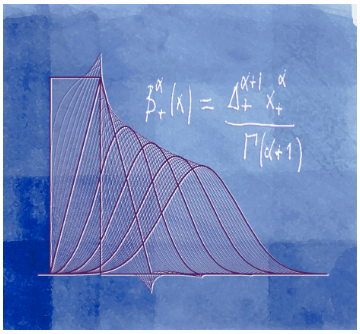
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Sparse stochastic processes

Stochastic processes and splines

Prof. Michael Unser, LIB



March 2017

EDEE Course

CONTENT

- **1. Preliminaries ✔**
- **2. Reproducing kernel Hilbert spaces ✔**
- **3. Variational splines and representer theorems ✔**
- **4. Gaussian processes**
  - Generalized stochastic processes (GSP)
  - Mean and covariance forms
  - The characteristic functional
  - Characterization of Gaussian processes
  - MMSE solution of inverse problems

Formalism of probability theory

- Probability space: $(\Omega,\ \Sigma(\Omega),\ \mathscr{P})$

All possible outcomes

Sigma algebra (all measurable sets)

Probability measure $\mathscr{P} : \Sigma(\Omega) \to [0,1]$

- Random variable as a map $X : \Omega \to \mathcal{X}$

$$\omega \mapsto X(\omega) \in \mathcal{X}$$

"Conceptual" indicator of outcome of ONE random experiment

What the observer sees: the realization

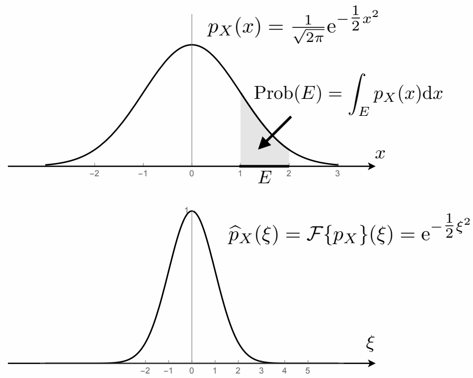
  - $\mathcal{X}$ = State space (assumed to be a vector space such as $\mathbb{R}$)
  - $Borel(\mathcal{X})$: Borel sigma-algebra of $\mathcal{X}$

    "Smallest sigma-algebra that contains all open sets of $\mathcal{X}$"
  - Induced probability measure

$$\mathscr{P}_X(E) = \mathscr{P}\{\omega \in \Omega : X(\omega) \in E\} \text{ for all } E \in Borel(\mathcal{X})$$

Review: real-valued random variable

$\mathcal{X} = \mathbb{R}$ $\qquad\qquad$ $\omega \mapsto X(\omega) \in \mathbb{R}$



- Probability density function: $p_X : \mathbb{R} \to \mathbb{R}^+$
- Probability measure: $Borel(\mathbb{R}) \to [0,1]$

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

- Expectation operator (f measurable function $\mathbb{R} \to \mathbb{R}$)

$$\mathbb{E}\{f(X)\} = \int_{\mathbb{R}} f(x)p_X(x)\mathrm{d}x = \int_{\mathbb{R}} f(x)\mathscr{P}_X(\mathrm{d}x)$$

Mean: $\mu_X = \mathbb{E}\{X\}$

Variance: $\sigma_X^2 = \mathbb{E}\{(X - \mu_X)^2\}$

- Characteristic function: $\mathbb{R} \to \mathbb{C}$

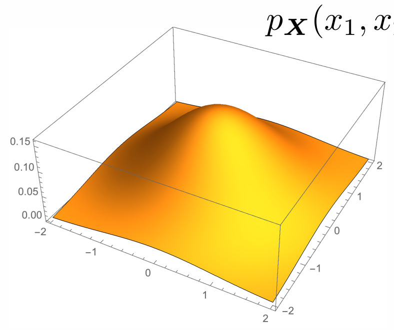
$$\widehat{\mathscr{P}}_X(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}X\xi}\} = \int_{\mathbb{R}} p_X(x)\mathrm{e}^{\mathrm{j}x\xi}\mathrm{d}x$$

Review: random vector in $\mathbb{R}^N$

$\mathcal{X} = \mathbb{R}^N$ $\qquad$ $\boldsymbol{X} = (X_1, \dots, X_N)$

$\omega \mapsto \boldsymbol{X}(\omega) \in \mathbb{R}$

$p_{\boldsymbol{X}}(x_1, x_2)$



- Probability density function $p_{\boldsymbol{X}} : \mathbb{R}^N \to \mathbb{R}^+$
- Probability measure: $Borel(\mathbb{R}^N) \to [0, 1]$

$$\mathscr{P}_{\boldsymbol{X}}(E) = \mathrm{Prob}(\boldsymbol{X}(\omega) \in E) = \int_E p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- Expectation operator (f measurable function $\mathbb{R}^N \to \mathbb{R}^M$)

$$\mathbb{E}\{f(\boldsymbol{X})\} = \int_{\mathbb{R}^N} f(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad = \int_{\mathbb{R}^N} f(\boldsymbol{x}) \mathscr{P}_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

Lebesgue–Stieltjes integral

Mean vector: $\boldsymbol{\mu}_{\boldsymbol{X}} = \mathbb{E}\{\boldsymbol{X}\} \quad \in \mathbb{R}^N$

