



## MODELO DE REGRESIÓN LINEAL: MÍNIMOS CUADRADOS PARCIALES (PLS) CON MÉTODO CUASI-NEWTON

(Linear regression model: Partial Least Squares (PLS) with Quasi-Newton method)

Recibido: 19/11/2014 Aceptado: 30/07/2015

Rodríguez, Eddy

Universidad del Zulia, Venezuela

[eddyjackeline@yahoo.es](mailto:eddyjackeline@yahoo.es)

### RESUMEN

El modelado de problemas de regresión lineal es de gran utilidad en áreas como la medicina, ingeniería y economía, entre otros. En particular, el método de mínimos cuadrados parciales (PLS) tiene sus inicios en los años sesenta con aplicaciones en la quimiométrica (Geladi y Kowalski, 1986a). Actualmente, son varios los estudios realizados sobre el desarrollo del algoritmo de este método y sobre su uso en el ajuste de problemas de regresión. Este trabajo tuvo por objetivo integrar al método PLS el método de Newton modificado para encontrar los vectores proyección durante el proceso, aquí se trabajó de manera específica con la propuesta de Davidon-Fletcher-Powell (Fletcher y Powell, 1963), por su propiedad de convergencia. La metodología empleada se basó en una investigación aplicada, partiendo de una revisión bibliográfica de libros y artículos de autores especialistas en el tema de estudio como: Rodríguez (2010), Hoskuldsson (1998) y Luenberger (1989), seguido del planteamiento, análisis y formulación del algoritmo propuesto. Se concluyó que es posible aplicar el método de Newton modificado para encontrar en el PLS los vectores proyección entre las matrices de entrada y respuesta y las variables latentes. Los datos experimentales verifican que con este método de optimización el PLS modela problemas de regresión lineal de forma satisfactoria y puede trabajar con más de un vector de salida al mismo tiempo.

**Palabras claves:** Mínimos Cuadrados Parciales (PLS), método de Newton modificado, propuesta de Davidon-Fletcher-Powell, algoritmo, vectores proyección.

### MODEL OF LINEAR REGRESSION: SQUARE PARTIAL MINIMUMS (PLS) WITH METHOD CUASI-NEWTON

#### ABSTRACT

Modeling linear regression problems is very useful in areas such as medicine, engineering and economics, among others. In particular, the method of Partial Least Squares Regression (PLS) has its beginnings in the sixties with applications in chemometrics (Geladi and Kowalski, 1986a). Currently, several studies on the development of the algorithm of this method and its use in adjusting regression problems. This study aimed to integrate the PLS method Newton's method to find the vectors modified during projection; here we worked specifically with Davidon-Fletcher-Powell (Fletcher and Powell, 1963) proposal for its convergence property. The methodology is based on applied research, based on a literature review of books and articles from author's specialists in the field of



study as: Rodriguez (2010), Hoskuldsson (1998) and Luenberger (1989) followed the approach, analysis and formulation of the proposed algorithm. It was concluded that it is possible to apply the modified Newton method to find the PLS the projection vectors between the input and response matrices and latent variables. Experimental data verifies that this optimization method PLS models linear regression problems satisfactorily and can work with more than one vector output simultaneously.

**Keywords:** Partial Least Squares Regression (PLS), Newton's method modified, Davidon-Fletcher-Powell's offer, algorithm, vectors projection.

## INTRODUCCIÓN

En el presente trabajo se aborda el estudio sobre el modelado de problemas de regresión lineal, de manera específica el llamado método de Mínimos Cuadrados Parciales (PLS, por sus siglas en inglés), que tiene por característica principal que puede ser aplicado para modelar cualquier tipo de problema aproximadamente lineales, aun cuando las variables que estén involucradas sean linealmente independiente o no.

El estudio del PLS fue desarrollado inicialmente para aplicaciones quimiométricas (Geladi y Kowalski, 1986a, 1986b; Hoskuldsson, 1998), sin embargo, se extendió a otras ramas de la ciencia por ser una técnica útil en el modelado de problemas de regresión y clasificación, en especial en situaciones en las que la data experimental se caracteriza por tener una alta dimensionalidad, elevada multi-colinealidad y tener pocas observaciones. Este método permite establecer estrategias de regularización, evitando sobre ajustes y solventando el mal condicionamiento de problemas con alta colinealidad.

El método de Newton se caracteriza por aproximar a la función de estudio a una función cuadrática, a través de una aproximación por serie de Taylor truncada de orden dos cerca del punto de estudio, el proceso iterativo en este método involucra la inversa de la matriz Hessiana, que no siempre es posible. Por ello, se estudian modificaciones de este método que presenta formas alternas de aproximar esta matriz Hessiana o su inversa directamente, conocida como métodos Cuasi-Newton.

En este trabajo se desarrolla el estudio optimal de dos funciones cuadráticas basadas en la teoría del PLS, aplicando el método Davidon-Fletcher-Powell (Fletcher y Powell, 1963), que es un método Cuasi-Newton de optimización sin restricción para encontrar los vectores proyección entre los vectores de las matrices de entrada y respuesta, así como también entre los vectores de la matriz respuesta y sus proyecciones sobre las variables latentes.

