CEA Comité Español de Automática
XLVII JORNADAS DE AUTOMÁTICA CÓRDOBA 2026
UNIVERSIDADE DA CORUÑA

Jornadas de Automática

Imitación de políticas NMPC mediante *universal kriging* regularizado

Carlos Moreno-Blazquez[a,*], Filiberto Fele[a], Teodoro Alamo[a]

[a] *Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla, Camino de los Descubrimientos s/n, 41092 Sevilla, España.*

To cite this article: Moreno-Blazquez, C., Fele, F., Alamo, T. 2026. NMPC policy imitation via regularized universal kriging. Jornadas de Automática, 47. https://doi.org/10.17979/ja-cea.2026.47.13770

Resumen

Este trabajo aproxima fuera de línea la ley de control de un controlador predictivo no lineal (NMPC, por sus siglas en inglés) mediante *universal kriging* (UK) para evitar la carga computacional en línea asociada a la optimización de horizonte deslizante del NMPC. A partir de pares estado-acción se construyen una política UK estándar y una versión regularizada con penalización ℓ_1 sobre los pesos, orientada a favorecer interpolaciones más dispersas e interpretables. Demostramos la efectividad de la metodología mediante un modelo simulado de swing-up de un péndulo invertido con restricción de par.

Palabras clave: Control predictivo, Control basado en datos, Aprendizaje para control, Control predictivo no lineal, Control con restricciones

NMPC policy imitation via regularized universal kriging

Abstract

This work approximates offline the feedback law of a nonlinear model predictive controller (NMPC) using universal kriging (UK) to avoid the online computational load associated with NMPC receding-horizon optimization. From state-action pairs, a standard UK policy and an ℓ_1-regularized variant on the kriging weights are built to promote sparser and more interpretable interpolations. We demonstrate the effectiveness of the methodology using a simulated swing-up model of a torque-constrained inverted pendulum.

Keywords: Predictive control, Data-based control, Learning for control, Nonlinear predictive control, Constrained control

1. Introducción

El control predictivo basado en modelo (*model predictive control*, MPC) es una metodología ampliamente utilizada para sistemas con restricciones, al permitir incorporar límites de operación y optimizar una función de coste sobre un horizonte móvil (Rawlings et al., 2017). Su versión no lineal (*nonlinear model predictive control*, NMPC) extiende esta idea a dinámicas no lineales, restricciones no convexas y maniobras alejadas del régimen lineal, siendo adecuada en energía, robótica, procesos y sistemas electromecánicos. Su principal limitación es computacional: en cada muestreo debe resolverse un problema de optimización no lineal cuya dimensión crece con los horizontes de predicción y control. Aunque existen algoritmos rápidos (Richter et al., 2012), el tiempo real sigue siendo crítico con periodos de muestreo reducidos, hardware limitado o muchas evaluaciones del controlador, lo que ha motivado métodos explícitos, aproximaciones locales y estrategias basadas en datos que sustituyen total o parcialmente la optimización en línea (Karg et al., 2021; Nadales et al., 2023).

Una vía habitual consiste en aproximar directamente la política generada por el NMPC. El controlador original proporciona acciones óptimas en distintos estados y, con ellas, se aprende una aproximación de la ley de control, cuya evaluación en línea es menos costosa que resolver el problema del NMPC. Esta idea aparece en aproximaciones neuronales de reguladores de horizonte deslizante no lineales (Parisini and Zoppoli, 1995), en el aprendizaje de leyes de MPC explícito con arquitecturas neuronales (Maddalena et al., 2020) y en el

*Autor para correspondencia: Carlos Moreno-Blazquez, cmb@us.es

uso de procesos gaussianos e interpolación kernel para controladores predictivos (Binder et al., 2019; Tokmak et al., 2025). El objetivo común es reducir la carga de la optimización repetida mediante la evaluación rápida de una política aprendida.

Estas alternativas, sin embargo, presentan compromisos. Las formulaciones explícitas pueden verse limitadas por el número de regiones y la memoria; las redes neuronales son rápidas, pero altamente parametrizadas y poco interpretables, y las garantías de estabilidad, factibilidad o satisfacción de restricciones suelen exigir mecanismos adicionales (Fabiani and Goulart, 2023); los procesos gaussianos y métodos kernel ofrecen una descripción más estructurada de la incertidumbre, aunque su evaluación puede depender mucho del número de muestras (Rose et al., 2025). En estos casos, interesa usar bases de datos moderadas y conservar interpretabilidad sobre la acción aproximada.

Dentro de esta familia, una alternativa natural para aproximar mapas estado-acción es *kriging*. Originado en geoestadística (Krige, 1962; Cressie, 1993), construye predictores lineales insesgados de mínima varianza y asigna pesos según la dependencia estimada entre muestras. Aunque está estrechamente relacionado con la regresión por procesos gaussianos (Williams and Rasmussen, 2006), su formulación modela la dependencia residual bajo la hipótesis intrínseca (Matheron, 1963), menos restrictiva que la estacionariedad de segundo orden, y sus pesos indican directamente qué datos intervienen en cada predicción.

