Lección n°1: Introducción y herramientas de base : procesos gaussianos, núcleos y RKHS

AMARUN 2024

1. Introducción

El modelamiento numérico de EDPs se realiza a través de distintas técnicas matemáticas de acuerdo al tipo de ecuación que se estudia. Existen varias particularidades a cada una de ellas, de tal manera que es necesario presentar métodos adaptados a la estructura, dinámica y dimensión de cada clase de ecuaciones. A manera general, entre los métodos deterministas más usados para simular soluciones de EDPs se encuentran los elementos finitos y volúmenes finitos. Estos métodos, si bien aseguran un alto grado de precisión de la solución final, son muy rígidos para adaptarse a distintas soluciones y muy costosos de un punto de vista computacional.

Por otro lado, existen métodos probabilísticos, como la regresión por procesos gaussianos (*Gaussian Process Regression*, GPR, o también llamada *kriging*, en inglés), que permiten aproximar soluciones de manera más intuitiva y con cálculos relativamente directos en comparación a métodos basados en mallas. Además, son más flexibles desde un punto de vista de adaptabilidad y permiten obtener modelos de distinto grado de fidelidad para los fenómenos de estudio. Dentro de las aplicaciones en ingeniería, pueden ser usados inclusive en la cuantificación de incertitudes, ya que existe acceso al cálculo de covarianzas condicionadas dentro de la estimación.

Así, el objetivo principal de este curso es presentar el cuadro matemático adecuado (principalmente basado en la teoría de espacios de Hilbert a núcleo reproductor) para tratar problemas de aproximación de EDP, a través de un enfoque probabilístico usando regresión guassiana. El contenido de estas lecciones se encuentra en la interfaz entre la teoría de probabilidades, análisis funcional y análisis de EDPs. Las principales referencias son [1, 2, 3, 4, 5, 6].

En esta lección, se presentan las definiciones y propiedades elementales sobre las funciones núcleo y los espacios de Hilbert a núcleo reproductor (RKHS). En la Lección 2, se estudiarán propiedades de estos espacios que permitirán establecer una equivalencia entre las funciones núcleo y los RKHS. De esta manera, se podrá manipular las propiedades de las funciones dentro del RKHS a partir de la definición de la función núcleo asociada. En la Lección 3, se detallarán los métodos de regresión gaussiana dentro de estos espacios para finalmente ejemplificar el uso de estas técnicas en el planteamiento de problemas de aproximación numérica de EDPs en la Lección 4.

2. Preliminares

2.1. Espacios de Lebesgue y de Sobolev

Dentro del estudio de las EDPs es fundamental la noción de integración y diferenciabilidad. Por ello es importante recordar la definición de los espacios de Lebesgue y de Sobolev ya que permiten clasificar las funciones de acuerdo a su nivel de regularidad, sin pedir que las funciones sean necesariamente continuas.

Para $p \in [1, +\infty[$, considerando la medida de Lebesgue dx, se define el espacio de Lebesgue en un conjunto $\Omega \subseteq \mathbb{R}^d$ como

$$L^p(\Omega) = \left\{ f : \Omega \longrightarrow \mathbb{R}^d, \ \textit{medible} \quad \middle| \quad \int_\Omega |f(x)|^p \, dx < +\infty \right\},$$

es decir el espacio de funciones medibles *p–integrables*. Estos espacios son normados y completos bajo la norma $\|f\|_{L^p(\Omega)} = \left(\int_\Omega |f(x)|^p \, dx \right)^{1/p}$. El espacio $L^2(\Omega)$ es además un espacio de Hilbert, dotado del producto interno definido por

$$\langle f, g \rangle_{L^2(\Omega)} = \int_\Omega f(x) g(x) \, dx, \quad f, g \in L^2(\Omega).$$

Así mismo, se pueden considerar condiciones no solamente sobre la función si no también sobre su derivada (en sentido distribucional) de la misma. Se dice entonces que una función u localmente integrable en Ω admite una derivada *distribucional* o *débil* en la dirección de x_j si se cumple la siguiente condición :

$$\text{existe } g \in L^1_{\text{loc}}(\Omega) \text{ tal que para todo } \varphi \in \mathcal{C}_c^\infty(\Omega) \ : \quad \int_\Omega u(x) \frac{\partial}{\partial x_j} \varphi(x) \ dx = - \int_\Omega g(x) \varphi(x) \ dx,$$

donde se dice que g es una derivada *distribucional* o *débil* de u, y la notaremos como $\partial_{x_j} u$ o $\dfrac{\partial}{\partial x_j} u$. Considerando ahora un multi-índice $\alpha = (\alpha_1 \ \cdots \ \alpha_d)$, la derivada distribucional de u con respecto a α (o de orden α) viene definida con la condición :

$$\text{existe } g \in L^1_{\text{loc}}(\Omega) \text{ tal que para todo } \varphi \in \mathcal{C}_c^\infty(\Omega) \ : \quad \int_\Omega u(x) \partial^\alpha \varphi(x) \ dx = (-1)^{|\alpha|} \int_\Omega g(x) \varphi(x) dx,$$

donde notaremos $\partial^\alpha \varphi = \dfrac{\partial^{|\alpha|}}{\partial x_1 \cdots \partial x_d} \varphi$. Por definición, la derivada distribucional es única y coincide con la derivada en el sentido clásico (o derivada *fuerte*) de la función.

Una vez presentada la noción de regularidad *débil*, se puede definir los espacios de Sobolev de orden $k \in \mathbb{N}$ y parámetro de integración $p \in [1, +\infty[$ como

$$W^{k,p}(\Omega) = \left\{ \ f \in L^p(\Omega) \quad \middle| \quad \partial^\alpha f \in L^p(\Omega) \ \text{ para todo multi-índice } \alpha \text{ tal que } \ |\alpha| \leq k \ \right\}.$$

En el caso particular $p = 2$, estos espacios se denotan como $H^k(\Omega) = W^{k,2}(\Omega)$ y son espacios de Hilbert con el producto interno definido por

$$\langle f, g \rangle_{H^k(\Omega)} = \langle f, g \rangle_{L^2(\Omega)} + \sum_{|\alpha| < k} \langle \partial^\alpha f, \partial^\alpha g \rangle_{L^2(\Omega)} \, .$$

Dependiendo de la dimensión del dominio (e incluso su geometría), existe cierto grado de regularidad que pueden tener estos espacios. En particular, cuando $k > d/p$, se tiene la *inyección* de Sobolev sobre el espacio de funciones continuas :

$$W^{k,p}(\Omega) \subset \mathcal{C}^m(\overline{\Omega}), \quad \text{con} \quad m = \left\lfloor k - \frac{d}{2p} \right\rfloor .$$

Para revisar más resultados de inyecciones en otros espacios se puede revisar la sección 9.3 de [7].

Estos espacios funcionales pueden generalizarse según la medida que se utilice en lugar de la medida de Lebesgue, como se presenta a continuación en la siguiente sección.

Más lecciones sobre los espacios de Sobolev pueden encontrarse en : *Curso de Verano Amarun 2021 (EDP)*, *Curso de Verano Amarun 2013 (Espacios Funcionales)*.

2.2. Procesos gaussianos

Considérese un espacio de probabilidades $(\mathcal{S}, \mathcal{A}, \mathbb{P})$[1]. Una función medible a valores reales de la forma $X : \mathcal{S} \longrightarrow \mathbb{R}$ se dice que es una *variable aleatoria* (v.a.). A cada variable aleatoria se puede asociar una *esperanza*, que viene definida como

$$\mathbb{E}(X) = \int_S X(\omega) \ d\mathbb{P}(\omega).$$

En el caso de que la variable aleatoria admita una función de densidad de probabilidad f_X, se tiene que $\mathbb{E}(X) = \displaystyle\int_{\mathbb{R}} x f_X(x) \ dx$, y para una función real g, se tiene que

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) \ dx.$$

[1]El espacio muestral se denota como $\mathcal{S}$, la σ-álgebra asociada como $\mathcal{A}$, y la medida de probabilidad como $\mathbb{P}$.

En este sentido, se pueden considerar así mismo espacios de variables aleatorias integrables con respecto a su medida de probabilidad. Más prácticamente, definimos el espacio de variables integrables de orden $p \in]1, +\infty[$ como

$$L^p(\mathbb{P}) = \left\{ X \quad v.a. \quad \middle| \quad \mathbb{E}(|X|^p) < +\infty \right\}.$$

Siendo así, considerando dos variables aleatorias reales X e Y, tales que sus momentos de segundo orden sean finitos, es decir tal que $X, Y \in L^2(\mathbb{P})$, su *covarianza* viene definida como

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right).$$

En particular, tenemos que la *varianza* de X viene definida como

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

Seguidamente, un proceso estocástico $G = (G_x)_{x \in \Omega}$ es un conjunto de variables aleatorias indexado en algún conjunto de índices, no necesariamente finito o numerable, que notaremos como Ω (usualmente en aplicaciones, puede ser un subconjunto de $\mathbb{R}^+$, $\mathbb{R}$, $\mathbb{R}^d$ o incluso $\mathbb{R}^+ \times \mathbb{R}^d$).

Además, se dice que este proceso es un proceso *gaussiano* si para todo subconjunto finito de elementos *índices* $x_1, x_2, \ldots, x_N \in \Omega$, el vector aleatorio $(G_{x_1}, G_{x_2}, \ldots, G_{x_N})$ es un vector gaussiano, es decir un vector que sigue alguna ley de probabilidad normal multi-variada $\mathcal{N}(\mu, \mathbf{\Sigma})$.

Para definir entonces completamente un proceso gaussiano, es necesario establecer un valor para la esperanza de cada variable indexada en el proceso y el valor de la covarianza entre todas los pares de variables indexadas. Así, consideremos una función de esperanza (o *media*) μ y una función de covarianza K. Notaremos $G \sim \mathcal{GP}(\mu, K)$ al proceso gaussiano G tal que se cumpla que

$$\begin{array}{rcl} \mu : \Omega & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \mu(x) = \mathbb{E}(G_x), \end{array} \qquad \text{y que} \qquad \begin{array}{rcl} K : \Omega \times \Omega & \longrightarrow & \mathbb{R} \\ (x, y) & \longmapsto & K(x, y) = \mathrm{Cov}(G_x, G_y), \end{array}$$

es decir que para cada conjunto de índices $x_1, \ldots, x_n \in \Omega$ se tenga que

$$\begin{pmatrix} G_{x_1} \\ \vdots \\ G_{x_n} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu(x_1) \\ \vdots \\ \mu(x_n) \end{pmatrix} , \mathbf{\Sigma} \right), \quad \text{con} \quad \mathbf{\Sigma} = \left(K\left(x_i, x_j\right)\right)_{1 \leq i,j \leq n}.$$

donde también se suele notar $\mu(X) = (\mu(x_1), \ldots, \mu(x_n))^\intercal$ y $\Sigma = K(X, X)$ para $X = (x_1, \ldots, x_n)^\intercal$. En este caso, la función de densidad de probabilidad del vector gaussiano está definida por la fórmula

$$f_{G_{x_1}, \ldots, G_{x_n}}(z_1, \ldots, z_n) = \frac{1}{(2\pi)^{n/2} \det(K(X, X))^{1/2}} \exp\left(\frac{1}{2} \left[(Z - \mu(X))^\intercal K(X, X)^{-1} (Z - \mu(X)) \right] \right),$$

para $Z = (z_1, \ldots, z_n)^\intercal$. Para que el proceso gaussiano esté bien definido, se puede observar que la función μ no requiere de más propiedades en particular. Sin embargo, para la función K, es necesario pedir ciertas otras condiciones a verificar de manera que las propiedades por definición de la covarianza se mantengan. A una función de este tipo la llamaremos función *núcleo* y su estudio sera sujeto de la siguiente sección.