Primero, se explica brevemente el método de Mínimos Cuadrados Parciales siguiendo a Rey (2009) y el método de Newton modificado con la propuesta de Davidon-Fletcher-Powell (Fletcher y Powell, 1963), y luego se presenta el aporte de este trabajo que es el desarrollo del PLS con adaptación del método de Newton modificado para encontrar los vectores de peso.



## MÉTODO DE CUADRADOS MÍNIMOS PARCIALES (PLS)

Este método puede ser usado para resolver problemas de regresión lineal donde la matriz  $X$  tenga más número de columnas que filas, lo que nos lleva a infinitas soluciones, como lo explican Mardia et al. (1979). Se basa en:

$$\min(e) = \min\|Y - XB\|^2 \quad (1)$$

El método PLS es un esquema aplicable a conjuntos multivariados y consiste en obtener de forma iterativa direcciones ortogonales en los espacios de variables, con el objeto de formar bases óptimas donde proyectarlas, seguido de una regresión de mínimos cuadrados sobre estas proyecciones o sobre un subconjunto de las mismas (Rey, 2009).

El método de Cuadrados Mínimos Parciales (PLS) relaciona la matriz de las variables de entrada  $x \in X \subset R^m$  y la matriz de las variables de salida  $y \in Y \subset R^p$  a través de las variables latentes de cada una, donde las variables latentes se pueden definir como un conjunto de variables artificiales que son ortogonales y que permiten hacer una reducción del espacio, como se ha demostrado en las investigaciones de Geladi y Kowalski (1986a) (1986b) y Hoskuldsson (1998):

$$\begin{aligned} X &= TP^T + E \\ Y &= UC^T + F \end{aligned} \quad Y = XB \quad (2) \quad \longrightarrow$$

Donde:  $T$  y  $U$  son las matrices compuestas por los vectores latentes.

$P$  y  $C$  son las matrices compuestas por los vectores de peso.

$E$  y  $F$  son las matrices residuales.

### FORMULACIÓN MATEMÁTICA DEL PLS

El método PLS tiene como objetivo la maximización de la covarianza muestral entre las variables latentes de cada matriz de datos, y la minimización del cuadrado del error como en la ecuación (1).

Paso 1: Encuéntrese los vectores “ $c$ ” y “ $w$ ”, mediante la minimización de la distancia entre los vectores de la matriz  $X$  y sus proyecciones sobre  $Y$ :

$$\min\|X - YW^T\|$$

Y la minimización de la distancia entre los vectores de la matriz  $Y$  con sus proyecciones sobre  $t$ :



$$\min \|Y - tc^T\|$$

Paso 2: se procede a encontrar el primer vector de las matrices T y U, a partir de:

$$t = Xw$$

$$u = Yc$$

En este paso es conveniente normalizar los vectores  $t$  y  $u$ , debido a posteriores cálculos con ellos.

Paso 3: se calculan las matrices residuales de X e Y, y si la matriz Y no es lo suficientemente pequeña, se procede a encontrar los segundos vectores latentes, siguiendo el mismo procedimiento anterior, así sucesivamente hasta que la matriz Y sea insignificante.

Para garantizar la ortogonalidad de la secuencia de variables latentes, a las matrices de datos originales se les va substrayendo en forma iterativa la parte de la información explicada por la variable latente obtenida en la presente iteración. Esto se logra restándole a las matrices residuales la matrices óptimas de rango 1, basadas en la variable latente  $t$  obtenida en la presente iteración. Estas matrices se obtienen resolviendo:

$$\min_p \|X - tp^T\| \text{ y } \min_q \|Y - tq^T\|$$

Donde:

$$p = X^T t \text{ y } q = Y^T t$$

De manera que las matrices se puede representar como:

$$X = X - tt^T X \quad \text{y} \quad Y = Y - tt^T Y$$

Llámesese a las matrices residuales de X e Y como E y F respectivamente, y calcúlense de la siguiente manera:

$$E = X - tp^T$$

- Donde  $tp^T = X$  si  $E \approx 0$ ,
- De esta manera  $p^T \approx t^T X$ , quedando que:
- $E = X - tt^T X$



$$F = Y - uc^T$$

- Y como  $c \approx Y^T t$ , queda:

$$F = Y - t t^T Y$$

Regresión en el PLS:

El modelo o función de regresión se deduce a partir de:

$$Y = XB$$

$$\min_c \|Y - TC^T\|^2$$

Donde:

$$Y = XW\alpha$$

$$B = W\alpha$$

Entonces:

$$Y = TP^T W\alpha$$

$$\min_\alpha \|TP^T W\alpha - TC^T\|^2 = \min_\alpha \|T(P^T W\alpha - C^T)\|^2$$

La manera de minimizar la expresión es hacer que:

$$P^T W\alpha = C^T$$

Luego al despejar  $\alpha$  y considerando que  $B = W\alpha$ , queda:

$$B = W(P^T W)^{-1} C^T$$

Como:

$$P^T = (T^T T)^{-1} T^T X$$

$$C^T = (T^T T)^{-1} T^T Y$$



Entonces:

$$B = X^T U (T^T X X^T U)^{-1} T^T Y$$

### MÉTODO CUASI- NEWTON: DAVIDON FLETCHER POWELL

Este método resuelve problemas de optimización sin restricciones, donde a través de un proceso iterativo se llega a la minimización del problema. Es una modificación del método de Newton, que por su sencillez y propiedades de convergencia asegura la obtención de la solución.

