REVISTA DE MATEMÁTICA: TEORÍA Y APLICACIONES 2016 **23**(2) : 475–488
CIMPA – UCR ISSN: 1409-2433 (PRINT), 2215-3373 (ONLINE)

SIMPLIFICACIÓN DEL GRADIENTE DE LA FUNCIÓN DE MÁXIMA VEROSIMILITUD DEL ANÁLISIS FACTORIAL CONFIRMATORIO

SIMPLIFICATION TO THE MAXIMUM LIKELIHOOD FUNCTION IN CONFIRMATORY FACTOR ANALYSIS

LUIS ROJAS–TORRES*

Received: 14 Sep 2015; Revised: 10 Apr 2016; Accepted: 18 Apr 2016

---

*Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica. E-Mail: luismiguel.rojas@ucr.ac.cr

Resumen

En este artículo se presenta el proceso de simplificación del gradiente de la función de máxima verosimilitud, utilizada en la estimación del Análisis Factorial Confirmatorio. El gradiente obtenido se presenta en función de las matrices tradicionales del AFC: $\mathbf{\Lambda}$, $\mathbf{\Phi}$ y $\mathbf{\Theta}_\varepsilon$ (coeficientes de regresión, varianzas de las variables latentes y varianzas de los errores). Esta simplificación se realizó mediante las leyes de derivación de matrices y permitió obtener una expresión para el gradiente de fácil programación.

Palabras clave: gradiente; función de máxima verosimilitud; análisis factorial confirmatorio; derivación de matrices.

Abstract

This paper present the process of simplification of the function maximum likelihood gradient, used in the estimation of confirmatory factor analysis. The gradient obtained is presented in function of CFA traditional matrix: $\mathbf{\Lambda}$, $\mathbf{\Phi}$ y $\mathbf{\Theta}_\varepsilon$ (regression coefficients, variances of latent variables and error variances). This simplification was performed using matrices derivation laws and yielded an expression for the gradient easy for programing.

Keywords: gradient; maximum likelihood function; confirmatory factor analysis; matrices derivation.

Mathematics Subject Classification: 62H25.

1 Introducción

El Análisis Factorial Confirmatorio (AFC) es un tipo de modelo de ecuaciones estructurales (SEM por sus siglas en inglés: *structural equation models*) que trabaja específicamente con modelos de medida, esto es, las relaciones entre variables observadas y variables no observadas [2]. El análisis de estas relaciones se basa en la reexpresión de la matriz de varianzas y covarianzas de las variables observadas, en matrices cuyos parámetros son las relaciones en estudio o las varianzas de las variables no observadas. Una vez planteada esta forma equivalente, se debe buscar los parámetros que aproximen esta forma a la matriz de varianzas observadas, de la manera más precisa; lo cual se resuelve mediante un proceso de optimización sobre una función que modele la discrepancia anterior.

Generalmente, la reexpresión de la matriz de varianzas y convarianzas de las variables observadas de un modelo de ecuaciones estructurales se presenta por bloques, los cuales se definen con base al tipo de variables observadas (exógenas o endógenas); en el caso del AFC la matriz de varianzas se reduce al bloque definido por las variables exógenas. En [6] se presenta una reexpresión de la matriz de varianzas de las variables observadas de un modelo SEM que no utiliza bloques, por lo cual hace uso de matrices poco tradicionales dentro del contexto del SEM, sin embargo ejemplifica paso a paso la obtención de las derivadas parciales de la función de discrepancia respecto a los parámetros planteados en la nueva forma de la matriz de varianzas.

Siguiendo la metodología de [6], en este artículo se calcula la forma simplificada del gradiente de la función de discrepancia de Máxima Verosimilitud de un Análisis Factorial Confirmatorio (AFC), pero utilizando la reexpresión tradicional de la matriz de varianzas y covarianzas con base en el bloque de variables observadas exógenas. Esta definición facilita a los investigadores el análisis detallado del proceso de estimación del AFC, ya que este proceso generalmente depende del cálculo del gradiente.

Además del cálculo detallado del gradiente de la función de discrepancia, el artículo presenta un ejemplo de cómo se utilizan las fórmulas y un código de R, para estimar los coeficientes de un AFC utilizando el método de minimización de steepest descend, en el cual se presenta la programación para el cálculo del gradiente.

2 Definición del AFC

La ecuación matricial básica del Análisis Factorial Confirmatorio está dada por

$$\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\varepsilon}$$

donde $\mathbf{x} = (x_i)$ es el vector $q \times 1$ de variables observadas, $\boldsymbol{\xi} = (\xi_i)$ es el vector $n \times 1$ de variables latentes, $\mathbf{\Lambda} = (\lambda_{ij})$ es la matriz $q \times n$ de coeficientes de regresión de las variables latentes sobre las variables observadas y $\boldsymbol{\varepsilon} = (\varepsilon_i)$ es el vector $q \times 1$ de errores asociados a las variables observadas. Para esta ecuación, se supone que las variables latentes y los errores son independientes.