Más contenido sobre variables aleatorias puede encontrarse en : *Curso de Verano Amarun 2013 (Probabilidades, en inglés).*

3. Funciones núcleo y espacios de Hilbert a núcleo reproductor (RKHS)

Definición 1 (Función núcleo) *Sea $\Omega \subset \mathbb{R}^d$. Una función $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ es una función* núcleo *(o* kernel*, en inglés) si es simétrica y además semi-definida positiva (s.d.p.), es decir que para todo conjunto de puntos $x_1, \ldots, x_n \in \Omega$ la matriz $K(X, X)$ sea s.d.p.*

En la definición anterior, para que la matriz $K(X, X)$ sea s.d.p., se tiene que cumplir que

$$\text{para todo } \beta = (\beta_1, \ldots, \beta_n)^\intercal \in \mathbb{R}^n \ : \quad \beta^\intercal K(X, X) \beta = \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j K(x_i, x_j) \geq 0. \tag{1}$$

Ejercicio 3.1 *Sea K la función de covarianza de un proceso gaussiano $G = (G(x))_{x\in\Omega} \sim \mathcal{GP}(\mu, K)$. Muestre que K cumple las condiciones de una función núcleo, y que por lo tanto la Definición 1 está bien planteada.*

En el ejercicio siguiente se pueden verificar algunas propiedades útiles de las funciones núcleo.

Ejercicio 3.2 *Sean K_1 y K_2 dos funciones núcleo sobre $\Omega \times \Omega$, muestre que :*

1. *para todo $\alpha \in \mathbb{R}^+$, la función αK_1 es una función núcleo;*
2. *la suma $K_1 + K_2$ es una función núcleo;*
3. *la multiplicación $K_1 K_2$ es una función núcleo;*
4. *para todo mapa $\phi : \mathcal{D} \longrightarrow \Omega$, la función compuesta K_ϕ definida por $K_\phi(s,t) = K_1(\phi(s), \phi(t)),\ s,t \in \mathcal{D}$, es una función núcleo.*

Además, se puede extender la noción de función núcleo para una función vectorial o matricial. En particular, este tipo de funciones núcleo serán de importancia al modelar variables como la velocidad o aceleración en varias dimensiones.

Observación 3.1 (Recurso digital) *Para comprender el comportamiento de las funciones núcleo y los procesos gaussianos de una manera visual, y revisar algunos ejemplos de funciones núcleo comúnmente utilizadas, se puede consultar el recurso digital en el enlace : Görtler et al. (2019). A Visual Exploration of Gaussian Processes*

Por otro lado, resulta muy útil revisar la relación entre una función núcleo y los espacios de Hilbert. En realidad, una función de este tipo puede ser considerada como un precursor de un producto interno. Esto además permite extender los resultados presentados para núcleos de la forma $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ hacia núcleos de la forma $K : \Omega \times \Omega \longrightarrow \mathbb{R}^n$.

Observación 3.2 *Para toda función núcleo K, es posible definir un espacio de Hilbert $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ y un mapa $Z : \Omega \longrightarrow \mathcal{H}$ tal que su producto interno satisfaga*

$$K(x,y) = \langle Z(x), Z(y)\rangle_{\mathcal{H}}, \quad x,y \in \Omega.$$

Este mapa puede ser, por ejemplo, un proceso gaussiano. En la Lección 2 se podrá ver más en detalle un resultado que explora la equivalencia (en un cierto sentido) de la relación entre funciones núcleos y ciertos espacios funcionales de Hilbert. Sin embargo, antes de ello, es necesario revisar algunas definiciones más precisas.

Definición 2 (Núcleo reproductor) *Sea $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ un espacio de Hilbert de funciones (a valores reales) que tienen como dominio $\Omega \subset \mathbb{R}^d$. Se dice que una función núcleo $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ es un* núcleo reproductor *(n.r.) de $\mathcal{H}$ si se cumple que para todo $x \in \Omega$:*

1. *la función marginal $K(x,\cdot) : \Omega \longrightarrow \mathbb{R}$ pertenece a $\mathcal{H}$; y que*
2. *para todo $f \in \mathcal{H}$: $\langle f, K(x,\cdot)\rangle_{\mathcal{H}} = f(x)$.*

La propiedad 2. es llamada también como propiedad de reproducción *de la función núcleo asociada a $\mathcal{H}$. Un espacio de Hilbert $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ que admita una función núcleo es llamado un* espacio de Hilbert a núcleo reproductor *(o* Reproducing Kernel Hilbert Space*, RKHS, como comúnmente conocido en inglés).*

Así, la relación entre los elementos de un espacio de Hilbert y las funciones núcleo queda más claramente establecida a través de su relación con el producto interno. Es decir, la evaluación de una función del espacio en un punto fijo puede accederse mediante la acción del producto interno en un elemento particular, el cual se obtiene a través de una evaluación parcial de la función núcleo. De esta manera, se *trasladan* las propiedades de la función núcleo K a las funciones elementos del espacio de Hilbert $\mathcal{H}$.

Finalmente, veamos una caracterización de los RKHS que nos sera útil en su manipulación.

Proposición 3.1 (Caracterización de los RKHS) *Un espacio de Hilbert $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ de funciones (a valores reales) con dominio en $\Omega \subset \mathbb{R}^d$ es un RKHS si y solamente si se cumple que para todo $x \in \Omega$, la distribución de Dirac δ_x es continua, con respecto a la topología de $\mathcal{H}$.*

Demostración. ($\Longrightarrow$) Por las propiedades de definición de un RKHS, se tiene directamente que gracias a la desigualdad de Cauchy-Schwartz, para todo $f \in \mathcal{H}$ y $x \in \Omega$,

$$|\langle \delta_x, f\rangle| = |\langle f, K(x,\cdot)\rangle_{\mathcal{H}}| \leq \|K(x,\cdot)\|_{\mathcal{H}} \|f\|_{\mathcal{H}},$$

y dado que $\|K(x,\cdot)\|_{\mathcal{H}} < +\infty$, se tiene que la norma en el sentido de los operadores de δ_x sobre $\mathcal{H}$ es finita. Por lo tanto, δ_x es continua en este sentido.

($\Longleftarrow$) Dado que el funcional lineal δ_x sobre $\mathcal{H}$ es continuo, por el teorema de representación de Riesz, existe un elemento, que notaremos como $h_x \in \mathcal{H}$ tal que éste representa la acción de δ_x a través del producto interno, es decir para cualquier $f \in \mathcal{H}$,

$$\langle f, h_x\rangle_{\mathcal{H}} = \langle \delta_x, f\rangle = f(x).$$

Gracias a la unicidad de este elemento h_x (el cual en realidad es una función) para cada uno de los $x \in \Omega$, se puede definir una función $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ tal que $K(x,y) = h_x(y)$, con $x, y \in \Omega$. Es directo verificar además, nuevamente por la unicidad de h_x, que la función candidata K es bien un núcleo reproductor de $\mathcal{H}$, y por tanto es un RKHS. ■

Ejercicio 3.3 *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de un espacio RKHS $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ convergente en el sentido de la norma de $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$. Demuestre que esta sucesión converge puntualmente.*

Algunos ejemplos de RKHS se presentan a continuación.

Ejercicio 3.4 *Muestre que todos los espacios de Hilbert de dimensión finita son RKHS y, a partir de una base ortonormal del espacio, encuentre una expresión para su núcleo reproductor.*

Ejercicio 3.5 *Para $T > 0$, considérese el espacio de Sobolev* anclado

$$H_0^1(0,T) = \left\{ f \in L^2(0,T) \quad \middle| \quad \frac{d}{dx} f \in L^2(0,T) \text{ y que } f(0) = 0 \right\}.$$

1. *Justifique el hecho de que para todo $f \in H_0^1(0,T)$ se tiene que $f(x) = \displaystyle\int_0^x \frac{d}{dx} f(y)\, dy$.*

2. *Verifique si el espacio de Hilbert $\left(H_0^1(0,T), \langle\cdot,\cdot\rangle_{H_0^1(0,T)}\right)$, con $\langle f, g\rangle_{H_0^1(0,T)} = \displaystyle\int_0^T \frac{d}{dx} f(x) \frac{d}{dx} g(x)\, dx$, es un RKHS. En caso de serlo, explicite una función núcleo que sea núcleo reproductor de $H_0^1(0,T)$.*

Ejercicio 3.6 *Estudie si el espacio $L^2(\mathbb{R})$ es un RKHS.*

A manera general, bajo ciertas condiciones de regularidad de la frontera de $\Omega \subset \mathbb{R}^d$, el espacio $H^k(\mathbb{R}^d)$ es un RKHS cuando $k > d/2$ (véase la Sección 1.3 de [6]).

Ahora bien, la caracterización que hemos visto permite definir un RKHS sin necesidad de considerar explícitamente una función núcleo, pues ésta puede ser definida directamente con el uso del teorema de representación de Riesz. Sin embargo, la interrogante natural que sigue : es posible definir un RKHS a partir de una función núcleo ? Esto será justamente lo que intentaremos responder en la siguiente lección.

Referencias

[1] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in Probability and Statistics.* Springer Science & Business Media, 2004.

[2] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447:110668, 2021.

[3] Iain Henderson, Pascal Noble, and Olivier Roustant. Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 29(4):3396–3422, 2023.

[4] Iain Henderson. Sobolev regularity of Gaussian random fields. *Journal of Functional Analysis*, 286(3):110241, 2024.

[5] Olivier Roustant. Metamodeling with Gaussian processes. September 2023. Lecture notes.

[6] Saburou Saitoh and Yoshihiro Sawano. *Theory of reproducing kernels and applications*. Springer Singapore, 2016.

[7] Haïm Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer Science & Business Media, 2011.

Lección n°2: Algunas propiedades de los RKHS AMARUN 2024

En la lección anterior, se presentó la definición de los espacios de Hilbert a núcleo reproductor o RKHS (resumida en la condición de que " $\delta_x \in \mathcal{H}$ ", para todo $x \in \mathcal{H}$), así como algunos ejemplos de estos espacios. Se puso en evidencia además la función de "núcleo reproductor" y algunas de sus propiedades.

En esta lección, nuestro interés es estudiar algunas propiedades más profundas de estos espacios funcionales. Primero, veremos los detalles de la equivalencia entre funciones núcleo y RKHS. Después, presentaremos condiciones para caracterizar la regularidad (diferenciabilidad) de estos espacios.

1 Teorema de Moore–Aronszajn

Teorema 1.1 (Moore–Aronszajn) *Sea K una función núcleo que toma valores en $\Omega \times \Omega$. Existe un único RKHS, denotado como $\mathcal{H}_K$, que posee a K como su respectivo núcleo reproductor.*

Demostración. La demostración se la realizará de manera constructiva, es decir, definiremos un espacio funcional a partir de la función núcleo, verificaremos que este espacio cumpla con las condiciones de un RKHS y estableceremos su unicidad.

