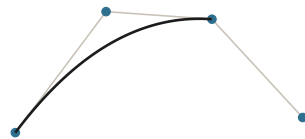


Universidad Nacional de Colombia · RC Trust - TU Dortmund



Introducción a los B-splines

*De los Fundamentos de Álgebra Lineal y Mínimos Cuadrados a la
Construcción de Curvas Paramétricas.*

Estudiante: Luismar Alvarez

Proyecto Pares Ordenados 2026

Mentor académico: Dr. Javier Aguilar

Agosto de 2026

Manizales, Caldas

Índice

1. Introducción y Marco del Proyecto	2
1.1. El problema del ajuste de curva	2
1.2. La naturaleza de la regresión	2
2. Fundamentos matemáticos: el problema de mínimos cuadrados y álgebra lineal esencial	2
2.1. El problema de mínimos cuadrados	2
2.2. Fundamentos Algebraicos: Conceptos Mínimos	4
2.3. El problema de optimización	7
2.4. La solución algebraica: convergencia del cálculo y la geometría	8
2.5. La evolución de la matriz de diseño: De lo global a lo local	8
2.5.1. Regresión polinomial clásica: El problema del soporte global	9
2.5.2. Splines lineales: Dividiendo el dominio	9
2.5.3. Splines polinomiales: Recuperando la suavidad	9
3. B-splines: Flexibilidad y control local	10
3.1. Definición Formal de Splines	10
3.2. Definición Formal de B-splines	11
3.3. Definición y construcción: La fórmula de Cox de Boor	11
3.4. Propiedades fundamentales	12
3.5. La matriz de diseño: La conclusión metodológica	13
4. Conclusiones:	13
4.1. Nota	14

1 Introducción y Marco del Proyecto

1.1 El problema del ajuste de curva

En el análisis de datos y la modelación matemática, uno de los objetivos principales es encontrar una función que describa de manera adecuada la relación entre un conjunto de variables observadas.

$$Y = f(X) + \varepsilon \quad (1)$$

Y es la variable respuesta, X la predictora, $f(X)$ la curva, y ε el ruido.

El método de mínimos cuadrados se introduce como una de las herramientas fundamentales para este propósito. Este método permite estimar los parámetros de un modelo al minimizar la suma de los errores cuadráticos, es decir, la diferencia al cuadrado entre los valores que realmente observamos y los que nuestro modelo predice.

1.2 La naturaleza de la regresión

Es un error común pensar que el método de mínimos cuadrados está estrictamente limitado a modelar relaciones que son geoméricamente líneas rectas. Es importante aclarar que la regresión se denomina lineal porque es lineal respecto a los parámetros que se desean estimar, pero no tiene que serlo respecto a los atributos o variables independientes.

Por ejemplo, si tenemos un conjunto de puntos $(x_1, y_1), \dots, (x_n, y_n)$ y suponemos que siguen un comportamiento parabólico, podemos aplicar mínimos cuadrados para hallar los coeficientes a , b , y c del modelo $y = a + bx + cx^2$. En esta expresión, aunque la variable x esté elevada al cuadrado, la relación sigue siendo lineal para los parámetros (a, b, c) . Por lo tanto, **el método de mínimos cuadrados es capaz de modelar curvas, con la única condición de que se fije previamente una forma funcional global y adecuada para todo el conjunto de datos.**

2 Fundamentos matemáticos: el problema de mínimos cuadrados y álgebra lineal esencial

En esta sección enunciaremos el problema de mínimos cuadrados y definiremos la teoría necesaria del álgebra lineal para entender este problema y su solución con la finalidad también de fijar una base para permitir un fácil entendimiento de manera posterior de la solución mediante B-splines y en qué se diferencia esta solución de una solución mediante una regresión lineal ordinaria.

2.1 El problema de mínimos cuadrados

Todo lo que cambia, de un modelo al siguiente, son las columnas de X . El álgebra que las resuelve no cambia nunca.

En la modelación estadística, el punto de partida es el modelo lineal general:

$$Y = X\beta + \varepsilon \quad (2)$$

donde $Y \in \mathbb{R}^n$ es el vector de respuestas observadas, $X \in \mathbb{R}^{n \times p}$ es la **matriz de diseño** (cuyas columnas representan las variables predictivas evaluadas en n observaciones), $\beta \in \mathbb{R}^p$ es el vector de parámetros a estimar, y $\varepsilon \in \mathbb{R}^n$ es el error aleatorio.

Estructuralmente, este modelo es la representación matricial del clásico sistema de ecuaciones lineales $Ax = b$. En el lenguaje de la regresión, la matriz de diseño X se define como matriz de coeficientes A , el vector de parámetros β es x y el vector de respuestas Y es b .

