



Multivariate Gaussian processes: definitions, examples and applications

Zexun Chen¹ · Jun Fan² · Kuo Wang³

Received: 18 July 2022 / Accepted: 7 January 2023
© The Author(s) 2023

Abstract

Gaussian processes occupy one of the leading places in modern statistics and probability theory due to their importance and a wealth of strong results. The common use of Gaussian processes is in connection with problems related to estimation, detection, and many statistical or machine learning models. In this paper, we propose a precise definition of multivariate Gaussian processes based on Gaussian measures on vector-valued function spaces, and provide an existence proof. In addition, several fundamental properties of multivariate Gaussian processes, such as stationarity and independence, are introduced. We further derive two special cases of multivariate Gaussian processes, including multivariate Gaussian white noise and multivariate Brownian motion, and present a brief introduction to multivariate Gaussian process regression as a useful statistical learning method for multi-output prediction problems.

Keywords Gaussian measure · Gaussian process · Multivariate Gaussian process · Multivariate Gaussian distribution · Matrix-variate Gaussian distribution · Brownian motion

1 Introduction

In probability theory and statistics, Gaussian processes are used in connection with problems such as estimation, detection, stochastic analysis, even modern statistical or machine learning. These problems are often effectively formulated in terms of Gaussian measures on appropriate real-valued function space. As an important extension, vector-valued Gaussian processes have not been adequately explained in the literature so that the real-valued function space may limit applications of classical Gaussian processes since the correlation between multiple random values is difficult to be considered in many problems. Therefore, this paper is to consolidate the fundamentals of vector-valued stochastic processes, especially extending classical Gaussian process to multivariate Gaussian process by introducing Gaussian measures on vector-valued

✉ Zexun Chen
Zexun.Chen@ed.ac.uk

¹ University of Edinburgh Business School, Edinburgh, UK

² School of Mathematical Science, University of Nottingham Ningbo China, Ningbo, China

³ College of Mathematics, Physics and Information Engineering, Jiaying university, Jiaying, China

function spaces. The main contribution of this paper is to provide a concise proof and a proper explanation of the multivariate Gaussian process followed by examples and applications such as the multivariate Brownian motion and multivariate Gaussian process regression.

The paper is organised as follows. Section 2 introduces some preliminaries, including classical Gaussian measures, matrix-variate Gaussian distribution, stationary process, and classical Gaussian process. Section 3 presents some theoretical definitions of multivariate Gaussian process with the proof of existence. Two examples and one application of multivariate Gaussian processes which show their usefulness is presented in Sect. 4 and Sect. 5. Conclusions and a discussion are given in Sect. 6.

2 Preliminary

2.1 Gaussian measure

Definition 1 (*Gaussian measure on \mathbb{R}* [1]) Let $\mathcal{B}(\mathbb{R})$ denote the completion of the Borel σ -algebra on \mathbb{R} . Let $\lambda : \mathcal{B}(\mathbb{R}) \mapsto [0, +\infty]$ denote the usual Lebesgue measure. Then the Borel probability measure $\gamma : \mathcal{B}(\mathbb{R}) \mapsto [0, 1]$ is Gaussian with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$,

$$\gamma(A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) d\lambda(x),$$

for any measurable set $A \in \mathcal{B}(\mathbb{R})$.

A random variable X on a probability space $(\Omega, \mathcal{B}, \mathcal{P})$ is Gaussian with mean μ and variance σ^2 if its distribution measure is Gaussian, i.e.

$$\mathcal{P}(X \in A) = \gamma(A).$$

In terms of random variable, we have a definition of Gaussian random variable.

Definition 2 An n -dimensional random vector $\mathbf{X} = (X_1, \dots, X_n)$ is Gaussian if and only if $\langle \mathbf{a}, \mathbf{X} \rangle := \mathbf{a}^\top \mathbf{X} = \sum a_i X_i$ is a Gaussian random variable for all $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$.

In terms of measure, we can naturally have the definition of Gaussian measure on \mathbb{R}^n .

Definition 3 (*Gaussian measure on \mathbb{R}^n* [1]) Let γ be a Borel probability measure on \mathbb{R}^n . For each $\mathbf{a} \in \mathbb{R}^n$, denote a random variable $Y(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$ as a mapping $\mathbf{x} \mapsto \langle \mathbf{a}, \mathbf{x} \rangle \in \mathbb{R}$ on the probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \gamma)$. The Borel probability measure γ is a Gaussian measure on \mathbb{R}^n if and only if the random variable Y is Gaussian for each \mathbf{a} .

2.2 Matrix-variate Gaussian distribution

In statistics, the matrix-variate Gaussian distribution is a probability distribution that is a generalization of the multivariate normal distribution to matrix-valued random variables.

Definition 4 (*Matrix-variate Gaussian distribution* [2]) The random matrix is said to be Gaussian:

$$\mathbf{X} \sim \mathcal{MN}_{n,d}(M, U, V),$$

if and only if

$$\text{vec}(\mathbf{X}) \sim \mathcal{N}_{nd}(\text{vec}(M), V \otimes U),$$

where \otimes denotes the Kronecker products and $\text{vec}(\mathbf{X})$ denotes the vectorisation of \mathbf{X} .

