- = e_k (\alpha_k \delta_n^{-2} \alpha_k - \alpha_k \delta_n^{-2} - \delta_n^{-2} \alpha_k + \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \delta_n^{-4}) e_k. \quad (\text{B10})$$

According to (22), the following inequality can be derived.

$$e_k (\alpha_k \delta_n^{-2} \alpha_k - \alpha_k \delta_n^{-2} - \delta_n^{-2} \alpha_k + \delta_n^{-2} \gamma_k) e_k \leq 0. \quad (\text{B11})$$

Pre- and post-multiplying the left side of (B4) by $\mathbf{z}(\mathbf{x}_k)^T$ and $\mathbf{z}(\mathbf{x}_k)$ generates the following equality.

$$\mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) = \mathbf{z}(\mathbf{x}_k)^T (\mathbf{I} - \delta_n^{-2} \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \mathbf{z}(\mathbf{x}_k)^T) \mathbf{P}_{k|k-1} \mathbf{z}(\mathbf{x}_k). \quad (\text{B12})$$

We re-arrange (B12) as

$$\mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) + \mathbf{z}(\mathbf{x}_k)^T \delta_n^{-2} \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_{k|k-1} \mathbf{z}(\mathbf{x}_k) = \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_{k|k-1} \mathbf{z}(\mathbf{x}_k). \quad (\text{B13})$$

Further, (B13) is simplified as

$$\delta_n^{-2} \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k) = \frac{\mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_{k|k-1} \mathbf{z}(\mathbf{x}_k)}{\mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_{k|k-1} \mathbf{z}(\mathbf{x}_k) + \delta_n^2}. \quad (\text{B14})$$

Based on (B14), (B11) can be rewritten as

$$e_k (\alpha_k \delta_n^{-2} \alpha_k - \alpha_k \delta_n^{-2} - \delta_n^{-2} \alpha_k + \delta_n^{-4} \mathbf{z}(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{z}(\mathbf{x}_k)) e_k \leq 0. \quad (\text{B15})$$

Substituting (B15) into (B10) gives $V_k - V_k^- \leq 0$. Since $V_k^- - V_{k-1} \leq 0$, we have $V_k \leq V_{k-1}$.

Remark 1. The trackability of online BRFF is improved by introducing the diffusion noise in (4). The inequalities in (20), (21), and (22) are the sufficient but not necessary conditions for the stability of online BRFF. Since the inequality (21) implies the boundary of variance δ_D^2 of the diffusion noise, which is shown in Appendix C, a bounded δ_D^2 is required to guarantee the stability of the online BRFF. Since \mathbf{P}_k is a positive definite matrix, the positive Lyapunov function $\tilde{\boldsymbol{\theta}}_k^T \mathbf{P}_k^{-1} \tilde{\boldsymbol{\theta}}_k$ can also guarantee $1 - \gamma_k > 0$ in (22). It can be seen from Theorem 1 that V_k is a decreasing sequence, and thus the stability of online BRFF is guaranteed.

Appendix C

This appendix shows the boundedness feature of variance δ_D^2 of diffusion noise \mathbf{q}_k on the condition of (21).

Combing (18) and (21), we have inequality

$$\mathbf{q}_{k-1,j} \tilde{\boldsymbol{\theta}}_{k-1,j} \leq 0, j = 1, \dots, D, \quad (\text{C1})$$

where $\mathbf{q}_{k-1,j}$ and $\tilde{\boldsymbol{\theta}}_{k-1,j}$ denote the j -th entries of \mathbf{q}_{k-1} and $\tilde{\boldsymbol{\theta}}_{k-1}$ respectively.

