

William Castillo - Javier Trejos  
*Editores Científicos*

Memorias del

**X SIMPOSIO INTERNACIONAL  
de  
MÉTODOS MATEMÁTICOS  
APLICADOS A LAS CIENCIAS**



Programa de Investigación en Modelos y Análisis de Datos  
Escuela de Matemática  
Universidad de Costa Rica

Departamento de Matemática  
Instituto Tecnológico de Costa Rica

4-7 de febrero de 1997  
Liberia, Costa Rica

X SIMPOSIO INTERNACIONAL DE MÉTODOS MATEMÁTICOS APLICADOS A LAS CIENCIAS

LIBERIA, COSTA RICA 3-7 FEBRERO 1997. *William Castillo - Javier Trejos, Editores Científicos*

Universidad de Costa Rica – Instituto Tecnológico de Costa Rica

## Contenido

---

### Conferencias invitadas

Similitudes, disimilitudes, preferencias: un enfoque unificado

*Serge Joly y Georges Le Calvé* ..... 1

Análisis procrusteano: fundamentos y aplicación para un análisis de tablas evolutivas

*Bernard Fichet* ..... 4

Geometrical tools in classification and pattern recognition

*Jean-Paul Rasson y Marcel Rémon* ..... 6

### Minicursos

Métodos de puntos interiores en optimización lineal

*Ramiro J. Cáceres E.* ..... 15

Redes neuronales: experimentos con MATHEMATICA<sup>©</sup>

*Ileana Castillo, José Luis Espinoza, Geovanni Figueroa y Walter Mora* . 17

Programación estocástica en optimización

*Sergio de los Cobos, Blanca Rosa Pérez y Miguel Ángel Gutiérrez* ..... 31

Análisis tiempo-frecuencia de señales cardiovasculares bajo diferentes condiciones hemodinámicas

*Raúl Jiménez, Bruno Günther y Eric Zabala* ..... 46

Métodos no paramétricos univariados y multivariados

*Belmer García Negrillo* ..... 56

### Ponencias

Implementación del ACP usando distancias relacionales en presencia de dos gru-

pos de variables	
<i>Marco Alfaro y Mario Villalobos</i> .....	57
Polinomios y spline interpolantes en $\mathbb{R}^n$	
<i>Vernor Arguedas y Roberto Mata</i> .....	65
Aplicación de los conjuntos aproximados a la simplificación de reglas en sistemas expertos y clasificación automática	
<i>Alcides Astorga</i> .....	66
Estudio de métricas en análisis en componentes principales mediante generadores primarios	
<i>Fabrizio Bolaños</i> .....	76
Estudio evolutivo de tablas de contingencia: aplicación a los egresos hospitalarios de mujeres entre 20 y 44 años	
<i>William Castillo y Jorge González</i> .....	78
Sistema de generación de particiones	
<i>Ben Chacón Pereira</i> .....	88
Búsqueda tabú: metaprocedimiento de mejora para el problema de inventarios multiproducto	
<i>Sergio de los Cobos, Blanca R. Pérez, Miguel A. Gutiérrez y Manuel Ordorica</i> .....	97
Controlabilidad direccional para sistemas no lineales en el plano a través de un proceso de linealización	
<i>Víctor Delgado</i> .....	107
Análisis en modelos de regresión con datos censurados	
<i>Jorge Domínguez</i> .....	112
Information et valeurs pronostiques de certaines variables dans le cancer du sein	
<i>F. Max Dorville, Jérôme Manuceau, Marylène Troupé y Jean Vaillant</i> .	121
Descripción de clases mediatne conjuntos aproximados	
<i>José Luis Espinoza, Geovanni Figueroa, Walter Mora y Javier Trejos</i> .	130
Sobre multirresoluciones, elementos finitos y un problema de dinámica de fluidos	
<i>Mijail Guillemard</i> .....	140
Multivariate elliptically contoured linear models	
<i>Arjun K. Gupta</i> .....	151

Optimización con recocido simulado para el problema de asignación cuadrática y comparaciones con otras técnicas de solución	
<i>Miguel Ángel Gutiérrez, Sergio de los Cobos y Blanca Rosa Pérez</i> . . . . .	152
Estudio de algunos aspectos relativos a las partículas suspendidas en la Ciudad de México	
<i>B. Hernández, F. Aldape y R. Hernández</i> . . . . .	164
Paquete didáctico de optimización en redes	
<i>Rafael López Bracho, M. Paula Ortuño, Felipe Carrillo y M. Teresa Rodríguez</i>	165
Un enfoque geométrico para el análisis de sensibilidad en programación lineal	
<i>Mario Marín</i> . . . . .	177
Comparación de índices en la clasificación jerárquica ascendente	
<i>Álex Murillo</i> . . . . .	187
Valores extremos de formas cuadráticas: una propiedad óptima de las componentes principales	
<i>Manuel Murillo y Miguel Alpízar</i> . . . . .	197
Métodos no paramétricos: consideraciones sobre ventajas y desventajas respecto a los métodos paramétricos	
<i>Graciela Navarro</i> . . . . .	198