2.3 Stationary process

Definition 5 (*Stationary process* [3]) Let $X = \{X_n(\omega)\}$ be a stochastic process on a probability space $(\Omega, \mathcal{B}, \mathcal{P})$. If for any integers n and h , two random vectors

$$(X_n, X_{n+1}, \dots, X_{n+k-1}) \text{ and } (X_{n+h}, X_{n+h+1}, \dots, X_{n+h+k-1})$$

have the same probability distribution, then X is said to be a *stationary process*.

2.4 Gaussian process

Let \mathbb{R}_T be function space of all \mathbb{R} -valued functions on T ¹. Consider the product topology on \mathbb{R}_T , which is defined as the smallest topology that makes the projection maps $\Pi_{t_1, \dots, t_n}(f) = [f(t_1), \dots, f(t_n)]$ from \mathbb{R}_T to \mathbb{R}^n measurable, and define \mathcal{F} as the Borel σ -algebra of this topology.

Definition 6 (*Gaussian measure on $(\mathbb{R}_T, \mathcal{F})$*) A measure γ on $(\mathbb{R}_T, \mathcal{F})$ is called as a Gaussian measure if for any $n \geq 1$ and $t_1, \dots, t_n \in T$, the push-forward measure $\gamma \circ \Pi_{t_1, \dots, t_n}^{-1}$ on \mathbb{R}^n is a Gaussian measure.

Briefly speaking, (multivariate) Gaussian distributions are Gaussian measures on \mathbb{R}^n , and Gaussian processes are Gaussian measures on the function space $(\mathbb{R}_T, \mathcal{F})$ due to the following theorem.

Theorem 1 (*Relationship between Gaussian process and Gaussian measure*) If $X = (X_t)_{t \in T}$ is a Gaussian process, then the push-forward measure $\gamma = \mathcal{P} \circ X^{-1}$ with $X : \Omega \mapsto \mathbb{R}_T$ is Gaussian on \mathbb{R}_T , namely, γ is a Gaussian measure on $(\mathbb{R}_T, \mathcal{F})$. Conversely, if γ is a Gaussian measure on $(\mathbb{R}_T, \mathcal{F})$, then on the probability space $(\mathbb{R}_T, \mathcal{F}, \gamma)$, the co-ordinate random variable $\Pi = (\Pi_t)_{t \in T}$ is from a Gaussian process.

The proof of the relationship between Gaussian process and Gaussian measure can be found in [4]. Motivated by the relationship between Gaussian measures and Gaussian processes, we properly define multivariate Gaussian processes by extending Gaussian measures on real-valued function space to vector-valued function space.

3 Multivariate Gaussian process

Gaussian processes (GPs) have been proven to be an effective statistical learning method for nonlinear problems due to many desirable properties, such as a clear structure with Bayesian interpretation, a simple integrated approach of obtaining and expressing uncertainty in predictions and the capability of capturing a wide variety of data feature by hyper-parameters

¹ Usually, T is a time space, however it could be considered as any arbitrary space (both one-dimensional and multi-dimensional, both discrete and continuous space) other than time.

[5, 6]. With the development of Gaussian processes related to machine learning algorithms, Gaussian process applications face a conspicuous limitation. The classical GP models can be only used to deal with a single output or single response problem because the process itself is defined on \mathbb{R} , and as a result the correlation between multiple tasks or responses cannot be taken into consideration [6, 7]. In order to overcome the drawback above, [8] proposed a framework to perform multi-output prediction using multivariate Gaussian processes (MV-GPs). Although [8] showed the usefulness of the proposed methods via data-driven examples, some theoretical issues of multivariate Gaussian processes, such as existence, independence condition, and stationary condition, are still not clear.

3.1 Definitions

Following the classical theory of Gaussian measure, matrix-variate Gaussian distribution, and Gaussian process, we introduce Gaussian measure on matrix space and Gaussian measure on \mathbb{R}^d -valued function space, and finally define the multivariate Gaussian process.

According to Definition 3 (equivalence between Gaussian measure on \mathbb{R}^n and random variable on \mathbb{R}^n) and Definition 4 (equivalence between random variable on \mathbb{R}^{nd} and random matrix on $\mathbb{R}^{n \times d}$), we can have a definition of Gaussian measure on $\mathbb{R}^{n \times d}$.

Definition 7 (*Gaussian measure on $\mathbb{R}^{n \times d}$*) Let γ be a Borel probability measure on $\mathbb{R}^{n \times d}$. For each $\mathbf{a} \in \mathbb{R}^{nd}$, denote a random variable $Y(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^{n \times d}$ as a mapping $\mathbf{x} \mapsto \langle \mathbf{a}, \text{vec}(\mathbf{x}) \rangle \in \mathbb{R}$ on the probability space $(\mathbb{R}^{n \times d}, \mathcal{B}(\mathbb{R}^{n \times d}), \gamma)$. The Borel probability measure γ is a Gaussian measure on $\mathbb{R}^{n \times d}$ if and only if the random variable Y is Gaussian for each \mathbf{a} .

