ISSN: 2773-7330

Alpha publicaciones

Vol. 3, N° 3.1, p. 83-97
agosto, 2021

Recibido: 13-06-2021 / Revisado: 22-06-2021 / Aceptado: 11-07-2021 / Publicado: 05-08-2021

Usando la Factorización de Rango Completo de Cholesky en la Solución de los Mínimos Cuadrados Ponderados

DOI: https://doi.org/10.33262/ap.v3i3.1.79

Full Rank Cholesky Factorization for the Re-Weighted Least Squares Applications

Zenaida Natividad Castillo Marrero. [1], Ernesto Antonio Ponsot Balaguer. [2], Franklin José Camacho. [3] & María Victoria León Sánchez. [4]

Abstract

Factorization of matrices are required frequently in sciences research, in order to simplify computations when solving linear or non-linear systems of equations. In particular, these systems arise in data analysis and also after modeling industrial process, playing an important role in the final solution. When the matrices involved are full rank and squared classical factorizations can be used efficiently. However, when one has to deal with rectangular and/or rank deficient matrices, extensions of these factorizations should be considered, if it is possible with the same good properties. **Objective**: In this work a full rank Cholesky factorization for rank deficient matrices is described, along with its use in a particular application arising in Statistics, where usually one has to solve overdetermined systems of equations via least squared method or its extensions. **Methodology**: We describe an algorithm to generate this factorization in an iteratively

[1] Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Escuela de Matemática, Grupo CIDED, Riobamba, Ecuador, zenaida.castillo@espoch.edu.ec, https://orcid.org/0000-0002-4424-8652.

[2] Data Science Consulting/Universidad de los Andes, Escuela de Estadística. Mérida, Venezuela, ernesto.pb@gmail.com, https://orcid.org/0000-0001-5221-1799.

[3] Universidad de Investigación de Tecnología Experimental Yachay, Escuela de Ciencias Matemáticas y Computacionales, Urcuquí,060150, Ecuador, fcamacho@yachaytech.edu.ec, https://orcid.org/0000-0001-7802-5687

[4] Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Escuela de Matemática, Riobamba, Ecuador, maria.leon@espoch.edu.ec, https://orcid.org/0000-0003-0612-7478.

reweighed least squared method applied to maximum likelihood parameters estimation. **Results**: Preliminary experiments are presented over synthetic data, supporting the use of this decomposition for rectangular or rank deficient matrices arising in similar applications. **Conclusions**: The proposal of using the full rank Cholesky factorization in the iteratively reweighted least squared to estimate maximum likehood parameters is promising and could be used to improve CPU time in the solution of similar problems.

Keywords: Matrix Factorizations, Rank deficiency, Full rank Cholesky, Weighted Least squares.

Resumen

Es frecuente encontrar en la investigación en las ciencias e ingenierías, que se requiere de factorizaciones de matrices para facilitar o simplificar los cálculos, ya sea para resolver sistemas de ecuaciones lineales, o no lineales, provenientes de análisis de datos o de modelado o simulación de procesos industriales. Cuando las matrices son cuadradas y de rango completo existen factorizaciones que permiten hacer este trabajo eficientemente; sin embargo, cuando las matrices son rectangulares y/o deficientes en rango, las extensiones de estas factorizaciones se hacen necesarias y deben ser tratadas cuidadosamente para que los resultados sean confiables. **Objetivo**: En este trabajo se describe la descomposición de rango completo de Cholesky y su uso en algunas aplicaciones en áreas como la Estadística, donde a menudo se requiere resolver sistemas sobredeterminados por el método de los mínimos cuadrados o extensiones del mismo. **Metodología**. Se describen los pasos para generar esta descomposición y su uso en la estimación máximo verosímil de parámetros por el método de los mínimos cuadrados reponderados. **Resultados**: Se presenta experimentación preliminar en datos sintéticos, con resultados que sustentan el uso de esta descomposición para matrices rectangulares o deficientes en rango. **Conclusión**: La factorización propuesta para resolver el problema de estimación de parámetros con la función de verosimilitud es satisfactoria y pudiera ser utilizada para optimizar el tiempo de cómputo de la solución de problemas similares.

Palabras claves: Factorización de Matrices, Deficiencia en rango, Factorización de rango completo, Cholesky, Mínimos Cuadrados Ponderados.