Bajo la definición de la ecuación matricial del AFC, se deduce que la matriz de variancias y covariancias de las variables observadas $\mathbf{\Sigma}$, es equivalente a la matriz de variancias de $\mathbf{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\varepsilon}$, la cual se puede reexpresar como $\mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}_\varepsilon$, donde $\mathbf{\Phi} = (\phi_{ij})$ representa la matriz de variancias y covariancias de las variables latentes, y $\mathbf{\Theta}_\varepsilon = (\theta_{ij})$, la matriz de variancias y covariancias de los errores. Esta última matriz se supone que es diagonal, esto es, que los errores no están correlacionados entre sí [4].

Ahora, sea $\mathbf{\Sigma}(\mathbf{\Omega})$ una función que evalúe la definición de $\mathbf{\Sigma}$, en el vector $\mathbf{\Omega}$ de parámetros desconocidos que aparecen en la definición. De esta manera, se tiene que $\mathbf{\Omega} = (\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\theta})$, donde $\boldsymbol{\lambda}$, $\boldsymbol{\phi}$ y $\boldsymbol{\theta}$, son los vectores de parámetros desconocidos que aparecen en las matrices $\mathbf{\Lambda}$, $\mathbf{\Phi}$ y $\mathbf{\Theta}_\varepsilon$, respectivamente.

La estimación de los valores de las variables del AFC, se basa en determinar el vector $\mathbf{\Omega}_0$, que mejor aproxime la matriz $\mathbf{\Sigma}(\mathbf{\Omega}_0)$, a la matriz de variancias y covariancias muestral $\mathbf{S}$. Para hallar el vector $\mathbf{\Omega}_0$, se define una función $F(\mathbf{\Omega})$ que permite determinar la discrepancia entre $\mathbf{S}$ y $\mathbf{\Sigma}(\mathbf{\Omega})$, y posteriormente se busca su mínimo, el cual correspondería a la solución buscada. En las siguientes líneas se utilizará la notación $\widehat{\mathbf{\Sigma}}$, para hacer referencia a la función $\mathbf{\Sigma}(\mathbf{\Omega})$.

Una función de discrepancia muy utilizada es la de máxima verosimilitud, la cual se basa en la hipótesis de que las variables observadas provienen de una distribución multinormal [3]; la fórmula de esta función es la siguiente:

$$F_{\text{ML}}(\mathbf{\Omega}) = \log|\widehat{\mathbf{\Sigma}}| + \text{tr}(\mathbf{S} * \widehat{\mathbf{\Sigma}}^{-1}) - \log|\mathbf{S}| - q.$$

Generalmente, para estimar el mínimo de esta función, se requiere de métodos numéricos de minimización. Estos, en su mayoría, dependen del vector gradiente, el cual se calculará en la siguiente sección. En los próximos apartados, cuando se haga mención de F, se estará refiriendo a la función de discrepancia de Máxima Verosimilitud.

3 Gradiente de la función de discrepancia de la máxima verosimilitud

3.1 Derivada de F respecto a cualquier variable

Primeramente, nótese que para cualquier variable ω de $\mathbf{\Omega}$, se cumple que

$$\frac{\partial F}{\partial \omega} = \text{tr}\left(\mathbf{Q}\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \omega}\right) \tag{1}$$

donde $\mathbf{Q} = \widehat{\mathbf{\Sigma}}^{-1} - \widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}\widehat{\mathbf{\Sigma}}^{-1}$, $\widehat{\mathbf{\Sigma}} = (\sigma_{ij})$ y (ver [6])

$$\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \omega} = \left(\frac{\partial \sigma_{ij}}{\partial \omega}\right).$$

Prueba

En esta prueba, hay que recordar que para $\mathbf{A}$, $\mathbf{B}$ y $\mathbf{C}$, matrices con tamaños adecuados para las operaciones indicadas, se cumplen las siguientes propiedades:

- $\dfrac{\partial(\mathbf{A}+\mathbf{B})}{\partial \omega} = \dfrac{\partial \mathbf{A}}{\partial \omega} + \dfrac{\partial \mathbf{B}}{\partial \omega}.$