If $\tilde{\boldsymbol{\theta}}_{k-1,j} < 0$, we obtain the following inequalities using (18), (21) and (C1).

$$\begin{cases} 0 \leq \mathbf{q}_{k-1,j} \leq -\tilde{\boldsymbol{\theta}}_{k-1,j}, & \text{for } 0 \leq \boldsymbol{\beta}_{k-1,j} \leq 1, \\ -\tilde{\boldsymbol{\theta}}_{k-1,j} \leq \mathbf{q}_{k-1,j} \leq -2\tilde{\boldsymbol{\theta}}_{k-1,j}, & \text{otherwise} \end{cases}. \quad (\text{C2})$$

Similarly, $\tilde{\boldsymbol{\theta}}_{k-1,j} \geq 0$ generates

$$\begin{cases} -\tilde{\boldsymbol{\theta}}_{k-1,j} \leq \mathbf{q}_{k-1,j} \leq 0, & \text{for } 0 \leq \beta_{k-1,j} \leq 1, \\ -2\tilde{\boldsymbol{\theta}}_{k-1,j} \leq \mathbf{q}_{k-1,j} \leq -\tilde{\boldsymbol{\theta}}_{k-1,j}, & \text{otherwise} \end{cases}. \quad (\text{C3})$$

When (21) is satisfied, combing (C2) and (C3) gives

$$|\mathbf{q}_{k-1}| \leq 2|\tilde{\boldsymbol{\theta}}_{k-1}|. \quad (\text{C4})$$

Therefore, according to (C4), we have

$$\begin{aligned} E(|\mathbf{q}_{k-1}||\mathbf{q}_{k-1}|^T) &\leq 2E(|\tilde{\boldsymbol{\theta}}_{k-1}||\mathbf{q}_{k-1}|^T), \\ E(|\tilde{\boldsymbol{\theta}}_{k-1}||\mathbf{q}_{k-1}|^T) &\leq 2E(|\tilde{\boldsymbol{\theta}}_{k-1}||\tilde{\boldsymbol{\theta}}_{k-1}|^T), \end{aligned} \quad (\text{C5})$$

which generates

$$E(|\mathbf{q}_{k-1}||\mathbf{q}_{k-1}|^T) \leq 4E(|\tilde{\boldsymbol{\theta}}_{k-1}||\tilde{\boldsymbol{\theta}}_{k-1}|^T). \quad (\text{C6})$$

Since \mathbf{q}_{k-1} follows the Gaussian distribution $\mathcal{N}(\mathbf{q}; \mathbf{0}, \delta_D^2 \mathbf{I})$ and the inequality $E(\mathbf{q}_{k-1} \mathbf{q}_{k-1}^T) \leq E(|\mathbf{q}_{k-1}||\mathbf{q}_{k-1}|^T)$ holds, we rewrite (C6) as

$$\delta_D^2 \mathbf{I} \leq 4E(|\tilde{\boldsymbol{\theta}}_{k-1}||\tilde{\boldsymbol{\theta}}_{k-1}|^T). \quad (\text{C7})$$

Therefore, there exists the boundary of δ_D^2 when (21) is satisfied.

Appendix D Simulation results

This Appendix shows three examples of nonlinear signal processing, namely the system identification, Mackey-Glass (MG) time series prediction, and nonlinear regression chosen for validating the proposed BRFF. In the system identification, the stability of online BRFF in terms of the Lyapunov function is verified. The stationary and nonstationary cases are discussed to comprehensively evaluate the performance of BRFF. For comparison, the chosen online learning algorithms [5] include KLMS based on Mercer Kernel [6–9] and OS-ELM [10, 11]. For each example, 50 Monte Carlo simulations are performed, and the results are averaged over these simulations.

Appendix D.1 System identification

Consider the identified system constructed using the nonlinear mapping (2) [12–14] and the linear model (5). In (2), the nonlinear mapping constructed by the random weight $\boldsymbol{\omega}$ and b first projects the d -dimensional input space onto the D -dimensional feature space. Then the linear form in (5) modeled by the coefficient $\boldsymbol{\theta}_k$ and the feature input can be implemented to learn the relationship built in the original input-output pairs.

