

Un Enfoque de Optimización Global para Regresión No Lineal mediante el Cubic Algorithm

Rodrigo Serna Pérez
(Dated: 6 de octubre de 2025)

Este trabajo aborda el problema de la regresión no lineal mediante técnicas de optimización global. Se presenta la reformulación de problemas de ajuste de modelos como la minimización de una función de pérdida, la cual puede ser no convexa y multimodal. Se introduce el marco teórico de las funciones Lipschitzianas, fundamental para garantizar la convergencia de algoritmos deterministas. Se detalla el funcionamiento del "Cubic Algorithm", un método de tipo Branch and Bound que localiza regiones de ϵ -óptimos. Finalmente, se aplican estos conceptos en una serie de experimentos numéricos con modelos de regresión logística, demostrando la superioridad del enfoque global frente a métodos locales como L-BFGS-B, no solo para encontrar mejores soluciones, sino también para diagnosticar la estructura del problema de optimización.

I. INTRODUCCIÓN

En el ámbito de la modelización matemática, la optimización ocupa un lugar central. Dentro de este contexto, la *optimización global* emerge como una disciplina especialmente relevante cuando el dominio de la función objetivo presenta múltiples mínimos locales, lo cual imposibilita la aplicación directa de técnicas convencionales.

Este trabajo se enmarca en el estudio de técnicas de optimización global aplicadas a problemas de regresión no lineal. La motivación principal surge de la constatación de que muchos modelos de regresión no son adecuadamente abordados mediante métodos de optimización local, debido a la no convexidad de las funciones de pérdida asociadas. Como consecuencia, el riesgo de quedar atrapado en mínimos locales subóptimos compromete tanto la calidad del ajuste como la interpretabilidad del modelo.

El principal desafío en la regresión no lineal es la estimación de parámetros que minimicen el error. Para ello, exploramos el "Cubic Algorithm", un método de tipo *Branch and Bound* que garantiza teóricamente la convergencia a una región de ϵ -óptimos con precisión arbitraria, siempre que se disponga de una constante de Lipschitz de la función objetivo.

II. REGRESIÓN COMO PROBLEMA DE OPTIMIZACIÓN GLOBAL

La regresión no lineal consiste en ajustar un modelo paramétrico $y = f(x; \theta)$ a un conjunto de datos $\{(x_i, y_i)\}_{i=1}^N$. El objetivo es encontrar el vector de parámetros θ que mejor se ajusta a los datos, lo cual supone resolver el sistema de ecuaciones $\{f(x_i; \theta) = y_i\}_{i=1}^N$.

Este sistema puede reducirse a un problema de optimización global mediante la incorporación de una función de pérdida. Lo más usual es buscar la minimización del error cuadrático, formulando el problema como:

$$\min_{\theta \in \Theta} \mathcal{L}(\theta) = \sum_{i=1}^N (f(x_i; \theta) - y_i)^2 \quad (1)$$

donde Θ es el dominio de búsqueda de los parámetros. A diferencia de los métodos locales, la optimización global no se limita a encontrar un único punto, sino que busca el mínimo absoluto de $\mathcal{L}(\theta)$, proporcionando además información sobre la estructura de la solución. Por ejemplo:

- Si la región de ϵ -óptimos es amplia, puede indicar que el modelo está mal condicionado o que los parámetros no son identificables.
- Si tiene componentes no conexas, puede sugerir que el modelo no captura adecuadamente la estructura del problema.

III. FUNCIONES LIPSCHITZIANAS

La noción de continuidad de Lipschitz es una condición más fuerte que la continuidad uniforme y es fundamental para los algoritmos de optimización global deterministas.

Definición 1 (Función Lipschitziana). *Sea una función $f : D \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$. Decimos que f es Lipschitziana si existe una constante $L \geq 0$ tal que para todo par de puntos $x, y \in D$:*

$$\|f(x) - f(y)\| \leq L\|x - y\|.$$

La constante L se denomina constante de Lipschitz de f .