- $\dfrac{\partial\left(\mathbf{AC}\right)}{\partial\omega}=\dfrac{\partial\mathbf{A}}{\partial\omega}\mathbf{C}$, donde $\mathbf{C}$ es una matriz constante.
- $\dfrac{\partial\left(\mathbf{AB}\right)}{\partial\omega}=\mathbf{A}\dfrac{\partial\mathbf{B}}{\partial\omega}+\dfrac{\partial\mathbf{A}}{\partial\omega}\mathbf{B}$.
- $\dfrac{\partial\mathbf{A}^{-1}}{\partial\omega}=-\mathbf{A}^{-1}\dfrac{\partial\mathbf{A}}{\partial\omega}\mathbf{A}^{-1}$.
- $\dfrac{\partial\mathrm{tr}\left(\mathbf{A}\right)}{\partial\omega}=\mathrm{tr}\left(\dfrac{\partial\mathbf{A}}{\partial\omega}\right)$.
- $\dfrac{\partial\log|\mathbf{A}|}{\partial\omega}=\mathrm{tr}\left(\mathbf{A}^{-1}\dfrac{\partial\mathbf{A}'}{\partial\omega}\right)$ (ver [1]).

Ahora, se procede a la derivación de la función F. Por ley de la derivada de una suma, se obtiene que

$$\frac{\partial F}{\partial\omega}=\frac{\partial}{\partial\omega}\log|\widehat{\mathbf{\Sigma}}|+\frac{\partial}{\partial\omega}\mathrm{tr}(\mathbf{S}*\widehat{\mathbf{\Sigma}}^{-1})-\frac{\partial}{\partial\omega}\log|\mathbf{S}|-\frac{\partial}{\partial\omega}q.$$

Luego, al primer sumando se le aplica la ley de la derivada del logaritmo, al segundo la ley de la traza y a los dos últimos sumandos se les anula por ser constantes; de esta manera

$$\frac{\partial F}{\partial\omega}=\mathrm{tr}\left(\widehat{\mathbf{\Sigma}}^{-1}\frac{\partial\widehat{\mathbf{\Sigma}}'}{\partial\omega}\right)+\mathrm{tr}\left(\mathbf{S}*\frac{\partial}{\partial\omega}\widehat{\mathbf{\Sigma}}^{-1}\right).$$

Ahora, dado que $\widehat{\mathbf{\Sigma}}$ es simétrica, se puede sustituir $\widehat{\mathbf{\Sigma}}'$ por $\widehat{\mathbf{\Sigma}}$; y además se puede aplicar la ley de la derivada de la inversa, con lo que se tiene que

$$\frac{\partial F}{\partial\omega}=\mathrm{tr}\left(\widehat{\mathbf{\Sigma}}^{-1}\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\omega}\right)+\mathrm{tr}\left(\mathbf{S}*-\widehat{\mathbf{\Sigma}}^{-1}\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\omega}\widehat{\mathbf{\Sigma}}^{-1}\right).$$

Seguidamente, se aplica la ley $\mathrm{tr}(\mathbf{ABC})=\mathrm{tr}(\mathbf{CAB})$ al segundo término, para que en ambos sumandos la derivada parcial quede a la derecha

$$\frac{\partial F}{\partial\omega}=\mathrm{tr}\left(\widehat{\mathbf{\Sigma}}^{-1}\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\omega}\right)+\mathrm{tr}\left(-\widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}\widehat{\mathbf{\Sigma}}^{-1}\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\omega}\right).$$

Luego, se unen las dos trazas en una sola, y se saca a factor común la derivada de $\widehat{\widehat{\mathbf{\Sigma}}}$ respecto a ω, lo cual implica que

$$\frac{\partial F}{\partial\omega}=\mathrm{tr}\left(\left[\widehat{\mathbf{\Sigma}}^{-1}-\widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}\widehat{\mathbf{\Sigma}}^{-1}\right]\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\omega}\right).$$

Finalmente, se sustituye $\widehat{\mathbf{\Sigma}}^{-1} - \widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}\widehat{\mathbf{\Sigma}}^{-1}$ por $\mathbf{Q}$, tal como se definió en el enunciado

$$\frac{\partial F}{\partial \omega} = \text{tr}\left(\mathbf{Q}\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \omega}\right).$$

3.2 Derivadas parciales de $\widehat{\Sigma}$

Debido al resultado mostrado en la sección anterior, para determinar el gradiente de F, se hace necesario determinar las derivadas de $\widehat{\mathbf{\Sigma}}$ respecto a las variables de $\mathbf{\Omega}$. A continuación, se procederá a calcular las derivadas parciales de $\widehat{\mathbf{\Sigma}}$ respecto a λ_{ij} en $\boldsymbol{\lambda}$, donde ij representa la entrada en que se ubica esa variable dentro de la matriz $\mathbf{\Lambda}$.

