



Matemáticas: Enseñanza Universitaria

ISSN: 0120-6788

reviserm@univalle.edu.co

Escuela Regional de Matemáticas

Colombia

Arévalo S., Alexander; Martínez R., Héctor J.; Sanabria R., Ana M.
Cálculo eficiente del estimador Jackknife agrupado para mínimos cuadrados lineales
Matemáticas: Enseñanza Universitaria, vol. XX, núm. 2, diciembre, 2012, pp. 55-68
Escuela Regional de Matemáticas
Cali, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=46826273005>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Cálculo eficiente del estimador Jackknife agrupado para mínimos cuadrados lineales

Alexander Arévalo S.
Universidad del Valle

Héctor J. Martínez R.
Universidad del Valle

Ana M. Sanabria R.
Universidad del Valle

Recibido Ago. 30, 2011

Aceptado Ago. 01, 2012

Abstract

In this paper, we generalize the results obtained by Martínez and Sanabria to calculate the Jackknife Estimator for Linear Least Squares, which express the estimates of the subproblems that result when calculating the Grouped Jackknife Estimator for the Linear Least Squares Problem (GJELLS) in terms of the initial estimate and other simple expressions to calculate, and thus modify the standard algorithm to calculate the GJELLS.

This modification reduces the number of operations of the order $O(\frac{m^2n^2}{h}) + O(\frac{m^2n}{h}) + O(\frac{mn^3}{h})$ to a number of operations of the order $O(mn^2) + O(hn) + O(mn) + O(mh^2)$, where m is the sample size, h a fixed number given by the Grouped Jackknife Estimator ($h \ll m$) and n is the number of parameters to estimate ($m \geq n$).

Keywords: Linear Least Square, Jackknife estimator, Complexity of computation

MSC(2000): 93E24, 62F40, 03D15

Resumen

En este artículo, hacemos una generalización de los resultados obtenidos por Martínez y Sanabria para el cálculo del Estimador Jackknife para Mínimos Cuadrados Lineales (EJMCL), los cuales permiten expresar los estimadores de los subproblemas que resultan al calcular el Estimador Jackknife Agrupado para Mínimos Cuadrados Lineales (EJAMCL) en términos del estimador inicial y otras expresiones sencillas de calcular, y así modificar el algoritmo estándar para calcular el EJAMCL.

Con esta modificación, se reduce el número de operaciones del orden $O(\frac{m^2n^2}{h}) + O(\frac{m^2n}{h}) + O(\frac{mn^3}{h})$ a un número de operaciones del orden $O(mn^2) + O(hn) + O(mn) + O(mh^2)$, donde m es el tamaño de la muestra, h un número fijo dado por el Estimador Jackknife Agrupado ($h \ll m$) y n es el número de parámetros a estimar ($m \geq n$).

Palabras y frases claves: Mínimos Cuadrados Lineales, Estimador Jackknife, Complejidad computacional

1 Introducción

El algoritmo estándar para el cálculo del Estimador Jackknife para Mínimos Cuadrados Lineales (EJMCL) requiere un número de operaciones del orden $O(m^2n^2) + O(mn^3)$, donde m es el tamaño de la muestra y n es el número de parámetros a estimar, lo cual hace que calcular el EJMCL sea muy costoso, computacionalmente hablando. Sin embargo, Martínez y Sanabria, usando convenientemente propiedades básicas del álgebra lineal, lograron obtener un algoritmo mucho más eficiente, disminuyendo el número de operaciones al orden $O(mn) + O(mn^2)$ bajo la condición de que el problema de estimación inicial y los subproblemas involucrados sean de rango completo [2]; posteriormente, lograron mantener el resultado

anterior sin la necesidad de que los subproblemas involucrados fuesen de rango completo [3]; y por último, lograron conservar la eficiencia del cálculo sin requerir condición alguna sobre el problema inicial, es decir, sin importar que el problema de estimación inicial sea de rango deficiente [4].

En este artículo, presentamos una generalización de los resultados obtenidos por Martínez y Sanabria, los cuales permiten realizar una modificación al algoritmo estándar para calcular el Estimador Jackknife Agrupado para Mínimos Cuadrados Lineales (EJAMCL), reduciendo el número de operaciones a realizar, generalizando así el algoritmo para calcular el EJMCL obtenido por Martínez y Sanabria para el caso del estimador Jackknife agrupado.

2 Estimador Jackknife Agrupado para Mínimos Cuadrados Lineales (EJAMCL)

Definición 1. Sea X_1, \dots, X_m una muestra aleatoria de una población caracterizada por un parámetro θ y $T = t_m(X_1, \dots, X_m)$ un estimador de dicho parámetro, basado en la muestra de tamaño m . Al dividir la muestra en g grupos de tamaño h ($m = gh$), si denotemos por T_j al estimador T evaluado para los $(m - h)$ elementos que quedan después de quitar el j -ésimo grupo de elementos de la muestra ($j = 1, \dots, g$), el Estimador Jackknife Agrupado (EJA) [1] es

$$T_{JA} = \frac{1}{g} \sum_{j=1}^g (gT - (g-1)T_j) = gT - \frac{(g-1)}{g} \sum_{j=1}^g T_j.$$