Los siguientes teoremas son cruciales para estimar la constante de Lipschitz en problemas de regresión.

Teorema 1. *Sea $f : D \subset \mathbb{R}^m \rightarrow \mathbb{R}$ una función diferenciable en un dominio convexo D . Entonces, f es Lipschitziana si y solo si su gradiente está acotado en D . La constante de Lipschitz es $L = \sup_{x \in D} \|\nabla f(x)\|$.*

Demostración. Por el Teorema del Valor Medio, para cualquier $x, y \in D$, existe c en el segmento que une x e y tal que $f(y) - f(x) = \nabla f(c) \cdot (y - x)$. Tomando el valor absoluto y aplicando la desigualdad de Cauchy-Schwarz:

$$|f(y) - f(x)| = |\nabla f(c) \cdot (y - x)| \leq \|\nabla f(c)\| \|y - x\|.$$

Si el gradiente está acotado por $L = \sup_{z \in D} \|\nabla f(z)\|$, entonces:

$$|f(y) - f(x)| \leq L\|y - x\|,$$

lo que prueba que f es Lipschitziana. Recíprocamente, si f es Lipschitziana, el cociente $|f(y) - f(x)|/\|y - x\|$ está acotado por L . Al tomar el límite cuando $y \rightarrow x$, este cociente se convierte en la norma del gradiente en x , demostrando que el gradiente está acotado. ■

Teorema 2. Sea $F(x) = (f_1(x), \dots, f_n(x))$ una función vectorial donde cada componente f_i es Lipschitziana con constante L_i . Entonces F es Lipschitziana con constante $L = \sqrt{\sum_{i=1}^n L_i^2}$.

Demostración. Consideramos la norma euclídea. La diferencia al cuadrado es:

$$\|F(x) - F(y)\|^2 = \sum_{i=1}^n (f_i(x) - f_i(y))^2.$$

Dado que cada f_i es Lipschitziana con constante L_i :

$$(f_i(x) - f_i(y))^2 \leq (L_i\|x - y\|)^2 = L_i^2\|x - y\|^2.$$

Sumando sobre todas las componentes:

$$\|F(x) - F(y)\|^2 \leq \sum_{i=1}^n L_i^2\|x - y\|^2 = \left(\sum_{i=1}^n L_i^2 \right) \|x - y\|^2.$$

Tomando la raíz cuadrada, se obtiene $\|F(x) - F(y)\| \leq \sqrt{\sum L_i^2}\|x - y\|$, con $L = \sqrt{\sum L_i^2}$. ■

Teorema 3. Sea $F: \mathbb{R}^m \rightarrow \mathbb{R}^n$ una función Lipschitziana con constante L y sea $l(x) = \|F(x)\|_p$. Entonces l es una función lipschitziana con la misma constante L .

Demostración. Por la desigualdad triangular inversa (o generalizada):

$$|\|F(x)\|_p - \|F(y)\|_p| \leq \|F(x) - F(y)\|_p.$$

Dado que F es Lipschitziana con constante L , tenemos:

$$\|F(x) - F(y)\|_p \leq L\|x - y\|.$$

Combinando ambas desigualdades:

$$|l(x) - l(y)| = |\|F(x)\|_p - \|F(y)\|_p| \leq L\|x - y\|.$$

Esto demuestra que $l(x)$ es Lipschitziana con la misma constante L . ■

IV. EL CUBIC ALGORITHM

El Cubic Algorithm es un método de tipo *Branch and Bound* (B&B) para la optimización global de funciones Lipschitzianas. Los algoritmos B&B dividen recursivamente el dominio de búsqueda en subdominios más pequeños (*branching*) y descartan aquellos que no pueden

contener una solución óptima mediante el uso de cotas (*bounding*).

El Cubic Algorithm opera sobre un cuboide inicial D_0 y requiere una constante de Lipschitz L para la función objetivo $f(\theta)$. Su funcionamiento se detalla a continuación. Para una descripción más exhaustiva, junto con la demostración de convergencia y la implementación, véase [1].