Siguiendo a Luenberger (1989) se describe a continuación:

### MÉTODO DE NEWTON

Este método se basa en que una función  $f$  se aproxima a una función cuadrática para ser minimizada, para ello cerca del punto de estudio la función se aproxima por la serie de Taylor truncada:

$$f(x) \approx f(x_k) + \nabla f(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^T H(x_k)(x - x_k)$$

El lado derecho se minimiza en:

$$x_{k+1} = x_k - (H(x_k))^{-1} \nabla f(x_k)^T$$

Considerando las condiciones de suficiencia de segundo orden para un punto mínimo, se supone que para un punto mínimo relativo,  $x^*$ , la matriz hessiana  $H(x^*)$  es definida positiva. De manera que si  $f$  tiene segundas derivadas parciales continuas,  $H(x^*)$  es definida positiva cerca de  $x^*$ , por lo que el método está bien definido cerca de la solución.

### MODIFICACIONES DEL MÉTODO DE NEWTON

El método de Newton puede ser modificado de manera de garantizar la convergencia desde cualquier punto de inicio

Dado un  $x_k$ , considere la dirección  $d = -B \nabla f(x_k)$ , donde  $B = (\epsilon I + H)^{-1}$  es una matriz simétrica definida positiva con  $H = H(x_k)$  y  $\epsilon \geq 0$ . La sucesión de puntos se calcula a través de  $x_{k+1} = x_k + \lambda d$ , donde  $\lambda$  es la solución optimal del problema de minimización  $f(x_k + \lambda d)$  sujeto a  $\lambda \geq 0$ .

El escalar  $\epsilon \geq 0$  puede calcularse de la siguiente manera:

Se fija un  $\delta > 0$ , de manera que  $\epsilon \geq 0$  sea el escalar mas pequeño que haga todos los valores propios de la matriz  $(\epsilon I + H)$  mayor o igual a  $\delta$ . Esta matriz es definida positiva, todos sus valores propios son positivos y tiene inversa. En particular, la matriz  $B$  es también definida positiva, ya que los valores propios de una matriz dependen continuamente de sus elementos,  $\epsilon$  es una función continua de  $x$  y de aquí el mapeo



punto a punto  $D: E^n \rightarrow E^n$  definido por  $D(x) = (x, d)$  es continuo. El algoritmo mapea  $A = MD$ , donde  $M$  es la búsqueda lineal usual  $\{\lambda: \lambda \geq 0\}$ .

Sea  $\Omega = \{x^*: \nabla f(x^*) = 0\}$ , y  $x \notin \Omega$ , mientras  $B$  sea definida positiva  $d \neq 0$ , entonces,  $M$  es cerrado en  $(x, d)$ , ya que  $D$  es una función continua, entonces  $A$  es cerrado sobre el complemento de  $\Omega$ . Además, como  $\nabla f(x)^T d = -\nabla f(x)^T B \nabla f(x) < 0$ ,  $B$  es definida positiva y  $\nabla f(x) \neq 0$ , entonces,  $d$  es una dirección de descenso de  $f$  en  $x$ . Defínase a  $y \in A(x)$  de manera que  $f(y) < f(x)$ , entonces  $f$  es una función de descenso y asumiendo que la secuencia generada por el algoritmo esta contenida en un conjunto compacto, se sigue que el algoritmo converge.

Esto demuestra que si el menor valor propio de la matriz  $H$  es mayor o igual que un  $\delta$ , entonces el conjunto de puntos  $\{x_k\}$  generado por el algoritmo alcanza el  $x^*$ . Es de observar que si  $\epsilon_k = 0$ , entonces  $d_k = -H(x_k)^{-1} \nabla f(x_k)$ , lo que nos lleva al algoritmo del método de Newton sin modificación.

La utilidad de este algoritmo presenta como obstáculo el cálculo de los valores propios de la matriz Hessiana. Para solventar esto, se aplican los llamados métodos Cuasi-Newton que forman una aproximación a la inversa del Hessiano. Esta forma puede variar desde las más sencillas, donde la matriz permanece fija durante el proceso iterativo, hasta los más avanzados, donde se construyen aproximaciones mejoradas basadas en la información del proceso.

### CONSTRUCCIÓN DE LA MATRIZ INVERSA

En los métodos Cuasi-Newton la inversa de la matriz Hessiana ( $H$ ) o una aproximación de ella se logran por la información almacenada durante el desarrollo del proceso de descenso. Idealmente, la aproximación converge a la inversa de  $H$  en el punto solución.