Consideremos el conjunto de funciones $\{K(x,\cdot)\}_{x\in\Omega}$ y estudiemos el conjunto $\mathcal{H}_0 = \operatorname{span}\{K(x,\cdot)\}_{x\in\Omega}$ de todas las combinaciones lineales de la forma $\displaystyle\sum_{i=1}^{n} \alpha_i K(x_i,\cdot)$. Definamos ahora una forma bilineal sobre funciones de $\mathcal{H}_0$ de tal forma que podamos dotarlo de un producto interno. Consideremos para $f = \displaystyle\sum_{i=1}^{n} \alpha_i K(x_i,\cdot)$ y $g = \displaystyle\sum_{j=1}^{m} \beta_j K(z_j,\cdot)$ de $\mathcal{H}_0$,

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \beta_j K(x_i, z_j)$$

Ejercicio 1.1 *Muestre que la definición de esta forma bilineal es independiente de las descomposiciones lineales de f y de g, y por tanto esta bien definida.*

Ejercicio 1.2 *Muestre que esta forma bilineal satisface la propiedad de reproducción de funciones del espacio $\mathcal{H}_0$, es decir que :*

$$\text{para todo } f \in \mathcal{H}_0 \quad \text{y para todo } x \in \Omega : \quad \langle f, K(x,\cdot)\rangle_{\mathcal{H}_0} = f(x). \tag{1}$$

En particular, se tiene para todo $x, y \in \Omega$ que $\langle K(x,\cdot),\ K(y,\cdot)\rangle_{\mathcal{H}_0} = K(x,y)$.

Ejercicio 1.3 *Verifique que la forma bilineal $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ es una forma simétrica y semi-positiva definida.*

Ejercicio 1.4 *Verifique que la desigualdad de Cauchy-Schwartz es válida para esta forma bilineal (la cual se asume que no necesariamente es definida positiva en el sentido estricto), es decir que :*

$$\text{para todo } f,\ g \in \mathcal{H}_0 \ : \quad \langle f, g\rangle_{\mathcal{H}_0} \leq \langle f, f\rangle_{\mathcal{H}_0}^{1/2} \langle g, g\rangle_{\mathcal{H}_0}^{1/2}. \tag{2}$$

Una vez que hemos vistas las propiedades precedentes de la forma bilineal $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$, para ver que esta forma es bien un producto interno, nos resta verificar que en realidad también es estricta definida positiva.

Consideramos entonces un $f \in \mathcal{H}_0$ tal que $\langle f, f\rangle_{\mathcal{H}_0} = 0$. Gracias a (1) y (2), tenemos para cualquier $y \in \Omega$,

$$|f(y)| = \left|\langle f, K(y,\cdot)\rangle_{\mathcal{H}_0}\right| \leq \langle f, f\rangle_{\mathcal{H}_0}^{1/2} \langle K(y,\cdot),\ K(y,\cdot)\rangle_{\mathcal{H}_0}^{1/2} = \langle f, f\rangle_{\mathcal{H}_0}^{1/2} K(y,y)^{1/2} = 0, \tag{3}$$

y por lo tanto f es la función nula, lo cual implica que $\left(\mathcal{H}_0,\ \langle\cdot,\cdot\rangle_{\mathcal{H}_0}\right)$ es un espacio dotado de producto interno $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$. Notaremos la norma inducida sobre $\mathcal{H}_0$ por este producto interno como $\|\cdot\|_{\mathcal{H}_0}$.

Ahora bien, una vez que hemos dotado a este espacio de un producto interno, podemos definir tomar la clausura topológica con respecto a $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$:

$$\mathcal{H}_K := \overline{\mathcal{H}_0}^{\|\cdot\|_{\mathcal{H}_0}}. \tag{4}$$

Se puede verificar que la extensión por continuidad de la topología inducida por $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ es única, y por tanto bien definida. Además, como se había mencionado, el espacio $\mathcal{H}_K$ está conformado por todas las funciones $f \in \mathcal{H}_0$ para las cuales existe respectivamente una sucesión $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_0$ que converge en el sentido de la norma, y por lo tanto también puntualmente, hacia f.

Finalmente, nos resta a ver que el espacio $\mathcal{H}_K$ cumple las propiedades de definición de un RKHS, y que en efecto es el único RKHS con núcleo reproductor asociado K.

Primero, tenemos por definición de $\mathcal{H}_K$, que este espacio es cerrado y completo para la topología inducida por el producto interno (extendido por continuidad a $\mathcal{H}_K$). Por lo tanto, nos hace falta verificar, usando la caracterización de los RKHS de la Lección 1, que para todo $x \in \Omega$, los funcionales lineales δ_x son continuos en $\mathcal{H}_K$.

Sea $f \in \mathcal{H}_0$. Para cualquier $x \in \Omega$, gracias a la propiedad de reproducción de funciones en $\mathcal{H}_0$ dada por (1) y a la desigualdad de Cauchy-Schwartz sobre $\mathcal{H}_0$, se tiene que

$$|\langle \delta_x, f \rangle| = \left|\langle f, K(x,\cdot)\rangle_{\mathcal{H}_0}\right| \leq K(x,x)^{1/2} \|f\|_{\mathcal{H}_0}^{1/2}. \tag{5}$$

Dado que $K(x,x) < +\infty$ en todo $x \in \Omega$, se tiene entonces que el funcional lineal $\delta_x : \mathcal{H}_0 \longrightarrow \mathbb{R}$ es continuo. Por definición de $\mathcal{H}_K$ se tiene además la continuidad de la propiedad de reproducción sobre todas las funciones de $\mathcal{H}_K$. En efecto, para una función f de $\mathcal{H}_K$, existe una sucesión $(f_n)_{n\in\mathbb{N}}$ en $\mathcal{H}_0$ que converge a f en el sentido de la norma $\|\cdot\|_{\mathcal{H}_0}$. Por continuidad de la extensión, se cumple para cualquier $x \in \Omega$,

$$f(x) = \lim_{n\to+\infty} f_n(x) = \lim_{n\to+\infty} \langle f_n, K(x,\cdot)\rangle_{\mathcal{H}_0} = \langle f, K(x,\cdot)\rangle_{\mathcal{H}_K},$$

es decir que por los mismos razonamientos anteriores, (5) es válida para $f, g \in \mathcal{H}_K$. Esto concluye la demostración de que K es por tanto el núcleo reproductor de $\mathcal{H}_K$.

Finalmente, veamos que este espacio $\mathcal{H}_K$ es único. Para ello, suponemos la existencia de otro espacio RKHS $\mathcal{H}'_K$ tal que K sea su núcleo reproductor. Por definición y completitud de este espacio, tenemos que $\mathcal{H}_0$ es denso también en $\mathcal{H}'_K$.

Ejercicio 1.5 *Muestre que para toda función h en $\mathcal{H}_0$ se tiene que*

$$\|h\|_{\mathcal{H}_K} = \|h\|_{\mathcal{H}'_K}. \tag{6}$$

Gracias al último ejercicio, se tiene que la convergencia en ambas normas de los RKHS es la misma y por lo tanto, todas las sucesiones de funciones en $\mathcal{H}_0$ que convergen hacia una función en $\mathcal{H}_K$, convergen también hacia una función en $\mathcal{H}'_K$, que resulta ser la misma (por la convergencia puntual para todo $x \in \Omega$). Con esto se establece que $\mathcal{H}_K = \mathcal{H}'_K$, y se concluye la demostración del Teorema 1.1.

■

Gracias a este teorema, basta con definir y estudiar directamente una función núcleo para poder establecer un espacio funcional con propiedades específicas que se pueda utilizar en problemas de modelamiento.

Ejercicio 1.6 *La demostración del Teorema 1.1 puede realizarse considerando una definición más general del RKHS propuesto. En lugar de usar* (4)*, defínase a $\mathcal{H}_K$ como el conjunto :*

$$\mathcal{H}_K := \left\{ g \ : \ \text{existe una sucesión de Cauchy } (g_n)_{n\in\mathbb{N}} \text{ en } \mathcal{H}_0 \text{ tal que } \lim_{n\to+\infty} g_n = g \text{ puntualmente} \right\},$$

y demuestre que :

1. *toda sucesión de Cauchy $(f_n)_{n\in\mathbb{N}}$ en $\mathcal{H}_0$ que converge puntualmente a 0, converge también en el sentido de la norma $\|\cdot\|_{\mathcal{H}_0}$ hacia $\mathbf{0} \in \mathcal{H}_0$;*

2. *siendo* $f, g \in \mathcal{H}_K$ *con respectivas sucesiones de Cauchy* $(f_n)_{n\in\mathbb{N}}$ *y* $(g_n)_{n\in\mathbb{N}}$ *en* $\mathcal{H}_0$ *que convergen puntualmente, la sucesión real* $\left(\langle f_n, g_n\rangle_{\mathcal{H}_0}\right)_{n\in\mathbb{N}}$ *también es convergente (esto nos permite definir directamente un producto interno en el espacio* $\mathcal{H}_K$*);*

3. *para todo* $g \in \mathcal{H}_K$*, la sucesión de Cauchy asociada (que converge puntualmente) converge también en el sentido de la norma* $\|\cdot\|_{\mathcal{H}_0}$ *hacia* g*.*

Este análisis complementario permite demostrar de una manera alternativa el Teorema 1.1, partiendo esta vez de un punto de vista más general en la definición propuesta para el espacio $\mathcal{H}_K$ *(véase el Teorema 3 de la Sección 1.3 de [1]).*

2 Regularidad de Sobolev de los RKHS

La regularidad de espacios funcionales es un concepto elemental en el estudio teórico y numérico de las EDPs. La ventaja que se tiene al modelar EDPs usando RKHS es que se puede dar cierto grado de regularidad a los RKHS a partir solamente de la definición de su núcleo reproductor.

Recientemente en [2], se demuestra una caracterización analítica para este tipo de regularidad en los RKHS. En particular, se tiene el siguiente teorema.

Teorema 2.1 (Regularidad de Sobolev) *Sea* Ω *un conjunto abierto de* $\mathbb{R}^d$ *y* K *una función núcleo sobre* $\Omega \times \Omega$*. Se tiene que el RKHS* $\mathcal{H}_K$ *asociado a* K *es un subconjunto del espacio de Sobolev* $H^m(\Omega)$*, con* $m \in \mathbb{N}$*, si y solamente si para todo multi-índice* α *tal que* $|\alpha| \leq m$*, la derivada cruzada en el sentido de las distribuciones*[1] $\partial_x^\alpha \partial_y^\alpha K$ *pertenece a* $L^2(\Omega \times \Omega)$*.*

Para la demostración de este teorema, es necesario un paso intermedio mediante una descomposición de Mercer del núcleo K y de sus derivadas. Este resultado es la herramienta central para poder caracterizar la regularidad de las derivadas de una manera adecuada, por medio de funciones propias (o *eigenfunctions*, en inglés) del núcleo, que a su vez permitirán encontrar el vínculo directo con la regularidad en el RKHS.

