Facultad de Ciencias Naturales y Exactas
Universidad del Valle

Structured BFGS for the Estimation of the Maximum Likelihood

Favián Arenas
Universidad del Cauca

Héctor Jairo Martínez
Universidad del Valle

Rosana Pérez
Universidad del Cauca

Received: September 2, 2016 Accepted: December 13, 2016

Pag. 39-54

Abstract

Given the special structure of the Hessian matrix of the log-likelihood function which is parallel to that found in nonlinear least-squares problems, we introduce the structured BFGS secant method for the maximum-likelihood estimation and for the development or the local and super-linear convergence theory for the algorithm, following the lines of [12, 13] and the theory about maximum-likelihood estimation given in [10]. We present the results of some numerical experiments which show a good performance of our algorithm.

Keywords: maximum likelihood estimation, likelihood function, structured secant method, nonlinear least-squares problem, super-linear convergence.

Método BFGS estructurado para la estimulación de máxima verosimilitud

Resumen

Teniendo en cuenta la estructura especial de la matriz hessiana del logaritmo de la función de verosimilitud análoga a la estructura encontrada en problemas de mínimos cuadrados no lineales, se propone el método BFGS estructurado para el problema de la estimación de máxima verosimilitud y se desarrolla su teoría de convergencia local y $q-$superlineal siguiendo los lineamientos generales de la teoría de convergencia desarrollada en [12, 13] para métodos secante estructurados y la teoría sobre estimación de la máxima verosimilitud dada en [10]. Además, se realizaron pruebas numéricas preliminares que muestran el buen comportamiento local del método propuesto.

Palabras clave: Estimación de máxima verosimilitud, función de verosimilitud, método secante estructurado, mínimos cuadrados no lineales, convergencia super lineal.

1 Introducción

Uno de los problemas que se presenta con más frecuencia en el campo estadístico es el de la estimación de parámetros de una población con función de densidad conocida; esto es, dada una muestra aleatoria $X_1, X_2, \ldots, X_n$, de una variable aleatoria X con función de densidad $f(x;\theta_*)$, estimar el parámetro $\theta_* \in \mathbb{R}^k$ [15].

Un procedimiento que permite encontrar un estimador del parámetro θ_* es el de *máxima verosimilitud*. En este método se considera la función de verosimilitud L, la cual, bajo el supuesto de que la muestra genere variables aleatorias independientes, está definida por:

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta). \tag{1}$$

Si X es discreta, $L(x_1, x_2, \ldots, x_n; \theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n)$; similarmente, si X es absolutamente continua, $L(x_1, x_2, \ldots, x_n; \theta)$ representa la función de densidad de probabilidad conjunta de $(X_1, X_2, \ldots, X_n)$ dado el parámetro θ.

El estimador de máxima verosimilitud (EMV) de θ_*, denotado $\hat{\theta}$, basado en una muestra aleatoria $X_1, X_2, \ldots, X_n$ se define como el valor de θ que maximiza la función de verosimilitud $L(\theta) = L(x_1, x_2, \ldots, x_n; \theta)$ dada en (1).

Con el fin de encontrar el EMV, se debe determinar el valor máximo de la función L, el cual, también es el valor máximo de la función $l(\theta) = \ln L(\theta)$ y, debido al carácter multiplicativo que tienen las probabilidades conjuntas para variables aleatorias independientes (ver (1)), en algunos casos puede ser más fácil trabajar con $l(\theta)$ en lugar de hacerlo con $L(\theta)$.

Como en todo problema de optimización, la maximización de $L(\theta)$, en algunos casos, presenta considerables dificultades numéricas, matemáticas y estadísticas. Estas últimas se presentan debido a que el sesgo y la varianza de los estimadores pueden ser indeseablemente grandes para tamaños de muestra prácticos.

La función $l(\theta)$ y por lo tanto $L(\theta)$ puede tener múltiples maximizadores tanto locales como globales o no tenerlos; en particular, $l(\theta)$ puede tener más de un maximizador global o no tenerlo.

La estimación de la máxima verosimilitud, generalmente, es la aproximación preferida para el problema de *estimación de mezcla finita*, en el cual, la función de densidad $f(x;\theta)$ es una *suma convexa de funciones de densidad* y se puede interpretar como la función asociada a una población estadística, la cual es una mezcla de m poblaciones con funciones de densidad asociadas $\{f_i\}_{i=1,\ldots,n}$ y proporciones de mezcla $\{\alpha_i\}_{i=1,\ldots,m}$.

Este problema se puede plantear de la siguiente forma: [17]. Dada una muestra aleatoria $X_1, \ldots, X_n$ de una variable aleatoria X con función de densidad dada por:

$$f(x;\theta_*) = \sum_{i=1}^{m} \alpha_{*_i} f_i(x;\theta_{*_i})$$

estimar el parámetro $\theta_* = (\alpha_{*_1}, \ldots, \alpha_{*_m}, \theta_{*_1}, \ldots, \theta_{*_m})$, donde los α_{*_i} son no negativos tales que $\sum_{i=1}^{m} \alpha_{*_i} = 1$ y cada función componente $f_i(x;\theta_{*_i})$ es también una función de densidad, con frecuencia, perteneciente a una de las familias paramétricas conocidas.

Esta suma convexa de funciones de densidad aparece con frecuencia en el estudio de poblaciones biológicas y procesos económicos. Otras aplicaciones en las cuales estas sumas convexas de densidades juegan un papel central son: la interpretación y clasificación de datos de satélites percibidos remotamente, la determinación de mezclas químicas vía espectroscopía de absorción, diagnósticos y pronósticos médicos, etc. [11].

Debe destacarse que, exceptuando casos en los que $f(x;\theta_*)$ es miembro de una de las bien conocidas familias paramétricas, los estimadores de máxima verosimilitud no se pueden obtener analíticamente y, por ello, se deben aproximar numéricamente.

En [10], el autor observa que $l(\theta)$ tiene una estructura similar a la de la suma de los cuadrados de los residuos en problemas de mínimos cuadrados no lineales [6], bosqueja y da resultados de convergencia local para métodos cuasi-Newton especiales sugeridos por la estructura presente en la matriz hessiana de $l(\theta)$, los cuales son análogos a métodos conocidos para mínimos cuadrados no lineales (método de Gauss-Newton, por ejemplo). A pesar del éxito de los resultados numéricos dados, el método secante BFGS estructurado [6, 7] no se considera para este problema. Teniendo en cuenta que la actualización BFGS es la mejor actualización a la matriz hessiana que se conoce para problemas de optimización en general, en [16], se considera dicha actualización y se propone por primera vez el método BFGS estructurado para el problema de estimación de máxima verosimilitud, pero a pesar de los resultados teóricos obtenidos, este trabajo no se publicó porque carecía de experimentación numérica sobre el método propuesto.