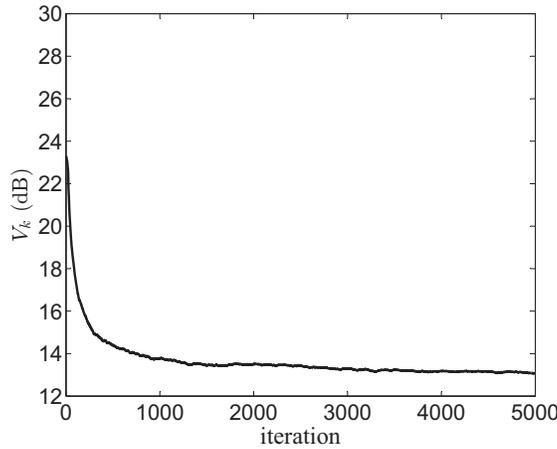


Figure D1 Lyapunov function in system identification.

In this simulation, we define the input matrix $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_M] \in \mathbb{R}^{d \times M}$ sampled from the distribution $\mathcal{N}(\mathbf{X}_n; \mathbf{0}, 2\mathbf{I}_d)$, where $d = 5$ and $M = 5100$ is the number of samples with 5000 for training and another 100 for testing. The desired outputs are generated by (5) associated with $\boldsymbol{\omega}$ and b drawn from Gaussian distribution $\mathcal{N}(\boldsymbol{\omega}; \mathbf{0}, 5\mathbf{I}_d)$ and uniform distribution in $[0, 2\pi]$, respectively. Moreover, the desired outputs are corrupted by Gaussian noise with zero mean and variance $\delta_n^2 = 0.01$. The variance of the diffusion noise in each direction δ_D^2 is 10^{-7} . Note that during the iteration, the coefficient $\boldsymbol{\theta}_k$ in (5) is unique and equal to the initialization $\boldsymbol{\theta}_1$ drawn from $\mathcal{N}(\boldsymbol{\theta}_1; \mathbf{0}, 2\mathbf{I}_D)$ with $D = 20$. For performance comparison, the mean square error (MSE) is defined as

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{f}(\mathbf{x}_k))^2, \quad (\text{D1})$$

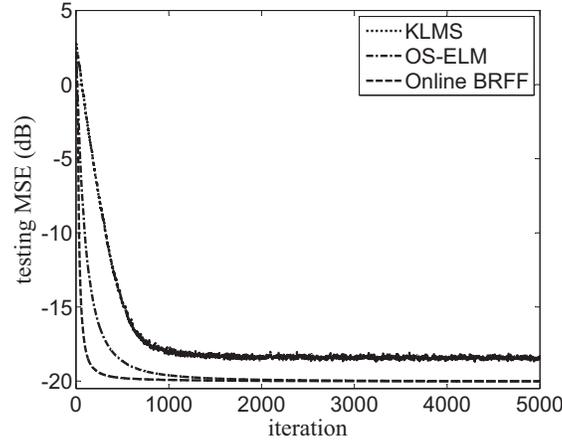


Figure D2 Learning curves of KLMS, OS-ELM and online BRFF in system identification.

where N is the total number of calculations. In the following, the learning curves of filters are plotted in terms of the testing MSE based on the test data.

For KLMS, the step size being 0.4 and the kernel size being 0.2 are configured. For the online BRFF, the random weight ω and b are used to construct the RFFS shown in (2). For fair comparison, the same Fourier basis as (2) is chosen as the activation function of OS-ELM, and learning is performed one by one in OS-ELM. The dimension of the feature space in OS-ELM is also set to 20.