Ahora, la derivada de $\widehat{\mathbf{\Sigma}}$ respecto a λ_{ij} está dada por

$$\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \lambda_{ij}} = \frac{\partial\left(\mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}_\varepsilon\right)}{\partial \lambda_{ij}}.$$

Usando las leyes de matrices se obtiene que,

$$\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \lambda_{ij}} = \mathbf{\Lambda\Phi}\frac{\partial \mathbf{\Lambda}'}{\partial \lambda_{ij}} + \frac{\partial \mathbf{\Lambda}}{\partial \lambda_{ij}}\mathbf{\Phi\Lambda}'.$$

Luego, dado que en la matriz $\boldsymbol{\lambda}$ solamente el elemento ij depende de λ_{ij}, se obtiene que

$$\frac{\partial \mathbf{\Lambda}}{\partial \lambda_{ij}} = \mathbf{1}^{ij}_{n\times q}$$

donde $\mathbf{1}^{ij}_{n\times q}$ es una matriz $n \times q$, con 0's en todas las entradas, a excepción de la entrada ij, que es igual a 1. Finalmente, la derivada de $\widehat{\mathbf{\Sigma}}$ respecto a λ_{ij}, equivale a

$$\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \lambda_{ij}} = \mathbf{\Lambda\Phi}\mathbf{1}^{ji}_{n\times q} + \mathbf{1}^{ij}_{q\times n}\mathbf{\Phi\Lambda}'. \tag{2}$$

De manera análoga, para ϕ_{ij} en $\boldsymbol{\phi}$ y θ_{ij} en $\boldsymbol{\theta}$ (donde ij representa la entrada que tienen esas variables en las matrices $\mathbf{\Phi}$ y $\mathbf{\Theta}_\varepsilon$, respectivamente), se obtiene

$$\begin{aligned}\frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \lambda_{ij}} &= \mathbf{\Lambda}\frac{\partial \mathbf{\Phi}}{\partial \phi_{ij}}\mathbf{\Lambda}' \\ \frac{\partial \widehat{\mathbf{\Sigma}}}{\partial \theta_{ij}} &= \frac{\partial \mathbf{\Theta}_\varepsilon}{\partial \theta_{ij}}.\end{aligned}$$

Luego, dado que las matrices $\mathbf{\Phi}$ y $\mathbf{\Theta}_\varepsilon$ son simétricas, sus derivadas respecto al elemento ij, poseen un 1 adicional en la posición ji. Finalmente, dado que la matriz de orden n con 1's en las posiciones ij y ji resulta de la operación $\mathbf{1}^{ij}_{n\times n} + \mathbf{1}^{ji}_{n\times n} - \mathbf{1}^{ij}_{n\times n}\mathbf{1}^{ij}_{n\times n}$, se obtiene que

$$\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\lambda_{ij}} = \mathbf{\Lambda}(\mathbf{1}^{ij}_{n\times n} + \mathbf{1}^{ji}_{n\times n} - \mathbf{1}^{ij}_{n\times n}\mathbf{1}^{ij}_{n\times n})\mathbf{\Lambda}' \tag{3}$$

$$\frac{\partial\widehat{\mathbf{\Sigma}}}{\partial\theta_{ij}} = \mathbf{1}^{ij}_{q\times q} + \mathbf{1}^{ji}_{q\times q} - \mathbf{1}^{ij}_{q\times q}\mathbf{1}^{ij}_{q\times q}. \tag{4}$$

3.3 Derivadas parciales de F

Con base a los resultados anteriores, se puede proceder a calcular las derivadas de F respecto a λ_{ij}, ϕ_{ij} y θ_{ij}. De esta manera, sustituyendo 2, 3 y 4 en 1, se obtiene que

$$\begin{aligned}
\frac{\partial F}{\partial\lambda_{ij}} &= \text{tr}\left[\mathbf{Q}\left(\mathbf{\Lambda\Phi}\mathbf{1}^{ji}_{n\times q} + \mathbf{1}^{ij}_{q\times n}\mathbf{\Phi\Lambda}'\right)\right]\\
\frac{\partial F}{\partial\phi_{ij}} &= \text{tr}\left[\mathbf{Q\Lambda}(\mathbf{1}^{ij}_{n\times n} + \mathbf{1}^{ji}_{n\times n} - \mathbf{1}^{ij}_{n\times n}\mathbf{1}^{ij}_{n\times n})\mathbf{\Lambda}'\right]\\
\frac{\partial F}{\partial\theta_{ij}} &= \text{tr}\left[\mathbf{Q}(\mathbf{1}^{ij}_{n\times n} + \mathbf{1}^{ji}_{n\times n} - \mathbf{1}^{ij}_{n\times n}\mathbf{1}^{ij}_{n\times n})\right].
\end{aligned}$$

