Resolviendo ecuaciones estocásticas en dimension infinita mediante redes neuronales

David Felipe, Alonso Urbina y Sebastian Pieringer

Fecha de la última edición: 06-06-2025

Departamento de Ingeniería Matemática, Universidad de Chile

Índice de contenidos

1 Introducción y algunas ecuaciones importantes

1.1 Introducción

El tipo de ecuaciones que se estudiaran en este trabajo son una generalización de las ecuaciones diferencias estocásticas (EDEs), aveces llamadas ecuaciones de Itô. La generalización a espacios de dimensión infinita, es útil y necesaria para, por ejemplo, resolver ecuaciones estocásticas de tipo parabolico, es decir, ecuaciones que involuacran derivadas parciales en tiempo y en espacio, además de una fuente de aleatoriedad. El estudio análitico de este tipo de ecuaciones es complejo y en la gran mayoría de los casos no se pueden encontrar soluciones explícitas, lo que hace necesario recurrir a métodos numéricos para simular sus soluciones.

En este trabajo simularemos la solución de una ecuación estocástica en dimensión infinita, utilizando redes neuronales. Para ello introduciremos formalmente el concepto de ecuación estocástica en dimensión infita y haremos uso de resultados importantes de la teoría de procesos estocásticos y ecuaciones diferenciales estocásticas. Además, presentaremos resultados teóricos de aproximación mediante redes neuronales y explicitaremos un algoritmo que permita construir una red neuronal para simular la solución de una ecuación dada. Para finalizar, mostraremos ejemplos de implementaciones en *Python* y discutiremos los resultados obtenidos.

1.2 Algunas ecuaciones importantes

Estas ecuaciones pueden describir fenómenos aleatorios en multiples áreas del conocimiento. Algunas aplicaciones incluyen modelos en física estadística, finanzas, ingeniería, química y biología.

1.2.1 Ecuación del calor estocástica con ruido aditivo

La ecuación del calor usual en dimensión 1 es una EDP de la siguiente forma:

$$\partial_t u(t,x) = \sigma \partial_{xx} u(t,x) + f(t,x), \quad t \in [0,T], x \in \mathcal{D}$$

donde f es un término fuente que puede depender del tiempo y del espacio, $\mathcal{D} \subseteq \mathbb{R}$ y $\sigma > 0$ es un coeficiente de difusividad térmica. Esta ecuación describe, por ejemplo, la difusión de calor en un medio, también es ampliamente utilizada en finanzas para modelar la evolución de precios de activos.

Su versión estocástica con ruido aditivo es una ecuación de la siguiente forma:

$$\partial_t u(t,x) = \sigma \partial_{xx} u(t,x) + f(t,x) + \dot{W}(t,x), \quad t \in [0,T], x \in \mathcal{D}$$

donde $\dot{W}(t,x)$ es un ruido blanco tiempo espacial, que representa una perturbación aleatoria en el sistema, que tiene media nula en cada punto espacio temporal y varianza unitaria, además, es independiente en el tiempo y en el espacio, esto puede representar, por ejeplo, una fuente de calor aleatoria que esta afectando al sistema.

Es importante notar que el término $\dot{W}(t,x)$ es un proceso estocástico y por lo tanto, esta asociado a un espacio de probabilidad subyacente $(\Omega, \mathcal{F}, \mathbb{P})$. Luego la ecuación anterior también se puede escribir como:

$$\partial_t u(t,x;\omega) = \sigma \partial_{xx} u(t,x;\omega) + f(t,x;\omega) + \dot{W}(t,x;\omega), \quad t \in [0,T], x \in \mathcal{D}, \omega \in \Omega$$

Entonces la solución de esta ecuación es un proceso estocástico $u(t, x; \omega)$ que depende del tiempo, del espacio y de la realización ω.

2 Ecuaciones estocásticas en dimensión infinita

2.1 Preliminares

Durante esta sección consideramos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

Definición 2.1 (Movimiento Browniano)**.** Un proceso $W := (W_t)_{t\in[0,\infty)} : [0,\infty) \times \Omega \to \mathbb{R}$ se denomina Movimiento Browniano si:

1. $W_0 = 0$,
2. W_t tiene trayectorias continuas,
3. **(Incrementos estacionarios)** Para $0 \leq s < t$, $W_t - W_s \sim \mathcal{N}(0, t-s)$.
4. **(Incrementos independientes)** Para $0 = t_0 < t_1 < \cdots < t_n$, $\left\{W_{t_{i+1}} - W_{t_i}\right\}_{i=0}^{n-1}$ son variables aleatorias independientes.

