

Literature Review: Sparse Variational Gaussian Processes

1. Gaussian Measures in Hilbert Spaces [wild2022generalized], [Kukush'2019]

Gaussian measures are typically defined as a Lebesgue measure on a physical probability space $(\Omega, \mathcal{A}, \mathbb{P})$. However, there does not exist an infinite-dimensional equivalent to the Lebesgue measure. This means that in most cases, we cannot assume the existence of an infinite-dimensional analog of a probability measure that exists in a finite-dimensional space. However, there are special cases where it is possible to define a push-forward measure from $(\Omega, \mathcal{A}, \mathbb{P})$ to an infinite-dimensional space. In the following, we define the infinite-dimensional analog of a Gaussian measure $\mathbb{P}^F(A)$, a push-forward measure constructed from an underlying finite-dimensional $(\Omega, \mathcal{A}, \mathbb{P})$. In particular, we will define $\mathbb{P}^F(A)$ on a Hilbert space \mathcal{H} with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

1.1 Gaussian Random Elements

Consider a random element $F \in \mathcal{H}$. F is a *Gaussian Random Element* (GRE) if $\forall h \in \mathcal{H}$:

$$\langle F, h \rangle_{\mathcal{H}} \sim \mathcal{N}(\mu(F, h), \sigma^2(F, h)) \quad (1)$$

In other words, for each $h \in \mathcal{H}$, $\langle F, h \rangle_{\mathcal{H}}$ is a Gaussian random variable (possibly with zero variance). The mean can be written as an inner product:

$$\mu(F, h) = \langle m, h \rangle_{\mathcal{H}} \quad (2)$$

where:

$$m := \int F(\omega) d\mathbb{P}(\omega) \quad (3)$$

the expectation of $F(\omega)$ with respect to some probability measure \mathbb{P} on a finite dimensional space Ω . The variance can be written as an inner product:

$$\sigma^2(F, h) = \langle C(h), h \rangle_{\mathcal{H}} \quad (4)$$

where C is the covariance operator:

$$C(h) := \int \langle F(\omega), h \rangle_{\mathcal{H}} F(\omega) d\mathbb{P}(\omega) - \langle m, h \rangle_{\mathcal{H}} m \quad (5)$$

Thus, we can denote:

$$\langle F, h \rangle_{\mathcal{H}} \sim \mathcal{N}(\langle m, h \rangle_{\mathcal{H}}, \langle C(h), h \rangle_{\mathcal{H}}) \quad (6)$$

and we denote that F is a GRE with:

$$F \sim \mathcal{N}(m, C) \quad (7)$$

1.2 Gaussian Measures [wild2022generalized]

We can define the push-forward measure of \mathbb{P} through the mapping $F : \Omega \rightarrow \mathcal{H}$, $F \in \mathcal{H}$ as $\mathbb{P}^F(A) := \mathbb{P}(F^{-1}(H))$ for all Borel-measurable $H \subset \mathcal{H}$. Moreover, if F is a Gaussian random element, then we can define the distribution $P := \mathbb{P}^F$, the corresponding push-forward measure, as a Gaussian measure. This formulation allows us to define a Gaussian distribution P over the infinite-dimensional Hilbert space \mathcal{H} . [JW: The ability to define $\mathbb{P}^F(A)$ seems to hinge on the existence of GREs in \mathcal{H} . Is there a reason why we can't define random elements in a similar fashion for other distributions that aren't Gaussian?]

2. Gaussian Processes [wild2022generalized], [wild2023connections]

A Gaussian process $F \sim GP(m, k)$, is a random function mapping $F : \mathcal{X} \rightarrow \mathbb{R}$, defined with respect to a mean function $m : \mathcal{X} \rightarrow \mathbb{R}$ and a positive definite kernel $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. In particular, for N points $\mathbf{X} = \{x_n\}_{n=1}^N$ such that $x_n \in \mathcal{X}$, the evaluation $F_X := [F(x_1) \cdots F(x_N)]^T \in \mathbb{R}^N$ will follow a Gaussian distribution:

$$F_X \sim P_{\mathbf{X}}(\mathbf{Y}) = \mathcal{N}(\mathbf{m}_X, \mathbf{K}_{X,X}) \quad (8)$$

where $\mathbf{m}_X = [m(x_1) \cdots m(x_N)]^T \in \mathbb{R}^N$ and $\mathbf{K}_{X,X} \in \mathbb{R}^{N \times N}$ having elements $[\mathbf{K}_{X,X}]_{n,n'} = k(x_n, x_{n'})$ for $n, n' = 1, \dots, N$.