To validate the stability of online BRFF in system identification, the corresponding Lyapunov function of $\hat{\theta}_k^T \mathbf{P}_k^{-1} \hat{\theta}_k$ established in (B9) is shown in Figure D1. It can be seen from Figure D1 that the Lyapunov function of online BRFF is decreasing and convergent provided that δ_D^2 is small enough, which guarantees its stability. Figure D2 shows the performance comparison of KLMS, OS-ELM and online BRFF in the system identification. From Figure D2, we see that the online BRFF can achieve a faster convergence rate than OS-ELM and KLMS. In addition, the online BRFF has a lower steady state mean square error than KLMS.

Appendix D.2 Mackey-Glass Time Series Prediction

The MG time series, which exhibits periodic and chaotic dynamics, is generated by the following nonlinear time-delay differential equation [15]:

$$\frac{dx(t)}{dt} = -0.1x(t) + \frac{0.2x(t-\tau)}{1+x(t-\tau)^{10}}, \quad (\text{D2})$$

where $\tau = 30$. The time series is first discretized using a sampling period of 6 s to obtain discrete variable $x(k)$. The obtained time series is corrupted by additive Gaussian noise with zero mean and a standard deviation of 0.04. The input $\mathbf{x}_k = [x(k-10), x(k-9), \dots, x(k-2), x(k-1)]^T$ is chosen to predict the current $x(k)$ that is the desired output y_k . A time series of 2000 samples is used as the training data and another 400 samples are used as the test data.

The dimension of the random feature space D is a crucial parameter in the online BRFF, which affects the filtering performance. Hence, the dependence of the performance on D is first studied by simulations. For the online BRFF, the variance of diffusion noise is set to $\delta_D^2 = 2 \times 10^{-5}$; the random weight ω is drawn from $\mathcal{N}(\omega; \mathbf{0}, 1.2 \times \mathbf{I}_D)$ and b uniformly from $[0, 2\pi]$.

Figure D3 shows the learning curves of online BRFF versus D in the MG time series prediction, and Figure D4 presents the averaged consumed time and steady-state MSEs over 50 Monte Carlo runs versus D , where the steady-state MSE is obtained by averaging the final 200 iterations. It can be seen from Figure D3 and Figure D4 that the filtering accuracy of online BRFF can be improved by increasing D at the expense of increasing the computational overhead, but is almost kept unchanged when D exceeds 400. Hence, to balance the performance and computational cost, an appropriate value of D should be used. Here, $D = 200$ is chosen in the following simulations.

For fair comparison, the activation function of OS-ELM is same as that used in the system identification, and learning is also performed one by one in OS-ELM. The dimension of the feature space in OS-ELM is also set to 200. For KLMS, the step size is 0.1 and the kernel size is 1.0. The learning curves of online and batch BRFF, OS-ELM, and KLMS are shown in Figure D5. We see that the online BRFF outperforms OS-ELM and KLMS, and approaches the batch BRFF from the aspect of filtering accuracy. The mean consumed time of online BRFF with 2.535 s, which is almost the same as that of OS-ELM, is far less than that of KLMS with 17.759 s that has a linearly increasing network size. Therefore, no sparsification is required in the online BRFF for online applications. Combining the filtering accuracy and computational time, the proposed online BRFF is more efficient than KLMS and OS-ELM.

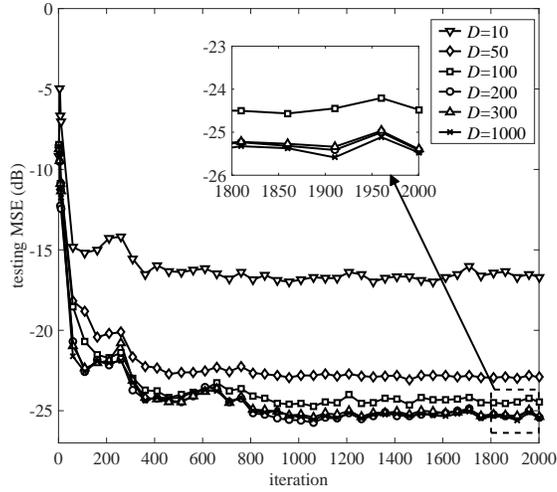


Figure D3 Learning curves of online BRFF for different dimensions of feature space in MG time series prediction.