Observación. Un Movimiento Browniano W se puede interpretar como una variable aleatoria:

$$X : \Omega \to \mathcal{C}([0,\infty))$$

con $X : \mathcal{F} - \mathcal{B}(\mathcal{C}([0,\infty)))$ medible donde $\mathcal{C}([0,\infty))$ está dotado de la topología de la convergencia uniforme en compactos.

Teorema 1 (Existencia y unicidad del Movimiento Browniano)**.** Existe una única medida de probabilidad $\mathbb{W}$ en $(\mathcal{C}([0,T]), \mathcal{B}(\mathcal{C}([0,T])))$ que genera un moviento Browniano. Es decir:

$$\mathbb{W}(A) = \mathbb{P}(W \in A)$$

con W un Movimiento Browniano.

Definición 2.2 (Movimiento Browniano en $\mathbb{R}^d$)**.** Si $W^1, W^2, \ldots, W^d$ son d movimientos Browniano independientes a $W = (W^1, W^2, \ldots, W^d)$ lo llamamos movimiento Browniano d-dimensional.

Definición 2.3 (Filtración estándar del Movimiento Browniano)**.** Sea W un ,ovimiento Browniano d-dimensional. Definamos para $t > 0$:

$$\mathcal{F}_t^+ = \bigcap_{\epsilon>0} \sigma\left(W_s : 0 \leq s \leq t+\epsilon\right).$$

Además, definimos la filtración estándar del moviemiento Browniano como la familia $\{\mathcal{F}_t\}_{t\geq 0}$ dada por $\mathcal{F}_t = \overline{\mathcal{F}_t^+}$. Donde la barra denota la completación de la σ-álgebra.

Desde este punto en adelante, consideraremos un movimiento Browniano W d-dimensional y la filtración estándar $\{\mathcal{F}_t\}_{t\geq 0}$ generada por W.

Definición 2.4 (Proceso simple)**.** Un proceso simple $\phi(t)$, $t \in [0,T]$, con valores en un espacio de Hilbert H es de la forma

$$\phi(t) = \sum_{m=0}^{k-1} \phi_m \mathbf{1}_{(t_m,t_{m+1}]}(t), \quad 0 = t_0 < t_1 < \ldots < t_k = T,$$

donde cada ϕ_m es una variable aleatoria a valores en H, que es $\mathcal{F}_{t_m}$-medible.

Definición 2.5 (Integral estocástica para procesos simples)**.** Sea un proceso simple $\phi(t)$, $t \in [0,T]$, con valores en $L^1(\mathbb{R}^d; H)$. Para dicho proceso simple ϕ, definimos la integral estocástica con respecto al movimiento Browniano W como:

$$\int_0^t \phi(s)\, dW(s) := \sum_{m=0}^{k-1} \phi_m \left(W_{t_{m+1}\wedge t} - W_{t_m \wedge t}\right), \quad t \in [0,T],$$

que comúnmente denotamos por $\phi \cdot W(t)$. Además, definimos una norma asociada al proceso ϕ por:

$$\|\phi\|_T := \left(\mathbb{E}\left[\int_0^T \|\phi(s)\|^2_{L^2(\mathbb{R}^d;H)}\, ds\right]\right)^{1/2}.$$

Definición 2.6 (Martingalas cuadrádo integrables)**.** Sea $\{\mathcal{G}_t\}_{t\geq 0}$ una filtración de $\mathcal{F}$. Un proceso estocástico $X = (X_t)_{t\in[0,T]}$ a valores en H es una martingala continua cuadrado integrable si:

1. X es $\mathcal{F}_t$-medible para cada $t \geq 0$.
2. $\sup_{t\in[0,T]} \mathbb{E}[\|X_t\|_H^2] < \infty$.
3. $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ para todo $0 \leq s < t \leq T$.
4. X tiene trayectorias continuas casi seguramente.