Proposición 2.1 (Descomposición de Mercer para funciones núcleo) *Sea* K *una función núcleo continua sobre* $\Omega \times \Omega$ *y perteneciente a* $L^2(\Omega \times \Omega)$*. Entonces, existe una sucesión de valores propios reales no–negativos* $(\lambda_n)_{n\in\mathbb{N}}$ *y una sucesión de vectores propios* $(\phi_n)_{n\in\mathbb{N}}$*, que constituyen una base ortonormal de* $L^2(\Omega)$*, y son tales que*

$$\text{para todo } x, y \in \Omega: \quad K(x,y) = \sum_{n=0}^{+\infty} \lambda_n \phi_n(x)\phi_n(y), \tag{7}$$

en donde la serie converge en el sentido de $L^2(\Omega)$*.*

La demostración de este resultado puede revisarse en [3] (véase el Teorema 2.30). Sin embargo, nuestro interés radica principalmente en la descomposición de las derivadas del núcleo como se presenta a continuación.

Proposición 2.2 (Derivadas de una función núcleo) *En el escenario de la proposición anterior, fijando* $m \in \mathbb{N}$ *y un multi-índice* α *tal que* $|\alpha| \leq m$*, la derivada distribucional cruzada* $\partial_x^\alpha \partial_y^\alpha K$ *pertenece a* $L^2(\Omega \times \Omega)$ *si y solamente si para todo* $n \in \mathbb{N}$*, la derivada* $\partial^\alpha \phi_n$ *está en* $L^2(\Omega)$ *y además*

$$\sum_{n=0}^{+\infty} \lambda_n \left\|\partial^\alpha \phi_n\right\|^2_{L^2(\Omega)} < +\infty. \tag{8}$$

En este caso se tiene que

$$\text{para todo } x, y \in \Omega: \quad \partial_x^\alpha \partial_y^\alpha K(x,y) = \sum_{n=0}^{+\infty} \lambda_n \partial^\alpha \phi_n(x) \partial^\alpha \phi_n(y), \tag{9}$$

en donde la serie converge en el sentido de $L^2(\Omega)$*.*

[1] La notación ∂_x^α y ∂_y^α hace referencia a la derivada parcial con respecto a la primera y segunda variable de K, respectivamente.

Observación 2.1 *La demostración de* (8) *no es trivial y se la puede consultar en [2] (véase la Proposición 4.4). No obstante, asumiendo esta convergencia, la astucia importante para verificar la fórmula* (9) *viene dada gracias a la separación de variables de la función núcleo que brinda la descomposición de Mercer en la Proposición 2.1.*

En efecto, veamos que para todo $n \in \mathbb{N}$*, tal que* $\lambda_n \neq 0$*, la función* $\partial^\alpha \phi_n$ *pertenece a* $L^2(\Omega)$*. Para ello, utilizaremos una caracterización variacional de la función* ϕ_n*. Para una función regular de prueba* $\varphi \in \mathcal{C}_c^\infty(\Omega)$ *cualquiera, se tiene que por el teorema de Fubini e integrando por partes,*

$$\begin{aligned}\lambda_n \left(\int_\Omega \phi_n(x) \partial^\alpha \varphi(x) \, dx \right)^2 &\leq \sum_{k=0}^{+\infty} \int_{\Omega\times\Omega} \lambda_k \phi_n(x)\phi_n(y) \partial^\alpha \varphi(x) \partial^\alpha \varphi(y) \, dxdy \\ &\leq \int_{\Omega\times\Omega} \partial_x^\alpha \partial_y^\alpha K(x,y)\varphi(x)\varphi(y) \, dxdy \\ &\leq \left\| \partial_x^\alpha \partial_y^\alpha K \right\|_{L^2(\Omega\times\Omega)} \|\varphi\|^2_{L^2(\Omega)} ,\end{aligned}$$

y por tanto

$$\int_\Omega \phi_n(x) \partial^\alpha \varphi(x) \, dx \leq C_{n,\alpha} \|\varphi\|_{L^2(\Omega)} , \qquad con \qquad C_{n,\alpha} = \left(\frac{1}{\lambda_n} \left\| \partial_x^\alpha \partial_y^\alpha K \right\|_{L^2(\Omega\times\Omega)} \right)^{1/2} < +\infty,$$

lo cual implica que $\partial^\alpha \phi_n$ *esta representada en un sentido distribucional por un elemento de* $L^2(\Omega)$*.*

Una vez vistas estas herramientas, se puede dar paso a la demostración del teorema de regularidad de Sobolev de los RKHS con la astucia de la separación de variables de la función núcleo.

Demostración (del Teorema 2.1). En esta lección, nos interesaremos solamente en la dirección ($\Longrightarrow$). Siendo $m \in \mathbb{N}$ fijo, se tiene que para todo multi-índice α tal que $|\alpha| \leq m$, las fórmulas (7) y (9) de las Proposiciones 2.1 y 2.2 son válidas. Nuestro objetivo es poder utilizar estas fórmulas para trasladar la propiedad de regularidad desde la función núcleo hasta las funciones del RKHS $\mathcal{H}_K$.

Para ello, considerando un elemento $h \in \Omega$, observemos la relación de reproducción con respecto al operador de diferencias finitas $\Delta_h : \mathcal{H}_K \longrightarrow \mathcal{H}_K$ definido como

$$\Delta_h[f](x) = f(x+h) - f(x), \quad f \in \mathcal{H}_K, \ x \in \Omega.$$

Tenemos que para todo $g \in \mathcal{H}_K$ y $x \in \Omega$,

$$\Delta_h[g](x) = g(x+h) - g(x) = \langle f, K(x+h,\cdot) - K(x,\cdot) \rangle_{\mathcal{H}_K} = \left\langle g, \Delta_h' K(x,\cdot) \right\rangle_{\mathcal{H}_K} ,$$

en donde Δ_h' es el operador de diferencias finitas que actúa sobre la primera variable de K. En este sentido, si se consideran sucesivos operadores de diferencias finitas para $h_1, \ldots, h_j \in \Omega$, notando $\Delta_{h_1,\ldots,h_j} = \Delta_{h_1} \circ \Delta_{h_2} \circ \ldots \circ \Delta_{h_j}$,

$$\Delta_{h_1,\ldots,h_j}[f](x) = \left\langle \Delta_{h_1,\ldots,h_j}[f], \ K(x,\cdot) \right\rangle_{\mathcal{H}_K} = \left\langle f, \ \Delta'_{h_1,\ldots,h_j} K(x,\cdot) \right\rangle_{\mathcal{H}_K} .$$

A continuación, para un $f \in \mathcal{H}_K$ genérico, se tiene que nuevamente por la propiedad de reproducción y la desigualdad de Cauchy-Schwartz en $\mathcal{H}_K$,

$$\left\| \Delta_{h_1,\ldots,h_j}[f] \right\|^2_{L^2(\Omega)} = \int_\Omega \left| \Delta_{h_1,\ldots,h_j}[f](x) \right|^2 \, dx \leq \int_\Omega \|f\|^2_{\mathcal{H}_K} \left\| \Delta'_{h_1,\ldots,h_j} K(x,\cdot) \right\|^2_{\mathcal{H}_K} \, dx \tag{10}$$

en donde se puede notar que a su vez

$$\begin{aligned}\left\| \Delta'_{h_1,\ldots,h_j} K(x,\cdot) \right\|^2_{\mathcal{H}_K} &= \left\langle \Delta'_{h_1,\ldots,h_j} K(x,\cdot), \ \Delta'_{h_1,\ldots,h_j} K(x,\cdot) \right\rangle_{\mathcal{H}_K} \\ &= \Delta_{h_1,\ldots,h_j}[\Delta'_{h_1,\ldots,h_j} K](x,x) \\ &= \sum_{n=0}^{+\infty} \lambda_n \left| \Delta_{h_1,\ldots,h_j}[\phi_n](x) \right|^2 .\end{aligned}$$

Aquí, $\Delta_{h_1,\ldots,h_j}$ actúa en el segundo argumento de K y se considera la descomposición dada por la fórmula (9).

Finalmente, volviendo a la desigualdad (10), obtenemos que :

$$\begin{aligned}\left\|\Delta_{h_1,\ldots,h_j}[f]\right\|^2_{L^2(\Omega)} &\leq \|f\|^2_{\mathcal{H}_K} \sum_{n=0}^{+\infty} \lambda_n \int_\Omega \left|\Delta_{h_1,\ldots,h_j}[\phi_n](x)\right|^2 \, dx \\ &\leq \|f\|^2_{\mathcal{H}_K} \sum_{n=0}^{+\infty} \lambda_n \left\|\Delta_{h_1,\ldots,h_j}[\phi_n]\right\|^2_{L^2(\Omega)} \cdot |h_1|^2 \cdots |h_j|^2\end{aligned}$$

dado que $\phi_n \in H^m(\Omega)$. Por la caracterización por diferencias finitas del espacio $H^m(\Omega)$, se puede concluir que $f \in H^m(\Omega)$ y que $\mathcal{H}_K \subset H^m(\Omega)$. ■

Observación 2.2 *El resultado del Teorema 2.1 sobre la regularidad en los espacios de Sobolev $H^m(\Omega)$ puede también generalizarse a los espacios de Sobolev $W^{k,p}(\Omega)$ (véase la sección 3 de [2]) .*

Referencias

[1] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in Probability and Statistics*. Springer Science & Business Media, 2004.

[2] Iain Henderson. Sobolev regularity of Gaussian random fields. *Journal of Functional Analysis*, 286(3):110241, 2024.

[3] Saburou Saitoh and Yoshihiro Sawano. *Theory of reproducing kernels and applications*. Springer Singapore, 2016.

Lección n°3: Optimización y RKHS en la regresión con procesos gaussianos AMARUN 2024

1. Introducción

La *regresión con procesos gaussianos* o *regresión gaussiana* (GPR, por sus siglas en inglés) es una técnica de estimación para modelar relaciones entre variables en situaciones donde los datos observados son ruidosos, escasos o cuando la estructura subyacente es compleja, como en el caso de las EDPs. En comparación con otros métodos tradicionales de regresión que asumen formas específicas para las funciones estimadas (como lineal o polinómica), la GPR utiliza un enfoque basado en distribuciones de probabilidad normales. Esta técnica es especialmente útil en problemas de estimación de incertidumbre, ya que no solo proporciona una interpolación puntual, sino también una medida de confianza asociada a cada predicción.

Sus aplicaciones más extendidas abarcan campos como el aprendizaje automático, optimización bayesiana, geofísica, robótica y bio–informática [1, 2, 3]. En esta lección, se presentan las nociones clásicas de definición y estimación a partir de GPR. Sin embargo, es preciso remarcar que su utilización dentro del modelamiento de EDPs no requiere necesariamente de la utilización de observaciones reales de un fenómeno asociado a la ecuación. Los avances más recientes tienen como objetivo intervenir justamente dentro de la aproximación de soluciones de EDPs en un sentido de simulación numérica de fenómenos físicos.

2. Regresión con procesos gaussianos

Antes de revisar los conceptos asociados a la GPR, algunos resultados importantes de recordar sobre las variables aleatorias normales que nos serán de utilidad vienen dados en los siguientes ejercicios.