Introducción

Uno de los cálculos que aparece con mayor frecuencia en la resolución de problemas de modelado industrial, o en el análisis de datos, es sin duda la solución de sistemas lineales de la forma $\mathbf{Ax} = \mathbf{b}$, con **m** ecuaciones y **n** incógnitas $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$, los cuales pueden resolverse fácilmente cuando la matriz **A** es cuadrada, de rango completo, de tamaño moderado, y quizás con algunas propiedades convenientes como el ser simétrica y/o definida positiva. Para efectos de precisión en los cálculos y optimización de recursos computacionales, estos sistemas pueden ser resueltos por métodos directos previa descomposición de la matriz **A** en el producto de dos o más matrices. Existen también diversos métodos iterativos que muestran excelentes características de convergencia generando aproximaciones a la solución tan cercanas como se desee. Muchos de estos métodos también hacen uso internamente de factorizaciones de matrices. Una descripción detallada de estos métodos puede encontrarse en Bjorck (1996).

Cuando la matriz es simétrica y definida positiva es conveniente usar la descomposición de Cholesky, ya que se disminuye considerablemente el tiempo de cálculo y se aprovecha la fortaleza de esta factorización. En la mayoría de los lenguajes de programación que se manejan en la actualidad, existen librerías con códigos robustos para estas descomposiciones, los cuales funcionan apropiadamente para aplicaciones generales densas o dispersas, ver Datta (2010) para mayor referencia.

Ahora bien, en ocasiones las matrices tienen una estructura particular que pudiera explotarse para ganar precisión o rapidez en los cálculos; en este caso los investigadores prefieren elaborar sus propios códigos.

Para matrices semidefinidas positivas y para matrices rectangulares de rango incompleto aún es posible hallar una extensión de la factorización de Cholesky con las mismas propiedades o el mismo potencial. En este trabajo se describen este tipo de factorizaciones y se presenta su uso en aplicaciones de interés en áreas de proyección en la actualidad. La experimentación preliminar muestra la efectividad de la propuesta en un ejemplo de aplicación.

Factorizaciones de matrices

En esta sección y para efectos de comparación se presenta un resumen de las factorizaciones clásicas de matrices, que han resultado de mucha utilidad para resolver sistemas lineales de tamaño moderado cuya matriz de coeficientes es de rango completo.

Factorización LU

La factorización $\mathbf{A} = \mathbf{LU}$, con $\mathbf{L}$ triangular inferior y $\mathbf{U}$ triangular superior, está garantizada cuando todas las submatrices principales de $\mathbf{A}$ son no-singulares, y se asocia a las operaciones elementales de fila que tienen lugar en el proceso de triangularización de la matriz $\mathbf{A}$ vía eliminación Gaussiana. El proceso culmina con la obtención de una matriz triangular superior $\mathbf{U}$, y los multiplicadores que permiten crear un cero en la posición $\mathbf{A}$ **(i, j)** se colocan en la posición **(i, j)** de la matriz $\mathbf{L}$.

Esta factorización es usada posteriormente para resolver el sistema $\mathbf{Ax} = \mathbf{b}$ en dos fases, que se muestran en la siguiente figura:

Figura 1.

$\mathbf{Ax} = \mathbf{b}$ vía factorización $\mathbf{LU}$ de $\mathbf{A}$

$$
\boxed{
\begin{aligned}
& A = LU \\
\Rightarrow\ & Ax = b \equiv L(Ux) = b \\
& \qquad\quad \equiv \begin{cases} Ly = b \\ Ux = y \end{cases}
\end{aligned}
}
$$

Fuente: Elaboración propia.

El esquema mostrado en la figura 1 puede ser usado para calcular la inversa de la matriz **A** en problemas de tamaño moderado.

Una de las debilidades de la factorización **LU** es su inestabilidad numérica; es por ello que en muchas aplicaciones es preferible usar, a un mayor costo computacional, otro tipo de factorización, como la factorización $\mathbf{A} = \mathbf{QR}$, conocida por su estabilidad numérica. Para mayor detalle sobre esta factorización refiérase a Trefethen (1997).

Cuando la matriz **A**, de orden **n**, es simétrica y definida positiva, una buena alternativa en términos de estabilidad, que resulta menos costosa, es la factorización de Cholesky, la cual descompone a la matriz **A** en el producto de una matriz triangular inferior **L** y su traspuesta $\mathbf{L}^{\mathbf{T}}$. Así $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$.