Covariance matrix: $\mathbf{C}_{\boldsymbol{X}} = \mathbb{E}\{(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})^T\} \quad \in \mathbb{R}^{N \times N}$

- Characteristic function: $\mathbb{R}^N \to \mathbb{C}$

$$\widehat{\mathscr{P}}_{\boldsymbol{X}}(\boldsymbol{\xi}) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \boldsymbol{X}, \boldsymbol{\xi} \rangle}\} = \int_{\mathbb{R}^N} p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\xi}, \boldsymbol{x} \rangle} \mathrm{d}\boldsymbol{x} \quad = \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\xi}, \boldsymbol{x} \rangle} \mathscr{P}_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

4.1 Generalized stochastic process in $\mathcal{S}'(\mathbb{R}^d)$

$\mathcal{X} = \mathcal{S}'(\mathbb{R}^d)$

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

- Probability density function $p_G : \mathcal{X} \to \mathbb{R}^+$???

infinite dimensional

- Probability **measure**: $Borel\big(\mathcal{S}'(\mathbb{R}^d)\big) \to [0, 1]$

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

- **Abstract expectation** operator (f measurable function $\mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$)

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

Abstract Lebesgue integral

Mean g-function: $\mu_G = \mathbb{E}\{G\} \quad \in \mathcal{S}'(\mathbb{R}^d)$

Covariance **operator** $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$

$$\mathrm{R}_G = \mathbb{E}\{(G - \mu_G) \otimes (G - \mu_G)\}$$

- Characteristic **functional**: $\mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$

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

GSP as a random linear functional on $\mathcal{S}(\mathbb{R}^d)$

Any realization $g = G(\omega)$ specifies a continuous linear map $\mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$

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

Definition

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.

Examples of GSP

- (Generalized) deterministic process

$$\omega \mapsto G_{\text{Const}}(\omega) = p_0 \quad \in \mathcal{S}'(\mathbb{R}^d)$$

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

- Linear process $\Rightarrow$ finite-dimensional entity

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

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

where $A_n \sim \mathcal{N}(0, \sigma^2)$: i.i.d. Gaussian, $(p_1, \dots, p_N)$: fixed elements of $\mathcal{S}'(\mathbb{R}^d)$.

- Gaussian white noise

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

$\Rightarrow$ $W_{\text{Gauss}}(\omega)$: infinite-dimensional entity (random counterpart of Dirac δ)

Properties of GSP: Definitions

- Independence

  The GSP G_1 and G_2 in $\mathcal{S}'(\mathbb{R}^d)$ are **mutually independent**

  $\Leftrightarrow \quad X_1 = \langle G_1, \varphi \rangle$ and $X_2 = \langle G_2, \varphi \rangle$ are mutually independent for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$

- Statistical properties

  The generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$ is said to be:

  - **Gaussian** if $X = \langle G, \varphi \rangle$ is Gaussian distributed for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$

    Special case of standardized **white Gaussian noise**:

    $X = \mathcal{N}(0, \sigma_X^2)$ with $\sigma_X^2 = \|\varphi\|^2_{L_2(\mathbb{R}^d)}$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ (resp., any $\varphi \in L_2(\mathbb{R}^d)$)

  - **Stationary** Shift-invariance: $G = G(\cdot - \boldsymbol{x}_0)$ (in law)

    if $X_{\boldsymbol{x}_0} = \langle G, \varphi(\cdot + \boldsymbol{x}_0) \rangle$ identically distributed for any $\boldsymbol{x}_0 \in \mathbb{R}^d$

  - **Self-similar** with Hurst index H Scale invariance: $G = a^H G(\cdot/a)$ (in law)

    if $X_a = a^H \langle G, |a|^d \varphi(a\cdot) \rangle$ identically distributed for any $a \in \mathbb{R}^+$

Linear transformation of GSP

Adjoint pair of continuous linear operators: $\mathrm{T} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d), \quad \mathrm{T}^* : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$

- Definition of linear transformation

  $\langle \mathrm{T}\{G\}, \varphi \rangle \triangleq \langle G, \mathrm{T}^*\{\varphi\} \rangle \qquad$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$

- Primary transformations of $g = G(\omega) \in \mathcal{S}'(\mathbb{R}^d)$

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

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

  - Rotation of coordinate system $\boldsymbol{x} \mapsto \mathbf{R}\boldsymbol{x}$ with $\mathbf{R}^{-1} = \mathbf{R}^T$:

    $\langle g(\mathbf{R}\cdot), \varphi \rangle \triangleq \langle g, \varphi(\mathbf{R}^{-1}\cdot) \rangle$

  - Partial derivative operator $\partial^{\mathbf{n}}$ of multi-order $\mathbf{n} = (n_1, \ldots, n_d)$:

    $\langle \partial^{\mathbf{n}} g, \varphi \rangle \triangleq \langle g, (-1)^{|\mathbf{n}|} \partial^{\mathbf{n}} \varphi \rangle$

    where $|\mathbf{n}| = n_1 + \cdots + n_d$ and $\partial^{\mathbf{n}} \varphi(\boldsymbol{x}) = \frac{\partial^{|\mathbf{n}|} \varphi(x_1, \ldots, x_d)}{\partial x_1^{n_1} \cdots \partial x_d^{n_d}}$