El conjunto de todas las martingalas continuas cuadrado integrables, en $[0,T]$ se denota por $\mathcal{M}_T^2$. Este espacio es un espacio de Banach con la norma:

$$\|X\|_{\mathcal{M}_T^2} = \left(\sup_{t\in[0,T]} \mathbb{E}\left[\|X_t\|_H^2\right]\right)^{1/2}.$$

Proposición 2.1. *Sea ϕ un proceso simple que satisface $\|\phi\|_T < \infty$. Entonces, el proceso integral estocástico $\phi \cdot W$ es una martingala continua en $[0,T]$, que verifica la siguiente identidad:*

$$\mathbb{E}\left[\|\phi \cdot W(t)\|_H^2\right] = \|\phi\|_t^2, \quad \text{para cada } t \in [0,T].$$

en particular, como $\|\phi\|_t^2$ es creciente en t, se tiene que:

$$\|\phi \cdot W\|_{\mathcal{M}_T^2} = \left(\sup_{t\in[0,T]} \mathbb{E}\left[\|\phi \cdot W(t)\|_H^2\right]\right)^{1/2} = \|\phi\|_T.$$

Observación. La integral estocástica $\phi \cdot W$ es un operador lineal desde el espacio de procesos simples con norma $\|\cdot\|_T$ al espacio de martingalas cuadrado integrables $\mathcal{M}_T^2$. Además, este operador es una isometría.

Definición 2.7 (Proceso predecible)**.** Un proceso ϕ es predecible si es medible respecto a la σ-álgebra generada por los conjuntos del tipo $A \times (s,t]$ con $0 \leq s \leq t$ y $A \in \mathcal{F}_s$.

Proposición 2.2 (Extensión por aproximación)**.** *Sea ϕ un proceso predecible con valores en $L^2(\mathbb{R}^d; H)$ que satisface:*

$$\|\phi\|_T = \left(\mathbb{E}\left[\int_0^T \|\phi(s)\|_{L^2}^2\, ds\right]\right)^{1/2} < \infty.$$

Entonces, existe una sucesión $\{\phi_n\}_n$ de procesos simples que converge a ϕ en la norma $\|\cdot\|_T$, es decir,

$$\|\phi - \phi_n\|_T \to 0, \quad \textit{cuando } n \to \infty.$$

Gracias a lo anterior, el siguiente límite está bien definido y se utiliza como definición de la integral estocástica del proceso ϕ respecto al movimiento Browniano W:

$$\phi \cdot W(t) := \lim_{n\to\infty} \phi_n \cdot W(t), \quad t \in [0,T].$$

Observación. Lo anterior, además nos permite estimar númericamente la integral estocástica de un proceso predecible ϕ mediante la aproximación por procesos simples. En particular, si ϕ es un proceso predecible con valores en $L^2(\mathbb{R}^d; H)$, podemos aproximar:

$$\phi \cdot W(t) \approx \sum_{m=0}^{k-1} \phi(t_m)\big(W_{t_{m+1}\wedge t} - W_{t_m \wedge t}\big),$$

Donde $\phi_m =$ para todo $m \in \{0,\ldots,k-1\}$.

2.2 Ecuaciones estocásticas en dimensión infinita y sus soluciones

Consideremos:

- $T > 0$.
- $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad.
- $(H, \langle\cdot,\cdot\rangle_H)$ un espacio de Hilbert separable.
- $X_t : [0,T] \times \Omega \to H$ es el proceso estocástico solución con trayectorias en H.
- $A : \mathrm{dom}(A) \subset H \to H$ una aplicación lineal continua.
- $F : [0,T] \to H$.
- $B : H \to L_2(\mathbb{R}^d; H)$.
- W_t es un Movimiento Browniano d-dimensional.
- $\chi_0 \in H$.

Definición 2.8 (Ecuación estocástica en dimensión infinita)**.** Una ecuación estocástica en dimensión infinita o ecuación estocástica con derivadas parciales (SPDE) es una ecuación de la forma:

$$dX_t = (AX_t + F(t))dt + B(X_t)dW_t, \quad t \in [0,T], \tag{1}$$

con condición inicial $X_0 = \chi_0 \in H$.
Decimos que X es una solución fuerte de la ecuación (1) si es un proceso estocástico predecible a valores en H, tal que para todo $t \in [0,T]$ se cumple que $X(t) \in \mathrm{dom}(A)$ y:

$$X_t - \chi_0 = \underbrace{\int_0^t AX_s + F(s)ds}_{\text{Integral de Bochner}} + \underbrace{\int_0^t B(X_s)dW_s}_{\text{Integral estocástica}}, \quad \mathbb{P} - c.s. \tag{2}$$

