

ANÁLISIS DE CORRELACIÓN CANÓNICA USANDO ALGORITMOS GENÉTICOS

Brenda Catalina Matías Castillo^{1,*}; María de Lourdes Sandoval Solís³, Gladys Linares Fleites², Hortensia Josefina Reyes Cervantes¹.

¹ Facultad de Ciencias Físico Matemáticas. BUAP, Puebla, México.

² Posgrado en Ciencias Ambientales, Instituto de Ciencias, BUAP, Puebla, México.

³ Facultad de Ciencias de la Computación, BUAP, Puebla, México.

ABSTRACT

Canonical Correlation Analysis (CCA) is an exploratory method of Multivariate Analysis, it studies the relationship between two sets of quantitative variables observed in the same set of individuals. It obtains new variables that are linear combination of the original variables of the two groups, such that the correlation between the projections of the data of this new variables is maximum. There are several proposals to determine canonical correlations and canonical vectors using techniques of numerical linear algebra, that is, in the formulation of Lagrange, CCA becomes a generalized eigenvalues and eigenvectors problem, assuming that the variance-covariance matrices are invertible. On the other hand, Genetic Algorithms (GA) are adaptive methods used to solve global optimization problems. When the invertibility condition is not satisfied by the variance-covariance matrices, we propose in this paper, use GA to solve the problem of CCA directly from the definition. We test the proposal with problems reported in the literature and also presents a real application of CCA to data of soil carbon in the area of Teziutlan, Puebla.

KEYWORDS: Canonical correlation, genetic algorithms, eigenvalue, eigenvector.

MSC: 62H20

RESUMEN

El Análisis de Correlación Canónica (ACC) es un método exploratorio del Análisis Multivariado, que estudia la relación entre dos conjuntos de variables cuantitativas observadas sobre el mismo conjunto de individuos. El ACC obtiene nuevas variables, que son combinaciones lineales de las variables originales de los dos grupos, tal que la correlación entre las proyecciones de los datos de estas nuevas variables sea máxima. Existen diferentes propuestas para determinar las correlaciones y vectores canónicos usando técnicas del álgebra lineal numérica, es decir, en la formulación de Lagrange el ACC se convierte a un problema de valores y vectores propios generalizado, suponiendo que la matrices de varianzas y covarianzas son invertibles. Por otro lado, los Algoritmos Genéticos (AG) son métodos adaptativos que se usan para resolver problemas de optimización global. Cuando la condición de invertibilidad no se cumple en las matrices de varianzas y covarianzas, se propone en este trabajo, usar AG para resolver el problema del ACC directamente de la definición. Se prueba la propuesta con problemas publicados en la literatura y se presenta además una aplicación del ACC a datos de carbono en el suelo de la zona de Teziutlán, Puebla.

1. INTRODUCCIÓN

El Análisis de Correlación Canónica (ACC) es un método del Análisis multivariado que se resuelve usando técnicas de Álgebra Lineal Numérica. Su objetivo principal es encontrar las correlaciones entre dos conjuntos de variables cuantitativas observadas sobre el mismo conjunto de unidades experimentales, a través de combinaciones lineales de las variables iniciales, lo que permite reducir la dimensionalidad [4]. El problema de las correlaciones Canónicas no es más que un problema de máximos con restricciones, por lo que puede resolverse aplicando el método de los multiplicadores de Lagrange que conlleva a un problema de valores y vectores propios generalizado según los trabajos de Mardia et al. [8], Seber [10] y Hardoon et al. [5]. Mardia et al. [8] calculan las matrices de varianzas y covarianzas de los datos y posteriormente realizan la

*caty_b26@hotmail.com

descomposición en valores singulares para obtener los valores y vectores propios (correlaciones y vectores canónicos). Por su parte, Seber [10] descompone directamente las matrices de datos usando descomposición QR para calcular los valores y vectores propios, mientras que Haroon et al. [5] una vez que llegan al problema de valores y vectores propios generalizado utilizan la descomposición de Cholesky para las matrices de varianzas y covarianzas resolviendo un problema de valores y vectores propios simple. Cabe resaltar que para poder encontrar las correlaciones Canónicas y los vectores canónicos mediante los métodos utilizados en la literatura es necesario hacer la suposición de invertibilidad en las matrices de varianzas y covarianzas de los datos. En este trabajo se tiene como propuesta utilizar el método de Algoritmos Genéticos (AG) para encontrar la correlación máxima entre dos grupos de variables sujeta a restricciones, directamente de la definición de correlación Canónica, ya que AG es un método de optimización que ayuda a aproximar los óptimos globales de una función que no requiere la invertibilidad de las matrices de varianzas y covarianzas.

En la primera sección se presenta al ACC clásico y tres enfoques para darle solución a partir del problema de valores y vectores propios generalizado. En la siguiente sección se presenta una alternativa a la solución del ACC usando algoritmos Genéticos en la ecuación de valores y vectores propios generalizado. Posteriormente, se presentan algunos ejemplos de la literatura para realizar comparaciones entre los resultados usando los diferentes métodos y un ejemplo real en el que una de las matrices de varianzas y covarianzas no es invertible. Después se usa el método AG para resolver directamente la ecuación de correlaciones Canónicas con restricciones que resuelve el problema de no invertibilidad, probándose con los ejemplos presentados en la sección anterior. Finalmente se presenta la discusión del trabajo y la bibliografía.

2. ANÁLISIS DE CORRELACIÓN CANÓNICA

En el Análisis de Correlación Canónica se está interesado en las medidas de asociación entre dos grupos de variables. El primer grupo, de p variables, está representado por el vector aleatorio \mathbf{X} de orden $(p \times 1)$ y el segundo grupo, de q variables, está representado por el vector aleatorio \mathbf{Y} de orden $(q \times 1)$. Se asume que, $p \leq q$.

