Statistical Mechanical Analysis of Gaussian Processes

Jun Tsuzurugi *

1-1 Ridai-cho, Kita-ku, Okayama-city,Okayama 700-0005, Japan

In this paper, we analyze Gaussian processes using statistical mechanics. Although the input variable is originally multidimensional, we simplify our model by considering the input variable as one-dimensional for statistical mechanical analysis. Furthermore, we employ periodic boundary conditions as an additional modeling approach. By using periodic boundary conditions, we can diagonalize the covariance matrix. The diagonalized covariance matrix is then applied to Gaussian processes. This allows for a statistical mechanical analysis of Gaussian processes using the derived diagonalized matrix. The results indicate that the analytical solutions obtained in this method closely match the results from simulations.

1. Introduction

With the publication of seminal texts on machine learning,[1] it can be considered that the field of machine learning is nearly mature. Within the field of machine learning, there exists a Gaussian process[2]-.[9] Although there are already studies that have analyzed Gaussian processes using statistical methods,[10] no research has yet solved Gaussian processes by deriving exact solutions. In this paper, we aim to perform a rigorous analysis of a Gaussian process using statistical mechanical analysis.

To perform a statistical mechanical analysis, we conduct mathematical calculations through modeling.[11] Specifically, we first impose periodic boundary conditions. Next, although the inputs of a Gaussian process are multidimensional, we consider them to be one-dimensional. These model assumptions facilitate discussions on Fourier space and allow for the diagonalization of the covariance matrix, i.e., the kernel matrix.

In this study, we focus on the case where the input variable points of a Gaussian process are equally spaced, and we derive an analytical solution for the mean squared error using a statistical mechanics approach. Here, we discuss the validity and significance of assuming equally spaced inputs.

*juntuzu@ous.ac.jp

Generally, the input variable points of a Gaussian process are assumed to be arbitrary. However, in certain application areas and problem settings, the input variable points are naturally equally spaced. Moreover, the assumption of equally spaced inputs can enable analytical tractability and provide valuable theoretical insights. The analytical solution derived in this study is particularly useful in such situations. Specifically, the following points support the validity of the equally spaced input assumption in our study:

1.1 Applicability to equally spaced sampling in practical applications

The results of this study are particularly useful when dealing with data sampled at equal intervals. For instance, in time series analysis, data are often acquired at regular intervals using sensors. Time series data such as temperature, stock prices, and audio signals are frequently sampled at equal intervals, and the analytical solution obtained in this study is applicable when applying Gaussian processes to these data. Furthermore, the assumption of equally spaced inputs can also be valid for modeling spatial fields using data obtained from sensors arranged at equal intervals on a spatial grid, such as meteorological observation and soil moisture data. In addition, cases where the input variable data are inherently arranged in a regular grid, such as pixel data in digital images, are also possible. The assumption is not limited to these examples and can be reasonable or approximated as such in a wide range of situations.

1.2 Enabling theoretical analysis

By assuming equally spaced inputs, the kernel matrix can be diagonalized, and Fourier transform techniques can be applied to the relevant equations, which enabled us to derive an analytical solution for the mean squared error in this study. This analytical solution provides valuable theoretical insights into the behavior of Gaussian processes. In particular, since it is generally difficult to obtain analytical solutions for arbitrary input points, the analytical solution obtained under the specific condition of equally spaced inputs serves as a stepping stone for understanding the properties of Gaussian processes.

1.3 Usefulness as a benchmark for numerical methods and approximations

The analytical solution derived in this study serves as a benchmark for evaluating the accuracy of numerical methods and approximations for arbitrary input points. In general, approximation methods such as variational inference and Markov Chain Monte Carlo methods are used to compute Gaussian processes for arbitrary inputs. The accuracy of these approximations can be assessed by comparing them with the analytical solution in the case of equally spaced inputs.

1.4 Limitations and future work

Note that the analytical solution in this study is derived under the condition of equally spaced inputs and is not directly applicable to cases with arbitrary input points. This is a limitation of the present study. However, as discussed above, there are application areas where the assumption of equally spaced inputs is reasonable, and the analytical solution also has theoretical significance.

Future work includes deriving analytical results for arbitrary input points without assuming equally spaced inputs. Although this is generally a challenging problem, the analytical solution obtained in this study for equally spaced inputs can provide a foundation for further theoretical development. Another important direction for future research is to apply the results of this study to various application areas and improve the accuracy of analysis for equally spaced sampled data.