En la práctica, particularmente al trabajar con datos empíricos, el número de observaciones supera el de parámetros ($n > p$). Esta sobredeterminación implica que el sistema $X\beta = Y$ es, por lo general, inconsistente (No tiene solución Exacta). La presencia del ruido aleatorio ε provoca que el vector Y casi nunca pertenezca al espacio columna de la matriz de diseño, $\text{Col}(X)$. Al no existir una solución exacta, el objetivo se transforma en un **problema de optimización**: encontrar la mejor aproximación posible a la solución dentro de $\text{Col}(X)$ que minimice la distancia euclidiana respecto a Y , es decir, resolver:

$$\min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|^2 \quad (3)$$

Definimos geométricamente, la mejor aproximación como la **proyección ortogonal** del vector Y sobre el subespacio $\text{Col}(X)$. Denotemos esta proyección como $\hat{Y} = X\hat{\beta}$, la cual representa la mejor predicción posible dada la estructura del modelo. La condición analítica que caracteriza a esta proyección es que el vector residual (ε), $e = Y - \hat{Y}$, debe ser ortogonal a todo el espacio $\text{Col}(X)$ y, por ende, ortogonal a cada una de las columnas de X . En términos matriciales, esto exige que el producto interno entre las columnas de X y el residuo sea cero:

$$X^T e = \mathbf{0} \quad \implies \quad X^T (Y - X\hat{\beta}) = \mathbf{0} \quad (4)$$

reemplazando el error por su definición como la diferencia entre el vector de observaciones y la mejor estimación de este; y distribuyendo los términos, llegamos a las fundamentales ecuaciones normales:

$$X^T X \hat{\beta} = X^T Y \quad (5)$$

Este nuevo sistema de ecuaciones siempre es consistente y garantiza al menos una solución $\hat{\beta}$, reduciendo el problema de mínimos cuadrados a su resolución algebraica.

La unicidad de dicha solución depende de las propiedades intrínsecas de la matriz de diseño, descritas por el Teorema del Rango - Nulidad (enunciamos este teorema mas adelante):

- **Solución única:** Si $\text{rango}(X) = p$, las columnas de X son linealmente independientes. En consecuencia, la matriz $X^T X$ (de dimensiones $p \times p$) es definida positiva y, por lo tanto, invertible. La solución óptima está unívocamente determinada por:

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

- **Infinitas soluciones:** Si $\text{rango}(X) < p$, las columnas de X son linealmente dependientes (escenario común en modelos sobreparametrizados). En este caso, $X^T X$ es una matriz singular y el núcleo de X , $\text{Nul}(X)$, no es trivial. Existirán infinitas soluciones para $\hat{\beta}$, todas de la forma $\hat{\beta}_0 + v$, donde $\hat{\beta}_0$ es una solución particular y $v \in \text{Nul}(X)$.

Es importante destacar que, incluso cuando los estimadores $\hat{\beta}$ no son únicos, la predicción $\hat{Y} = X\hat{\beta}$ y el vector residual $e = Y - \hat{Y}$ sí lo son. La proyección ortogonal de un vector sobre un subespacio vectorial está determinada de manera única, independientemente de qué base geométrica o combinación de columnas se utilice para generar dicho subespacio.

2.2 Fundamentos Algebraicos: Conceptos Mínimos

Lo que sigue no es un repaso completo de álgebra lineal, sino algunos conceptos básicos para poder plantear, entender y, más adelante, resolver el problema de mínimos cuadrados, y para poder describir con precisión qué cambia entre el spline lineal, el polinomial y los B-Splines. Cada concepto se define de forma rigurosa, sin demostraciones, y se contextualiza hacia el uso que se le va a dar.

Definición 2.1 (Sistema lineal y eliminación de Gauss-Jordan). Resolver un sistema de ecuaciones lineales como el que tenemos significa preguntarse si el vector b de términos constantes es expresable como combinación lineal de las columnas de la matriz A : dado $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$, buscamos x tal que

$$Ax = b.$$

La eliminación de Gauss-Jordan responde a esta pregunta: al reducir la matriz aumentada $[A \mid b]$ a su forma escalonada reducida por renglones (RREF), y en esa forma la existencia y unicidad de x quedan a simple vista.

la pregunta de si b cae dentro o fuera del espacio que ellas generan es exactamente el que reaparece cuando se ajusta un modelo a datos. La diferencia es que ahí b casi nunca coincide con ese espacio: el ruido lo empuja fuera, el sistema se vuelve inconsistente, y es precisamente de esa inconsistencia de no poder resolver $Ax = b$ exactamente que nace el problema de mínimos cuadrados.