Este trabajo estudia *universal kriging* (UK) para imitar una política NMPC a partir de un número finito de acciones conocidas. Además del UK estándar, se considera una versión regularizada con penalización ℓ_1 sobre los pesos para favorecer predicciones más dispersas, interpretables y menos dependientes de extrapolaciones. Esta estrategia resulta especialmente útil cuando la formulación del NMPC debe actualizarse, por ejemplo ante nuevas referencias o puntos de operación (Limon et al., 2018), criterios económicos variables (Ferramosca et al., 2014), reajustes de pesos entre objetivos (Bemporad and Muñoz de la Peña, 2009) o cambios en su prioridad (He et al., 2021), ya que permite reconstruir la política aprendida a partir de un número moderado de nuevas resoluciones del controlador de referencia. La metodología se valida en el *swing-up* de un péndulo invertido, tomando como referencia un NMPC y analizando la fidelidad y coste computacional de la política aprendida.

2. Formulación NMPC

Considérese un sistema no lineal discreto descrito por $x_{k+1} = f(x_k, u_k)$ y $y_k = \psi(x_k, u_k)$, donde $x_k \in \mathbb{R}^{n_x}$ es el estado, $u_k \in \mathbb{R}^{n_u}$ la entrada manipulable, $y_k \in \mathbb{R}^{n_y}$ la salida controlada, $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ la dinámica discreta y $\psi : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$ el mapa de salida. Las restricciones, representadas mediante conjuntos admisibles $x_k \in \mathcal{X}$, $u_k \in \mathcal{U}$, $y_k \in \mathcal{Y}$, pueden incorporar límites físicos, restricciones de seguridad o saturaciones del actuador.

En un instante k, el NMPC calcula una secuencia de entradas futuras $\mathbf{u}_k = \{u_{0|k}, u_{1|k}, \ldots, u_{N_c-1|k}\}$, para un horizonte de control N_c, y predice la trayectoria $\mathbf{x}_k = \{x_{0|k}, x_{1|k}, \ldots, x_{N_p|k}\}$, para un horizonte de predicción $N_p \geq N_c$. La condición inicial del problema predictivo es $x_{0|k} = x_k$.

Para $i = 0, \ldots, N_p - 1$, las variables predichas satisfacen $x_{i+1|k} = f(x_{i|k}, u_{i|k})$ y $y_{i|k} = \psi(x_{i|k}, u_{i|k})$, donde, si $i \geq N_c$, se emplea habitualmente la última entrada libre, $u_{i|k} = u_{N_c-1|k}$.

La formulación cuadrática general del problema NMPC es

$$\begin{aligned} \min_{\mathbf{u}_k} \quad & \sum_{i=0}^{N_p-1} \ell(y_{i|k}, u_{i|k}, \Delta u_{i|k}) + V_f(x_{N_p|k}) \qquad (1) \\ \text{s.a.} \quad & x_{i+1|k} = f(x_{i|k}, u_{i|k}), \\ & y_{i|k} = \psi(x_{i|k}, u_{i|k}), \\ & x_{i|k} \in \mathcal{X},\ y_{i|k} \in \mathcal{Y},\ u_{i|k} \in \mathcal{U}, \\ & x_{0|k} = x_k, \end{aligned}$$

con V_f como coste terminal opcional, $\Delta u_{i|k} = u_{i|k} - u_{i-1|k}$ (tomando $u_{-1|k} = u_{k-1}$), y

$$\ell(y, u, \Delta u) = \|y - y^{\text{ref}}\|_{\text{Q}}^2 + \|u - u^{\text{ref}}\|_{\text{R}}^2 + \|\Delta u\|_{\text{S}}^2 \qquad (2)$$

donde $\|v\|_M^2 = v^\top M v$. La función de etapa penaliza el error de salida, el esfuerzo de control y el incremento de entrada. Las matrices $\text{Q} \succeq 0$, $\text{R} \succeq 0$ y $\text{S} \succeq 0$ fijan el compromiso entre seguimiento, energía y suavidad de actuación (Rawlings et al., 2017).

Tras resolver (1), solo se aplica la primera acción óptima $u_k = u_{0|k}^\star$. En el siguiente instante se mide de nuevo el estado y se repite el procedimiento. Por tanto, el NMPC define una política implícita

$$\pi_{\text{NMPC}}(x) = u_0^\star(x), \qquad (3)$$

que no está disponible en forma cerrada, sino como resultado de un problema de optimización. La idea de este trabajo es aproximar (3) con datos generados fuera de línea, aprendiendo directamente la acción que el NMPC completo aplicaría para un estado dado, haciendo para ello uso de *kriging*.

3. Interpolación mediante *universal kriging*

Sea la base de datos $\mathcal{D} = \{(z_i, \xi_i)\}_{i=1}^N$, donde $z_i \in \mathbb{R}^{n_z}$ representa el regresor e $\xi_i \in \mathbb{R}$ la magnitud observada. En el problema de imitación de política, z_i será el estado del sistema e ξ_i una componente de la acción óptima del NMPC. Para un nuevo punto z_0, *kriging* aproxima el valor desconocido ξ_0 mediante una combinación lineal de las muestras disponibles:

$$\hat{\xi}_0 = \sum_{i=1}^{N} \lambda_i \xi_i, \qquad (4)$$

donde $\lambda = (\lambda_i)_{i=1}^N \in \mathbb{R}^N$ son los pesos de interpolación. Estos pesos se eligen imponiendo que el predictor sea *insesgado* y que la varianza del error de predicción sea *mínima* (Cressie and Wikle, 2011, Sec. 4.1.2).