In this paper, we perform the following case distinctions to ensure that the model satisfies the periodic boundary conditions. First, let N denote the number of input elements. In making distinctions, as shown in Fig. 1(a), if the distance $|a - b|$ between two points does not exceed $\frac{N}{2}$, the distance is kept as is. On the other hand, as shown in Fig. 1(b), if the distance $|c - d|$ between two points exceeds $\frac{N}{2}$, the distance between the two points is taken as $N - |c - d|$, as illustrated in Fig. 1(c).

In the study by Tsuzurugi and Eiho,[12] periodic boundary conditions were also used for natural image processing. The absence of notable artifacts in the output images further supports the validity of using periodic boundary conditions. In this paper, we conduct statistical mechanical calculations using a model with periodic boundary conditions and one-dimensional input, as illustrated in Fig. 1. By applying the covariance matrix obtained under these modeled conditions to a Gaussian process, we derive the exact analytical solution of a Gaussian process. We then demonstrate that this analytical solution is consistent with the simulation results.

2. Formulation

We denote N as the number of observed (training) data points and M as the number of discretized test inputs used to approximate the continuous domain. The index i $(1 \leq i \leq N)$ is assigned to training data (x_i, t_i), whereas the index j $(1 \leq j \leq M)$ is used for unobserved inputs x_j and their corresponding target values S_j or predicted values $u_{\mathbf{out}}$. Throughout this paper,

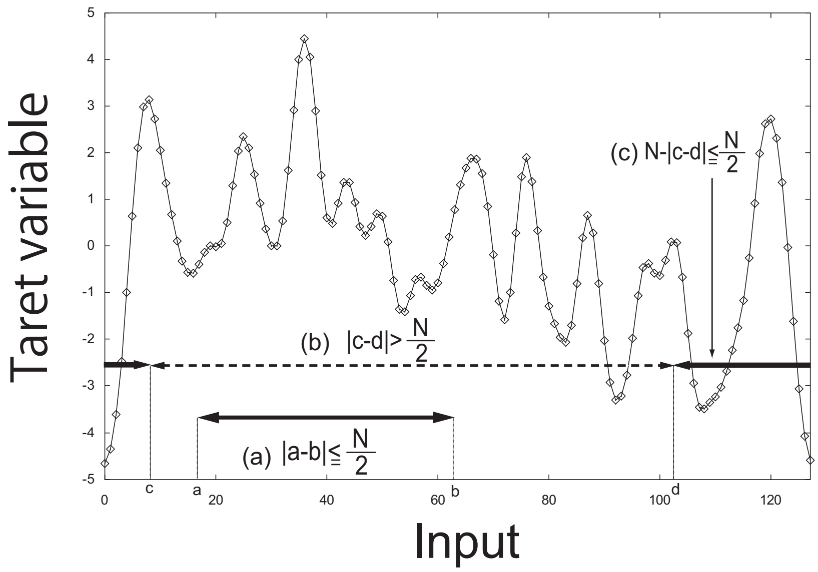

Fig. 1. Periodic boundary conditions are imposed on $N = 128$ evenly spaced discrete input values. The target variable follows a Gaussian distribution. Under the periodic boundary condition, when the distance between two points exceeds $\frac{N}{2}$, the alternative distance between the two points spans across the 0 point.

we consistently use i for observed variables (training data) and j for unobserved variables (test data) to avoid any notational ambiguity.

In this section, we define the key variables and notations used throughout the paper to clearly distinguish between observed and unobserved quantities. These definitions are fundamental for formulating the statistical mechanical analysis of Gaussian processes.

- **Observed input**: $\{x_i\}_{i=1}^N$ — Inputs in the training dataset.
- **Target variable**: $\boldsymbol{t}^* = \{t_i^*\}_{i=1}^N$ — The true (noise-free) values of the outputs corresponding to the observed inputs. These values are not observed directly.
- **Noisy observations**: $\boldsymbol{t} = \{t_i\}_{i=1}^N$ — The observed values of the target variable, which are discrete and include noise.
- **Unobserved input**: $\{x_j\}_{j=1}^M$ — New input values not included in the training set. In this study, these inputs are assumed to be equally spaced.
- **Unobserved output**: $\boldsymbol{S} = \{S_j\}_{j=1}^M$ — The true (noise-free) outputs corresponding to the unobserved inputs. These are the prediction targets and are not directly observable.
- **Predicted value**: $u_{\mathrm{out}} = \{u_j\}_{j=1}^M$ — The model's predicted values for the unobserved outputs, providing a continuous estimate of $\boldsymbol{S}$.