Definición 2.2 (Multiplicación matriz-vector y multiplicación de matrices). Dada $A \in \mathbb{R}^{m \times n}$ y $x \in \mathbb{R}^n$, el producto Ax definido como el producto punto entre cada fila de la matriz y el vector columna pero para fines de este estudio lo trabajaremos como **combinación lineal de las columnas** de A :

$$Ax = \sum_{j=1}^n x_j A_{:,j} \quad (6)$$

La multiplicación de dos matrices AB se entiende, bajo esta misma lógica, columna a columna: cada columna de AB es el producto de A por la columna correspondiente de B . Esta lectura columnar es la que hace visible, más adelante, que las predicciones de un modelo de regresión son combinaciones de las columnas de la matriz de diseño.

Definición 2.3 (Transformación lineal). Una función $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una transformación lineal si preserva sumas y escalares:

- $T(u + v) = T(u) + T(v)$
- $T(cu) = cT(u)$

Toda matriz $A \in \mathbb{R}^{m \times n}$ induce una transformación lineal mediante $T(x) = Ax$, y toda transformación lineal entre espacios de dimensión finita se representa así, para alguna matriz A . Pensar en A como una transformación y no solo como una tabla llena de números es lo que nos permite hablar, sin ambigüedad, del conjunto de todas las salidas posibles de un modelo.

Definición 2.4 (Subespacio, independencia lineal, base y dimensión). Un subconjunto no vacío $V \subseteq \mathbb{R}^n$ es un subespacio de $\mathbb{R}^n$ si y solo si cumple con las siguientes dos condiciones de cerradura:

1. $\forall x, y \in V \implies x + y \in V$ (Cerrado bajo la suma).
2. $\forall x \in V$ y $\forall a \in \mathbb{R} \implies ax \in V$ (Cerrado bajo la multiplicación escalar).

Los vectores $v_1, v_2, \dots, v_m \in V$ son linealmente independientes si son vectores únicos, no una combinación lineal de los demás vectores de su conjunto. Formalmente, el conjunto es *linealmente independiente* si la única forma de anular su combinación lineal es anulando todos los escalares:

$$\sum_{i=1}^m a_i v_i = 0 \implies a_1 = a_2 = \dots = a_m = 0, \quad a_i \in \mathbb{F}.$$

Si en cambio existen escalares $a_1, \dots, a_m \in \mathbb{F}$, no todos nulos, para los cuales la combinación se anula, el conjunto es *linealmente dependiente*: al menos uno de los vectores se puede expresar en términos de los otros.

Definiremos una **Base** $B = \{v_1, v_2, \dots, v_n\} \subset V$ como el conjunto mínimo de vectores que definen la estructura y dimensión de un espacio vectorial V sobre un cuerpo $\mathbb{F}$ y debe cumplir simultáneamente dos propiedades.

Primero, independencia lineal: el conjunto B no contiene redundancias, es decir la única combinación lineal que producen es el vector nulo.

Segundo, generación del espacio: B basta para construir cualquier vector de V ,

$$\text{span}(v_1, \dots, v_n) = V,$$

es decir, para todo $v \in V$ existen escalares $c_1, \dots, c_n \in \mathbb{F}$ tales que $v = \sum_{k=1}^n c_k v_k$.

Esta noción de base es la que distingue entre una base global de potencias y una base local de funciones spline: ambas generan un mismo espacio de curvas posibles, pero con propiedades muy distintas.

La dimensión de un espacio vectorial de dimensión finita V (denotada como $\dim V$) se define de la siguiente manera:

- $\dim V$ es la longitud (el número de vectores) de cualquiera de sus bases.

Si en cambio un espacio vectorial no puede ser generado por una lista finita de vectores, se clasifica como de dimensión infinita.

Definición 2.5 (Espacio columna y espacio nulo). Para $A \in \mathbb{R}^{m \times n}$:

- El **espacio columna** $\text{Col}(A) = \{Ax : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$ es el conjunto de todas las combinaciones lineales de las columnas de A el conjunto de todas las respuestas que el sistema puede producir.
- **Espacio nulo** (o kernel) de una matriz A . Es el conjunto de todos los vectores x que, al multiplicarlos por A , dan como resultado el vector cero:

$$\text{Nul}(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

El **rango** de A , $\text{rango}(A) = \dim(\text{Col}(A))$, es decir, el número máximo de columnas linealmente independientes de la matriz, indicando así cuantas columnas aportan información verdaderamente nueva.

Teorema 2.6 (Teorema de Rango-Nulidad / Teorema Fundamental de las Aplicaciones Lineales). Sea $A \in \mathbb{R}^{m \times n}$ la matriz representante de una transformación lineal $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ definida por $T(x) = Ax$. Se cumple que:

$$\text{rango}(A) + \dim(\text{Nul}(A)) = n$$

donde:

- $\text{rango}(A) = \dim(\text{Col}(A))$ representa la dimensión del espacio de columnas (o rango de la matriz).
- $\dim(\text{Nul}(A))$ representa la dimensión del espacio nulo (o nulidad de la matriz).
- n es el número de columnas de A (la dimensión del espacio de partida o dominio, equivalente al número de variables o incógnitas del sistema).