Para formular estas condiciones, *universal kriging* modela la variable de interés como

$$\xi(z) = \mu(z) + \delta(z), \qquad (5)$$

donde $\mu(z)$ recoge una tendencia determinista y $\delta(z)$ representa la componente residual, asumida de media nula. En este trabajo, la tendencia se aproxima mediante una base de funciones $\mu(z) = r(z)^\top \beta$, con $r(z) \in \mathbb{R}^{n_r}$ y coeficientes desconocidos $\beta \in \mathbb{R}^{n_r}$. Como dichos coeficientes no se conocen a priori, la

insesgadez $\mathbb{E}[\xi_0 - \hat{\xi}_0] = 0$ de (4) debe cumplirse para cualquier valor de β. Esto exige que los pesos reproduzcan en z_0 cualquier tendencia perteneciente al espacio generado por $r(z)$, lo que conduce a la restricción lineal

$$R\lambda = r(z_0), \tag{6}$$

donde $R = \begin{bmatrix} r(z_1) & r(z_2) & \cdots & r(z_N) \end{bmatrix} \in \mathbb{R}^{n_r \times N}$. En particular, si la base incluye un término constante, (6) contiene la condición $\sum_{i=1}^{N} \lambda_i = 1$.

Una vez impuesta la insesgadez, la contribución de la tendencia se cancela en el error de predicción, de modo que

$$e(z_0) = \xi(z_0) - \hat{\xi}(z_0) = \delta(z_0) - \sum_{i=1}^{N} \lambda_i \delta(z_i).$$

Dado que el predictor es insesgado, minimizar la varianza del error es equivalente a minimizar $\mathbb{E}[e^2(z_0)]$:

$$\mathbb{E}\left[e^2(z_0)\right] = \sum_{i=1}^{N} \lambda_i \mathbb{E}\left[(\delta_0 - \delta_i)^2\right] - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \mathbb{E}\left[(\delta_i - \delta_j)^2\right], \tag{7}$$

donde $\delta_0 = \delta(z_0)$ y $\delta_i = \delta(z_i)$. Por tanto, la minimización de la varianza del error queda determinada por términos de la forma $\mathbb{E}[(\delta_i - \delta_j)^2]$. Esta dependencia la describe el semivariograma

$$\gamma(h) := \frac{1}{2}\mathrm{Var}\{\delta(z+h) - \delta(z)\},$$

donde h representa la separación entre dos puntos del espacio de regresores. En el caso isotrópico considerado, h se identifica con la distancia euclídea entre regresores. A partir de los datos se estima primero un semivariograma empírico agrupando pares de muestras por distancia:

$$\hat{\gamma}(h) = \frac{1}{2|\mathcal{N}(h)|} \sum_{(i,j)\in\mathcal{N}(h)} (\delta_i - \delta_j)^2,$$

donde $\mathcal{N}(h)$ es el conjunto de pares cuya distancia $\|z_i - z_j\|_2$ cae en el intervalo asociado a la distancia h (Cressie, 1993; Schabenberger and Gotway, 2005). Después se ajusta un modelo teórico admisible, que permite evaluar la dependencia entre el punto de consulta y las muestras conocidas.

A partir del semivariograma ajustado, se definen el vector $\Gamma_0 \in \mathbb{R}^N$ y la matriz simétrica $\Gamma_{\mathcal{D}} \in \mathbb{R}^{N\times N}$ como

$$[\Gamma_0]_i = \gamma(z_0 - z_i), \qquad [\Gamma_{\mathcal{D}}]_{ij} = \gamma(z_i - z_j).$$

Entonces (7) puede escribirse como $\mathbb{E}[e^2(z_0)] = -\lambda^\top \Gamma_{\mathcal{D}} \lambda + 2\Gamma_0^\top \lambda$, lo que conduce al siguiente problema cuadrático con restricciones de igualdad:

$$\begin{aligned} \lambda^\star_{\mathrm{UK}}(z_0) = \underset{\lambda\in\mathbb{R}^N}{\arg\min} \quad & -\lambda^\top \Gamma_{\mathcal{D}} \lambda + 2\Gamma_0^\top \lambda \\ \text{s.a.} \quad & R\lambda = r(z_0). \end{aligned} \tag{8}$$

Para todo variograma admisible, $-\Gamma_{\mathcal{D}}$ es condicionalmente semidefinido positivo y el funcional anterior resulta convexo sobre el subespacio factible. Si además, los puntos disponibles en $\mathcal{D}$ son distintos, $-\Gamma_{\mathcal{D}}$ es condicionalmente *definido* positivo en dicho subespacio, y el problema admite un único mínimo. Finalmente, la predicción se obtiene sustituyendo $\lambda^\star_{\mathrm{UK}}(z_0)$ en (4).