Sea  $f \in E^n$  con segundas derivadas parciales continuas si para dos puntos  $x_{k+1}, x_k$ , se define:  $p_k = x_{k+1} - x_k$  como una dirección, entonces:

$$\nabla f(x_{k+1}) - \nabla f(x_k) \cong H(x_k)p_k$$

Si  $H$  es constante, queda:

$$q_k \equiv \nabla f(x_{k+1}) - \nabla f(x_k) = Hp_k$$

De manera que la evaluación del gradiente en dos puntos suministra información sobre  $H$ . Si se determinan "n" direcciones linealmente independientes  $p_k$  con  $k = 0, 1, \dots, n-1$ , de manera que se puede definir a  $P$  y  $Q$  como matrices  $n \times n$  con columnas  $p_k$  y  $q_k$  respectivamente entonces:  $H = QP^{-1}$

Si se define a  $F$  como la aproximación a la inversa del Hessiano, de manera que:

$$F_{k+1}q_i = p_i \quad 0 \leq i \leq k$$



Luego de “n” pasos linealmente independientes se tendrá  $F_n = H^{-1}$ . Sin embargo, para un  $k < n$  hay infinidad de soluciones, pues hay más grados de libertad que restricciones y se tendrían que tener presente algunas consideraciones adicionales, tal como la corrección de rango dos propuesta por Davidon-Fletcher-Powell (Fletcher y Powell, 1963).

### PROPUESTA POR DAVIDON-FLETCHER-POWELL

La inversa de la matriz Hessiana es actualizada en cada paso con la suma de dos matrices de rango uno. Se inicia con cualquier matriz  $F(x_0)$  definida positiva simétrica, con cualquier punto  $x_0$ , y se calculan:

La dirección  $d_k = -F_k \nabla f(x_k)$

Se minimiza  $f(x_k + \alpha d_k)$  respecto  $\alpha \geq 0$  y se determinan

$$x_{k+1} = x_k + \alpha_k d_k, p_k = \alpha_k d_k, \nabla f(x_{k+1}).$$

Se determina:

$$F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k}$$

### ALGORITMO DEL MÉTODO DE NEWTON CON DAVIDON-FLETCHER-POWELL

Dado  $x_0$

Para  $k = 0$  hasta  $n$

Calcúlese:

$$d(x_k) = d_k: \text{la dirección de descenso}, d_k = -F_k \nabla f(x_k)$$

$\alpha_k$  es un escalar que minimiza a  $f(x_k + \alpha_k d_k)$

$$x_{k+1} = x_k + \alpha_k d_k \quad F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k}$$

### ENFOQUE OPTIMAL DEL MÉTODO PLS CON MÉTODO CUASI-NEWTON

- Funciones Objetivos PLS:

El método PLS involucra dos problemas de minimización: el primero expone la minimización a través de  $w$  de la distancia entre los vectores de la matriz  $X$  y sus proyecciones sobre la matriz  $Y$ , es decir:  $\min \|X - Yw^T\|$ . El segundo consiste en minimizar mediante  $c$  la distancia entre los vectores de la matriz  $Y$  con sus proyecciones sobre  $T$ , es decir:  $\min \|Y - Tc^T\|$ .



Estos problemas se pueden expresar como las funciones objetivos siguientes, de acuerdo a lo expresado por Rodríguez (2010):

$$F1(w) = (X - Yw^T)^T(X - Yw^T) = (X^T - wY^T)(X - Yw^T) \\ = X^T X - X^T Y w^T - w Y^T X + w Y^T Y w^T = K - 2X^T Y w^T + w Y^T Y w^T$$

Donde  $K = X^T X$  es una constante, ya que no contiene la variable a minimizar  $w$ .

Si se divide la ecuación entre dos y se escribe como una ecuación cuadrática queda:

$$F1 = \frac{K}{2} - bw^T + \frac{1}{2}wQw^T \quad (3)$$

Con:

$$b = X^T Y, \quad Q = Y^T Y$$

$$F2(c) = (Y - Tc^T)^T(Y - Tc^T) = (Y^T - ct^T)(Y - Tc^T) = Y^T Y - Y^T T c^T - c T^T Y + c T^T T c^T \\ = K - 2Y^T T c^T + c T^T T c^T$$

Donde  $K = Y^T Y$  es una constante, ya que no contiene la variable a minimizar  $c$ .

Si se divide la ecuación entre dos y se escribe como una ecuación cuadrática queda:

$$F2 = \frac{K}{2} - bc^T + \frac{1}{2}cQc^T \quad (4)$$

Con:

$$b = Y^T T, \quad Q = T^T T$$

- PLS y Método de Davidon-Fletcher-Powell (Fletcher y Powell, 1963):

De la función F1: La ecuación (3):

$$F1 = \frac{K}{2} - bw^T + \frac{1}{2}wQw^T$$

Con:

$$b = X^T Y, \quad Q = Y^T Y$$

El gradiente es:  $\nabla F1(w) = Qw - b$

La matriz Hessiana es:  $H(w) = Q$

De manera que el mínimo se encuentra con la sucesión generada por:  
 $w_{k+1} = w_k - (Q)^{-1} \nabla F1(w_k)^T \quad (5)$



Para garantizar la convergencia se aplica el método de Davidon-Fletcher-Powell (Fletcher y Powell, 1963), es decir:

$$w_{k+1} = w_k + \alpha d \quad (6)$$

Con dirección:  $d_k = -F_k \nabla f(w_k)$

Se minimiza  $f(w_k + \alpha d_k)$  respecto  $\alpha \geq 0$  y se determinan  $w_{k+1} = w_k + \alpha_k d_k, p_k = \alpha_k d_k, \nabla f(w_{k+1})$ .