Definición 2.7 (Norma inducida por el producto interno). Sea V un espacio con producto interno sobre $\mathbb{F}$. La *norma inducida* de $v \in V$ es

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

En $\mathbb{R}^n$ con el producto interno canónico esto da $\|v\| = \sqrt{\sum_{i=1}^n v_i^2}$; en $L^2(a, b)$ da $\|f\| = \sqrt{\int_a^b f(x)^2 dx}$: la misma fórmula mide vectores discretos y funciones continuas. Como proviene de un producto interno, esta norma satisface

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2 \iff \langle u, v \rangle = 0.$$

Mínimos cuadrados minimiza, esta identidad es lo que permite reemplazar ese problema de minimización por una condición de ortogonalidad: en el óptimo, el residuo $e = y - X\hat{\beta}$ es perpendicular al espacio de columnas de X , y de ahí salen directamente las ecuaciones normales $X^T X \hat{\beta} = X^T y$.

Definición 2.8 (Proyección ortogonal y proyección sobre una base ortonormal). Dado un subespacio $W \subseteq \mathbb{R}^m$, todo vector y se descompone de manera única como $y = \hat{y} + z$, con $\hat{y} \in W$ y $z \in W^\perp$ (es decir, z ortogonal a todo vector de W).

El vector $\hat{y}$ es la **proyección ortogonal** de y sobre W , y es, entre todos los vectores de W , el más cercano a y . Cuando W tiene una **base ortonormal** $\{q_1, \dots, q_k\}$ vectores unitarios y mutuamente ortogonales, la proyección se calcula sin necesidad de resolver ningún sistema:

$$\hat{y} = \sum_{i=1}^k \langle y, q_i \rangle q_i.$$

Esta es, exactamente, la operación geométrica detrás del ajuste por mínimos cuadrados: buscar el punto del espacio columna de la matriz de diseño más cercano al vector de datos observado.

Definición 2.9 (Autovalores y autovectores). Un vector $x \neq 0$ es **autovector** de una matriz cuadrada $A \in \mathbb{R}^{n \times n}$, asociado al **autovalor** $\lambda \in \mathbb{R}$, si $Ax = \lambda x$: una dirección que A no rota, solo expande o contrae. Cuando A es simétrica como lo es siempre $A^T A$, sus autovalores son reales y no negativos.

Basta con tener esa propiedad: es la que explica por qué una matriz de la forma $A^T A$ se comporta de forma numéricamente estable o inestable según qué tan chicos sean sus autovalores más pequeños.

Ahora que hemos definido algunos conceptos importantes para comprender el problema de mínimos cuadrados, analizaremos este problema desde la perspectiva de la optimización, tratándolo como un problema de optimización.

2.3 El problema de optimización

Tomando los conceptos previamente definidos en la sección de fundamentación, en especial el de la norma inducida, minimizar la distancia $\|Y - X\beta\|$ es equivalente a minimizar su cuadrado lo que nos ahorra trabajar con raíces cuadradas al derivar.

$$\min_{\beta} \|Y - X\beta\| \iff \min_{\beta} \|Y - X\beta\|^2 \quad (7)$$

ahora definimos la función de pérdida (RSS - Suma de cuadrados de los residuos) como el cuadrado de la norma, es decir, el producto interior del vector de residuos consigo mismo:

$$S(\beta) = \|Y - X\beta\|^2 = (Y - X\beta)^T (Y - X\beta) \quad (8)$$

al desarrollar el producto de los binomios tenemos que la traspuesta del primer binomio es:

$$(Y - X\beta)^T = Y^T - \beta^T X^T.$$

luego multiplicamos los binomios

$$S(\beta) = (Y^T - \beta^T X^T)(Y - X\beta) = Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta.$$

ahora podemos simplificar los términos cruzados utilizando que todo escalar es igual a su propia traspuesta. Los términos $Y^T X\beta$ y $\beta^T X^T Y$ son escalares (de dimensión 1×1), por lo tanto

$$Y^T X\beta = (Y^T X\beta)^T = \beta^T X^T Y.$$

Así,

$$-Y^T X\beta - \beta^T X^T Y = -2\beta^T X^T Y.$$

llegando a la simplificación de la función de RSS

$$S(\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta.$$

2.4 La solución algebraica: convergencia del cálculo y la geometría

Al tratarse de una función cuadrática, encontrar el mínimo global de $S(\beta)$ requiere identificar su punto estacionario. Esto significa que debemos derivar la función con respecto al vector β e igualamos a 0.