3.1. Regularización ℓ_1 de los pesos

La solución de (8) es densa y, además, puede asignar pesos negativos de magnitud relevante, especialmente en presencia de muestras cercanas o redundantes, fenómeno relacionado con el efecto de apantallamiento (Stein, 1999). En aplicaciones de imitación de políticas de control, este comportamiento puede inducir extrapolaciones locales de una acción que se desea reconstruir principalmente a partir de acciones óptimas del controlador asociadas a estados próximos. Para favorecer soluciones más dispersas, robustas e interpretables, se incorpora una penalización ℓ_1 al problema (8), obteniendo

$$\begin{aligned} \lambda^\star_\alpha(z_0) = \underset{\lambda\in\mathbb{R}^N}{\arg\min} \quad & -\lambda^\top \Gamma_{\mathcal{D}} \lambda + 2\Gamma_0^\top \lambda + \alpha\|\lambda\|_1 \\ \text{s.a.} \quad & R\lambda = r(z_0), \end{aligned} \tag{9}$$

donde $\alpha > 0$ controla la intensidad de la regularización y $\|\lambda\|_1 = \sum_{i=1}^{N} |\lambda_i|$.

La penalización ℓ_1 promueve soluciones dispersas, de forma análoga a los métodos tipo Lasso (Hastie et al., 2015). Además, al combinarse con la restricción de insesgadez $\sum_i \lambda_i = 1$, penaliza indirectamente pesos negativos de gran magnitud, favoreciendo predicciones dominadas por contribuciones positivas de muestras relevantes (Carnerero et al., 2023; Matsui et al., 2025). En este sentido, el predictor regularizado queda sesgado hacia la *interpolación* frente a la extrapolación, lo cual resulta coherente con el uso local de los datos en *kriging* (Mariethoz and Caers, 2014, Cap. 2).

El problema (9) no admite la misma solución cerrada que (8), debido a la no diferenciabilidad de la norma ℓ_1. Para su solución eficiente, en este trabajo adoptamos el algoritmo K-ADMM propuesto en Moreno-Blazquez et al. (2026).

4. UK como ley de control

La estrategia propuesta reduce el coste computacional en línea del NMPC a costa de un cálculo previo fuera de línea. Para ello, se resuelven múltiples problemas NMPC asociados a diferentes estados considerados y se construye la base de datos $\mathcal{D}$ con las primeras acciones óptimas obtenidas. Después, en lazo cerrado, la acción se obtiene evaluando rápidamente el interpolante de *kriging*, sin resolver de nuevo el problema predictivo completo en cada instante de muestreo. Sin embargo, este enfoque solo es fiable siempre que la política π_{NMPC} sea suficientemente regular en el dominio de interés.

Sea $\mathcal{X}_{\mathrm{tr}} \subset \mathcal{X}$ un conjunto de estados conocidos. Para cada $x_i \in \mathcal{X}_{\mathrm{tr}}$ se resuelve el NMPC (1) y se almacena la primera acción óptima $u_i^\star = \pi_{\mathrm{NMPC}}(x_i)$. Por tanto, en la notación de *kriging* usada en la sección anterior, la base de datos de imitación es $\mathcal{D} = \{(z_i, \xi_i)\}_{i=1}^N = \{(x_i, u_i^\star)\}_{i=1}^N$, el regresor es $z_i = x_i$ y la salida interpolada es $\xi_i = u_i^\star$. En el caso escalar, $u_i^\star \in \mathbb{R}$; para sistemas multi-entrada, se puede entrenar un interpolante por componente o construir una extensión vectorial.

El controlador aprendido se define como $\pi_K(x) = \sum_{i=1}^{N} \lambda_i(x) u_i^\star$, donde los pesos $\lambda_i(x)$ se calculan con (8) o (9), usando $z = x$ como regresor. La acción aplicada al proceso es $u_k = \Pi_{\mathcal{U}}(\pi_K(x_k))$, donde $\Pi_{\mathcal{U}}$ proyecta la acción sobre el conjunto admisible de entradas.

5. Caso de estudio

La metodología se evalúa en el problema de *swing-up* de un péndulo invertido actuado por par. El objetivo es llevar el péndulo desde la posición inferior hasta el equilibrio vertical superior, manteniendo el par dentro de límites físicos. Este problema es representativo porque combina no linealidad, saturaciones durante la maniobra y un equilibrio inestable (Muskinja and Tovornik, 2006).

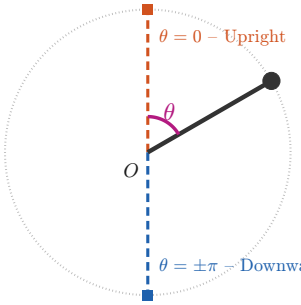

Figura 1: Definición del ángulo del péndulo. El origen $\theta = 0$ corresponde a la posición vertical superior, mientras que $\theta = \pm\pi$ representa la posición inferior.

La Fig. 1 fija la convención angular usada en toda la simulación. El estado es $x = \begin{bmatrix}\theta & \dot\theta\end{bmatrix}^\top$, donde θ es el ángulo y $\dot\theta$ la velocidad angular. La entrada u es el par aplicado en el pivote. La dinámica continua considerada es

$$\dot x_1 = x_2, \quad \dot x_2 = -\frac{u}{mL^2} + \frac{g}{L}\sin(x_1) - \frac{b}{mL}x_2,$$

con $m = 1$ kg, $L = 1{,}5$ m, $g = 9{,}81$ m/s^2 y $b = 0{,}001$ N · s. La entrada está limitada por $-15 \le u \le 15$ N · m. La discretización se realiza con periodo de muestreo $T_s = 0{,}1$ s.