We denote vectors in bold and use the asterisk notation (*) to indicate true values without observation noise. This distinction allows us to formally separate learning from prediction, which is essential for the analytical formulation in later sections. We distinguish between two types of true output value in this paper:

- $\boldsymbol{t}^*$: The true (noise-free) output values corresponding to the observed inputs $\{x_i\}$ in the training dataset. Although these values represent the underlying function we aim to model, they are not directly observable owing to the presence of noise. Instead, we use the noisy observations $\boldsymbol{t}$ for training.
- $\boldsymbol{S}$: The true (noise-free) output values corresponding to the unobserved inputs $\{x_i\}$, which are not part of the training data. These values serve as the prediction targets and are also unobservable.

This distinction is fundamental for separating the theoretical learning objective from the prediction targets and is essential for the statistical mechanical formulation developed in this paper. Table I summarizes the differences between $\boldsymbol{t}^*$ and $\boldsymbol{S}$ used in this study. In general,

Symbol	**Input Type**	**Observable**	**Noise**	**Role**
$\boldsymbol{t}^*$	Observed input (x_i)	No	No	True value for learning
$\boldsymbol{S}$	Unobserved input (x_j)	No	No	True target for prediction

Table I. Comparison between $\boldsymbol{t}^*$ and $\boldsymbol{S}$

when the number of inputs is N, the unobserved inputs of a Gaussian process $\boldsymbol{x}_j (1 \le j \le N)$ are given in D dimensions. On the other hand, the target variable $t_i(1 \le j \le N)$, which is the output corresponding to the unobserved inputs, is one-dimensional. In this case, the dataset for training is defined as follows.

$$\begin{cases} \boldsymbol{x}_i & (1 \le i \le N) \\ t_i & (1 \le i \le N) \end{cases} \tag{1}$$

In a Gaussian process, it is assumed that the target variable t_i already has noise superimposed, following a Gaussian distribution with a mean of 0 and a variance of b^2. Additionally, when the unobserved input in D dimensions is denoted as $\boldsymbol{x}_{\mathbf{in}}$, its one-dimensional target variable is denoted as $t_{\mathbf{out}}$. Here, $x_{\mathbf{in}}$ denotes an unobserved input point for prediction that is not included in the training set. In this case, the unobserved dataset is defined as follows.

$$\begin{cases} \boldsymbol{x}_{\mathbf{in}} \\ t_{\mathbf{out}} \end{cases} \tag{2}$$

Here, x_j denotes equally spaced one-dimensional input points and t_j is the corresponding observed noisy output. In this paper, to further simplify the analysis, we model the inputs of a Gaussian process, as given in the equation, as equally spaced discrete values $\boldsymbol{j} = j$ and assume periodic boundary conditions for the target values $\boldsymbol{t} = t_j$. Therefore, the possible values of the elements of the unobserved input vector $\boldsymbol{j}$ are one-dimensional, and the training dataset is defined as follows.

$$\begin{cases} x_i & (1 \leq i \leq N) \\ t_j & (1 \leq j \leq N) \end{cases} \tag{3}$$

Let N be an even number. Additionally, let the one-dimensional unobserved input be $x_{\mathbf{in}}$ and its target variable be $u_{\mathbf{out}}$. Then, the unobserved dataset can be defined as follows.

$$\begin{cases} x_{\mathbf{in}} \\ u_{\mathbf{out}} \end{cases} \tag{4}$$

In a Gaussian process, both the true target variables without noise and the noise itself are given by Gaussian distributions. First, the probability distribution of the unobserved outputs without noise, $\boldsymbol{S} = S_j$, is given as follows.

$$P(\boldsymbol{S}) = \frac{1}{Z_{\boldsymbol{S}}} \exp\left(-\frac{1}{2}\boldsymbol{S}^T A^{-1} \boldsymbol{S}\right) \tag{5}$$

$$Z_{\boldsymbol{S}} = (2\pi)^{\frac{N}{2}} |A|^{\frac{1}{2}} \tag{6}$$

Here, A is the covariance matrix, and to satisfy the periodic boundary conditions, its elements A_{fj} are given by the following Eq. (7) and (9).

$$A_{fj} = a^2 \exp[-v^2 |f - j|^2] \tag{7}$$

$$(|f - j| \leq \frac{N}{2}) \tag{8}$$

$$A_{fj} = a^2 \exp[-v^2 (N - |f - j|)^2] \tag{9}$$

$$(|f - j| > \frac{N}{2}) \tag{10}$$

Here, a and v are real numbers. The larger the value of v, the greater the correlation between neighboring target variables, leading them to take similar values. Conversely, when v is small, the target variables tend to take independent values. The subscript f is used for A_{fj} instead of i to avoid confusion with the imaginary unit i used in the Fourier transform discussed later.