Procedemos a calcular el gradiente de la función simplificada, analizando término por término:

$$\begin{aligned} S(\beta) &= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta, \\ \frac{\partial S}{\partial \beta} &= \frac{\partial}{\partial \beta}(Y^T Y) - 2\frac{\partial}{\partial \beta}(\beta^T X^T Y) + \frac{\partial}{\partial \beta}(\beta^T X^T X \beta) \\ &= 0 - 2X^T Y + 2X^T X \beta. \\ \frac{\partial S}{\partial \beta} &= 0 \implies -2X^T Y + 2X^T X \beta = 0 \end{aligned}$$

Agrupando estos resultados parciales, planteamos la ecuación de primer orden para la minimización:

$$\frac{\partial S}{\partial \beta} = -2X^T Y + 2X^T X \beta = 0.$$

Luego, al dividir toda la expresión por 2 y reorganizar los términos, llegamos a la siguiente ecuación, que es la misma a la que se llegó anteriormente al plantear el problema de mínimos cuadrados bajo el concepto de proyección ortogonal.

$$X^T X \beta = X^T Y.$$

Llegando de dos maneras a las mismas **ecuaciones normales**.

A la solución teórica de este sistema la denotaremos como $\hat{\beta}$. Para concluir y encontrar el valor explícito de $\hat{\beta}$, usando el Teorema de Rango-Nulidad definido antes. Si asumimos que las columnas de la matriz de diseño X son linealmente independientes, entonces la matriz cuadrada $X^T X$ es simétrica y definida positiva, lo cual garantiza de forma absoluta que es invertible.

Multiplicando ambos lados de la ecuación normal por la izquierda por la inversa $(X^T X)^{-1}$, logramos despejar el vector de parámetros y concluimos que, con esta ecuación, podemos resolver por completo nuestro problema de optimización por mínimos cuadrados :

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

2.5 La evolución de la matriz de diseño: De lo global a lo local

Hemos establecido el motor algebraico de mínimos cuadrados: la solución siempre será la misma. La diferencia fundamental entre un ajuste rígido y uno flexible radica exclusivamente en cómo construimos las columnas de la matriz de diseño X .

2.5.1. Regresión polinomial clásica: El problema del soporte global

El primer enfoque que podemos utilizar para ajustar datos que no siguen una línea recta es aumentar el grado del modelo utilizando un polinomio global de grado d :

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d.$$

En este modelo, las columnas de la matriz X están formadas por las potencias de la variable predictora: $1, x, x^2, \dots, x^d$. Aunque el álgebra de mínimos cuadrados resuelve este sistema sin problemas, esta base geométrica tiene una limitación importante que es el **soporte global**.

Dado que la función x^d está definida y es distinta de cero en todo el dominio, alterar un solo punto de los datos en el extremo izquierdo de la gráfica modificará los coeficientes β , lo cual alterará la curva en el extremo derecho. Además, al intentar capturar variaciones complejas aumentando el grado d , los polinomios tienden a oscilar violentamente en los extremos del intervalo. Necesitamos, por tanto, una estrategia que nos brinde control local.

2.5.2. Splines lineales: Dividiendo el dominio

Para solucionar el problema del soporte global, la estrategia es simple: dividir el problema en partes más pequeñas. Introducimos una serie de puntos de corte en el dominio del eje x llamados **nodos** (o *knots*), denotados como $\tau_1, \tau_2, \dots, \tau_k$.

Un spline lineal ajusta una recta diferente en cada intervalo definido por estos nodos, pero con la restricción matemática de que las rectas deben conectarse exactamente en los nodos para garantizar que la función final sea continua.

Al hacer esto, las columnas de nuestra matriz X cambian. Ya no usamos potencias globales, sino funciones que solo se activan o son diferentes de 0 en ciertas regiones del dominio. Esto nos permite **control local**: si los datos cambian en un intervalo, solo cambian las pendientes de ese vecindario, dejando el resto de la curva intacta. Sin embargo, surge un nuevo problema geométrico: aunque la curva es continua, las uniones en los nodos forman picos. La función no es diferenciable en los nodos, lo que definiremos como falta de suavidad.

2.5.3. Splines polinomiales: Recuperando la suavidad

Para eliminar los picos de los splines lineales sin perder el control local, **elevamos el grado matemático de los tramos. En lugar de conectar líneas rectas, conectamos polinomios de grado superior (generalmente grado 3, conocidos como splines cúbicos).**

El salto cualitativo aquí es que no solo seguimos teniendo la restricción de que los tramos polinomiales se unan en los nodos (manteniendo la continuidad de la función, $f(x)$), sino que también tenemos que sus derivadas coincidan en esos puntos de unión. Para un spline cúbico, forzamos a que la primera derivada (la pendiente, $f'(x)$) y la segunda derivada (la concavidad, $f''(x)$) sean iguales al pasar de un intervalo al siguiente por el nodo τ_i .