El NMPC emplea $N_p = 10$ y $N_c = 5$, y regula el sistema hacia $x^{\text{ref}} = \begin{bmatrix}0 & 0\end{bmatrix}^\top$. La función de etapa usada para evaluar todos los controladores es

$$\ell(x_k, u_k, u_{k-1}) = \|x_k - x^{\text{ref}}\|^2_{\text{Q}} + \|u_k - u^{\text{ref}}\|^2_{\text{R}} + \|u_k - u_{k-1}\|^2_{\text{S}},$$

donde $\text{Q} = \text{diag}(9, 0{,}09)$, $\text{R} = 0{,}04$, $\text{S} = 0{,}01$, y $u^{\text{ref}} = 0$.

$$J = \sum_{k=1}^{N_{\text{sim}}} \ell(x_k, u_k, u_{k-1}), \tag{10}$$

con $T_{\text{sim}} = 10$ s ($N_{\text{sim}} = T_{\text{sim}}/T_s = 100$ instantes de control) y condición inicial $x_0 = [-\pi, 0]^\top$.

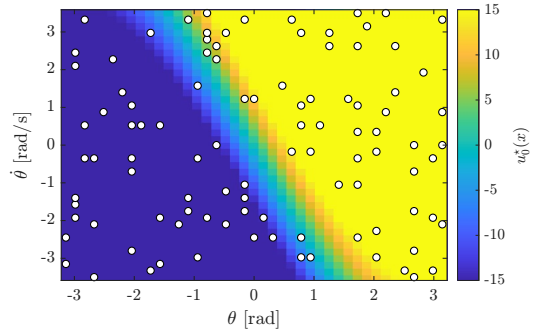

Figura 2: Mapa de acción obtenido resolviendo el NMPC en la malla densa y subconjunto de 100 muestras conocidas (puntos blancos) por los métodos *kriging*. El resto de puntos son utilizados para validación.

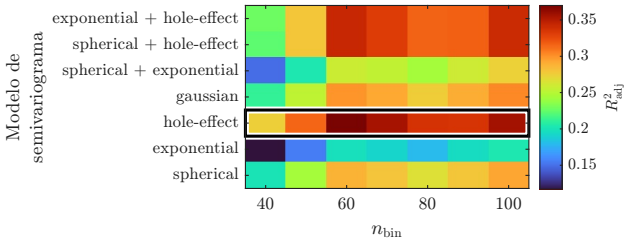

Figura 3: Barrido de selección del semivariograma. El eje horizontal indica n_{bin}, número de contenedores empleados para discretizar la distancia h en el semivariograma empírico. El rectángulo negro indica la combinación con mayor R^2_{adj}, correspondiente al modelo *hole-effect*. Puede apreciarse que este modelo es el mejor independientemente del número de contenedores.

5.1. Base de datos y selección del variograma

Para construir la política del NMPC se generó una malla densa de 41×41 estados en el plano $(\theta, \dot\theta)$. En cada punto se resolvió una vez el NMPC y se almacenó la primera acción óptima. En este caso, el regresor y la salida aprendida son:

$$z = x = \begin{bmatrix}\theta & \dot\theta\end{bmatrix}^\top, \quad \xi = \pi_{\text{NMPC}}(x) = u_0^\star(x).$$

Posteriormente se seleccionaron aleatoriamente 100 puntos de dicha malla como base de datos disponibles para la interpolación *kriging*, como se aprecia en la Fig. 2. El resto de puntos se mantiene para evaluar la calidad de la reconstrucción del mapa. En este caso, generar la malla densa de 41×41 estados requiere 1681 resoluciones del NMPC y tarda 37,45 s. En cambio, generar las 100 muestras usadas por *kriging* y ajustar la tendencia y el semivariograma tarda 4,24 s, reduciendo el coste previo en un factor 8,8.

En este caso, la tendencia de (5) se estima mediante un polinomio completo de cuarto orden en las variables de estado. En particular, para $z = [\theta\ \dot\theta]^\top$, el vector de funciones de tendencia se define como

$$r(z) = \begin{bmatrix}1, & \theta, & \dot\theta, & \theta^2, & \theta\dot\theta, & \dot\theta^2, & \ldots, & \theta^c\dot\theta^d\end{bmatrix}^\top, \qquad 1 \le c + d \le 4.$$

Así, la parte determinista se escribe como $\mu(z) = r(z)^\top\beta$, y el semivariograma se ajusta sobre los residuos $\delta_i = \xi_i - \mu(z_i)$, donde $\xi_i = u_0^\star(z_i)$. En la implementación, la base de datos se normaliza previamente, incluyendo tanto las variables de estado como la salida aprendida, con el fin de mejorar el acondicionamiento numérico del ajuste de la tendencia y del semivariograma. La estimación final de la acción, $\pi_K(x)$, se obtiene desnormalizando la predicción de *kriging*.