Esta descomposición, en forma similar a la factorización **LU**, utiliza el proceso basado en la eliminación Gaussiana para producir la matriz triangular **L**; la diferencia es que al hallar **L** ya se tiene que $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$, los cálculos son más estables y la simetría garantiza un menor costo computacional. La solución de **Ax=b** usando la factorización de Cholesky se plantea en forma similar a la mostrada en la figura 1 para la factorización **LU**:

$$\mathbf{Ax} = \mathbf{b} \equiv \mathbf{L}\mathbf{L}^{\mathbf{T}}\mathbf{x} = \mathbf{b} \equiv \mathbf{Ly} = \mathbf{b} \quad \wedge \quad \mathbf{y} = \mathbf{L}^{\mathbf{T}}\mathbf{x}$$

Toda vez que la factorización de Cholesky posee características de estabilidad, se reduce el costo computacional porque no se requiere pivoteo en la eliminación Gaussiana. Así, en aplicaciones donde la matriz de los coeficientes es simétrica y definida positiva se prefiere la factorización de Cholesky para resolver **Ax=b**. Más aún, en la práctica los algoritmos más eficientes no usan la eliminación Gaussiana para hacer esta descomposición, sino que se aprovechan de la estructura de la factorización, reduciendo aún más el costo computacional.

Algoritmo de Cholesky

La estructura de la factorización de Cholesky de una matriz **A**, de orden **n**, se muestra en la siguiente figura:

Figura 2.

Estructura de la factorización de Cholesky

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n} & a_{n2} & \cdots & a_{nn} \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0\cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}}_{\mathbf{L}} \underbrace{\begin{pmatrix} l_{11} & l_{21} & \cdots & l_{1n} \\ 0 & l_{22} & \cdots & l_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & l_{nn} \end{pmatrix}}_{\mathbf{L}^{\mathbf{T}}}$$

Fuente: Elaboración propia.

Una simple inspección de la figura 3 nos lleva al algoritmo de Cholesky, el cual se basa en hallar los elementos de **L** en forma iterativa. Conociendo el valor de $\boldsymbol{a_{11}}$, la primera entrada de la matriz **A**, podemos inferir que el valor de $\mathbf{l_{11}}$, la primera entrada de **L**; luego se usa este valor de $\mathbf{l_{11}}$ para calcular el resto de las entradas de la primera columna. Este procedimiento se repite para el resto de las columnas. El esquema algorítmico es presentado en la figura 3.

Figura 3.

Algoritmo básico de Cholesky

Cholesky Algorithm
for $k = 1, 2, \cdots n$ **do**
$\quad l_{ki} = \frac{1}{l_{ii}} \left(a_{ki} - \sum_{j=1}^{i-1} l_{ij} l_{kj} \right)$
end for

Fuente: Elaboración propia.

Cuando hablamos de matrices de $m \times n$, con $m > n$, que representan grandes volúmenes de datos, mal condicionadas o de partida con columnas dependientes, debemos conformarnos con hallar una buena aproximación del vector $\mathbf{Ax}$ al vector $\mathbf{b}$; es decir lograr que el tamaño de $\|\mathbf{Ax} - \mathbf{b}\|$ satisfaga una tolerancia aceptable para la aplicación que se analiza. Particularmente, cuando el sistema tiene más ecuaciones que incógnitas ($m > n$), es posible que el sistema $\mathbf{Ax} = \mathbf{b}$ no tenga solución, y en este caso se busca la mejor solución en el sentido de los mínimos cuadrados.

El método de los mínimos cuadrados es ampliamente usado en aplicaciones de análisis de datos; en particular se propone en este trabajo, el estudio de una de sus extensiones como lo es el método de los mínimos cuadrados reponderados, el cual es una extensión del conocido método de los mínimos cuadrados ponderados que se describe en la próxima sección.

Métodos de los Mínimos Cuadrados Ponderados

Esta extensión del método de los mínimos cuadrados, también conocido como el método GSK, fue propuesto por Grizzie, Starmer y Koch en 1969, para ser usado en la estimación de parámetros de un modelo lineal generalizado para variables de respuesta con distribución multinomial, o de Poisson.

El método introduce una variante en el estudio de variables categóricas, una vez que permite analizarlas a través de una función de respuesta en lugar de sus frecuencias absolutas. Esta función de respuesta se escogerá de acuerdo a la naturaleza de los datos y los objetivos que persiga la investigación. Para mayor detalle sobre la evaluión y cálculo de estas funciones se recomienda Forthofer y Lehnen (1981).

De forma análoga al método de los mínimos cuadrados, este método parte de un conjunto de observaciones $(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), \ldots, (\mathbf{x_n}, \mathbf{y_n})$, y se busca la recta $\mathbf{y} = \mathbf{ax} + \mathbf{b}$ que minimice la suma de los cuadrados de los errores entre los $\mathbf{y_i}$ y los obtenidos en la recta de los valores observados, que satisface $\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{a}\boldsymbol{x}_{\boldsymbol{i}} + \mathbf{b}$, **con** $\mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}$, de acuerdo a una norma ponderada con pesos $\boldsymbol{w_i}$, $con\ i = 1, \ldots, n$. El objetivo entonces es minimizar:

$$\sum_{i=1}^{n} \mathbf{w}_i (\hat{\mathbf{y}}_\mathbf{i} - \mathbf{y_i})^2$$

La escogencia de los pesos obedece a criterios para reducir o amplificar la influencia de las observaciones en la recta de ajuste; en la práctica se suele asociar a cada observación **i** un peso inversamente proporcional a la variabilidad de $\mathbf{y_i}$, la cual se aproxima en función de $\mathbf{x_i}$. En cualquier caso, si la suma de los pesos en igual a 1, la solución del problema se basa en los cálculos que se presentan en la figura 4.