Por otro lado, decimos que X es una solución débil de la ecuación (1) si es un proceso estocástico predecible a valores en H, tal que las trayectorias de X son $\mathbb{P} - c.s.$ Bochner integrable y se cumple que para todo $z \in \mathrm{dom}(A^*)$ y para todo $t \in [0,T]$ se tiene que:

$$\langle X_t, z\rangle_H - \langle \chi_0, z\rangle_H = \int_0^t \langle X_s, A^*z\rangle + \langle F(s), z\rangle\, ds + \left\langle z, \int_0^t B(X(s))dW_s \right\rangle_H, \quad \mathbb{P} - c.s. \tag{3}$$

Proposición 2.3. *Toda solución fuerte de la ecuación estocástica en dimensión infinita* (1) *es una solución débil de la misma ecuación.*

Observación. Esta no es la forma más general de una SPDE, pero es lo suficientemente general para el propósito de este trabajo.

2.3 Soluciones Mild de las SPDEs

En esta sección, introduciremos el concepto de solución más débil de una SPDE, conocido como solución mild.

Definición 2.9 (C_0-Semigrupo)**.** Una familia de operadores $(S_t)_{t\in[0,T]} \subseteq \mathcal{L}(H;H)$ es llamada un C_0-semigrupo si:

1. $S_0 = Id_H$, donde Id_H es el operador identidad en H.
2. $S_s \circ S_t = S_{s+t}$, $\forall s,t \in T$ tales que $s + t \leq t$.
3. $\lim_{t\to 0} S_t x = x$ para todo $x \in H$.

Ejemplo (Semigrupo del Laplaciano). Sea $A = \Delta : H^2(\mathcal{D}) \subseteq L^2(\mathcal{D}) \to L^2(\mathcal{D})$, con $\mathcal{D} \subseteq \mathbb{R}^n$ el operador Laplaciano. Entonces, el semigrupo asociado al operador Laplaciano esta dado por:

$$S_t x = \begin{cases} Id, & \text{si } t = 0, \\ (u \mapsto \int_{\mathbb{R}^n} \phi_t(u-v)x(v)dv), & \text{si } t > 0, \end{cases}$$

para $t \in [0,T]$, con $\phi_t(y) = \frac{1}{(4\pi t)^{n/2}} e^{-\frac{\|y\|^2}{4t}}$ el núcleo del calor.

Desde este punto en adelante, suponemos que:

1. $A : \mathrm{dom}(A) \subseteq H \to H$ es el generador de un semigrupo C_0-semigroup $(S_t)_{t\in[0,T]}$.
2. El mapeo $B : H \to L_2\left(\mathbb{R}^d; H\right)$ es $\mathcal{B}(H)/\mathcal{B}\left(L_2\left(\mathbb{R}^d; H\right)\right)$-medible.
3. Existe $C_B > 0$ tal que $\forall x, y \in H$:

$$\|B(x) - B(y)\|_{L_2(\mathbb{R}^d;H)} \leq C_B\|x - y\|_H,$$

   y:

$$\|B(x)\|^2_{L_2(\mathbb{R}^d;H)} \leq C_B^2(1 + \|x\|_H^2).$$

4. La condición inicial $\chi_0 \in H$ es determinista.

Definición 2.10 (Solución mild)**.** Una función $X : [0, T] \times \Omega \to H$ se llama mild solution de la ecuación (1) si es $\mathcal{F}$-predecible y cumple:

$$\mathbb{P}\left(\int_0^T \|X_t\|_H^2 dt < \infty\right) = 1,$$

y para todo $t \in [0, T]$,

$$X_t = S_t\chi_0 + \int_0^t S_{t-s}B(X_s)\, dW_s, \quad \mathbb{P}\text{-c.s.} \tag{4}$$

Teorema 2 (Teorema de Existencia y Unicidad)**.** Bajo las suposiciones anteriores, existe una única solución mild $X : [0, T] \times \Omega \to H$ de la ecuación (1) que tiene trayectorias continuas en H. Además, existe una constante $C > 0$ tal que para todo $p \in [1, \infty)$:

$$\mathbb{E}\left[\sup_{t\in[0,T]} \|X_t\|_H^p\right] \leq C(1 + \|\chi_0\|_H^p).$$

Proposición 2.4. *Una solución débil de la ecuación* (1) *es una solución mild de la misma ecuación.*

2.4 Expansión en caos de Wiener y Teorema de Cameron Martin

Esta sección veremos el resultado clave que nos permite aproximar la solución de una SPDE mediante redes neuronales. Este resultado es conocido como el teorema de Cameron Martin.