Here, B is the covariance matrix of the noise distribution, and its elements B_{fj} are given by

$$B_{fj} = b^2 \delta_{f,j} \; , \tag{11}$$

where b is a real number and $\delta_{f,j}$ is the Kronecker delta given by

$$\delta_{f,j} = \begin{cases} 1, & f = j \\ 0, & f \neq j \; . \end{cases} \tag{12}$$

In this case, the noise is the difference between the noisy observation $\boldsymbol{t}$ and the unobserved output $\boldsymbol{S}$, and the probability distribution of the noise is given by

$$P_{\mathbf{out}}(\boldsymbol{t}|\boldsymbol{S}) \quad = \quad \frac{1}{Z_{\mathbf{noise}}} \exp\left(-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{S})^T B^{-1}(\boldsymbol{t}-\boldsymbol{S})\right) \; , \tag{13}$$

$$Z_{\mathbf{noise}} = (2\pi)^{\frac{ND}{2}} |B|^{\frac{1}{2}} \; . \tag{14}$$

Since noisy observations are obtained by superimposing noise onto the unobserved output, and C_{fj} is also the covariance matrix, C_{fj} is, using A_{fj} and B_{fj}, determined as

$$C_{fj} = A_{fj} + B_{fj} \; . \tag{15}$$

The probability distribution of the target variables with noise (noisy observations) $\boldsymbol{t} = t_i$ is given as

$$P(\boldsymbol{t}|C) = \frac{1}{Z_C} \exp\left(-\frac{1}{2}\boldsymbol{t}^T C^{-1} \boldsymbol{t}\right) \; , \tag{16}$$

$$Z_C = (2\pi)^{\frac{ND}{2}} |C|^{\frac{1}{2}} \; . \tag{17}$$

Here, the posterior probability function of the predicted value $u_{\mathbf{out}}$ can be calculated given the target variables.

$$P(S\,|\boldsymbol{t}, a^2, v^2, b^2) = \frac{P(S, \boldsymbol{t}|a^2, v^2, b^2)}{P(\boldsymbol{t}|a^2, v^2, b^2)} \tag{18}$$

The predicted value $u_{\mathbf{out}}$ is defined as the posterior expectation of the random variable S given the observed data $\boldsymbol{t}$.

$$u_{\mathbf{out}} = \mathbb{E}[S \mid \boldsymbol{t}] \; , \tag{19}$$

where $\mathbb{E}[\cdot]$ denotes the expectation over the posterior distribution $P(S \mid \boldsymbol{t})$.

This is because determining $u_{\mathbf{out}}$ requires the degraded dataset $\boldsymbol{t}$. Furthermore, from C_{fj}, a^2 and v^2 required for A_{fj}, and b^2 required for B_{fj}, are also conditions.

The correlation between the unobserved input $j_{\mathbf{in}}$ and the elements of the training input j

is given as

$$k(j, j_{\mathbf{in}}) = a^2 \exp\left[-v^2(j - j_{\mathbf{in}})^2\right] . \tag{20}$$

Furthermore, the following can be stated:

$$\begin{aligned} & k_{\mathbf{in}}(j_{\mathbf{in}}, j_{\mathbf{in}}) \\ = \quad & a^2 \exp\left[-v^2(j_{\mathbf{in}} - j_{\mathbf{in}})^2\right] + b^2\delta_{\mathbf{in},\mathbf{in}} \\ = \quad & a^2 + b^2 . \end{aligned} \tag{21}$$

Here, we define the vector $\boldsymbol{k}$ with the element $k(j, j_{\mathbf{in}})$ as

$$\begin{aligned} \boldsymbol{k}^T &= \{k(1, j_{\mathbf{in}}), \cdots, k(j, j_{\mathbf{in}}), \cdots, k(N, j_{\mathbf{in}})\} \\ &= \{k_f\} . \end{aligned} \tag{22}$$

The mean of the unobserved outputs is given by[2)-?)]

$$\begin{aligned} u_{\mathbf{out}} &= \boldsymbol{k}^T C^{-1} \boldsymbol{t} \\ &= \sum_f \sum_j k_f C^{-1}_{fj} t_j . \end{aligned} \tag{23}$$

3. Diagonalization of the covariance matrix

By imposing discretization and periodic boundary conditions on the probabilistic model under consideration, we focus on cases where the covariance matrix becomes a symmetric circulant matrix, i.e., a translationally symmetric matrix. Under these conditions, the Fourier transform of S_j becomes possible.

