4 Gaussian processes

From a conceptual point of view, the theory of RKHS can be seen as the infinite-dimensional generalization of the concepts of quadratic forms and symmetric positive-definite matrices in linear algebra. Similarly, one may argue that the theory of generalized stochastic processes constitutes the infinite-dimensional generalization of multivariate probability theory, with Gaussian processes being the natural extension of Gaussian random variables.

Under construction: Since we do not expect all readers to be at ease with probability theory and multivariate statistics, we have provided a brief summary of the basic concepts and ideas in Appendix C. The other very useful feature of this appendix is that it actually yields a road map to guide our generalization.

For our discussion of stochastic processes, we assume that the reader is reasonably familiar with measure theory and the classical theory of probability, including the multivariate Gaussian distribution.

In a nutshell, we shall parallel the definition of random variable and probability measures in Appendix C.1 to introduce generalized stochastic processes in Section 4.1.

4.1 Generalized stochastic processes (GSP)

In the classical theory of probability, a real-valued random variable X is defined as a measurable function $\omega \mapsto X(\omega)$ from the probability space $(\Omega, \Sigma(\Omega), \mathscr{P})$ to the state space $\mathcal{X} = \mathbb{R}$ endowed with the Borelian σ-field $B(\mathbb{R})$. Here, Ω is the sample space of possible outcomes indexed by ω with native probability measure $\mathscr{P}$, while $B(\mathbb{R})$ is the collection of all Borel-measurable subsets of $\mathbb{R}$. The elements of $B(\mathbb{R})$ are called *events*—they typically correspond to some interval of $\mathbb{R}$ such as $E = (-\infty, x]$. The induced probability measure for X is $\mathscr{P}_X(E) = \mathscr{P}\{\omega \in \Omega : X(\omega) \in E\}$ for all $E \in B(\mathbb{R})$. The latter is also given by

$$\mathscr{P}_X(E) = \mathrm{Prob}(X(\omega) \in E) = \int_E p_X(x)\mathrm{d}x, \tag{125}$$

where p_X is the pdf of X and E is any Borel subset of $\mathbb{R}$. The other fundamental point is that this probability measure is in one-to-one correspondence

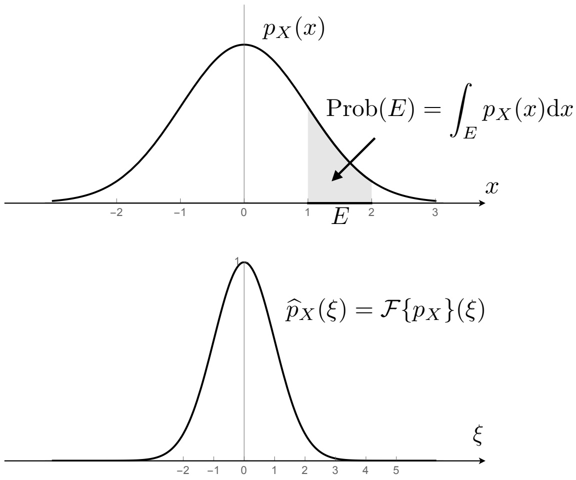

Figure 3: Probability density and characteristic function of a univariate Gaussian random variable.

with the characteristic function

$$\widehat{\mathscr{P}}_X(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}X\xi}\} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{j}x\xi}\mathscr{P}_X(\mathrm{d}x)$$
$$= \int_{\mathbb{R}} p_X(x)\mathrm{e}^{\mathrm{j}x\xi}\mathrm{d}x = \mathcal{F}^*\{p_X\}(\xi) \tag{126}$$

which is the (conjugate) Fourier transform of the pdf p_X. These quantities are illustrated in Fig. 3 for the special case of a standardized Gaussian random variable X_{Gauss} with mean $\mu = \mathbb{E}\{X_{\mathrm{Gauss}}\} = 0$ and variance $\sigma^2 = \mathbb{E}\{(X_{\mathrm{Gauss}} - \mu)^2\} = 1$.

The transition from the univariate setting to an infinite number of dimensions is achieved by replacing the state space $\mathbb{R}$ by some topological vector space $\mathcal{X}$. Moreover, there are theoretical advantages in taking $\mathcal{X}$ to be the dual of a nuclear space—the case of our interest being $\mathcal{X} = \mathcal{S}'(\mathbb{R}^d)$: Schwartz's space of tempered distributions. Within that extended framework, there are three possible ways to conceptualize a generalized stochastic process (GSP):

1. as a map that translates the outcome $\omega \in \Omega$ of a random experiment into a generalized function $G(\omega) \in \mathcal{S}'(\mathbb{R}^d)$;

2. as a random continuous linear functional $\varphi \mapsto G(\varphi)$ that associates a

real-valued random variable $G(\varphi) = \langle G, \varphi \rangle$ to any test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$;

3. as a global mechanism for generating collections of multivariate random variables $\boldsymbol{X} = (\langle G, \varphi_1 \rangle, \dots, \langle G, \varphi_N \rangle)$ with compatible statistical distributions.

These three complementary point of views are explained next. Remarkably, they turn out to be equivalent, thanks to the nuclear structure of the dual pair of spaces $\big(\mathcal{S}(\mathbb{R}^d), \mathcal{S}'(\mathbb{R}^d)\big)$.

4.1.1 GSPs as random generalized functions

A conventional random variable is a real variable that takes different values depending on the outcome of a random phenomenon. Likewise, one can think of a generalized stochastic process as a functional variable G that returns different generalized functions $g = G(\omega) \in \mathcal{S}'(\mathbb{R}^d)$ depending on the outcome $\omega \in \Omega$ of some random experiment.

This extension relies on the fundamental property that Borel σ-fields and probability measures can also be defined on general (infinite-dimensional) vector spaces (see Appendix F). This allows for the use of the same mapping technique as in the univariate case to specify generalized stochastic processes whose realizations are randomly selected elements of the extended state space $\mathcal{X} = \mathcal{S}'(\mathbb{R}^d)$. Such processes are then characterized by their probability measure $\mathscr{P}_G : B\big(\mathcal{S}'(\mathbb{R}^d)\big) \to [0,1]$ where $B\big(\mathcal{S}'(\mathbb{R}^d)$ is the Borel σ-field of $\mathcal{S}'(\mathbb{R}^d)$. This is stated formally as follows.

Definition 18 (Generalized random process)**.** *Let $(\Omega, \Sigma(\Omega), \mathscr{P})$ be our universal* [4] *probability space. Then, a generalized random process in $\mathcal{S}'(\mathbb{R}^d)$ is a* measurable *function*

$$G : \omega \mapsto G(\omega) \in \mathcal{S}'(\mathbb{R}^d)$$

from the space of possible outcomes Ω to the (state) space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$. It is fully characterized by the induced probability measure on $\mathcal{S}'(\mathbb{R}^d)$; that is,

$$\mathscr{P}_G(E) = \mathscr{P}\{\omega \in \Omega : G(\omega) \in E\},$$

for any Borel subset $E \subseteq \mathcal{S}'(\mathbb{R}^d)$.

[4] $\big(\Omega, \Sigma(\Omega)\big)$ should be sufficiently rich to map into $\big(\mathcal{S}'(\mathbb{R}^d), B\big(\mathcal{S}'(\mathbb{R}^d)\big)\big)$ which is implicit in the requirement that the map $\omega \mapsto G(\omega)$ should be measurable. In other words, the pre-image by G of any measurable subset E of $\mathcal{S}'(\mathbb{R}^d)$ should be included in $\Sigma(\Omega)$.

Concretely, this means that each realization $g = G(\omega)$ of G is a generalized function $g \in \mathcal{S}'(\mathbb{R}^d)$, while the random generation mechanism—i.e., the drawing of an outcome ω—is governed by the probability measure $\mathscr{P}_G$ on $\mathcal{S}'(\mathbb{R}^d)$ with

$$\mathrm{Prob}(G(\omega) \in E) = \mathscr{P}_G(E)$$

for any Borelian subset $E \in \mathcal{S}'(\mathbb{R}^d)$. This is to say that the measure $\mathscr{P}_G$ provides us with the knowledge of the probability associated with any region (or subset) E within the extended space of tempered distributions. This is very similar to the univariate scenario described by (125), except that there is no direct "infinite-dimensional" counterpart of the probability density function (pdf) that appears on the right-hand side of this equation. However, we shall see in Section 4.3 that there exists an infinite-dimensional analog of the characteristic function $\widehat{\mathscr{P}}_X(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\xi X}\}$ under the condition that the underlying topological vector space is nuclear, which is precisely the case for $\mathcal{S}'(\mathbb{R}^d)$. While this infinite-dimensional generalization results in a characteristic functional $\widehat{\mathscr{P}_G} : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ rather than a function, the good news is that it is endowed with the same properties as its finite-dimensional counterpart and that it is (almost) as easy to use in practice.

4.1.2 GSPs as random linear functionals

Instead of describing a generalized random process G through its "infinite-dimensional" probability measure, it is possible to characterize its effect on some generic test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Indeed, since the members of $\mathcal{S}'(\mathbb{R}^d)$ are linear functionals on $\mathcal{S}(\mathbb{R}^d)$, any realization $g = G(\omega)$ specifies a continuous linear map $\mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$

$$\varphi \mapsto \langle G(\omega), \varphi \rangle$$

for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Accordingly, if we fix φ, we can interpret $\omega \mapsto X_\varphi = \langle G(\omega), \varphi \rangle$ as a conventional real-valued random variable whose pdf $p_{X_\varphi} : \mathbb{R} \to \mathbb{R}^+$ can be determined, at least in principle, from $\mathscr{P}_G$. This suggest the following operational definition.

Definition 19 (Generalized stochastic process as a random functional). *A generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ is a random linear functional $\varphi \mapsto \langle G, \varphi \rangle$ on $\mathcal{S}(\mathbb{R}^d)$ with the following properties:*

- *Generation mechanism: for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the quantity $X_\varphi = \langle G, \varphi \rangle$ is an ordinary scalar random variable whose pdf p_{X_φ} is parametrized by φ.*

- *Linearity:* $\langle G, a_1\varphi_1 + a_2\varphi_2\rangle = a_1\langle G, \varphi_1\rangle + a_2\langle G, \varphi_2\rangle$ *in law for any* $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$ *and* $a_1, a_2 \in \mathbb{R}$.
- *Continuity: If the sequence* (φ_n) *is converging in* $\mathcal{S}(\mathbb{R}^d)$ *then* $\lim_{n\to\infty}\langle G, \varphi_n\rangle = \langle G, \lim_{n\to\infty}\varphi_n\rangle$ *in law.*

Definition 19 provides us with a convenient link between the theory of stochastic processes and the classical univariate setting of probability theory. One can then draw on this connection to define a number of basic properties of generalized stochastic processes.

Definition 20 (Statistical independence)**.** *Two generalized stochastic processes* G_1 *and* G_2 *in* $\mathcal{S}'(\mathbb{R}^d)$ *are said to be mutually independent if the random variables* $X_1 = \langle G_1, \varphi\rangle$ *and* $X_2 = \langle G_2, \varphi\rangle$ *are mutually independent for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$.

Definition 21 (Gaussian processes)**.** *A generalized stochastic process* G *in* $\mathcal{S}'(\mathbb{R}^d)$ *is said to be Gaussian if the random variable* $X = \langle G, \varphi\rangle$ *has a Gaussian distribution for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$*. In particular, it is a standardized Gaussian white noise (or Gaussian innovation process) if* X *has zero mean and variance* $\sigma_X^2 = \|\varphi\|^2_{L_2(\mathbb{R}^d)}$ *for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$ *and, by extension, for any* $\varphi \in L_2(\mathbb{R}^d)$.

Definition 22. *A generalized stochastic process* $G : \varphi \mapsto \langle G, \varphi\rangle$ *is*

- *Infinite divisible: if the random variable* $X = \langle G, \varphi\rangle$ *is infinitely divisible for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$.
- *Second-order: if* $\mathbb{E}\{|\langle G, \varphi\rangle|^2\} < \infty$ *for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$.
- *Stationary: if, for any given* $\varphi \in \mathcal{S}(\mathbb{R}^d)$*, the random variables* $X_{\boldsymbol{x}_0} = \langle G, \varphi(\cdot + \boldsymbol{x}_0)\rangle$ *are identically distributed for any* $\boldsymbol{x}_0 \in \mathbb{R}^d$.
- *Self-similar with Hurst index* H*: if, for any given* $\varphi \in \mathcal{S}(\mathbb{R}^d)$*, the random variables* $X_a = a^H\langle G, |a|^d\varphi(a\cdot)\rangle$ *are identically distributed for any contraction factor* $a \in \mathbb{R}^+$.

Definition 23 (Linear transform of a generalized stochastic process)**.** *For any given continuous linear operator* $\mathrm{T} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ *whose adjoint* T^* *is continuous* $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$*, the linear transformation of the generalized stochastic process* G *in* $\mathcal{S}'(\mathbb{R}^d)$ *is defined as*

$$\langle \mathrm{T}\{G\}, \varphi\rangle \triangleq \langle G, \mathrm{T}^*\{\varphi\}\rangle$$

for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$.

Note that the effect of such a transformation on a realization $\omega \mapsto g = G(\omega)$ translates into $\langle \mathrm{T}\{G\}(\omega), \varphi\rangle = \langle \mathrm{T}\{g\}, \varphi\rangle = \langle g, \mathrm{T}^*\{\varphi\}\rangle$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, which is compatible with the standard definition (by duality) of the linear transformation of a generalized function $g \in \mathcal{S}'(\mathbb{R}^d)$. The most important linear transformations for our purpose are:

1. Translation by $\boldsymbol{x}_0 \in \mathbb{R}^d$:
$$\langle g(\cdot - \boldsymbol{x}_0), \varphi\rangle \triangleq \langle g, \varphi(\cdot + \boldsymbol{x}_0)\rangle$$

2. Dilation (or scaling) by $a \in \mathbb{R}^+$:
$$\langle g(\cdot/a), \varphi\rangle \triangleq \langle g, |a|^d \varphi(a\cdot)\rangle$$

3. Rotation of coordinate system $\boldsymbol{x} \mapsto \mathbf{R}\boldsymbol{x}$ with $\mathbf{R}^{-1} = \mathbf{R}^T$:
$$\langle g(\mathbf{R}\cdot), \varphi\rangle \triangleq \langle g, \varphi(\mathbf{R}^{-1}\cdot)\rangle$$

4. Partial derivative operator $\partial^{\mathbf{n}}$ of multi-order $\mathbf{n} = (n_1, \dots, n_d)$:
$$\langle \partial^{\mathbf{n}} g, \varphi\rangle \triangleq \langle g, (-1)^{|\mathbf{n}|}\partial^{\mathbf{n}}\varphi\rangle$$
where $|\mathbf{n}| = n_1 + \cdots + n_d$ and
$$\partial^{\mathbf{n}}\varphi(\boldsymbol{x}) = \frac{\partial^{|\mathbf{n}|}\varphi(x_1, \dots, x_d)}{\partial x_1^{n_1} \cdots \partial x_d^{n_d}}$$

Hence another equivalent interpretation of the stationarity property in Definition 22 is that the generalized stochastic process G is undistinguishable in law from its translated version $G(\cdot - \boldsymbol{x}_0)$ for all $\boldsymbol{x}_0 \in \mathbb{R}^d$. Similarly, a self-similar process with Hurst exponent H is a process that is statistically undistinguishable from its dilated and renormalized version $a^H G(\cdot/a)$ for any $a > 0$.