En el presente artículo, presentamos el método BFGS estructurado para la estimación de máxima verosimilitud, así como su teoría local de convergencia. Además, realizamos pruebas numéricas preliminares que muestran el buen comportamiento local del método.

Organizamos la presentación de este documento de la siguiente forma: en la sección 2, se presenta la estructura especial del logaritmo de la función de verosimilitud. En la sección 3, se hace una descripción general del problema de los mínimos cuadrados no lineales así como de los métodos cuasi-Newton para resolverlo, principalmente el método de Gauss-Newton y los métodos secante

estructurados. En la sección 4, teniendo en cuenta la analogía del problema de los mínimos cuadrados no lineales con el problema de máxima verosimilitud, se extiende la teoría cuasi-Newton a la estimación máximo verosímil; esto incluye una descripción del método de *scoring* y de los métodos de actualización secante propuestos en [10]. En la sección 5, se presenta el método secante estructurado para estimación de máxima verosimilitud y su teoría de convergencia local. En la sección 6, se presentan algunas pruebas numéricas preliminares que nos permiten apreciar el comportamiento numérico del método propuesto. Finalmente, en la sección 7, se hacen algunos comentarios finales y propuestas de trabajos futuros sobre el tema.

2 Estructura especial del logaritmo de la función de verosimilitud

Se considera a continuación la estructura especial de la función l definida por:

$$l(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(f(x_i;\theta)\right), \tag{2}$$

donde el parámetro $\theta = (\theta_1\ \theta_2\ \ldots\ \theta_k)^T \in \mathbb{R}^k$, con x_i, $i = 1,\ldots,n$ variables mutuamente independientes y el factor $1/n$ se incluye para simplificar algunos términos que resultan en los cálculos del vector gradiente y la matriz hessiana de l en θ, denotadas por $\nabla_\theta\, l(\theta)$ y $\nabla^2_\theta\, l(\theta)$, respectivamente.

Si f es una función dos veces continuamente diferenciable, se obtienen las siguientes expresiones para $\nabla_\theta\, l(\theta)$ y $\nabla^2_\theta\, l(\theta)$:

$$\nabla_\theta\, l(\theta) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla_\theta\, f(x_i;\theta)}{f(x_i;\theta)} = \frac{1}{n}\, J(\theta)^T \vec{\mathbf{1}} \tag{3}$$

$$\nabla^2_\theta\, l(\theta) \quad = \quad -\frac{1}{n}\, J(\theta)^T J\theta) + \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla^2_\theta\, f(x_i;\theta)}{f(x_i;\theta)} \tag{4}$$

donde

$$J(\theta) = \begin{pmatrix} \frac{\nabla_\theta\, f(x_1;\theta)^T}{f(x_1;\theta)} \\ \vdots \\ \frac{\nabla_\theta\, f(x_n;\theta)^T}{f(x_n;\theta)} \end{pmatrix} \qquad \text{y} \qquad \vec{\mathbf{1}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Luego, la matriz hessiana de l en θ se puede expresar en la forma:

$$\nabla^2_\theta\, l(\theta) = C(\theta) + S(\theta), \tag{5}$$

con

$$C(\theta) = -\frac{1}{n}\, J(\theta)^T J(\theta) \qquad\qquad S(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla^2_\theta\, f(x_i;\theta)}{f(x_i;\theta)}.$$

De esta forma, la estructura especial de la matriz hessiana es clara; la información de primer orden la tiene el término $C(\theta)$, el cual es fácil de calcular una vez se haya calculado $\nabla_\theta\, l(\theta)$, y, además, siempre es semidefinido negativo y en algunas aplicaciones puede ser definido negativo, si el tamaño de la muestra es suficientemente grande [6]. El término $S(\theta)$ contiene la información de segundo orden y es costoso de evaluar.

La estructura especial de la matriz hessiana dada por (4) es análoga a la estructura particular de la matriz hessiana de las sumas residuales de cuadrados en problemas de mínimos cuadrados no lineales [6]. Esta analogía es muy importante, puesto que permite extender toda la teoría existente para el problema de los mínimos cuadrados no lineales al problema de máxima verosimilitud.

3 El problema de los mínimos cuadrados no líneales

El problema de *mínimos cuadrados no lineales* consiste en resolver el siguiente problema de minimización sin restricciones:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{Minimizar} \qquad g(\boldsymbol{x}) = \frac{1}{2}R(\boldsymbol{x})^T R(\boldsymbol{x}) \;=\; \frac{1}{2}\sum_{i=1}^{m} r_i(\boldsymbol{x})^2, \tag{6}$$

donde la función residual $R\colon \mathbb{R}^n \to \mathbb{R}^m$ es no lineal y $r_i(\boldsymbol{x})$ es la $i-$ésima componente de $R(\boldsymbol{x})$.

La estructura particular del problema (6) se observa claramente en las expresiones para el *vector gradiente* y *la matriz hessiana* de g, en $\boldsymbol{x}$. En efecto, después de realizar algunos cálculos algebraicos, se tienen:

$$\nabla g(\boldsymbol{x}) = J(\boldsymbol{x})^T R(\boldsymbol{x}) \qquad \text{y} \qquad \nabla^2 g(\boldsymbol{x}) = J(\boldsymbol{x})^T J(\boldsymbol{x}) + S(\boldsymbol{x}), \tag{7}$$

donde, $J(\boldsymbol{x}) \in \mathbb{R}^{m\times n}$ es la *matriz jacobiana* de R en $\boldsymbol{x}$, dada por

$$J(\boldsymbol{x}) = \begin{pmatrix} \nabla r_1(\boldsymbol{x})^T \\ \vdots \\ \nabla r_m(\boldsymbol{x})^T \end{pmatrix} \qquad \text{y} \qquad S(\boldsymbol{x}) = \sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x}).$$

La matriz $J(\boldsymbol{x})$ contiene solamente información de *primer orden* (primeras derivadas parciales) y $S(\boldsymbol{x})$ contiene información de *segundo orden* (es una combinación lineal de m matrices hessianas). Esta estructura especial es la que se aprovecha en algunos de los métodos usados para resolver el problema (6) y es la razón por la cual no se usan métodos de propósito general para resolver el mismo [6].