La ecuación de la derivada de F respecto a λ_{ij}, se puede expresar como $\text{tr}\left[\mathbf{Q\Lambda\Phi}\mathbf{1}^{ji}_{n\times q}\right] + \text{tr}\left[\mathbf{Q}\mathbf{1}^{ij}_{q\times n}\mathbf{\Phi\Lambda}'\right]$. Luego, al aplicar leyes de la traza se obtiene que el segundo elemento de la suma es igual $\text{tr}\left[\mathbf{1}^{ij}_{q\times n}\mathbf{\Phi\Lambda}'\mathbf{Q}\right] = \text{tr}\left[\left(\mathbf{Q\Lambda\Phi}\mathbf{1}^{ji}_{n\times q}\right)'\right]$, y por leyes de la traza, este sumando es igual a $\text{tr}\left[\mathbf{Q\Lambda\Phi}\mathbf{1}^{ji}_{n\times q}\right]$. De esta manera se obtiene que

$$\frac{\partial F}{\partial\lambda_{ij}} = 2\cdot\text{tr}\left[\mathbf{Q\Lambda\Phi}\mathbf{1}^{ji}_{n\times q}\right].$$

Esta derivada se simplifica aún más, si se utiliza el hecho de que $\text{tr}\left[\mathbf{A}\mathbf{1}^{ij}\right]$ es igual a $[\mathbf{A}]_{ji}$. Por tanto, se concluye que,

$$\frac{\partial F}{\partial\lambda_{ij}} = 2\cdot[\mathbf{Q\Lambda\Phi}]_{ij}\,. \tag{5}$$

La derivada de F respecto a ϕ_{ij} se puede reexpresar como $\text{tr}\left[\mathbf{\Lambda}'\mathbf{Q\Lambda}(\mathbf{1}^{ij}_{n\times n} + \mathbf{1}^{ji}_{n\times n} - \mathbf{1}^{ij}_{n\times n}\mathbf{1}^{ij}_{n\times n})\right]$. Ahora, dado que para una matriz simétrica $\mathbf{A}$, de orden n, se cumple que

$$\text{tr}\left[\mathbf{A}(\mathbf{1}_{n\times n}^{ij}+\mathbf{1}_{n\times n}^{ji}-\mathbf{1}_{n\times n}^{ij}\mathbf{1}_{n\times n}^{ij})\right]=(2-[\mathbf{I}_n]_{ij})\,[\mathbf{A}]_{ij}$$

entonces, dado que $\mathbf{\Lambda}'\mathbf{Q}\mathbf{\Lambda}$ es simétrica, se llega a

$$\frac{\partial F}{\partial \phi_{ij}}=(2-[\mathbf{I}_n]_{ij})\left[\mathbf{\Lambda}'\mathbf{Q}\mathbf{\Lambda}\right]_{ij}. \tag{6}$$

Mediante el mismo razonamiento, se obtiene que

$$\frac{\partial F}{\partial \theta_{ij}}=(2-[\mathbf{I}_q]_{ij})\,[\mathbf{Q}]_{ij}\,.$$

Finalmente, dado a que en la definición clásica del AFC, las variables de $\mathbf{\Theta}_\varepsilon$ se encuentran en la diagonal, esta definición se simplifica a

$$\frac{\partial F}{\partial \theta_{ii}}=[\mathbf{Q}]_{ii}\,. \tag{7}$$

3.4 Ejemplo del cálculo del gradiente

A continuación se presenta el cálculo del gradiente de la función F para un AFC específico en un punto particular. Considere las siguientes matrices asociadas a un AFC de 4 variables observadas y 2 variables latentes, cuya métrica se define fijando las variancias de las variables latentes a 1:

$$\mathbf{\Lambda}=\begin{pmatrix}\lambda_{11} & 0\\ \lambda_{21} & 0\\ 0 & \lambda_{32}\\ 0 & \lambda_{42}\end{pmatrix},\ \mathbf{\Phi}=\begin{pmatrix}1 & \\ \phi_{21} & 1\end{pmatrix},$$

$$\mathbf{\Theta}_\varepsilon=\begin{pmatrix}\theta_{11} & & & \\ 0 & \theta_{22} & & \\ 0 & 0 & \theta_{33} & \\ 0 & 0 & 0 & \theta_{44}\end{pmatrix}\ \text{y}\ \mathbf{S}=\begin{pmatrix}1.202 & & & \\ 0.265 & 1.298 & & \\ 0.167 & 0.034 & 1.439 & \\ 0.169 & 0.164 & 0.190 & 1.170\end{pmatrix}.$$

De esta manera, el vector de parámetros está dado por

$$\mathbf{\Omega}=(\lambda_{11},\lambda_{21},\lambda_{32},\lambda_{42},\phi_{21},\theta_{11},\theta_{22},\theta_{33},\theta_{44}).$$