4.1.3 GSPs as consistent generators of random variables

Under construction: Implicit in the definition of GSP is the mutual consistency of the underlying statistical distributions which follows from the linearity property. For instance, by considering a series of test functions $\varphi_1, \dots, \varphi_N$, we may define the random vector $\boldsymbol{X} = (X_1, \dots, X_N)$ with $X_n = \langle G, \varphi_n\rangle$.

4.1.4 Examples

The functional counterpart of the constant random variable is the (generalized) deterministic process:

$$\omega \mapsto G_{\mathrm{Const}}(\omega) = p_0$$
$$\varphi \mapsto G_{\mathrm{Const}}(\varphi) = \langle p_0, \varphi\rangle$$

where p_0 is a fixed element of $\mathcal{S}'(\mathbb{R}^d)$. A slightly more involved example is the linear process

$$\omega \mapsto G_{N_0}(\omega) = \sum_{n=1}^{N_0} A_n(\omega) p_n$$
$$\varphi \mapsto G_{N_0}(\varphi) = \langle \sum_{n=1}^{N_0} A_n p_n, \varphi\rangle = \sum_{n=1}^{N_0} A_n \langle p_n, \varphi\rangle$$

where the $A_n \sim \mathcal{N}(0, \sigma_n^2)$ are independent Gaussian random variables with zero mean and variance σ_n^2 and where the p_n are fixed elements of $\mathcal{S}'(\mathbb{R}^d)$. Clearly, this generalized process is finite-dimensional since its realizations live in $\mathrm{span}\{p_n\}_{n=1}^{N_0} \subseteq \mathcal{S}'(\mathbb{R}^d)$. Moreover, it is Gaussian with $G_{N_0}(\varphi) \sim \mathcal{N}(0, \sum_{m=1}^{N_0} \sigma_n^2 |\langle p_n, \varphi\rangle|^2)$ since it results from a linear combination of N_0 (independent) Gaussian random variables A_n.

A more ambitious construction is the standardized white Gaussian noise of Definition 21:

$$\varphi \mapsto W_{\mathrm{Gauss}}(\varphi) = \langle W_{\mathrm{Gauss}}, \varphi\rangle \sim \mathcal{N}(0, \|\varphi\|^2_{L_2(\mathbb{R}^d)}),$$

which yields a series of Gaussian random variables with zero mean and standard deviation $\sigma = \|\varphi\|_{L_2(\mathbb{R}^d)}$. This is a true infinite-dimensional object that is the natural functional extension of the normal random variable. By contrast with the first two examples, it is much less obvious there to give a functional description of the individual realizations $W_{\mathrm{Gauss}}(\omega) \in \mathcal{S}'(\mathbb{R}^d)$. In fact, these signals are too rough to have a pointwise interpretation. One should think of them as some kind of stochastic counterpart of the Dirac distribution because their average spectral density is flat.

4.2 Mean and covariance forms

We have seen that the observation $G(\varphi) = \langle G, \varphi\rangle$ of a generalized random process in $\mathcal{S}'(\mathbb{R}^d)$ is an ordinary random variable in $\mathbb{R}$. The expected value of $G(\varphi)$ is then given by

$$\mathbb{E}\{\langle G, \varphi\rangle\} = \mathbb{E}\{G(\varphi)\} = \int_{\mathbb{R}} x p_{G(\varphi)}(x)\mathrm{d}x$$

where $p_{G(\varphi)}(x)$ is the underlying pdf. The key observation, which follows from the axioms in Definition 19, is that the map $\varphi \mapsto \mathbb{E}\{\langle G, \varphi\rangle\}$ is a continuous linear functional on $\mathcal{S}(\mathbb{R}^d)$ so that there exits a unique element $\mu_G \in \mathcal{S}'(\mathbb{R}^d)$ such that

$$\varphi \mapsto \mathbb{E}\{\langle G, \varphi\rangle\} = \langle \mu_G, \varphi\rangle.$$

The generalized function μ_G is called the mean of the stochastic process G; it is also written as $\mu_G = \mathbb{E}\{G\}$. The latter may also be formally defined by the functional integral

$$\mu_G = \mathbb{E}\{G\} = \int_{\mathcal{S}'(\mathbb{R}^d)} g \mathscr{P}_G(\mathrm{d}g).$$

For instance, the mean of the deterministic process $\omega \mapsto G(\omega) = p_0$ of Subsection 4.1.4 is simply p_0, while the mean of the white Gaussian noise process W_{Gauss} is $\mathbb{E}\{W_{\mathrm{Gauss}}\} = 0$.

Instead of a single random variable $G(\varphi)$, we may also consider any pair of observations $\boldsymbol{X} = \big(G(\varphi_1), G(\varphi_2)\big)$ of the generalized process G, which admits some joint pdf $p_{\boldsymbol{X}}$ parameterized by φ_1 and φ_2. The second-order dependencies of the process are then measured by the covariance form

$$\begin{aligned} C_G(\varphi_1, \varphi_2) &= \mathbb{E}\{\langle G - \mu_G, \varphi_1\rangle\langle G - \mu_G, \varphi_2\rangle\} \\ &= \int_{\mathbb{R}^2} (x_1 - \mu_1)(x_2 - \mu_2)\, p_{\boldsymbol{X}}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \end{aligned}$$

with $\mu_1 = \langle \mu_G, \varphi_1\rangle$ and $\mu_2 = \langle \mu_G, \varphi_2\rangle$. We then rely on the axioms in Definition 19 once more to derive the functional properties of the covariance functional $C_G : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$.

Theorem 22 (Properties of the covariance form)**.** *Let G be a generalized stochastic process in $\mathcal{S}'(\mathbb{R}^d)$ with mean $\mathbb{E}\{G\} = \mu_G$ and the second-order property $\mathbb{E}\{\langle G, \varphi\rangle^2\} < \infty$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Then, its covariance form, which is defined by*

$$\begin{aligned} C_G(\varphi_1, \varphi_2) &= \mathbb{E}\{\langle G - \mu_G, \varphi_1\rangle\langle G - \mu_G, \varphi_2\rangle\} \\ &= \int_{\mathcal{S}'(\mathbb{R}^d)} \langle g - \mu_G, \varphi_1\rangle\langle g - \mu_G, \varphi_2\rangle \mathscr{P}_G(\mathrm{d}g) \end{aligned} \quad (127)$$

for any $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$, has the following properties:

- *Symmetry:* $C_G(\varphi_1,\varphi_2)=C_G(\varphi_2,\varphi_1)$.
- *Bilinearity: The functional* $C_G:(\varphi_1,\varphi_2)\mapsto C_G(\varphi_1,\varphi_2)$ *is linear in each of its arguments.*
- *Continuity:* C_G *continuously maps* $\mathcal{S}(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)\to\mathbb{R}$.
- *positive definiteness:* $C_G(\varphi_1,\varphi_1)\geq 0$.
- *Link with covariance operator: There exists a unique continuous linear operator* $\mathrm{R}_G:\mathcal{S}(\mathbb{R}^d)\to\mathcal{S}'(\mathbb{R}^d)$ *such that*

$$C_G(\varphi_1,\varphi_2)=\langle \mathrm{R}_G\{\varphi_1\},\varphi_2\rangle=\langle \mathrm{R}_G\{\varphi_2\},\varphi_1\rangle.$$

- *Kernel representation: There exists a unique symmetric kernel* $r_G\in\mathcal{S}'(\mathbb{R}^d\times\mathbb{R}^d)$ *such that*

$$C_G(\varphi_1,\varphi_2)=\langle r_G,\varphi_1\otimes\varphi_2\rangle=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} r_G(\boldsymbol{x},\boldsymbol{y})\varphi_1(\boldsymbol{x})\varphi_2(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.$$

Proof. Let us fix $\varphi_1,\varphi_2\in\mathcal{S}(\mathbb{R}^d)$ and consider the random variables $X_1=\langle G,\varphi_1\rangle$, $X_2=\langle G,\varphi_2\rangle$, as well as their linear combination $X_3=a_1X_1+X_2=\langle G,a_1\varphi_1+\varphi_2\rangle$ with $a_1\in\mathbb{R}$. The corresponding means are $\mathbb{E}\{X_1\}=\mu_1$, $\mathbb{E}\{X_2\}=\mu_2$, and $\mathbb{E}\{X_3\}=\mathbb{E}\{a_1X_1+X_2\}=a_1\mu_1+\mu_2$ with $\mu_i=\langle\mu_G,\varphi_i\rangle$ (as direct consequence of the linearity assumption in Definition 19). Thanks to the second-order condition, the covariance matrix of $\boldsymbol{X}=(X_1,X_2)$ is well-defined for any $\varphi_1,\varphi_2\in\mathcal{S}(\mathbb{R}^d)$; it is given by

$$\mathbf{C}_{\boldsymbol{X}}=\begin{pmatrix} C_G(\varphi_1,\varphi_1) & C_G(\varphi_1,\varphi_2)\\ C_G(\varphi_2,\varphi_1) & C_G(\varphi_2,\varphi_2)\end{pmatrix}$$

with $C_G(\varphi_i,\varphi_j)=\mathbb{E}\{(X_i-\mu_i)(X_j-\mu_j)\}=\mathbb{E}\{X_iX_j\}-\mu_i\mu_j=C_G(\varphi_j,\varphi_i)$. Since $\mathbf{C}_{\boldsymbol{X}}$ is symmetric and positive-definite by construction, these properties carry over to the form $C_G:\mathcal{S}(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)\to\mathbb{R}$. To establish the bilinearity property, we evaluate the covariance between X_3 and X_1 as

$$\begin{aligned}C_G(a_1\varphi_1+\varphi_2,\varphi_1)&=\mathbb{E}\{X_3X_1\}-\mu_3\mu_1\\&=\mathbb{E}\{(a_1X_1+X_2)X_1\}-(a_1\mu_1+\mu_2)\mu_1\\&=a_1\big(\mathbb{E}\{X_1X_1\}-\mu_1\mu_1\big)+\big(\mathbb{E}\{X_2X_1\}-\mu_2\mu_1\big)\\&=a_1C_G(\varphi_1,\varphi_1)+C_G(\varphi_2,\varphi_1).\end{aligned}$$

The continuity assumption in Definition 19 implies that $C_G(\cdot, \varphi_2)$ with φ_2 fixed is continuous $\mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ in its first argument, while the same obviously holds true if one switches the role of the two arguments. Because of the linearity property, the separate continuity of C_G in each of its arguments is equivalent to its joint continuity (see proof of Proposition 23 in Appendix B for a more detailed explanation). Ultimately, this translates into the existence of a directed Schwartz norm $\|\cdot\|_m$ and a constant $A > 0$ such that

$$C_G(\varphi_1, \varphi_2) \leq A\|\varphi_1\|_m \|\varphi_2\|_m \tag{128}$$

for all $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$—as implied by the boundedness condition in Schwartz' kernel theorem (Theorem 37) with $m = \min(m_0, n_0)$. The same theorem also ensures the existence of a unique kernel $r_G \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ and a corresponding linear operator R_G such that the remaining equivalences are met. □

The linear map $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ in Proposition 22 is the so-called *covariance operator* whose equivalent "integral" representation is

$$\mathrm{R}_G\{\varphi\}(\boldsymbol{x}) = \langle r_G(\boldsymbol{x}, \cdot), \varphi\rangle = \int_{\mathbb{R}^d} r_G(\boldsymbol{x}, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}. \tag{129}$$

Its kernel $r_G(\boldsymbol{x}, \boldsymbol{y})$, which is formally identified as

$$r_G(\boldsymbol{x}, \boldsymbol{y}) = C_G\big(\delta(\cdot - \boldsymbol{x}), \delta(\cdot - \boldsymbol{y})\big) = \mathrm{R}_G\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x}),$$

is the *covariance function* of G. Since $\langle \mathrm{R}_G\{\varphi\}, \varphi\rangle = C_G\big(\varphi, \varphi\big) \geq 0$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, both entities are symmetric, positive-definite by construction (see Definition 30 in Appendix A).

Three further remarks are in order. First, the knowledge of the map $\varphi \mapsto \mathrm{Var}\{\langle G, \varphi\rangle\} = C_G(\varphi, \varphi)$ that returns the variance of the scalar variable $\langle G, \varphi\rangle$ is sufficient to determine C_G uniquely. Indeed, we have that

$$C_G(\varphi_1, \varphi_2) = \frac{1}{4}\big(\mathrm{Var}\{\langle G, \varphi_1 + \varphi_2\rangle\} - \mathrm{Var}\{\langle G, \varphi_1 - \varphi_2\rangle\big), \tag{130}$$

as a consequence of (bi)linearity. Another remarkable implication of the latter property is that the continuity of C_G at $(0, 0)$ $\big($which is equivalent to the boundedness of C_G in (128)$\big)$ is sufficient to ensure its continuity over $\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d)$. Finally, the covariance form C_G satisfies the Cauchy-Schwarz-like inequality

$$|C_G(\varphi_1, \varphi_2)| \leq \sqrt{C_G(\varphi_1, \varphi_1)}\sqrt{C_G(\varphi_2, \varphi_2)} < \infty, \tag{131}$$

which is equivalent to the positive definiteness of the matrix $\mathbf{C}_{\boldsymbol{X}}$ in the proof of Proposition 22.