De otro lado, dado el conjunto de observaciones (a_i^T, α_i) , donde $a_i \in \mathbb{R}^n$ y $\alpha_i \in \mathbb{R}$ para $i = 1, \dots, m$ con $m \geq n$, el problema de estimar x tal que $\alpha_i = a_i^T x$ por el método de los mínimos cuadrados lineales, se reduce a encontrar \hat{x} tal que

$$\|A\hat{x} - y\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - y\|_2,$$

donde $A = [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ y $y = (\alpha_1, \dots, \alpha_m)^T$. Al vector \hat{x} se le denomina *Estimador de Mínimos Cuadrados Lineales (EMCL)*. Ahora, dividiendo las m observaciones de la muestra en g grupos de tamaño h ($m = gh$) y aplicando el método de estimación Jackknife agrupado a \hat{x} , obtenemos el *Estimador Jackknife Agrupado para Mínimos Cuadrados Lineales (EJAMCL)*

$$\hat{x}_{JA} = g\hat{x} - (g-1) \sum_{j=1}^g \frac{\hat{x}_j}{g},$$

donde \hat{x}_j es la solución del subproblema

$$\|A_j \hat{x}_j - y_j\|_2 = \min_{x \in \mathbb{R}^n} \|A_j x - y_j\|_2, \quad j = 1, \dots, g,$$

donde A_j es la matriz resultante de extraer el grupo j -ésimo de filas de la matriz A y y_j es el vector resultante de extraer el grupo j -ésimo de componentes del vector y .

Para efectos de un mayor entendimiento, daremos a continuación un pequeño ejemplo del cálculo del EJAMCL.

Ejemplo 1. Dados los puntos $(-6, -9)$, $(-5, -6)$, $(-4, -5)$, $(-3, -3)$, $(-2, 1)$, $(-1, 0)$, $(0, 3)$, $(1, 5)$, $(2, 7)$, $(3, 9)$, $(4, 12)$, $(5, 13)$, deseamos estimar la mejor recta que los aproxime, en el sentido de los mínimos cuadrados; es decir, necesitamos estimar M y b tales que, si $z_i = Mt_i + b$, $\|z - y\|_2$ sea mínimo, siendo (t_i, y_i) los puntos dados.

En términos de la notación utilizada, tenemos que $m = 12$ y $n = 2$, y la matriz A , el vector y y el vector x estarían dados por

$$A = \begin{pmatrix} -6 & 1 \\ -5 & 1 \\ -4 & 1 \\ -3 & 1 \\ -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} -9 \\ -6 \\ -5 \\ -3 \\ 1 \\ 0 \\ 3 \\ 5 \\ 7 \\ 9 \\ 12 \\ 13 \end{pmatrix} \quad \text{y} \quad x = \begin{pmatrix} M \\ b \end{pmatrix}$$

Así, el estimador de mínimos cuadrados

$$\hat{x} = \left(\frac{3402}{1716}, \frac{927}{286} \right)^T \approx (1, 98, 3, 24)^T$$

resuelve el problema $\min_{x \in \mathbb{R}^2} \|Ax - y\|_2$.

De igual forma, tomando $g = 3$, $h = 4$, podemos calcular el EJAMCL, calculando los respectivos \hat{x}_j , para $j = 1, 2, 3$, que resuelven $\min_{x \in \mathbb{R}^2} \|A_j x - y_j\|_2$ para las siguientes matrices y vectores:

$$y_1 = (1, 0, 3, 5, 7, 9, 12, 13)^T$$

$$A_1 = \begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -6 & 1 \\ -5 & 1 \\ -4 & 1 \\ -3 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{pmatrix}, \quad y_2 = \begin{pmatrix} -9 \\ -6 \\ -5 \\ -3 \\ 7 \\ 9 \\ 12 \\ 13 \end{pmatrix}, \quad A_3 = \begin{pmatrix} -6 & 1 \\ -5 & 1 \\ -4 & 1 \\ -3 & 1 \\ -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad y_3 = \begin{pmatrix} -9 \\ -6 \\ -5 \\ -3 \\ 1 \\ 0 \\ 3 \\ 5 \end{pmatrix}$$

Los estimadores de las submuestras son:

$$\widehat{x}_1 = \begin{pmatrix} \frac{41}{21} \\ \frac{93}{28} \end{pmatrix} \approx \begin{pmatrix} 1,95 \\ 3,32 \end{pmatrix}, \quad \widehat{x}_2 = \begin{pmatrix} \frac{128}{69} \\ \frac{877}{276} \end{pmatrix} \approx \begin{pmatrix} 1,85 \\ 3,17 \end{pmatrix}, \quad \widehat{x}_3 = \begin{pmatrix} \frac{27}{14} \\ \frac{43}{14} \end{pmatrix} \approx \begin{pmatrix} 1,92 \\ 3,07 \end{pmatrix}$$

Ahora, utilizando la fórmula para \widehat{x}_{JA} , tenemos el EJAMCL así,

$$\widehat{x}_{JA} = 3 \begin{pmatrix} \frac{3402}{1716} \\ \frac{927}{286} \end{pmatrix} - \frac{3-1}{3} \left\{ \begin{pmatrix} \frac{41}{21} \\ \frac{93}{28} \end{pmatrix} + \begin{pmatrix} \frac{128}{69} \\ \frac{877}{276} \end{pmatrix} + \begin{pmatrix} \frac{27}{14} \\ \frac{43}{14} \end{pmatrix} \right\} = \begin{pmatrix} \frac{212}{100} \\ \frac{334}{100} \end{pmatrix} \approx \begin{pmatrix} 2,12 \\ 3,34 \end{pmatrix},$$

el cual es un estimador de x que posee mejores propiedades estadísticas que \widehat{x} , calculado inicialmente.