Definición 2.11 (Polinomios de Hermite)**.** La familia de polinomios de Hermite $(h_n)_{n\in\mathbb{N}_0}$ esta definida por:

$$h_n : s \in \mathbb{R} \mapsto h_n(s) = (-1)^n \cdot e^{\frac{s^2}{2}} \frac{d^n}{ds^n}(e^{-\frac{s^2}{2}}) \in \mathbb{R}, \quad n \in \mathbb{N}_0$$

Para representar la solución en función del ruido, usamos la expansión en caos de Wiener. Sea $\left(g_j\right)_{j\in\mathbb{N}}$ una base ortonormal de $L^2([0, T])$.

Ejemplo (Ejemplo de base ortonormal). Una base ortonormal de $L^2([0,T])$ frecuentemente utilizada es la base $\mathcal{B} := (g_j)_{j\in\mathbb{N}^*}$ dada por:

$$
\begin{aligned}
g_1(t) : [0,T] \to \mathbb{R}, \quad g_1(t) &= \frac{1}{\sqrt{T}}, \\
g_j(t) : [0,T] \to \mathbb{R}, \quad g_j(t) &= \sqrt{\frac{2}{T}} \cos\left(\frac{(j-1)\pi t}{T}\right), \quad j \geq 2.
\end{aligned}
$$

Esta base es la que frencuentemente se utiliza en esta aplicación, más adelante veremos para que es útil.

Notación. Denotemos

$$
\xi_{i,j} := \int_0^T g_j(s)\, dW_s^i, \quad i \in [d], j \in \mathbb{N} \tag{5}
$$

Proposición 2.5. *Se tiene que* $(\xi_{i,j})_{i\in[d],j\in\mathbb{N}} \sim \mathcal{N}(0,1)$ *i.i.d.*

Lema 2.1. *Sea* $t \in [0,T], i \in \mathbb{N}^*$. *Entonces, se tiene que:*

$$
W_t = \sum_{i=1}^{d} \sum_{j=1}^{\infty} \xi_{i,j} \left(\int_0^t g_j(s)\, ds\right) e_i
$$

donde la suma converge con respecto a la norma $\|\cdot\|_{L^2(\Omega,\mathcal{F},\mathbb{P})}$. *Aquí* e_i *es el i-ésimo vector canónico de* $\mathbb{R}^d$.

Ejemplo. Para las funciones de la base ortonormal $\mathcal{B}$ del ejemplo anterior, tenemos que, para $0 \leq t' < t \leq T$:

$$
\begin{aligned}
\int_{t'}^{t} g_1(s)\, ds &= \frac{t-t'}{\sqrt{T}}, \\
\int_{t'}^{t} g_j(s)\, ds &= \frac{\sqrt{2T}}{(j-1)\pi}\left(\sin\left(\frac{\pi(j-1)t}{T}\right) - \sin\left(\frac{\pi(j-1)t'}{T}\right)\right), \quad j \geq 2.
\end{aligned}
$$

Observación. El lema anterior nos permite expresar el movimiento Browniano W en termino de las variables aleatorias $\xi_{i,j}$, las cuales son independientes y tienen distribución normal estándar. Es decir, conciendo el valor de $\xi_{i,j}$, podemos reconstruir el movimiento Browniano W.

Ahora vamos a definir algunas notaciones que nos serán utiles para definir los polinomios de Wick y posteriormente expresar una aproximación de la solución de la SPDE.