Se determina:

$$F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k}$$

De la función F2: La ecuación (4):

$$F2 = \frac{K}{2} - bc^T + \frac{1}{2} cQc^T$$

Con:

$$b = Y^T T, \quad Q = T^T T$$

El gradiente es:  $\nabla F2(c) = Qc - b$

La matriz Hessiana es:  $H(c) = Q$

De manera que el mínimo se encuentra con la sucesión generada por:

$$c_{k+1} = c_k - (Q)^{-1} \nabla F2(c_k)^T \quad (7)$$

Nuevamente, para garantizar la convergencia se aplica el método de Davidon-Fletcher-Powell (Fletcher y Powell, 1963):

$$c_{k+1} = c_k + \alpha d \quad (8)$$

De dirección:  $d_k = -F_k \nabla f(c_k)$

Se minimiza  $f(c_k + \alpha d_k)$ , respecto  $\alpha \geq 0$  y se determinan  $c_{k+1} = c_k + \alpha_k d_k, p_k = \alpha_k d_k, \nabla f(c_{k+1})$ .

Se determina:

$$F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k}$$



- Algoritmo del PLS con método de Davidon-Fletcher-Powell (Fletcher y Powell, 1963):

Con las matrices  $X$  y  $Y$  centradas se procede con los siguientes pasos:

1. Obtener los vectores: se aplica el algoritmo del método de Newton con Davidon-Fletcher-Powell a las siguientes funciones:

$W$ : a partir de:

$$F1 = \frac{K}{2} - bw^T + \frac{1}{2}wQw^T$$

$C$ : a partir de

$$F2 = \frac{K}{2} - bc^T + \frac{1}{2}cQc^T$$

2. Se calcula:

$$t = Xw \text{ hacer } t = \frac{t}{\|t\|}$$

3. Se calcula:

$$u = Yc \text{ hacer } u = \frac{u}{\|u\|}$$

4. Se obtienen las nuevas matrices:

$$X = X - tt^T X \quad Y = Y - tt^T Y$$

5. Se repite hasta que se cumple un criterio de parada preestablecido:

Los vectores  $t$  y  $u$  son almacenados para construir las matrices  $T$  y  $U$  respectivamente.

6. Se calcula:

$$B = X^T U (T^T X X^T U)^{-1} T^T Y$$

7. Se obtiene el Modelo de Regresión:

$$Y = BX$$

Utilizando este algoritmo se obtiene la reconstrucción de la matriz predictiva  $X$  y la matriz de respuestas  $Y$  como las sumas de matrices de componentes de rango 1 y un conjunto reducido de variables latentes artificiales que se utilizan para describir las variaciones más significativas de estas matrices, mientras que los residuales se van



obteniendo por el proceso de reconstrucción y está relacionado con las variaciones insignificantes de ambas matrices.

## CONCLUSIÓN Y ESTUDIO DE RESULTADOS

En este trabajo se propusieron dos tipos de ejemplos con los que se pretende medir la efectividad y robustez del algoritmo.

### Metodología Experimental:

Se crearon los ejemplos a través del computador usando el programa MATLAB (R2012a), diseñando los mismos bajo un esquema de problema línea multivariable,  $Y = XB + E$ , donde  $X$ : es la matriz que contiene las variables de entrada,  $Y$ : es la matriz que contiene las variables de salida,  $B$ : es la matriz de los coeficientes de regresión y  $E$ : es la matriz de error o ruido aleatorio.

Trabajando en el ejemplo 1 con una gran cantidad de muestras y planteando el ejemplo 2 con una pequeña cantidad de muestras, ambos con más de una variable de salida y con variables de entrada independientes y dependientes, se tuvo que son problemas de alta dimensionalidad y multicolinealidad.

Para el ejemplo 1, los datos se generaron con 100 muestras y se representan en cuatro variables de salidas  $Y = [Y_1 Y_2 Y_3 Y_4]$  de tamaño  $100 \times 4$ , seis variables de entrada  $X = [X_1 X_2 X_3 X_4 X_5 X_6]$  de tamaño  $100 \times 6$ , una matriz de error  $E = [E_1 E_2 E_3 E_4]$  de tamaño  $100 \times 4$  y la matriz de los coeficientes de regresión  $B = [B_1 B_2 B_3 B_4]$  de tamaño  $6 \times 4$ . Para este ejemplo, se diseñaron a  $X_1, X_2$  y  $X_3$  como variables independientes, generadas de forma aleatoria con distribución normal, y el resto de las variables de entrada son dependientes basadas en las siguientes formulas:

$$X_4 = 2X_1 + X_3$$

$$X_5 = 3X_2$$

$$X_6 = X_1 + X_2 + X_3$$

Para el ejemplo 2: los datos se crearon con 20 muestras, 5 variables de salida  $Y = [Y_1 Y_2 Y_3 Y_4 Y_5]$  de tamaño  $20 \times 5$ , 10 variables de entrada  $X = [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10}]$  de tamaño  $20 \times 10$ , una matriz de error  $E = [E_1 E_2 E_3 E_4 E_5]$  de tamaño  $20 \times 5$  y la matriz de los coeficientes de regresión  $B = [B_1 B_2 B_3 B_4 B_5]$  de tamaño  $10 \times 5$ . Este ejemplo se diseñó con  $X_1, X_2, X_3, X_4, X_5$  y  $X_6$  como variables independientes generadas de forma aleatoria con distribución normal y el resto de las variables de entrada como dependientes basada en las formulas siguientes:

$$X_7 = 2X_1$$

$$X_8 = 3X_1 + 2X_3$$

$$X_9 = X_1 + X_2 + X_4$$

$$X_{10} = X_4 + 2X_5$$

La presencia de la matriz de errores ( $E$ ) evidenció la estabilidad del modelo y la variabilidad en el número de observaciones entre el primer ejemplo (100 muestras) y el segundo ejemplo (20 muestras). También, mostró que el algoritmo se puede usar tanto en los casos en los que es fácil la obtención de muestras como cuando hay limitaciones para la obtención de las mismas.

Los datos se dividieron en dos grupos: la mitad de los datos para entrenar el modelo y la otra mitad para validar el modelo, considerando como criterios para la validación la suma de los cuadrados de los errores (SSE), el promedio de la suma de los cuadrados de los errores (PSSE) y el coeficiente de regresión lineal entre los  $y$  dados y los  $y$  estimados ( $R$ ).

Los valores aquí expresados son los resultados de la validación de los datos:

**Tabla 1. Resultados de la validación del ejemplo 1. Número de variables latentes: 1**

Variable de salida	SSE	PSSE	$R^2$
Y1	6.2989	0.0840	0.9217
Y2	2.7189	0.0363	0.9290
Y3	23.0761	0.3077	0.6394
Y4	3.2723	0.0436	0.9380

Fuente: elaboración propia.

**Tabla 2. Resultados de la validación del ejemplo 1. Número de variables latentes: 3**

Variable de salida	SSE	PSSE	$R^2$
		( $10^{-3}$ )	
Y1	0.0056	0.0751	0.9999
Y2	0.0086	0.1151	0.9998
Y3	0.0064	0.0847	0.9999
Y4	0.0094	0.1254	0.9998

Fuente: elaboración propia.

**Tabla 3. Resultados de la validación del ejemplo 1. Número de variables latentes: 6**

Variable de salida	SSE	PSSE	$R^2$
Y1	31.6330	0.4218	0.6066
Y2	13.4723	0.1796	0.6480
Y3	23.1982	0.3093	0.6375
Y4	20.5347	0.2738	0.6112

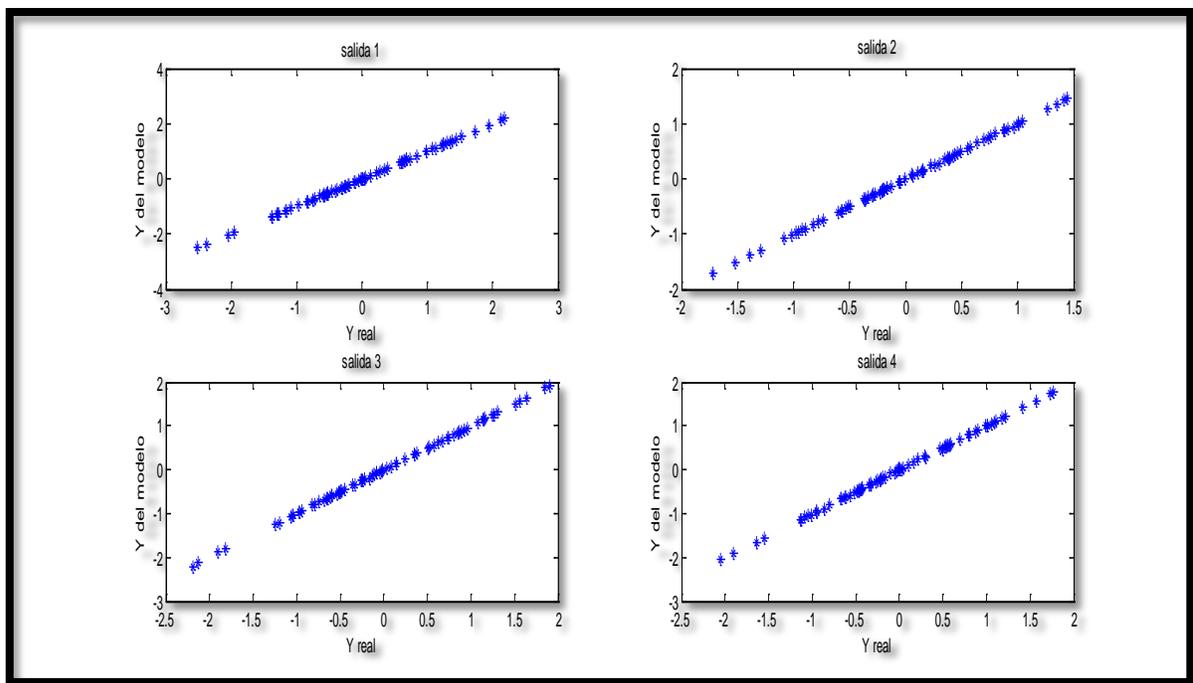
Fuente: elaboración propia.