El algoritmo estándar para calcular el EJAMCL lo podemos expresar en los siguientes cuatro pasos:

1. Dados $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$.
2. Resolver $\min_{x \in \mathbb{R}^n} \|Ax - y\|_2^2$. \Rightarrow Salida: \widehat{x} .
3. Para $j = 1, \dots, g$. Resolver $\min_{x \in \mathbb{R}^n} \|A_j x - y_j\|_2^2$. \Rightarrow Salida: \widehat{x}_j .
4. Calcular $\widehat{x}_{JA} = g\widehat{x} - (g-1) \sum_{j=1}^g \frac{\widehat{x}_j}{g}$. \Rightarrow Salida: \widehat{x}_{JA} .

Si las matrices A y A_j , para $j = 1, \dots, g$, son de rango completo, los pasos 1 y 2 se reducen a encontrar las soluciones únicas de los sistemas de ecuaciones

$$A^T A x = A^T y \quad \text{y} \quad A_j^T A_j x_j = A_j^T y_j, \quad \text{para } j = 1, \dots, g.$$

En otras palabras, se reduce a calcular \widehat{x} y los \widehat{x}_j , tal que

$$\widehat{x} = (A^T A)^{-1} A^T y \quad \text{y} \quad \widehat{x}_j = (A_j^T A_j)^{-1} A_j^T y_j, \quad \text{para } j = 1, \dots, g.$$

Resolviendo los problemas planteados en los pasos 1 y 2 por el método de las ecuaciones normales, un algoritmo más detallado para el cálculo de EJAMCL es el siguiente.

3 Algoritmo estándar detallado para calcular el EJAMCL

1. Dados $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$.
2. $\{ \text{Resolver } A^T A x = A^T y. \}$

- Calcular $C = A^T A$.
 - Calcular $d = A^T y$.
 - Resolver $Cx = d$. \Rightarrow Salida: \hat{x} .
3. Para $j = 1, \dots, g$. $\{ \text{Resolver } A_j^T A_j x = A_j^T y_j. \}$
- Calcular $C_j = A_j^T A_j$.
 - Calcular $d_j = A_j^T y_j$.
 - Resolver $C_j x_j = d_j$. \Rightarrow Salida: \hat{x}_j .
4. Calcular $\hat{x}_{JA} = g\hat{x} - (g-1) \sum_{j=1}^g \frac{\hat{x}_j}{g}$. \Rightarrow Salida: \hat{x}_{JA} .

En el algoritmo detallado anteriormente, nótese que la cantidad de operaciones necesarias para resolver los sistemas de ecuaciones lineales¹ en el paso 1 es aproximadamente $[mn^2 + mn + \frac{n^3}{3}]$ y, en el paso 2, es aproximadamente $[\frac{m}{h}((m-h)n^2 + (m-h)n + \frac{n^3}{6})]$, donde m es el tamaño de la muestra, h un número fijo dado por el Estimador Jackknife Agrupado ($h \ll m$) y n es el número de parámetros a estimar ($m \geq n$); por tanto, lo más costoso del algoritmo es el paso 2, donde se deben calcular los respectivos \hat{x}_j .

En adelante, usaremos la siguiente notación. Dada $A = [a_1, \dots, a_m]^T$ con $a_i \in \mathbb{R}^n$, $i = 1, \dots, m$, la matriz $A_j \in \mathbb{R}^{(m-h) \times n}$ es la matriz que resulta de quitar h filas a la matriz A , que, sin pérdida de generalidad, suponemos que son seguidas. Así

$$A_j = [a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m]^T.$$

Denotaremos por $B_j \in \mathbb{R}^{(h \times n)}$ la matriz formada por las h filas que se le quitaron a la matriz A . Así

$$B_j = [a_k, \dots, a_{k+h-1}]^T.$$

De igual forma, dado $y = (\alpha_1, \dots, \alpha_m)^T$ con $\alpha_i \in \mathbb{R}$, $i = 1, \dots, m$, el vector $y_j \in \mathbb{R}^{m-h}$ es el vector que resulta de quitar las h componentes correspondientes al vector y . Así

$$y_j = (\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+h}, \dots, \alpha_m)^T.$$

Denotaremos por $b_j \in \mathbb{R}^h$ el vector formado por las h componentes que se le quitaron al vector y . Así

$$b_j = (\alpha_k, \dots, \alpha_{k+h-1})^T.$$

¹Se asume que el sistema $Cx = d$ se resuelve con un algoritmo como Cholesky. Otros algoritmos como QR, para resolver este sistema, no requieren el cálculo de la matriz C pero resultan ser más costosos, aunque proporcionan una mayor estabilidad numérica.