En cuanto al semivariograma, se ajustó a partir de las 100 muestras disponibles. Se probaron modelos esféricos, exponenciales, gaussianos, *hole-effect* y combinaciones anidadas, para distintos números de contenedores n_{bin} asociados a la distancia h, como se aprecia en la Fig. 3. El criterio para escoger la mejor parametrizacion fue el coeficiente de determinación ajustado, R^2_{adj}, calculado sobre el semivariograma empírico. Este indicador cuantifica la variabilidad explicada por el modelo e introduce una penalización por el número de parámetros, permitiendo comparar modelos variográficos de distinta complejidad. Con este criterio, la mejor configuración fue un modelo *hole-effect*, tomando en este caso $n_{\text{bin}} = 60$ de forma coherente con el número de muestras a disposición.

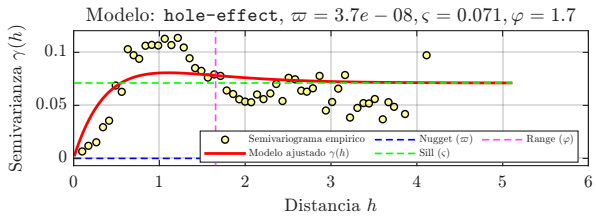

Figura 4: Semivariograma empírico del mapa de acción y modelo ajustado usado para construir las matrices de *universal kriging*.

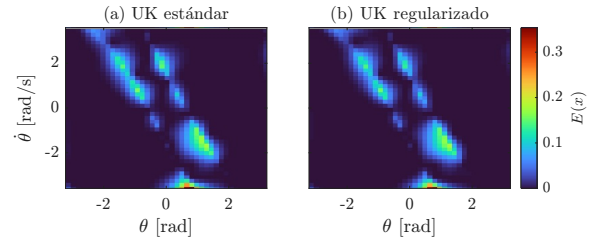

Figura 5: Mapas de $E(x)$ (eq. (11)) de las políticas aproximadas mediante UK estándar y UK regularizado respecto al mapa del NMPC. La magnitud interpolada es la primera acción óptima $u_0^\star(x)$ del NMPC. La similitud entre ambos mapas es esperable: la regularización preserva la fidelidad y actúa principalmente sobre la dispersión de los pesos.

La Fig. 4 muestra el ajuste finalmente empleado sobre el semivariograma empírico, el modelo *hole-effect*, cuya forma se expresa como:

$$\gamma(h) = \varpi + (\varsigma - \varpi)\left[1 - \frac{\sin(h/\varphi)}{h/\varphi}\right],$$

donde ϖ es el efecto pepita o *nugget*, ς es la meseta o *sill*, y φ es el parámetro de alcance o *range*. Este modelo es adecuado cuando la dependencia no decrece de forma estrictamente monótona con la distancia. En este caso, la acción óptima del péndulo presenta simetrías, saturaciones y transiciones propias de una maniobra de *swing-up*, por lo que pueden existir estados separados que requieran acciones similares.

Para la versión regularizada se empleó un peso uniforme de regularización $\alpha = 0{,}4$, escogido mediante validación sobre 100 puntos aleatorios de la malla densa.

5.2. *Reconstrucción de la política NMPC*

Para evaluar la fidelidad de las políticas estimadas respecto al controlador de referencia, la Fig. 5 muestra los mapas de error cuadrático normalizado obtenidos para UK estándar y UK regularizado. Si $\pi_K(x)$ denota la política estimada y $\pi_{\text{NMPC}}(x)$ la política real generada por el NMPC, se define

$$E(x) = \min\left\{1, \left(\frac{\pi_K(x) - \pi_{\text{NMPC}}(x)}{\max_{x_j} |\pi_{\text{NMPC}}(x_j)|}\right)^2\right\}, \qquad (11)$$

pudiendo comparar ambos métodos con una referencia común. Aunque la métrica queda acotada en $[0, 1]$, en la Fig. 5 la escala de color se ajusta al intervalo $[0, \max_x E(x)]$, considerando conjuntamente los errores de ambos métodos, con el fin de mejorar la resolución visual de las diferencias.

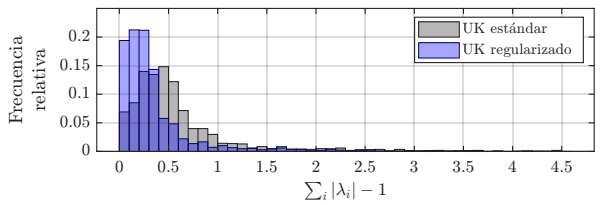

Figura 6: Distribución de pesos negativos efectivos sobre la malla de validación. La frecuencia relativa se expresa en tanto por uno.

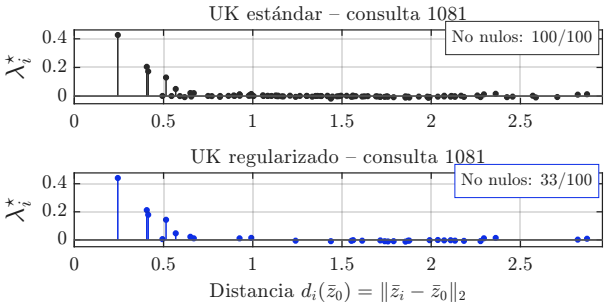

Figura 7: Pesos de *kriging* en función de la distancia normalizada al estado de consulta para un punto de validación.