Let $(\mathbb{R}^d)_T$ be all \mathbb{R}^d -valued functions on T and consider the smallest topology on \mathbb{R}^d that makes the projection mappings $\Xi_{t_1, \dots, t_n}(\mathbf{f}) = [\mathbf{f}(t_1), \dots, \mathbf{f}(t_n)]$ from $(\mathbb{R}^d)_T$ to $\mathbb{R}^{n \times d}$ measurable, and define \mathcal{G} as the Borel σ -algebra of this topology. Thus we can have a definition of Gaussian measure on \mathbb{R}^d -valued function space.

Definition 8 (*Gaussian measure on $((\mathbb{R}^d)_T, \mathcal{G})$*) A measure γ on $((\mathbb{R}^d)_T, \mathcal{G})$ is called as a Gaussian measure if for any $n \geq 1$ and $t_1, \dots, t_n \in T$, the push-forward measure $\gamma \circ \Xi_{t_1, \dots, t_n}^{-1}$ on $\mathbb{R}^{n \times d}$ is a Gaussian measure.

Inspired by the relationship between Gaussian process and Gaussian measure in Theorem 1, we can appropriately define *multivariate Gaussian processes* (MV-GPs).

Definition 9 (*d-variate Gaussian process*) Given a Gaussian measure on $((\mathbb{R}^d)_T, \mathcal{G})$, $d \geq 1$, the co-ordinate random vector $\Xi = (\Xi_t)_{t \in T}$ on the probability space $((\mathbb{R}^d)_T, \mathcal{G}, \gamma)$ is said to be from a *d-variate Gaussian process*².

Given the definition of MV-GPs, it is essential to show the existence before proceeding with further research. Inspired by the fact that GPs can be specified by mean function and covariance function (thus they can be denoted as $\mathcal{GP}(\mu, k)$), we present the proof of existence of MV-GPs by applying Daniell-Kolmogorov theorem.

Theorem 2 (Existence of *d-variate Gaussian process*³) For any index set T , any vector-valued mean function $\mathbf{u} : T \mapsto \mathbb{R}^{1 \times d}$, any covariance function $k : T \times T \mapsto \mathbb{R}$ and any

² The co-ordinate random vector can be either column and row. For simplicity, we use row random vector only in our discussions of the *d-variate Gaussian process*.

³ In order to clarify the shape of random vector, we use the notation, \mathbb{R}^d for column vector only, and use the notation, $\mathbb{R}^{1 \times d}$, for row vector only.

positive semi-definite parameter matrix $\Lambda \in \mathbb{R}^{d \times d}$, there exists a probability space $(\Omega, \mathcal{G}, \mathcal{P})$ and a d -variate Gaussian process $\mathbf{f}(t) \in \mathbb{R}^{1 \times d}$, $t \in T$ on this space, whose mean function is \mathbf{u} , covariance function is k and parameter matrix is Λ , such that,

- $\mathbb{E}[\mathbf{f}(t)] = \mathbf{u}(t), \quad \forall t \in T,$
- $\mathbb{E}[(\mathbf{f}(t_s) - \mathbf{u}(t_s))(\mathbf{f}(t_l) - \mathbf{u}(t_l))^T] = \text{tr}(\Lambda)k(t_s, t_l), \quad \forall t_s, t_l \in T,$
- $\mathbb{E}[(\mathbf{F}_{t_1, \dots, t_n} - \mathbf{M}_{t_1, \dots, t_n})^T(\mathbf{F}_{t_1, \dots, t_n} - \mathbf{M}_{t_1, \dots, t_n})] = \text{tr}(K_{t_1, \dots, t_n})\Lambda, \quad \forall n \geq 1, t_1, \dots, t_n \in T,$ where

$$\begin{aligned} \mathbf{M}_{t_1, \dots, t_n} &= [\mathbf{u}(t_1)^T, \dots, \mathbf{u}(t_n)^T]^T \\ \mathbf{F}_{t_1, \dots, t_n} &= [\mathbf{f}(t_1)^T, \dots, \mathbf{f}(t_n)^T]^T \\ K_{t_1, \dots, t_n} &= \begin{bmatrix} k(t_1, t_1) & \dots & k(t_1, t_n) \\ \vdots & \ddots & \vdots \\ k(t_n, t_1) & \dots & k(t_n, t_n) \end{bmatrix}. \end{aligned}$$

We denote $\mathbf{f} \sim \mathcal{MG}\mathcal{P}_d(\mathbf{u}, k, \Lambda)$.