4 Algunos resultados del álgebra lineal

A continuación, veremos una serie de resultados que, aunque no sean requeridos de manera directa, son importantes para el diseño y demostración de los resultados que respaldan los algoritmos.

En primer lugar, con el objetivo de rebajar el costo del cálculo de $A_j^T A_j$, obtuvimos una generalización del Lema de Vargas [2].

Lema 1. Dada una matriz $A = [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$, si la matriz $A_j = [a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m]^T$, es la matriz que resulta de quitar h filas a la matriz A y la matriz B_j es la matriz que se forma con las h filas que se quitaron a A ; entonces

$$A_j^T A_j = A^T A - B_j^T B_j.$$

Demostración.

$$\begin{aligned} A_j^T A_j &= [a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m][a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m]^T \\ &= \sum_{i=1}^{k-1} a_i a_i^T + \sum_{i=k+h}^m a_i a_i^T \\ &= \sum_{i=1}^{k-1} a_i a_i^T + \sum_{i=k}^{k+h-1} a_i a_i^T - \sum_{i=k}^{k+h-1} a_i a_i^T + \sum_{i=k+h}^m a_i a_i^T \\ &= \sum_{i=1}^m a_i a_i^T - \sum_{i=k}^{k+h-1} a_i a_i^T \\ &= [a_1, \dots, a_m][a_1, \dots, a_m]^T - [a_k, \dots, a_{k+h-1}][a_k, \dots, a_{k+h-1}]^T \\ &= A^T A - B_j^T B_j. \end{aligned}$$

□

De igual manera, generalizamos el Lema 2 dado en [2], que nos permite simplificar el cálculo de $A_j^T y_j$.

Lema 2. Dada una matriz $A = [a_1, \dots, a_m]^T$ y un vector $y = (\alpha_1, \dots, \alpha_m)^T$, si la matriz $A_j = [a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m]^T$ y el vector

$$y_j = (\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+h}, \dots, \alpha_m)^T,$$

son, respectivamente, la matriz y el vector que resultan de quitar h filas a la matriz A y las h componentes correspondientes del vector y , y la matriz B_j es la matriz que se forma con las h filas que se quitaron a A y el vector b_j es el vector que se forma con las h componentes que se quitaron a y ; entonces

$$A_j^T y_j = A^T y - B_j^T b_j.$$

Demostración.

$$\begin{aligned}
A_j^T y_j &= [a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m](\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+h}, \dots, \alpha_m)^T \\
&= \alpha_1 a_1 + \dots + \alpha_{k-1} a_{k-1} + \alpha_{k+h} a_{k+h} + \dots + \alpha_m a_m \\
&= \alpha_1 a_1 + \dots + \alpha_m a_m - \alpha_k a_k - \dots - \alpha_{k+h-1} a_{k+h-1} \\
&= \sum_{i=1}^m \alpha_i a_i - [a_k, \dots, a_{k+h-1}](\alpha_k, \dots, \alpha_{k+h-1})^T \\
&= A^T y - B_j^T b_j.
\end{aligned}$$

□

Por último, un resultado clave para el logro de nuestro objetivo, conocido como la fórmula general de Sherman-Morrison-Woodbury.

Lema 3. (*Sherman-Morrison-Woodbury*) Dadas las matrices $W \in \mathbb{R}^{n \times n}$ no singular, $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{m \times n}$ y la idéntica $I_k \in \mathbb{R}^{k \times k}$. $(I_m + VW^{-1}U)$ es no singular, si y sólo si, $(W + UV)$ es no singular. Además, si $(I_m + VW^{-1}U)$ es no singular,

$$(W + UV)^{-1} = W^{-1} - W^{-1}U(I_m + VW^{-1}U)^{-1}VW^{-1}$$

Demostración.

\Rightarrow) Si $(I_m + VW^{-1}U)$ es no singular, tenemos que

$$\begin{aligned}
(W + UV)(W^{-1} - W^{-1}U(I_m + VW^{-1}U)^{-1}VW^{-1}) \\
&= I_n + UVW^{-1} - (U(I_m + VW^{-1}U)^{-1}VW^{-1} + \\
&\quad UVW^{-1}U(I_m + VW^{-1}U)^{-1}VW^{-1}) \\
&= I_n + UVW^{-1} - (U + UVW^{-1}U)(I_m + VW^{-1}U)^{-1}VW^{-1} \\
&= I_n + UVW^{-1} - U(I_m + VW^{-1}U)(I_m + VW^{-1}U)^{-1}VW^{-1} \\
&= I_n + UVW^{-1} - UVW^{-1} = I_n.
\end{aligned}$$

\Leftarrow) Sea $Z = W^{-1}U$. Si $(I_m + VW^{-1}U) = (I_m + VZ)$ es singular, entonces existe $x \in \mathbb{R}^m$ diferente de cero, tal que $(I + VZ)x = 0$. Demostremos que $y = Zx \in \mathbb{R}^n$ es diferente de cero y $(I_n + ZV)y = 0$.