Luego, el gradiente de F en un punto específico, supóngase

$$\mathbf{\Omega}_0=(1,1,1,1,0.5,0.7,0.7,0.7,0.7),$$

depende de las matrices

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \mathbf{\Phi} = \begin{pmatrix} 1 & \\ 0.5 & 1 \end{pmatrix} \text{ y } \mathbf{\Theta}_\varepsilon = \begin{pmatrix} 0.7 & & & \\ 0 & 0.7 & & \\ 0 & 0 & 0.7 & \\ 0 & 0 & 0 & 0.7 \end{pmatrix}.$$

A partir de estas matrices se obtienen las matrices $\widehat{\mathbf{\Sigma}} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}_\varepsilon$ y $\mathbf{Q} = \widehat{\mathbf{\Sigma}}^{-1} - \widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}\widehat{\mathbf{\Sigma}}^{-1}$. Luego, utilizando las ecuaciones 5, 6 y 7, se procede a calcular las matrices de las cuales dependen las derivadas de F respecto a λ_{ij}, ϕ_{ij} y θ_{ii}, las cuales son:

$$\mathbf{Q\Lambda}\Phi = \begin{pmatrix} 0.162 & 0.034 \\ 0.125 & 0.095 \\ 0.088 & 0.085 \\ 0.044 & 0.209 \end{pmatrix}, \mathbf{\Lambda}'\mathbf{Q\Lambda} = \begin{pmatrix} 0.297 & \\ -0.019 & 0.304 \end{pmatrix}$$

y

$$\mathbf{Q} = \begin{pmatrix} -0.171 & & & \\ 0.365 & -0.261 & & \\ -0.061 & 0.122 & -0.444 & \\ -0.001 & -0.078 & 0.498 & -0.249 \end{pmatrix}.$$

Finalmente, la derivada de F respecto a λ_{ij} evaluada en $\mathbf{\Omega}_0$, corresponde al doble de la entrada ij de la matriz $\mathbf{Q\Lambda}\Phi$; con respecto a ϕ_{ij} corresponde al doble de la entrada ij de la matriz $\mathbf{\Lambda}'\mathbf{Q\Lambda}$, a menos que sea una entrada de la diagonal ($i = j$), en cuyo caso corresponde únicamente a la entrada ii de la matriz; y por último, las derivadas respecto a θ_{ii}, corresponden al elemento ii de la matriz $\mathbf{Q}$.

De esta manera se obtiene que el gradiente de F evaluado en $\mathbf{\Omega}_0$ es:

$$\begin{bmatrix} \partial F(\mathbf{\Omega}_0)/\partial\lambda_{11} \\ \partial F(\mathbf{\Omega}_0)/\partial\lambda_{21} \\ \partial F(\mathbf{\Omega}_0)/\partial\lambda_{32} \\ \partial F(\mathbf{\Omega}_0)/\partial\lambda_{42} \\ \partial F(\mathbf{\Omega}_0)/\partial\phi_{21} \\ \partial F(\mathbf{\Omega}_0)/\partial\theta_{11} \\ \partial F(\mathbf{\Omega}_0)/\partial\theta_{22} \\ \partial F(\mathbf{\Omega}_0)/\partial\theta_{33} \\ \partial F(\mathbf{\Omega}_0)/\partial\theta_{44} \end{bmatrix} = \begin{bmatrix} 2\cdot[\mathbf{Q\Lambda}\Phi]_{11} \\ 2\cdot[\mathbf{Q\Lambda}\Phi]_{21} \\ 2\cdot[\mathbf{Q\Lambda}\Phi]_{32} \\ 2\cdot[\mathbf{Q\Lambda}\Phi]_{42} \\ 2\cdot[\mathbf{\Lambda}'\mathbf{Q\Lambda}]_{21} \\ [\mathbf{Q}]_{11} \\ [\mathbf{Q}]_{22} \\ [\mathbf{Q}]_{33} \\ [\mathbf{Q}]_{44} \end{bmatrix} = \begin{bmatrix} 0.324 \\ 0.250 \\ 0.170 \\ 0.418 \\ -0.038 \\ -0.171 \\ -0.261 \\ -0.444 \\ -0.249 \end{bmatrix}.$$

4 Estimación de los coeficientes del AFC

Existen varios métodos de minimización en funciones de varias variables, uno de ellos es el método de *steepest descend*. Los pasos para encontrar el valor mínimo de una función multivariable F a partir de este método son:

- Seleccione una solución inicial $\mathbf{x}_0$.
- Calcule el vector director $\mathbf{d} = -\nabla f(\mathbf{x}_0)/||\nabla f(\mathbf{x}_0)||$.
- Calcule el punto auxiliar $\mathbf{x}_0 + \mathbf{d}$.
- Estime $\beta_1 = \min\left[1, \frac{-2F(\mathbf{x}_0)}{\nabla(\mathbf{x}_0)'\mathbf{d}}\right]$.
- Calcule $\beta = \frac{\nabla(\mathbf{x}_0)' d\,\beta_1}{\nabla(\mathbf{x}_1)'\mathbf{d} - \nabla(\mathbf{x}_0)'\mathbf{d}}$
- Calcule el punto $\mathbf{x}_1 = \mathbf{x}_0 + \beta\mathbf{d}$, el cual es una aproximación al punto donde la función F halla su mínimo sobre la recta que pasa por $\mathbf{x}_0$, en la dirección $\mathbf{d}$.
- Calcule $|F(\mathbf{x}_1) - F(\mathbf{x}_0)|$ y verifique si esa diferencia satisface la condición de convergencia, en caso afirmativo tome $\mathbf{x}_1$ como la solución del modelo.
- Si no se satiface, tome $\mathbf{x}_0 = \mathbf{x}_1$ y vuelva a repetir el proceso [6].

Ahora, dado a que en la sección anterior se estimó la forma explícita del gradiente en un punto, la programación de un AFC bajo este método se facilita. En los anexos se presenta un código de R [7], que permite estimar los coeficientes de un AFC con el método de *steepest descend*, el cual puede ser modificado por el lector para las necesidades de su investigación, por ejemplo, utilizar otro método de minimización u otro valor inicial.

Si se estima la solución del AFC presentado en el ejemplo de la sección 3.4, utilizando el código presentado y el software *lavaan* [8] se obtienen soluciones muy similares. Estas soluciones se presentan en la Tabla 1.

Tabla 1: Soluciones del AFC presentado en la sección 3.4.

Código	λ_{11}	λ_{12}	λ_{32}	λ_{42}	ϕ_{12}	δ_{11}	δ_{22}	δ_{33}	δ_{44}
propuesto	0.608	0.436	0.370	0.514	0.594	0.832	1.108	1.302	0.906
lavaan	0.608	0.435	0.370	0.513	0.594	0.831	1.107	1.301	0.905

5 Conclusión

El artículo presentó un método accesible para el cálculo del gradiente de la función de discrepancia de máxima verosimilitud para el AFC. Se considera que el estudio detallado de la fórmula, puede facilitar la comprensión del gradiente generalizado a modelos de ecuaciones estructurales, presentado en [6], además de la programación de códigos para el estudio de estos modelos.

A partir de la forma simplificada del gradiente y con los teoremas presentados en este artículo, se puede generar una forma simplificada de la Matriz Hessiana de la función de discrepancia, la cual permite calcular los errores estándar de las estimaciones y brinda la posibilidad de utilizar otros métodos de estimación, como el de Newton-Raphson [5].

Referencias

[1] Bock, R.D.; Bargmann, R.E. (1966) "Analysis of covariance structures", *Psychometrika* **31**(4): 507–534.

[2] Brown (2006) *Confirmatory Factor Analysis for Applied Research*. The Guilford Press, New York.

[3] Cea, M.A. (2002) *Análisis Multivariable: Teoría y Práctica en la Investigación Social*. Editorial Síntesis, Madrid.

[4] Jöreskog, K.G.; Sörbom, D.; du Toit, S.; du Toit, M. (2000) *LISREL 8: New Statistical Features*. Scietific Software International, Lincolnwood IL.

[5] Kaplan, D. (2009) *Structural Equation Modeling. Foundations and Extensions*. Sage Publications, Thousand Oaks CA.

[6] Mulaik, S.A. (2009) *Linear Causal Modeling with Structural Equations*. Chapman & Hall/CRC, Boca Raton FL.

[7] R Core Team (2014) "R: A language and environment for statistical computing. R Foundation for Statistical Computing", Vienna, Austria. `http://www.R-project.org/`

[8] Rossell, Y. (2012) "lavaan: An R package for structural equation modeling", *Journal of Statistical Software* **48**(2): 1–36.

Anexos

A continuación se presenta el código de R para la estimación de un AFC, los parámetros de la función son los siguientes:

- S: Matriz $q \times q$ de variancias y covariancias muestral.
- Lambda: Matriz $q \times n$ de 1's y 0's, donde las entradas ij iguales a 1 indican que existe una regresión de la variable observada i sobre la variable latente j. La variancia muestral de las variable observada i debe presentarse en la entrada ii de la matriz de variancias y covariancias muestral.
- std.lat: Si es verdadero la métrica de las variables latentes se define igualando sus variancias a 1; si es falso, esta se define fijando a 1, el primer coeficiente de regresión sobre la variable latente, que aparece en la matriz Lambda. Por default es Verdadero.
- Tol: Es el umbral para determinar la convergencia del ciclo de minimización. Por defecto es igual a 1×10^{-5}.

```
AFC<-function(S, Lambda, std.lat="T",Tol=1*10^-5)
{
##vector de parametros
q=dim(Lambda)[1];n=dim(Lambda)[2]
i=c()
j=c()
k=1
if(std.lat=="T")
 {
 for(s in 1:q){for (t in 1:n){if(Lambda[s,t]==1){i[k]=s;j[k]=t;k=k+1}}}
 n1=k-1
 if(n>1){for(s in 2:n){for (t in 1:(s-1)){i[k]=s;j[k]=t;k=k+1 }}}
 n2=k-1
 for(s in 1:q){i[k]=s;j[k]=s;k=k+1}
 n3=k-1
 }
 else
 {
 Lambda_aux=Lambda
 for (t in 1:n){for(s in 1:q){if(Lambda[s,t]==1){Lambda_aux[s,t]=0;break}}}
 for(s in 1:q){for (t in 1:n){if(Lambda_aux[s,t]==1){i[k]=s;j[k]=t;k=k+1}}}
 n1=k-1
 for(s in 1:n){for (t in 1:s){i[k]=s;j[k]=t;k=k+1 }}
 n2=k-1
 for(s in 1:q){i[k]=s;j[k]=s;k=k+1}
 n3=k-1}
##Funcion de discrepancia y Gradiente
F<-function(x)
 {
 ##Matrices base
```

```
 Phi=matrix(rep(1,n*n),n,n)
 Theta=matrix(rep(0,q*q),q,q)
 for(k in 1:n1){Lambda[i[k],j[k]]=x[k]}
 if(std.lat=="F"|(std.lat=="T"&n>1))
  {
  for(k in (n1+1):n2){Phi[i[k],j[k]]=x[k];Phi[j[k],i[k]]=x[k]}
  }
 for(k in (n2+1):n3){Theta[i[k],j[k]]=x[k];Theta[j[k],i[k]]=x[k]}
 Sigma=Lambda%*%Phi%*%t(Lambda)+Theta
 Q=solve(Sigma)-solve(Sigma)%*%S%*%solve(Sigma)
 dLambda=Q%*%Lambda%*%Phi
 dPhi=t(Lambda)%*%Q%*%Lambda
 #Funcion de discrepancia
 f<-log(det(Sigma))+sum(diag(S%*%solve(Sigma)))-log(det(S))-q
 ##Gradiente
 Gra=c()
 for(k in 1:n1){Gra[k]=2*dLambda[i[k],j[k]]}
 if(std.lat=="F"|(std.lat=="T"&n>1))
  {
  for(k in (n1+1):n2){Gra[k]=(2-diag(rep(1,n))[i[k],j[k]])*dPhi[i[k],j[k]]}
  }
 for(k in (n2+1):n3){Gra[k]=(2-diag(rep(1,q))[i[k],j[k]])*Q[i[k],j[k]]}
 ##Salida
 return(list(f,Gra))
 }
##Proceso de minimizacion
##Valores iniciales
x0=c(rep(1,n1),rep(.5,(n2-n1)),diag(S)/2)
if(std.lat=="F"){for (k in (n1+1):n2){if(i[k]==j[k]){x0[k]=1}}}
Par_x0=F(x0)
f0<-Par_x0[[1]]
Gra0<-Par_x0[[2]]
delta=1
##Desarrollo del ciclo
while(delta>Tol)
 {
 d=-Gra0/sqrt(sum(Gra0^2))
 beta1=-2*f0/sum(Gra0*d)
 beta1=min(1,beta1)
 x1=x0+beta1*d
 Gra1<-F(x1)[[2]]
 beta=sum(Gra0*d)*beta1/(sum(Gra1*d)-sum(Gra0*d))
 xnew=x0+beta*Gra0
 ##AnÃ¡lisis de condiciÃ³n de salida
 Par_xnew=F(xnew)
 f_xnew=Par_xnew[[1]]
 delta=abs(f_xnew-f0)
 x0=xnew
 Gra0=Par_xnew[[2]]
 f0=f_xnew
 }
##Salida
rel=rep(NA,n3)
for(k in 1:n1)
 {
 a=toString(i[k]);b=toString(j[k]);
```

```
 rel[k]=paste(c('x',a,'~','F',b),collapse="")
 }
if(std.lat=="F"|(std.lat=="T"&n>1))
 {
 for(k in (n1+1):n2)
  {
  a=toString(i[k]);b=toString(j[k]);
rel[k]=paste(c('F',a,'~','F',b),collapse="")
  }
 }
for(k in (n2+1):n3)
 {
 a=toString(i[k]);b=toString(j[k]);
 rel[k]=paste(c('x',a,'~','x',b),collapse="")
 }
coef=xnew
X=data.frame(rel,coef)
return(X)
}
```