Notación.

$$
\mathcal{J} \ := \ \left\{ \alpha = (\alpha_{i,j})_{i,j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}\times\mathbb{N}} \ \middle| \ |\alpha| \ := \ \sum_{i,j=1}^{\infty} \alpha_{i,j} < \infty \right\}
$$

y definamos, para cada $\alpha \in \mathcal{J}$,

$$
\alpha! \ := \ \prod_{i,j=1}^{\infty} \alpha_{i,j}!
$$

También consideremos, el conjunto de índices

$$
\mathcal{J}_{I,J,K} := \left\{ \alpha = (\alpha_{i,j})_{i,j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}\times\mathbb{N}} \ \middle| \ \alpha_{i,j} = 0 \text{ si } i > I \text{ o } j > J,\ |\alpha| \leqslant K \right\} \subseteq \mathcal{J},
$$

Observación. Se satisface que:

$$|\mathcal{J}_{I,J,K}| = \sum_{k=0}^{K} \binom{IJ+k-1}{k} = \frac{(IJ+K)!}{(IJ)!K!}$$

Definición 2.12 (Polinomios de Wick)**.** La familia de polinomios de Wick $(\xi_\alpha)_{\alpha\in\mathcal{J}}$, esta definida por

$$\xi_\alpha : \omega \in \Omega \mapsto \xi_\alpha(\omega) = \frac{1}{\sqrt{\alpha!}} \prod_{i,j} h_{\alpha_{i,j}}(\xi_{i,j}(\omega)), \quad \alpha \in \mathcal{J} \tag{6}$$

donde $h_n(\cdot)$ es el n-ésimo polinomio de Hermite, y α es un multiíndice con soporte finito.

Proposición 2.6. *Los polinomios de Wick forman una base ortonormal en $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$.*

Teorema 3 (Teorema de Cameron Martin)**.** Sea X la solución mild de la SPDE. Entonces, para todo $\varepsilon > 0$, existen $I, J, K \in \mathbb{N}$ y funciones $(x_\alpha)_{\alpha\in\mathcal{J}_{I,J,K}} \subset C([0,T]; H)$ tales que:

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left\| X_t - \sum_{\alpha\in\mathcal{J}_{I,J,K}} x_\alpha(t)\xi_\alpha \right\|_H^p\right]^{1/p} < \varepsilon.$$

En el caso $p = 2$, la expansión converge en $L^2(\Omega; H)$ y los coeficientes se obtienen como:

$$x_\alpha(t) = \mathbb{E}[X_t\xi_\alpha].$$

La expansión se puede dividir en:

$$X_t \approx x_0(t) + \sum_{|\alpha|=1} x_\alpha(t)\xi_\alpha + \sum_{|\alpha|\geq 2} x_\alpha(t)\xi_\alpha,$$

correspondiendo a una parte determinista, una parte gaussiana y una parte no-gaussiana, respectivamente.

De esta forma, encontramos una manera de aproximar la solución utilizando un truncamiento sobre las sumas anteriores.

3 Aproximación por redes neuronales

3.1 Redes neuronales feedforward

Definición 3.1 (Red neuronal)**.** Para $T > 0$ y $U \subseteq \mathbb{R}^m$, una red neuronal es de la forma

$$[0,T] \times U \ni (t,u) \quad \mapsto \quad \varphi(t,u) := \sum_{n=1}^{N} y_n \rho \left(a_{0,n} t + a_{1,n}^\top u - b_n \right) \in \mathbb{R}^d$$

para algún $N \in \mathbb{N}$ el número de neuronas y algún $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ La función de activación. Los parámetros consisten en los pesos $a_{0,1}, \ldots, a_{0,N} \in \mathbb{R}$ y $a_{1,1}, \ldots, a_{1,N} \in \mathbb{R}^m$, los sesgos $b_1, \ldots, b_N \in \mathbb{R}$, y las funciones lineales de salida $y_1, \ldots, y_N \in \mathbb{R}^d$. Consideraremos esta estructura de aquí en adelante, es decir, redes neuronales con solo una capa oculta.
En el caso de que no exista una dependencia temporal, podemos definir la red neuronal de la forma

$$U \ni u \mapsto \varphi(u) := \sum_{n=1}^{N} y_n \rho \left(a_{1,n}^\top u_n - b_n \right) \in \mathbb{R}^d$$

.

3.2 Teorema de aproximación universal

El teorema de aproximación universal de redes neuronales establece que toda función continua puede aproximarse a partir de redes neuronales como las definidas anteriormente. En particular, podemos aproximar los propagadores $(x_\alpha)_{\alpha \in \mathcal{J}_{I,J,K}} \subseteq C^0([0,T]; H)$. En base a esto y al teorema de Cameron-Martin tenemos el siguiente resultado.