De las hipótesis sobre x , tenemos que $x = -VZx = -Vy$, por lo tanto, si $y = 0$, x sería 0; por lo cual, concluimos que $y \neq 0$. Además,

$$(I_n + ZV)y = (I_n + ZV)Zx = (Z + ZVZ)x = Z(I_m + VZ)x = Z0 = 0,$$

por lo tanto, $(I_n + ZV)$ es singular, y como $(W + UV) = W(I_n + W^{-1}UV) = W(I_n + ZV)$, concluimos que $(W + UV)$ es singular. □

Ahora, haciendo uso de los lemas anteriores y bajo el supuesto que el problema inicial y los subproblemas respectivos son de rango completo, podemos expresar las soluciones de los sistemas $A_j^T A_j x_j = A_j^T y_j$ del segundo paso del algoritmo, en términos de la solución de $A^T A x = A^T y$ y otras expresiones sencillas de calcular, como se muestra en el siguiente teorema.

Teorema 1. Dada una matriz $A = [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ y un vector $y = (\alpha_1, \dots, \alpha_m)^T$, si la matriz $A_j = [a_1, \dots, a_{k-1}, a_{k+h}, \dots, a_m]^T$ y el vector $y_j = (\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+h}, \dots, \alpha_m)$ son, respectivamente, la matriz y el vector que resultan de quitar h filas a la matriz A y las h componentes correspondientes del vector y ; y además, A y A_j son matrices de rango completo, entonces \hat{x}_j , la solución de $A_j^T A_j x_j = A_j^T y_j$, está dada por

$$\hat{x}_j = \hat{x} + Z_j(w_j - b_j),$$

donde \hat{x} es la solución de $A^T A x = A^T y$, Z_j es la solución de $A^T A Z_j = B_j^T$, w_j es la solución de $(I - B_j Z_j)w = Z_j^T d_j$ con $d_j = A_j^T y_j$, y además, B_j y b_j son la matriz y vector formado con las h filas quitadas a A y las h componentes quitadas a y , respectivamente.

Demostración. Sean $C = A^T A$, $C_j = A_j^T A_j$ y $d = A^T y$. Como A y A_j son de rango completo, entonces C y C_j son invertibles. Por el Lema 1, $C_j = C - B_j^T B_j$, por el Lema 2, $d_j = d - B_j^T b_j$ y por el Lema 3, tenemos que

$$C_j^{-1} = C^{-1} + C^{-1} B_j^T (I - B_j C^{-1} B_j^T)^{-1} B_j C^{-1}.$$

Ahora

$$\begin{aligned} \hat{x}_j &= C_j^{-1} d_j \\ &= [C^{-1} + C^{-1} B_j^T (I - B_j C^{-1} B_j^T)^{-1} B_j C^{-1}] d_j \\ &= C^{-1} d_j + C^{-1} B_j^T (I - B_j C^{-1} B_j^T)^{-1} B_j C^{-1} d_j \\ &= C^{-1} (d - B_j^T b_j) + C^{-1} B_j^T (I - B_j C^{-1} B_j^T)^{-1} B_j C^{-1} d_j \\ &= C^{-1} d - C^{-1} B_j^T b_j + C^{-1} B_j^T (I - B_j C^{-1} B_j^T)^{-1} B_j C^{-1} d_j. \end{aligned}$$

Sea Z_j la solución de $C Z_j = B_j^T$, entonces

$$\hat{x}_j = \hat{x} - Z_j b_j + Z_j (I - B_j Z_j)^{-1} Z_j^T d_j.$$

Sea w_j la solución de $(I - B_j Z_j)w = Z_j^T d_j$, entonces

$$\begin{aligned} \hat{x}_j &= \hat{x} - Z_j b_j + Z_j w_j \\ &= \hat{x} + Z_j (w_j - b_j). \end{aligned}$$

□

Con base en el Teorema 1, podemos decir que para la solución de los $(m - h)$ sistemas de los subproblemas no es necesario calcular $A_j^T A_j$ ni $A_j^T y_j$, y que la solución del sistema $A_j^T A_j x = A_j^T y_j$ se reduce a la solución de sistemas con la matriz $A^T A$ ² y al cálculo de algunos productos internos, haciendo que el costo del algoritmo estándar del EJMCL sea mucho menor. Por tanto, proponemos modificar el paso 2 del algoritmo estándar de la siguiente manera.

Para $j = 1, \dots, g$. {Resolver $A_j^T A_j x = A_j^T y_j$.}

- Resolver $CZ_j = B_j^T$.
- Calcular $S_j = B_j Z_j$.
- Calcular $r_j = Z_j^T d_j$.
- Resolver $(I - S_j)w_j = r_j$.
- Calcular $\hat{x}_j = \hat{x} - Z_j(w_j - b_j)$.

\Rightarrow Salida: \hat{x}_j .

Haciendo esta modificación en el paso 2 del algoritmo estándar del EJMCL, bajo el supuesto que A y A_j son de rango completo, reducimos el número de operaciones del algoritmo de $[\frac{m}{h}((m-h)n^2 + (m-h)n + \frac{n^3}{6})]$ a $[n^2 + hn + 2n + \frac{h^2}{3}]$, donde m es el tamaño de la muestra, h un número fijo dado por el estimador Jackknife Agrupado ($h \ll m$) y n es el número de parámetros a estimar ($m \geq n$).