4.2.1 Reproducing kernels and mean-square continuity

A classical stochastic process on $\mathbb{R}^d$—that is, an indexed collection of random variables $\{G(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$—can be viewed as a special instance of a generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ for which the space of test functions has been extended to include the sampling functionals $\delta(\cdot - \boldsymbol{x}_0)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$. In other words, the generalized process must be such that its sample values

$$G(\boldsymbol{x}) \triangleq \langle G, \delta(\cdot - \boldsymbol{x})\rangle$$

are well-defined random variables for all $\boldsymbol{x} \in \mathbb{R}^d$. In particular, this implies that the mean of the process is an ordinary function of $\boldsymbol{x}$

$$\mathbb{E}\{G(\boldsymbol{x})\} = \langle \mathbb{E}\{G\}, \delta(\cdot - \boldsymbol{x})\rangle = \mu_G(\boldsymbol{x}).$$

Likewise, under the second-order (or finite variance) hypothesis, the covariance function of G is an "ordinary" bivariate function, which is given by

$$r_G(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}\{\big(G(\boldsymbol{x}) - \mu_G(\boldsymbol{x})\big(G(\boldsymbol{y}) - \mu_G(\boldsymbol{y})\big)\}, \tag{132}$$

an expression that is compatible with the more general definition of the kernel r_G of the covariance form C_G in Theorem 22. We also note that the covariance function $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is symmetric and positive-definite; in particular, it satisfies

$$|r_G(\boldsymbol{x}, \boldsymbol{y})| \le \sqrt{r_G(\boldsymbol{x}, \boldsymbol{x})}\sqrt{r_G(\boldsymbol{y}, \boldsymbol{y})}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ (see Proposition 21 in Appendix A), which is the kernel counterpart of (131).

In addition to their variance being finite, a minimum requirement for classical stochastic processes is that they be continuous in the mean-square sense.

Definition 24 (Mean-square continuity)**.** *A real-valued stochastic process* $\{G(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$ *is said to be mean-square continuous at* $\boldsymbol{x}_0 \in \mathbb{R}^d$ *if* $\mathbb{E}\{[G(\boldsymbol{x}_0)]^2\} < \infty$ *and*

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \mathbb{E}\{\big[G(\boldsymbol{x}) - G(\boldsymbol{x}_0)\big]^2\} = 0.$$

Let us assume, for simplicity, that the process is centered; i.e., $\mathbb{E}\{G(\boldsymbol{x})\} = \mu_G(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{R}^d$. We can then expand the above expectation as

$$\begin{aligned}\mathbb{E}\{\big[G(\boldsymbol{x}) - G(\boldsymbol{x}_0)\big]^2\} &= \mathbb{E}\{G(\boldsymbol{x})^2 + G(\boldsymbol{x}_0)^2 - 2G(\boldsymbol{x})G(\boldsymbol{x}_0)\}\\ &= r_G(\boldsymbol{x}, \boldsymbol{x}) + r_G(\boldsymbol{x}_0, \boldsymbol{x}_0) - 2r_G(\boldsymbol{x}, \boldsymbol{x}_0),\end{aligned}$$

which shows that G is mean-square continuous if and only if its covariance function $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is (jointly) continuous at the point $(\boldsymbol{x}_0, \boldsymbol{x}_0)$. For a non-centered process, we obviously also require the mean function $\mu_G : \mathbb{R}^d \to \mathbb{R}$ to be continuous at $\boldsymbol{x}_0$.

Thus, to ensure the mean-square continuity of G over $\mathbb{R}^d$, we need the continuity of $(\boldsymbol{x}, \boldsymbol{y}) \mapsto r_G(\boldsymbol{x}, \boldsymbol{y})$ over the diagonal of its domain—i.e., for any $\boldsymbol{x} = \boldsymbol{y} \in \mathbb{R}^d$—which, somewhat remarkably, is equivalent to its continuity over $\mathbb{R}^d \times \mathbb{R}^d$.

Theorem 23 (Mean-square continuity of stochastic processes)**.** *A second-order stochastic process G over $\mathbb{R}^d$ ismean-square continuous if and only if its mean and covariance functions, μ_G and r_G, are continuous over $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^d$, respectively. This also implies that $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a valid reproducing kernel.*

Proof. We shall assume that $G(\boldsymbol{x})$ is zero-mean; otherwise, we recenter the process by taking $G(\boldsymbol{x}) - \mu_G(\boldsymbol{x})$. In light of the above discussion, the only delicate issue is to prove that the continuity of r_G over the diagonal implies continuity everywhere.

Using the property that the covariance matrix of $\boldsymbol{X} = (X_1, X_2)$ with $X_1 = G(\boldsymbol{x})$ and $X_2 = G(\boldsymbol{y}) - G(\boldsymbol{y}_0)$ is positive-definite, we first show that

$$\begin{aligned}|r_G(\boldsymbol{x}, \boldsymbol{y}) - r_G(\boldsymbol{x}, \boldsymbol{y}_0)|^2 &= |\mathbb{E}\{X_1 X_2\}|^2 \\ &\le \mathbb{E}\{X_1^2\}\,\mathbb{E}\{X_2^2\} \qquad \text{(by positive definiteness)} \\ &= r_G(\boldsymbol{x}, \boldsymbol{x})\big[r_G(\boldsymbol{y}, \boldsymbol{y}) + r_G(\boldsymbol{y}_0, \boldsymbol{y}_0) - 2r_G(\boldsymbol{y}, \boldsymbol{y}_0)\big].\end{aligned}$$

The continuity of r_G over the diagonal implies that both $r_G(\boldsymbol{y}, \boldsymbol{y})$ and $r_G(\boldsymbol{y}, \boldsymbol{y}_0)$ tend to $r_G(\boldsymbol{y}_0, \boldsymbol{y}_0)$ as $\boldsymbol{y} \to \boldsymbol{y}_0$, so that $\lim_{\boldsymbol{y}\to\boldsymbol{y}_0} |r_G(\boldsymbol{x}, \boldsymbol{y}) - r_G(\boldsymbol{x}, \boldsymbol{y}_0)| = 0$. Since r_G is symmetric, the latter is equivalent to its separate continuity in each argument. By applying the same method once more, we find that

$$\begin{aligned}|r_G(\boldsymbol{x}, \boldsymbol{y}) - r_G(\boldsymbol{x}_0, \boldsymbol{y}_0)| &\le |r_G(\boldsymbol{x}, \boldsymbol{y}) - r_G(\boldsymbol{x}, \boldsymbol{y}_0)| + |r_G(\boldsymbol{x}, \boldsymbol{y}_0) - r_G(\boldsymbol{x}_0, \boldsymbol{y}_0)| \\ &\le \sqrt{r_G(\boldsymbol{x}, \boldsymbol{x})}\big[r_G(\boldsymbol{y}, \boldsymbol{y}) + r_G(\boldsymbol{y}_0, \boldsymbol{y}_0) - 2r_G(\boldsymbol{y}, \boldsymbol{y}_0)\big]^{\frac{1}{2}} \\ &\quad + \sqrt{r_G(\boldsymbol{y}_0, \boldsymbol{y}_0)}\big[r_G(\boldsymbol{x}, \boldsymbol{x}) + r_G(\boldsymbol{x}_0, \boldsymbol{x}_0) - 2r_G(\boldsymbol{x}, \boldsymbol{x}_0)\big]^{\frac{1}{2}},\end{aligned}$$

which implies the continuity of r_G over $\mathbb{R}^d \times \mathbb{R}^d$. The final point is that the continuity and the positive definiteness of r_G are sufficient for the reproducing kernel property (see Theorem 6). $\square$

Note that the joint continuity assumption of Theorem 23 together with the requirement that $r_G \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ in Theorem 22 is equivalent to the existence of some $\alpha \in \mathbb{R}$ such that $r_G \in C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$. It then follows from Theorem 6 that the covariance function $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the reproducing kernel of a unique RHKS $\mathcal{H} \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$. In such a scenario, the covariance operator R_G is the Riesz map $\mathcal{H}' \to \mathcal{H}$, which has some pleasing consequences for the theory of Gaussian processes and the link with variational splines.

To conclude this discussion of continuity properties, we remark that the mean-square continuity at $\boldsymbol{x}_0$ implies the continuity of G in probability. Indeed, we have that, for any $\epsilon > 0$,

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \mathrm{Prob}\left(|G(\boldsymbol{x}) - G(\boldsymbol{x}_0)| > \epsilon\right) = \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{\mathbb{E}\{\big(G(\boldsymbol{x}+\boldsymbol{\epsilon}) - G(\boldsymbol{x})\big)^2\}}{\epsilon^2} = 0,$$

which follows directly from Tchebyshev's inequality.

4.2.2 Effect of a linear transformation

The mean μ_G and the covariance form $C_G(\varphi_1, \varphi_2)$ provide a complete characterization of the first and second-order moments of the generalized process G. It is not difficult to infer the effect of a linear transformation $\mathrm{T} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ on those quantities. Specifically, we have that

$$\mathbb{E}\{\langle \mathrm{T}\{G\}, \varphi\rangle\} = \langle \mu_G, \mathrm{T}^*\{\varphi\}\rangle = \langle \mathrm{T}\{\mu_G\}, \varphi\rangle$$

which shows that $\mu_{\mathrm{T}G} = \mathrm{T}\mu_G$. Likewise,

$$C_{\mathrm{T}G}(\varphi_1, \varphi_2) = C_G(\mathrm{T}^*\varphi_1, \mathrm{T}^*\varphi_2) = \langle \varphi_1, \mathrm{T}\mathrm{R}_G\mathrm{T}^*\varphi_2\rangle$$

so that $r_{\mathrm{T}G}(\boldsymbol{x}, \boldsymbol{y}) = C_G(\mathrm{T}^*\{\delta(\cdot - \boldsymbol{x})\}, \mathrm{T}^*\{\delta(\cdot - \boldsymbol{y})\})$ and $\mathrm{R}_{\mathrm{T}G} = \mathrm{T}\mathrm{R}_G\mathrm{T}^*$, so that there is a direct parallel with the properties of the covariance matrix in finite-dimensional statistics.

4.2.3 Stationary processes

A case of special interest is when the generalized process G is centered (zero-mean) and stationary. The direct implication of the stationarity property in Definition 22 is that

$$\begin{aligned} r_G(\boldsymbol{x}, \boldsymbol{y}) &= C_G\big(\delta(\cdot - \boldsymbol{x}), \delta(\cdot - \boldsymbol{y})\big) \\ &= C_G\big(\delta, \delta(\cdot - (\boldsymbol{y} - \boldsymbol{x})\big) = r_G\big(\boldsymbol{0}, (\boldsymbol{y} - \boldsymbol{x})\big) = a_G(\boldsymbol{y} - \boldsymbol{x}) \end{aligned}$$

so that the correlation function only depends on the relative displacement $(\boldsymbol{y}-\boldsymbol{x})$. The univariate function $a_G : \mathbb{R}^d \to \mathbb{R}$ is the *autocorrelation function* of the stationary process G; its classical definition is

$$a_G(\boldsymbol{\tau}) \stackrel{\triangle}{=} \mathbb{E}\{G(\boldsymbol{x})G(\boldsymbol{x}+\boldsymbol{\tau})\}.$$

4.3 The characteristic functional

We have seen that a generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ is completely characterized by its probability measure $\mathscr{P}_G$. An alternative representation is provided by its characteristic functional $\widehat{\mathscr{P}}_G$, which can be interpreted as the infinite-dimensional Fourier transform of this probability measure.

Definition 25 (Characteristic functional)**.** *The characteristic functional* $\widehat{\mathscr{P}}_G : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ *of the generalized stochastic process* G *in* $\mathcal{S}'(\mathbb{R}^d)$ *is given by*

$$\widehat{\mathscr{P}}_G(\varphi) \stackrel{\triangle}{=} \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\varphi\rangle}\} = \int_{\mathcal{S}'(\mathbb{R}^d)} \mathrm{e}^{\mathrm{j}\langle g,\varphi\rangle}\, \mathscr{P}_G(\mathrm{d}g)$$

where the right-hand side is an abstract Lebesgue integral over the function space $\mathcal{S}'(\mathbb{R}^d)$.

Since G is an infinite-dimensional entity, $\widehat{\mathscr{P}}_G$ is indexed by a function φ rather than the usual scalar or vectorial Fourier-domain variable $\boldsymbol{\xi}$.

To demystify the concept, we shall now determine the characteristic functional of the stochastic processes of Section 4.1.4. The simplest instance is

$$\widehat{\mathscr{P}}_{G_{\mathrm{Const}}}(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G_{\mathrm{Const}},\varphi\rangle}\} = \mathrm{e}^{\mathrm{j}\langle p_0,\varphi\rangle}, \tag{133}$$

which characterizes the deterministic process $G_{\mathrm{Const}} = p_0$. As for the finite-dimensional Gaussian process $G_{N_0} = \sum_{n=1}^{N_0} A_n p_n$, we first write the characteristic function of the Gaussian random variable $Y = \langle G_{N_0}, \varphi\rangle \sim \mathcal{N}(0, \sigma_Y^2)$ with $\sigma_Y^2 = \sum_{n=1}^{N_0} \sigma_n^2 |\langle p_n, \varphi\rangle|^2$ as

$$\hat{p}_Y(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\xi Y}\} = \exp\left(-\tfrac{1}{2}\xi^2\sigma_Y^2\right)$$

Next, we observe that $\hat{p}_Y(\xi)|_{\xi=1} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}Y}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G_{N_0},\varphi\rangle}\} = \widehat{\mathscr{P}}_{G_{N_0}}(\varphi)$, which yields

$$\widehat{\mathscr{P}}_{G_{N_0}}(\varphi) = \exp\left(-\tfrac{1}{2}\sum_{n=1}^{N_0} \sigma_n^2 |\langle p_n, \varphi\rangle|^2\right).$$

We use the same identification technique to obtain the characteristic functional of the Gaussian white noise process

$$\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}(\varphi) = \mathrm{e}^{-\frac{1}{2}\|\varphi\|^2_{L_2(\mathbb{R}^d)}}. \tag{134}$$

We observe that the latter is the infinite-dimensional counterpart of $\hat{p}_{\mathrm{Gauss}}(\boldsymbol{\xi}) = \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{\xi}\|_2^2}$: the characteristic function of the N-dimensional multivariate normal distribution $p_{\mathrm{Gauss}}(\boldsymbol{x}) = (2\pi)^{-N/2}\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{x}\|_2^2}$ (see derivation of (162) in Appendix E.3).