Para los vectores \mathbf{X} y \mathbf{Y} se calculan:

$$\begin{aligned} E[\mathbf{X}] &= \mu_X, \text{Var}(\mathbf{X}) = \Sigma_{XX} \\ E[\mathbf{Y}] &= \mu_Y, \text{Var}(\mathbf{Y}) = \Sigma_{YY} \\ \text{Cov}(\mathbf{X}, \mathbf{Y}) &= \Sigma_{XY} = \Sigma'_{YX}. \end{aligned}$$

Sean U y V combinaciones lineales que proporcionan medidas simples de los conjuntos de variables.

$$\begin{aligned} U &= \mathbf{a}'\mathbf{X}, \\ V &= \mathbf{b}'\mathbf{Y} \end{aligned}$$

donde \mathbf{a} y \mathbf{b} son vectores de coeficientes. Además, sean

$$\begin{aligned} \text{Var}(U) &= \mathbf{a}'\Sigma_{XX}\mathbf{a} \\ \text{Var}(V) &= \mathbf{b}'\Sigma_{YY}\mathbf{b} \\ \text{Cov}(U, V) &= \mathbf{a}'\Sigma_{XY}\mathbf{b}. \end{aligned}$$

Se requiere buscar los vectores de coeficientes \mathbf{a} y \mathbf{b} tales que:

$$\rho = \text{Corr}(U, V) = \frac{\mathbf{a}'\Sigma_{XY}\mathbf{b}}{\sqrt{\mathbf{a}'\Sigma_{XX}\mathbf{a}}\sqrt{\mathbf{b}'\Sigma_{YY}\mathbf{b}}} \quad (2..1)$$

sea máxima. Sujeto a

$$\mathbf{a}'\Sigma_{XX}\mathbf{a} = \mathbf{b}'\Sigma_{YY}\mathbf{b} = 1. \quad (2..2)$$

Los vectores de coeficientes \mathbf{a} y \mathbf{b} que cumplen esta condición son los primeros vectores canónicos. La máxima correlación entre (U, V) es la primera correlación Canónica $\rho(1)$ [3].

Como se puede observar, las ecuaciones (2.1) y (2.2) plantean un problema de máximos con restricciones, entonces para encontrar su solución se ha utilizado, tradicionalmente, el método de multiplicadores de Lagrange. Considérese la función de Lagrange

$$F(\mathbf{a}, \mathbf{b}, \lambda, \beta) = \mathbf{a}'\Sigma_{XY}\mathbf{b} - \frac{\lambda}{2}(\mathbf{a}'\Sigma_{XX}\mathbf{a} - 1) - \frac{\beta}{2}(\mathbf{b}'\Sigma_{YY}\mathbf{b} - 1)$$

donde λ y μ son multiplicadores de Lagrange. Para obtener el valor máximo se deben cumplir las condiciones de Karush-Kuhn-Tucker, obteniéndose:

$$\frac{\partial F}{\partial \mathbf{a}} = \Sigma_{XY}\mathbf{b} - \lambda\Sigma_{XX}\mathbf{a} = 0 \quad (2.3)$$

$$\frac{\partial F}{\partial \mathbf{b}} = \Sigma_{YX}\mathbf{a} - \beta\Sigma_{YY}\mathbf{b} = 0 \quad (2.4)$$

$$\frac{\partial F}{\partial \lambda} = -\frac{1}{2}(\mathbf{a}'\Sigma_{XX}\mathbf{a} - 1) = 0, \quad \frac{\partial F}{\partial \beta} = -\frac{1}{2}(\mathbf{b}'\Sigma_{YY}\mathbf{b} - 1) = 0. \quad (2.5)$$

Las ecuaciones (2.5) recuperan las restricciones. Multiplicando (2.3) por \mathbf{a}' y (2.4) por \mathbf{b}' , se tiene

$$\begin{aligned} \mathbf{a}'\Sigma_{XY}\mathbf{b} - \lambda\mathbf{a}'\Sigma_{XX}\mathbf{a} &= 0 \\ \mathbf{b}'\Sigma_{YX}\mathbf{a} - \beta\mathbf{b}'\Sigma_{YY}\mathbf{b} &= 0 \end{aligned}$$

entonces $\lambda = \beta = \mathbf{a}'\Sigma_{XY}\mathbf{b}$.

Despejando a \mathbf{b} de (2.4) se obtiene

$$\lambda\Sigma_{YY}\mathbf{b} = \Sigma_{YX}\mathbf{a} \Rightarrow \mathbf{b} = \lambda^{-1}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} \quad (2.6)$$

suponiendo que Σ_{YY} es invertible. Ahora se sustituye \mathbf{b} en la ecuación (2.3), entonces se obtiene

$$\begin{aligned} \Sigma_{XY}(\lambda^{-1}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}) - \lambda\Sigma_{XX}\mathbf{a} &= 0 \\ \Rightarrow \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a} &= \lambda^2\Sigma_{XX}\mathbf{a}. \end{aligned} \quad (2.7)$$

Como se puede observar, es un problema de valores y vectores propios generalizado, el cuál requiere que Σ_{YY} sea invertible. De manera similar se trabaja para \mathbf{b} , despejando \mathbf{a} de (2.3) y sustituyendo en (2.4), se obtiene

$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{b} = \lambda^2\Sigma_{YY}\mathbf{b}. \quad (2.8)$$

Note que una vez obtenidos los valores de \mathbf{a} y λ se calcula el valor de \mathbf{b} mediante la ecuación (2.6).

A continuación se presentan tres métodos, mencionados en la literatura, para determinar las correlaciones Canónicas y vectores canónicos a partir de la ecuación (2.7).

2.1. ACC usando descomposición en valores singulares

Este método es presentado por Mardia et al. [8], quien define una matriz \mathbf{K} , le aplica el teorema de descomposición en valores singulares llegando a un problema de vectores y valores propios.

Bajo la suposición de que Σ_{XX} y Σ_{YY} son invertibles, sea

$$\mathbf{K} = \Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2},$$

tomando las inversas de Σ_{XX} y Σ_{YY} en las ecuaciones (2.7) y (2.8) respectivamente, se definen,

$$\mathbf{M}_1 = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} = \Sigma_{XX}^{-1/2}\mathbf{N}_1\Sigma_{XX}^{1/2}, \quad (2.9)$$

$$\mathbf{M}_2 = \Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} = \Sigma_{YY}^{-1/2}\mathbf{N}_2\Sigma_{YY}^{1/2}. \quad (2.10)$$

donde $\mathbf{N}_1 = \mathbf{K}\mathbf{K}'$ y $\mathbf{N}_2 = \mathbf{K}'\mathbf{K}$. Se tiene que \mathbf{N}_1 es definida positiva y tiene los mismos valores propios que \mathbf{N}_2 y estos son positivos. Sea $k = \text{rank}(\mathbf{K}) = \text{rank}(\Sigma_{XY})$ el número de valores propios diferentes de cero. Por el teorema de descomposición en valores singulares, se tiene que:

$$\mathbf{K} = (\mathbf{x}_1, \dots, \mathbf{x}_k)\mathbf{D}(\mathbf{y}_1, \dots, \mathbf{y}_k)'$$

donde \mathbf{x}_i y \mathbf{y}_i son vectores propios ortogonales estandarizados de \mathbf{N}_1 y \mathbf{N}_2 , respectivamente, ya que