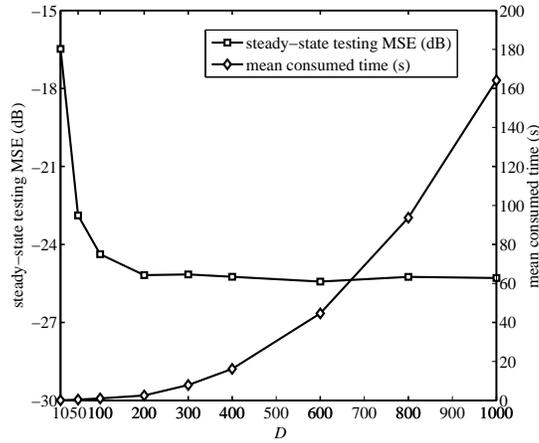


Figure D4 Mean consumed time versus dimension D of feature space in MG time series prediction.

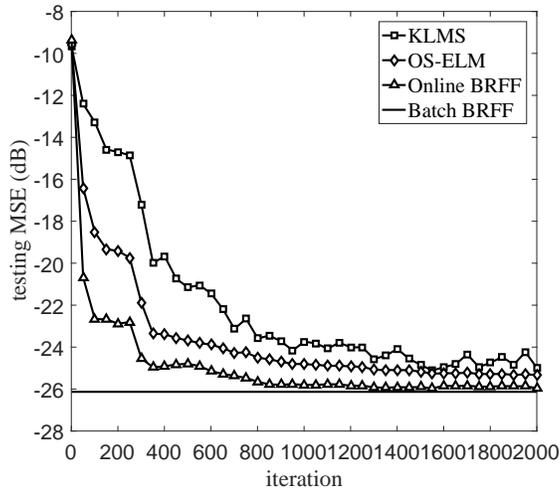


Figure D5 Learning curves of KLMS, OS-ELM, online BRFF and batch BRFF in MG time series prediction.

Appendix D.3 Nonlinear regression

In order to test the tracking performance of online BRFF, the nonlinear regression is considered here. The representative nonlinear system is [16]

$$y_k = y_{k-1}(\mathbf{a}(1) - \mathbf{a}(2) \exp(-y_{k-1}^2)) - y_{k-2}(\mathbf{a}(3) + \mathbf{a}(4) \exp(-y_{k-1}^2)) + \mathbf{a}(5) \sin(y_{k-1} \pi), \quad (\text{D3})$$

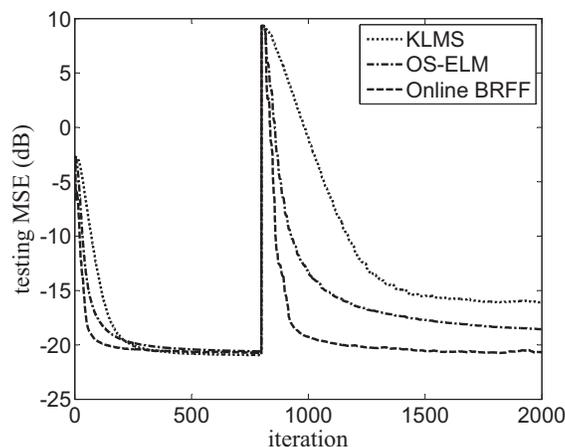


Figure D6 Learning curves of KLMS, OS-ELM and online BRFF in nonlinear regression.

where y_k is the output at discrete time k ; $y_{-1} = 0.1$ and $y_{-2} = 0.1$ are set as the initial values; and $\mathbf{a} = [\mathbf{a}(1), \dots, \mathbf{a}(5)]^T$ denotes the coefficient vector. According to (D3), the nonlinear regression can be described using $\mathbf{x}(k) = [y_{k-1}, y_{k-2}]^T$ and y_k as the input vector and the desired output, respectively.