Proof Given $n > 1$, for every $t_1, \dots, t_n \in T$, a Gaussian measure γ_{t_1, \dots, t_n} on $\mathbb{R}^{n \times d}$ satisfies the assumptions of Daniell-Kolmogorov theorem because the projection of a matrix Gaussian distribution on $\mathbb{R}^{n \times d}$ with $[\mathbf{u}(t_1)^T, \dots, \mathbf{u}(t_n)^T]^T \in \mathbb{R}^{n \times d}$, $n \times n$ column covariance matrix $K = (k_{i,j}) \in \mathbb{R}^{n \times n}$, and $d \times d$ row covariance matrix $\Lambda \in \mathbb{R}^{d \times d}$, to the first $n - 1$ coordinates, is precisely the Gaussian distribution with $[\mathbf{u}(t_1)^T, \dots, \mathbf{u}(t_{n-1})^T]^T \in \mathbb{R}^{(n-1) \times d}$, $(n - 1) \times (n - 1)$ column covariance matrix $K = (k_{i,j}) \in \mathbb{R}^{(n-1) \times (n-1)}$, and row covariance matrix $\Lambda \in \mathbb{R}^{d \times d}$ due to the conditional property of matrix Gaussian distribution [2, 8]. By the Daniell-Kolmogorov theorem, there exists a probability space $(\Omega, \mathcal{G}, \mathcal{P})$ as well as a d -variate Gaussian process $X = (X_t)_{t \in T} \sim \mathcal{MG}\mathcal{P}_d(\mathbf{u}, k, \Lambda)$ defined on this space such that any finite dimensional distribution of $[X_{t_1}, \dots, X_{t_n}]$ is given by the measure γ_{t_1, \dots, t_n} . \square

3.2 Properties

Following the existence of d -variate GPs, we also achieve some properties as follow.

Proposition 3 (Stationary) A d -variate Gaussian process $\mathcal{MG}\mathcal{P}_d(\mathbf{u}, k, \Lambda)$ is said to be stationary if

$$\mathbf{u}(t) = \mathbf{u}(t + h), \quad k(t_s + h, t_l + h) = k(t_s, t_l), \forall t, t_s, t_l, h \in T.$$

Proof Let $\mathbf{f} \sim \mathcal{MG}\mathcal{P}_d(\mathbf{u}, k, \Lambda)$, then for $\forall n \geq 1, t_1, \dots, t_n \in T$,

$$[\mathbf{f}(t_1)^T, \dots, \mathbf{f}(t_n)^T]^T \sim \mathcal{MN}(\mathbf{u}_{t_1, \dots, t_n}, K_{t_1, \dots, t_n}, \Lambda),$$

where $\mathcal{MN}(\cdot, \cdot, \cdot)$ is matrix-variate Gaussian distribution,

$$\mathbf{u}_{t_1, \dots, t_n} = \begin{bmatrix} \mathbf{u}(t_1) \\ \vdots \\ \mathbf{u}(t_n) \end{bmatrix}, \quad K_{t_1, \dots, t_n} = \begin{bmatrix} k(t_1, t_1) & \dots & k(t_1, t_n) \\ \vdots & \ddots & \vdots \\ k(t_n, t_1) & \dots & k(t_n, t_n) \end{bmatrix}.$$

Given any time increment $h \in T$, there also exists,

$$[\mathbf{f}(t_1 + h)^T, \dots, \mathbf{f}(t_n + h)^T]^T \sim \mathcal{MN}(\mathbf{u}_{t_1+h, \dots, t_n+h}, K_{t_1+h, \dots, t_n+h}, \Lambda),$$

where

$$\mathbf{u}_{t_1+h, \dots, t_n+h} = \begin{bmatrix} \mathbf{u}(t_1+h) \\ \vdots \\ \mathbf{u}(t_n+h) \end{bmatrix},$$

$$K_{t_1+h, \dots, t_n+h} = \begin{bmatrix} k(t_1+h, t_1+h) & \cdots & k(t_1+h, t_n+h) \\ \vdots & \ddots & \vdots \\ k(t_n+h, t_1+h) & \cdots & k(t_n+h, t_n+h) \end{bmatrix}.$$

Since $\mathbf{u}(t) = \mathbf{u}(t+h)$, $k(t_s+h, t_l+h) = k(t_s, t_l)$, $\forall t, t_s, t_l, h \in T$,

$$\mathbf{u}_{t_1+h, \dots, t_n+h} = \begin{bmatrix} \mathbf{u}(t_1+h) \\ \vdots \\ \mathbf{u}(t_n+h) \end{bmatrix} = \begin{bmatrix} \mathbf{u}(t_1) \\ \vdots \\ \mathbf{u}(t_n) \end{bmatrix} = \mathbf{u}_{t_1, \dots, t_n},$$

$$K_{t_1+h, \dots, t_n+h} = \begin{bmatrix} k(t_1+h, t_1+h) & \cdots & k(t_1+h, t_n+h) \\ \vdots & \ddots & \vdots \\ k(t_n+h, t_1+h) & \cdots & k(t_n+h, t_n+h) \end{bmatrix} = \begin{bmatrix} k(t_1, t_1) & \cdots & k(t_1, t_n) \\ \vdots & \ddots & \vdots \\ k(t_n, t_1) & \cdots & k(t_n, t_n) \end{bmatrix}$$

$$= K_{t_1, \dots, t_n}.$$

Therefore, $[\mathbf{f}(t_1+h)^\top, \dots, \mathbf{f}(t_n+h)^\top]^\top$ has the same probability distribution as $[\mathbf{f}(t_1)^\top, \dots, \mathbf{f}(t_n)^\top]^\top$. Due to the arbitrary choice of $n > 1$ and $t_1, \dots, t_n \in T$, $\mathbf{f} \sim \mathcal{MGPD}_d(\mathbf{u}, k, \Lambda)$ is a stationary process according to Definition 5. □

Proposition 4 (Independence) *A d -collection of functions $\{\mathbf{f}_i\}_{i=1,2,\dots,d}$ identically independently follows a Gaussian process $\mathcal{GP}(\mu, k)$ if and only if*

$$\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_d] \sim \mathcal{MGPD}_d(\mathbf{u}, k, \Lambda),$$

where $\mathbf{u} = [\mu, \dots, \mu] \in \mathbb{R}^d$ and Λ is any diagonal positive semi-definite matrix.