Ejercicio 2.1 *Sean X, Z dos variables aleatorias conjuntamente distribuidas en una distribución normal multivariada. Muestre que se tiene entonces la equivalencia entre los siguientes puntos :*

1. *X y Z son variables aleatorias independientes entre si;*
2. $\mathrm{Cov}(X, Z) = 0$.

Ejercicio 2.2 *Sean X, Z dos variables aleatorias. Verifique las siguientes propiedades con respecto a la esperanza condicional :*

1. $\mathbb{E}(Z|Z) = Z$;
2. $\mathbb{E}\left(\mathbb{E}(X|Z)\right) = \mathbb{E}(X)$;
3. *si X y Z son variables independientes, entonces* $\mathbb{E}(X|Z) = \mathbb{E}(X)$.

El objetivo de esta lección es formular la GPR desde un punto de vista de teoría de probabilidades y otro desde el análisis funcional, los RKHS y la optimización. Así, estudiando la relación entre estos dos enfoques, se establecerá un camino suficiente para plantear los problemas de modelamiento de EDPs en la siguiente lección.

2.1. Enfoque probabilístico

Sean X y Z dos variables aleatorias conjuntamente distribuidas en una distribución normal multivariada

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Z \end{pmatrix}, \Sigma\right), \quad \text{con} \quad \Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Z) \\ \mathrm{Cov}(Z,X) & \mathrm{Var}(Z) \end{pmatrix}. \tag{1}$$

Suponiendo que se tiene acceso a la variable Z, es decir que sus realizaciones se pueden medir o que pueden ser objeto de una variable auxiliar dentro de un proceso de modelamiento, nuestro interés es de poder

simular la variable aleatoria $\mathbb{E}(X|Z)$. Más precisamente, vamos a querer encontrar una fórmula explícita para $\mathbb{E}(X \mid Z = w)$ cuando se conoce el valor de $w \in \mathbb{R}$. Este procedimiento es posible gracias a la hipótesis (1), como se detalla en la Proposición 2.1. Sin embargo, si no se tiene esta condición de normalidad para X y Z, el cálculo de la esperanza condicional puede ser significativamente complicado, y en varias aplicaciones, se lo realiza de manera numérica.

Observación 2.1 *Siendo $\{X_1, \ldots, X_N\}$ un conjunto de variables aleatorias, cuyos valores de realizaciones pueden ser medidos. Para establecer una fórmula de regresión clásica, se puede revisar de igual manera el valor obtenido de la esperanza condicional de Y conociendo las variables $X_1, \ldots, X_N$. Si en lugar de considerar a $\mathbb{E}(Y|X_1, \ldots, X_N)$ como la proyección de Y sobre $X_1, \ldots, X_N$ en el sentido de $L^2(\mathbb{P})$, se considera la proyección sobre el espacio de funciones obtenidas a partir de combinaciones lineales afines de $X_1, \ldots, X_N$, se obtiene que*

$$\mathbb{E}_L(Y|X_1, \ldots, X_N) = \beta_0 + \beta^\intercal X = \beta_0 + \sum_{i=1} \beta_i X_i,$$

donde $\beta_i \in \mathbb{R}$, $i = 0, 1, \ldots, N$, son coeficientes que representan la proyección de Y en este espacio funcional. Esta proyección se la notará como $\mathbb{E}_L$, y no siempre coincide con la esperanza condicional usual.

La siguiente proposición describe explícitamente a la esperanza condicional de X, conociendo Z, en el caso de variables normales distribuidas conjuntamente.

Proposición 2.1 *Sean X, Z dos variables aleatorias conjuntamente distribuidas en una distribución normal multivariada, según el enunciado* (1)*. Considerando $\mathbb{E}(X|Z)$ la proyección de Y sobre la variable Z, en el sentido de $L^2(\mathbb{P})$, se tiene que :*

$$\mathbb{E}(X|Z) = \mu_X + \mathrm{Cov}(X, Z) \cdot \mathrm{Var}(Z)^{-1} \cdot (Z - \mu_Z)\,. \tag{2}$$

Nótese que $E(X|Z)$ es una variable aleatoria, de tal manera que, para un valor fijo $w \in \mathbb{R}$,

$$\mathbb{E}(X|Z = w) = \mu_X + \mathrm{Cov}(X, Z) \cdot \mathrm{Var}(Z)^{-1} \cdot (w - \mu_Z)$$

es un número real. Siendo así, verifiquemos la fórmula (2).

Demostración de la Proposición 2.1 . Definamos la variable aleatoria $Y = X - cZ$, donde $c \in \mathbb{R}$ es determinista. Nuestro interés es conocer el valor de c para que esta variable no sea correlacionada con Z. Para ello, se debe cumplir que, por la bi-linealidad de la covarianza,

$$0 = \mathrm{Cov}(Y, Z) = \mathrm{Cov}(X, Z) - c\,\mathrm{Var}(Z).$$

De manera que podemos fijar $c = \mathrm{Cov}(X, Z) \cdot \mathrm{Var}(Z)^{-1}$ (en el caso no trivial que $\mathrm{Var}(Z) \neq 0$) y afirmar que Y y Z son no correlacionadas y conjuntamente distribuidas en una distribución normal. Por ello, gracias a los Ejercicios 2.1 y 2.2, son variables independientes y se tiene que $\mathbb{E}(Y|Z) = \mathbb{E}(Y)$. Finalmente, por la linealidad de la esperanza condicional,

$$\begin{aligned}
\mathbb{E}(X|Z) &= \mathbb{E}(Y + cZ|Z) \\
&= \mathbb{E}(Y|Z) + c\mathbb{E}(Z|Z) \\
&= \mathbb{E}(Y) + cZ \\
&= \mathbb{E}(X - cZ) + cZ \\
&= \mathbb{E}(X) + c(Z - \mathbb{E}(Z)) \\
&= \mu_X + \mathrm{Cov}(X, Z) \cdot \mathrm{Var}(Z)^{-1} \cdot (Z - \mu_Z)\,.
\end{aligned}$$

■

Como corolario de esta fórmula de esperanza condicional, se tiene la fórmula denominada de *regresión gaussiana* utilizada en las aplicaciones de métodos basados en funciones núcleo.

Corolario 2.1 *Sea G un proceso gaussiano indexado en Ω y distribuido según $\mathcal{GP}(\mu, K)$. Para $x, y \in \Omega$ y $w \in \mathbb{R}$, se tiene que la esperanza de $G(x)$ condicionada en que $G(y) = w$ está dada por*

$$\mathbb{E}(G(x)|G(y) = w) = \mu(x) + K(x,y)K(y,y)^{-1}\left(w - \mu(y)\right). \tag{3}$$

En este punto es importante notar que una vez que se fija la variable aleatoria $G(y) = w$ con un valor real $w \in \mathbb{R}$, la función $h : \Omega \longrightarrow \mathbb{R}$ definida por

$$h(x) = \mathbb{E}(G(x)|G(y) = w), \quad x \in \Omega,$$

pertenece al RKHS $\mathcal{H}_K$ asociado a la función núcleo K si y solamente si la función media μ también pertenece a este espacio $\mathcal{H}_K$. Nótese también que si $x = y$, se obtiene exactamente la condición de *interpolación*

$$\mathbb{E}(G(y)|G(y) = w) = w.$$

Ahora bien, en un caso más general, dado el vector $Y = (y_1, \ldots, y_N)^\intercal$ de puntos de Ω y el vector de valores reales $W = (w_1, \ldots, w_N)^\intercal$. Para cualquier $x \in \Omega$, se tiene que la esperanza de $G(x)$ condicionada en $G(y_i) = w_i$, para todo $i = 1, \ldots, N$, viene dada como una generalización de la fórmula (3) en un sentido matricial :

$$\mathbb{E}(G(x)|\ G(Y) = W\ ) = \mu(x) + K(x,Y)K(Y,Y)^{-1}\left(W - \mu(Y)\right), \tag{4}$$

en donde $K(x,Y)^\intercal = (K(x,y_j))_{1\leq j\leq N}$ y $\mu(Y) = (\mu(y_j))_{1\leq j\leq N}$ son vectores columna y $K(Y,Y) = (K(y_i,y_j))_{1\leq i,j\leq N}$ es un matriz de orden $N \times N$.

Esta notación puede entenderse también como una extensión del operador de covarianza hacia valores matriciales

$$\mathbb{E}(G(x)|\ G(Y) = W\ ) = \mathbb{E}(G(x)) + \mathrm{Cov}(G(x), G(Y))\,\mathrm{Var}(G(Y))^{-1}\left(W - \mathbb{E}(G(Y))\right),$$

donde, más explícitamente, se tiene que

$$\begin{pmatrix} G(x) \\ G(Y) \end{pmatrix} = \begin{pmatrix} G(x) \\ G(y_1) \\ \vdots \\ G(y_N) \end{pmatrix} \sim \mathcal{N}\left(\mu\begin{pmatrix} x \\ Y \end{pmatrix},\ \Sigma\right) = \mathcal{N}\left(\begin{pmatrix} \mu(x) \\ \mu(y_1) \\ \vdots \\ \mu(y_N) \end{pmatrix},\ \Sigma\right),$$

con

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \mathrm{Var}(G(x)) & \mathrm{Cov}(G(x), G(Y)) \\ \mathrm{Cov}(G(Y), G(x)) & \mathrm{Var}(G(Y)) \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Var}(G(x)) & \mathrm{Cov}(G(x), G(y_1)) & \cdots & \mathrm{Cov}(G(x), G(y_N)) \\ \mathrm{Cov}(G(y_1), G(x)) & \mathrm{Var}(G(y_1)) & \cdots & \mathrm{Cov}(G(y_1), G(y_N)) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(G(y_N), G(x)) & \mathrm{Cov}(G(y_N), G(y_1)) & \cdots & \mathrm{Var}(G(y_N)) \end{pmatrix} \\
&= \begin{pmatrix} K(x,x) & K(x,y_1) & \cdots & K(x,y_N) \\ K(y_1,x) & K(y_1,y_1) & \cdots & K(y_1,y_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(y_N,x) & K(y_N,y_1) & \cdots & K(y_N,y_N) \end{pmatrix}.
\end{aligned}$$

Observación 2.2 *Haciendo un alcance a la Observación 2.1, se puede verificar que el el caso de considerar variables normales conjuntamente distribuidas, las definiciones planteadas de $\mathbb{E}(G(x)|G(Y))$ y $\mathbb{E}_L(G(x)|G(Y))$, sí coinciden. En efecto, de manera similar a la fórmula (4), se tiene que*

$$\mathbb{E}(G(x)|G(Y)) = \beta_0 + \sum_{j=1}^{N} \beta_j G(y_j), \qquad con \quad \beta_0 = \mu(x) - K(x,Y)K(Y,Y)^{-1}\mu(Y),$$

y con $\beta_j = \left(K(x,Y)K(Y,Y)^{-1}\right)_j$ como la j–ésima componente del vector fila $K(x,Y)K(Y,Y)^{-1}$, para $j = 1, \ldots, N$.

Asimismo, se pueden plantear incluso condiciones que dependan de un operador lineal aplicado al proceso gaussiano. Por ejemplo, siendo $\mathcal{L}_y$ un operador lineal que actúa sobre el índice del proceso gaussiano G, se tiene que

$$\operatorname{Cov}(G(x), \mathcal{L}_y G(y)) = \mathcal{L}'_y \operatorname{Cov}(Gx, G(y)),$$

en donde $\mathcal{L}'_y$ esta vez es un operador lineal que actual con respecto a la segunda variable del operador de covarianza.