Uno de los primeros métodos cuasi-Newton[1] diseñados especialmente para el problema (6) fue el método de *Gauss-Newton*, cuya iteración básica es de la forma

[1] La idea subyacente en los métodos cuasi-Newton es utilizar una aproximación de la matriz hessiana en lugar de la hessiana misma, dado que la evaluación y utilización de esta matriz es en muchos casos imposible o impráctica por el número de operaciones involucrado en su cálculo.

cuasi-Newton [6].

$$\begin{aligned} B_k \boldsymbol{s}_k &= -\nabla g(\boldsymbol{x}_k) \\ \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \boldsymbol{s}_k, \end{aligned} \tag{8}$$

con $B_k = J(\boldsymbol{x}_k)^T J(\boldsymbol{x}_k)$. Es decir, el método de *Gauss-Newton* usa, como aproximación de la matriz $\nabla^2 g(\boldsymbol{x})$, la matriz que contiene solo la información de primer orden en dicha matriz hessiana. Aún cuando el método de Gauss-Newton presenta algunos problemas, él es la base de algunos métodos prácticos y exitosos para problemas de mínimos cuadrados no lineales [6]. Entre las ventajas del método, están su convergencia local y $q-$cuadrática en problemas de residuo cero; su convergencia local y $q-$lineal rápida en problemas casi lineales y problemas de residuos pequeños. Además, resuelve problemas de mínimos cuadrados lineales en una iteración.

En resumen, el método de *Gauss-Newton* es exitoso en situaciones en las cuales se puede ignorar el término que contiene la información de segundo orden en $\nabla^2 g(\boldsymbol{x})$, pero pierde su efectividad cuando este término no es insignificante comparado con el término que contiene la información de primer orden.

Una alternativa para evitar las dificultades del *método Gauss-Newton* cuando la matriz Jacobiana no tiene *rango completo* o está *mal condicionada* consiste en usar la iteración (8) con $B_k = J(\boldsymbol{x}_k)^T J(\boldsymbol{x}_k) + \mu_k I$. Así, en cada iteración, la matriz $S(\boldsymbol{x}_k)$ se aproxima por un múltiplo de la matriz identidad. Cuando $\mu_k = 0$, se tiene el método de *Gauss-Newton*. Este método del cual existen varias versiones que varían según la estrategia para escoger μ_k, fue propuesto por *Levenberg* en (1944) y *Marquardt* en (1963) y se le conoce como el método de *Levenberg-Marquard* [6]. Las propiedades de convergencia de este método son similares a las del método de *Gauss-Newton*, sin embargo, algunas implementaciones de su algoritmo, como la dada por *Moré* y usada en MINPACK[2] han resultado muy exitosas en la práctica.

Un método cuasi-Newton, para el problema (6), que considera la estructura de la matriz hessiana de g en $\boldsymbol{x}$, es el método cuasi-Newton (o secante) estructurado [7, 12, 13]. En general, la idea básica de estos métodos es utilizar iterativamente *una aproximación estructurada* de la matriz hessiana, es decir, una aproximación de la forma

$$B_k = C(\boldsymbol{x}_k) + A_k, \tag{9}$$

donde A_k es una aproximación de $S(\boldsymbol{x}_k)$.

Una aproximación *secante* estructurada de la matriz hessiana es una actualización de la matriz B_k de la forma

$$B_{k+1} = C(\boldsymbol{x}_{k+1}) + A_{k+1}$$

[2]MINPACK es una biblioteca de subrutinas en FORTRAN altamente portátil, robusta y confiable, utilizada para resolver sistemas de ecuaciones no lineales o *problemas de mínimos cuadrados lineales* y *no lineales* [14].

donde $A_{k+1} = A_k + \Delta(\boldsymbol{s}_k, \boldsymbol{y}_k^{\#}, A_k, \boldsymbol{v}_k) \equiv A_k + \Delta_k$, con

$$\Delta_k = \frac{(\boldsymbol{y}_k^{\#} - A_k \boldsymbol{s}_k)\boldsymbol{v}_k^T + \boldsymbol{v}_k(\boldsymbol{y}_k^{\#} - A_k \boldsymbol{s}_k)^T}{\boldsymbol{v}_k^T \boldsymbol{s}_k} - \frac{\boldsymbol{s}_k^T(\boldsymbol{y}_k^{\#} - A_k \boldsymbol{s}_k)\boldsymbol{v}_k \boldsymbol{v}_k^T}{(\boldsymbol{v}_k^T \boldsymbol{s}_k)^2}, \qquad (10)$$

$\boldsymbol{v}_k \equiv \boldsymbol{v}_k(\boldsymbol{s}_k, \boldsymbol{y}_k, B_k)$, llamado *la escala* y, $\boldsymbol{y}_k^{\#}$ y $\boldsymbol{y}_k$ son aproximaciones a $S(\boldsymbol{x}_{k+1})\boldsymbol{s}_k$ y $\nabla^2 g(\boldsymbol{x}_{k+1})s_k$, respectivamente. La matriz A_{k+1} satisface *la ecuación secante* (*estructurada*)

$$A_{k+1}\, \boldsymbol{s}_k \;=\; \boldsymbol{y}_k^{\#}. \qquad (11)$$

En particular, para el problema (6), la *aproximación secante estructurada* de la matriz hessiana es

$$B_k \;=\; J(\boldsymbol{x}_k)^T J(\boldsymbol{x}_k) \;+\; A_k, \qquad (12)$$

donde la matriz A_k que se adiciona a la aproximación *Gauss-Newton* de la matriz hessiana se obtiene por actualización. Se espera que este método disfrute de convergencia local superlineal con un costo menor por iteración frente al que resultaría si B_k fuera la matriz hessiana completa de g.