Claramente, el hecho que una matriz A sea de rango completo no implica que las matrices A_j también lo sean. Por ello, nos propusimos encontrar una caracterización de las soluciones de $A_j^T A_j x = A_j^T y_j$ basada en la solución de $A^T A x = A^T y$, independientemente si las respectivas A_j son o no de rango completo. Para ello, necesitamos probar que el sistema $(I - B_j Z_j)p = B_j \hat{x} - b_j$ tiene solución, aún cuando $(I - B_j Z_j)$ sea singular, como lo demostramos en el siguiente lema.

Lema 4. Dada una matriz $A = [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ de rango completo, un vector $y = (\alpha_1, \dots, \alpha_m)^T$, \hat{x} , la solución de $A^T A x = A^T y$, y Z_j la solución de $A^T A Z = B_j^T$, entonces

$$B_j \hat{x} - b_j \quad \text{es solución de} \quad (I - B_j Z_j)p = B_j \hat{x} - b_j,$$

donde B_j y b_j son la matriz y el vector formado con las h filas quitadas a A y las h componentes quitadas a y , respectivamente.

Demostración. Sean $C = A^T A$, $C_j = A_j^T A_j$, $d = A^T y$ y $d_j = A_j^T y_j$, entonces

$$\begin{aligned} B_j \hat{x} - b_j &= B_j C^{-1} d - b_j \\ &= B_j C^{-1} (d_j + B_j^T b_j) - b_j \\ &= B_j C^{-1} d_j + B_j C^{-1} B_j^T b_j - b_j \\ &= Z_j^T d_j + (Z_j^T B_j^T - I) b_j \end{aligned}$$

²Recordemos que resolver un segundo sistema con la misma matriz, resulta menos costoso puesto que ya se tiene la factorización de la matriz calculada al resolver el primer sistema.

$$\begin{aligned}
B_j \hat{x} - b_j &= Z_j^T C_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= Z_j^T (C - B_j^T B_j) \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= Z_j^T C \hat{x}_j - Z_j^T B_j^T B_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= B_j \hat{x}_j - Z_j^T B_j^T B_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= (I - Z_j^T B_j^T) B_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= (I - Z_j^T B_j^T) (B_j \hat{x}_j - b_j) \\
&= (I - B_j Z_j) (B_j \hat{x}_j - b_j),
\end{aligned}$$

puesto que $Z_j^T B_j^T = Z_j^T C Z_j$ es una matriz simétrica. \square

Veamos ahora, que para el caso en que la matriz $(I - B_j Z_j)$ es singular, gracias al lema anterior, podemos determinar un conjunto solución de $A_j^T A_j x = A_j^T y_j$.

Teorema 2. Dada una matriz $A = [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ de rango completo, un vector $y = (\alpha_1, \dots, \alpha_m)^T$ y \hat{x} , la solución de $A^T A x = A^T y$, entonces un conjunto solución de $A_j^T A_j x = A_j^T y_j$ es

$$\hat{x}_j = \hat{x} + Z_j u_j,$$

para todo $u_j \in \mathbb{R}^h$, tal que u_j sea solución de $(I - B_j Z_j)u = B_j \hat{x} - b_j$, donde Z_j es la solución de $A^T A Z = B_j^T$, y además, B_j y b_j son la matriz y el vector formado con las h filas quitadas a A y las h componentes correspondientes quitadas a y , respectivamente.

Demostración. Sean $C = A^T A$, $C_j = A_j^T A_j$, $d = A^T y$ y $d_j = A_j^T y_j$, entonces

$$\begin{aligned}
C_j(\hat{x} + Z_j u_j) &= (C - B_j^T B_j)(\hat{x} + Z_j u_j) \\
&= C \hat{x} + C Z_j u_j - B_j^T B_j \hat{x} - B_j^T B_j Z_j u_j \\
&= d + B_j^T u_j - B_j^T B_j \hat{x} - B_j^T B_j Z_j u_j \\
&= d - B_j^T B_j \hat{x} + B_j^T (I - B_j Z_j) u_j \\
&= d - B_j^T B_j \hat{x} + B_j^T (B_j \hat{x} - b_j) \\
&= d - B_j^T b_j \\
&= d_j.
\end{aligned}$$

\square

Por el teorema anterior, garantizamos que todo elemento de la forma $\hat{x} + Z_j u_j$, donde $(I - B_j Z_j)u_j = B_j \hat{x} - b_j$, es solución de $A_j^T A_j x = A_j^T y_j$. Ahora, para completar una caracterización del conjunto solución de $A_j^T A_j x_j = A_j^T y_j$, necesitamos ver que toda solución \hat{x}_j es de la forma $\hat{x} + Z_j u_j$, lo cual establecemos en el siguiente resultado.