4.2. Mean and covariance forms

$G(\varphi) = \langle G, \varphi \rangle$ is an ordinary random variable for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$

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

- Mean as a linear functional

$\mathbb{E}\{\langle G, \varphi \rangle\} : \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ (linear and continuous map)

$\Leftrightarrow$ there exists a unique $\mu_G \in \mathcal{S}'(\mathbb{R}^d)$ (the mean of the GSP G) such that

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

- Examples

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

$\Rightarrow$ $\mu_{G_{\mathrm{Const}}} = p_0$

Gaussian white noise: $\varphi \mapsto W_{\mathrm{Gauss}}(\varphi) = \mathcal{N}(0, \|\varphi\|_{L_2}^2)$

$\Rightarrow$ $\mu_{G_{\mathrm{Gauss}}} = 0 : \varphi \mapsto \langle \mu_{G_{\mathrm{Gauss}}}, \varphi \rangle = 0$

Covariance form / operator

Extraction of second-order statistics with $X_1 = \langle G, \varphi_1 \rangle$ and $X_2 = \langle G, \varphi_2 \rangle$

$$C_G(\varphi_1, \varphi_2) \triangleq \mathbb{E}\{\langle G - \mu_G, \varphi_1 \rangle \langle G - \mu_G, \varphi_2 \rangle\} = \mathrm{Cov}(X_1 X_2)$$

Theorem (Properties of covariance form)

Let G be a GSP 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 $C_G : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ has the following properties:

- **Symmetry**: $C_G(\varphi_1, \varphi_2) = C_G(\varphi_2, \varphi_1)$.
- **Bilinearity**: $C_G : (\varphi_1, \varphi_2) \mapsto C_G(\varphi_1, \varphi_2)$ 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**: Unique $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ s.t.

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

- Kernel representation: Unique **symmetric kernel** $r_G \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ s.t.

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

Positive-definite covariance form / kernel / operator

$$C_G(\varphi,\varphi) = \langle \mathrm{R}_G\{\varphi\},\varphi\rangle \geq 0$$

- Covariance function: $r_G(\boldsymbol{x},\boldsymbol{y}) = C_G\big(\delta(\cdot-\boldsymbol{x}),\delta(\cdot-\boldsymbol{y})\big)$
- Covariance operator: $\mathrm{R}_G\{\varphi\}(\boldsymbol{x}) = \langle r_G(\boldsymbol{x},\cdot),\varphi\rangle = \int_{\mathbb{R}^d} r_G(\boldsymbol{x},\boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$
- Effect of a linear transformation $\mathrm{T}: \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$

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

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

- Special case of a stationary processes

$$r_G(\boldsymbol{x},\boldsymbol{y}) = r_G\big(\boldsymbol{0},(\boldsymbol{y}-\boldsymbol{x})\big) = a_G(\boldsymbol{y}-\boldsymbol{x})$$

$$a_G(\boldsymbol{\tau}) \triangleq \mathbb{E}\{G(\boldsymbol{x})G(\boldsymbol{x}+\boldsymbol{\tau})\}$$ (classical autocorrelation function)

Mean and Covariance of a random vector in $\mathbb{R}^N$

$$\boldsymbol{X} = (X_1,\dots,X_N)$$

- Mean vector: $\boldsymbol{\mu_X} = \mathbb{E}\{\boldsymbol{X}\} \quad \in \mathbb{R}^N$

$$\Rightarrow \quad \mathbb{E}\{Y\} = \langle \boldsymbol{\mu_X}, \mathbf{u}\rangle$$

- Covariance matrix: $\mathbf{C}_{\boldsymbol{X}} = \mathbb{E}\{(\boldsymbol{X}-\boldsymbol{\mu_X})(\boldsymbol{X}-\boldsymbol{\mu_X})^T\} \quad \in \mathbb{R}^{N\times N}$

$$\Rightarrow \quad \mathrm{Cov}\{X_m,X_n\} = [\mathbf{C}_{\boldsymbol{X}}]_{m,n}$$

- Covariance operator $\mathbb{R}^N \to \mathbb{R}^N$: $\quad \mathbf{u} \mapsto \mathbf{v} = \mathbf{C}_{\boldsymbol{X}}\mathbf{u}$

Positive definiteness: $\langle \mathbf{u}, \mathbf{C}_{\boldsymbol{X}}\mathbf{u}\rangle = \mathrm{Var}\{Y\} \geq 0$

- Effect of a linear transformation $\mathbf{T}: \mathbb{R}^N \to \mathbb{R}^M$

$$\boldsymbol{Y} = \mathbf{T}\boldsymbol{X} \in \mathbb{R}^M$$

$$\boldsymbol{\mu_Y} = \mathbb{E}\{\boldsymbol{Y}\} = \mathbf{T}\boldsymbol{\mu_X} \in \mathbb{R}^M$$

$$\mathbf{C}_{\boldsymbol{Y}} = \exp\{(\boldsymbol{Y}-\boldsymbol{\mu_Y})(\boldsymbol{Y}-\boldsymbol{\mu_Y})^T\} = \mathbf{T}\mathbf{C}_{\boldsymbol{Y}}\mathbf{T}^T \quad \in \mathbb{R}^{M\times N}$$