La fórmula de actualización para A_k, que originalmente se utilizó en el código NL2SOL (Non Linear Square Solver) [5], fue:

$$A_{k+1} = A_k + \Delta(\boldsymbol{s}_k, \boldsymbol{y}_k^{\#}, A_k, \boldsymbol{v}_k) \equiv A_k + \Delta_k, \qquad (13)$$

donde Δ_k está dado por (10) con $\boldsymbol{v}_k = \boldsymbol{y}_k = \nabla g(\boldsymbol{x}_{k+1}) - \nabla g(\boldsymbol{x}_k)$ y $\boldsymbol{y}_k^{\#} = (J(\boldsymbol{x}_{k+1}) - J(\boldsymbol{x}_k))^T R(\boldsymbol{x}_{k+1})$. En 1981, Dennis y Walker [8] demostraron que bajo condiciones generales sobre A_0 y $g(\boldsymbol{x})$, la iteración (8) con A_k, actualizada de esta forma, converge local y q superlinealmente al minimizador $\boldsymbol{x}_*$. Actualmente, el código NL2SOL utiliza la fórmula de actualización BFGS (Broyden, Fletcher Golfard y Shanno) estructurada en la cual

$$\boldsymbol{v}_k = \boldsymbol{y}_k + \left(\frac{\boldsymbol{y}_k^T \boldsymbol{s}_k}{\boldsymbol{s}_k^T B_k \boldsymbol{s}_k}\right)^{1/2} B_k \boldsymbol{s}_k, \qquad (14)$$

y $\boldsymbol{y}_k^{\#} = (J(\boldsymbol{x}_{k+1}) - J(\boldsymbol{x}_k))^T R(\boldsymbol{x}_{k+1})$. debido a su conveniencia numérica, al tiempo que se conserva la convergencia local y super lineal del minimizador $\boldsymbol{x}_*$ [7, 12, 13].

4 El problema de máxima verosimilitud

Debido a la analogía en la estructura de $\nabla^2 g(\boldsymbol{x})$ en el problema de mínimos cuadrados no lineales y la estructura de $\nabla^2 l(\theta)$ en el problema de máxima verosimilitud, dadas en (7) y (5), respectivamente, se han considerado para máxima verosimilitud, métodos análogos al de Gauss-Newton y a los métodos utilizados en el algoritmo NL2SOL [5, 6, 10] para problemas de mínimos cuadrados no lineales.

4.1 Método de *acoring*

Este método se considera análogo, en el problema de máxima verosimilitud, al método de Gauss-Newton. El método se le atribuye a Fischer [9] y su forma

clásica está dada por la forma cuasi-Newton

$$\begin{aligned} B_k \boldsymbol{s}_k &= \nabla_\theta \, l(\theta_k) \\ \theta_{k+1} &= \theta_k + \boldsymbol{s}_k, \end{aligned} \tag{15}$$

donde B_k, la aproximación a $\nabla^2 l(\theta)$, es la *matriz información de Fischer* definida por

$$I(\theta) = \int_{\mathbb{R}^n} [\nabla_\theta \, \ln \, f(\boldsymbol{x};\theta)][\nabla_\theta \, \ln \, f(\boldsymbol{x};\theta)]^T \, f(\boldsymbol{x};\theta) \, d\mu,$$

siendo μ una medida sobre $\mathbb{R}^n$ apropiada para $f(x;\theta)$.

El uso de la *matriz información de Fisher* como aproximación a $\nabla^2_\theta \, l(\theta)$ presenta inconvenientes en muchas aplicaciones debido a que su evaluación es muy costosa numéricamente, y en algunos casos, impráctica. Una alternativa efectiva para esta situación es considerar como aproximación a $\nabla^2_\theta \, l(\theta)$, la matriz que contiene la información de primer orden en (5), al igual que en el método de Gauss-Newton, lo cual constituye el método de *scoring modificado*, cuya iteración básica es dada por (15) con $B_k = -\frac{1}{n} J(\theta_k)^T J(\theta_k)$.

En [10], se dan condiciones generales bajo las cuales, el *método de scoring modificado* converge local y $q-$linealmente en un maximizador de l. También, se presenta un análisis detallado de su convergencia, análogo al realizado en [6] para el método de Gauss-Newton.

El *método de scoring modificado* trabaja muy bien cuando el término $S(\theta)$ en $\nabla^2 l(\theta)$ es relativamente pequeño comparado con el término $C(\theta)$. Al igual que en el método de Gauss-Newton, el *método de scoring modificado* pierde su efectividad cuando el término $S(\theta)$ no es despreciable comparado con $C(\theta)$. Una alternativa a esta situación se considera a continuación.

4.2 Método de *Levenberg- Marquard*

Este método, descrito en la sección anterior en su versión estructurada, también es usado en el contexto de la estimación de máxima verosimilitud, en la cual su iteración básica tiene la forma cuasi-Newton (15) con $B_k = \nabla^2 l(\theta_k) + \mu_k \, I$. La elección de μ_k se realiza de tal forma que, inicialmente domine el método del gradiente (μ_k grande) y, a medida que μ_k tienda a cero, domine el método de Newton. Es decir, el método se comporta como un híbrido entre los métodos del gradiente y el de Newton. En general, para la aplicación del método, se usa un paquete estadístico que tenga incorporado este proceso iterativo [1].

4.3 Métodos de actualización secante

Estos métodos se consideran análogos, en el problema de estimación de máxima verosimilitud, a los métodos tipo NL2SOL utilizados para resolver problemas de mínimos cuadrados no lineales en situaciones en las cuales el término que contiene la información de segundo orden no es insignificante comparado con el término que contiene la información de primer orden y, por lo tanto, no pueden ser ignorados

fácilmente.

Estos métodos, llamados en [10], métodos de actualización de estimadores de máxima verosimilitud, tienen la forma cuasi-Newton estructurada (15), con, $B_k = C(\theta_k) + A_k$, donde A_k es una aproximación a $S(\theta_k)$. La actualización de A_k se hace de tal forma que satisfaga la *ecuación secante* (11).

En [10], se considera la fórmula de actualización $A_{k+1} = A_k + \Delta_k$, donde Δ_k está dado por (10), con $\boldsymbol{v}_k = \boldsymbol{y}_k = \nabla_\theta\, l(\theta_{k+1}) - \nabla_\theta\, l(\theta_k)$ y $\boldsymbol{y}_k^{\#}$ dado por

$$\boldsymbol{y}_1^{(k)} = \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla_\theta\, f(x_k;\theta_{k+1}) - \nabla_\theta\, f(x_k;\theta_k)}{f(x_k;\theta_{k+1})},$$

aunque presentan otras fórmulas que se pueden usar; a saber:

$$\begin{aligned} \boldsymbol{y}_2^{(k)} &= \boldsymbol{y}_k + \frac{1}{n} J(\theta_{k+1})^T J(\theta_{k+1})\boldsymbol{s}_k. \\ \boldsymbol{y}_3^{(k)} &= \boldsymbol{y}_k + \frac{1}{n} J(\theta_k)^T J(\theta_k)\boldsymbol{s}_k. \end{aligned}$$

Además, en [10], se demuestra que el método converge local y $q-$superlinealmente en un maximizador local de l.