Figura 4.

Solución general de los mínimos cuadrados ponderados

$$\hat{a} = \frac{\left(\sum_{i=1}^{n} w_i x_i y_i\right) - \left(\sum_{i=1}^{n} w_i y_i\right)\left(\sum_{i=1}^{n} w_i x_i\right)}{\left(\sum_{i=1}^{n} w_i x_i^2\right) - \left(\sum_{i=1}^{n} w_i x_i\right)^2}$$

$$\hat{b} = \frac{\left(\sum_{i=1}^{n} w_i y_i\right)\left(\sum_{i=1}^{n} w_i x_i^2\right) - \left(\sum_{i=1}^{n} w_i x_i y_i\right)\left(\sum_{i=1}^{n} w_i x_i\right)}{\left(\sum_{i=1}^{n} w_i x_i^2\right) - \left(\sum_{i=1}^{n} w_i x_i\right)^2}$$

Fuente: Elaboración propia.

Ahora bien, tal como se ha dicho, las funciones de peso pueden variar de acuerdo a una ponderación conveniente; de esta manera pudiéramos tener funciones como:

1) Ponderación uniforme: $\mathbf{w_i} = \frac{\mathbf{1}}{\mathbf{n}}$, i =1, …, n, que nos conduce al método de los mínimos cuadrados ordinarios.

2) Ponderación Bisquare: $\mathbf{w_i} = \left[\mathbf{1} - \left(\frac{x_i - \bar{x}}{\mathbf{d}}\right)^2\right]^2$ si $\lceil x_i - \bar{x} \rceil < d$, y $\mathbf{w_i} = \mathbf{0}$ de lo contrario. Donde **d** es un factor de escala que garantiza que la función es positiva; el máximo valor de **d** definirá el número máximo de puntos con peso distinto de cero, y garantiza una distribución de pesos.

3) Ponderación de Huber: $\mathbf{w_i} = \frac{\mathbf{1}}{\lceil \mathbf{x_i} - \bar{\mathbf{x}} \rceil}$.

4) Ponderación exponencial $w_i = \boldsymbol{exp}\left[-\left(\frac{x_i - \bar{x}}{\alpha.\mathbf{d}}\right)\right]$ si $\lceil \mathbf{x_i} - \bar{\mathbf{x}} \rceil < \mathbf{d}$, y $\mathbf{w_i} = \mathbf{0}$ de lo contrario. La variable es seleccionada con un criterio de proporcionalidad, en el cual $\boldsymbol{\alpha} > \mathbf{0}$, y **d** es un factor de escala que garantiza que la función sea positiva.

Existen otras funciones de ponderación que son más apropiadas en ciertas aplicaciones, tales como la ponderación Gaussiana, el spline cúbico, y otras que conllevan cálculos más pesados. En la práctica, las más usadas son las ponderaciones de Bisquare y la de Huber. Para mayores detalles sobre estas funciones refiérase a Torkzi (2011).

El método iterativo de los mínimos cuadrados reponderados (IRLS por sus siglas en inglés) resuelve en cada paso un problema usando mínimos cuadrados ponderados, y resulta de mucho interés en aplicaciones donde se busca minimizar una norma p.

En particular el método es muy útil en el análisis de modelos lineales generalizados, en los cuales se estudia la relación:

$$\mathbf{y_{ij}} = \mathbf{x'_{ij}}\boldsymbol{\beta} + \mathbf{e_{ij}}, \qquad con \quad j = 1,2,\ldots,n_i, \quad i = 1,2,\ldots,m.$$

El problema general que se resuelve es:

$$\underbrace{\boldsymbol{arg\ min}}_{\boldsymbol{\beta}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} |\, \boldsymbol{y_i} - \boldsymbol{f_i}\,(\boldsymbol{\beta})|^{\mathbf{2}}$$

En este modelo, $\mathbf{y_{ij}}$ denota las variables de respuestas, $\mathbf{x'_{ij}}$ los valores de la traspuesta de la matriz de covarianza $\mathbf{x_{ij}}$, $\mathbf{e_{ij}}$ los errores aleatorios que usualmente se asumen independientes y $\mathbf{var}(\mathbf{e_{ij}}) = \boldsymbol{\sigma}^{\mathbf{2}}$. Usualmente se aplica el método ordinario de los mínimos cuadrados para estimar algún parámetro $\boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\beta})$ de interés.