A classical result of probability theory is that all characteristic functions are continuous, positive-definite and normalized (see Theorems 38 and 39 in Appendix E). We shall now see that the same holds true in infinite-dimensions, the fundamental difference being that one is dealing with functionals rather than functions.

Definition 26 (Continuous functional)**.** *A functional $F : \mathcal{X} \to \mathbb{C}$ is said to be continuous (with respect to the topology of the function space $\mathcal{X}$) if, for any convergent sequence (φ_n) in $\mathcal{X}$ with limit $\varphi \in \mathcal{X}$, the sequence $F(\varphi_n)$ converges to $F(\varphi)$; that is,*

$$\lim_n F(\varphi_n) = F(\lim_n \varphi_n).$$

Definition 27 (Positive-definite functional)**.** *A complex-valued functional $F : \mathcal{X} \to \mathbb{C}$ defined over the function space $\mathcal{X}$ is said to be* positive-definite *if*

$$\sum_{m=1}^{N}\sum_{n=1}^{N} z_m F(\varphi_m - \varphi_n)\overline{z}_n \geq 0 \tag{135}$$

for every possible choice of $\varphi_1, \dots, \varphi_N \in \mathcal{X}$, $z_1, \dots, z_N \in \mathbb{C}$, and $N \in \mathbb{N}^+$. Likewise, it is said to be conditionally positive-definite *if* (135) *holds subject to the constraint $\sum_{n=1}^{N} z_n = 0$.*

Example 5. *Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Then, $F(\varphi) = \mathrm{e}^{-\frac{1}{2}\|\varphi\|^2_{\mathcal{H}}}$ is positive-definite over $\mathcal{H}$, while $G(\varphi) = \log F(\varphi) = -\frac{1}{2}\|\varphi\|^2_{\mathcal{H}}$ is conditionally positive-definite. We readily verify the last state-*

ment by noting that

$$-\tfrac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N} z_m\overline{z}_n\|\varphi_m-\varphi_n\|_{\mathcal{H}}^2 =$$

$$= -\tfrac{1}{2}\underbrace{\sum_{n=1}^{N}\overline{z}_n}_{=0}\sum_{m=1}^{N} z_m\|\varphi_m\|_{\mathcal{H}}^2 - \tfrac{1}{2}\underbrace{\sum_{m=1}^{N} z_m}_{=0}\sum_{n=1}^{N}\overline{z}_n\|\varphi_n\|_{\mathcal{H}}^2 + \sum_{m=1}^{N}\sum_{m=1}^{N} z_m\overline{z}_n\langle\varphi_m,\varphi_n\rangle_{\mathcal{H}}$$

$$= \sum_{m=1}^{N}\sum_{n=1}^{N} z_m\overline{z}_n\langle\varphi_m,\varphi_n\rangle_{\mathcal{H}} = \Big\|\sum_{n=1}^{N} z_n\varphi_n\Big\|_{\mathcal{H}}^2 \geq 0,$$

for any $\varphi_1,\dots,\varphi_N \in \mathcal{H}$ and $z_1,\dots,z_N \in \mathbb{C}$ such that $\sum_{n=1}^N z_n = 0$. Unfortunately, establishing the positive definiteness of $F(\varphi) = \mathrm{e}^{G(\varphi)}$ is not as easy. We bypass the difficulty by invoking Schoenberg's correspondence principle (Theorem 27).

We now proceed with the derivation the key properties of the characteristic functional, in direct analogy with Theorem 38 of Appendix E, which lists the corresponding properties of the characteristic function of a finite-dimensional random variable $\boldsymbol{X}$ in $\mathbb{R}^N$.

Theorem 24. *The characteristic functional $\widehat{\mathscr{P}}_G : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ of a generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ enjoys the following properties:*

1. *$\widehat{\mathscr{P}}_G$ is continuous, bounded (i.e. $|\widehat{\mathscr{P}}_G(\varphi)| \leq 1$), Hermitian-symmetric (i.e., $\widehat{\mathscr{P}}_G(-\varphi) = \overline{\widehat{\mathscr{P}}_G(\varphi)}$) and normalized such that $\widehat{\mathscr{P}}_G(0) = 1$.*

2. *$\widehat{\mathscr{P}}_G$ is positive-definite in the sense of Definition 27.*

3. *Connection with joint pdf: Let $\varphi_1,\dots,\varphi_N \in \mathcal{S}(\mathbb{R}^d)$ be any fixed collection of test functions. Then, the joint pdf of the random vector $\boldsymbol{G} = (\langle G,\varphi_1\rangle,\dots,\langle G,\varphi_N\rangle)$ is given by the following finite-dimensional inverse Fourier transform*

$$p_{\boldsymbol{G}}(\boldsymbol{x}) = \int_{\mathbb{R}^N}\widehat{\mathscr{P}}_G(\xi_1\varphi_1+\cdots+\xi_N\varphi_N)\mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\xi},\boldsymbol{x}\rangle}\frac{\mathrm{d}\boldsymbol{\xi}}{(2\pi)^N}.$$

   with Fourier-domain variable $\boldsymbol{\xi} = (\xi_1,\cdots,\xi_N)$.

4. *Linear transformation: Let T be a continuous linear operator $\mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ and $\mu_0 \in \mathcal{S}'(\mathbb{R}^d)$ some constant generalized function. Then, the characteristic functional of the transformed process $Q = \mathrm{T}\{G\} + \mu_0$ is*

$$\widehat{\mathscr{P}}_Q(\varphi) = \widehat{\mathscr{P}}_{\mathrm{T}\{G\}+\mu_0}(\varphi) = \widehat{\mathscr{P}}_G(\mathrm{T}^*\varphi)\mathrm{e}^{\mathrm{j}\langle\mu_0,\varphi\rangle}$$

where $\mathrm{T}^* : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ *is the (continuous) adjoint of* T.

5. *Sum of independent stochastic processes: Let* G_1 *and* G_2 *be two independent generalized stochastic processes with characteristic functionals* $\widehat{\mathscr{P}}_{G_1}$ *and* $\widehat{\mathscr{P}}_{G_2}$, *respectively. Then, the characteristic functional of* $G = G_1 + G_2$ *is*

$$\widehat{\mathscr{P}}_{G_1+G_2}(\varphi) = \widehat{\mathscr{P}}_{G_1}(\varphi)\widehat{\mathscr{P}}_{G_2}(\varphi).$$

Let $\widehat{\mathscr{P}}_G(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\varphi\rangle}\}$ *be the characteristic functional of the generalized stochastic process* G *in* $\mathcal{S}'(\mathbb{R}^d)$. *Then,* G *is*

1. stationary *iff.* $\widehat{\mathscr{P}}_G(\varphi) = \widehat{\mathscr{P}}_G\big(\varphi(\cdot + \boldsymbol{x}_0)\big)$ *for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$ *and any* $\boldsymbol{x}_0 \in \mathbb{R}$.
2. self-similar *with Hurst exponent* H *iff.* $\widehat{\mathscr{P}}_G(\varphi) = \widehat{\mathscr{P}}_G\big(a^{H+d}\varphi(a\cdot)\big)$ *for any* $\varphi \in \mathcal{S}(\mathbb{R}^d)$ *and any contraction factor* $a \in \mathbb{R}^+$.

Proof. The boundedness, Hermitian symmetry and normalization properties are readily deduced from the measure-theoretic definition of the characteristic functional; i.e.,

$$\begin{aligned}\left|\widehat{\mathscr{P}}_G(\varphi)\right| &= \left|\int_{\mathcal{S}'(\mathbb{R}^d)} \mathrm{e}^{\mathrm{j}\langle g,\varphi\rangle}\,\mathscr{P}_G(\mathrm{d}g)\right| \\ &\le \int_{\mathcal{S}'(\mathbb{R}^d)} \mathscr{P}_G(\mathrm{d}g) = \widehat{\mathscr{P}}_G(0) = \mathscr{P}_G\big(\mathcal{S}'(\mathbb{R}^d)\big) = 1.\end{aligned}$$

This representation also provides the basis for establishing the positive definiteness by performing the same manipulation as in the finite-dimensional case (see Definition 42 in Appendix E and subsequent derivation).

To derive the linear transformation property, it is convenient to switch back to the expectation notation, which yields

$$\begin{aligned}\widehat{\mathscr{P}}_{\mathrm{T}\{G\}+\mu_0}(\varphi) &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \mathrm{T}\{G\}+\mu_0,\varphi\rangle}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \mathrm{T}\{G\},\varphi\rangle}\mathrm{e}^{\mathrm{j}\langle \mu_0,\varphi\rangle}\} \\ &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\mathrm{T}^*\{\varphi\}\rangle}\}\,\mathrm{e}^{\mathrm{j}\langle \mu_0,\varphi\rangle} \\ &= \widehat{\mathscr{P}}_G(\mathrm{T}^*\{\varphi\})\,\mathrm{e}^{\mathrm{j}\langle \mu_0,\varphi\rangle}.\end{aligned}$$

Let $X = \langle G, \varphi_0\rangle$ where $\varphi_0 \in \mathcal{S}(\mathbb{R}^d)$ is fixed. Then, the characteristic function of X is given by

$$\hat{p}_X(\omega) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega\langle G,\varphi_0\rangle}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\omega\varphi_0\rangle}\} = \widehat{\mathscr{P}}_G(\omega\varphi_0)$$

where we have simply used the linearity of the duality product and applied the definition of the characteristic functional with $\varphi = \omega\varphi_0$.

Now, the stationary property in Definition 22 yields $\widehat{\mathscr{P}}_G(\omega\varphi_0) = \widehat{\mathscr{P}}_G\big(\omega\varphi_0(\cdot - \boldsymbol{x}_0)\big)$ for any $\varphi_0 \in \mathcal{S}(\mathbb{R}^d)$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$. The claim in Item 1 then follows by taking $\omega = 1$ and $\varphi_0 = \varphi$. The argument also carries over for the self-similarity property (Item 2).

Similarly, let $X_1 = \langle G_1, \varphi_0\rangle$ and $X_2 = \langle G_2, \varphi_0\rangle$ whose characteristic functions are given by $\hat{p}_{X_1}(\omega) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X_1}\} = \widehat{\mathscr{P}}_{G_1}(\omega\varphi_0)$ and $\hat{p}_{X_2}(\omega) = \widehat{\mathscr{P}}_{G_2}(\omega\varphi_0)$, respectively. By Definition 20, the independence of the generalized processes G_1 and G_1 is equivalent to the independence of the random variables X_1 and X_2 so that the characteristic function of X_1+X_2 (see Property 5 in Theorem 38) is equal to the product of the individual functions

$$\hat{p}_{X_1+X_1}(\omega) = \hat{p}_{X_1}(\omega)\hat{p}_{X_2}(\omega) = \widehat{\mathscr{P}}_{G_1}(\omega\varphi_0)\widehat{\mathscr{P}}_{G_2}(\omega\varphi_0)$$

Moreover, we have that $X_1 + X_2 = \langle G_1 + G_2, \varphi_0\rangle$ (by the linearity property of generalized stochastic processes) so that

$$\widehat{\mathscr{P}}_{G_1+G_2}(\omega\varphi_0) = \hat{p}_{X_1+X_1}(\omega) = \widehat{\mathscr{P}}_{G_1}(\omega\varphi_0)\widehat{\mathscr{P}}_{G_2}(\omega\varphi_0),$$

which yields the desired factorization by setting $\varphi = \varphi_0$ and $\omega = 1$.

Finally, we show that the continuity of $\widehat{\mathscr{P}}_G : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ follows from the continuity requirement in Definition 19. To that end, we consider a series of random variables $X_n = \langle G, \varphi_n\rangle$ where (φ_n) is a converging sequence in $\mathcal{S}(\mathbb{R}^d)$. The convergence of the X_n in law is equivalent to $\lim_{n\to\infty} \hat{p}_{X_n}(\omega) = \lim_{n\to\infty} \widehat{\mathscr{P}}_{G_2}(\omega\varphi_n) = \widehat{\mathscr{P}}_G(\omega \lim_{n\to\infty} \varphi_n)$ for any $\omega \in \mathbb{R}$ (by Lévy's continuity theorem); in particular, for $\omega = 1$, which proves the claim.

□

We recall that the continuity, positive definiteness and normalization of the characteristic function are central to the finite-dimensional theory of probability because the implication also goes the other way around; i.e., if a function $\hat{p} : \mathbb{R}^N \to \mathbb{C}$ displays these three properties, then its inverse Fourier transform $p = \mathcal{F}^{-1}\{\hat{p}\}$ is guaranteed to be a valid probability density function (or probability measure) over $\mathbb{R}^N$, as stated in Theorem 39. It turns out that Bochner's theorem can be extended to the functional setting $F : \mathcal{X} \to \mathbb{C}$, but only if the space $\mathcal{X}$ is nuclear, which rules out[5] all infinite-dimensional Hilbert or Banach spaces. This is the fundamental reason why

[5]The classical counterexample is the functional $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}(\varphi) = \exp(-\frac{1}{2}\|\varphi\|^2_{L_2(\mathbb{R}^d)})$ whose corresponding Gaussian white noise measure is well defined over the (large) space of tempered distribution $\mathcal{S}'(\mathbb{R}^d)$, which is the dual of the nuclear space $\mathcal{X} = \mathcal{S}(\mathbb{R}^d)$. While

we are constraining the domain of the characteristic functional in Definition 25 to the nuclear space $\mathcal{X} = \mathcal{S}(\mathbb{R}^d)$, which corresponds to the specification of a measure on its continuous dual: the space of tempered distribution $\mathcal{S}'(\mathbb{R}^d)$. The foundational result that supports this characterization is the extended version of Bochner's theorem for generalized functions.