5 El método BFGS estructurado en la estimación de máxima verosimilitud

En esta sección, se propone un método de actualización de estimadores de máxima verosimilitud cuya forma de actualización usa la fórmula BFGS estructurada (13), donde $\Delta(\boldsymbol{s}_k, \boldsymbol{y}_k^{\#}, A_k, \boldsymbol{v}_k)$ está dada por (10), con $\boldsymbol{v}_k$ dada por (14). Además, se presenta el desarrollo de la teoría de convergencia local y $q-$superlineal [16].

Una iteración básica del método BFGS estructurado para el problema

$$\underset{\theta \in \mathbb{R}^k}{Maximizar} \quad l(\theta), \tag{16}$$

donde l es la función definida en (2), está dada por

$$\begin{aligned} (C(\theta_k) + A_k)\, \boldsymbol{s}_k &= \nabla_\theta\, l(\theta_k) \\ \theta_{k+1} &= \boldsymbol{x}_k + \boldsymbol{s}_k, \end{aligned} \tag{17}$$

donde la matriz A_k se actualiza de acuerdo a (13) y (14). Los vectores $\boldsymbol{y}_k^{\#}$ y $\boldsymbol{y}_k$ están dados por:

$$\boldsymbol{y}_k^{\#} = \frac{1}{n} J(\theta_{k+1})^T \hat{\mathbf{1}} - \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla f(x_i;\theta_k)}{f(x_i;\theta_{k+1})}, \tag{18}$$

$$\boldsymbol{y}_k \;=\; \boldsymbol{y}_k^{\#} \;-\; \frac{1}{n}\, J(\theta_{k+1})^T J(\theta_{k+1})\boldsymbol{s}_k \tag{19}$$

donde $J(\theta_{k+1})$ está dado por (5). La escogencia de $\boldsymbol{y}_k$ es análoga a la sugerida en [2] para el problema de los mínimos cuadrados no lineales, para calcular el vector $\boldsymbol{v}$, introduciendo, de esta forma, la estructura del problema en la escala de la fórmula de actualización.

La escogencia de $y_k^{\#}$ es análoga a la dada, independientemente en [4, 3] para el problema de los mínimos cuadrados no lineales y es la escogencia que se usa actualmente en el código NL2SOL.

5.1 Teoría de convergencia

Dennis y Walker [8] desarrollaron una teoría de convergencia para métodos secantes estructurados la cual se puede utilizar para los métodos PSB (Powell-Simetric Broyden) y DFP (Davidon, Fletcher, Powell), pero no para el método BFGS estructurado. Antes de estos resultados, muy pocos desarrollos teóricos sobre métodos secante estructurados fueron dados, aunque muchos algoritmos que usaban estas ideas fueron propuestos y experimentados numéricamente.

En [7, 12, 13], se demuestra la convergencia local y q-superlineal del método secante BFGS estructurado, llenando así el vacío existente sobre la teoría de convergencia de los métodos secante estructurados. Esta teoría será la que se empleará para demostrar la convergencia local y q-superlineal del método BFGS estructurado propuesto para el problema de máxima verosimilitud.

Para el desarrollo de la teoría de convergencia local del método propuesto, es conveniente escribir algunas expresiones dadas anteriormente en forma tal que permita realizar algunos cálculos de manera sencilla. Es fácil demostrar que las expresiones para el gradiente y la matriz hessiana de l, dadas por (3) y (4), así como la expresión de $\boldsymbol{y}_k^{\#}$ dada por (18) son equivalentes a las siguientes expresiones:

$$\nabla_\theta l(\theta) = \frac{1}{n}\, \hat{J}(\theta)^T R(\theta). \tag{20}$$

$$\nabla^2_\theta l(\theta) = -\frac{1}{n}\, \hat{J}(\theta)^T \bar{J}(\theta) \;+\; S(\theta). \tag{21}$$

$$y_k^{\#} = \frac{1}{n}\, \hat{J}(\theta_{k+1})^T R(\theta_{k+1}) - \frac{1}{n}\, \hat{J}(\theta_k)^T R(\theta_{k+1}), \tag{22}$$

donde $\hat{J}(\theta)^T \in \mathbb{R}^{n\times n}$, $\bar{J}(\theta) \in \mathbb{R}^{n\times n}$ y $R(\theta) \in \mathbb{R}^{n\times 1}$ están dadas por:

$$\hat{J}(\theta)^T = \begin{pmatrix} \nabla_\theta\, f(x_1;\theta)^T \\ \vdots \\ \nabla_\theta\, f(x_n;\theta)^T \end{pmatrix} \quad R(\theta) \;=\; \begin{pmatrix} \frac{1}{f(x_1;\theta)} \\ \vdots \\ \frac{1}{f(x_n;\theta)} \end{pmatrix} \quad \bar{J}(\theta) = \begin{pmatrix} \frac{\nabla_\theta\, f(x_1;\theta)^T}{f(x_1;\theta)^2} \\ \vdots \\ \frac{\nabla_\theta\, f(x_n;\theta)^T}{f(x_n;\theta)^2} \end{pmatrix}.$$

Las hipótesis generales, bajo las cuales se desarrollará la teoría de convergencia, son las siguientes:

H1. El problema (16) tiene una solución θ_*.

H2. La función l es dos veces continuamente diferenciable y $\hat{J}$, $\bar{J}$ y $\nabla^2 l$ son localmente lipschitz continuas en θ_*, es decir, existen constantes positivas β_1, β_2, β_3 y ϵ tales que

$$\begin{aligned} \|\hat{J}(\theta) - \hat{J}(\theta_*)\| &\leq \beta_1 \|\theta - \theta_*\| \\ \|\bar{J}(\theta) - \bar{J}(\theta_*)\| &\leq \beta_2 \|\theta - \theta_*\| \\ \|\nabla^2_\theta l(\theta) - \nabla^2_\theta l(\theta_*)\| &\leq \beta_3 \|\theta - \theta_*\|, \end{aligned}$$

para todo $\theta \in D = \{\theta : \|\theta - \theta_*\| \leq \epsilon\}$.