Proof Necessity: if $\mathbf{f} \sim \mathcal{MGPD}_d(\mathbf{u}, k, \Lambda)$, then for $\forall n \geq 1, t_1, \dots, t_n \in T$,

$$[\mathbf{f}(t_1)^\top, \dots, \mathbf{f}(t_n)^\top]^\top \sim \mathcal{MN}(\mathbf{u}_{t_1, \dots, t_n}, K_{t_1, \dots, t_n}, \Lambda),$$

where,

$$\mathbf{u}_{t_1, \dots, t_n} = \begin{bmatrix} \mathbf{u}(t_1) \\ \vdots \\ \mathbf{u}(t_n) \end{bmatrix}, \quad K_{t_1, \dots, t_n} = \begin{bmatrix} k(t_1, t_1) & \cdots & k(t_1, t_n) \\ \vdots & \ddots & \vdots \\ k(t_n, t_1) & \cdots & k(t_n, t_n) \end{bmatrix}.$$

Rewrite the left, we obtain

$$[\xi_1, \xi_2, \dots, \xi_d] \sim \mathcal{MN}(\mathbf{u}_{t_1, \dots, t_n}, K_{t_1, \dots, t_n}, \Lambda),$$

where $\xi_i = [f_i(t_1), f_i(t_2), \dots, f_i(t_n)]^\top$. Since Λ is a diagonal matrix, for any $i \neq j$

$$\mathbb{E}[\xi_i^\top \xi_j] = \text{tr}(K_{t_1, \dots, t_n} \Lambda_{ij}) = \text{tr}(K_{t_1, \dots, t_n}) \cdot 0 = 0.$$

Because ξ_i and ξ_j are any finite number of realisations of \mathbf{f}_i and \mathbf{f}_j respectively from the same Gaussian process, \mathbf{f}_i and \mathbf{f}_j are uncorrelated. Due to joint finite realisations of \mathbf{f}_i and \mathbf{f}_j follow Gaussian, non-correlation implies independence.

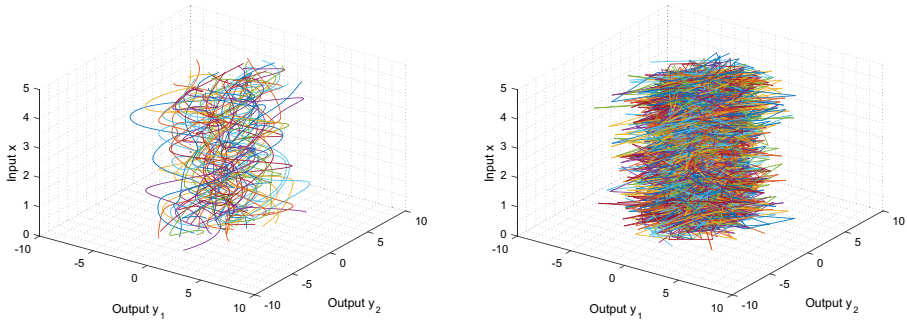


Fig. 1 The 50 random realisation sample points generated from of 2-variate Gaussian process. Left: A centred 2-variate Gaussian process with Gaussian covariance function $k(t_s, t_l) = 1.5 \exp(-(t_s - t_l)^2/2/0.5^2)$. Right: A 2-variate Gaussian white noise as a 2-variate Gaussian process with covariance function $k(t_s, t_l) = 1.5\delta(t_s, t_l)$

Sufficiency: if $\{f_i\}_{i=1,2,\dots,d} \sim \mathcal{GP}(0, k)$ are independent, for $\forall n \geq 1, t_1, \dots, t_n \in T$ and for any $i \neq j$,

$$0 = \mathbb{E}[\xi_i^\top \xi_j] = \text{tr}(K_{t_1, \dots, t_n}) \Lambda_{ij}.$$

Since $\text{tr}(K_{t_1, \dots, t_n})$ is non-zero, Λ_{ij} must be 0. Due the arbitrary choices of i, j , Λ must be diagonal. That is to say, $\xi_i = [f_i(t_1), \dots, f_i(t_n)]^\top$ can be written as a matrix Gaussian distribution $\mathcal{MN}(\mathbf{u}_{t_1, \dots, t_n}, K_{t_1, \dots, t_n}, \Lambda)$ where Λ is a diagonal positive semi-definite matrix. Since for $\forall n \geq 1, t_1, \dots, t_n \in T$ the above result holds, $\{f_i\}_{i=1, \dots, d}$ are identically independent Gaussian processes $\mathcal{GP}(\mu, k)$. □

4 Example: special cases

Instinctively, a special case is *centred multivariate Gaussian process* where vector-valued mean function $\mu = \mathbf{0}$. The 50 realisation samples generated from a centred multivariate Gaussian process are demonstrated in Fig. 1: Left. Furthermore, we can derive the multivariate Gaussian white noise and the multivariate Brownian motion.