Como se observa en la Fig. 5, los errores más elevados se concentran en las regiones de transición entre acciones positivas y negativas, donde la política del NMPC cambia de régimen debido a la saturación de la entrada y a la naturaleza no lineal del péndulo. Fuera de estas zonas, ambas aproximaciones preservan la estructura principal de la política de referencia.

5.3. *Análisis de pesos de kriging*

Para cuantificar el efecto de la regularización se analiza $\sum_{i=1}^{N} |\lambda_i(x)| - 1$; como la restricción $\sum_i \lambda_i = 1$ se cumple, valores próximos a cero indican pesos predominantemente no negativos, mientras que valores altos revelan combinaciones con pesos negativos de mayor magnitud. La Fig. 6 muestra que la versión regularizada concentra las predicciones en valores menores de $\sum_{i=1}^{N} |\lambda_i(x)| - 1$, lo que confirma que la penalización favorece interpolaciones más próximas a combinaciones convexas. En términos de control, esta propiedad es relevante porque reduce la dependencia de extrapolaciones algebraicas entre acciones óptimas del NMPC asociadas a estados alejados. La Fig. 7 muestra una consulta concreta. En UK estándar se utilizan prácticamente todos los pesos, mientras que el método regularizado selecciona un subconjunto menor de muestras activas. Además, los pesos activos se concentran en regiones más cercanas al estado de consulta, lo que proporciona una lectura geométrica de la regularización.

5.4. *Simulación en lazo cerrado*

La validación principal se realiza en lazo cerrado desde $x_0 = [-\pi, 0]^\top$ durante 10 s. Se comparan tres controladores: el NMPC, la política UK estándar y la política UK regularizada.

Tabla 1: Resultados de los controladores en bucle cerrado desde $x_0 = [-\pi, 0]^\top$: J es el coste acumulado (eq. (10)), $\bar{t}$ el tiempo medio de cálculo por acción en milisegundos, $\|x_f\|$ la norma del estado final.

Controlador	J	$\bar{t}$ [ms]	$\|x_f\|$
NMPC original 10/5	585.19	30.52	0.000
Política UK estándar	599.76	0.38	0.029
Política UK regularizada	596.94	0.90	0.018

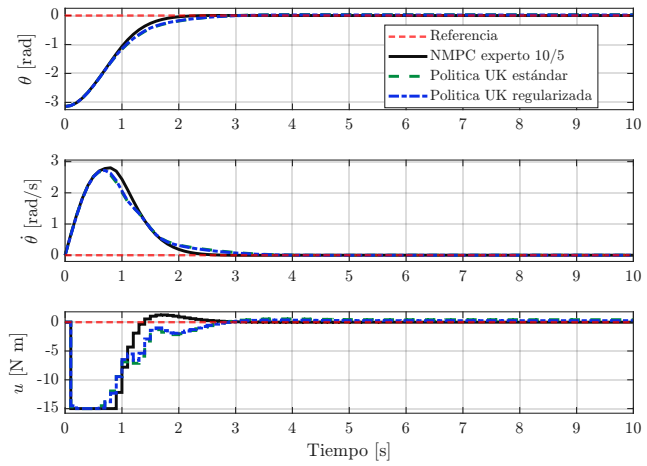

Figura 8: Evolución temporal de estados y entrada para el NMPC y las dos políticas imitadas.

La Fig. 8 muestra que las dos políticas imitadas reproducen la maniobra de swing-up y estabilización del NMPC con trayectorias muy próximas. La Tabla 1 recoge los resultados cuantitativos. Esta resume el compromiso entre coste cerrado y tiempo de cálculo. El NMPC obtiene $J = 585{,}19$ con un tiempo medio por acción de 30,52 ms. La política UK estándar obtiene $J = 599{,}76$, pero reduce el tiempo medio por acción a 0,38 ms. La política UK regularizada obtiene $J = 596{,}94$, con un tiempo medio de 0,90 ms. Ambas políticas estabilizan el péndulo cerca del equilibrio. La norma del estado final es $2{,}9 \cdot 10^{-2}$ para UK estándar y $1{,}8 \cdot 10^{-2}$ para UK regularizado, frente a $2{,}8 \cdot 10^{-7}$ en el NMPC. Por tanto, el NMPC conserva la mayor precisión final. No obstante, las políticas aprendidas reproducen la maniobra principal con un coste muy próximo y con una reducción sustancial de tiempo de cómputo.

6. Conclusiones

Este trabajo ha presentado una estrategia de imitación de políticas NMPC basada en *universal kriging*. La metodología usa el NMPC completo como controlador de referencia fuera de línea, almacena pares estado-acción y construye un interpolante *kriging* que aproxima directamente la primera acción óptima. En línea, el controlador resultante aplica la acción predicha por *kriging*, saturada dentro de los límites físicos del actuador, evitando resolver en cada instante problemas NMPC computacionalmente más exigentes.

En este caso, con pocas muestras conocidas, las políticas basadas en *kriging* pueden aproximar con alta fidelidad la acción del NMPC, mientras reducen el tiempo medio de cálculo por acción en varios órdenes de magnitud. La versión regularizada mantiene el rendimiento en bucle cerrado al tiempo que favorece interpolaciones más dispersas e interpretables.

Agradecimientos

Expresar nuestro agradecimiento a los revisores anónimos por sus comentarios para la mejora del artículo.