H3. La matriz $\nabla^2_\theta\, l(\theta_*)$ es no singular.

Además, en el desarrollo de la teoría de convergencia local, se hará uso de un resultado bastante útil del cálculo en varias variables [6] y del teorema de convergencia del método secante BFGS estructurado dado en [7, 12, 13]. Por completez los dos resultados aparecen a continuación:

Lema 1 *[6]: Sean $F : \mathbb{R}^n \to \mathbb{R}^m$ continuamente diferenciable en un conjunto convexo $D \subset \mathbb{R}^n$, $\boldsymbol{x} \in D$ y J, la matriz jacobiana de F; lipschitz sea continua en $\boldsymbol{x}$, con constante γ, usando una norma matricial y su correspondiente norma matricial inducida, se tiene que para cualquier $\boldsymbol{u}$ y $\boldsymbol{v}$ en D, se cumple que:*

$$\|F(\boldsymbol{v}) - F(\boldsymbol{u}) - J(\boldsymbol{x})(\boldsymbol{v} - \boldsymbol{u})\| \leq \gamma \frac{\|\boldsymbol{v} - \boldsymbol{x}\| + \|\boldsymbol{u} - \boldsymbol{x}\|}{2} \|\boldsymbol{v} - \boldsymbol{u}\|. \qquad (23)$$

Teorema 1 *[7, 12, 13] Si las hipótesis* **H1** *a* **H3** *se verifican, $\boldsymbol{s} = \boldsymbol{x}_1 - \boldsymbol{x}_2$ y $\boldsymbol{y}^\#$ es una aproximación a $S(\boldsymbol{x}_*)\boldsymbol{s}$, la cual satisface*

$$\|\boldsymbol{y}^\# - S(\boldsymbol{x}_*)\boldsymbol{s}\| \leq K_2\, \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2)\|\boldsymbol{s}\|$$

para $\boldsymbol{x}_1$ y $\boldsymbol{x}_2$ en D y alguna constante $K_2 \geq 0$, entonces existen constantes positivas ϵ y δ tales que, si

$$\|\boldsymbol{x}_0 - \boldsymbol{x}_*\| < \epsilon \quad y \quad \|A_0 - S(\boldsymbol{x}_*)\| < \delta,$$

entonces, la sucesión de iteraciones $\{x_k\}$ generada por el método BFGS estructurado para un problema de minimización sin restricciones converge $q-$superlinealmente en $\boldsymbol{x}_*$.

El siguiente lema es fundamental para el resultado de convergencia local y $q-$superlineal del método:

Lema 2 *Si las hipótesis* **H1** *a* **H3** *se verifican, entonces existe una constante positiva K, tal que:*

$$\|\boldsymbol{y}_k^\# - S(\boldsymbol{x}_*)\boldsymbol{s}_k\| \leq K\, \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2)\|\boldsymbol{s}_k\|$$

donde $\boldsymbol{y}^\#$ está dado por (22), $\{\theta_k\} \subset D$, $\boldsymbol{s}_k = \theta_{k+1} - \theta_k$ y

$$\sigma(\theta_k, \theta_{k+1}) = \text{máx}\{\|\theta_k - \theta_*\|, \|\theta_{k+1} - \theta_*\|\}.$$

Demostración. Usando la definición de $\boldsymbol{y}_k^{\#}$, tenemos que

$$\boldsymbol{y}_k^{\#} - S(\theta_*)\boldsymbol{s}_k \;=\; \frac{1}{n}\,\hat{J}(\theta_{k+1})^T R(\theta_{k+1}) - \frac{1}{n}\,\hat{J}(\theta_k)^T R(\theta_{k+1}) - S(\theta_*)\boldsymbol{s}_k.$$

Sumando y restando algunos términos y usando las expresiones para el gradiente de l en θ, dada por (20), se tiene:

$$\begin{aligned}
\boldsymbol{y}_k^{\#} - S(\theta_*)\boldsymbol{s}_k &= \nabla_\theta l(\theta_{k+1}) - \nabla_\theta l(\theta_k) - \frac{1}{n}\hat{J}(\theta_k)^T R(\theta_{k+1}) + \frac{1}{n}\,\hat{J}(\theta_k)^T R(\theta_k) \\
&\quad -\frac{1}{n}\left[\hat{J}(\theta_k)^T \bar{J}(\theta_*)\boldsymbol{s}_k - \hat{J}(\theta_k)^T \bar{J}(\theta_*)\boldsymbol{s}_k + \hat{J}(\theta_*)^T \bar{J}(\theta_*)\boldsymbol{s}_k\right] \\
&\quad -\nabla_\theta^2 l(\theta_*)\boldsymbol{s}_k, \\
&= \nabla_\theta l(\theta_{k+1}) - \nabla_\theta l(\theta_k) - \nabla_\theta^2 l(\theta_*)\boldsymbol{s}_k \\
&\quad +\frac{1}{n}\hat{J}(\theta_k)^T \left[R(\theta_k) - R(\theta_{k+1}) - \bar{J}(\theta_*)^T \boldsymbol{s}_k\right] \\
&\quad +\frac{1}{n}\left[\hat{J}(\theta_k) - \hat{J}(\theta_*)\right]^T \bar{J}(\theta_*)^T \boldsymbol{s}_k. \qquad (24)
\end{aligned}$$

Usando la hipótesis **H2** y el **Lema 1**, se tienen las dos desigualdades siguientes:

$$\begin{aligned}
\|\nabla_\theta l(\theta_{k+1}) - \nabla_\theta l(\theta_k) - \nabla_\theta^2 l(\theta_*)\boldsymbol{s}_k\| &\leq \beta_3\,\sigma(\theta_k,\theta_{k+1})\|\boldsymbol{s}_k\|. \\
\|R(\theta_k) - R(\theta_{k+1}) - \bar{J}(\theta_*)^T \boldsymbol{s}_k\| &\leq \beta_2\,\sigma(\theta_k,\theta_{k+1})\|\boldsymbol{s}_k\|.
\end{aligned}$$