Teorema 3. Dada una matriz $A = [a_1, \dots, a_m]^T \in \mathbb{R}^{m \times n}$ de rango completo, un vector $y = (\alpha_1, \dots, \alpha_m)^T$ y \hat{x} , la solución de $A^T A x = A^T y$. Si \hat{x}_j es una solución de $A_j^T A_j x_j = A_j^T y_j$, entonces

$$\hat{x}_j = \hat{x} + Z_j u_j,$$

para algún $u_j \in \mathbb{R}^h$, solución del sistema $(I - B_j Z_j)u = B_j \hat{x} - b_j$, donde Z_j es la solución de $A^T A Z = B_j^T$, y además, B_j y b_j son la matriz y el vector formado con las h filas quitadas a A y las h componentes correspondientes quitadas a y , respectivamente.

Demostración. Sean $C = A^T A$, $C_j = A_j^T A_j$, $d = A^T y$ y $d_j = A_j^T y_j$, entonces

$$\begin{aligned} C_j \hat{x}_j &= d_j \\ (C - B_j^T B_j) \hat{x}_j &= d - B_j^T b_j \\ C \hat{x}_j - B_j^T B_j \hat{x}_j &= d - B_j^T b_j \\ C \hat{x}_j &= d - B_j^T b_j + B_j^T B_j \hat{x}_j. \end{aligned}$$

Sea $v_j = B_j \hat{x}_j$, entonces

$$\begin{aligned} C \hat{x}_j &= d - B_j^T b_j + B_j^T v_j \\ &= d + B_j^T (v_j - b_j). \end{aligned}$$

Sea $u_j = v_j - b_j$, entonces

$$\begin{aligned} C \hat{x}_j &= d + B_j^T u_j \\ \hat{x}_j &= C^{-1} (d + B_j^T u_j) \\ &= C^{-1} d + C^{-1} B_j^T u_j \\ &= \hat{x} + Z_j u_j. \end{aligned}$$

Además,

$$(I - B_j Z_j) u_j = (I - B_j Z_j) (v_j - b_j) = (I - B_j Z_j) (B_j \hat{x}_j - b_j) = B_j \hat{x} - b_j.$$

□

Así, hemos logrado caracterizar el conjunto solución de $A_j^T A_j x_j = A_j^T y_j$, aún para cuando A_j no es de rango completo.

De los resultados anteriores, se puede notar que u_j es de la forma $w_j - b_j$; donde w_j es solución de $(I - B_j Z_j)w = Z_j^T d_j$. Veamos que $u_j = w_j - b_j$ es solución

de $(I - B_j Z_j)u = B_j \hat{x} - b_j$.

$$\begin{aligned}
 (I - B_j Z_j)u_j &= (I - B_j Z_j)(w_j - b_j) &= (I - B_j Z_j)w_j - (I - B_j Z_j)b_j \\
 &= Z_j^T d_j + B_j Z_j b_j - b_j \\
 &= Z_j^T (d - B_j^T b_j) + B_j Z_j b_j - b_j \\
 &= Z_j^T d - Z_j^T B_j^T b_j + B_j Z_j b_j - b_j \\
 &= B_j C^{-1} d - b_j \\
 &= B_j \hat{x} - b_j
 \end{aligned}$$

Ahora, en el caso en que A_j es de rango completo, la matriz $(I - B_j Z_j)$ es no singular y así el vector u_j es único e igual a $(w_j - b_j)$, donde w_j es la solución de $(I - B_j Z_j)w = Z_j^T d_j$, como se demostró en el Teorema 1.

Dados los resultados anteriores, ahora podemos realizar una pequeña, pero significativa modificación al algoritmo dado en la sección anterior, manteniendo su eficiencia y sin la condición de que las matrices A_j sean de rango completo; es decir, con la única condición de que sólo A sea de rango completo. Como existe la posibilidad que alguno de los subproblemas (A_j) sean de rango deficiente ($I - B_j Z_j$ singular), el subproblema j tendría infinitas soluciones. En tal caso, proponemos tomar uno de los u_j que sean solución de $(I - B_j Z_j)u = B_j \hat{x} - b_j$ y tomar a $\hat{x}_j = \hat{x} + Z_j u_j$ como solución del subproblema.

Finalmente, generalizaremos nuestro resultado, quitando la condición sobre el problema inicial; es decir, sin importar si A es o no de rango completo. Para ello, veremos un lema que nos permitirá prescindir de esta condición.

Lema 5. Dada una matriz $A = [a_1, \dots, a_m]^T$, un vector $y = (\alpha_1, \dots, \alpha_m)^T$ y \hat{x} , una solución de $A^T A x = A^T y$, entonces

$$B_j \hat{x}_j - b_j \quad \text{es solución de} \quad (I - B_j Z_j)p = B_j \hat{x} - b_j,$$

donde Z_j es una solución de $A^T A Z = B_j^T$, y además, B_j y b_j son la matriz y el vector formado con las h filas quitadas a A y las h componentes correspondientes quitadas a y , respectivamente..