Esta publicación es parte del proyecto de investigación e innovación aplicada 2024/00000800, cofinanciado por la UE - Ministerio de Hacienda y Función Pública - Fondos Europeos - Junta de Andalucía – Consejería de Universidad, Investigación e Innovación, y del proyecto PID2022-142946NA-I00 financiado por MICIU/AEI /10.13039/501100011033 y por FEDER, UE. F. Fele también agradece la ayuda RYC2021-033960-I financiada por MICIU/AEI /10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR.

Referencias

Bemporad, A., Muñoz de la Peña, D., 2009. Multiobjective model predictive control. Automatica 45 (12), 2823–2830.

Binder, M., Darivianakis, G., Eichler, A., Lygeros, J., 2019. Approximate explicit model predictive controller using Gaussian processes. In: 2019 IEEE 58th Conference on Decision and Control (CDC). pp. 841–846.

Carnerero, A., Ramirez, D., Lucia, S., Alamo, T., 2023. Prediction regions based on dissimilarity functions. ISA Transactions 139, 49–59.

Cressie, N., Wikle, C. K., 2011. Statistics for spatio-temporal data. John Wiley & Sons.

Cressie, N. A., 1993. Statistics for Spatial Data. John Wiley & Sons, Inc.

Fabiani, F., Goulart, P. J., 2023. Reliably-stabilizing piecewise-affine neural network controllers. IEEE Transactions on Automatic Control 68 (9), 5201–5215.

Ferramosca, A., Limon, D., Camacho, E. F., 2014. Economic MPC for a changing economic criterion for linear systems. IEEE Transactions on Automatic Control 59 (10), 2657–2667.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical learning with sparsity: the lasso and generalizations. Chapman & Hall/CRC.

He, D., Li, H., Du, H., 2021. Lexicographic multi-objective MPC for constrained nonlinear systems with changing objective prioritization. Automatica 125, 109433.

Karg, B., Alamo, T., Lucia, S., 2021. Probabilistic performance validation of deep learning-based robust NMPC controllers. International Journal of Robust and Nonlinear Control 31 (18), 8855–8876.

Krige, D., 1962. Effective pay limits for selective mining. Journal of the South African Institute of Mining and Metallurgy,, 345–362.

Limon, D., Ferramosca, A., Alvarado, I., Alamo, T., 2018. Nonlinear MPC for tracking piece-wise constant reference signals. IEEE Transactions on Automatic Control 63 (11), 3735–3750.

Maddalena, E., da S. Moraes, C., Waltrich, G., Jones, C., 2020. A neural network architecture to learn explicit MPC controllers from data. IFAC-PapersOnLine 53 (2), 11362–11367, 21st IFAC World Congress.

Mariethoz, G., Caers, J., 2014. Multiple-point geostatistics. John Wiley & Sons, Ltd.

Matheron, G., 1963. Principles of geostatistics. Economic geology 58, 1246–1266.

Matsui, H., Yamamoto, K., Yamakawa, Y., Jun. 2025. Sparse estimation in kriging for functional data. Stochastic Environmental Research and Risk Assessment 39, 2413–2425.

Moreno-Blazquez, C., Fele, F., Alamo, T., 2026. Accelerated kriging interpolation for real-time grid frequency forecasting. Sustainable Energy, Grids and Networks 47, 102403.

Muskinja, N., Tovornik, B., 2006. Swinging up and stabilization of a real inverted pendulum. IEEE transactions on industrial electronics 53 (2), 631–639.

Nadales, J. M., Carnerero, A. D., Moreno-Blazquez, C., Haes-Ellis, R. M., Limon, D., 2023. Learning-based NMPC on SoC platforms for real-time applications using parallel lipschitz interpolation. IFAC-PapersOnLine 56 (2), 6298–6303, 22nd IFAC World Congress.

Parisini, T., Zoppoli, R., 1995. A receding-horizon regulator for nonlinear systems and a neural approximation. Automatica 31 (10), 1443–1451.

Rawlings, J. B., Mayne, D. Q., Diehl, M. M., 2017. Model Predictive Control: Theory, Computation, and Design, 2nd Edition. Nob Hill Publishing, Madison, Wisconsin.

Richter, S., Jones, C. N., Morari, M., 2012. Computational complexity certification for real-time MPC with input constraints based on the fast gradient method. IEEE Transactions on Automatic Control 57 (6), 1391–1403.

Rose, A., Schaub, P., Findeisen, R., 2025. Safe and high-performance learning of model predictive control using kernel-based interpolation. In: 2025 American Control Conference (ACC). pp. 91–96.

Schabenberger, O., Gotway, C. A., 2005. Statistical methods for spatial data analysis. Chapman and Hall/CRC.

Stein, M. L., 1999. Interpolation of spatial data: some theory for kriging. Springer.

Tokmak, A., Fiedler, C., Zeilinger, M. N., Trimpe, S., Köhler, J., 2025. Automatic nonlinear MPC approximation with closed-loop guarantees. IEEE Transactions on Automatic Control 70 (10), 6388–6403.

Williams, C. K., Rasmussen, C. E., 2006. Gaussian processes for machine learning. Vol. 2. MIT press Cambridge, MA.