Sin embargo, existen situaciones donde los errores se distribuyen por grupos y entre grupos difieren; en este caso el método de los mínimos cuadrados ponderados es una mejor herramienta para estimar $\boldsymbol{\theta}$.

También podemos encontrar situaciones que no se relacionen con las $\mathbf{x_{ij}}$, y en ese caso el método de los mínimos cuadrados ponderados no funcionaría, y se sugiere usar un proceso iterativo en el cual se estime en cada paso un valor para $\boldsymbol{\beta}$ que permita obtener los estimadores para $\boldsymbol{\sigma}^{\mathbf{2}}$ y aplicar los mínimos cuadrados ponderados. Para mayor entendimiento de estas situaciones ver Fuller y Rao (1978), y Davidian y Carroll (1987).

La figura 5 muestra los pasos básicos que contempla el algoritmo iterativo de los mínimos cuadrados reponderados.

Figura 5.

Esquema de iteración de los mínimos cuadrados ponderados

1) $\hat{\theta}_w^{(m)} = g\left(\hat{\beta}_w^{(m)}\right)$ $\qquad$ m = 1, 2, ...

2) $\hat{\beta}_w^{(m)} = \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} w_i^{(m)} x_{ij}x'_{ij}\right)^{-1} \sum_{i=1}^{k}\sum_{j=1}^{n_i} w_i^{(m)} x_{ij}y_{ij}$

3) $w_i^{(m)} = \left[v_i\left(\hat{\beta}_w^{(m-1)}\right)\right]^{-1}, \quad i = 1, 2, \ldots, k$

Fuente: Elaboración propia.

El valor inicial $\boldsymbol{\beta}_w^{(0)}$ puede tomarse como aquel que arroja el método de los mínimos cuadrados ordinarios, de esta manera, el primer iterado $\boldsymbol{\beta}_w^{(1)}$ será el de los mínimos cuadrados ponderados. Para una descripción detallada del método se recomienda Chen (1993).

En la práctica, todos estos métodos tienen una forma matricial y la solución se halla resolviendo el correspondiente sistema de ecuaciones normales de $m \times n$, que trataremos en la próxima sección en su forma matricial $\mathbf{A^T\, A\, x} = \mathbf{A^T\, b}$.

Factorización de rango completo de Cholesky

En Ponsot (2011), el autor estudia la agregación de niveles de factores en un modelo logit binomial, y una de las propuestas es la utilización del método de los mínimos cuadrados iterativos reponderados para estimar los parámetros a través de la función de verosimilitud. Las pruebas realizadas con la implementación muestran el éxito de esa propuesta en la aplicación de estudio.

Así como esta aplicación surgen otras en las cuales el método de los mínimos cuadrados iterativo ponderado resulta de utilidad. Es por ello que proponemos en este trabajo, una implementación robusta de la factorización de rango completo de Cholesky que mejore los tiempos de cómputo en las aproximaciones a los parámetros en cada iteración.

Una idea es implementar esquemas numéricos de la factorización rectangular sobre la matriz de los coeficientes del sistema de ecuaciones normales que se resuelve en cada iteración, en los cuales se eviten cálculos de inversas y la construcción de matrices de gran tamaño como es el caso de $\mathbf{A^T\, A}$ que para el caso de estudio no es de rango completo.

Una matriz **A**, deficiente en rango por columnas, puede descomponerse en el producto de dos matrices rectangulares **G** y **U**, tales que $\mathbf{A} = \mathbf{GU}$, donde $\mathbf{G} \in \mathfrak{R}^{\mathbf{mxr}}$ y $\mathbf{U} \in \mathfrak{R}^{\mathbf{rxn}}$ es escalonada reducida con diagonal unitaria por bloques, con $\mathbf{r} = \mathbf{rango}(\mathbf{G}) = \mathbf{rango}(\mathbf{U})$. En este caso, la matriz **G** tiene rango completo por columnas y $\mathbf{G^TG}$ es simétrica definida positiva, y en consecuencia el cálculo del factor $\mathbf{L_G}$ de Cholesky no requiere la formación del producto $\mathbf{G^TG}$. La matriz $\mathbf{L_G}$ es entonces utilizada para calcular el factor **L** escalonado de **A**, el cual resulta ser $\mathbf{L} = \mathbf{U^T L_G}$.