Teorema 4. Supongamos que para $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ (abierto, si $k \geqslant 1$), y $\gamma \in (0, \infty)$, sea $(H, \langle \cdot, \cdot \rangle_H)$ un espacio de Hilbert de funciones $f : U \to \mathbb{R}^d$ tales que el mapeo de restricción

$$\left(C_b^k \left(\mathbb{R}^m; \mathbb{R}^d \right), \| \cdot \|_{C^k_{pol,\gamma}(\mathbb{R}^m; \mathbb{R}^d)} \right) \ni f \quad \mapsto \quad f \Big|_U \in (H, \| \cdot \|_H)$$

es un embedding continuo y denso, i.e. es continuo y su imagen es densa en $(H, \| \cdot \|_H)$.
Sea $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ no polinomial, y sea $p \in [1, \infty)$. Más aún, supongamos que las suposiciones del teorema de existencia y unicidad se tienen, y sea $X : [0,T] \times \Omega \to H$ la mild solution de la (SPDE). Entonces, para todo $\varepsilon > 0$ existen $I, J, K \in \mathbb{N}$ y $(\varphi_\alpha)_{\alpha \in \mathcal{J}_{I,J,K}} \subseteq \mathcal{NN}^\rho_{[0,T] \times U, d}$ tales que

$$\mathbb{E} \left[\sup_{t \in [0,T]} \left\| X_t - \sum_{\alpha \in \mathcal{J}_{I,J,K}} \varphi_\alpha(t, \cdot) \xi_\alpha \right\|_H^p \right]^{\frac{1}{p}} < \varepsilon$$

3.3 Algoritmo de aproximación de solución de SPDE.

Como se explico en la sección anterior , podemos aproximar la solución mild de la SPDE de la siguiente forma:

$$X_t \approx \sum_{\alpha \in \mathcal{J}_{I,J,K}} \varphi_\alpha(t, \cdot) \xi_\alpha$$

Con $I, J, K \in \mathbb{N}$ y $(\varphi_\alpha)_{\alpha \in \mathcal{J}_{I,J,K}} \subseteq \mathcal{NN}^{\rho}_{[0,T]\times U,d}$. Donde buscamos encontrar los parámetros adecuados para cada φ_α, las redes las entrenaremos desde un enfoque no supervisado (sin utilizar puntos conocidos del grafo de la solución, pero de quererse tambien es posible integrar esta informacion) definiendo una función de perdida que nos informe de que tan bien se ajustan los parámetros a la SPDE. Aproximaremos nuestra solución por el proceso $X^{(I,J,K)} : [0,T] \times \Omega \to H$, dado por:

$$X_t^{(I,J,K)} := \sum_{\alpha \in \mathcal{J}_{I,J,K}} \varphi_\alpha(t,\cdot)\xi_\alpha$$

Utilizamos como función de perdida el error empírico de Sobolev (con pesos) y con los hiperparámetros $M_1, M_2, M_3 \in \mathbb{N}$, $(\omega_{m_1})_{m_1=1}^{M_1} \subseteq \Omega$, $0 \leq t_1 < t_2 < \cdots < t_{M_2} \leq T$, $(u_{m_3})_{m_3=1}^{M_3} \subseteq U \subseteq \mathbb{R}^m$ y por ultimo los coeficientes que determinan la contribución de las derivadas, dependientes de $\beta \in \mathbb{N}^m_{0,k}, m_1 \in [M_1], m_2 \in [M_2]_0, m_3 \in [M_3]$, $\quad c_{\beta,m_1,m_2,m_3} \in [0,\infty)$ (donde k es el numero de veces que podemos diferenciar débilmente la solución)

$$L(X^{(I,J,K)};\Theta) := (\sum_{m_1=1}^{M_1}\sum_{m_2=0}^{M_2}\sum_{m_3=1}^{M_3}\sum_{\beta\in\mathbb{N}^m_{0,k}} c_{\beta,m_1,m_2,m_3}|l(X^{(I,J,K)},\omega_{m1},t_{m_2},u_{m_3},\beta)|^2)^{\frac{1}{2}}$$