$$\tilde{S}_p = \frac{1}{\sqrt{N}} \sum_j S_j e^{-i(pj)} \tag{24}$$

Here, i is the imaginary unit. The inverse Fourier transform of Eq. (24) is given by

$$S_j = \frac{1}{\sqrt{N}} \sum_p \tilde{S}_p e^{i(pj)} . \tag{25}$$

The possible values of the elements of the vector $\boldsymbol{p}$ are given as

$$0, \frac{2}{N}\pi, \frac{4}{N}\pi, \cdots, \frac{2(N-1)}{N}\pi . \tag{26}$$

In the case of a general two-dimensional array, the Fourier transform $\tilde{R}_{hk}$ of an arbitrary circulant matrix , R_{fj} , is given by

$$\tilde{R}_{hk} = \frac{1}{N} \sum_f \sum_j e^{i(hj-kf)} R_{fj} . \tag{27}$$

Here, we consider the Fourier transform of the matrix elements commonly used as covariance matrices, given by Eq. (27) . Owing to the imposed periodic boundary conditions, we distinguish the cases where $0 \leq l \leq \frac{N}{2}$. When f and j denote the positions of the elements, the original Gaussian process[2)-?)] defines A_{fj} as follows.

$$A_{fj} = a^2 \exp[-\sum_{d=1}^{D} v_d^2 |f^d - j^d|^2] \tag{28}$$

Here, d is not an exponent but an index of the input variable dimension. For simplicity, in this paper we consider the input variable dimension to be one-dimensional, hence, using Eq.(28) with $D = 1$. The result of calculating A_{fj} by distinguishing cases is given by the following equation:

$$\begin{aligned}\tilde{A}_k \equiv \tilde{A}_{kk} = a^2 \sum_{l=0}^{\frac{N}{2}} e^{-v^2 l^2} \cos(lk) \\ +a^2 \sum_{l=\frac{N}{2}+1}^{N-1} e^{-v^2 (N-l)^2} \cos(lk)\,.\end{aligned} \tag{29}$$

Similarly, by diagonalizing Eq. (11) using the Fourier transform, we obtain the following equation:

$$\tilde{B}_p = b^2\,. \tag{30}$$

Here, by applying the Fourier representation to Equation, we obtain the following equation:

$$u_{\mathbf{out}} = \sum_q \tilde{k}_{-q} \tilde{C}_q^{-1} \tilde{t}_q\,. \tag{31}$$

In this paper, the mean of the unknown target variable $u_{\mathbf{out}}$ is used as the restored data. Since $u_{\mathbf{out}}$ is a real number, the imaginary part of Eq. (31) can be disregarded. $\tilde{k}{-q}$ is the complex conjugate of $\tilde{k}q$. Furthermore, since $\tilde{C}q$ is diagonalized by the Fourier transform, to obtain $\tilde{C}^{-1}q$, it is sufficient to take the reciprocal of the diagonal elements of $\tilde{C}_q$. These are the advantages of introducing the Fourier transform under periodic boundary conditions.

4. Explanation of the theory

From here, we refer to the target value without noise S_j as "true value data" and the unobserved data $u_{\mathbf{out}}$ as "estimated value data". The noise, given by $\left|S_j - t_j\right|$, represents the difference between the degraded data t_j and the true target value S_j. We will briefly explain the method of calculating the mean squared error E_1 between the restored data $u_{\mathbf{out}}$ obtained in the previous section and the true value data S_j without noise. In this section, variables with a hat symbol ˆ are defined as unknown hyperparameters. We define E_1 as follows. Since E_1 is the mean squared error between the original data S_j and the restored data $u_{\mathbf{out}}$, the smaller the value, the better the model's output. Although Eq. (31) represents the target variable derived from arbitrary inputs and typically takes continuous values, in practice, when calculating it using the equation below, we will use discrete values and replace the subscript of $u_{\mathbf{out}}$ with j.

$$E_1 = \left\| \frac{1}{N} \sum_j \left(S_j - u_j \right)^2 \right\| \tag{32}$$

Here, $\|\cdot\|$ denotes the average with respect to the joint probability distribution $P(\boldsymbol{t}, \boldsymbol{S}) = P_{\mathbf{out}}(\boldsymbol{t}|\boldsymbol{S})P(\boldsymbol{S})$. Note that in this context, $\boldsymbol{S}$ denotes the true (noise-free) output values corresponding to the unobserved inputs, and $u_{\mathbf{out}}$ denotes the model's predicted output values for those same inputs. Therefore, the mean squared error given by Eq. (31) represents the generalization error, not the training error.