The tracking performance is evaluated in a nonstationary environment, where two different coefficient vectors are used for data generation in (D3). We generate Data A and Data B in (D3) by using $\mathbf{a} = [0.8, 0.5, 0.3, 0.9, 0.1]^T$ and $\mathbf{a} = [0.2, 0.7, 0.8, 0.8, 0.2]^T$, respectively. All the data are corrupted by a zero-mean Gaussian noise with variance $\sigma_n^2 = 0.004$. Therefore, nonstationary data are generated by Data A for $0 \leq k \leq 800$ and Data B for $801 \leq k \leq 2000$. For KLMS, the kernel size of the Gaussian kernel is 1.5 and the step size is 0.1. The parameters for the online BRFF and OS-ELM are the same as those used in the aforementioned MG time series prediction. The averaged MSEs of KLMS, OS-ELM and online BRFF are plotted in Figure D6. It can be seen from Figure D6 that compared with KLMS and OS-ELM, the online BRFF can provide not only a faster convergent rate but also better tracking performance with smaller steady-state MSE in a nonstationary scenario.

References

- 1 Rasmussen C E, Williams C K I. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press, 2006
- 2 Särkkä S. Bayesian Filtering and Smoothing. Cambridge: Cambridge University Press, 2013
- 3 Wu Z, Shi J, Zhang X, et al. Kernel recursive maximum correntropy. *Signal Process*, 2015, 117: 11-16
- 4 Xiong K, Zhang H, Chan C. Performance evaluation of UKF-based nonlinear filtering. *Automatica*, 2006, 42: 261-270
- 5 Kivinen J, Smola A J, Williamson R C. Online learning with kernels. *IEEE Trans Signal Process*, 2004, 52: 2165-2176
- 6 Ma W, Duan J, Man W, et al. Robust kernel adaptive filters based on mean p-power error for noisy chaotic time series prediction. *Eng Appl Artif Intell*, 2017, 58: 101-110
- 7 Pokharel P P, Liu W, Príncipe J C. Kernel least mean square algorithm with constrained growth. *Signal Process*, 2009, 89: 257-265
- 8 Tao J W, Chung F L, Wang S T. A kernel learning framework for domain adaptation learning. *Sci China Inf Sci*, 2012, 55: 1983-2007
- 9 Van V S, Lazaro-Gredilla M, Santamaria I. Kernel recursive least-squares tracker for time-varying regression. *IEEE Trans Neural Network Learn Syst*, 2012, 23: 1313-1326
- 10 Liang N Y, Huang G B, Saratchandran P, et al. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Trans Neural Netw*, 2006, 17: 1411-1423
- 11 Deng C W, Huang G B, Xu J, et al. Extreme learning machines: New trends and applications. *Sci China Inf Sci*, 2015, 58: 1-16
- 12 Rahimi A, Recht B. Random features for large-scale kernel machines. *Adv Neural Inf Process Syst*, 2007, 1177-1184
- 13 Hu Z, Lin M, Zhang C S. Dependent online kernel learning with constant number of random Fourier features. *IEEE Trans Neural Network Learn Syst*, 2015, 26: 2464-2476
- 14 Bouboulis P, Pougkakiotis S, Theodoridis S. Efficient KLMS and KRLS algorithms: A random Fourier feature perspective. *Proc. IEEE Statist. Signal Process. Workshop 2016*, 1-5
- 15 Liu W, Príncipe J C, Haykin S. *Kernel Adaptive Filtering: A Comprehensive Introduction*. New York: John Wiley&Sons, 2010
- 16 Takizawa M, Yukawa M. Adaptive nonlinear estimation based on parallel projection along affine subspaces in reproducing kernel Hilbert space. *IEEE Trans Signal Process*, 2015, 63: 4257-4269