Mean-square continuity and RKHS

Classical stochastic process on $\mathbb{R}^d$ = indexed collection of random variables $\{G(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$

GSP with **extended space** of test functions that includes $\delta(\cdot - \boldsymbol{x}_0)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$

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

$$\mathbb{E}\{G(\boldsymbol{x})\} = \mu_G(\boldsymbol{x}) \quad \text{and} \quad r_G(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}\{\big(G(\boldsymbol{x}) - \mu_G(\boldsymbol{x})\big)\big(G(\boldsymbol{y}) - \mu_G(\boldsymbol{y})\big)\}$$

Definition

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

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

Theorem (Mean-square continuity)

A second-order stochastic process G on $\mathbb{R}^d$ is **mean-square continuous** over $\mathbb{R}^d$ **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 is a valid **reproducing kernel**.

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

4.3 Characteristic functional

Definition

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

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

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

- Examples
  - $\widehat{\mathscr{P}}_{G_{\text{Const}}}(\varphi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G_{\text{Const}},\varphi\rangle}\} = \mathrm{e}^{\mathrm{j}\langle p_0,\varphi\rangle}$
  - Finite-dimensional Gaussian process $G_{N_0} = \sum_{n=1}^{N_0} A_n p_n$ with $A_n \sim \mathcal{N}(0, \sigma_n^2)$

$$\Rightarrow \quad Y = \langle G_{N_0}, \varphi\rangle \sim \mathcal{N}(0, \sigma_Y^2) \text{ with } \sigma_Y^2 = \sum_{n=1}^{N_0} \sigma_n^2 |\langle p_n, \varphi\rangle|^2$$

$$\Rightarrow \quad \hat{p}_Y(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\xi Y}\} = \exp\left(-\tfrac{1}{2}\xi^2\sigma_Y^2\right) \quad \Rightarrow \quad \widehat{\mathscr{P}}_{G_{N_0}}(\varphi) = \exp\left(-\tfrac{1}{2}\sum_{n=1}^{N_0} \sigma_n^2 |\langle p_n, \varphi\rangle|^2\right)$$

Continuity and positive definiteness

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

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

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

Positive-definite functionals: fundamental examples

$\mathcal{H}$: Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

- $F(\varphi) = \mathrm{e}^{-\frac{1}{2}\|\varphi\|^2_{\mathcal{H}}}$ is positive-definite over $\mathcal{H}$
- $G(\varphi) = \log F(\varphi) = -\frac{1}{2}\|\varphi\|^2_{\mathcal{H}}$ is conditionally positive-definite over $\mathcal{H}$

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

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

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

- Schoenberg's correspondence theorem

  $G(\varphi)$ conditionally positive-definite over $\mathcal{X}$

  $\Leftrightarrow \quad F(\varphi) = \mathrm{e}^{\tau G(\varphi)}$ positive-definite over $\mathcal{X}$ for any $\tau \in \mathbb{R}^+$

Characteristic functional: key properties

Theorem

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. $|\widehat{\mathscr{P}}_G(\varphi)| \le 1$), 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**.
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 $Q = \mathrm{T}\{G\} + \mu_0$ 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 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
$$\widehat{\mathscr{P}}_{G_1+G_2}(\varphi) = \widehat{\mathscr{P}}_{G_1}(\varphi)\widehat{\mathscr{P}}_{G_2}(\varphi).$$

Characteristic function: key properties

Random vector in $\mathbb{R}^N$: $\boldsymbol{X} = (X_1, \dots, X_N)$ with pdf $p_{\boldsymbol{X}} : \mathbb{R}^N \to \mathbb{R}^+$

Theorem

The characteristic function $\hat{p}_{\boldsymbol{X}}(\boldsymbol{\xi}) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\xi}, \boldsymbol{X}\rangle}\} = \int_{\mathbb{R}^N} p_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{e}^{\mathrm{j}\langle \boldsymbol{\xi}, \boldsymbol{x}\rangle}\mathrm{d}\boldsymbol{x}$ enjoys the following properties:

1. $\hat{p}_{\boldsymbol{X}} : \mathbb{R}^N \to \mathbb{C}$ is **continuous**, bounded (i.e., $|\hat{p}_{\boldsymbol{X}}(\boldsymbol{\xi})| \le 1$), Hermitian-symmetric, and **normalized** such that $\hat{p}_{\boldsymbol{X}}(\boldsymbol{0}) = 1$.
2. $\hat{p}_{\boldsymbol{X}}$ is **positive-definite**.
3. Invertibility: $p_{\boldsymbol{X}}(\boldsymbol{x}) = \mathcal{F}^{*-1}\{\hat{p}_{\boldsymbol{X}}\}(\boldsymbol{x}) = \int_{\mathbb{R}^N} \hat{p}_{\boldsymbol{X}}(\boldsymbol{\xi})\mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\xi}, \boldsymbol{x}\rangle} \frac{\mathrm{d}\boldsymbol{\xi}}{(2\pi)^N}$.
4. Linear transformation: Let $\mathbf{H} \in \mathbb{R}^{M\times N}$ be an arbitrary transformation matrix and $\mathbf{b} \in \mathbb{R}^M$ some constant offset vector. Then, the characteristic function of $\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} + \mathbf{b} \in \mathbb{R}^M$ is
$$\hat{p}_{\boldsymbol{Y}}(\boldsymbol{\xi}) = \hat{p}_{\mathbf{H}\boldsymbol{X}+\mathbf{b}}(\boldsymbol{\xi}) = \hat{p}_{\boldsymbol{X}}(\mathbf{H}^T\boldsymbol{\xi})\mathrm{e}^{\mathrm{j}\mathbf{b}^T\boldsymbol{\xi}}$$
with Fourier-domain variable $\boldsymbol{\xi} \in \mathbb{R}^M$ and $\mathbf{H}^T\boldsymbol{\xi} \in \mathbb{R}^N$.
5. Sum of independent random variables: Let $\boldsymbol{X}_1 \in \mathbb{R}^N$ and $\boldsymbol{X}_2 \in \mathbb{R}^N$ be two independent random vectors with cfs $\hat{p}_{\boldsymbol{X}_1}$ and $\hat{p}_{\boldsymbol{X}_1}$, respectively. Then, the characteristic function of $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2$ is
$$\hat{p}_{\boldsymbol{X}_1+\boldsymbol{X}_2}(\boldsymbol{\xi}) = \hat{p}_{\boldsymbol{X}_1}(\boldsymbol{\xi})\hat{p}_{\boldsymbol{X}_2}(\boldsymbol{\xi}).$$
6. Preservation of separability (or joint cf of a collection of independent random variables). Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ with $p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) = p_{\boldsymbol{X}_1}(\boldsymbol{x}_1)p_{\boldsymbol{X}_2}(\boldsymbol{x}_2)$. Then,
$$\hat{p}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}(\boldsymbol{\xi}) = \hat{p}_{\boldsymbol{X}_1}(\boldsymbol{\xi}_1)\hat{p}_{\boldsymbol{X}_2}(\boldsymbol{\xi}_2) \quad \text{with } \boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2).$$

Functional characterization theorems

Theorem (Minlos-Bochner)

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

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

Theorem (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 $\phi \in \mathcal{X}$.

- Transfer of positive-definiteness: denseness of $\mathcal{S}(\mathbb{R}^d)$ in $\mathcal{X}$ + continuity of $\widehat{\mathscr{P}}_G$
- Let $X = \langle G, \phi_0\rangle$ with $\phi_0 \in \mathcal{X}$ fixed. Then, the map $\mathbb{R} \to \mathbb{C}$

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

is continuous, positive-definite and normalized. Hence, $\mathcal{X}$ is a bona-fide random variable (by Bochner).

Distributional extension of Bochner's theorem

The (complex-valued) generalized function $g : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ is said to be positive-definite if

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \overline{\varphi(\boldsymbol{x})} g(\boldsymbol{x}-\boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} = \langle g * \varphi, \overline{\varphi}\rangle = \langle g, (\overline{\varphi}^\vee * \varphi)\rangle \geq 0$$

where $\varphi^\vee$ denotes the reversed version φ; i.e., $\varphi^\vee(\boldsymbol{x}) \triangleq \varphi(-\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathbb{R}^d$.

Theorem (Bochner-Schwartz)

A generalized function $\hat{g} \in \mathcal{S}'(\mathbb{R}^d)$ is **positive-definite if and only if** it is the generalized Fourier transform of a **positive distribution** $g \geq 0$; that is,

$$\langle \hat{g}, \varphi\rangle = \langle g, \hat{\varphi}\rangle = \int_{\mathbb{R}^d} \hat{\varphi}(\boldsymbol{x}) g(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

where $\hat{\varphi}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\omega},\boldsymbol{x}\rangle}\mathrm{d}\boldsymbol{x}$ is the Fourier transform of the test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Moreover, if $\hat{g}$ is continuous at the origin with $\hat{g}(0) = 1$, then it is the (ordinary) Fourier transform of a Borel measure $g \geq 0$ with $\int_{\mathbb{R}^d} g(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$.

Key idea (Generalized Parseval's relation)

$$\langle \hat{g}, (\varphi^\vee * \varphi)\rangle = \langle \hat{g}, \phi\rangle = \langle g, \hat{\phi}\rangle = \langle g, |\hat{\varphi}|^2\rangle \geq 0 \quad \Rightarrow \quad g \geq 0$$

4.4 Characterization of Gaussian processes

Theorem (Generalized Gaussian processes)

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

- Special case: Gaussian white noise
  - $C_{W_{\mathrm{Gauss}}}(\varphi_1,\varphi_2) = \langle\varphi_1,\varphi_2\rangle_{L_2}$ or, equivalently, $\mathrm{R}_{W_{\mathrm{Gauss}}} = \mathrm{Identity}$
  - $\widehat{\mathscr{P}}_{W_{\mathrm{Gauss}}}(\varphi) = \exp\left(-\frac{1}{2}\|\varphi\|^2_{L_2(\mathbb{R}^d)}\right) \quad \Leftrightarrow \quad W_{\mathrm{Gauss}} \sim \mathcal{N}(0, \mathrm{Identity})$ in $\mathcal{S}'(\mathbb{R}^d)$