En Urbano et al. (2011), los autores proponen una modificación a este esquema que evita la factorización inicial de **A** en **G** y **U**, ganando así precisión en los cálculos y mejor utilización del tiempo de CPU. El proceso consiste en ir descartando las columnas que no aportan al rango a medida que se anulan las entradas de **A**. Para una explicación detallada de las propiedades que satisface este esquema de factorización ver Cantó, et. al (2009, 2015).

Se presenta en la figura 6 una adaptación del algoritmo descrito en Cantó et al (2015) para la obtención de la factorización de rango completo de una matriz **A** de *mxm*.

Figura 6.

Algoritmo para la factorización rango completo de Cholesky

Algorithm 1 Factorización rango completo de Cholesky

Require: A: Matriz simétrica semidefinida positiva, tol: tolerancia exigida.

```
L_11 = sqrt(a_11); rango = 1;
iter = 0; c = [1];
for each i = 2 : m do
    for each j = 2 : i − iter − 1 do
        l_{i,j} = (a_{i,c_j} − Σ_{t=1}^{j−1} l_{c_j,t} * l_{i,t}) / l_{c_j,j}
    end for
    aux := sqrt(a_{i,i} − Σ_{t=1}^{i− iter-1} l_{i,t}^2)
    if aux < tol then
        iter = iter + 1;
    else
        l_{i,i−iter} = aux;
        c = [c  i];
        r = r + 1;
    end if
end for
L = L(c);
```

Ensure: L, r, c

Fuente: Adaptación de Cantó (2015).

La salida del algoritmo contiene el factor **L** de la factorización, el número de columnas linealmente independientes (rango) y un vector **c** que contiene los índices de estas columnas.

Note que el algoritmo recibe una matriz cuadrada; pero una adaptación del mismo puede ser usada en matrices rectangulares. Resultados en esta dirección pueden encontrarse en Cantó (2015), Gustavson (2009) y Camero (2018).

Experimentación numérica

En esta sección se presenta un código, con propósito particular, del método de los mínimos cuadrados iterativo reponderado, usado en Ponsot (2011) en un estudio estadístico de estimación de parámetros por máxima verosimilitud en un modelo logit binomial. Se discuten las operaciones susceptibles a mejora computacional en términos de uso de memoria y tiempo de CPU.

En el trabajo de Ponsot (2011) se plantea el método de los mínimos cuadrados reponderados para hallar la estimación máximo-verosímil de parámetros de acuerdo a la siguiente iteración:

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} + (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{y}-\boldsymbol{\mu})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^r} \qquad \boldsymbol{r} = \boldsymbol{0}, \boldsymbol{1}, \ldots$$

- β: vector de parámetros a estimar.
- X: matriz de diseño de la experimentación, con valores no aleatorios.
- μ: vector de las medias de las observaciones **y** observadas de la muestra.

El vector de parámetros inicial $\boldsymbol{\beta}^{(0)}$ se toma como aquel que resuelve $\mathbf{g}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}^{(0)}$, donde $\mathbf{g}(\boldsymbol{\mu})$ representa una transformación biyectiva de la media.

El algoritmo para los mínimos cuadrados iterativo reponderado utilizado fue implementado en Matlab y se presenta en la figura 7.

Figura 7.

Estimación de máxima verosimilitud con MCRP

Algorithm 2 Mínimos Cuadrados Iterativo Reponderado (MCRP)

1: function $[B, InfF, InvG\_InfF, iter] = MCRP(X, y, n, tol)$
2: $y_0 = logitMod(y./n)$
3: $B_0 = pinv(X' * X) * X' * y0$
4: $conv = 0$; $iter = 0$;
5: **while** $conv == 0$ **do**
6: $\quad iter = iter + 1$;
7: $\quad XB = X * B_0$;
8: $\quad p = InvLogit(XB)$;
9: $\quad \mu = n. * p$;
10: $\quad W = diag(\mu. * (1 - p))$;
11: $\quad InfF = X' * W * X$;
12: $\quad InvG\_InfF = pinv(InfF)$;
13: $\quad$ **if** $norm(B - B_0) < tol$ **then**
14: $\qquad conv = 1$;
15: $\quad$ **else**
16: $\qquad B_0 = B$
17: $\quad$ **end if**
18: **end while**

Fuente: Elaboración propia.

Tal como puede observarse es una iteración que se detiene cuando se satisface una tolerancia del error entre dos estimaciones sucesivas. El código utiliza dos funciones auxiliares para aproximar la función $logit(p) = log\left(\frac{p}{1-p}\right)$, y su inversa:

Lo primero que notamos en el código presentado es el cálculo de inversas generalizadas o pseudoinversa, en las líneas 3 y 12. Posteriormente en la línea 7 se resuelve un sistema de ecuaciones con esta pseudoinversa. Se propone en este trabajo sustituir estos cálculos por otros equivalentes numéricamente que se ejecutan sobre la una factorización de rango completo de las matrices involucradas. La resolución de sistemas de ecuaciones puede hacerse en dos fases de menor costo con los factores de Cholesky.