De esta manera, considerando por ejemplo un operador diferencial $\mathcal{L} = \dfrac{d}{dx}$ (donde x hace referencia a la variable índice del proceso gaussiano G), se tiene que la distribución conjunta de $G(z)$ y $\dfrac{d}{dx}G(y)$ viene dada como

$$\begin{pmatrix} G(z) \\ \dfrac{d}{dx}G(y) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu(z) \\ \dfrac{d}{dx}\mu(y) \end{pmatrix}, \begin{pmatrix} K(z,z) & \dfrac{\partial}{\partial y}K(z,y) \\ \dfrac{\partial}{\partial x}K(y,z) & \dfrac{\partial}{\partial x}\dfrac{\partial}{\partial_y}K(z,y) \end{pmatrix} \right),$$

en donde $\dfrac{\partial}{\partial x}$ y $\dfrac{\partial}{\partial_y}$ son dos operadores lineales que hacen referencia a las derivadas parciales con respecto a la primera y a la segunda variable de K, respectivamente. En este sentido, se muestra en evidencia la relación entre los elementos del proceso gaussiano y las propiedades de regularidad de su función de covarianza asociada. Así, el manejo de las propiedades de una variable de interés se traslada a la definición particular de cada función núcleo y de la forma de las condiciones que intervienen en la esperanza condicional.

2.2. Enfoque funcional : Teorema del Representante

Ahora bien, el interés principal de usar la GPR para modelar ecuaciones diferenciales es justamente el acceso a fórmulas explícitas de cálculo que se pueden obtener como aquellas dadas en la Proposición 2.1 y el Corolario 2.1. El objetivo en esta sección será de estudiar su relación con problemas de optimización en RKHS.

Consideremos una variable de interés $u : \Omega \longrightarrow \mathbb{R}$, para la cual conocemos un conjunto de puntos $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset \Omega \times \mathbb{R}$ en los cuales se tiene que

$$u(x_i) = y_i + \varepsilon, \quad i = 1, \ldots, N,$$

con $\varepsilon \sim \mathcal{N}(0, \lambda)$, modelado como un error de observación normal. Dada una función núcleo $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ que define un RKHS denotado como $\mathcal{H}_K$, nuestro interés es encontrar una función dentro de dicho espacio funcional que minimice el funcional objetivo dado por la fórmula

$$J_{\mathcal{D},\lambda}(u) = \sum_{i=1}^{N} (u(x_i) - y_i)^2 + \lambda \|u\|^2_{\mathcal{H}_K}. \tag{5}$$

La respuesta a esta interrogante viene dada en el siguiente teorema y muestra además que el problema de optimización en un espacio funcional de dimensión infinita puede estar *representado* por un problema de minimización solamente sobre $\mathbb{R}^N$.

Teorema 2.1 (del Representante) *Dada la información $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset \Omega \times \mathbb{R}$, $\lambda > 0$, y una función núcleo $K : \Omega \times \Omega \longrightarrow \mathbb{R}$, con RKHS asociado $\mathcal{H}_K$, se tiene que*

$$\min_{u \in \mathcal{H}_K} J_{\mathcal{D},\lambda}(u) = \min_{z \in \mathbb{R}^N} \widetilde{J_{\mathcal{D},\lambda}}(z) \tag{6}$$

donde el funcional $J_{\mathcal{D},\lambda}$ se define como en (5), y $\widetilde{J_{\mathcal{D},\lambda}}$ se define a partir de $J_{\mathcal{D},\lambda}$.

Demostración. La estrategia para transformar el problema de minimización en $\mathcal{H}_K$ de (6) en un problema en dimensión finita es usar las propiedades de representación de las funciones en $\mathcal{H}_K$ para separar el funcional $J_{\mathcal{D},\lambda}$ en varias partes.

En este sentido, consideremos el espacio vectorial $V = \operatorname{span}\{K(\cdot, x_i)\}_{i=1}^N \subset \mathcal{H}_K$. Este espacio es cerrado en el sentido topológico con respecto a la norma $\|\cdot\|_{\mathcal{H}_K}$ y se tiene la descomposición de $\mathcal{H}_K$ como

$$\mathcal{H}_K = V \oplus V^{\intercal},$$

con $V^\intercal$, el complemento ortogonal a V en $\mathcal{H}_K$ el sentido de su producto interno. Así, toda función $u \in \mathcal{H}_K$ puede ser descompuesta como $u = h + g$ con $h \in V$ y $g \in V^\intercal$.

Ahora, gracias a la propiedad de reproducción del RKHS y a la ortogonalidad de $V^\intercal$, se tiene que para cada punto $x_i \in \Omega$, $i = 1, \ldots, N$,

$$g(x_i) = \langle g, K(\cdot, x_i) \rangle_{\mathcal{H}_K} = 0.$$

Y además, nuevamente por ortogonalidad,

$$\|u\|^2_{\mathcal{H}_K} = \|h\|^2_{\mathcal{H}_K} + \|g\|^2_{\mathcal{H}_K}.$$

Esto nos permite observar que

$$J_{\mathcal{D},\lambda}(u) = \sum_{i=1}^{N} (h(x_i) + g(x_i) - y_i)^2 + \lambda \left(\|h\|^2_{\mathcal{H}_K} + \|g\|^2_{\mathcal{H}_K} \right) = J_{\mathcal{D},\lambda}(h) + \lambda \|g\|^2_{\mathcal{H}_K},$$

y por lo tanto, el problema de optimización puede ser separado,

$$\min_{u \in \mathcal{H}_K} J_{\mathcal{D},\lambda}(u) = \min_{\substack{u = h+g \in \mathcal{H}_K, \\ h \in V,\ g \in V^\intercal}} \left(J_{\mathcal{D},\lambda}(h) + \lambda \|g\|^2_{\mathcal{H}_K} \right) = \min_{h \in V} J_{\mathcal{D},\lambda}(h) + \min_{g \in V^\intercal} \lambda \|g\|^2_{\mathcal{H}_K} = \min_{h \in V} J_{\mathcal{D},\lambda}(h),$$

siendo representado por un problema en un espacio de dimensión finita.

Para finalizar la demostración, hace falta ver que el funcional $J_{\mathcal{D},\lambda}$ restringido a V tiene en efecto una forma reducida correspondiente al problema del lado derecho de (6). Para ello, denotemos $X = (x_1, \ldots, x_N)^\intercal$ e $Y = (y_1, \ldots, y_N)^\intercal$ y con la misma notación matricial para K que la sección anterior, se tiene que un elemento $h \in V$ puede escribirse como

$$h = \sum_{j=1}^{N} z_j K(\cdot, x_j) = K(\cdot, X) z.$$

para algún $z \in \mathbb{R}^N$. Re-escribiendo el funcional $J_{\mathcal{D},\lambda}$ en términos de z, usando una vez más las propiedades de reproducción de $\mathcal{H}_K$,

$$\begin{aligned} J_{\mathcal{D},\lambda}(h) &= \sum_{i=1}^{N} (h(x_i) - y_i)^2 + \lambda \|h\|^2_{\mathcal{H}_K} \\ &= \sum_{i=1}^{N} (K(x_i, X) z - y_i)^2 + \lambda \|K(\cdot, X) z\|^2_{\mathcal{H}_K} \\ &= (K(X, X) z - Y)^\intercal (K(X, X) z - Y) + \lambda \cdot z^\intercal K(X, X) z \\ &=: \widetilde{J_{\mathcal{D},\lambda}}(z) \end{aligned}$$

con $\widetilde{J_{\mathcal{D},\lambda}}$ actuando sobre $\mathbb{R}^N$ en lugar de V. ■

Para caracterizar un mínimo de este funcional, podemos usar una condición de nulidad de su derivada total.

Ejercicio 2.3 *Sea $J : \mathbb{R}^N \longrightarrow \mathbb{R}$ el funcional definido por $J(z) = (Az - Y)^\intercal (Az - Y) + \lambda \cdot z^\intercal Az$, donde $A \in \mathbb{R}^{N \times N}$ es una matriz simétrica, $Y \in \mathbb{R}^N$ y $\lambda \in \mathbb{R}$. Verifique que su derivada total (o derivada en el sentido de Fréchet) viene dada por*

$$DJ(z)\beta = 2 \left[(A + \lambda I) z - Y \right]^\intercal A\beta, \qquad \text{para } z, \beta \in \mathbb{R}^N. \tag{7}$$

De esta manera, para un elemento minimal $\widehat{z} \in \mathbb{R}^N$, la condición de que $D\widetilde{J_{\mathcal{D},\lambda}}(\widehat{z})\beta = 0$ para todo $\beta \in \mathbb{R}^N$ es equivalente a que la matriz $D\widetilde{J_{\mathcal{D},\lambda}}(\widehat{z})$ sea la matriz nula. De acuerdo con la fórmula (7), se tiene que

$$\begin{gathered} D\widetilde{J_{\mathcal{D},\lambda}}(\widehat{z}) = 2 \left[(K(X, X) + \lambda I) \widehat{z} - Y \right]^\intercal K(X, X) = \mathbf{0} \\ \Longrightarrow \quad \widehat{z} = (K(X, X) + \lambda I)^{-1} Y. \end{gathered}$$

Se concluye entonces que el funcional $J_{\mathcal{D},\lambda}$ tiene un mínimo dado por

$$\widehat{h} = K(\cdot, X)\widehat{z} = K(\cdot, X)\left(K(X,X) + \lambda I\right)^{-1} Y, \tag{8}$$

el cual es similar a la fórmula de regresión gaussiana sobre las condiciones $\widehat{h}(x_i) = y_i$. En efecto, si se considera $\lambda = 0$, las condiciones de interpolación $\widehat{h}(x_i) = y_i$ son estrictas y la fórmula (8) coincide con la fórmula de esperanza condicional (4) para un proceso gaussiano centrado. Así, podemos observar a la esperanza condicional como un elemento minimal de un problema de optimización sobre un RKHS.

Observación 2.3 *En la práctica, la inversión de la matriz $K(X,X) + \lambda I$ es más simple computacionalmente que la matriz $K(X,X)$, en particular si existen puntos x_i y x_j con con valores muy similares. Esto tiene una explicación en el modelamiento de la regresión : al permitir un error de observación, no es necesario que la interpolación pase exactamente por la condición, flexibilizando así la definición de la esperanza condicional y* regularizándola *(de ahí que el parámetro λ suela denominarse también como parámetro de* regularización*).*

Desde un punto de vista de optimización, el parámetro λ determina el nivel de importancia de la norma $\|\cdot\|_{\mathcal{H}_K}$ dentro del del problema de minimización. De manera que además de requerir las condiciones de interpolación, también se busca un elemento dentro el espacio funcional que tenga una regularidad minimal. Esto puede relacionarse con la regularidad del espacio RKHS según lo estudiado en la Lección 2.

En la siguiente lección podremos valernos de la fórmula (8) para plantear problemas de minimización en RKHS cuyos elementos minimales sean aproximaciones de soluciones de una EDP.

Referencias

[1] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in Probability and Statistics*. Springer Science & Business Media, 2004.