Note: Domain extendable from $\mathcal{S}(\mathbb{R}^d)$ to $L_2(\mathbb{R}^d)$

Multivariate Gaussian distribution

Definition

The characteristic function of a multivariate Gaussian random vector of dimension N with mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and symmetric positive-definite covariance matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$ is

$$\hat{p}_{\mathrm{Gauss}}(\boldsymbol{\xi}|\boldsymbol{\mu},\mathbf{C}) = \exp\left(-\tfrac{1}{2}\boldsymbol{\xi}^T\mathbf{C}\boldsymbol{\xi} + j\boldsymbol{\mu}^T\boldsymbol{\xi}\right).$$

Bilinear form $\mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$: $\quad (\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \mapsto \langle\boldsymbol{\xi}_1, \mathbf{C}\boldsymbol{\xi}_2\rangle$

Linear form $\mathbb{R}^N \to \mathbb{R}$: $\quad \boldsymbol{\xi} \mapsto \langle\boldsymbol{\mu},\boldsymbol{\xi}\rangle$

- Notation

$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{C})$: The random vector $\boldsymbol{X} = (X_1,\dots,X_N)$ is multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$

$$p_{\mathrm{Gauss}}(\boldsymbol{X}|\boldsymbol{\mu},\mathbf{C}) = \frac{1}{\sqrt{(2\pi)^N|\det(\mathbf{C})|}}\exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\mathbf{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right)$$

Proposition (Invariance by affine transformation)

Consider some fixed matrix $\mathbf{H} \in \mathbb{R}^{N_2\times N_1}$, an offset vector $\mathbf{b} \in \mathbb{R}^{N_1}$ and some N_1-dimensional Gaussian random vector $\boldsymbol{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1,\mathbf{C}_1)$. Then, $\boldsymbol{X}_2 = \mathbf{H}\boldsymbol{X}_1 + \mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu}_2,\mathbf{C}_2)$ with

$$\boldsymbol{\mu}_2 = \mathbf{H}\boldsymbol{\mu}_1 + \mathbf{b} \qquad \text{and} \qquad \mathbf{C}_2 = \mathbf{H}\mathbf{C}_1\mathbf{H}^T.$$

Proof of Gaussian characterization theorem

- Existence and unicity of generalized stochastic process G in $\mathcal{S}'(\mathbb{R}^d)$

  $\varphi \mapsto \exp\left(-\frac{1}{2}\langle \mathrm{R}_G\{\varphi\}, \varphi\rangle + \mathrm{j}\langle \mu_G, \varphi\rangle\right)$ is positive-definite, continuous, and normalized

  (due to the positive-definiteness of R_G and Schoenberg's correspondence)

  $\Rightarrow$ G is a GSP in $\mathcal{S}'(\mathbb{R}^d)$ (by Bochner-Minlos' theorem)

- Determination of characteristic function of $X = \langle G, \varphi\rangle$:

$$\begin{aligned}\mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\} &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle G, \omega\varphi\rangle}\} = \widehat{\mathscr{P}}_G(\omega\varphi)\\ &= \exp\left(-\tfrac{1}{2}\omega^2 C_G(\varphi,\varphi) + \mathrm{j}\omega\langle \mu_G, \varphi\rangle\right) = \mathrm{e}^{-\frac{1}{2}\omega^2\sigma^2}\mathrm{e}^{\mathrm{j}\omega\mu}\end{aligned}$$

  $\Rightarrow$ Gaussian cf with mean $\mu = \langle \mu_G, \varphi\rangle$ and variance $\sigma^2 = C_G(\varphi,\varphi) = \mathrm{Var}\{X\}$

- Identification of covariance form (using bilinearity) with $X_1 = \langle G, \varphi_1\rangle$ and $X_2 = \langle G, \varphi_2\rangle$

$$\mathrm{Cov}(X_1, X_2) = \frac{1}{4}\Big(\overbrace{C_G(\varphi_1+\varphi_2, \varphi_1+\varphi_2)}^{\mathrm{Var}(X_1+X_2)} + \overbrace{C_G(\varphi_1-\varphi_2, \varphi_1-\varphi_2)}^{\mathrm{Var}(X_1-X_2)}\Big) = C_G(\varphi_1, \varphi_2)$$

- Continuity of $C_G : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ is a necessary condition (see covariance theorem)

Invariance to coordinate transformations

Continuous linear functional on $\mathcal{S}(\mathbb{R}^d)$: $g \in \mathcal{S}'(\mathbb{R}^d)$

Continuous linear operator $\mathrm{T} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$

- Definition

  g and T respectively are said to be

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

$$\langle g(\cdot - \boldsymbol{x}_0), \varphi\rangle \triangleq \langle g, \varphi(\cdot + \boldsymbol{x}_0)\rangle = \langle g, \varphi\rangle$$

$$\mathrm{T}\{\varphi(\cdot - \boldsymbol{x}_0)\} = \mathrm{T}\{\varphi\}(\cdot - \boldsymbol{x}_0)$$

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

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

$$\mathrm{T}\{\varphi(a\cdot)\} = a^\gamma\mathrm{T}\{\varphi\}(a\cdot)$$

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

$$\langle g(\mathbf{R}\cdot), \varphi\rangle \triangleq \langle g, \varphi(\mathbf{R}^{-1}\cdot)\rangle = \langle g, \varphi\rangle$$

$$\mathrm{T}\{\varphi(\mathbf{R}\cdot)\} = \mathrm{T}\{\varphi\}(\mathbf{R}\cdot).$$

Categorization of Gaussian processes

Proposition (Types 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 $\mathrm{R}_G : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$. Then, depending on the properties of μ_G and R_G, the process G is:

- **stationary** iff. both μ_G and R_G are shift-invariant; that is, when $\mu_G = \mathrm{Const}$ 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 $2H$;
- **isotropic** iff. both μ_G and R_G are rotation-invariant;
- **mean-square continuous** 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; i.e, the mean and the covariance functions are both continuous and of slow growth.