$$\mathbf{N}_1 = \mathbf{K}\mathbf{K}' = (\mathbf{x}_1, \dots, \mathbf{x}_k)\mathbf{D}^2(\mathbf{x}_1, \dots, \mathbf{x}_k)'$$

e igualmente para \mathbf{N}_2 . Además \mathbf{K} tiene valores propios λ_i , contenidos en \mathbf{D} , donde los valores propios para \mathbf{N}_1 y \mathbf{N}_2 , son λ_i^2 . Se tiene el siguiente problema de valores y vectores propios,

$$\begin{aligned} \mathbf{N}_1\mathbf{x}_i &= \lambda_i^2\mathbf{x}_i \\ \Sigma_{XX}^{-1/2}\mathbf{N}_1\Sigma_{XX}^{1/2}\Sigma_{XX}^{-1/2}\mathbf{x}_i &= \lambda_i^2\Sigma_{XX}^{-1/2}\mathbf{x}_i \\ \mathbf{M}_1\mathbf{a}_i &= \lambda_i^2\mathbf{a}_i \end{aligned}$$

donde

$$\mathbf{a}_i = \Sigma_{XX}^{-1/2}\mathbf{x}_i.$$

se deduce que \mathbf{N}_1 y \mathbf{M}_1 tienen los mismos valores propios y además los vectores propios de \mathbf{N}_1 y \mathbf{M}_1 estén relacionadas por las ecuaciones anteriores. De igual forma se obtiene $\mathbf{b}_i = \Sigma_{YY}^{-1/2}\mathbf{y}_i$. En general, se resolvieron dos problemas de valores y vectores propios de la forma $\mathbf{A}\theta = \rho\theta$ donde en uno $\mathbf{A} = \mathbf{M}_1$, $\rho = \lambda^2$ y $\theta_i = \mathbf{a}_i = \Sigma_{XX}^{-1/2}\mathbf{x}_i$ y para el otro $\mathbf{A} = \mathbf{M}_2$, $\rho = \lambda^2$ y $\theta_i = \mathbf{b}_i = \Sigma_{XX}^{-1/2}\mathbf{y}_i$.

2.2. ACC usando descomposición de Cholesky

A continuación se presenta la propuesta de Haroon et al. [5] que da solución a la ecuación (2.7) que representa un problema de valores y vectores propios generalizado y para ello se propone utilizar el método de descomposición de Cholesky sobre las matrices de covarianzas para resolver un problema de valores y vectores propios simple. Supóngase que Σ_{XX} es invertible, entonces dado que Σ_{XX} es simétrica y definida positiva, se puede descomponer, de acuerdo al Teorema de Cholesky, como:

$$\Sigma_{XX} = \mathbf{R}_{XX}\mathbf{R}'_{XX} \quad (2.11)$$

donde \mathbf{R}_{XX} es una matriz triangular inferior. Se define

$$\mathbf{w}_x = \mathbf{R}'_{XX}\mathbf{a} \Rightarrow \mathbf{R}_{XX}^{-1'}\mathbf{w}_x = \mathbf{a}. \quad (2.12)$$

Entonces sustituyendo (2.12) en (2.7) se obtiene

$$\Rightarrow \mathbf{R}_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{R}_{XX}^{-1'}\mathbf{w}_x = \lambda^2\mathbf{w}_x,$$

Similarmente se realiza el mismo procedimiento para determinar a \mathbf{b} usando la ecuación (2.8). Por lo tanto, se resolvió un problema de valores y vectores propios simple de la forma $\mathbf{A}\theta = \rho\theta$ donde $\mathbf{A} = \mathbf{R}_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{R}_{XX}^{-1'}$, $\rho = \lambda^2$ y $\theta = \mathbf{w}_x = \mathbf{R}'_{XX}\mathbf{a}$.

2.3. ACC usando factorización QR

Otra forma de obtener los valores y vectores propios es mediante el uso de un algoritmo de descomposición **QR**, presentado en Seber [10].

Como primer paso, usando (2.11) se tiene que $\Sigma_{XX} = \mathbf{R}_{XX}\mathbf{R}'_{XX}$ y sea $\Sigma_{YY} = \mathbf{R}_{YY}\mathbf{R}'_{YY}$ donde \mathbf{R}_{XX} y \mathbf{R}_{YY} son matrices triangulares inferiores, se define

$$\mathbf{C} = \mathbf{R}_{XX}^{-1}\Sigma_{XY}(\mathbf{R}'_{YY})^{-1} \quad (2.13)$$

entonces (2.9) y (2.10) se pueden reescribir como sigue

$$\mathbf{M}_1 = \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} = (\mathbf{R}'_{XX})^{-1} \mathbf{C} \mathbf{C}' \mathbf{R}'_{XX}$$

$$\mathbf{M}_2 = \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} = (\mathbf{R}'_{YY})^{-1} \mathbf{C}' \mathbf{C} \mathbf{R}'_{YY}.$$

Dado que \mathbf{M}_1 con $\mathbf{C} \mathbf{C}'$ y \mathbf{M}_2 con $\mathbf{C}' \mathbf{C}$ son transformaciones de similitud, entonces tienen los mismos valores propios.

Por otro lado, si se toma una muestra de observaciones \mathbf{x} y \mathbf{y} , donde $E[\mathbf{x}] = \bar{\mathbf{x}}$ y $E[\mathbf{y}] = \bar{\mathbf{y}}$, los vectores centrados son: $\bar{\mathbf{X}} = (\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}})$ y $\bar{\mathbf{Y}} = (\mathbf{y}_1 - \bar{\mathbf{y}}, \mathbf{y}_2 - \bar{\mathbf{y}}, \dots, \mathbf{y}_n - \bar{\mathbf{y}})$. Además se tiene que

$$\begin{aligned} Var(\bar{\mathbf{X}}) &= \bar{\mathbf{X}}' \bar{\mathbf{X}} = \mathbf{Q}_{XX}, & Var(\bar{\mathbf{Y}}) &= \bar{\mathbf{Y}}' \bar{\mathbf{Y}} = \mathbf{Q}_{YY}, \\ Cov(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) &= \bar{\mathbf{X}}' \bar{\mathbf{Y}} = (\bar{\mathbf{Y}}' \bar{\mathbf{X}})' = \mathbf{Q}_{XY}. \end{aligned}$$