Finalmente, una vez que en esta aplicación la pseudoinversa de la matriz que se calcula en la línea 12 es requerida al final de este cálculo, se sugiere sustituir por una factorización rectangular de Cholesky de rango completo, a fin de que solo sea necesario invertir el factor.

Los cambios sugeridos fueron implementados sobre el código original, usando la versión rectangular del algoritmo propuesto en la figura 6 para hallar el factor de Cholesky de rango completo. La figura 8 muestra el código en Matlab con las modificaciones sugeridas.

Figura 8.

MCRP modificado

```
function [B, InfF, InvG_InfF, iter] = MCRP_Mod(X,y,n,tol)
conv = 0; iter = 0; M = zeros(size(X));
[mx,nx] = size(X);
y0 = logitMod(y./n);
[Lx] = Rectangular_Cholesky(X);
z = Lx\(X'*y0); B0 = Lx'\z;
while conv == 0
    iter = iter+1;
    XB = X*B0;
    p = Inv_Logit(XB);
    mu = n.*p;
    w = mu.*(1-p);
    InfF=X'*(w.*X);
    [LInfF] = Rectangular_Cholesky(InfF);
    if rank(LInfF) < nx
        conv = 1;
    else
        invL = inv(LInfF,tol);
        InvG_InfF = invL'*invL*InfF;
        B = B0 + InvG_InfF*X'*(y-mu);
        if norm(B-B0) < tol
            conv = 1;
        else
            B0 = B;
        end
    end
end
end
```

Fuente: Elaboración propia.

Como puede observarse, en la versión propuesta se sustituye el cálculo de las pseudoinversas en las líneas 3 y 12 del código original. También se sustituye el producto con la pseudoinversa de la línea 7 del código original, por la resolución de sistemas triangulares equivalentes con el factor **L** y su traspuesta. Por último, la pseudoinversa que se calcula en la línea 12 del código original, que es necesaria para un futuro cálculo en esta aplicación, se ha sustituido por el cálculo de la inversa del factor de Cholesky.

Se llevó a cabo una experimentación numérica tomando matrices sintéticas **X** de $n \times m$ variando el número de filas entre **n=500** y **n**=**30.000**, y manteniendo **m =70**. Se seleccionó el vector **y** de respuestas aleatoriamente en cada experimento. Los resultados comparativos entre la estimación hallada por el código original y la hallada por el código modificado, así como también los tiempos de ejecución medidos se muestran en la figura 9.

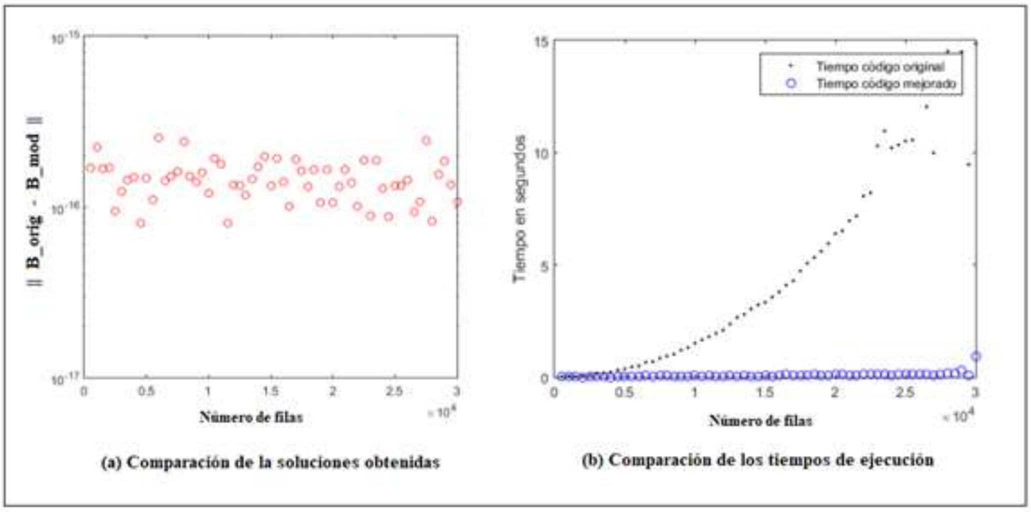
Existen otras propuestas efectivas de carácter general para acelerar el cálculo de estas factorizaciones, ver por ejemplo el trabajo de Camero (2018). También pueden encontrarse eficientes implementaciones en librerías desarrolladas en Matlab, Fortran, C, y C++, tales como las presentadas por Molt (2021) y Gustavson (2018). Sin embargo, para aplicaciones específicas, como la que se estudia, la implementación propuesta saca provecho de la estructura del problema y las características de la solución buscada.