- Examples
  - Gaussian white noise $W_{\mathrm{Gauss}} \sim \mathcal{N}(0, \mathrm{Identity})$ in $\mathcal{S}'(\mathbb{R}^d)$: stationary, isotropic, self-similar
  - Brownian motion $G_{\mathrm{Wiener}} \sim \mathcal{N}(0, \mathrm{R}_{\mathrm{D}})$ in $\mathcal{S}'(\mathbb{R})$: self-similar, mean-square continuous
    Covariance function: $r_{\mathrm{Wiener}}(x, y) = h_{\mathrm{D}}(x, y) = \frac{1}{2}\big(|x| + |y| - |x - y|\big)$

Classical Gaussian processes and RKHS

- Preliminary observations
  - Domain of $F(\varphi) = \mathrm{e}^{-\frac{1}{2}\|\varphi\|^2_{\mathcal{H}'}}$ can be extended from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{H}'$
  - To recover a classical process on $\mathbb{R}^d$, $\mathcal{H}'$ should include $\delta(\cdot - \boldsymbol{x}_0)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$

Corollary (Equivalence between Gaussian processes and RKHS)

A GSP 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

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

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 $\mathcal{H} \subseteq C_{\mathrm{b},\alpha}(\mathbb{R}^d)$.

Gaussian marginals

Proposition

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, $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu_Y}, \mathbf{C}_{\boldsymbol{Y}})$ with mean vector

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

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

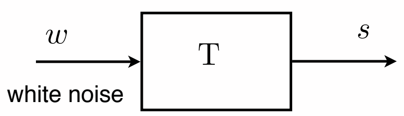
$$\begin{aligned}[\mathbf{C}_{\boldsymbol{Y}}]_{m,n} &= \langle \mathrm{R}_G\{\varphi_m\}, \varphi_n\rangle = \langle \varphi_m, \varphi_n\rangle_{\mathcal{H}'} \\ &= \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \varphi_m(\boldsymbol{x}) r_G(\boldsymbol{x}, \boldsymbol{y})\varphi_n(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.\end{aligned}$$

4.5 Gaussian solutions of SDE

Adjoint pair of continuous linear operators: $\mathrm{T} : \mathcal{S}'(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ and $\mathrm{T}^* : \mathcal{S}(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$

- Linear transformation of a white noise

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



$$S \sim \mathcal{N}(\mu_G, \mathrm{TT}^*)$$

Generation of Gaussian process with factorizable covariance operator: $\mathrm{R}_S = \mathrm{TT}^*$

- Innovation model = stochastic differential equation

$$\mathrm{L}s = w \qquad \Rightarrow \qquad s = \mathrm{L}^{-1}w$$

Coercivity hypothesis: Continuity of $\mathrm{L}^{-1*} = \mathrm{T}^* : \mathcal{S}(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$

Operators with non-trivial null space

L is **spline-admissible** with finite-dimensional null space $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$

Biorthogonal boundary functionals $\boldsymbol{\phi} : \mathcal{H}'_{\mathrm{L}}(\mathbb{R}^d) \to \mathbb{R}^{N_0}$ s.t. $\boldsymbol{\phi}(p_n) = \mathbf{e}_n$

- Solution of linear stochastic differential equation

Imposing N_0 boundary conditions

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

Generalization:

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

a_n: realizations of independent Gaussian variables A_n with zero mean and variance σ_n^2

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

Explicit solutions of linear SDE

w: realization of white Gaussian noise

a_n: realizations of independent Gaussian random variables $A_n \sim \mathcal{N}(0, \sigma_n^2)$

$$\mathrm{L}s = w \quad \text{s.t.} \quad \boldsymbol{\phi}(s) = (a_1, \dots, a_{N_0}) \qquad \Rightarrow \qquad s = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + \sum_{n=1}^{N_0} a_n p_n$$

- Characterization of underlying stochastic process

$$S = \mathrm{L}_{\boldsymbol{\phi}}^{-1} W_{\mathrm{Gauss}} + \sum_{n=1}^{N_0} A_n p_n \quad \sim \mathcal{N}(0, \mathrm{R}_S)$$

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

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

Covariance operator: $\mathrm{R}_S = \mathrm{A}_{\boldsymbol{\phi}} + \sum_{n=1}^{N_0} \sigma_n^2 \mathrm{P}_{p_n}$ with $\mathrm{P}_u : \varphi \mapsto u \langle u, \varphi \rangle$

$\Rightarrow$ same form as in Section 2 on RKHS !!!!