2.1 Gaussian Process Regression [wild2023connections]

Consider the regression task where we have N observation pairs $\{(x_n, y_n)\}_{n=1}^N$ such that $x_n \in \mathcal{X}$ and $y_n \in \mathbb{R}$. Using the Gaussian Process formulation in (8), we can ‘predict’ for M points $\{x_m\}_{m=1}^M$ by using Bayes’ Rule to calculate the posterior distribution of \mathbf{Y}_M :

[JW: ↓ Need to fix this ↓]

$$F_{X_M}|Y_N \sim P_{\mathbf{X}_M}(\mathbf{Y}_M|\mathbf{Y}_N = \{y_n\}_{n=1}^N) = \frac{P_{\mathbf{X}_M}(\mathbf{Y}_M)P_{\mathbf{X}_N}(\mathbf{Y}_N = \{y_n\}_{n=1}^N)}{\int_{\mathbb{R}^N} P_{\mathbf{X}_N}(\mathbf{Y}_N)d\mathbf{Y}_N} \quad (9)$$

With all terms in (9) being Gaussian, the posterior has the closed form expression:

$$P(\mathbf{X}_M|\mathbf{Y}_N, \mathbf{X}_N) = \mathcal{N}(\bar{\mathbf{m}}_{X_M}, \bar{\mathbf{K}}_{X_M, X_M}) \quad (10)$$

[JW: ↑ Need to fix this ↑]

where:

$$\bar{\mathbf{m}}_{X_M} = \mathbf{m}_{X_M} + \mathbf{K}_{X_M, X_N} (\mathbf{K}_{X_N, X_N} + \sigma^2 \mathbf{I}_N)^{-1} (\mathbf{Y}_N - \mathbf{m}_{X_N}) \quad (11)$$

$$\bar{\mathbf{K}}_{X_M, X_M} = \mathbf{K}_{X_M, X_M} - \mathbf{K}_{X_M, X_N} (\mathbf{K}_{X_N, X_N} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{K}_{X_N, X_M} \quad (12)$$

2.2 Gaussian Measures → Gaussian Processes [wild2022generalized]

A Gaussian process $F \sim GP(m, k)$ can have many Gaussian measure formulations. To define a corresponding Gaussian *measure* for a GP requires specification of an appropriate

measureable space. As shown in [wild2022generalized], a Gaussian process can be specified as a Gaussian measure P on a Hilbert space of square-integrable functions $L^2(\mathcal{X}, \rho, \mathbb{R})$ when the mean $m \in L^2(\mathcal{X}, \rho, \mathbb{R})$ and the kernel k satisfies:

$$\int_{\mathcal{X}} k(x, x) d\rho(x) < \infty \quad (13)$$

These conditions guarantee sample functions from F to have square-integrable paths, allowing for a corresponding Gaussian measure $P := \mathbb{P}^F \sim \mathcal{N}(m, C)$ defined on the *measureable space* $L^2(\mathcal{X}, \rho, \mathbb{R})$ with the same mean m and a covariance operator C :

$$C(f(\cdot)) := \int k(\cdot, x') f(x') d\rho(x') \quad (14)$$

for any function $f \in L^2(\mathcal{X}, \rho, \mathbb{R})$.

Although Gaussian processes and Gaussian measures are distinct concepts, (13) is satisfied for most GPs, ensuring the existence of corresponding Gaussian measures. As such, Gaussian measures and Gaussian processes can often be used interchangeably. Working with statistical divergences in the following sections, it will be more appropriate to use Gaussian measures instead of Gaussian processes.

[JW: Is it trivial to show that the NNGP kernel will satisfy (13)?]

3. Sparse Variational Gaussian Processes [wild2023connections]

3.1 Inducing Points [terenin2022numerically]

References