The Fourier representation of Eq. (32) is given as follows.

$$E_1 = \frac{1}{N} \sum_{\boldsymbol{k}} \left\| \left(\tilde{S}_{\boldsymbol{k}} - \tilde{u}_{\boldsymbol{k}} \right) \left(\tilde{S}_{-\boldsymbol{k}} - \tilde{u}_{-\boldsymbol{k}} \right) \right\| \tag{33}$$

Since $P(\boldsymbol{t}, \boldsymbol{S})$ is diagonalized in the Fourier representation, it can be easily computed.

Here, by setting as in the following equation,

$$\tilde{A}_{\boldsymbol{k}}^{-1} + \tilde{B}_{\boldsymbol{k}}^{-1} = \tilde{C}'_{\boldsymbol{k}} , \tag{34}$$

we obtain the following equation:

$$\begin{aligned}
&\left\| \left(\tilde{S}_{\boldsymbol{k}} - \tilde{u}_{\boldsymbol{k}} \right) \left(\tilde{S}_{-\boldsymbol{k}} - \tilde{u}_{-\boldsymbol{k}} \right) \right\| \\
= \; & \frac{1}{Z} \int \int d\tilde{S}_{\boldsymbol{k}} d\tilde{t}_{\boldsymbol{k}} \left| \tilde{S}_{\boldsymbol{k}} - \tilde{u}_{\boldsymbol{k}} \right|^2 \\
& \times \exp\left[- \frac{\tilde{C}'_{\boldsymbol{k}}}{2} \left| \tilde{S}_{\boldsymbol{k}} - \frac{\tilde{B}_{\boldsymbol{k}}^{-1}}{\tilde{C}'_{\boldsymbol{k}}} \tilde{t}_{\boldsymbol{k}} \right|^2 \right. \\
& \qquad \left. - \left\{ \frac{\tilde{B}_{\boldsymbol{k}}^{-1}}{2} - \frac{\left(\tilde{B}_{\boldsymbol{k}}^{-1} \right)^2}{2\tilde{C}'_{\boldsymbol{k}}} \right\} |\tilde{t}_{\boldsymbol{k}}|^2 \right] \\
= \; & \frac{1}{\tilde{C}'_{\boldsymbol{k}}} + \left(\frac{\tilde{B}_{\boldsymbol{k}}^{-1}}{\tilde{C}'_{\boldsymbol{k}}} - \hat{\tilde{A}}_{-\boldsymbol{k}} \hat{\tilde{C}}_{\boldsymbol{k}}^{-1} \right)^2 \frac{1}{\tilde{B}_{\boldsymbol{k}}^{-1} - \frac{\left(\tilde{B}_{\boldsymbol{k}}^{-1} \right)^2}{\tilde{C}'_{\boldsymbol{k}}}} .
\end{aligned} \tag{35}$$

Therefore, when the following condition is satisfied,

$$\hat{a} = a, \quad \hat{b} = b, \quad \hat{v} = v , \tag{36}$$

The mean squared error $E_{1_{min}}$ obtained using Eq. (34) reaches the following minimum value:

$$E_{1_{min}} = \frac{1}{N}\sum_k \frac{1}{\tilde{C}'_k} = \frac{1}{N}\sum_k \frac{1}{\tilde{A}_k^{-1} + \tilde{B}_k^{-1}} . \tag{37}$$

Thus, $E_{1_{\mathrm{min}}}$ is given by

$$E_{1_{min}} = \sum_k \frac{1}{N\left(\frac{1}{a^2 \sum_l e^{-v^2 l^2} e^{i(k\cdot l)} + a^2 \sum_l e^{-v^2 (N-l)^2} e^{i(k\cdot l)}} + \frac{1}{b^2}\right)} . \tag{38}$$

Similarly, the mean squared error between the true target variable without noise and the noisy target variable before restoration can be obtained as follows.

$$E_2 = \left\| \frac{1}{N}\sum_j \left(S_j - t_j\right)^2 \right\| = \frac{1}{N}\sum_k \tilde{B}_k = \frac{1}{N}\sum_k b^2 = b^2 \tag{39}$$

In general, for a Gaussian process, if the input variable values of the model are equally spaced and satisfy periodic boundary conditions, and the kernel is a symmetric circulant matrix, then A_{fj} and B_{fj} can be diagonalized by the discrete Fourier transform, and the general solution can be obtained as described above.