Usando el algoritmo de descomposición \mathbf{QR} , se tiene que

$$\bar{\mathbf{X}} = \mathbf{Q}_{d1} \mathbf{T}_{XX} \quad \text{y} \quad \bar{\mathbf{Y}} = \mathbf{Q}_{d2} \mathbf{T}_{YY},$$

donde \mathbf{Q}_{d1} y \mathbf{Q}_{d2} , son matrices ortogonales y \mathbf{T}_{XX} y \mathbf{T}_{YY} son matrices triangulares superiores. Las matrices de varianzas y covarianzas se reescriben de la siguiente forma

$$\mathbf{Q}_{XX} = \bar{\mathbf{X}}' \bar{\mathbf{X}} = (\mathbf{Q}_{d1} \mathbf{T}_{XX})' (\mathbf{Q}_{d1} \mathbf{T}_{XX}) = \mathbf{T}'_{XX} \mathbf{T}_{XX}$$

y

$$\mathbf{Q}_{YY} = \bar{\mathbf{Y}}' \bar{\mathbf{Y}} = (\mathbf{Q}_{d2} \mathbf{T}_{YY})' (\mathbf{Q}_{d2} \mathbf{T}_{YY}) = \mathbf{T}'_{YY} \mathbf{T}_{YY}.$$

Por otra parte, se toma a \mathbf{C} como en (2.13), entonces se determina a $\bar{\mathbf{C}}$ cuando se sustituye \mathbf{T}_{XX} en lugar de \mathbf{R}'_{XX} y cuando se usan las matrices de covarianzas muestrales. Entonces

$$\begin{aligned} \bar{\mathbf{C}} &= (\mathbf{T}'_{XX})^{-1} \mathbf{Q}_{XY} \mathbf{T}_{YY}^{-1} \\ &= \mathbf{Q}'_{d1} \mathbf{Q}_{d2}. \end{aligned}$$

Ahora considere la descomposición en valores singulares para $\bar{\mathbf{C}}$ y $\bar{\mathbf{C}}'$

$$\begin{aligned} \bar{\mathbf{C}} &= \mathbf{S} \mathbf{D} \mathbf{V} \\ \bar{\mathbf{C}}' &= \mathbf{V}' \mathbf{D}' \mathbf{S}' \end{aligned}$$

donde \mathbf{V} y \mathbf{S} son matrices ortogonales y \mathbf{D} es una matriz diagonal. Entonces los valores propios de $\bar{\mathbf{C}}$ son $\lambda = \mathbf{D}$. Como se puede observar, los valores propios de $\bar{\mathbf{C}}'$ son los mismos que $\bar{\mathbf{C}}$, además se tiene que los valores propios de $\bar{\mathbf{C}} \bar{\mathbf{C}}'$ y $\bar{\mathbf{C}}' \bar{\mathbf{C}}$ son \mathbf{D}^2 . Los vectores propios de $\bar{\mathbf{C}} \bar{\mathbf{C}}'$ son los representados por la matriz \mathbf{S} , ya que

$$\bar{\mathbf{C}} \bar{\mathbf{C}}' = \mathbf{S} \mathbf{D} \mathbf{V} \mathbf{V}' \mathbf{D}' \mathbf{S}' = \mathbf{S} \mathbf{D}^2 \mathbf{S}',$$

se argumenta lo mismo para $\bar{\mathbf{C}}' \bar{\mathbf{C}}$, y sus vectores propios estén dados por \mathbf{V} . Entonces se resolvieron dos problemas de valores y vectores propios de la forma $\mathbf{A} \theta = \rho \theta$ donde en uno $\mathbf{A} = \mathbf{C} \mathbf{C}'$, $\rho = \lambda^2$ y $\theta = \mathbf{v}_i$ y para el otro $\mathbf{A} = \mathbf{C}' \mathbf{C}$, $\rho = \lambda^2$ y $\theta = \mathbf{s}$.

Para cada uno de los métodos presentados en las subsecciones anteriores se realizó un programa usando el software OCTAVE [9]. Posteriormente se presentan ejemplos en donde se hace una comparación de los resultados obtenidos mediante el programa de cada método.

3. SOLUCIÓN DEL ACC MEDIANTE ALGORITMOS GENÉTICOS

A continuación se presenta una alternativa para la solución del Análisis de correlaciones Canónicas clásico, esta es, usando el método de algoritmos genéticos. En el ACC se requieren encontrar los vectores ρ , \mathbf{a} y \mathbf{b} . Siguiendo la teoría presentada en la sección anterior, se llegó a un problema de valores y vectores propios generalizado, dado en la ecuación (2.7). Para trabajar con algoritmos Genéticos se debe tener una función objetivo. Para el problema de ACC, se tendría la siguiente función objetivo

$$F(\mathbf{a}, \lambda) = \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \mathbf{a} - \lambda^2 \Sigma_{XX} \mathbf{a} = 0 \quad (3.14)$$

que es equivalente a resolver

$$\min F'(\mathbf{a}, \lambda) F(\mathbf{a}, \lambda) \quad (3.15)$$

donde se debe cumplir la siguiente condición $0 < \lambda \leq 1$.

3.1. Algoritmos genéticos

Un algoritmo genético es un método que imita la teoría de la evolución biológica de Darwin y es usado para resolver problemas de optimización global. Esto es, el contenido genético de una población contiene potencialmente la solución, o una solución mejor, a un problema dado de adaptación. Los genes mejor adaptados representarían la mejor solución al problema de optimización. Para encontrarla se debe cambiar la información genética usando mutación y cruza, y seleccionar la mejor población [2].

De manera general en el mecanismo evolutivo se tiene lo siguiente:

Cromosoma. Los individuos (posibles soluciones del problema), pueden representarse como un conjunto de parámetros, denominados genes, los cuales agrupados forman una cadena o serie de valores, referida como cromosoma. Un ejemplo de cromosoma es el siguiente

$$[6 \ 1 \ 9 \ 4 \ 7 \ 0 \ 0 \ 7 \ 5 \ 7 \ 0 \ 3 \ 5 \ 4 \ 7]$$

Función de adaptación o fitness. La función de adaptación debe ser diseñada para cada problema de manera específica. Esta función asigna un número real a un cromosoma particular, el cual refleja el nivel de adaptación del individuo al problema representado por el cromosoma.

Fenotipo. Es el conjunto de números reales representando la solución del problema de optimización.

Genotipo. Es la representación de la solución real codificada en un cromosoma.

La transformación fenotipo genotipo y viceversa es mostrada en la Figura 1.