El resultado es una curva suave a la vista y al cálculo, que no cambia violentamente en los extremos y que mantiene un comportamiento local. Sin embargo, a nivel computacional, construir las columnas de la matriz X para que respeten todas estas condiciones de continuidad usando bases tradicionales (como las potencias truncadas) genera matrices $X^T X$ extremadamente inestables y casi singulares (Teóricamente imposible de invertir).

Debido a este problema de mal condicionamiento, es obligatorio recurrir a métodos numéricos de factorización matricial para aproximar la solución del sistema, tales como la Descomposición QR o la Descomposición en Valores Singulares (SVD).

En este punto nos planteamos una mejor solución también desde la parte computacional. Necesitamos una base matemática diferente que genere exactamente este mismo espacio de splines polinomiales suaves, pero cuyas columnas sean numéricamente estables, estrictamente locales y fáciles de computar. Esa base definitiva, así introducimos el método de **B-splines**.

Nota: Dado que el desarrollo e implementación de estos algoritmos numéricos desde cero no es el objetivo de este trabajo, para nuestro análisis práctico, Usamos una implementación computacional ya hecha en Python.

3 B-splines: Flexibilidad y control local

Como hemos establecido al final de la sección anterior, el problema analítico de los splines polinomiales tradicionales radica en la base matemática que se elige para construir el espacio de funciones. Con nuestros conocimientos anteriores de álgebra lineal, la teoría de espacios vectoriales y la geometría podemos dar paso a una solución mucho más robusta y teóricamente mejor estructurada: los **B-splines** (cuyo nombre proviene de *Basis splines* o splines base).

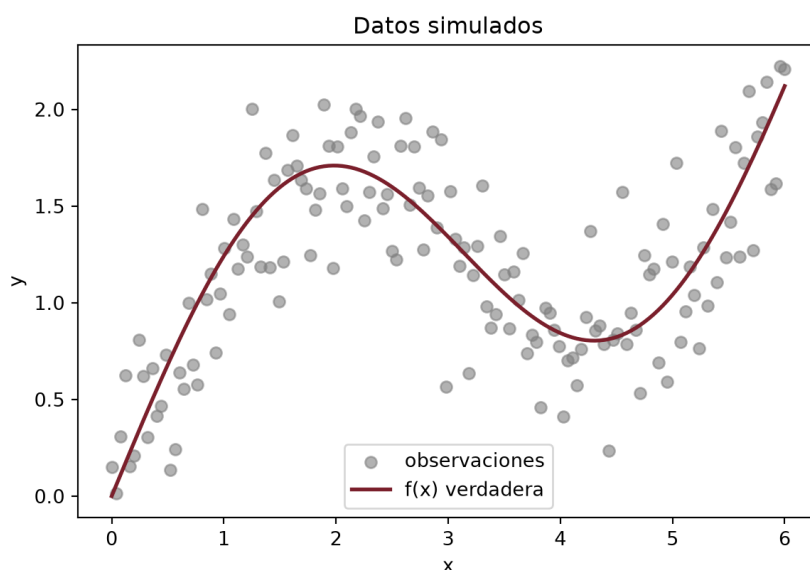


Figura 1: Datos simulados con curva de ajuste B-spline

3.1 Definición Formal de Splines

Sean una partición o secuencia de nodos $\tau_0 < \tau_1 < \dots < \tau_m$ sobre un dominio real. Un **spline** $s(u)$ de grado n es una función polinómica por trozos definida de manera que en cada intervalo intermodal $[a_i, a_{i+1}]$ es un polinomio de grado a lo sumo n , y tiene derivadas continuas hasta el orden k en los nodos (es decir, es de clase C^k).

3.2 Definición Formal de B-splines

Los **B-splines** (Basis splines) constituyen una base óptima para el espacio de los splines descritos anteriormente. Cada función base $N_i^n(u)$ de grado n se define formalmente como una función polinómica por tramos caracterizada por un **soporte compacto**:

$$\text{supp } N_i^n = [a_i, a_{i+n+1}]$$

Esto significa rigurosamente que la función base satisface:

- $N_i^n(u) > 0$ para todo $u \in (a_i, a_{i+n+1})$.
- $N_i^n(u) = 0$ para todo $u \notin [a_i, a_{i+n+1}]$.

Adicionalmente, los B-splines cumplen con la propiedad de **partición de la unidad** en el dominio de interés, asegurando que:

$$\sum_i N_i^n(u) = 1$$

Los B-splines no representan un nuevo tipo de curva, sino una **base matemática óptima** para el espacio de funciones spline. Su construcción teórica resuelve de manera simultánea la exigencia de suavidad global y la necesidad de contar con un conjunto generador linealmente independiente y bien acondicionado.