Theorem 25 (Minlos-Bochner for generalized stochastic processes)**.** *A functional $\widehat{\mathscr{P}}_G : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ is the characteristic functional of some generalized stochastic G in $\mathcal{S}'(\mathbb{R}^d)$ if and only if it is positive-definite, continuous $\mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ and normalized such that $\widehat{\mathscr{P}}_G(0) = 1$. This is equivalent to the existence of a unique probability measure $\mathscr{P}_G$ on $\mathcal{S}'(\mathbb{R}^d)$, such that*

$$\widehat{\mathscr{P}}_G(\varphi) = \int_{\mathcal{S}'(\mathbb{R}^d)} \mathrm{e}^{-\mathrm{j}\langle g,\varphi\rangle} \mathscr{P}_G(\mathrm{d}g) = \mathbb{E}\{\mathrm{e}^{-\mathrm{j}\langle G,\varphi\rangle}\}.$$

It often of interest to observe generalized stochastic processes with test functions that not in $\mathcal{S}(\mathbb{R}^d)$—the most notable example being $\langle G, \delta(\cdot - \boldsymbol{x}_0)\rangle$, which returns the "sampled" value $G(\boldsymbol{x}_0)$. This is feasible provided that the domain of the characteristic functional is extendable to some appropriate function space $\mathcal{X} \subseteq \mathcal{S}'(\mathbb{R}^d)$. The next result shows that the crucial ingredient for this extension is the continuity of the map $\widehat{\mathscr{P}}_G : \mathcal{X} \to \mathbb{C}$.

Theorem 26 (Extension of the domain)**.** *Let $\widehat{\mathscr{P}}_G$ be a valid characteristic functional whose domain of continuity is extendable to some topological vector space $\mathcal{X}$ with the property that $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{X} \subseteq \mathcal{S}'(\mathbb{R}^d)$. Then, the extended functional $\widehat{\mathscr{P}}_G : \mathcal{X} \to \mathbb{C}$ is continuous, positive-definite and normalized, which implies that the random variable $G(\phi) = \langle G, \phi\rangle$ is well-defined for any $\phi \in \mathcal{X}$.*

Proof. The key for this transfer of positive definiteness is that $\mathcal{S}(\mathbb{R}^d)$ is dense in $\mathcal{X}$, a property that is inherited from the denseness of $\mathcal{S}(\mathbb{R}^d)$ in $\mathcal{S}'(\mathbb{R}^d)$. Specifically, any $\phi_m \in \mathcal{X}$ can be approached as closely as desired by the sequence of functions $\varphi_{m,k} = (\phi_m * \hat{u}_k)u_k \in \mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{X}$ where (u_k) is a series of window functions in $\mathcal{S}(\mathbb{R}^d)$ such that $\lim_{k\to\infty} u_k = 1$ (e.g., $u_k(\boldsymbol{x}) = \mathrm{e}^{-(\boldsymbol{x}/k)^2}$). Then, thanks to the continuity of $\widehat{\mathscr{P}}_G$ on $\mathcal{X}$ and the property that $\lim_{k\to\infty} \varphi_{m,k} = \phi_m$, we have that

$$\sum_{m=1}^{N}\sum_{n=1}^{N} z_m \widehat{\mathscr{P}}_G(\phi_m - \phi_n)\overline{z}_n = \lim_{k\to\infty} \sum_{m=1}^{N}\sum_{n=1}^{N} z_m \widehat{\mathscr{P}}_G(\varphi_{m,k} - \varphi_{n,k})\overline{z}_n \geq 0$$

we shall also see that $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}$ is continuous, positive-definite over $L_2(\mathbb{R}^d)$, one has the less intuitive property that $\mathscr{P}_{W_{\mathrm{Gauss}}}(L_2(\mathbb{R}^d)) = 0$, which reflects the fact that none of its realization $W_{\mathrm{Gauss}}(\omega)$ has a finite energy.

for any $\phi_1, \dots, \phi_N \in \mathcal{X}$ and $z_1, \dots, z_N \in \mathbb{C}$, which proves that $\widehat{\mathscr{P}}_G$ is positive definite.

Next, let $X = \langle G, \phi_0 \rangle$ with $\phi_0 \in \mathcal{X}$ fixed. If we now consider the restricted family of test functions $\phi = \xi\phi_0 \in \mathcal{X}$ with ξ ranging over $\mathbb{R}$, we can specify the map

$$\xi \mapsto \widehat{\mathscr{P}}_G(\xi\phi_0) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\xi\langle G,\phi_0\rangle}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\xi X}\} = \hat{p}_X(\xi),$$

which is continuous, positive-definite over $\mathbb{R}$ and such that $\hat{p}_X(0) = \widehat{\mathscr{P}}_G(0) = 1$, by construction. Hence, it is a valid characteristic function (by Bochner's theorem), which proves that X is a well-defined random variable. □

A very useful application of Theorem 26 is the extension of Property 4 in Theorem 38 to the much broader class of linear operators T whose adjoint T^* continuously maps $\mathcal{S}(\mathbb{R}^d) \to \mathcal{X} \subseteq \mathcal{S}'(\mathbb{R}^d)$. Indeed, the extended continuity of $\widehat{\mathscr{P}}_G : \mathcal{X} \to \mathbb{C}$ implies the continuity and positive definiteness of $\varphi \mapsto \widehat{\mathscr{P}}_G(\mathrm{T}^*\varphi)$ over $\mathcal{S}(\mathbb{R}^d)$. We then invoke Theorem 25, which ensures that $\mathrm{T}\{G\}$ is a well-defined generalized stochastic process in $\mathcal{S}'(\mathbb{R}^d)$.

4.4 Characterization of Gaussian processes

We shall now rely on the proposed functional framework to derive the key properties of generalized Gaussian processes, starting from their axiomatic Definition 21.

Proposition 17 (Gaussian white noise)**.** *The generalized Gaussian innovation process W_{Gauss} in $\mathcal{S}'(\mathbb{R}^d)$ has the following characteristics:*

- $\mathbb{E}\{W_{\mathrm{Gauss}}\} = 0 \quad \Leftrightarrow \quad \mathbb{E}\{\langle W_{\mathrm{Gauss}}, \varphi\rangle\} = 0$ *for any* $\varphi \in L_2(\mathbb{R}^d)$.
- $\mathbb{E}\{\langle W_{\mathrm{Gauss}}, \varphi_1\rangle\langle W_{\mathrm{Gauss}}, \varphi_2\rangle\} = \langle \varphi_1, \varphi_2\rangle_{L_2}$ *for any* $\varphi_1, \varphi_2 \in L_2(\mathbb{R}^d)$*, or, equivalently,* $\mathrm{R}_{W_{\mathrm{Gauss}}} = \mathrm{I}$ *(identity operator).*
- $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle W_{\mathrm{Gauss}},\varphi\rangle}\} = \exp\left(-\frac{1}{2}\|\varphi\|^2_{L_2(\mathbb{R}^d)}\right)$ *for all* $\varphi \in L_2(\mathbb{R}^d)$ *where the characteristic functional* $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}} : L_2(\mathbb{R}^d) \to \mathbb{R}$ *is continuous and positive-definite.*

Proof. The first statement is simply the re-transcription of the defining zero-mean property. For any given $\varphi_1, \varphi_2 \in L_2(\mathbb{R}^d)$, we define the random variables

$$Y_1 = \langle W_{\mathrm{Gauss}}, \frac{\varphi_1 + \varphi_2}{\sqrt{2}}\rangle \quad \text{and} \quad Y_2 = \langle W_{\mathrm{Gauss}}, \frac{\varphi_1 - \varphi_2}{\sqrt{2}}\rangle$$

which, by definition of the Gaussian innovation process, are zero-mean Gaussian with variance $\sigma_{Y_1}^2 = \frac{1}{2}\|\varphi_1 + \varphi_2\|_{L_2(\mathbb{R}^d)}^2$ and $\sigma_{Y_2}^2 = \frac{1}{2}\|\varphi_1 - \varphi_2\|_{L_2(\mathbb{R}^d)}^2$, respectively. Likewise, we consider

$$X_1 = \langle W_{\mathrm{Gauss}}, \varphi_1 \rangle = \frac{Y_1 + Y_2}{\sqrt{2}} \quad \text{and} \quad X_2 = \langle W_{\mathrm{Gauss}}, \varphi_2 \rangle = \frac{Y_1 - Y_2}{\sqrt{2}}.$$

We then observe that $X_1 X_2 = \frac{1}{2}(Y_1^2 - Y_2^2)$, from which we deduce that

$$\mathbb{E}\{\langle W_{\mathrm{Gauss}}, \varphi_1 \rangle \langle W_{\mathrm{Gauss}}, \varphi_2 \rangle\} = \tfrac{1}{2}\,\mathbb{E}\{Y_1^2\} - \tfrac{1}{2}\,\mathbb{E}\{Y_2^2\} = \tfrac{1}{2}\sigma_{Y_1}^2 - \tfrac{1}{2}\sigma_{Y_2}^2 = \langle \varphi_1, \varphi_2 \rangle_{L_2(\mathbb{R}^d)}.$$

By definition, the probability density of X is $p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{x^2}{2\sigma_X^2}\right)$. The characteristic function of X is the conjugate Fourier transform of p_X, which is given by

$$\hat{p}_X(\omega) = \mathcal{F}^*\{p_X\}(\omega) = \int_{\mathbb{R}} p_X(x) \mathrm{e}^{\mathrm{j}\omega x} \mathrm{d}x = \mathrm{e}^{-\frac{1}{2}\omega^2 \sigma_X^2}.$$

In terms of expected values, this is equivalent to

$$\mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega \langle W_{\mathrm{Gauss}}, \varphi \rangle}\} = \exp\left(-\tfrac{1}{2}\omega^2 \|\varphi\|_{L_2(\mathbb{R}^d)}^2\right).$$

Since the latter identity holds for any $\varphi \in L_2(\mathbb{R}^d)$ and $\omega \in \mathbb{R}$, we obtain the desired result by setting $\omega = 1$.

To establish the positive definiteness of $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}} : L_2(\mathbb{R}^d) \to \mathbb{C}$, we use the property that the functional $G : \varphi \mapsto -\frac{1}{2}\|\varphi\|_{L_2(\mathbb{R}^d)}^2$ is continuous and conditionally positive-definite on $L_2(\mathbb{R}^d)$ (see Example 5). We then invoke Theorem 27 below with $\tau = 1$, which is the functional counterpart of Schoenberg's correspondence principle [2, 4].

Theorem 27 (Schoenberg's correspondence [?, Lemma 2.3])**.** *Let $G : \mathcal{H} \to \mathbb{C}$ be a complex-valued functional defined over a Hilbert space $\mathcal{H}$ such that $G(0) = 0$. Then, G is conditionally positive-definite if and only if $F(\varphi) = \exp\big(\tau G(\varphi)\big)$ is positive-definite on $\mathcal{H}$ for any $\tau \in \mathbb{R}^+$.*

The proof of this correspondence is essentially the same as the one for the scalar case with $\mathcal{H} = \mathbb{R}$ [2, 3].

As $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}$ is the composition of $F(\cdot) = -\frac{1}{2}\|\cdot\|_{L_2(\mathbb{R}^d)}^2$, which is obviously continuous $L_2(\mathbb{R}^d) \to \mathbb{R}$, and the exponential function, which is continuous $\mathbb{R} \to \mathbb{R}$, we readily deduce that $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}} = \exp \circ F$ continuously maps $L_2(\mathbb{R}^d) \to \mathbb{R}$. $\square$

Theorem 28. *A generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ is Gaussian if and only if $\widehat{\mathscr{P}}_G(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\varphi\rangle}\} = \exp\big(-\frac{1}{2}C_G(\varphi,\varphi) + \mathrm{j}\langle \mu_G, \varphi\rangle\big)$ where C_G is a continuous positive-definite bilinear form : $\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ and $\mu_G \in \mathcal{S}'(\mathbb{R}^d)$. This generalized Gaussian process is uniquely characterized by its mean*

$$\mathbb{E}\{G\} = \mu_G$$

and its covariance operator $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ defined as

$$\varphi \mapsto \langle \mathrm{R}_G\{\varphi\}, \cdot\rangle = C_G(\varphi, \cdot),$$

which is indicated as $G \sim \mathcal{N}(\mu_G, \mathrm{R}_G)$ in $\mathcal{S}'(\mathbb{R}^d)$, whereas the covariance form of the process is C_G, as the notation suggests.

Proof. We use the same arguments as in the proof of Proposition 17 to first show that the functional $\varphi \mapsto -\frac{1}{2}C_G(\varphi,\varphi) + \mathrm{j}\langle \mu_G, \varphi\rangle$ is continuous and conditionally-positive-definite over $\mathcal{S}(\mathbb{R}^d)$. This allows us to deduce that $\widehat{\mathscr{P}}_G(\varphi)$ is a valid characteristic functional, which proves that G is a generalized stochastic process in $\mathcal{S}'(\mathbb{R}^d)$.

We then apply the definition $\widehat{\mathscr{P}}_G(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\varphi\rangle}\}$ to determine the characteristic function of the scalar random variable $X_1 = \langle G, \varphi_1\rangle$ as

$$\begin{aligned}
\mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X_1}\} &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\omega\varphi_1\rangle}\} = \widehat{\mathscr{P}}_G(\omega\varphi_1) \\
&= \exp\big(-\tfrac{1}{2}C_G(\omega\varphi_1, \omega\varphi_1) + \mathrm{j}\langle \mu_G, \omega\varphi_1\rangle\big) \\
&= \exp\big(-\tfrac{1}{2}\omega^2 C_G(\varphi_1, \varphi_1) + \mathrm{j}\omega\langle \mu_G, \varphi_1\rangle\big) = \mathrm{e}^{-\frac{1}{2}\omega^2\sigma_1^2}\mathrm{e}^{\mathrm{j}\omega\mu_1} \qquad (136)
\end{aligned}$$

with $\mu_1 = \langle \mu_G, \varphi_1\rangle$ and $\sigma_1^2 = C_G(\varphi_1, \varphi_1)$, which is recognized as the (conjugate) 1D Fourier transform of a Gaussian with mean μ_1 and variance σ_1^2. In other words, we have established that $\langle G, \varphi\rangle \sim \mathcal{N}\big(\langle \mu_G, \varphi\rangle, C_G(\varphi,\varphi)\big)$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$. This proves that G is a generalized Gaussian process in $\mathcal{S}'(\mathbb{R}^d)$ and allows us to identify μ_G as the mean of the stochastic process and C_G as its covariance form—i.e., $C_G(\varphi_1, \varphi_2) = \mathbb{E}\{\langle G - \mu_G, \varphi_1\rangle\langle G - \mu_G, \varphi_2\rangle\}$.

For the converse part of the claim, we invoke Theorem 22 to show that the continuity and positive definiteness of $C_G : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ is not only sufficient but also necessary for this construction. The same is obviously true for the condition $\mu_G = \mathbb{E}\{G\} \in \mathcal{S}'(\mathbb{R}^d)$. In addition, Theorem 22 ensures the existence of the symmetric, positive-definite covariance operator $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ such that $C_G(\varphi_1, \varphi_2) = \langle \mathrm{R}_G\{\varphi_1\}, \varphi_2\rangle$.