4.1 Multivariate Gaussian white noise

Let $f = [f_1, \dots, f_d] \sim \mathcal{MGPD}(\mathbf{0}, \sigma^2 \mathbb{1}(t_s, t_l), \Lambda)$, where $\mathbb{1}(\cdot)$ is an indicator function that equals 1 if $t_s = t_l$, otherwise 0, thus

$$\begin{aligned} \mathbb{E}[f(t)] &= \mathbf{0}, \quad \forall t \in T, \\ \mathbb{E}\left[f(t_s) f(t_l)^\top\right] &= \text{tr}(\Lambda) \sigma^2 \mathbb{1}(t_s, t_l) = \begin{cases} 0 & \text{if } t_s \neq t_l \\ \sigma^2 \text{tr}(\Lambda) & \text{if } t_s = t_l \end{cases}, \quad \forall t_s, t_l \in T. \end{aligned}$$

Furthermore, $\forall n \geq 1, t_1, \dots, t_n \in T$, there exists,

$$\mathbb{E}[\mathbf{F}_{t_1, \dots, t_n}^\top \mathbf{F}_{t_1, \dots, t_n}] = \text{tr}(K_{t_1, \dots, t_n}) \Lambda = \text{tr}(\sigma^2 \mathbf{I}_{d \times d}) \Lambda = d \sigma^2 \Lambda,$$

where $F_{t_1, \dots, t_n} = [f(t_1)^\top, \dots, f(t_n)^\top]^\top$.

Therefore, it is nature to have a definition of the d -variate Gaussian white noise.

Definition 10 (*d-variate Gaussian white noise*) A d -variate Gaussian process $\mathcal{MG}\mathcal{P}_d(\mathbf{u}, k, \Lambda)$ is said to be a d -variate Gaussian white noise if $\mathbf{u} = \mathbf{0}$ and $k(t_s, t_t) = \sigma^2 \mathbb{1}(t_s, t_t)$.

Remark 1 We observe that d -variate Gaussian white noise has independence property as white noise along with T , but it has correlation along with d -variate dimension. Therefore, d -variate Gaussian white noise is also called as variate-dependent Gaussian white noise or variate-correlated Gaussian white noise, which is distinct from the traditional d -dimensional independent Gaussian white noise. Here are 50 realisation samples generated from a multi-variate Gaussian white noise shown in Fig. 1: Right.

4.2 Multivariate Brownian motion

According to the Chapter 2 of the book [9], there is a definition of Brownian motion, which is a Gaussian white noise whose intensity is Lebesgue measure. Since Brownian motion is a special case of Gaussian process with continuous sample paths, mean function $u = 0$ and covariance function $k(s, t) = \min(s, t)$, we propose an example, d -variate Brownian motion, as a special case of d -variate Gaussian process with vector-valued mean function $\mathbf{u} = \mathbf{0}$, covariance function $k(s, t) = \min(s, t)$ and parameter matrix Λ . Based on the Theorem 2, we derive some properties of the traditional Brownian motion to a more general vector-valued case.

Definition 11 (*d-variate Brownian motion*) A d -variate Gaussian process $\mathcal{MG}\mathcal{P}_d(\mathbf{u}, k, \Lambda)$ is said to be d -variate Brownian motion⁴ if all sample paths are continuous, $\mathbf{u} = \mathbf{0}$ and $k(t_s, t_t) = \min(t_s, t_t)$.

Let B_t be a d -variate Brownian motion, which means for all $0 \leq t_1 \leq \dots \leq t_n$ the random matrix $Z = (B_{t_1}^\top, \dots, B_{t_n}^\top)^\top \in \mathbb{R}^{n \times d}$ has a Gaussian distribution on the probability space $(\Omega, \mathcal{G}, \mathcal{P})$ mentioned in Theorem 2. There exists a matrix $M \in \mathbb{R}^{n \times d}$ and two non-negative definite matrices $C = [c]_{jm} \in \mathbb{R}^{n \times n}$ and $\Lambda = [\lambda]_{ab} \in \mathbb{R}^{d \times d}$ such that

$$\begin{aligned} \mathbb{E} \left[\exp \left(i \sum_{j=1}^n W_{j,\cdot} Z_{j,\cdot}^\top \right) \right] &= \exp \left(-\frac{1}{2} \sum_{j,m} W_{j,\cdot} c_{jm} W_{m,\cdot}^\top + i \sum_j W_{j,\cdot} M_{j,\cdot}^\top \right), \\ \mathbb{E} \left[\exp \left(i \sum_{a=1}^d W_{\cdot,a} Z_{\cdot,a}^\top \right) \right] &= \exp \left(-\frac{1}{2} \sum_{a,b} W_{\cdot,a} \lambda_{ab} W_{\cdot,b}^\top + i \sum_a W_{\cdot,a} M_{\cdot,a}^\top \right), \end{aligned}$$

where $W = [w]_{ja} \in \mathbb{R}^{n \times d}$ and i is the imaginaryary unit. Moreover, we also have the mean value $M = \mathbb{E}[Z]$ and two covariance matrices

$$\begin{aligned} c_{jm} &= \mathbb{E}[(Z_{j,\cdot} - M_{j,\cdot})(Z_{m,\cdot} - M_{m,\cdot})^\top], \\ \lambda_{ab} &= \mathbb{E}[(Z_{\cdot,a} - M_{\cdot,a})(Z_{\cdot,b} - M_{\cdot,b})^\top]. \end{aligned}$$