3.3 Definición y construcción: La fórmula de Cox de Boor

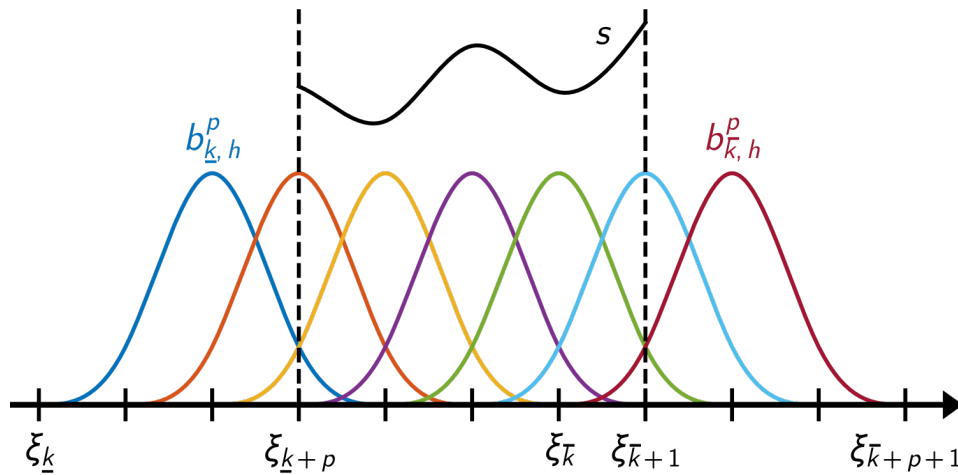


Figura 2: Representación de las funciones base B-spline utilizadas en la construcción del espacio de splines.

Un B-spline es una función polinómica por tramos que sirve como función base. A diferencia de los polinomios globales como x^2 o x^3 , que tienen valores distintos de cero en casi toda la recta real, un B-spline se define de manera estrictamente local sobre una secuencia de nodos τ .

La particularidad analítica de los B-splines radica en que no se definen a través de la resolución simultánea de un extenso sistema de ecuaciones acopladas, sino mediante la recursión matemática conocida como la **fórmula de Cox-de Boor**. de esta forma se construyen splines de grado superior combinando linealmente splines de grado inferior.

Para un conjunto de nodos no decrecientes $\tau_0 \leq \tau_1 \leq \dots \leq \tau_m$, el B-spline de grado 0 (el caso base, que es un spline constante a trozos) se define formalmente como una función indicadora simple:

$$B_{i,0}(x) = \begin{cases} 1 & \text{si } \tau_i \leq x < \tau_{i+1}, \\ 0 & \text{en otro caso.} \end{cases} \quad (9)$$

Para cualquier grado $p \geq 1$, el B-spline se construye de forma recursiva como una combinación convexa de dos B-splines de grado $p - 1$:

$$B_{i,p}(x) = \frac{x - \tau_i}{\tau_{i+p} - \tau_i} B_{i,p-1}(x) + \frac{\tau_{i+p+1} - x}{\tau_{i+p+1} - \tau_{i+1}} B_{i+1,p-1}(x).$$

Convención algebraica: Si el denominador de alguna de las fracciones es cero (lo cual ocurre analíticamente cuando hay nodos de multiplicidad mayor a uno), ese término completo se define por convención como cero.

3.4 Propiedades fundamentales

La formulación de Cox de Boor dota a los B-splines de propiedades estructurales que los hacen mejores frente a las bases polinomiales ordinarias para propósitos de análisis funcional y teoría de aproximación:

- **Soporte compacto (Control estrictamente local):** Un B-spline $B_{i,p}(x)$ es estrictamente mayor que cero solo en el intervalo abierto (τ_i, τ_{i+p+1}) y es exactamente cero en cualquier otro lugar del dominio. Esto significa que la alteración de un coeficiente lineal asociado a este B-spline modificará la curva final únicamente en ese subintervalo.
- **No negatividad:** Para todo x y para todo i, p , se cumple la desigualdad $B_{i,p}(x) \geq 0$. La positividad estricta dentro de su soporte garantiza un comportamiento analítico monótono en la combinación convexa y asegura el buen condicionamiento de las matrices de diseño asociadas.
- **Partición de la unidad:** La suma de todos los B-splines de grado p evaluados en cualquier punto x del dominio es exactamente 1:

$$\sum_i B_{i,p}(x) = 1.$$

Esta propiedad geométrica asegura que la curva aproximada siempre permanecerá dentro de la zona definida por sus coeficientes, garantizando un comportamiento invariante bajo transformaciones afines y estrictamente acotado.