Como era de esperarse, se observa en las Tablas 1, 2 y 3 que el mejor resultado se obtiene con tres variables latentes, debido a que el número de variables independientes

son tres. Al trabajar con un menor o mayor número de variables se da un ajuste deficiente.

También se observa que el modelo es robusto en presencia de una matriz de error con distribución normal. En la Figura 1 se muestra el ajuste lineal que hay entre las variables de salida reales y las variables de salida obtenidas por el modelo, donde se observa que se forma una línea recta con pendiente casi uno, corroborando la cercanía entre los valores obtenidos por el modelo y los reales.

**Figura 1. Representación de las variables de salida reales versus las variables de salida del modelo del ejemplo1**



Fuente: elaboración propia.

**Tabla 4. Resultados de la validación del ejemplo 2. Número de variables latentes: 1**

Variable de salida	SSE	PSSE	$R^2$
Y1	0.4774	0.0477	0.9288
Y2	0.7670	0.0767	0.8501
Y3	0.7493	0.0749	0.9019
Y4	3.0414	0.3041	0.4388
Y5	1.3108	0.1311	0.8929

Fuente: elaboración propia.



**Tabla 5. Resultados de la validación del ejemplo 2. Número de variables latentes: 3**

Variable de salida	SSE	PSSE	$R^2$
Y1	0.7382	0.0738	0.8900
Y2	0.0881	0.0088	0.9828
Y3	0.5615	0.0561	0.9265
Y4	0.3268	0.0327	0.9397
Y5	0.0684	0.0068	0.9944

Fuente: elaboración propia.

**Tabla 6. Resultados de la validación del ejemplo 2. Número de variables latentes: 6**

Variable de salida	SSE	PSSE ( $10^{-3}$ )	$R^2$
Y1	0.0012	0.1235	0.9998
Y2	0.0026	0.2560	0.9995
Y3	0.0008	0.0840	0.9999
Y4	0.0039	0.3948	0.9993
Y5	0.0012	0.1159	0.9999

Fuente: elaboración propia.

**Tabla 7. Resultados de la validación del ejemplo 2. Número de variables latentes: 10**

Variable de salida	SSE	PSSE	$R^2$
Y1	0.3932	0.0393	0.9414
Y2	0.7622	0.0762	0.8511
Y3	0.8242	0.0824	0.8921
Y4	0.5482	0.0548	0.8989
Y5	1.7480	0.1748	0.8571

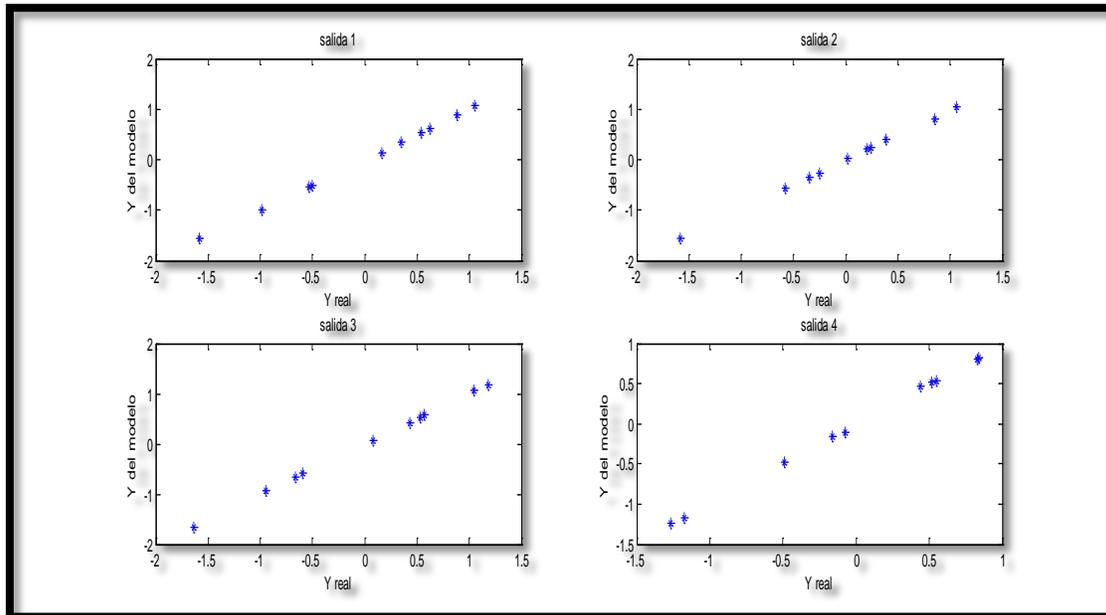
Fuente: elaboración propia.