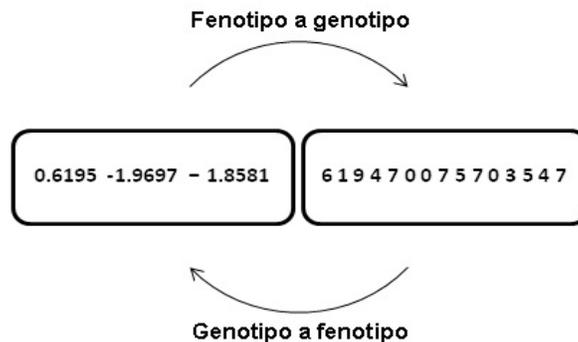


Figure 1: Transformación fenotipo-genotipo.

Sobre esta población elegida, algunos individuos son seleccionados aleatoriamente para reproducirse. Durante la reproducción, los nuevos individuos de la población resultan de modificaciones e intercambio genético de los padres. Hay esencialmente dos operadores genéticos.

Mutación. Consiste de simplemente cambiar un gen en el cromosoma por algún otro elegido aleatoriamente (usualmente con probabilidad pequeña), como en el siguiente ejemplo

$$\begin{bmatrix} 6 & 1 & 9 & 4 & 7 & 0 & 0 & 7 & 5 & 7 & 0 & 3 & 5 & 4 & 7 \\ & & & \downarrow & & & & & & & & & & & \\ 6 & 1 & 9 & 8 & 7 & 0 & 0 & 7 & 5 & 7 & 0 & 3 & 5 & 4 & 7 \end{bmatrix}.$$

Cruza. Es una recombinación de la información durante la reproducción de los individuos seleccionados. Se eligen dos padres seleccionados y se cortan sus cadenas de cromosomas en una posición elegida al azar, para producir dos subcadenas iniciales y dos subcadenas finales. Después se intercambian las subcadenas iniciales, produciéndose dos nuevos cromosomas completos. Ambos descendientes heredan genes de cada uno de los padres. Esto es presentado a continuación.

Antes de la cruce

$$\left[\begin{array}{cccccc|cccccccc} 6 & 1 & 9 & 4 & 7 & & 0 & 0 & 7 & 5 & 7 & 0 & 3 & 5 & 4 & 7 \\ 7 & 2 & 5 & 8 & 3 & & 1 & 0 & 6 & 2 & 3 & 9 & 7 & 2 & 1 & 7 \end{array} \right].$$

Después de la cruce

$$\left[\begin{array}{cccccc|cccccccc} 7 & 2 & 5 & 8 & 3 & & 0 & 0 & 7 & 5 & 7 & 0 & 3 & 5 & 4 & 7 \\ 6 & 1 & 9 & 4 & 7 & & 1 & 0 & 6 & 2 & 3 & 9 & 7 & 2 & 1 & 7 \end{array} \right].$$

Una vez que se renueva la población, el proceso recomienza.

El presente trabajo se basa en la propuesta de Ison et al. [6] donde introducen los algoritmos Genéticos que resuelven el problema de minimización de funciones de dos variables utilizando MATLAB. El programa es llamado GENETICO con las siguientes características: se define como función de adaptación al recíproco de la función objetivo; el espacio de soluciones son todas las parejas ordenadas (x,y) las cuales serán los fenotipos que posteriormente son representados por los genotipos que son vectores de 8 entradas y cada entrada es un dígito entre 0 y 9, este genotipo representa nuestro cromosoma y posible solución del problema. En concreto se tiene una transformación de un vector de dos variables de números reales a un vector de 8 entradas con dígitos del 0 al 9. A estos cromosomas se les aplica la mutación y cruce. Una vez mejorada la evaluación se aplica la transformación inversa para obtener la solución aproximada (fenotipo solución).

3.2. Programa elaborado para resolver el ACC mediante AG en el formalismo de Lagrange

Como se vió en (2..7), se quiere determinar los valores de ρ , \mathbf{a} y \mathbf{b} , en donde no es necesario incluir a \mathbf{b} , ya que se encuentra en términos de \mathbf{a} . En este caso, para resolver el ACC, se define a la función (3..14), de la cual se requiere minimizar (3..15). Como se puede observar, es necesario determinar las matrices de varianzas y covarianzas entre los grupos de variables ya que son necesarios para la función. En particular, en el problema se requiere encontrar valores ρ y \mathbf{a} tal que (3..14) sea máxima, para ello se toma como función de adaptación $1/F'F$. El programa GENETICO como tal no se puede aplicar directamente al problema, ya que este es para encontrar los valores óptimos de funciones de dos variables (es decir, se toma un cromosoma con 2 genes), sin embargo, en el ACC los dos grupos de variables X y Y tienen 2 o más variables, por lo que si N es el Número de variables de X , entonces el cromosoma tendrá $N + 1$ genes.

Posteriormente se elige una población inicial aleatoria y se le realiza el procedimiento descrito en la subsección anterior. Cabe aclarar que para determinar los mejores resultados o los mejores valores se utilizaron dos criterios, el primero es tomar los valores con menor error de $F'F$ y el segundo es verificar las restricciones dadas en la ecuación (2..2), es decir, se tomarán los valores que tengan un error pequeño.

A continuación, en la Tabla 1 se presenta el pseudocódigo utilizado para este trabajo

Se realizó un algoritmo en el software OCTAVE [9] para encontrar las correlaciones Canónicas y los vectores canónicos usando algoritmos Genéticos siguiendo el pseudocódigo anterior. El programa se corrió en una computadora con procesador Intel(R)Core(TM)i525000 @ 3.30 GHz con memoria 8.00 GB RAM. Para obtener los mejores resultados se llevó a cabo un procedimiento de calibración para los parámetros: tamaño de la población, número de generación, razón de mutación y razón de cruce. El procedimiento consiste en fijar tres parámetros mientras el cuarto es corrido en un rango de valores de posibles mejores soluciones. Por ejemplo, primero se fijan los parámetros

Número de generaciones = 300

Razón de mutación = 0.2

Pseudocódigo
1) Definir los parámetros (función de adaptación, tamaño de la población, razón de cruza, razón de mutación, número de generaciones).
2) Generar una población inicial aleatoriamente de posibles soluciones.
3) Evaluar la función de adaptación.
4) Obtener los genotipos de la población.
5) Realizar mutación y cruza de acuerdo a la razón de cruza y mutación.
6) Determinar la nueva población.
7) Transformar el genotipo al fenotipo de la nueva población.
8) Se regresa al paso 3 hasta llegar al Número de generaciones.

Table 1: Pseudocódigo del algoritmo genético.