3.5 La matriz de diseño: La conclusión metodológica

Es este el punto más importante que identifica y diferencia los B-splines de sus intentos de soluciones previas. Habíamos establecido que todo lo que cambia, de un espacio de representación a otro, son las columnas de X , el motor algebraico de proyección ortogonal que las resuelve permanece invariante.

Si buscamos aproximar una curva suave a un conjunto de observaciones utilizando B-splines, el problema de optimización en el espacio euclidiano sigue siendo exactamente el mismo problema de minimización de la norma (El mismo problema de mínimos cuadrados):

$$\min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|^2.$$

Y la solución analítica sigue siendo la misma proyección ortogonal deducida mediante las ecuaciones normales:

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

¿Qué es lo que cambia de manera definitiva desde la geometría del problema?

La estructura del subespacio generado por las columnas de la matriz de diseño X . En lugar de construir las columnas de X con una base monomial global (como $1, x, x^2, \dots$), evaluamos nuestras funciones base de soporte compacto en cada uno de los puntos x_j . La matriz X adopta entonces la siguiente estructura:

$$X = \begin{bmatrix} B_{1,p}(x_1) & B_{2,p}(x_1) & \dots & B_{k,p}(x_1) \\ B_{1,p}(x_2) & B_{2,p}(x_2) & \dots & B_{k,p}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_{1,p}(x_n) & B_{2,p}(x_n) & \dots & B_{k,p}(x_n) \end{bmatrix}.$$

Debido a la propiedad de soporte finito, la mayoría de los B-splines se anulan en cualquier punto dado x_j donde no tenga soporte. Por esta razón, esta nueva matriz X está en su mayoría llena de ceros, dotando al producto $X^T X$ de una estructura de banda fuertemente diagonal, simétrica y definida positiva.

4 Conclusiones:

Al cambiar de el enfoque rígido de los subespacios generados por polinomios globales en favor de bases de soporte compacto (B-splines), adquirimos la capacidad de proyectar el vector de respuestas Y sobre un subespacio cuyas bases capturan variaciones estrictamente locales. Esto nos da una enorme libertad geométrica para aproximar formas funcionales altamente complejas sin que una perturbación en una región del dominio altere globalmente la función.

Sin embargo, a medida que dotamos al modelo de mayor flexibilidad funcional aumentando la dimensión del espacio de parámetros β , corremos el riesgo de que, si el número de funciones base crece de manera desproporcionada respecto a los puntos evaluados, o si su soporte induce

solapamientos redundantes (porque sus nodos están demasiado juntos), la matriz de diseño X pueda no tener rango completo por columnas.

Cuando X pierde su rango completo, la matriz producto $X^T X$ deja de ser definida positiva y, en consecuencia, carece de inversa. En este escenario de cuasisingularidad, el sistema de ecuaciones normales $X^T X \hat{\beta} = X^T Y$ ya no posee una solución única, lo que implica que los valores del vector de coeficientes $\hat{\beta}$ pierden su significado unívoco e intrínseco. Como establece la teoría matemática definida al inicio de este trabajo, sin imponer restricciones adicionales sobre el modelo, la solución deja de ser única y los parámetros se vuelven no identificables.

4.1 Nota

Nota computacional: El núcleo de este proyecto es el desarrollo teórico. Sin embargo, como material de apoyo para visualizar estos conceptos, se ha elaborado una implementación básica de B-splines en Python. El notebook está disponible en el siguiente enlace: luismath.me/Pares-Ordenados/notebook.

Referencias

- [1] Strang, G. (1988). *Linear Algebra and Its Applications* (3.a ed.). Thomson Learning.
- [2] Paluszny, M., Prautzsch, H., & Boehm, W. (2005). *Métodos de Bézier y B-splines*. Universitätsverlag Karlsruhe.
- [3] Floater, M. S. (2023). *An Introduction to Spline Theory*. Notas de curso, Universidad de Oslo.
- [4] van de Geer, S. A. (2005). Least Squares Estimation. En *Encyclopedia of Statistics in Behavioral Science* (pp. 1041–1045). John Wiley & Sons.
- [5] Miller, S. J. *The Method of Least Squares*. Brown University.
- [6] Jørgensen, B. *Theory of Linear Models*.
- [7] Nugent, J. (s.f.). *Splines*. Disponible en: <https://joshua-nugent.github.io/splines/>. Accedido: 15 de agosto de 2026.
- [8] Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., y Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, 19(46).
- [9] Denoeux, T. (2016). Lecture 7: Splines and Generalized Additive Models. *Computational Statistics* (notas de clase).
- [10] Fernandes, K., Vinagre, P., y Cortez, P. (2015). A proactive intelligent decision support system for predicting the popularity of online news. En *Proceedings of the 17th EPIA Conference on Artificial Intelligence* (pp. 535–546). Springer.