Para este ejemplo, cabe destacar que el número de observaciones es pequeño y cercano al número de variables de entrada, lo que verificó la eficiencia y robustez del modelo.

El mejor resultado, como se observa en las Tablas 4, 5, 6 y 7, se obtiene con seis variables latentes, lo cual corresponde al número de variables independientes de entrada. Sin embargo, se observa que con tres variables latentes los resultados son aceptables, lo que indica que ese número de variables latentes aporta la mayor información para el ajuste del modelo.

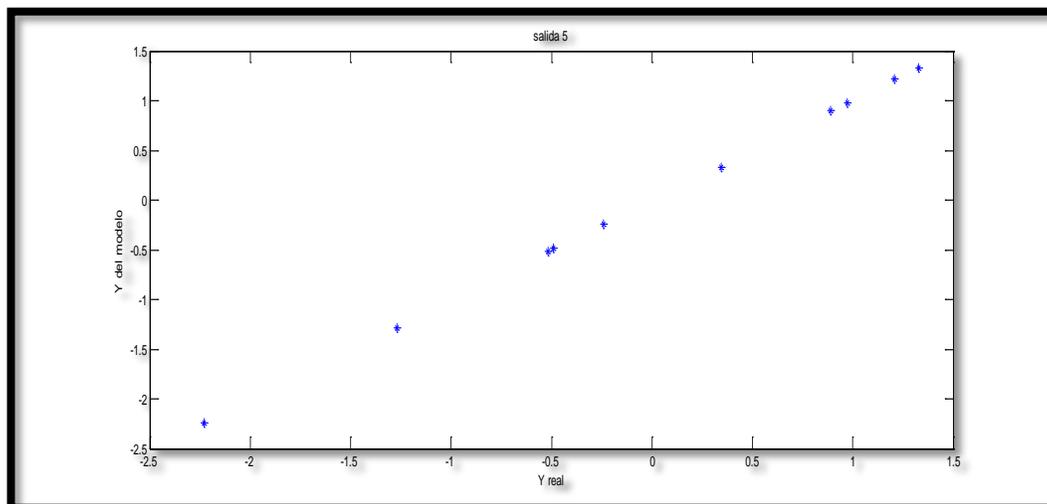
En las Figuras 2 y 3 se presenta el ajuste lineal que hay entre las variables de salida reales y las variables de salida obtenidas por el modelo, en el caso de seis variables latentes. Aquí se observa que se forma una línea recta con pendiente casi uno, lo que certifica la cercanía entre los valores obtenidos por el modelo y los reales.

**Figura 2. Representación de las variables de salida reales versus las variables de salida del modelo del ejemplo2 (de la primera a la cuarta)**



Fuente: elaboración propia.

**Figura 3. Representación de las variables de salida reales versus las variables de salida del modelo del ejemplo2 (la quinta)**



Fuente: elaboración propia.



Se concluye que el algoritmo aquí presentado, el cual integra el método de Mínimos Cuadrados Parciales y el método de optimización sin restricción Cuasi-Newton con la propuesta de Davidon-Fletcher-Powell (Fletcher y Powell, 1963) es eficiente y robusto en el uso del modelado de problemas de regresión lineal y es aplicado a problemas donde existen más de una salida.

### REFERENCIAS BIBLIOGRÁFICAS

- Fletcher, R. y Powell, M. (1963). A Rapidly Convergent Descent Method for Minimization. *The Computer Journal*. Volumen 6, número 2. (Pp. 163-168).
- Geladi P. y Kowalski B. (1986a). Partial Least-Squares Regression: a tutorial. *Analytica Chimica Acta*. Volumen 185. (Pp. 1-17).
- Geladi P. y Kowalski B. (1986b). An example of 2-block predictive partial least- squares regression with simulated data. *Analytica Chimica Acta*. Volumen 185. (Pp. 19-32).
- Hoskuldsson A. (1998). PLS Regression Methods. *Journal of Chemometrics*. Volumen 2. (Pp. 211-228).
- Luenberger D. (1989). Programación lineal y no lineal. México. Addison-Wesley Iberoamericana.
- Mardia, K.; Kent, J. y Bibby J. (1979). *Multivariate analysis*. Estados Unidos. Editorial Elsevier.
- Rey E. (2009). Estudio de Mezclas por Resonancia Magnética Nuclear. Tesis de Grado. Maestría en Ciencias–Química. Universidad Nacional de Colombia. Colombia.
- Rodríguez E. (2010). Desarrollo del método Cuadrados Mínimos Parciales (PLS) a partir del método de Factores Latentes Fortalecidos (BLF) y de algunos Métodos de Optimización sin Restricción. Trabajo de Ascenso para optar a la categoría de profesor titular. Universidad del Zulia. Venezuela.