5. Results

Then, we present the analysis results for Eq. (38) and Eq. (39).

Fig. 2 shows the mean squared error for $N = 8192$ as an approximation of $N \to \infty$. Here, $a = 1.0$ and $b = 1.0$. The dashed line in (A) represents the theoretical value $E_{1_{\mathrm{min}}}$, which is the mean squared error after restoration according to Eq. (38). The solid line in (A) represents the mean squared error after restoration obtained from simulations. The dashed line in (B) represents the theoretical value E_2, which is the mean squared error before restoration according to equation. The solid line in (B) represents the mean squared error before restoration obtained from simulations. From Fig. 2, it can be confirmed that the simulation results after restoration are in close agreement with the theoretical values. This suggests that for $N \to \infty$, the simulation results will perfectly match the theoretical values. Additionally, $N \to \infty$ is equivalent to the case where periodic boundary conditions are not applied.

Here, the behavior of a Gaussian process for $v = 0.01$ is shown in Fig. 3. For $v = 0.01$, $N = 8192$, $a = 1.0$, and $b = 1.0$. The horizontal axis represents the input variable and the vertical axis represents the target variable. Although the original dataset contains $N = 8291$ points, the input range is truncated at 2000 to improve the figure's readability. The solid line denotes the

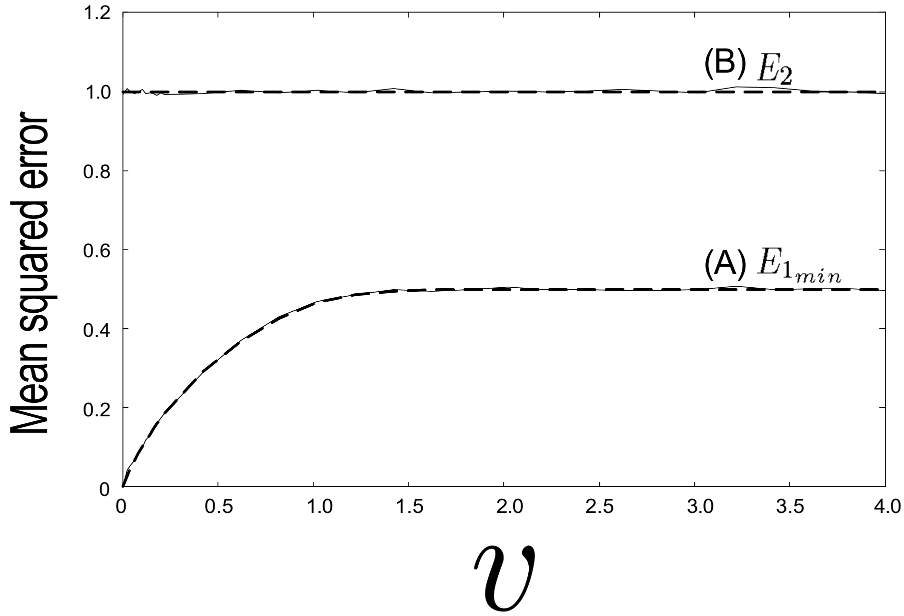

Fig. 2. The horizontal axis represents v and the vertical axis represents the mean squared error, in the case of $N = 8192$, $a = 1.0$, and $b = 1.0$. The dotted line (A) shows the mean squared error of the critical value, computed from Eq. (38). The solid line (A) shows the mean squared error obtained numerically after restoration. The dotted line (B) shows the mean squared error before restoration, computed from Eq. (39). The solid line (B) shows the mean squared error obtained numerically before restoration.

original data, the dashed line represents the restored data, and the diamond symbols indicate the degraded data. As shown in Fig. 3, the restored data are in close agreemennt with the original data, demonstrating the effectiveness of the restoration.

6. Conclusion

In this study, we performed precise diagonalization by case-separating a translationally symmetric matrix (translationally symmetric covariance matrix). Then, we applied the precisely diagonalized covariance matrix to a Gaussian process. As a result, we found that the analytical solution and the simulation results matched well in terms of the mean squared error before restoration (E_2) and after restoration (E_1).