1. Se inicializa una lista de cuboides activos, $\mathcal{C}$, con D_0 .
2. En cada iteración, se calcula una cota inferior para f en cada cuboide $C \in \mathcal{C}$ usando la propiedad de Lipschitz: $f(\theta) \geq f(c) - L \cdot \text{diag}(C)/2$, donde c es el centro de C .
3. Se actualiza el valor mínimo global encontrado, $f_{\min}$.
4. Se podan todos los cuboides cuya cota inferior sea mayor que $f_{\min}$.
5. Los cuboides restantes se subdividen en k^n piezas más pequeñas, donde n es la dimensión y k es un parámetro.
6. El proceso se repite hasta que el error estimado, $L \cdot \text{diag}(C)$, es menor que una tolerancia ϵ .

El resultado del algoritmo es un conjunto de cuboides que conforman la región de ϵ -óptimos.

V. EXPERIMENTOS

En esta sección se presentan los resultados de aplicar el Cubic Algorithm a diferentes problemas de regresión no lineal. El objetivo es doble: por un lado, validar la eficacia del método para encontrar óptimos globales; por otro, compararlo con un algoritmo de optimización local estándar, L-BFGS-B [4], para resaltar las ventajas del enfoque global. Todos los experimentos se centran en el ajuste de un modelo de regresión logística, un caso de estudio relevante por su amplio uso y su naturaleza no lineal.

A. Experimento 1: Ajuste Exacto con 2 Puntos

Se generaron dos puntos sin ruido con $\theta_{\text{real}} = (1, 4, 2)$. Se buscó la solución en la caja $[0, 3] \times [0, 25]$.

1. Estimación de la Constante de Lipschitz

El gradiente de $f(x; \theta) = \frac{\theta_1}{1 + e^{-\theta_2 x}}$ es $\nabla f = (\text{sig}(\theta_2 x), \theta_1 x \text{sig}'(\theta_2 x))$. Usando que $|\text{sig}(t)| \leq 1$ y $|\text{sig}'(t)| \leq 1/4$, y dado que $\theta_1 \leq 3$, la cota para la norma del gradiente en un punto x es $L(x) = \sqrt{1 + 3^2 x^2 / 16}$.

Para $x_1 = 1, x_2 = 2$, y usando el Teorema 2, la constante global es $L = \sqrt{L(x_1)^2 + L(x_2)^2} \approx 2,19$.

2. Resultados

El Cubic Algorithm encontró el mínimo global en $\theta \approx (1,4, 2)$ con error casi nulo. En cambio, L-BFGS-B convergió a un mínimo local en $\theta = (1,3, 12,5)$ con un error de 0,1. La Tabla I y la Fig. 1 resumen el proceso.

Cuadro I: Resultados del Experimento 1.

Iter.	Diag.	Cubos Num.	Cubos	Mínimo
1	12.589	4	0.294778	
5	0.786	904	0.053344	
10	0.024	2024	0.000793	
15	0.0007	1924	0.000050	
24	0.000003	504	0.000000	

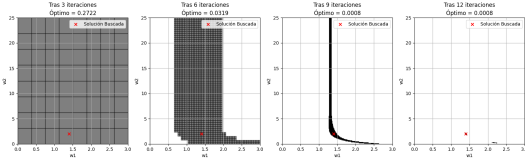


Fig. 1: Primeras iteraciones en el Experimento 1.

B. Experimento 2: Ajuste con 20 Puntos y Ruido

Se añadieron 20 muestras con ruido gaussiano.

1. Resultados

La constante de Lipschitz se estimó en $L = 16,6$. El algoritmo local L-BFGS-B encontró un mínimo de 0,14. El Cubic Algorithm encontró una región de óptimos con un mínimo de 0,037. El centro de esta región, $\theta^* = (1,397, 2,06)$, es una estimación puntual muy cercana a la real (ver Fig. 2 y 3).