Assume that the mean matrix M here is a zero matrix, i.e. $\mathbb{E}[Z] = \mathbb{E}[Z|t = 0] = 0, I_d = \Lambda$, and

$$C = \begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \vdots & & \vdots \\ t_1 & t_2 & \dots & t_n \end{bmatrix}.$$

⁴ For clarification, we only use row random vector in our discussions of the d -variate Brownian motion.

Hence, $\mathbb{E}[B_t] = 0$ for all $t \geq 0$ and

$$\begin{aligned} \mathbb{E}[B_t B_t^\top] &= dt, \quad \mathbb{E}[B_t B_s^\top] = d \min(s, t), \\ \mathbb{E}[B_t^\top B_t] &= t \Lambda, \quad \mathbb{E}[B_t^\top B_s] = \min(s, t) \Lambda. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \mathbb{E}[(B_t - B_s)(B_t - B_s)^\top] &= \mathbb{E}[B_t B_t^\top - 2B_s B_t^\top + B_s B_s^\top] = d|t - s|, \\ \mathbb{E}[(B_t - B_s)^\top (B_t - B_s)] &= \mathbb{E}[B_t^\top B_t - 2B_s^\top B_t + B_s^\top B_s] = |t - s| \Lambda. \end{aligned}$$

Note that this d -variate Brownian motion B_t still has independent increments since $\mathbb{E}[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})^\top] = 0$ and $\mathbb{E}[(B_{t_i} - B_{t_{i-1}})^\top (B_{t_j} - B_{t_{j-1}})] = 0$ when $t_i < t_j$ holds for all $0 < t_1 < \dots < t_n$.

Remark 2 Similar to d -variate Gaussian white noise, d -variate Brownian motion also has independence property along with T , but it has correlation along with d -variate dimension. Therefore, d -variate Brownian motion is also called as variate-dependent Brownian motion or variate-correlated Brownian motion, which is distinct from the "traditional" d -dimensional Brownian motion. Actually, the "traditional" d -dimensional Brownian motion is a special case of d -variate Brownian motion with diagonal matrix Λ .

As a Brownian motion, we then introduce Itô lemma for the d -variate Brownian motion.

Let $B_t = [B_1(t), \dots, B_d(t)]$ be the d -variate Brownian motion derived in Section 4.2. Then, we have the following lemma.

Lemma 5 (Itô lemma for the d -variate Brownian motion) *Let F be a twice continuously differentiable real function on \mathbb{R}^{d+1} and let $\Lambda = [\lambda]_{i,j} \in \mathbb{R}^{d \times d}$ be the covariance matrix for the d -variate dimension. Then,*

$$\begin{aligned} F(t, B_1(t), \dots, B_d(t)) &= F(0, B_1(0), \dots, B_d(0)) + \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial B_i}(s, B_1(s), \dots, B_d(s)) dB_i(s) \\ &+ \int_0^t \left\{ \frac{\partial F}{\partial s}(s, B_1(s), \dots, B_d(s)) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 F}{\partial B_i \partial B_j}(s, B_1(s), \dots, B_d(s)) \lambda_{i,j} \right\} ds. \end{aligned}$$

Proof By Itô lemma and the definition of the d -variate Brownian motion, we obtain

$$\begin{aligned} F(t, B_1(t), \dots, B_d(t)) &= F(0, B_1(0), \dots, B_d(0)) + \int_0^t \frac{\partial F}{\partial s}(s, B_1(s), \dots, B_d(s)) ds \\ &+ \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial B_i}(s, B_1(s), \dots, B_d(s)) dB_i(s) \\ &+ \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 F}{\partial B_i \partial B_j}(s, B_1(s), \dots, B_d(s)) d\langle B_i, B_j \rangle(s). \end{aligned}$$

The proof is complete by $d\langle B_i, B_j \rangle(s) = \lambda_{i,j} ds$. □

5 Application: multivariate Gaussian process regression

Multi-output prediction is a good example of a practical application of multivariate Gaussian processes. Multivariate Gaussian process modelling provides a solid and unified framework

for predicting multiple responses or tasks by exploiting their correlations. As a regression problem, multivariate Gaussian process regression (MV-GPR) has closed-form expressions for marginal likelihoods and predictive distributions, so parameter estimations can employ the same optimization techniques as conventional Gaussian process modelling [8].