Razón de cruza = 0.8

y el programa es ejecutado iterando el tamaño de la población en el rango entre 10 y 100 (con lapsos de 10 en 10). Supóngase que la mejor aproximación a la solución fue obtenida en

Tamaño de la población = 20,

posteriormente se realizará lo mismo para los otros parámetros corriendo el programa con: Número de generación entre 100 y 1500 (con lapsos de 100); ,de mutación entre 0.1 y 0.3 (con lapsos de 0.01); razón de cruza entre 0.7 y 0.9 (con lapsos de 0.01), obteniéndose los mejores valores para cada parámetro.

3.3. Ejemplos

Se utilizaron diversos ejemplos para determinar las correlaciones y vectores canónicos usando algoritmos Genéticos y se compararon con los resultados para los tres métodos aplicados anteriormente (SVD, Cholesky y QR).

3.3.1. Ejemplo datos de cabezas

A continuación se presenta un ejemplo publicado por Mardia et al. [8]. Se consideran los datos de $n = 25$ familias para las variables en donde se tomaron medidas de la cabeza del primer y segundo hijo (Tabla 2).

Primer grupo de variables (X)	Segundo grupo de variables (Y)
Longitud de la cabeza del primer hijo	Ancho de la cabeza del primer hijo
Longitud de la cabeza del segundo hijo	Ancho de la cabeza del segundo hijo

Table 2: Variables de los datos de cabezas

Como se puede ver, cada grupo tiene dos variables, entonces el cromosoma tendrá tres genes representando a λ y al vector propio propio \mathbf{a} con dos entradas. Por ejemplo, el fenotipo

$$[\lambda \quad \mathbf{a}_1 \quad \mathbf{a}_2] = [0.4368 \quad 1.4397 \quad 1.0553]$$

es transformado al siguiente genotipo

$$[\lambda \quad \mathbf{a}_1 \quad \mathbf{a}_2] = [4 \quad 3 \quad 6 \quad 7 \quad 8 \quad 5 \quad 0 \quad 7 \quad 1 \quad 9 \quad 5 \quad 0 \quad 5 \quad 2 \quad 7]$$

haciendo una codificación de cinco dígitos por gen.

Se calibró el algoritmo, obteniéndose lo siguientes parámetros

Número de generaciones = 1400

Tamaño de la población = 20

Razón de mutación = 0.06

Razón de cruza = 0.65

Una vez calibrado el programa, este se corrió para obtener la correlación Canónica y los vectores canónicos. Se aplicaron los tres métodos del ACC a este ejemplo para calcular las correlaciones y vectores canónicos y además se aplicó el método de algoritmos genéticos; para llevar a cabo este Análisis se usaron los programas en OCTAVE de cada método. Los resultados son presentados en la Tabla 3 en donde ρ_1 y ρ_2 representan la primera y segunda correlación Canónica, \mathbf{a}_1 con \mathbf{b}_1 los primeros vectores canónicos y \mathbf{a}_2 con \mathbf{b}_2 los segundos vectores canónicos.

Método	Cor. Canónica		Vectores canónicos							
	ρ_1	ρ_2	\mathbf{a}_1		\mathbf{b}_1		\mathbf{a}_2		\mathbf{b}_2	
ACC	0.78855	0.05382	0.55215	0.52155	0.50450	0.53824	-1.36650	1.37846	-1.76827	1.75829
Chol	0.78855	0.05382	-0.55215	-0.52155	-0.50450	-0.53824	-1.36650	1.37846	-1.76827	1.75829
SVD	0.78850	0.05373	-0.55218	-0.52153	-0.50444	-0.53828	-1.36637	1.37836	-1.76856	1.75856
QR	0.78866	0.05310	-0.55800	-0.52799	-0.51016	-0.54436	1.33800	-1.35000	1.75302	-1.74716

Table 3: Correlaciones y vectores canónicos de los datos de cabezas.

Como se puede observar, los resultados entre los métodos propuestos en la literatura con los resultados obtenidos mediante el programa de algoritmos Genéticos son similares, y además concuerdan con los publicados por Mardia [8].

3.3.2. Ejemplo de datos mmreg

Se presenta a continuación un ejemplo aplicado a datos almacenados en el archivo mmreg.dat¹. Un investigador ha recopilado datos sobre tres variables psicológicas y cuatro variables académicas para 600 estudiantes de primer año de universidad. Está interesado en la forma en que el conjunto de variables psicológicas se relaciona a las variables académicas. En particular, el investigador está interesado en cuántas dimensiones (variables Canónicas) son necesarios para entender la asociación entre los dos conjuntos de variables. Las variables están en la Tabla 4

Primer grupo de variables (X)	Segundo grupo de variables (Y)
Centro de control	Lectura
Autoconcepto	Escritura
Motivación	Matemáticas
	Ciencia

Table 4: Variables de los datos mmreg

Nuevamente se calibró el algoritmo para estos datos y se obtuvieron los parámetros siguientes

Número de generaciones=200

Tamaño de la población=60

Razón de mutación=0.17

Razón de cruce=0.75

A continuación se presentan las tres correlaciones Canónicas y sus vectores canónicos obtenidos usando los métodos clásicos para el ACC y los mejores resultados obtenidos usando algoritmos Genéticos (ver Tabla 5). Se observa que los resultados obtenidos mediante cada método son similares con errores muy pequeños.

En los dos ejemplos presentados se realizan las comparaciones con cada método para verificar el buen funcionamiento del programa de algoritmos genéticos. A continuación se presenta un problema real al que se requiere encontrar el grado de relación entre los dos grupos de variables.

3.3.3. Ejemplo de datos de carbono

Dado que no existe información sobre secuestro de carbono en sistemas forestales en el Estado de Puebla, México, y en particular, en la zona de Teziutlán, se hizo necesario iniciar estudios investigativos pertinentes.

¹<http://www.ats.ucla.edu/stat/r/dae/canonical.htm>

Método	Cor. Canónica		Vectores canónicos						
	ρ		a			b			
Chol	ρ_1	0.4462	0.8388	-0.1664	0.4265	0.4457	0.5337	0.1832	-0.0358
	ρ_2	0.1525	-0.5116	-0.5955	0.9035	-0.0214	-0.8751	-0.0323	1.2101
	ρ_3	0.0228	-0.3332	0.8493	0.3761	-0.8921	0.9401	-0.8263	0.8525
SVD	ρ_1	0.4462	-0.8388	0.1664	-0.4265	-0.4457	-0.5337	-0.1832	0.0358
	ρ_2	0.1525	-0.5116	-0.5955	0.9035	0.0214	0.8751	0.0323	-1.2101
	ρ_3	0.0228	-0.3332	0.8493	0.3761	0.8921	-0.9401	0.8263	-0.8525
QR	ρ_1	0.4464	0.8379	-0.1670	0.4281	0.4450	-0.5358	0.1826	-0.0368
	ρ_2	0.1533	-0.5134	-0.5941	0.9034	0.0160	0.8794	0.0278	-1.2055
	ρ_3	0.0225	-0.3328	0.8502	0.3747	0.8924	-0.9349	0.8268	-0.8589
AG	ρ_1	0.4472	-0.8380	0.1680	-0.4260	-0.4441	-0.5322	-0.1826	0.0359
	ρ_2	0.1525	0.5200	0.5980	-0.9180	-0.0233	-0.8866	-0.0338	1.2269
	ρ_3	0.0228	-0.3280	0.8480	0.3680	0.8869	-0.9755	0.8214	-0.7969

Table 5: Correlaciones y vectores canónicos de los datos mmreg.