C. Experimento 3: Modelo con Tres Parámetros

Se usó el modelo $f(x; \theta) = \theta_1 / (1 + e^{\theta_3 - \theta_2 x})$ en la caja $[0, 3] \times [0, 5] \times [0, 5]$.

1. Estimación de la Constante de Lipschitz

El gradiente es $\nabla f = (\text{sig}(\cdot), -\theta_1 x \text{sig}'(\cdot), \theta_1 \text{sig}'(\cdot))$. La cota para la norma del gradiente es $L(x) =$

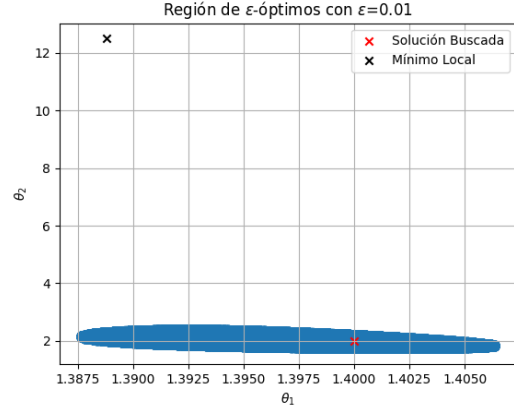


Fig. 2: Región de ϵ -mínimos en el Experimento 2.

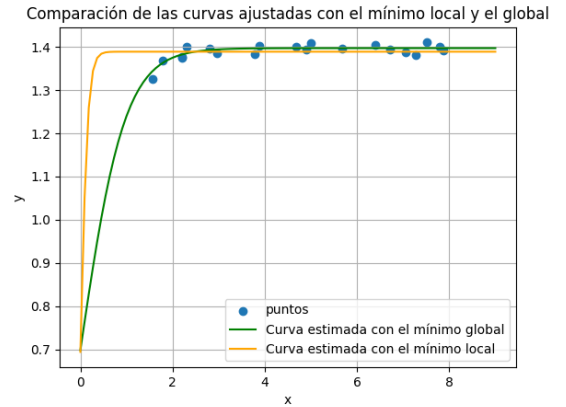


Fig. 3: Comparación de ajustes en el Experimento 2.

$\sqrt{1 + \frac{b^2}{16}(1 + x^2)}$. Con los 20 puntos del experimento se obtuvo $L = 11,06$.

2. Resultados

El mínimo local fue de 0,14, mientras que el global fue de 0,137. Lo más revelador fue que la región de ϵ -óptimos (Fig. 4) era muy extensa, indicando que el modelo estaba mal condicionado y los parámetros no eran identificables. Esta conclusión crucial es inalcanzable para un método de optimización local.

VI. CONCLUSIONES

Este trabajo ha demostrado la eficacia de aplicar el Cubic Algorithm a problemas de regresión no lineal. El enfoque global no solo encuentra consistentemente mejores soluciones que los métodos locales, sino que también ofrece una visión más profunda del problema. La capacidad de identificar regiones de óptimos permite diagnos-

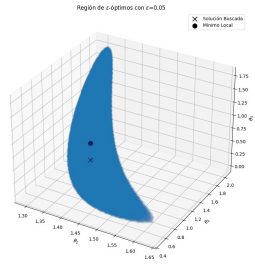


Fig. 4: Región de ϵ -óptimos en el experimento 3.

ticar problemas como la no identificabilidad de parámetros, una ventaja fundamental sobre los optimizadores locales.

-
- [1] R. Serna Pérez, *Optimización Global Aplicada a Regresión No Lineal* (Trabajo de Fin de Grado, Universidad Ficticia de España, 2024).
 - [2] A. Pazmiño, M. T. González, and O. C. Medina, INNOVA Research Journal **1**, 1 (2015).
 - [3] N. Lepore et al., Forecasting **6**, 128 (2024).
 - [4] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, SIAM J. Sci. Comput. **16**, 1190 (1995).
 - [5] P. Virtanen et al., Nature Methods **17**, 261 (2020).