4.6 MMSE solution of linear inverse problems

Linear measurement model: $s \mapsto \mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n} \in \mathbb{R}^M$

- Hypotheses
  - the unknown signal $s = S(\omega) \in \mathcal{S}'(\mathbb{R}^d)$ is a realization of a **stochastic process** S;
  - $r_S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ 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 $\mathbb{E}\left\{\big(S(\boldsymbol{x}) - \mu_S(\boldsymbol{x})\big)\big(S(\boldsymbol{y}) - \mu_S(\boldsymbol{y})\big)\right\} = r_S(\boldsymbol{x},\boldsymbol{y})$;
  - $\boldsymbol{\nu} : s \mapsto \boldsymbol{\nu}(s) = (\langle \nu_1, s\rangle, \dots, \langle \nu_M, s\rangle)$ 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.

MMSE estimator at location $\boldsymbol{x}$

Generalized Gauss-Markov theorem
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, $\boldsymbol{\nu}^* = (\nu_1^*, \dots, \nu_M^*)$ 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.

$$\nu_m^*(\boldsymbol{x}) = \int_{\mathbb{R}^d} r_S(\boldsymbol{x},\boldsymbol{y})\nu_m(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$$

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

Proof of generalized Gauss-Markov theorem

Linear measurement model: $s \mapsto \mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n} \in \mathbb{R}^M$

- MMSE solution: $s_{\mathrm{MMSE}}(\boldsymbol{x}|\mathbf{y}) = \mathbb{E}\{s(\boldsymbol{x})|\mathbf{y}\} \quad \Rightarrow$ determination of $p\big(s(\boldsymbol{x})|\mathbf{y}\big)$

- Distribution of $\boldsymbol{\nu}(S)$ (see Gaussian marginals theorem)

  $\Rightarrow \quad \boldsymbol{\nu}(S) \sim \mathcal{N}(\boldsymbol{\nu}(\mu_S), \mathbf{G})$ with $\mathbf{G} \in \mathbb{R}^{M \times M}$

  where $[\mathbf{G}]_{m,n} = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \nu_m(\boldsymbol{x}) r_S(\boldsymbol{x},\boldsymbol{y}) \nu_n(\boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \quad \Rightarrow \quad \mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n} \ \sim \ \mathcal{N}(\boldsymbol{\nu}(\mu_S), \mathbf{G} + \sigma_0^2 \mathbf{I}_M)$

- Joint pdf of $\boldsymbol{Z} = (s(\boldsymbol{x}), \mathbf{y}) \ \sim \ \mathcal{N}(\mathbf{m}_{\boldsymbol{Z}}, \mathbf{C}_{\boldsymbol{Z}})$ with $\mathbf{m}_{\boldsymbol{Z}} = \big(\mu_S(\boldsymbol{x}), \boldsymbol{\nu}(\mu_S)\big)$,

$$\mathbf{C}_{\boldsymbol{Z}} = \begin{pmatrix} r_S(\boldsymbol{x},\boldsymbol{x}) & \boldsymbol{\nu}^*(\boldsymbol{x})^T \\ \boldsymbol{\nu}^*(\boldsymbol{x}) & \mathbf{C}_{\boldsymbol{Y}} \end{pmatrix} \text{ where } \boldsymbol{\nu}^*(\boldsymbol{x}) = \begin{pmatrix} \nu_1^*(\boldsymbol{x}) \\ \vdots \\ \nu_M^*(\boldsymbol{x}) \end{pmatrix}$$

  with $\nu_m^*(\boldsymbol{x}) = \mathbb{E}\left\{S(\boldsymbol{x}) Y_m\right\} = \int_{\mathbb{R}^d} r_S(\boldsymbol{x},\boldsymbol{y}) \nu_m(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$

- Bayes rule + algebra $\Rightarrow p\big(s(\boldsymbol{x})|\mathbf{y}\big) = p_{\boldsymbol{Z}}(\mathbf{z})/p_{\boldsymbol{Y}}(\mathbf{y})$ 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 $\sigma^2_{s(\boldsymbol{x})|\mathbf{y}} = r_S(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{\nu}^*(\boldsymbol{x})^T (\mathbf{G} + \sigma_0^2 \mathbf{I})^{-1} \boldsymbol{\nu}^*(\boldsymbol{x})$

Equivalence with variational solution

- Case of zero-mean signal

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

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

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

- Formal equivalence with smoothing spline problem
  - $\mathcal{H}$: RKHS induced by covariance function $r_S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$
  - R_S: covariance operator is the Riesz map $\mathcal{H}' \to \mathcal{H}$
  - $\lambda = \sigma_0^2$: optimal choice of regularization parameter

$$s_{\mathrm{MMSE}}(\cdot|\mathbf{y}) = \arg\min_{f \in \mathcal{H}} \left(\sum_{m=1}^{M} \big|y_m - \langle \nu_m, f\rangle\big|^2 + \lambda \|f\|_{\mathcal{H}}^2 \right)$$

Exact discretization: $\nu_m^* = \mathrm{R}_S\{\nu_m\}$ and $[\mathbf{G}]_{m,n} = \langle \nu_m, \, \nu_n^* \rangle = \langle \nu_m^*, \nu_n^* \rangle_{\mathcal{H}}$