Resultados

Tal como se puede observar en la figura 8, en términos de precisión en la soluciones obtenidas, tanto el código original como el modificado generan numéricamente la misma respuesta en la estimación de los parámetros, en tanto que la versión modificada reduce significativamente el tiempo de ejecución. Estos resultados son alentadores y se espera que puedan corroborarse con datos experimentales reales.

Figura 9.

Comparación de resultados entre el código original y el código modificado



Fuente: Elaboración propia.

Conclusiones

- Hemos presentado en este trabajo una propuesta algorítmica para la resolución de los mínimos cuadrados iterativos reponderados, usando una factorización de rango completo de Cholesky cuando la matriz es rectangular o deficiente en rango. Se resaltan las ventajas de su aplicación en algoritmos estadísticos que requieren la resolución de sistemas de ecuaciones normales donde la matriz de los coeficientes es rectangular o posiblemente con columnas dependientes; situación que se presenta con frecuencia en análisis de datos categóricos.

- En particular, como un estudio de caso, se analizó la factorización de Cholesky de rango completo en una aplicación de los mínimos cuadrados iterativos reponderados para la estimación estadística máximo verosímil de parámetros en un modelo logit binomial. Los resultados sugieren que puede mantenerse la precisión en la solución obtenida al mismo tiempo que se reduce considerablemente el tiempo de cómputo.
- Tomando en cuenta que el algoritmo propuesto está basado mayormente en operaciones con matrices y vectores, los cuales son altamente paralelizables, el próximo paso en esta dirección será la evaluación numérico-computacional de la implementación de esta descomposición en máquinas de alto rendimiento o con múltiples unidades de procesamiento.

Referencias bibliográficas

Bjorck, A. (1996). Numerical Method for Least Squares Problems, Society for Industrial and Applied Mathematics; 1a. Edición.

Camero, C. (2018). Simple, Fast and Practicable Algorithms for Cholesky, LU and QR Decomposition using fast rectangular matrix multiplications, Department of Computer Science and Electronics, Universidad de Cantabria, UNICAN, Spain.

Cantó, R., Ricarte B., Urbano. A.M. (2009). Full rank factorization and Flanders Theorem, Electronic Journal of Linear Algebra, vol. 18, 352-363.

Cantó, R., Peláez M., Urbano. A.M. (2015). Full rank Cholesky factorization for rank deficient matrices. Applied Mathematics Letters, 40, 17-22.

Chen, J. (1993). Iterative Weighted Least Square Estimators, The Annals of Statistics, 21(2).

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. J. Amer. Statist. Assoc. Vol. 82, No. 400, pp. 1079-1091.

Datta, B.N (2010). Numerical Linear Algebra and Applications. Society for Industrial and Applied Mathematics; 2a. Edición.

Forthofer R.N., Lehnen R.G. (1981). One Response and Two Factor Variables. In: Public Program Analysis. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-6683-6_4.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. Ann. Statist., 6, 1149 - 1158.

Gustavson, F., Wasniewski, J., Dongarra, J., Langou, J. (2009). Rectangular Dull Packed Format for Cholesky's Algorithm: Factorization, Solution and Inversion ACM Transaction on Mathematical Software.

Molt, J. (2021). Cholesky decomposition. Recuperado August 6, 2021 de MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/71304-cholesky-decomposition.

Ponsot, E. (2011). Estudio de la agregación de niveles de los factores en el modelo logit binomial. Tesis Doctoral. Instituto de Estadística Aplicada, Universidad de los Andes, Venezuela.

Torkzi, K. (2011) Aproximaciones por mínimos cuadrados ponderados en el entorno de multirresolución por valores puntuales. Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, España.

Trefethen, L. N. & Bau III, D. (1997). Numerical linear algebra. Siam, Philadelphia.

Urbano, A.M., Cantó, R., Ricarte, B., (2011) Factorización de rango completo y aplicaciones XXI Congreso de Ecuaciones Diferenciales y Aplicaciones/XI Congreso de Matemática Aplicada, Cedya (2009), Ciudad Real, España, 21-25 septiembre, pp 1-8.

PARA CITAR EL ARTÍCULO INDEXADO.

Castillo Marrero, Z. N., Ponsot Balaguer, E. A., José Camacho, F., & León Sánchez, M. V. . (2021). Usando la Factorización de Rango Completo de Cholesky en la Solución de los Mínimos Cuadrados Ponderados . AlfaPublicaciones, 3(3.1), 83-97. https://doi.org/10.33262/ap.v3i3.1.79