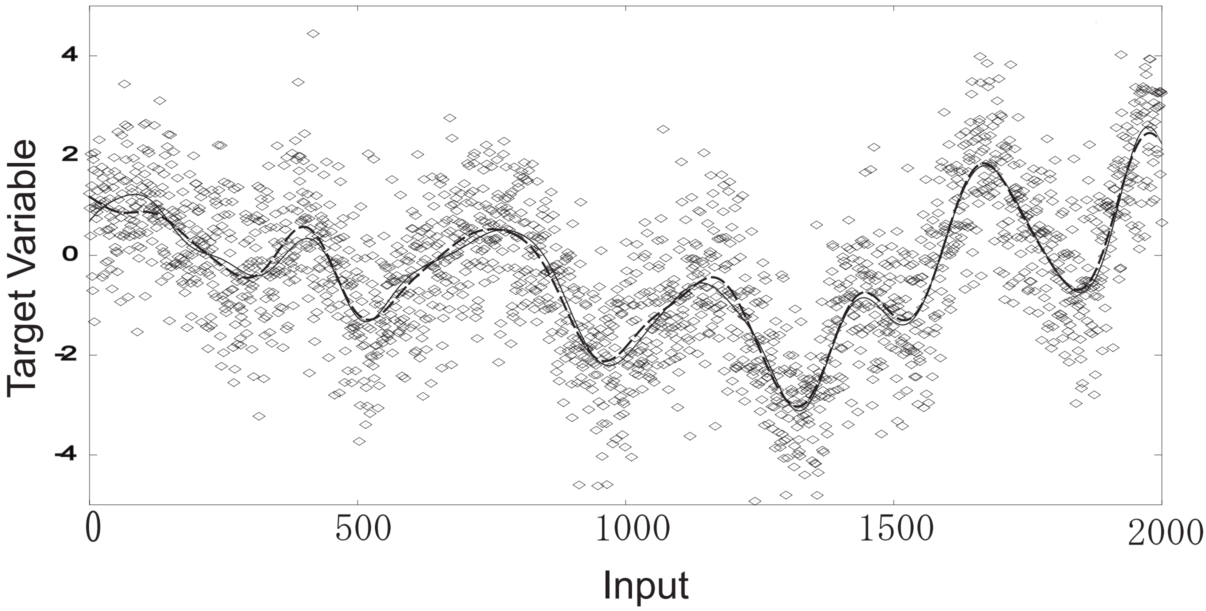

Fig. 3. The horizontal axis represents the input variable and the vertical axis represents the output variable, with $v = 0.08$, $N = 8192$, $a = 1.0$, and $b = 1.0$. The figure illustrates the relationship between the unobserved inputs $x_{\mathbf{in}}$ and their corresponding predicted outputs $u_{\mathbf{out}}$ obtained from the Gaussian process model. The solid line represents the predicted outputs $u_{\mathbf{out}}$, the dashed line represents the true (noise-free) values S, and the diamond symbols denote the noisy observations t. This visualization demonstrates how the model generalizes from the training data to estimate the target values for previously unseen inputs.

References

1) C. M. Bishop: *Pattern Recognition and Machine Learning* (Springer, New York, 2006).

2) C. K. I. Williams and C. E. Rasmussen: "Gaussian Processes for Regression," *Adv. Neural Inf. Process. Syst.* **8**, 514 (1996).

3) C. E. Rasmussen: "Evaluation of Gaussian Processes and Other Methods for Non-linear Regression," Ph.D. thesis, University of Toronto (1996).

4) M. Gibbs and D. J. C. MacKay: *Efficient Implementation of Gaussian Processes*, Technical Report, Cavendish Laboratory, University of Cambridge, UK (1997).

5) D. J. C. MacKay: "Introduction to Gaussian Processes," in *Neural Networks and Machine Learning*, ed. C. M. Bishop (Springer, New York, 1998) pp. 133–166.

6) C. K. I. Williams: "Prediction with Gaussian Processes: From Linear Prediction to Beyond," in *Learning in Graphical Models*, ed. M. I. Jordan (MIT Press, Cambridge, MA, 1999) pp. 599–621.

7) A. J. Smola and P. Bartlett: "Sparse Greedy Gaussian Process Regression," *Adv. Neural Inf. Process. Syst.* **13**, 619 (2001).

8) T. Yoshioka, S. Oba, and S. Ishii: "Gaussian Process Regression Using Representing Data," *Workshop on Information-Based Induction Science*, pp. 217–222 (July 2001).

9) D. J. C. MacKay: *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, UK, 2003).

10) N. Cressie: *Statistics for Spatial Data* (Wiley, New York, 1993).

11) J. Tsuzurugi and M. Okada: "Statistical Mechanics of Bayesian Image Restoration under Spatially Correlated Noise," *Phys. Rev. E* **66**, 066704 (2002).

12) J. Tsuzurugi and S. Eiho: "Image Restoration of Natural Images under Spatially Correlated Noise," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **E92-A**, 3, 809 (2009).