Por lo tanto,

$$\begin{aligned}
\|\boldsymbol{y}^{\#} - S(\theta_*)\boldsymbol{s}_k\| &\leq \beta_3\,\sigma(\theta_k,\theta_{k+1})\,\|\boldsymbol{s}_k\| \,+\, \frac{1}{n}\|\hat{J}(\theta_k)\|\,\beta_2\,\sigma(\theta_k,\theta_{k+1})\|\boldsymbol{s}_k\| \\
&\quad +\, \frac{1}{n}\beta_1\,\sigma(\theta_k,\theta_{k+1})\,\|\bar{J}(\theta_*)\|\|\boldsymbol{s}_k\|. \qquad (25)
\end{aligned}$$

Finalmente, usando nuevamente (24) y (25), se concluye

$$\|\boldsymbol{y}^{\#} - S(\theta_*)\boldsymbol{s}_k\| \;\leq\; K\,\sigma(\theta_k,\theta_{k+1})\|\boldsymbol{s}_k\|.$$

donde $K = \left(\beta_3 + \frac{1}{n}(\beta_1\epsilon + \bar{\beta}_*)\beta_2 + \frac{1}{n}\beta_1\,\hat{\beta}_*\right)$ con $\hat{\beta}_* = \|\bar{J}(\theta_*)\|$.

El siguiente teorema garantiza la convergencia local y $q-$superlineal del método BFGS estructurado para el problema de estimación de máxima verosimilitud:

Teorema 2 *Si las hipótesis* **H1** *a* **H3** *se verifican, entonces existen constantes positivas ϵ y δ tales que si $\|\theta_0 - \theta_*\| < \epsilon \quad y \quad \|A_0 - S(\theta_*)\| < \delta$, entonces la sucesión $\{\theta_k\}$ generada por el método BFGS estructurado para el problema de máxima verosimilitud converge $q-$superlinealmente a θ_*.*

Demostración. Es una aplicación directa del **Teorema 1** y **Lema 1**.

6 Pruebas numéricas

En esta sección, se analiza numéricamente el *comportamiento local* del *método BFGS estructurado* propuesto en la sección anterior. Para ello, se compara su desempeño con el desempeño de los métodos de *Scoring modificados*, *cuasi-Newton* (no estructurados) BFGS y *DFP*, y los *cuasi-Newton estructurados DFP*, *PSB* y *SR1* [6].

Estos métodos fueron aplicados a un *problema de estimación de mezcla finita* cuya función de densidad es la mezcla de dos *funciones de densidad de probabilidad normal univariada*, es decir,

$$f(\boldsymbol{x};\,\theta) = \frac{\alpha_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)}{\sigma_1^2}} + \frac{\alpha_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)}{\sigma_2^2}}, \tag{26}$$

donde $\theta = (\alpha_1,\, \alpha_2,\, \mu_1,\, \mu_2,\, \sigma_1,\, \sigma_2)$ es el vector de parámetros a estimar; α_1 y α_2 son constantes no negativas tales que $\alpha_1 + \alpha_2 = 1$.

Para escribir los códigos de los algoritmos se usó el *software* MATLAB®. Todos los experimentos numéricos fueron realizados en un computador con procesador Intel(R) Core(TM) i5-3450S CPU de 2.80 GHz, en un sistema operativo de 64 bits.

Para este experimento, se generó una muestra aleatoria de 400 puntos a partir de una mezcla con parámetros $\alpha_1^* = \alpha_2^* = 0.5;\ \mu_1^* = 0;\ \mu_2^* = 2.5;\ \ \sigma_1^* = \sigma_2^* = 1$ mediante la transformación de Box-Muller estándar [16],

$$x_k = \mu_i^* + \sigma_i^* \sqrt{-2\ln(r_1)}\cos(2\pi r_2), \quad i = 1, 2, \tag{27}$$

donde r_1 y r_2 son números aleatorios uniformemente distribuidos en el intervalo $[0, 1]$, los cuales fueron generados con el comando `rand(·)` de MATLAB. Para $k = 1, \ldots, 200$ se usa $i = 1$ y para generar los 200 puntos restantes $i = 2$.

Por conveniencia numérica, se reemplazó α_2 por $1-\alpha_1$ en (26). Así, la mezcla fue vista dependiendo de cinco parámetros: una *proporción de mezcla* α_1 en el intervalo $[0, 1]$, dos *medias* μ_1 y μ_2, y *dos desviaciones estándar* σ_1 y σ_2. Éstos fueron tratados como parámetros irrestrictos en los algoritmos, excepto que las iteraciones fueron detenidas si cualquier restricción se alcanzaba o se violaba.

Se generaron 100 puntos iniciales $\theta_0 = (t_1\alpha^*,\, t_2\mu_1^*,\, t_3\sigma_1^*,\, t_4\mu_2^*,\, t_5\sigma_2^*)$, donde $t_j = 0.5 + rand(\cdot)$, $j = 1, \ldots, 5$, con el objetivo de iniciar en una vecindad del vector de parámetros verdaderos.

Para cada uno de los 100 puntos iniciales con la misma muestra y la matriz hessiana inicial $C(\theta_0) = |l(\theta_0)|\, I$, donde I denota la matriz identidad, se usaron los algoritmos mencionados para encontrar un maximizador de L. Para cada método, se calculó el número promedio de iteraciones empleado hasta lograr convergencia.

Método	Iteraciones	Procesos exitosos
Scoring modificado	9.51190	84 %
BFGS	19.68000	100 %
PSB	25.76000	100 %
DFP	52.01266	79 %
SR1	19.68687	99 %
BFGS estructurado	8.96875	96 %
PSB estructurado	10.78125	96 %
DFP estructurado	9.71765	85 %
SR1 estructurado	12.47000	100 %

Tabla 1: Desempeño local del algoritmo local con una muestra de 400 puntos y 100 puntos iniciales.

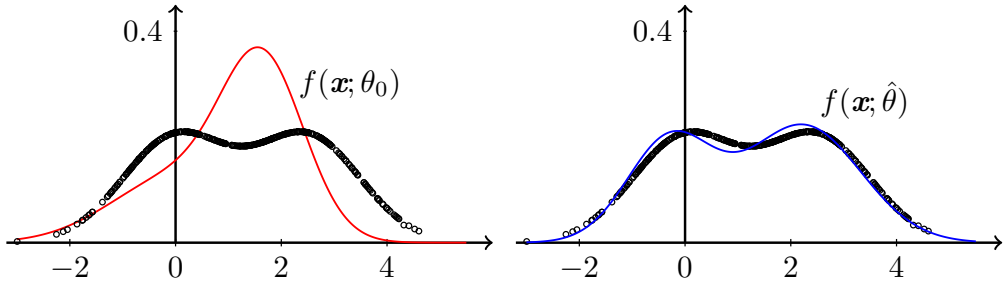

Figura 1: Relación de las distribuciones $f(\boldsymbol{x};\theta_0)$ y $f(\boldsymbol{x};\hat{\theta})$ con respecto a los puntos de la muestra ($f(\boldsymbol{x};\theta_*)$).