[2] Olivier Roustant. Metamodeling with Gaussian processes. September 2023. Lecture notes.

[3] Saburou Saitoh and Yoshihiro Sawano. *Theory of reproducing kernels and applications*. Springer Singapore, 2016.

Lección n°4: Modelamiento de ecuaciones lineales y no–lineales AMARUN 2024

En esta lección, se presentan ciertas estrategias de modelamiento de EDPs usando la teoría presentada en las lecciones anteriores sobre los RKHS y la GPR. Esta presentación no es una revisión exhaustiva de los métodos que existen para simular EDPs usando GPR pues éstos han estado en continuo desarrollo durante los últimos años.

1. Modelamiento de EDPs lineales con GPR

Recientemente, existen trabajos como [1], que presentan condiciones necesarias y suficientes sobre una función núcleo para obtener RKHS asociados que satisfagan en cierto sentido (usualmente en el sentido de una *solución débil*) una EDP lineal.

Considérese una función de interés $u : \Omega \longrightarrow \mathbb{R}$ sobre un dominio $\Omega \subset \mathbb{R}^d$. Se pide que esta función satisfaga una EDP lineal homogénea de la forma :

$$\mathcal{L}(u) = 0, \quad \text{donde} \quad \mathcal{L} = \sum_{|\gamma| \leq n} c_\gamma \frac{\partial^{|\gamma|}}{\partial x^\gamma} \tag{1}$$

para todos los multi–índices γ tales que $|\gamma| \leq n$ y coeficientes $c_\gamma \in \mathcal{C}^{|\gamma|}(\Omega)$. Una solución *débil* de esta ecuación viene definida como $\widetilde{u}$ (perteneciente a un espacio funcional adecuado, de manera general $L^1_{loc}(\Omega)$) tal que

$$\text{para toda } \varphi \in \mathcal{C}^\infty_c(\Omega) : \quad \int_\Omega \widetilde{u}(x) \sum_{|\gamma| \leq n} (-1)^{|\gamma|} \frac{\partial^{|\gamma|}}{\partial x^\gamma} \left(c_\gamma(x)\varphi(x)\right) \, dx = 0.$$

Esta solución débil puede ser interpretada como una realización de un proceso gaussiano a condición de que la función de covarianza de este proceso satisfaga la condición débil. En efecto, se tiene el siguiente teorema.

Teorema 1.1 *En el contexto de la ecuación* (1) *y considerando un proceso gaussiano* $G = (G(x))_{x\in\Omega}$ *con función de covarianza* $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ *tal que* $\sigma : x \longmapsto K(x,x)^{1/2}$ *pertenece a* $L^1_{loc}(\Omega)$*, entonces se tiene la equivalencia entre las siguientes afirmaciones :*

1. *las realizaciones de* G *satisfacen la EDP* (1) *en un sentido débil :*

$$\mathbb{P}\left(\mathcal{L}(G) = 0 \text{ en el sentido de las distribuciones }\right) = 1;$$

2. *para todo* $x \in \Omega$*, se tiene que* $\mathcal{L}(K(\cdot, x)) = 0$ *en el sentido débil.*

La demostración de este teorema se la puede revisar en [1] (ver la Proposición 3.5). De esta manera, una vez que se fija un conjunto de observaciones $u(x_i) = y_i$, $i = 1, \ldots, N$, se puede acceder a una fórmula de reconstrucción de u usando GPR a partir de K, como se reviso en la lección anterior.

Tomando el ejemplo particular de [1], considérese la ecuación de onda sobre $\Omega = \mathbb{R}^3 \times [0, \infty[$, con $c > 0$:

$$\begin{cases} \dfrac{1}{c^2}\dfrac{\partial^2}{\partial t^2}u(x,t) - \Delta u(x,t) = 0, & (x,t) \in \Omega, \\ u(x,0) = u_0(x), \quad \dfrac{\partial}{\partial t}u(x,0) = v_0(x), & x \in \mathbb{R}^3. \end{cases}$$

para u la variable de interés escalar[1] en (x,t), u_0 y v_0, condiciones iniciales que pueden venir dadas en un sentido distribucional. La solución general en el sentido distribucional de esta ecuación puede expresarse de manera compacta como

$$\widetilde{u}(x,t) = \left(F_t * v_0\right)(x) + \left(\dot{F} * u_0\right)(x),$$

[1] como por ejemplo, la presión de un gas o el desplazamiento con respecto a una referencia vertical

donde $F_t = \dfrac{\sigma_{ct}}{4\pi c^2 t}$, con σ_{ct} la medida esférica sobre la esfera centrada de radio ct.

Así, la definición de una función núcleo que verifique la ecuación de ondas en el sentido débil viene dada por

$$K((x,t),(y,s)) = \left[(F_t \otimes F_s) * K_2 + \left(\dot{F}_t \otimes \dot{F}_s\right) * K_1\right]$$

donde se considera que u_0 y v_0 vienen modelizadas como procesos gaussianos centrados de funciones de covarianza K_1 y K_2, respectivamente. Con la definición de este núcleo, y dado un conjunto de observaciones $\{X,T,Y\} = \{(x_i,t_i,y_i)\}_{i=1}^N$ tales que $u(x_i,t_i) = y_i$, $i = 1,\ldots,N$, donde u es una solución regular de la ecuación de ondas, la fórmula

$$\widehat{u}(x,t) = K((x,t),(X,T))K((X,T),(X,T))^{-1}Y$$

permite reconstruir la velocidad u a partir de los datos observados $\{X,T,Y\} = \{(x_i,t_i,y_i)\}_{i=1}^N$. En particular, por el Teorema 1.1, se pueden estimar $\widehat{u}(\cdot,0)$ y $\dfrac{\partial}{\partial t}\widehat{u}(\cdot,0)$ como reconstrucciones de u_0 y v_0.

Para profundizar más en el estudio de EDPs lineales usando técnicas de GPR, se puede revisar [1] (ver Sección 1).

2. Modelamiento de EDPs no–lineales con GPR

En contraste con los avances realizados para las EDP lineales, las estrategias de modelamiento para EDPs no–lineales son más limitados. Sin embargo, existen trabajos que han permitido sentar un buen punto de inicio para explorar técnicas de modelamiento asociadas.

A continuación, se presenta el caso de una ecuación elíptica. La generalización de este método hacia otro tipo de EDP no–lineales puede revisarse en [2].

2.1. Una ecuación elíptica

Para un dominio regular acotado $\Omega \subset \mathbb{R}^d$, consideremos una EDP de la forma

$$\begin{cases} -\Delta u(x) + h(u(x)) = f(x), & x \in \Omega \\ u(x) = g(x), & x \in \partial\Omega \end{cases} \tag{2}$$

donde $u : \Omega \longrightarrow \mathbb{R}$ es la variable de interés; $h : \Omega \longrightarrow \mathbb{R}$ es una función no–lineal continua; f es un término independiente continuo, y g es una función de condiciones de frontera en $\partial\Omega$.

La estrategia es encontrar una solución aproximada de la ecuación (2) a través de un problema de optimización con restricciones en puntos de colocación den la EDP sobre un RKHS. En otras palabras, sean K una función núcleo (regular) $K : \overline{\Omega}\times\overline{\Omega} \longrightarrow \mathbb{R}$ con RKHS asociado $\mathcal{H}_K$, $\{x_i\}_{i=1}^{N_\Omega} \subset \Omega$ puntos de colocación en Ω, y $\{\widetilde{x}_k\}_{k=1}^{N_{\partial\Omega}} \subset \partial\Omega$ puntos de colocación en la frontera $\partial\Omega$. Se tiene entonces el problema de optimización

$$\min_{u\in\mathcal{H}_K} \|u\|^2_{\mathcal{H}_K} \quad \text{sujeto a} \quad \begin{cases} -\Delta u(x_i) + \mathcal{A}u(x_i) = f(x_i), & i = 1,\ldots,N_\Omega \\ u(\widetilde{x}_k) = g(\widetilde{x}_k), & k = 1,\ldots,N_{\partial\Omega} \end{cases}. \tag{3}$$

donde la regularidad del RKHS (y por ende de la función núcleo K) depende del grado de la EDP en estudio.

A diferencia de una gran cantidad de EDP lineales, no siempre existen soluciones débiles de EDPs no–lineales disponibles para poder aproximar un mínimo local de un sistema como (3). Por ello, es necesario *simplificarlo* de manera que se pueda hacer uso de las herramientas de las anteriores secciones para poder formular un sistema de optimización más *directo*. En este caso se pedirá que $\mathcal{H}_K \in \mathcal{C}^2\Omega \cap \mathcal{C}(\overline{\Omega})$, lo cual puede satisfacerse directamente a través de condiciones de regularidad en K para enseguida usar el teorema de regularidad de Sobolev presentado en la Lección 3 en conjunto con inyecciones de Sobolev en espacios de funciones continuas.

En este sentido reformulemos (3) en el sistema

$$\min_{\substack{W\in\mathbb{R}^{2\times N_\Omega},\\ z\in\mathbb{R}^{N_{\partial\Omega}} \ \&\ \text{(B)}}} \left(\min_{\substack{u\in\mathcal{H}_K,\\ \&\ \text{(A)}}} \|u\|^2_{\mathcal{H}_K} \right) \tag{4}$$

donde

$$\begin{cases} u(\widetilde{x}_k) = z_k, & \text{para } \ k = 1, \dots, N_{\partial\Omega}, \\ u(x_i) = w_{1,i}, & \\ -\Delta u(x_i) = w_{2,i}, & \text{para } \ i = 1, \dots, N_\Omega, \end{cases} \tag{A}$$

$$\begin{cases} w_{2,i} + h(w_{1,i}) = f(x_i), & \text{para } i = 1, \dots, N_\Omega, \\ z_k = g(\widetilde{x}_i) & \text{para } \ k = 1, \dots, N_{\partial\Omega}. \end{cases} \tag{B}$$

Ejercicio 2.1 *Verifique que los sistemas de optimización* (3) *y* (4) *son sistemas equivalentes.*