□

Observe the striking similarity between this characterization of generalized Gaussian processes and the description of multivariate Gaussian distributions of Appendix E.4. In particular, there is a direct correspondence between the covariance operator $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ and the covariance matrix $\mathbf{C}$ of a multivariate Gaussian, which can also be viewed as a positive-definite operator $\mathbb{R}^N \to \mathbb{R}^N$. In effect, we have substituted the finite-dimensional linear functional $\boldsymbol{\xi} \mapsto \boldsymbol{\mu}^T\boldsymbol{\xi} = \langle \boldsymbol{\mu}, \boldsymbol{\xi}\rangle$ and the quadratic form $\boldsymbol{\xi} \mapsto \boldsymbol{\xi}^T\mathbf{C}\boldsymbol{\xi} = \langle \mathbf{C}\boldsymbol{\xi}, \boldsymbol{\xi}\rangle$ in Definition 43 by their infinite-dimensional counterparts $\varphi \mapsto \langle \mu_G, \varphi\rangle$ and $\varphi \mapsto \langle \mathrm{R}_G\{\varphi\}, \varphi\rangle$, respectively. While this seems quite reasonable retrospectively, the important point is that the construction is mathematically sound, thanks to Schwartz's kernel theorem and the nuclearity of the dual pair of spaces $\big(\mathcal{S}(\mathbb{R}^d), \mathcal{S}'(\mathbb{R}^d)\big)$. Based on this analogy, one may also wonder if there is an infinite-dimensional counterpart to (164) (the equation of the multivariate Gaussian pdf). Unfortunately, this is bound to failure because of the curse of dimensionality, the main point being that (164) vanishes as the dimension N tends to infinity (because of the term $(2\pi)^{N/2}$ in the denominator). This confirms the claim that is not possible to write the pdf of an infinite-dimensional random object—by contrast with its probability measure, which is always well defined.

In order to categorize stochastic processes, it is of interest to investigate the effect of simple coordinate transformations. As preliminary step, we identify classes of operators whose action commutes with the primary coordinate transformations: translation, scaling and rotation.

Definition 28 (Invariances). *Let $g \in \mathcal{S}'(\mathbb{R}^d)$ be a continuous linear functional on $\mathcal{S}(\mathbb{R}^d)$ and T a continuous linear operator $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$. Then, g and T respectively are said to be*

- shift-invariant *if, for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$,*

$$\langle g(\cdot - \boldsymbol{x}_0), \varphi\rangle \stackrel{\triangle}{=} \langle g, \varphi(\cdot + \boldsymbol{x}_0)\rangle = \langle g, \varphi\rangle$$
$$\mathrm{T}\{\varphi(\cdot - \boldsymbol{x}_0)\} = \mathrm{T}\{\varphi\}(\cdot - \boldsymbol{x}_0)$$

- scale-invariant *of order γ if, for any $a \in \mathbb{R}^+$,*

$$\langle g(a\cdot), \varphi\rangle \stackrel{\triangle}{=} \langle g, |a|^{-d}\varphi(\cdot/a)\rangle = a^{\gamma}\langle g, \varphi\rangle$$
$$\mathrm{T}\{\varphi(a\cdot)\} = a^{\gamma}\mathrm{T}\{\varphi\}(a\cdot)$$

- rotation-invariant *if, for any rotation matrix $\boldsymbol{R} : \mathbb{R}^d \to \mathbb{R}^d$,*

$$\langle g(\boldsymbol{R}\cdot), \varphi\rangle \stackrel{\triangle}{=} \langle g, \varphi(\boldsymbol{R}^{-1}\cdot)\rangle = \langle g, \varphi\rangle$$
$$\mathrm{T}\{\varphi(\boldsymbol{R}\cdot)\} = \mathrm{T}\{\varphi\}(\boldsymbol{R}\cdot).$$

While the complete characterization of the linear maps that meet the above properties is a research topic on his own right, we can at least provide some guidelines. The simplest case is shift-invariance, which is met iff. $g = \mathrm{Const}$ and T is a generalized convolution operator such that $\mathrm{T}\{\varphi\} = h*\varphi$ where $h = \mathrm{T}\{\delta\} \in \mathcal{S}'(\mathbb{R}^d)$ is the impulse response of the underlying filter. The scale-invariant functionals correspond to the class of so-called homogeneous distributions which is closed under the generalized Fourier transformation [?]. When the functional $g : \varphi \mapsto \langle g, \varphi\rangle = \int_{\mathbb{R}^d} g(\boldsymbol{x})\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ is associated with an ordinary function $g : \mathbb{R}^d \to \mathbb{R}$, g is rotation-invariant iff. it is a purely radial function; i.e., $g(\boldsymbol{x}) = g_{\mathrm{pol}}(\|\boldsymbol{x}\|)$ for any $\boldsymbol{x} \in \mathbb{R}^d$. The tempered distributions that are both scale- and rotation-invariant reduces to the family $h_\gamma(\boldsymbol{x}) = \|\boldsymbol{x}\|^\gamma$ with $\gamma \in \mathbb{R}$. This allows us to identify the family of operators that are simultaneously shift, scale and rotation-invariant as $\varphi \mapsto h_\gamma * \varphi$, where the latter convolution is defined in the distributional sense; it is such that

$$h_\gamma * \varphi = \mathcal{F}^{-1}\{\hat{h}_\gamma \hat{\varphi}\},$$

where $\hat{h}_\gamma = \mathcal{F}\{h_\gamma\} \in \mathcal{S}'(\mathbb{R}^d)$ and $\hat{\varphi}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle}\mathrm{d}\boldsymbol{x} \in \mathcal{S}(\mathbb{R}^d)$ are the generalized and ordinary Fourier transforms of $h_\gamma \in \mathcal{S}'(\mathbb{R}^d)$ and $\varphi \in \mathcal{S}(\mathbb{R}^d)$, respectively.

Since a Gaussian process in $\mathcal{S}'(\mathbb{R}^d)$ is uniquely specified by its mean and covariance operator—or, in more abstract terms, by a pair of linear and bilinear functionals on $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d)$, respectively—one can expect some form of equivalence (and transfer) between the properties of the process and the properties of these functionals. This is now made explicit for the invariance and continuity properties encountered so far.

Proposition 18 (Properties of Gaussian processes)**.** *Let $G \sim \mathcal{N}(\mu_G, \mathrm{R}_G)$ be a generalized Gaussian stochastic process in $\mathcal{S}'(\mathbb{R}^d)$ with mean $\mu_G \in \mathcal{S}'(\mathbb{R}^d)$ and covariance operator $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$, as specified in Theorem 28. Then, depending on the properties of μ_G and R_G, the process G is:*

- *stationnary iff. both μ_G and R_G are shift-invariant; that is, when $\mu_G =$ Const and R_G is a (positive-definite) convolution operator.*
- *self-similar with Hurst exponent H iff. μ_G and R_G are scale-invariant of order H and $2H$, respectively;*
- *isotropic iff. both μ_G and R_G are rotation-invariant;*
- *continuous in the mean-square sense on $\mathbb{R}^d$ iff. there exists some $\alpha \in \mathbb{R}$ such that $\mu_G = \mathbb{E}\{G\} \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ and $r_G \in C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$ where r_G*

is the kernel of the covariance operator R_G*. In other words, the mean and the covariance functions of* G *both need to be continuous and of slow growth.*

Proof. We focus on the reverse implication (necessity) since the direct part of these statements is obvious. According to Definition 22, G is stationary iff. $\mu = \langle \mu_G, \varphi\rangle = \langle \mu_G, \varphi(\cdot + \boldsymbol{x}_0)\rangle$ and $\mathrm{Var}\{G(\varphi)\} = C_G\big(\varphi, \varphi\big) = C_G\big(\varphi(\cdot + \boldsymbol{x}_0), \varphi(\cdot + \boldsymbol{x}_0)\big) = \mathrm{Var}\{G\big(\varphi(\cdot + \boldsymbol{x}_0)\big)\}$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\boldsymbol{x}_0 \in \mathbb{R}^d$. The first condition yields $\mu_G = $ Const, while the second is equivalent to $C_G\big(\varphi_1(\cdot - \boldsymbol{x}_0), \varphi_2(\cdot - \boldsymbol{x}_0)\big) = C_G\big(\varphi_1, \varphi_2\big)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$, in view of (130) and the linearity of the shift operator. By taking $\varphi = \varphi_2(\cdot - \boldsymbol{x}_0)$, the latter is rewritten as $\langle \mathrm{R}_G\{\varphi_1(\cdot - \boldsymbol{x}_0)\}, \varphi\rangle = \langle \mathrm{R}_G\{\varphi_1\}, \varphi(\cdot + \boldsymbol{x}_0)\rangle$, which is equivalent to the shift-invariance of R_G.

Similar considerations apply for the self-similarity and isotropy properties. For instance, the definition of a self-similar process with Hurst index H yields $a^H\langle \mu_G, |a|^d\varphi(a\cdot)\rangle = \langle \mu_G, \varphi\rangle$ and $a^{2H} C_G\big(|a|^d\varphi_1(a\cdot), |a|^d\varphi_2(a\cdot)\big) = C_G(\varphi_1, \varphi_2)$ for any $a \in \mathbb{R}^+$. With the (surface preserving) change of variable $\varphi = |a|^d\varphi_2(a\cdot)$, the latter identity translates into

$$a^{2H}\langle \mathrm{R}_G\{\varphi_1(a\cdot)\}, \varphi\rangle = \langle \mathrm{R}_G\{\varphi_1\}, |a|^{-d}\varphi(\cdot/a)\rangle = \langle \mathrm{R}_G\{\varphi_1\}(a\cdot), \varphi\rangle\big),$$

which is equivalent to the announced scale-invariance of the covariance operator. Likewise, the isotropy property implies that

$$C_G\big(\varphi_1(\boldsymbol{R}^{-1}\cdot), \varphi_2(\boldsymbol{R}^{-1}\cdot)\big) = C_G\big(\varphi_1, \varphi_2\big) = \langle \mathrm{R}_G\{\varphi_1\}, \varphi_2\rangle$$

for any rotation matrix $\boldsymbol{R}$, from which we deduce the rotation-invariance of R_G by taking $\varphi = \varphi_2(\boldsymbol{R}^{-1}\cdot)$.

The last statement on mean-square continuity follows directly from Theorem 23. Since the two functions $\mu_G : \mathbb{R}^d \to \mathbb{R}$ and $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are constrained to be of slow growth, this is equivalent to the existence of some $\alpha \in \mathbb{R}$ such that $\mu_G \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ and $r_G \in C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$. □

Let us recall that the kernel of the covariance operator coincides with the covariance function $r_G(\boldsymbol{x}, \boldsymbol{y})$. When the latter is the reproducing kernel of a RHKS $\mathcal{H} \subseteq \mathcal{S}'(\mathbb{R}^d)$, we can rewrite the characteristic functional of G in Theorem 28 as $\widehat{\mathscr{P}}_G(\varphi) = \exp\big(-\frac{1}{2}\|\varphi\|^2_{\mathcal{H}'} + \mathrm{j}\langle \mu_G, \varphi\rangle\big)$ where $\mathcal{H}'$ is the dual of the RKHS generated by r_G. Accordingly, we may safely extend the domain of $\widehat{\mathscr{P}}_G$ from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{H}'$ (see Theorem 26). The advantage of this configuration is that $\mathcal{H}'$ (as the dual of a RKHS) includes the sampling functionals $\delta(\cdot - \boldsymbol{x}_0)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$ (see Definition 6). This allows us to give a meaning to the random variables $\{G(\boldsymbol{x}) = \langle G, \delta(\cdot - \boldsymbol{x})\rangle : \boldsymbol{x} \in \mathbb{R}^d\}$, thereby drawing the connection with the classical formulation of stochastic processes. This leads to the following result, the last part of which incorporates the characterization of RKHS in Theorem 6.

Corollary 5 (Classical Gaussian process)**.** *A generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ is equivalent to a "classical" Gaussian process on $\mathbb{R}^d$ if and only if its characteristic functional is of the form*

$$\widehat{\mathscr{P}}_G(\varphi) = \exp\left(-\tfrac{1}{2}\|\varphi\|^2_{\mathcal{H}'} + \mathrm{j}\langle \mu_G, \varphi\rangle\right) \tag{137}$$

with

$$\|\varphi\|^2_{\mathcal{H}'} = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) r_G(\boldsymbol{x}, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} = \langle \varphi, \mathrm{R}_G\{\varphi\}\rangle$$

and $\mu_G \in \mathcal{H}$, where $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the reproducing kernel of some RKHS $\mathcal{H} \subseteq \mathcal{S}'(\mathbb{R}^d)$. This means that $G \sim \mathcal{N}(\mu_G, \mathrm{R}_G)$ and that its sample values, $\{G(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$, are well-defined Gaussian random variables with mean

$$\mathbb{E}\{G(\boldsymbol{x})\} = \mu_G(\boldsymbol{x})$$

and covariance function

$$\mathbb{E}\left\{\big(G(\boldsymbol{x}) - \mu_G(\boldsymbol{x})\big)\big(G(\boldsymbol{y}) - \mu_G(\boldsymbol{y})\big)\right\} = r_G(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{R}_G\{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x}).$$

Finally, G is mean-square continuous if and only if $r_G \in C_{\mathrm{b},\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$ for some $\alpha \in \mathbb{R}$, which implies that $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$.

Moreover, based on Proposition 18, we immediately deduce that such a Gaussian process is

- stationary iff. $\mu_G(\boldsymbol{x}) = \mathbb{E}\{G(\boldsymbol{x})\} = \mathrm{Const}$ and $r_G(\boldsymbol{x}, \boldsymbol{y}) = a_G(\boldsymbol{x} - \boldsymbol{y})$ where $a_G \in C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ for some $\alpha \in \mathbb{R}$;
- self-similar of order $H \geq 0$ iff. $\mu_G(a\boldsymbol{x}) = a^H \mu_G(\boldsymbol{x})$ and $r_G(a\boldsymbol{x}, a\boldsymbol{y}) = a^{2H} r_G(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ and any $a \in \mathbb{R}^+$;
- isotropic iff. $\mu_G(\boldsymbol{x}) = \mu_{\mathrm{pol}}(\|\boldsymbol{x}\|)$ and $r_G(\boldsymbol{x}, \boldsymbol{y}) = r_{\mathrm{pol}}(\|\boldsymbol{x}\|, \|\boldsymbol{y}\|)$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$; that is, iff. the mean and covariance functions are purely radial functions.