Presentamos los resultados en la Tabla 1, cuyas columnas contienen la siguiente información: la primera columna (método) indica el método utilizado; la segunda columna (iteraciones) contienen el número promedio de iteraciones empleado por cada método hasta lograr convergencia; la tercera columna (procesos exitosos) indica el porcentaje de experimentos en los que hubo convergencia.

Con respecto a cada uno de los métodos, la actualización DFP es la menos eficiente, puesto que realiza más iteraciones. En general, los métodos estructurados son más eficientes, pues alcanzan convergencia en un menor número de iteraciones, destacándose el método propuesto BFGS estructurado.

Finalmente, para ilustrar el funcionamiento del algoritmo BFGS estructurado, ilustramos los datos generados en una ejecución del algoritmo. En la Figura 1, los círculos representan los datos generados, usando (27); la curva de color rojo es la gráfica de la función de densidad de probabilidad (26) con el parámetro inicial θ_0; La curva de color azul es la gráfica de la función de densidad de probabilidad con $\hat{\theta}$, la solución encontrada por el algoritmo.

7 Comentarios finales

En muchos problemas de optimización sin restricciones, la matriz hessiana se puede expresar como la suma de dos partes: una, que es disponible y la otra parte que se desea aproximar. Se han realizado muchos trabajos en el área [7, 12, 13] los cuales, sugieren que en estos casos se debe aprovechar al máximo la estructura presente en el problema y no aproximar la información que ya ha sido calculada.

En este trabajo, se consideró el problema de máxima verosimilitud. Teniendo en cuenta la estructura especial de la matriz hessiana del logaritmo de la función de verosimilitud, se presenta el método BFGS estructurado para el problema de máxima verosimilitud y se desarrolla la teoría de convergencia local y $q-$superlineal de este nuevo método.

Se realizaron pruebas numéricas que permitieron analizar el comportamiento local del método BFGS estructurado para el problema de máxima verosimilitud. Se compara el comportamiento local de este método estructurado con el comportamiento local de otros métodos estructurados y no estructurados. Estas pruebas numéricas preliminares muestran un buen desempeño del método propuesto. Se recomienda incorporar estrategias de convergencia global y realizar el correspondiente análisis teórico y práctico del método.

Agradecimientos. Los autores agradecen a la Universidad del Cauca por el tiempo concedido para este trabajo mediante el proyecto de investigación VRI ID 4189.

Referencias bibliográficas

[1] Trinidad, A. (2014). *Modelos de crecimiento en Biología, su significado biológico y selección del modelo por su ajuste*. Tesis de Maestría. Universidad Autónoma Metropolitana Iztapalapa, División de Ciencias Básicas e Ingenierías, México.

[2] Al-Baaali, M. F. and Fletcher, R. (1985). Variational methods for nonlinear least squares. *Journal of the Operational Research Society* , *36*, 405-421.

[3] Bartholomew-Biggs, M. C. (1977). The estimation of the hessian matrix in nonlinear least squares problems with nonzero residuals.*Mathematical programming. 12*, 67-80.

[4] Dennis, J. E. (1976). A brief survey of convergence results for quasi-Newton methods. *SIAM-AMS Proceedings*, *9*, 186-201.

[5] Dennis, J. E., Gay, D. M. and Welsch, R. E. (1981). Algorithm 573:NL2SOL-An adaptive nonlinear Least Squares Algorithm. *Journal ACM transactions on mathematical Software*, 369-383.

[6] Dennis, J. E. and Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations.* New Jersey: Prentice-Hall.

[7] Dennis, J. E., Martínez, H. J. and Tapia, R. A. (1989). Covergence theory for structured BFGS secant method with an application to nonlinear least squares. *Journal of Optimization Theory and Applications*, *61*, 161-178.

[8] Dennis, J. E. and Walker, H. F. (1981). Convergence theorems for least change secant update methods. *SIAM Journal of Numerical Analysis*, *61*(6), 949-987.

[9] Fisher, R. A. (1946). A system of scoring linkage data, with special reference to pied factors in mice. American Naturalist, *80*(794), 568-578.

[10] Gonglewski, J. D. (1988). *On quasi- Newton methods for maximum-likelihood estimates with applications to the mixture density problem.* Ph.D. dissertation, Rice University, Houston, TX.

[11] Hathaway, R. J. (1983). Constrained maximum-likelihood estimation for a mixture of m univariate normal distributions. *Statistics Tech. Rep. 92*, 62F10-2, Univ. of South Carolina, Columbia, SC.

[12] Martínez R., H. J. (1988). *Local and superlinear convergence for structured secant methods from the convex class.* Ph.D. dissertation, Rice University, Houston, TX.

[13] Martínez R., H. J. and Engels J. (1991). Local and superlinear convergence for partially known quasi-Newton methods. *Siam Journal on Optimization*, *1*, 42-56.

[14] Moré, J. J., Garbow, B. S., and Hillstrom, K. E. (1980). User guide for MINPACK-1.*Argonne National Labs Report ANL*. 80-74.

[15] Mood, A. Graybill, F. and Boes, D., (1974). Introduction to the Theory of Statistics. *McGraw-Hill: tercera edición.*

[16] Pérez, R. (1991). *El método BFGS estructurado para estimación de máxima verosimilitud*. Tesis de Maestría, Universidad del Valle, Cali, Valle.

[17] Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review. 1*(2), 195-239

Dirección de los autores

Favián Arenas
Departamento de Matemáticas, Universidad del Cauca, Popayán - Colombia
farenas@unicauca.edu.co

Héctor Jairo Martínez
Departamento de Matemáticas, Universidad del Valle, Cali - Colombia
hector.martinez@correounivalle.edu.co

Rosana Pérez
Departamento de Matemáticas, Universidad del Cauca, Popayán - Colombia
rosana@unicauca.edu.co