Demostración. Sean $C = A^T A$, $C_j = A_j^T A_j$, $d = A^T y$ y $d_j = A_j^T y_j$. Como $C \hat{x} = d = d_j + B_j b_j$, entonces

$$Z_j^T C \hat{x} = Z_j^T d_j + Z_j^T B_j b_j.$$

Ahora, como $C Z_j = B_j^T$, entonces $B_j \hat{x} = Z_j^T C \hat{x} = Z_j^T d_j + Z_j^T B_j b_j$. Así, tenemos

que

$$\begin{aligned}
B_j \hat{x} - b_j &= Z_j^T d_j + (Z_j^T B_j^T - I) b_j \\
&= Z_j^T C_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= Z_j^T (C - B_j^T B_j) \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= Z_j^T C \hat{x}_j - Z_j^T B_j^T B_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= B_j \hat{x}_j - Z_j^T B_j^T B_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= (I - Z_j^T B_j^T) B_j \hat{x}_j + (Z_j^T B_j^T - I) b_j \\
&= (I - Z_j^T B_j^T) (B_j \hat{x}_j - b_j) \\
&= (I - B_j Z_j) (B_j \hat{x}_j - b_j).
\end{aligned}$$

□

Nótese que el sistema $A^T A Z = B_j^T$ siempre tiene solución (Z_j), puesto que el sistema $A^T A Z = B_j^T$ puede verse como h sistemas de la forma $A^T A z_i = a_{k+i-1}$, donde z_i y a_{k+i-1} con $i = 1, \dots, h$, son los vectores columna de las matrices Z y B_j^T , respectivamente. Ahora, cada sistema $A^T A z_i = a_{k+i-1}$ siempre tiene solución, puesto que, solucionar este sistema es equivalente a solucionar el sistema $A^T A z_i = A^T e_{k+i-1}$ que es el sistema de ecuaciones normales correspondiente al problema de mínimos cuadrados lineales $\min \|Az - e_{k+i-1}\|$, el cual siempre tiene solución (e_r es el r -ésimo vector canónico de \mathbb{R}^h). Así, con el Lema 5, garantizamos el mismo resultado del Lema 4; con la diferencia que ahora no necesitamos que la matriz A sea de rango completo, logrando garantizar que el Teorema 2 siga siendo válido aún cuando la matriz A no sea de rango completo. De esta manera, obtenemos un conjunto solución de $A_j^T A_j x = A_j^T y_j$ basado en una solución de $A^T A x = A^T y$, sin condición alguna sobre el problema inicial o los subproblemas requeridos; es decir, que la matriz A y las matrices A_j no necesariamente sean de rango completo.

Cabe anotar que el Teorema 3 no es válido sin la hipótesis de rango completo para la matriz A , por lo tanto, sin esta hipótesis no es posible caracterizar el conjunto solución de $A_j^T A_j x = A_j^T y_j$.

Dados los anteriores resultados, tenemos el soporte teórico para garantizar que el algoritmo propuesto es válido aún para cuando A y A_j sean de rango deficiente. Así, si el problema inicial (A) es de rango deficiente, entonces el problema inicial ($A^T A x = A^T y$) tendría infinitas soluciones. En tal caso, tomamos una de sus soluciones (un \hat{x}), una de las soluciones (Z_j) de $A^T A Z = B_j^T$ y un vector u_j que sea solución de $(I - B_j Z_j)u = B_j \hat{x} - b_j$ y la solución de $A_j^T A_j x_j = A_j^T y_j$ sigue siendo $\hat{x}_j = \hat{x} + Z_j u_j$, como se propuso anteriormente. Cabe resaltar que esta última modificación no altera la eficiencia lograda, puesto que la eficiencia se da al reducir el costo de solución de los subproblemas con base en la solución del problema inicial, aún así los subproblemas y el problema inicial sean de rango deficiente.

Créditos

Este artículo hace parte del Trabajo de Grado dirigido por H.J. Martínez y A.M. Sanabria y presentado por A. Arévalo como requisito parcial para optar al título de Matemático en la Universidad del Valle. Una versión inicial de este artículo fue presentado en el XVIII Congreso Colombiano de Matemáticas celebrado en Bucaramanga (Santander), en julio de 2011.

Referencias

- [1] Behar, R. y Yepes, M. Sobre algunas técnicas de remuestreo: El método Jackknife. *Heurística* 5, No 6, 1991.
- [2] Martínez, H. J. y Sanabria, A. M. Cálculo eficiente del estimador jackknife para mínimos cuadrados lineales bajo condiciones de unicidad. *Matemáticas: Enseñanza Universitaria*, Vol III, No 1 y 2, 2000.
- [3] Martínez, H. J. y Sanabria, A. M. Cálculo eficiente del estimador jackknife para mínimos cuadrados lineales de rango completo. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, Vol XXX, 2006.
- [4] Martínez, H. J. y Sanabria, A. M. Cálculo eficiente del estimador jackknife para mínimos cuadrados lineales de rango deficiente. Aceptado para publicación en la *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 2012.

Dirección de los autores

Alexander Arévalo S. — Departamento de Matemáticas, Universidad del Valle, Cali-Colombia

e-mail: alexanderrvl@hotmail.com

Héctor J. Martínez R. — Departamento de Matemáticas, Universidad del Valle, Cali-Colombia

e-mail: hector.martínez@correounivalle.edu.co

Ana M. Sanabria R. — Departamento de Matemáticas, Universidad del Valle, Cali-Colombia

e-mail: ana.sanabria@correounivalle.edu.co