As a summary of MV-GPR in [8], the noise-free multi-output regression model is considered and the noise term is incorporated into the kernel function. Given n pairs of observations $\{(x_i, \mathbf{y}_i)\}_{i=1}^n$, $x_i \in \mathbb{R}^p$, $\mathbf{y}_i \in \mathbb{R}^d$, we assume the following model

$$\mathbf{f} \sim \mathcal{MGPD}(\mathbf{0}, k', \Lambda), \quad \mathbf{y}_i = \mathbf{f}(x_i), \text{ for } i = 1, \dots, n,$$

where Λ is an undetermined covariance matrix (the relationship between different outputs), $k' = k(x_i, x_j) + \delta_{ij}\sigma_n^2$, and δ_{ij} is Kronecker delta. According to multivariate Gaussian process, it yields that the collection of functions $[\mathbf{f}(x_1), \dots, \mathbf{f}(x_n)]$ follows a matrix-variate Gaussian distribution

$$[\mathbf{f}(x_1)^\top, \dots, \mathbf{f}(x_n)^\top]^\top \sim \mathcal{MN}(\mathbf{0}, K', \Lambda),$$

where K' is the $n \times n$ covariance matrix of which the (i, j) -th element $[K']_{ij} = k'(x_i, x_j)$. Therefore, the predictive targets $\mathbf{f}_* = [f_{*1}, \dots, f_{*m}]^\top$ at the test locations $X_* = [x_{n+1}, \dots, x_{n+m}]^\top$ is given by

$$p(\mathbf{f}_* | X, Y, X_*) = \mathcal{MN}(\hat{M}, \hat{\Sigma}, \hat{\Lambda}),$$

where

$$\begin{aligned} \hat{M} &= K'(X_*, X)^\top K'(X, X)^{-1} Y, \\ \hat{\Sigma} &= K'(X_*, X_*) - K'(X_*, X)^\top K'(X, X)^{-1} K'(X, X_*), \end{aligned}$$

and

$$\hat{\Lambda} = \Lambda.$$

Here $K'(X, X)$ is an $n \times n$ matrix of which the (i, j) -th element $[K'(X, X)]_{ij} = k'(x_i, x_j)$, $K'(X_*, X)$ is an $m \times n$ matrix of which the (i, j) -th element $[K'(X_*, X)]_{ij} = k'(x_{n+i}, x_j)$, and $K'(X_*, X_*)$ is an $m \times m$ matrix with the (i, j) -th element $[K'(X_*, X_*)]_{ij} = k'(x_{n+i}, x_{n+j})$. In addition, the expectation and the covariance are obtained,

$$\begin{aligned} \mathbb{E}[\mathbf{f}_*] &= \hat{M} = K'(X_*, X)^\top K'(X, X)^{-1} Y, \\ \text{cov}(\text{vec}(\mathbf{f}_*^\top)) &= \hat{\Sigma} \otimes \hat{\Lambda} = [K'(X_*, X_*) - K'(X_*, X)^\top K'(X, X)^{-1} K'(X, X_*)] \otimes \Lambda. \end{aligned}$$

From a data science perspective, the hyperparameters involved in the covariance function (kernel) $k'(\cdot, \cdot)$ and the row covariance matrix of MV-GPR needs to be estimated from the training data using a variety of methods [10], including maximum likelihood estimation, maximum a posteriori and Markov Chain Monte Carlo [11].

6 Conclusion

In this paper, we provide a formal definition of the multivariate Gaussian process as well as several of its related properties, including existence, stationarity and independence. Additionally, we present some special cases and examples of multivariate Gaussian processes, such as multivariate Brownian motion and multivariate Gaussian white noise. Finally, we present an useful application of multivariate Gaussian process regression in statistical learning.

Acknowledgements The authors would like to thank Dr. Youssef El-Khatib for his comments and Dr. Gregory Markowsky for his kind proofreading and very helpful comments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Eldredge, N.: Analysis and probability on infinite-dimensional spaces. arXiv preprint [arXiv:1607.03591](https://arxiv.org/abs/1607.03591) (2016)
2. Gupta, A.K., Nagar, D.K.: Matrix variate distributions, vol. 104. Chapman and Hall/CRC, New York (1999)
3. Hida, T., Hitsuda, M.: Gaussian processes, vol. 120. American Mathematical Soc, Providence, Rhode Island (1993)
4. Rajput, B.S., Cambanis, S.: Gaussian processes and Gaussian measures. The annals of mathematical statistics, 1944–1952 (1972)
5. Rasmussen, C.E.: Evaluation of gaussian processes and other methods for non-linear regression. PhD thesis, University of Toronto, Toronto, Canada (1997)
6. Boyle, P., Frean, M.: Dependent Gaussian processes. Adv. Neural Inf. Process. Syst. **17**, 217–224 (2005)
7. Wang, B., Chen, T.: Gaussian process regression with multiple response variables. Chemom. Intell. Lab. Syst. **142**, 159–165 (2015)
8. Chen, Z., Wang, B., Gorban, A.N.: Multivariate Gaussian and student-t process regression for multi-output prediction. Neural Comput. Appl. **32**(8), 3005–3028 (2020)
9. Le Gall, J.-F.: Brownian motion, Martingales, and stochastic calculus, vol. 274. Springer, Berlin, Heidelberg (2016)
10. Williams, C.K., Barber, D.: Bayesian classification with Gaussian processes. IEEE Trans. Pattern Anal. Mach. Intell. **20**(12), 1342–1351 (1998)
11. Rasmussen, C.E., Williams, C.K.: Gaussian Processes for machine learning, vol. 1. MIT press, Cambridge, Massachusetts (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.