Donde,

$$\begin{aligned} l(X^{(I,J,K)},\omega_{m1},t_{m_2},u_{m_3},\beta) := \partial_\beta X^{(I,J,K)}_{t_{m_2}}(\omega_{m1})(u_{m_3}) - \partial_\beta[&\chi_0 \\ &+ \sum_{l=1}^{m_2} AX^{(I,J,K)}_{t_{l-1}}(\omega_{m1})(t_l - t_{l-1}) \\ &+ \sum_{l=1}^{m_2}\langle B(X^{(I,J,K)}_{t_{l-1}}(\omega_{m1})); W^{(I,J)}_{t_l}(\omega_{m1}) - W^{(I,J)}_{t_{l-1}}(\omega_{m1})\rangle](u_{m_3}) \end{aligned}$$

Algorithm 1 Algoritmo de aproximacion de solucion de la SPDE

Require: $I, J, K \in \mathbb{N}$, la C.I. $\chi_0 \in H$ y los coeficientes A, B de la SPDE.
Ensure: Aproximación de la solucion $X^{(I,J,K)} : [0,T] \times \Omega \to H$

Simular $I \cdot J \cdot M_1$ realizaciones de $\xi_{i,j}(w_{m_1})$ de v.a. i.i.d. $\mathcal{N}(0,1)$
for $m_1 = 1,\ldots,M_1$ y $m_2 = 1,\ldots,M_2$ calcular el MB aproximado **do** ▷ Calcular el MB aproximado.
  $W^{(I,J)}_{t_{m_2}}(\omega_{m1}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\xi_{i,j}(w_{m_1})(\int_0^{t_{m_2}} g_j(s)ds)e_i$
end for
for $m_1 = 1,\ldots,M_1$ y $\alpha \in \mathcal{J}_{(I,J,K)}$ **do** ▷ Calcular los polinomios de Wick.
  $\xi_\alpha(\omega_{m1}) = \frac{1}{\sqrt{\alpha!}}\prod_{i=1,j=1}^{\infty} h_{\alpha_{i,j}}(\xi_{i,j}(w_{m_1}))$
end for
escoger ρ función de activación.
for $\alpha \in \mathcal{J}_{(I,J,K)}$ **do**
  Inicializar las redes descritas anteriormente $\varphi_\alpha(t,\cdot)$
end for
Inicializar el proceso $X_t^{(I,J,K)}(\omega) := \sum_{\alpha\in\mathcal{J}_{I,J,K}} \varphi_\alpha(t,\cdot)\xi_\alpha(\omega)$
Minimizar $L(X^{(I,J,K)};\Theta)$ con Θ los parametros de las redes usando descenso de gradiente estocástico.
return $X^{(I,J,K)} : [0,T] \times \Omega \to H$

---

4 Experimentos numéricos y resultados para ecuaciones particulares

(...)

A Integral de Bochner

Work in progress.

B Variables aletorias en espacios de Hilbert

Consideremos $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $(E, \|\cdot\|)$ un espacio de Banach separable, con dual E^* y $\mathcal{B}(E)$ su σ-álgebra boreliana.

Una función $X : \Omega \to E$, $\mathcal{F} - \mathcal{B}(E)$ medible, es una variable aleatoria a valores en E. Además, definimos su ley como la medida de probabilidad $\mathcal{L}(X)$ sobre $\mathcal{B}(E)$ dada por,

$$\mathcal{L}(X)(A) = \mathbb{P}(\omega \in \Omega \mid X(\omega) \in A), \forall A \in \mathcal{B}(E)$$

Además, sea G un espacio de Banach y $\Psi : E \to G$ una función $\mathcal{B}(E) - \mathcal{B}(G)$ medible, se define

$$\mathbb{E}[\Psi(X)] := \int_E \Psi(x) d\mathbb{P}(x)$$

Donde la integral corresponde a una integral de Bochner.

Ahora sea $p > 1$, se define

$$\|X\|_p := (\mathbb{E}[\|X\|^p])^{1/p}$$

Si $\|X\|_p < \infty$, decimos que $X \in L^p(\Omega, \mathcal{F}, \mathbb{P}; E)$.

Sea ahora, $X \in L^2(\Omega, \mathcal{F}, \mathbb{P}; H)$ con H un espacio de Hilbert con producto interno $\langle \cdot, \cdot \rangle$. Se define su operador de covarianza por,

$$\mathrm{Cov}(X) : h \in H \to \mathbb{E}[X - \mathbb{E}[X]] \langle X - \mathbb{E}[X], h \rangle \in H$$

C Algunas demostraciones

Work in progress.

D Códigos en Python

Work in progress.

Referencias