We conclude this section on Gaussian processes with the determination of their finite-dimensional probability density functions (or marginals).

Proposition 19 (Gaussian marginals). *Let $G \sim \mathcal{N}(\mu_G, \mathrm{R}_G)$ with $\mathrm{R}_G : \varphi \mapsto \int_{\mathbb{R}^d} r_G(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ be a Gaussian process on $\mathbb{R}^d$ whose covariance function $r_G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the reproducing kernel of a RKHS $\mathcal{H} \subseteq \mathcal{S}'(\mathbb{R}^d)$ and such that $\mu_G \in \mathcal{H}$. Then, $\boldsymbol{Y} = (\langle G, \varphi_1\rangle, \dots, \langle G, \varphi_N\rangle)$ is a well-defined multivariate Gaussian vector if and only if $\varphi_1, \dots, \varphi_N \in \mathcal{H}'$. Specifically, $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{Y}}, \mathbf{C}_{\boldsymbol{Y}})$ with mean vector*

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = (\langle \mu_G, \varphi_1\rangle, \dots, \langle \mu_G, \varphi_N\rangle) \in \mathbb{R}^N$$

and covariance matrix $\mathbf{C}_{\boldsymbol{Y}} \in \mathbb{R}^{N\times N}$ such that

$$\begin{aligned}[\mathbf{C}_{\boldsymbol{Y}}]_{m,n} &= \langle \mathrm{R}_G\{\varphi_m\}, \varphi_n\rangle = \langle \varphi_m, \varphi_n\rangle_{\mathcal{H}'} \\ &= \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \varphi_m(\boldsymbol{x}) r_G(\boldsymbol{x}, \boldsymbol{y})\varphi_n(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.\end{aligned}$$

Proof. Theorem 28 yields the characteristic functional of the process, which is rewritten as

$$\widehat{\mathscr{P}}_G(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G,\varphi\rangle}\} = \exp\left(-\tfrac{1}{2}\langle \varphi, \varphi\rangle_{\mathcal{H}'} + \mathrm{j}\langle \mu_G, \varphi\rangle\right).$$

Since the domain of continuity of $\varphi \mapsto \langle \varphi, \varphi\rangle_{\mathcal{H}'}$ and $\varphi \mapsto \langle \mu_G, \varphi\rangle$ for any fixed $\mu_G \in \mathcal{H}$ is precisely $\mathcal{H}'$ (from the definition of the inner and duality products), the property carries over to the characteristic functional; i.e., $\widehat{\mathscr{P}}_G$ is continuous and positive-definite over $\mathcal{H}'$ (sufficiency). The characteristic function of $Y_n = \langle G, \varphi_n\rangle$ is then given by

$$\xi \mapsto \mathbb{E}\{\mathrm{e}^{\mathrm{j}\xi Y_n}\} = \widehat{\mathscr{P}}_G(\xi\varphi_n) = \exp\left(-\tfrac{1}{2}\xi^2\langle \varphi_n, \varphi_n\rangle_{\mathcal{H}'} + \mathrm{j}\xi\langle \mu_G, \varphi_n\rangle\right), \quad (138)$$

from which we immediately deduce that $Y_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ with $\mu_n = \langle \mu_G, \varphi_n\rangle$ and $\sigma_n^2 = \|\varphi_n\|^2_{\mathcal{H}'}$. Conversely, the formal expression on the rhs of (138) is continuous positive-definite—and hence a valid characteristic function (by Bochner's theorem)—iff. $\langle \varphi_n, \varphi_n\rangle_{\mathcal{H}'} < \infty$, which establishes the necessity of the condition $\varphi_n \in \mathcal{H}'$ for all n. The same mapping technique applies to the determination of the joint characteristic function of $\boldsymbol{Y}$. Specifically, we have that

$$\begin{aligned}\hat{p}_{\boldsymbol{Y}}(\boldsymbol{\xi}) &= \widehat{\mathscr{P}}_G(\xi_1\varphi_1 + \cdots + \xi_M\varphi_M) \\ &= \exp\left(-\tfrac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N}\xi_m\xi_n\langle \varphi_m, \varphi_n\rangle_{\mathcal{H}'} + \mathrm{j}\sum_{n=1}^{N}\xi_n\langle \mu_G, \varphi_n\rangle\right) \\ &= \exp\left(-\tfrac{1}{2}\boldsymbol{\xi}^T\mathbf{C}_{\boldsymbol{Y}}\boldsymbol{\xi} + \mathrm{j}\boldsymbol{\mu}_{\boldsymbol{Y}}^T\boldsymbol{\xi}\right)\end{aligned}$$

where we have made use of the (bi)linearity of the inner- and duality products and identified the underlying mean vector $\boldsymbol{\mu_Y}$ and covariance matrix $\mathbf{C_Y}$. Finally, we refer to Definition 43 for the general expression of the characteristic function of a multivariate Gaussian. $\square$

4.5 Gaussian solutions of stochastic differential equations

Let us consider a generalized Gaussian process S in $\mathcal{S}'(\mathbb{R}^d)$ whose covariance operator is factorizable as $\mathrm{R}_S = \mathrm{TT}^*$ where the continuous linear operators $\mathrm{T} : \mathcal{S}'(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ and $\mathrm{T}^* : \mathcal{S}(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ form an adjoint pair. We shall now show that $S \sim \mathcal{N}(\mu_S, \mathrm{R}_S)$ can be synthesized by linear transformation (or filtering) of the Gaussian white noise (or innovation process) W_{Gauss} specified in Proposition 17. We shall also investigate variations around this theme that involve the solution of linear stochastic differential equations.

The proposed construction is described by the following (deterministic) map

$$\omega \mapsto w = W_{\mathrm{Gauss}}(\omega) \mapsto s = S(\omega) = \mathrm{T}\{w\} + \mu_S, \tag{139}$$

where s and w denote the realizations of the Gaussian stochastic processes S and W_{Gauss} for the same outcome ω in our universal sample space Ω. Since $s, w \in \mathcal{S}'(\mathbb{R}^d)$ by assumption, the explicit transcription of the above rhs. equality is

$$\langle s, \varphi\rangle = \langle \mathrm{T}\{w\} + \mu_S, \varphi\rangle = \langle w, \mathrm{T}^*\varphi\rangle + \langle \mu_S, \varphi\rangle \tag{140}$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Implicit to this manipulation is the continuity of $\varphi \mapsto \langle w, \mathrm{T}^*\varphi\rangle$, which follows from our assumptions; namely,

1. $\mathrm{T}^*\varphi \in L_2(\mathbb{R}^d)$ since T^* continuously maps $\mathcal{S}(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$;
2. the domain of the random functional $\varphi \mapsto \langle W_{\mathrm{Gauss}}, \varphi\rangle$ is extendable to $L_2(\mathbb{R}^d)$, as shown in Proposition 17.

Based on the right-hand side of (140) and the interpretation of $\langle \mu_S, \varphi\rangle$ as a constant process whose characteristic form is given by (133) with $p_0 = \mu_S \in \mathcal{S}'(\mathbb{R}^d)$, we immediately deduce that

$$\widehat{\mathscr{P}}_S(\varphi) = \widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}(\mathrm{T}^*\varphi)\mathrm{e}^{\mathrm{j}\langle \mu_S, \varphi\rangle} = \exp\left(-\tfrac{1}{2}\|\mathrm{T}^*\varphi\|^2_{L_2(\mathbb{R}^d} + \mathrm{j}\langle \mu_G, \varphi\rangle\right).$$

This expression is compatible with the generic form of a generalized Gaussian process given in Theorem 28, which proves that $S \sim \mathcal{N}(\mu_G, \mathrm{TT}^*)$. The covariance form of the process is specified by

$$C_S(\varphi, \varphi) = \langle \mathrm{T}^*\varphi, \mathrm{T}^*\varphi\rangle = \|\mathrm{T}^*\varphi\|^2_{L_2(\mathbb{R}^d} = \langle \mathrm{TT}^*\varphi, \varphi\rangle \geq 0$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$. The central equality clearly shows that the operator $\mathrm{TT}^* : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is positive-definite, while it also suggests that the factorization is not unique. In fact, there is a whole equivalence class of possible transformations given by $\tilde{\mathrm{T}} = \mathrm{TU}$ where U is an arbitrary unitary operator $L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ with the property that $\mathrm{U}^{-1} = \mathrm{U}^*$.

We can relate the above procedure to the class of reproducing kernels investigated in Section 2.7 by considering the innovation model

$$\mathrm{L}s = w \tag{141}$$

where $w = W_{\mathrm{Gauss}}(\omega)$ is a realization of the Gaussian innovation process and L a spline-admissible operator in the sense of Definition 13.

In the easy case where L is invertible (coercive scenario), the solution of (141) is given by $s = \mathrm{L}^{-1}w$, which is compatible with the above linear generation mechanism if we set $\mathrm{T}^* = \mathrm{L}^{-1*}$. This leads to the conclusion that (141) uniquely specifies a generalized Gaussian process $S \sim \mathcal{N}(0, \mathrm{L}^{-1*}\mathrm{L}^{-1})$ in $\mathcal{S}'(\mathbb{R}^d)$ iff. L^{-1*} continuously maps $\mathcal{S}(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$.

The more interesting case is when L is a differential operator with a non-trivial null space of dimension N_0. To solve the corresponding *linear stochastic differential equation*, one then needs to add N_0 boundary conditions. A first possibility is to take

$$\mathrm{L}s = w \quad \text{s.t.} \quad \boldsymbol{\phi}(s) = \mathbf{0}$$

where the linear operator $\boldsymbol{\phi} : \mathcal{S}'(\mathbb{R}^d) \to \mathbb{R}^{N_0}$ fullfills the admissibility conditions of Definition 8. Alternatively, one may consider the more general model

$$\mathrm{L}s = w \quad \text{s.t.} \quad \boldsymbol{\phi}(s) = (a_1, \dots, a_{N_0}) \tag{142}$$

where the a_n are realizations of a series of independent Gaussian random variables A_n with zero mean and variance σ_n^2. The solution of (142) is then given by

$$s = \mathrm{L}_{\boldsymbol{\phi}}^{-1}w + \sum_{n=1}^{N_0} a_n p_n \tag{143}$$

where $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is the stable right-inverse of L specified in Theorem 10. At the level of the random process itself, this translates into

$$S = \mathrm{L}_{\boldsymbol{\phi}}^{-1}W_{\mathrm{Gauss}} + \sum_{n=1}^{N_0} A_n p_n$$

where the individual component processes $\mathrm{L}_{\boldsymbol{\phi}}^{-1}W_{\mathrm{Gauss}}, A_1p_1, \dots, A_{N_0}p_{N_0}$ are independent. By using the property that $\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$ continuously maps $\mathcal{S}(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ together with Property 5 in Theorem 38, we find that

$$\widehat{\mathscr{P}}_S(\varphi) = \exp\left(-\tfrac{1}{2}\|\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\varphi\|_{L_2}^2 - \tfrac{1}{2}\sum_{n=1}^{N_0}\sigma_n^2|\langle p_n, \varphi\rangle|^2\right).$$

Likewise, we readily calculate the covariance function of S as

$$r_S(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}\{S(\boldsymbol{x})S(\boldsymbol{y})\} = a_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{n=1}^{N_0}\sigma_n^2 p_n(\boldsymbol{x})p_n(\boldsymbol{y})$$

where $a_{\boldsymbol{\phi}}$ is the kernel of the symmetric, positive-definite operator $\mathrm{A}_{\boldsymbol{\phi}} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}$ whose explicit formula is provided by Theorem 11. Interestingly, we observe that the resulting composite kernel r_S has the same parametric form as $a_{\boldsymbol{\phi}}$ in (59) modulo some adjustment of the constants $r_{n,n} \to r_{n,n} + \sigma_n^2$ that accounts for the variability of the A_n. One can also easily generalize the model by taking $\boldsymbol{A} = (A_1, \dots, A_{N_0}) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\boldsymbol{A}})$ to be multivariate Gaussian with covariance matrix $\mathbf{C}_{\boldsymbol{A}} \in \mathbb{R}^{N_0 \times N_0}$.

To obtain the description of the process in the generic form of Corollary 5, we define the symmetric rank-one operator

$$\mathrm{P}_u : \varphi \mapsto u\langle u, \varphi\rangle$$

which is parameterized by $u \in \mathcal{S}'(\mathbb{R}^d)$. We then express the covariance operator of S as

$$\mathrm{R}_S = \mathrm{A}_{\boldsymbol{\phi}} + \sum_{n=1}^{N_0}\sigma_n^2\mathrm{P}_{p_n},$$

which allows us to conclude that $S \sim \mathcal{N}(0, \mathrm{R}_S)$. The extension over the linear generation mechanism described by (139) is the inclusion in (143) of the additional "boundary" components $a_n p_n$. While the elementary rank-one operators $\mathrm{P}_{p_n} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ are continuous and positive-definite by construction, they are not factorizable through $L_2(\mathbb{R}^d)$ unless $p_n \in L_2(\mathbb{R}^d)$, which is never the case in practice—remember that the null-space components of a differential operator is made up of polynomials and/or complex sinusoids whose L_2-norm is infinite.

At any rate, the remarkable outcome is that the covariance functions $r_S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ associated with the general class of Gaussian processes that are solution of the stochastic differential equation (142) are in direct correspondence with the reproducing kernels investigated in Section 2.7; in particular, we have an exact equivalence with the kernel in Theorem 11 (resp., the kernel in Theorem 12) by setting $\sigma_n^2 = 0$ (resp., $\sigma_n^2 = 1$) for $n = 1, \dots, N_0$.