De esta forma, considerando el problema de optimización interior del sistema (4), es decir, dados $W \in \mathbb{R}^{2\times N_\Omega}$ y $z \in \mathbb{R}^{N_{\partial\Omega}}$ fijos, considerando

$$\min_{u\in\mathcal{H}_K} \|u\|^2_{\mathcal{H}_K} \quad \text{sujeto a} \quad \begin{cases} u(\widetilde{x}_k) = z_k, & \text{para } \ k = 1, \dots, N_{\partial\Omega}, \\ u(x_i) = w_{1,i}, & \\ -\Delta u(x_i) = w_{2,i}, & \text{para } \ i = 1, \dots, N_\Omega, \end{cases},$$

vemos que de manera similar a las fórmulas obtenidas por el Teorema del Representante en la Lección 3, se puede obtener condiciones sobre el elemento minimal de este problema, tal que, para $x \in \Omega$, este elemento venga dado por

$$\widehat{u}(x) = \widetilde{K}\left(x, X_{\overline{\Omega}}\right) \widehat{K}\left(X_{\overline{\Omega}}, X_{\overline{\Omega}}\right)^{-1} Y, \tag{5}$$

donde

$$X_{\overline{\Omega}} = \begin{pmatrix} \widetilde{X}_{\partial\Omega} \\ X_\Omega \end{pmatrix} = \begin{pmatrix} \widetilde{x}_1 \\ \vdots \\ \widetilde{x}_{N_{\partial\Omega}} \\ x_1 \\ \vdots \\ x_{N_\Omega} \end{pmatrix}, \quad Y = \begin{pmatrix} Z \\ W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_{N_{\partial\Omega}} \\ w_{1,1} \\ \vdots \\ w_{1,N_\Omega} \\ w_{2,1} \\ \vdots \\ w_{2,N_\Omega} \end{pmatrix}, \quad \widetilde{K}\left(x, X_{\overline{\Omega}}\right) = \begin{pmatrix} K(x, \widetilde{x}_1) \\ \vdots \\ K\left(x, \widetilde{x}_{N_{\partial\Omega}}\right) \\ K(x, x_1) \\ \vdots \\ K\left(x, x_{N_\Omega}\right) \\ \Delta_y K(x, x_1) \\ \vdots \\ \Delta_y K\left(x, x_{N_\Omega}\right) \end{pmatrix}^\intercal,$$

$$\widehat{K}\left(X_{\overline{\Omega}}, X_{\overline{\Omega}}\right) = \begin{pmatrix} K\left(\widetilde{X}_{\partial\Omega}, \widetilde{X}_{\partial\Omega}\right) & K\left(\widetilde{X}_{\partial\Omega}, X_\Omega\right) & \Delta_y K\left(\widetilde{X}_{\partial\Omega}, X_\Omega\right) \\ K\left(X_\Omega, \widetilde{X}_{\partial\Omega}\right) & K\left(X_\Omega, X_\Omega\right) & \Delta_y K\left(X_\Omega, X_\Omega\right) \\ \Delta_x K\left(X_\Omega, \widetilde{X}_{\partial\Omega}\right) & \Delta_x K\left(X_\Omega, X_\Omega\right) & \Delta_x \Delta_y K\left(X_\Omega, X_\Omega\right) \end{pmatrix} =$$

$$\begin{pmatrix} K(\widetilde{x}_1, \widetilde{x}_1) & \cdots & K(\widetilde{x}_1, \widetilde{x}_{N_{\partial\Omega}}) & K(\widetilde{x}_1, x_1) & \cdots & K(\widetilde{x}_1, x_{N_\Omega}) & -\Delta_y K(\widetilde{x}_1, x_1) & \cdots & -\Delta_y K(\widetilde{x}_1, x_{N_\Omega}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ K\left(\widetilde{x}_{N_{\partial\Omega}}, \widetilde{x}_1\right) & \cdots & K\left(\widetilde{x}_{N_{\partial\Omega}}, \widetilde{x}_{N_{\partial\Omega}}\right) & K\left(\widetilde{x}_{N_{\partial\Omega}}, x_1\right) & \cdots & K\left(\widetilde{x}_{N_{\partial\Omega}}, x_{N_\Omega}\right) & -\Delta_y K\left(\widetilde{x}_{N_{\partial\Omega}}, x_1\right) & \cdots & -\Delta_y K\left(\widetilde{x}_{N_{\partial\Omega}}, x_{N_\Omega}\right) \\ K(x_1, \widetilde{x}_1) & \cdots & K(x_1, \widetilde{x}_{N_{\partial\Omega}}) & K(x_1, x_1) & \cdots & K(x_1, x_{N_\Omega}) & -\Delta_y K(x_1, x_1) & \cdots & -\Delta_y K(x_1, x_{N_\Omega}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ K\left(x_{N_\Omega}, \widetilde{x}_1\right) & \cdots & K\left(x_{N_\Omega}, \widetilde{x}_{N_{\partial\Omega}}\right) & K\left(x_{N_\Omega}, x_1\right) & \cdots & K\left(x_{N_\Omega}, x_{N_\Omega}\right) & -\Delta_y K\left(x_{N_\Omega}, x_1\right) & \cdots & -\Delta_y K\left(x_{N_\Omega}, x_{N_\Omega}\right) \\ -\Delta_x K(x_1, \widetilde{x}_1) & \cdots & -\Delta_x K(x_1, \widetilde{x}_{N_{\partial\Omega}}) & -\Delta_x K(x_1, x_1) & \cdots & -\Delta_x K(x_1, x_{N_\Omega}) & \Delta_x \Delta_y K(x_1, x_1) & \cdots & \Delta_x \Delta_y K(x_1, x_{N_\Omega}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\Delta_x K\left(x_{N_\Omega}, \widetilde{x}_1\right) & \cdots & -\Delta_x K\left(x_{N_\Omega}, \widetilde{x}_{N_{\partial\Omega}}\right) & -\Delta_x K\left(x_{N_\Omega}, x_1\right) & \cdots & -\Delta_x K\left(x_{N_\Omega}, x_{N_\Omega}\right) & \Delta_x \Delta_y K\left(x_{N_\Omega}, x_1\right) & \cdots & \Delta_x \Delta_y K\left(x_{N_\Omega}, x_{N_\Omega}\right) \end{pmatrix}.$$

donde Δ_x y Δ_y hacen referencia al operador laplaciano actuando sobre la primera y segunda variable de K, respectivamente. Esta expresión permite observar que el elemento minimal es justamente

$$\widehat{u}(x) = \mathbb{E}\left(G(x) \;\middle|\; \begin{matrix} G(\widetilde{x}_k) = z_k, & \text{para } \ k = 1, \dots, N_{\partial\Omega}, \\ G(x_i) = w_{1,i}, & \text{para } \ i = 1, \dots, N_\Omega, \\ -\Delta G(x_i) = w_{2,i}, & \end{matrix} \right),$$

la esperanza condicional en los puntos de colocación sobre el dominio y la frontera, donde $G = (G(x))_{x\in\overline{\Omega}}$ es un proceso gaussiano centrado de función de covarianza K.

Ejercicio 2.2 *En el contexto anterior, describa explícitamente la forma de la distribución normal conjunta del vector gaussiano*

$$\left(\; G(x),\; G(\widetilde{x}_1),\; G(x_1),\; -\Delta G(x_1) \; \right)^{\intercal}.$$

Ejercicio 2.3 *Considérese una función núcleo regular $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ y considérese un proceso gaussiano $G \sim \mathcal{GP}(\mu, K)$, indexado en $x \in \Omega$, tal que $G(x), \dfrac{d}{dx}G(x) \in \mathcal{H}_K$ para todo $x \in \Omega$. Siguiendo las demostraciones Y razonamientos presentados en las Lección 3 y en el presente lección, encuentre y justifique la expresión explícita de*

$$\mathbb{E}\left(\; G(x) \;\middle|\; G(y),\; \frac{d}{dx}G(y),\; \frac{d}{dx}G(z) \; \right),$$

para $y, z \in \Omega$ fijos.

Una vez que se ha encontrado la fórmula (5), se puede observar que

$$\|\widehat{u}\|^2_{\mathcal{H}_K} = Y^{\intercal}\widehat{K}\left(X_{\overline{\Omega}}, X_{\overline{\Omega}}\right)^{-1} Y$$

y por lo tanto, se puede simplificar el problema de optimización (4) de manera que, para $M = N_{\partial\Omega} + 2N_{\Omega}$, se obtiene el problema

$$\min_{Y \in \mathbb{R}^M} Y^{\intercal}\widehat{K}\left(X_{\overline{\Omega}}, X_{\overline{\Omega}}\right)^{-1} Y \quad \text{sujeto a} \quad \begin{cases} y_i = g(\widetilde{x}_i) & \text{para } \; k = 1, \ldots, N_{\partial\Omega}, \\ y_{i+N_{\partial\Omega}+N_{\Omega}} + h\left(y_{i+N_{\partial\Omega}}\right) = f(x_i), & \text{para } i = 1, \ldots, N_{\Omega}. \end{cases}$$

el cual es un problema de optimización en un espacio euclidiano $\mathbb{R}^M$ con restricciones, donde h es una función continua. Con esta formulación, se puede aún más reducir la dimensión del problema de manera que

$$\min_{\widetilde{Y} \in \mathbb{R}^{N_{\Omega}}} \left(g\left(\widetilde{X}_{\partial\Omega}\right),\; \widetilde{Y},\; f\left(X_{\Omega}\right) - h(\widetilde{Y}) \right) \widehat{K}\left(X_{\overline{\Omega}}, X_{\overline{\Omega}}\right)^{-1} \begin{pmatrix} g\left(\widetilde{X}_{\partial\Omega}\right) \\ \widetilde{Y} \\ f\left(X_{\Omega}\right) - h(\widetilde{Y}) \end{pmatrix}, \tag{6}$$

donde usamos la notación

$$g\left(\widetilde{X}_{\partial\Omega}\right) = \begin{pmatrix} g(\widetilde{x}_1) \\ \vdots \\ g\left(\widetilde{x}_{N_{\partial\Omega}}\right) \end{pmatrix}, \quad f\left(X_{\Omega}\right) - h(\widetilde{Y}) = \begin{pmatrix} f(x_1) - h(\widetilde{y}_1) \\ \vdots \\ f(x_{N_{\Omega}}) - h(\widetilde{y}_{N_{\Omega}}) \end{pmatrix}.$$

A diferencia de los sistemas estudiados en la Lección 3, este sistema depende significativamente de la no–linealidad de h. Por lo cual, para poder encontrar una solución numérica, es necesario plantear una estrategia de optimización adecuada, como variantes del algoritmo de Gauss-Newton (véase [2]).

En la práctica, se puede usar la fórmula de reconstrucción (5) de la solución de la EDP para realizar un estudio numérico según la forma e interés de aplicación. De igual manera, este enfoque basado en GPR permite acceder a valores de incertiud de esta estimación dada la información en los puntos de colocación, lo cual permite a su vez desarrollar análisis sobre la cuantificación de incertitudes (*uncertainty quantification* o UQ, en inglés) de la EDP. No obstante, nótese que esta estrategia de modelamiento requiere una alta regularidad de la función K puesto que se accede a valores puntuales de la misma.

Observación 2.1 *La extensión de esta metodología a casos generales de EDP no–lineales sigue el planteamiento de sistemas de optimización con restricciones en puntos de colocación (véase la sección 3 de [2]). Este enfoque además es similar al usado para aproximar soluciones de EDPs en un contexto de aprendizaje automático por redes neuronales (véase [3] sobre* Physics-Informed Neural Networks *o PINNs, en inglés). Sin embargo, en ese caso a pesar de ser técnicas estadísticas que otorgan resultados satisfactorios, no se tiene un marco matemático completo con respecto a su estudio, lo cual es una ventaja de los enfoques basados en GPR a través de la teoría de RKHS. Las líneas de investigación actuales se enfocan en como diseñar las funciones núcleo K de manera que el RKHS asociado integre la mayor cantidad de información física de la solución de la EDP (por ejemplo, condiciones de divergencia nula, periodicidad, escalamiento o incluso turbulencia, véase [4]).*

Referencias

[1] Iain Henderson, Pascal Noble, and Olivier Roustant. Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 29(4):3396–3422, 2023.

[2] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447:110668, 2021.

[3] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[4] Houman Owhadi. Gaussian process hydrodynamics. *Applied Mathematics and Mechanics*, 44(7):1175–1198, 2023.