El problema que se planteó investigar es que, como consecuencia de las variaciones climáticas ocurridas en los suelos que soportan vegetación forestal, en la región de Teziutlán, se considera que también han ocurrido variaciones en el contenido del carbono orgánico en esos suelos. Por ello, se aplicaron determinadas técnicas para realizar la medición correspondiente del contenido de carbono orgánico, en esos suelos, en muestras seleccionadas durante los años 1987 y 2009 [1], [7]. Se midieron variables que caracterizan la textura de los suelos (grupo X en la Tabla 6) y otras variables que caracterizan la fertilidad de los suelos (grupo Y en la Tabla 6) con el objetivo de evaluar el grado de asociación entre esos grupos de variables.

Primer grupo de variables (X)	Segundo grupo de variables (Y)
Densidad aparente	Porcentaje de materia orgánica
Porcentaje de arena	Porcentaje de carbono orgánico
Porcentaje de limo	Carbono orgánico en suelos
Porcentaje de arcilla	Porcentaje de nitrógeno total
	Relación carbono nitrógeno

Table 6: Variables de los datos de carbono.

Como se desea tener una medida del grado de asociación entre estos dos grupos de variables, se decide realizar el ACC usando los programas de los métodos de Cholesky, SVD y descomposición QR, sin embargo, no se pudo realizar este análisis, ya que la matriz de varianzas y covarianzas de los datos del grupo X no es definida positiva y por lo tanto no se puede aplicar el teorema de Cholesky a la matriz. Por otro lado, la matriz de varianzas y covarianzas del grupo Y no es invertible, como se puede ver en la versión corta de la matriz de varianzas y covarianzas en (3.16), mismo efecto que se puede observar en la matriz de correlaciones (3.17) del grupo Y (denotada por R_{yy}). Este hecho hace que la matriz de varianzas y covarianzas no cumpla con la condición de invertibilidad que debe cumplirse para cada uno de los métodos.

$$\Sigma_{yy} = \begin{bmatrix} 2.0292e+001 & 1.1770e+001 & 1.6611e+002 & 6.7708e-001 & 1.8419e+000 \\ 1.1770e+001 & 6.8273e+000 & 9.6352e+001 & 3.9274e-001 & 1.0684e+000 \\ 1.6611e+002 & 9.6352e+001 & 8.5190e+003 & 1.8940e+000 & -1.5539e+001 \\ 6.7708e-001 & 3.9274e-001 & 1.8940e+000 & 4.1738e-001 & 3.1020e-001 \\ 1.8419e+000 & 1.0684e+000 & -1.5539e+001 & 3.1020e-001 & 9.3903e+000 \end{bmatrix}, \quad (3.16)$$

$$R_{yy} = \begin{bmatrix} 1.0000 & 1.0000 & 0.3995 & 0.2327 & 0.1334 \\ 1.0000 & 1.0000 & 0.3995 & 0.2327 & 0.1334 \\ 0.3995 & 0.3995 & 1.0000 & 0.0318 & -0.0549 \\ 0.2327 & 0.2327 & 0.0318 & 1.0000 & 0.1567 \\ 0.1334 & 0.1334 & -0.0549 & 0.1567 & 1.0000 \end{bmatrix}. \quad (3.17)$$

Dado este problema, se propone entonces resolver el ACC usando algoritmos Genéticos directamente de la ecuación (2.1), ya que debido a la definición de esta ecuación no es necesario encontrar la inversa de las matrices de varianza y covarianzas de los datos. Este procedimiento es presentado en la siguiente sección.

4. SOLUCIÓN DIRECTA DEL ACC USANDO ALGORITMOS GENÉTICOS

A continuación se presenta una solución directa del ACC usando el método de algoritmos genéticos. En el ACC se requieren encontrar los vectores ρ , \mathbf{a} y \mathbf{b} . Se propone encontrar la solución directamente de la ecuación (2.1), que se denotará de ahora en adelante como

$$F(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}'\Sigma_{XY}\mathbf{b}}{\sqrt{\mathbf{a}'\Sigma_{XX}\mathbf{a}}\sqrt{\mathbf{b}'\Sigma_{YY}\mathbf{b}}}, \quad (4.18)$$

y mediante la cual se determinarán \mathbf{a} y \mathbf{b} tal que al sustituirse en la ecuación se obtenga un máximo global, que será la primera correlación Canónica ρ , todo esto sujeto a las restricciones (2.2) y además se debe cumplir la siguiente condición $0 < F(\mathbf{a}, \mathbf{b}) \leq 1$.

En el problema se requiere encontrar valores \mathbf{a} y \mathbf{b} tal que (4.18) sea máxima. Para el programa se tomó como función objetivo $-F$, para que el problema sea visto como un problema de mínimos y se toma $1/F$ como función de adaptación o fitness. Para este programa, si N_1 es el Número de variables de X y N_2 es el Número de variables de Y , entonces el cromosoma tendrá $N_1 + N_2$ genes. Para determinar el valor óptimo que maximiza (4.18), se realizó de manera análoga el procedimiento de algoritmos Genéticos descrito en la sección anterior. Para determinar los mejores resultados o los mejores valores se verificaron las restricciones dadas en la ecuación (2.2), es decir, se tomaron los valores con un error pequeño.

4.1. Ejemplo de datos de cabezas

Se aplicó este algoritmo al ejemplo de datos de cabezas de la sección 3. Como primer paso se llevó a cabo un procedimiento de calibración en donde se obtuvieron los parámetros siguientes

Número de generaciones=1100
 Tamaño de la población=80
 Razón de mutación=0.044
 Razón de cruce=0.69

y posteriormente, mediante el programa se obtuvieron los vectores canónicos \mathbf{a} y \mathbf{b} tal que la correlación Canónica ρ es máxima, estos valores están dados en la Tabla 7. Observe que los resultados obtenidos mediante este método son similares a los presentados en la Tabla 3, solo que en este método sólo se logró obtener la correlación máxima que representa la primera correlación Canónica.