4.6 MMSE solution of linear inverse problems

We now have all the elements to close the circle by making the connection between the minimum-error estimation of a signal under the Gaussian hypothesis (also known as the Wiener estimator) and the spline reconstruction techniques investigated in Section 3.1. To that end, we consider the linear measurement model:

$$s \mapsto \mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n} \in \mathbb{R}^M \tag{144}$$

where $s = S(\omega)$ (our so-called signal) is a realization of a Gaussian process $S \sim \mathcal{N}(\mu_S, \mathrm{R}_S)$ on $\mathbb{R}^d$ and $\boldsymbol{\nu}(s) = (\langle \nu_1, s\rangle, \dots, \langle \nu_M, s\rangle)$ is a linear measurement operator $\mathcal{S}'(\mathbb{R}^d) \to \mathbb{R}^M$ that returns the values of M (noise-free) linear measurements of the signal. The second component of the model, $\mathbf{n} \in \mathbb{R}^M$, is an additive disturbance term (discrete measurement noise) whose components are assumed to be i.i.d. Gaussian with zero mean and variance σ_0^2. The signal and noise components of the model are assumed to be mutually independent.

Given the prior knowledge of μ_S and R_S, the problem is to reconstruct the unknown signal s from its noisy measurements $\mathbf{y} \in \mathbb{R}^M$. Our working assumption is that the mean and covariance forms of the process satisfy the mean-square continuity hypotheses of Corollary 5, so that the stochastic process $\{S(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$ is well-defined in an ordinary pointwise fashion. In such a "classical" scenario, the Gaussian process is completely (and uniquely) specified by its mean and covariance functions

$$\mu_S(\boldsymbol{x}) = \mathbb{E}\{S(\boldsymbol{x})\} \tag{145}$$

$$\mathrm{r}_S(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}\left\{\big(S(\boldsymbol{x}) - \mu_S(\boldsymbol{x})\big)\big(S(\boldsymbol{y}) - \mu_S(\boldsymbol{y})\big)\right\} \tag{146}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. Moreover, we have the guarantee that r_S is the reproducing kernel of a RKHS $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$ for some $\alpha \in \mathbb{R}$, while the covariance operator $\mathrm{R}_S : \varphi \mapsto \int_{\mathbb{R}^d} r_S(\cdot, \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ is the Riesz map $\mathcal{H}' \to \mathcal{H}$ (see Proposition 7). The covariance form of the process is then given by

$$\begin{aligned} C_S(\varphi_1, \varphi_2) \triangleq \langle \mathrm{R}_S\{\varphi_1\}, \varphi_2\rangle &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi_1(\boldsymbol{x})\mathrm{r}_S(\boldsymbol{x}, \boldsymbol{y})\varphi_2(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\boldsymbol{y} \\ &= \langle \varphi_1, \varphi_2\rangle_{\mathcal{H}'} \end{aligned}$$

and it specifies the inner product for the Hilbert space $\mathcal{H}'$ (the continuous dual of the RKHS $\mathcal{H}$). This also means that its domain of continuity has been extended from $C_S : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ to $C_S : \mathcal{H}' \times \mathcal{H}' \to \mathbb{R}$, which as we shall see, has a direct incidence on the hypotheses on $\boldsymbol{\nu}$ for the well-posedness of our signal recovery problem.

Rather then attempting to reconstruct s as a whole, we shall focus on the optimal reconstruction of its sample value $s(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathbb{R}^d$. Since $S(\boldsymbol{x})$ with $\boldsymbol{x}$ fixed is an ordinary scalar random variable, we can then invoke a classical result in estimation theory, which states that the minimum-mean-square-error (MMSE) estimator of $s(\boldsymbol{x})$ given $\mathbf{y}$ is provided by the conditional mean; i.e.

$$s_{\mathrm{MMSE}}(\boldsymbol{x}|\mathbf{y}) = \mathbb{E}\{s(\boldsymbol{x})|\mathbf{y}\} = \int_{\mathbb{R}} s\, p(s|\mathbf{y})\mathrm{d}s$$

where $p(s|\mathbf{y})$ is the conditional probability of $s(\boldsymbol{x})$ given the noisy measurement $\mathbf{y} \in \mathbb{R}^M$. Hence, the work that needs to be accomplished here is the determination of $p(s(\boldsymbol{x})|\mathbf{y})$, which can already be predicted to be Gaussian.

First, we get the statistical distribution of the signal component $\boldsymbol{\nu}(S)$ by invoking Proposition 19, which yields $\boldsymbol{\nu}(S) \sim \mathcal{N}(\boldsymbol{\nu}(\mu_S), \mathbf{G})$ where the mean vector $\boldsymbol{\nu}(\mu_S) \in \mathbb{R}^M$ and the covariance matrix $\mathbf{G} \in \mathbb{R}^{M\times M}$ are given by

$$\boldsymbol{\nu}(\mu_S) = (\langle \mu_S, \nu_1\rangle, \dots, \langle \mu_S, \nu_1\rangle) \tag{147}$$

$$[\mathbf{G}]_{m,n} = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \nu_m(\boldsymbol{x}) r_S(\boldsymbol{x}, \boldsymbol{y})\nu_n(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}. \tag{148}$$

Thanks to the independence of the signal and noise components, we then deduce that the measurements $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{Y}}, \mathbf{C}_{\boldsymbol{Y}})$ have a multivariate Gaussian distribution with mean $\boldsymbol{\mu}_{\boldsymbol{Y}} = \boldsymbol{\nu}(\mu_S)$ and covariance matrix $\mathbf{C}_{\boldsymbol{Y}} = \mathbf{G} + \sigma_0^2\mathbf{I}_M$ where $\mathbf{I}_M$ is the $M \times M$ identity matrix.

Next, for some fixed $\boldsymbol{x} \in \mathbb{R}^d$, we consider $S(\boldsymbol{x}) = \langle S, \delta(\cdot - \boldsymbol{x})\rangle$, which is a scalar Gaussian random variable with mean $\mathbb{E}\{S(\boldsymbol{x})\} = \mu_S(\boldsymbol{x})$ and variance $r_S(\boldsymbol{x}, \boldsymbol{x})$. To characterize the joint dependency between $S(\boldsymbol{x})$ and the noisy measurements $\boldsymbol{Y} = (Y_1, \dots, Y_M)$, we define the augmented random variable $\boldsymbol{Z} = (S(\boldsymbol{x}), \boldsymbol{Y})$, which is multivariate Gaussian with mean vector $\mathbf{m}_{\boldsymbol{Z}} = \big(\mu_S(\boldsymbol{x}), \boldsymbol{\nu}(\mu_S)\big)$ and covariance matrix

$$\mathbf{C}_{\boldsymbol{Z}} = \begin{pmatrix} r_S(\boldsymbol{x}, \boldsymbol{x}) & \boldsymbol{\nu}^*(\boldsymbol{x})^T \\ \boldsymbol{\nu}^*(\boldsymbol{x}) & \mathbf{C}_{\boldsymbol{Y}} \end{pmatrix}$$

where

$$\boldsymbol{\nu}^*(\boldsymbol{x}) = \begin{pmatrix} \nu_1^*(\boldsymbol{x}) \\ \vdots \\ \nu_M^*(\boldsymbol{x}) \end{pmatrix}$$

with

$$\nu_m^*(\boldsymbol{x}) = \mathbb{E}\left\{S(\boldsymbol{x})Y_m\right\}.$$

Here, our use of the conjugate symbolism $\boldsymbol{\nu}^*$ is justified by the following manipulation

$$\begin{aligned}
\nu_m^*(\boldsymbol{x}) &= \mathbb{E}\left\{S(\boldsymbol{x})\langle S, \nu_m\rangle\right\} + \overbrace{\mathbb{E}\left\{S(\boldsymbol{x})N_m\right\}}^{=0} \\
&= \mathbb{E}\left\{\langle S, \delta(\cdot - \boldsymbol{x})\rangle\langle S, \nu_m\rangle\right\} = C_S\big(\delta(\cdot - \boldsymbol{x}), \nu_m\big) \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \delta(\boldsymbol{\tau} - \boldsymbol{x}) r_S(\boldsymbol{\tau}, \boldsymbol{y})\nu_m(\boldsymbol{y})\mathrm{d}\boldsymbol{\tau}\mathrm{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^d} r_S(\boldsymbol{x}, \boldsymbol{y})\nu_m(\boldsymbol{y})\mathrm{d}\boldsymbol{y},
\end{aligned} \tag{149}$$

which shows that $\nu_m^* = \mathrm{R}_S\{\nu_m\} \in \mathcal{H}$ where the covariance operator R_S is also known to be the Riesz map $\mathcal{H}' \to \mathcal{H}$.

By invoking Bayes' rule $p\big(s(\boldsymbol{x})|\mathbf{y}\big) = p\big(s(\boldsymbol{x}), \mathbf{y}\big)/p_{\boldsymbol{Y}}(\mathbf{y}) = p_{\boldsymbol{Z}}(\mathbf{z})/p_{\boldsymbol{Y}}(\mathbf{y})$ (see Appendix E.5 for the details), we then find that the conditional probability $p\big(s(\boldsymbol{x})|\mathbf{y}\big)$ is univariate Gaussian with mean

$$\mathbb{E}\{s(\boldsymbol{x})|\mathbf{y}\} = \mu_S(\boldsymbol{x}) + \boldsymbol{\nu}^*(\boldsymbol{x})^T(\mathbf{G} + \sigma_0^2\mathbf{I})^{-1}\big(\mathbf{y} - \boldsymbol{\nu}(\mu_S)\big)$$

and variance

$$\sigma^2_{s(\boldsymbol{x})|\mathbf{y}} = r_S(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{\nu}^*(\boldsymbol{x})^T(\mathbf{G} + \sigma_0^2\mathbf{I})^{-1}\boldsymbol{\nu}^*(\boldsymbol{x}).$$

We now summarize the outcome of this derivation in relation to our initial signal recovery problem.

Theorem 29 (Generalized Gauss-Markov theorem)**.** *Let us consider the following:*

- $r_S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *is the reproducing kernel of a RKHS* $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$*;*
- S *is a Gaussian process on* $\mathbb{R}^d$ *with mean* $\mathbb{E}\{S(\boldsymbol{x})\} = \mu_S(\boldsymbol{x}) \in \mathcal{H}$ *and covariance function* $\mathbb{E}\left\{\big(S(\boldsymbol{x}) - \mu_S(\boldsymbol{x})\big)\big(S(\boldsymbol{y}) - \mu_S(\boldsymbol{y})\big)\right\} = r_S(\boldsymbol{x}, \boldsymbol{y})$*;*
- *the unknown signal* $s = S(\omega) \in \mathcal{S}'(\mathbb{R}^d)$ *is a realization of* S*;*
- $\boldsymbol{\nu} : s \mapsto \boldsymbol{\nu}(s) = (\langle \nu_1, s\rangle, \dots, \langle \nu_M, s\rangle)$ *with* $\nu_m \in \mathcal{H}'$ *is a linear operator that extracts* M *measurements from the signal* s*;*

- $\mathbf{n} \in \mathbb{R}^M$ *is an independent additive white Gaussian noise (AWGN) component whose entries are i.i.d. with zero-mean and variance* σ_0^2.

Then, for any $\boldsymbol{x} \in \mathbb{R}^d$*, the minimum mean-square error (MMSE) estimation of* $s(\boldsymbol{x})$ *given the noisy linear observation* $\mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n}$ *of* s *is*

$$s_{\mathrm{MMSE}}(\boldsymbol{x}|\mathbf{y}) = \mathbb{E}\{s(\boldsymbol{x})|\mathbf{y}\} = \mu_S(\boldsymbol{x}) + \boldsymbol{\nu}^*(\boldsymbol{x})^T(\mathbf{G} + \sigma_0^2\mathbf{I}_M)^{-1}\big(\mathbf{y} - \boldsymbol{\nu}(\mu_S)\big),$$

while the corresponding estimation error is

$$\mathbb{E}\left\{\big(s_{\mathrm{MMSE}}(\boldsymbol{x}|\mathbf{y}) - s(\boldsymbol{x})\big)^2\right\} = r_S(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{\nu}^*(\boldsymbol{x})^T(\mathbf{G} + \sigma_0^2\mathbf{I}_M)^{-1}\boldsymbol{\nu}^*(\boldsymbol{x}).$$

Here, $\boldsymbol{\nu}^* = (\nu_1^*, \dots, \nu_M^*)$*, with* ν_m^* *specified by* (149)*, is the Riesz conjugate of the measurement operator* $\boldsymbol{\nu}$*, while* $\mathbf{G} \in \mathbb{R}^{M\times M}$ *is the corresponding Gram/covariance matrix whose entries are given by* (148).

In particular, if s is zero-mean (i.e., $\mu_S = 0$), then the expression of the conditional mean simplifies to

$$\mathbb{E}\{s(\boldsymbol{x})|\mathbf{y}\} = \boldsymbol{\nu}^*(\boldsymbol{x})^T(\mathbf{G} + \sigma_0^2\mathbf{I}_M)^{-1}\mathbf{y},$$

which is equivalent to

$$s_{\mathrm{MMSE}}(\boldsymbol{x}|\mathbf{y}) = \sum_{m=1}^{M} a_m \nu_m^*(\boldsymbol{x}) \tag{150}$$

with

$$\mathbf{a} = (a_1, \dots, a_M) = (\mathbf{G} + \sigma_0^2\mathbf{I}_M)^{-1}\mathbf{y}. \tag{151}$$

Remarkably, the statistical estimator defined by (150)-(151) happens to be identical to the solution of the generalized smoothing spline problem (99) in Proposition 15 for $\lambda = \sigma_0^2$ and the choice of reproducing kernel $r_{\mathcal{H}} = r_S$. In other words, we have a perfect equivalence between generalized smoothing splines and the hybrid form of Wiener reconstruction where the measurements are finite and the reconstruction is achieved in the continuous domain.

An interesting consequence of this equivalence is that the MMSE reconstruction of s is such that $s_{\mathrm{MMSE}} \in \mathcal{H}$ (because it also minimizes the spline functional (99)), while this inclusion property is typically not met by the underlying signal realization $s \in \mathcal{S}'(\mathbb{R}^d)$. In fact, in the case where s is the solution of a stochastic differential equation such as (142), we even suspect that $\mathrm{Prob}(s \in \mathcal{H}) = 0$, because of the known property that $\mathscr{P}_{W_{\mathrm{Gauss}}}\big(L_2(\mathbb{R}^d)\big) = 0$, which reflects the fact that w_{Gauss} cannot have a finite energy. The same holds true (almost surely) for all stationary signals due to their lack of decay at infinity. At any rate, it remains that the spline reconstruction of the signal provided by Theorem 29 is the best solution to our inverse problem in the absolute as it minimizes the statistical mean-square estimation error at every location $\boldsymbol{x} \in \mathbb{R}^d$.