Cor ρ	Vectores Canónicos				Error 1	Error 2
	\mathbf{a}		\mathbf{b}			
0.7885	0.5539	0.5202	0.5219	0.5434	0.0007	0.0437

Table 7: Correlación Canónica y vectores canónicos de los datos de cabezas.

Note que con este método \mathbf{a} y \mathbf{b} son determinados con los AG directamente de la definición de ACC para obtener a ρ , a diferencia del método de AG anterior, en donde \mathbf{b} era determinado a partir de \mathbf{a} .

4.2. Ejemplo de datos mmreg

Se aplicó este algoritmo al ejemplo de datos mmreg. Se llevó a cabo un procedimiento de calibración en donde se obtuvieron los parámetros siguientes

Número de generaciones=100
 Tamaño de la población=90
 Razón de mutación=0.18
 Razón de cruce=0.66.

Mediante el programa se obtuvieron los primeros vectores canónicos **a** y **b** con primera correlación Canónica ρ , estos valores están dados en la Tabla 8. Al comparar estos valores con los de la Tabla 5, se observa que son parecidos.

Cor. ρ	Vectores Canónicos								Error 1	Error 2
	a				b					
0.4439	-0.7687	0.2115	-0.5052	-0.4300	-0.5588	-0.2033	0.0124	-0.0361	0.0847	

Table 8: Correlación Canónica y vectores canónicos de los datos mmreg.

Nuevamente, estos dos ejemplos ayudan a verificar el buen funcionamiento de este segundo programa de algoritmos genéticos.

4.3. Ejemplo de datos de carbono

Para determinar la correlación Canónica máxima y sus vectores canónicos en el problema de los datos de carbono se utilizó el programa de AG que encuentra estos valores directamente de la definición. Como primer paso se realizó la calibración de los parámetros del programa, obteniéndose

Número de generaciones=900

Tamaño de la población=80

Razón de mutación=0.055

Razón de cruza=0.72.

Posteriormente, mediante el programa se determinaron los valores de los vectores canónicos del tal manera que al sustituirlos en (4.18) se obtiene el valor máximo, que representa a la primera correlación Canónica. El mejor valor obtenido de la correlación y los vectores canónicos están presentados en la Tabla 9; se presentan además los errores en las restricciones que fueron los más pequeños que se obtuvieron.

Cor. ρ	Vectores Canónicos									Error 1	Error 2
	a				b						
0.5741	-0.9252	-1.5064	-0.9940	-0.6400	1.1248	-0.1228	-0.1932	-0.3232	0.3048	-0.0584	-0.0057

Table 9: Correlación Canónica y vectores canónicos de los datos de carbono.

Mediante este programa se logró obtener el valor de la correlación Canónica, el cuál es 0.5741, que nos indica que los grupos de variables están correlacionados pero el grado de relación no es fuerte. Se observa además en el primer vector canónico, en el primer grupo, las variables que más aportan información son densidad aparente, porcentaje de arena y porcentaje de limo con signo negativo, mientras que para el segundo grupo solamente el porcentaje de materia orgánica con signo positivo. El resultado refleja cómo en estos suelos con elevada producción de materia orgánica se presenta con densidades aparentes bajas.

5. CONCLUSIONES

- Se realizó un programa usando algoritmos Genéticos para calcular los valores de λ y **a**, de la ecuación (2.7), mediante el formalismo de Lagrange en el método ACC.
- Para algunos ejemplos, se compararon los resultados obtenidos mediante el método AG con los métodos de valores y vectores propios generalizados, obteniéndose los mismos valores de la correlación Canónica y los vectores canónicos.
- Se realizó un programa usando algoritmos Genéticos para calcular los valores de **a** y **b**, de la ecuación (2.1), para aproximar el máximo de la correlación Canónica. Para la primera correlación Canónica y sus vectores canónicos, se compararon los resultados en los ejemplos trabajados en la fase anterior, obteniéndose resultados similares.

- Para el problema de datos de carbono en el suelo, donde la matriz de varianzas y covarianzas del segundo grupo no es invertible, se resolvió usando el programa que aproxima el máximo de las correlaciones Canónicas, obteniéndose **a** y **b** de la ecuación (2.1), calculando la primera correlación Canónica. De estos resultados se deduce que los grupos de variables estén medianamente correlacionadas.

RECEIVED: JUNE, 2015.

REVISED: MAY, 2016.

REFERENCES

- [1] CASTILLO, M., LINARES, G., VALERA, M. A. y GARCIA, N. E. (2009): Modelación de la materia orgánica en suelos volcánicos de la región de Teziutlán, Puebla, México. **Revista Latinoamericana de Recursos Naturales**, 5(2), 148-154. ISSN 1870-0667
- [2] COELLO, C. A. (1995): Introducción a los Algoritmos Genéticos. **Soluciones Avanzadas. Tecnologías de Información y Estrategias de Negocios**, 3(17), 5-11.
- [3] CUADRAS, C.M. (1981): **Métodos de Análisis Multivariante**. Editorial Universitaria. Barcelona, España.
- [4] GONZÁLEZ, I., DÉJEAN, S., MARTIN, P. y BACCINI, A. (2008): CCA: An R Package to Extend Canonical Correlation Analysis. **Journal of Statistical Software**, 23(12).
- [5] HARDOON, D., SZEDMAK, S. y SHAWE-TAYLOR, J. (2004): Canonical correlation analysis: an overview with application to learning methods. **Neural Computation**, 16(12), 2639-2641.
- [6] ISON, M., SITT, J, y TREVISAN, M. (2005): Algoritmos genéticos: aplicación en MATLAB. Guía de la materia Sistemas Complejos. Disponible en www.df.uba.ar/users/mison/genetico.tar.gz.
- [7] LINARES, G., VALERA, M. A. y CASTILLO, M. (2007): Análisis de datos de suelos forestales en la caldera de Teziutlán, Puebla, por componentes principales y técnicas geoestadísticas. **Memorias del XXI Foro Nacional de Estadística**. 63-68. Instituto Nacional de Estadística, Geografía e Informática. ISBN 978-970-13-4930-4
- [8] MARDIA, K., KENT, J. y BIBBY, J. (1979): **Multivariate Analysis**. Academic Press. Great Britain.
- [9] OCTAVE. Disponible en www.gnu.org/software/octave/
- [10] SEBER, G. A. F. (1984): **Multivariate Observations**. First edition. Wiley. New Jersey.