Simulação numérica da modelagem dos poluentes aerossóis, segunda parte: uma primeira abordagem <i>Sonia E. Palomino Castro</i> .....	199
Operacionalización del concepto de desarrollo sostenible a través de modelos bioeconómicos <i>José Francisco Pastrana</i> .....	205
Regresión lineal y análisis de la varianza usando el paquete estadístico Lineal <i>Blanca Rosa Pérez, Sergio de los Cobos y Miguel Ángel Gutiérrez</i> .....	206
Análisis de regresión de Gini <i>Blanca Rosa Pérez, Sergio de los Cobos y Miguel Ángel Gutiérrez</i> .....	210
Análisis de estructuras de opinión pública <i>Jorge Poltronieri</i> .....	211
PIMAD 2.1: los algoritmos <i>Oldemar Rodríguez</i> .....	212
PIMAD 2.1: guía al usuario <i>Oldemar Rodríguez</i> .....	222
La ecuación de Euler para problemas de cálculo variacional con retardo <i>Lorena Salazar</i> .....	232
Enfoque compartimental para contar particiones generalizadas de números enteros positivos <i>Oswaldo Skliar, Guillermo Oviedo y Víctor Medina</i> .....	241
Nuevo enfoque compartimental para la solución del problema del isomorfismo de grafos <i>Oswaldo Skliar, Guillermo Oviedo y Víctor Medina</i> .....	250
Estrategias computacionales aplicadas a la evaluación animal <i>Yasnay Teijeiro y Reinaldo López</i> .....	251
Análisis discriminante aplicado a un estudio taxonómico de <i>Scirpus</i> sección <i>Monocephales</i> <i>Marta Torres, Lidia Benítez y Graciela Ponessa</i> .....	252
Algoritmo de la distribución normal <i>Juan José Vargas</i> .....	258
Factores de diferenciación de la mortalidad en la provincia de Buenos Aires en	

la década del 80

*Guillermo Velazquez, Hernán Otero, Claudia Marinelli y Sebastián Torcida*  
259

Aplicación del análisis factorial de correspondencias al estudio de la contaminación del aire en Guadalajara, Jalisco (México)

*Luz María Zúñiga, Jesús Sánchez e Irina Tereschenko* ..... 271

Petri Nets methodology applied for process and product modeling

*Milena Didic, Ulrich Mampel y Karl-Heinz Sternemann* ..... 272

Modelado de máquinas asíncronas e identificación óptima

*Salvador De Lara y Jean Faucher* ..... 273

## SIMILITUDES, DISIMILITUDES, PREFERENCIAS: UN ENFOQUE UNIFICADO

SERGE JOLY<sup>1</sup> - GEORGES LE CALVÉ<sup>1</sup>

---

### Resumen

Los ligámenes entre variables usualmente se estudian en términos de similitudes (correlación) o disimilitudes (distancias). Representaciones geométricas se adaptan a estas nociones. Usualmente también, las nociones de preferencias se estudian en términos de órdenes.

Proponemos asociar la noción de posicionamiento multidimensional a los índices de preferencia. Más precisamente, dos variables se suponen muy distantes si una de ellas es fuertemente preferida a la otra, y muy cercanas si los dos índices de preferencia son débiles.

Proponemos modelos en los que las nociones de similitud, disimilitud y preferencia se ligan de manera simple: euclidiana,  $L_1$ , sup, árboles jerárquicos, árboles aditivos. Se proponen nuevas representaciones asociadas a los modelos y se dan algoritmos.

### Abstract

Links between variables are usually studied in terms of similarities (correlation) or dissimilarities (distances). Geometric representations are adapted to this notions. Usually too, notions of preferences are studied in terms of orders.

We propose to associate notions of multidimensional scaling to preferences indices. More precisely, two variables are supposed very distant if one is strongly preferred to the other, and very near if the the two preference indices are weak.

We propose models where notions of similarity, dissimilarity and preference are linked in a very simple way: euclidean,  $L_1$ , sup, hierarchical trees, additive trees. New representations, associated to these models are proposed. Algorithms are given.

Un panel de jueces ha dado una apreciación numérica sobre un cierto número de productos. La preferencia del jurado para el producto  $A$  respecto al producto  $B$ , se puede medir sumando las diferencias de notas para todos los jueces que hayan preferido  $A$  a  $B$ , lo que se denota  $P_{AB}$ . La distancia entre  $A$  y  $B$  se expresa entonces naturalmente como la suma  $P_{AB} + P_{BA}$ , y la similitud entre  $A$  y  $B$  al sumar para todos los jueces aquélla de las dos notas que es menor. Este enfoque es de naturaleza lineal. Se puede, por supuesto, generalizarlo.

---

<sup>1</sup>UNIVERSITÉ DE RENNES 2, 6 AVENUE GASTON BERGER, 35043 RENNES, FRANCIA

**Definición 1** Se dirá que una similitud  $S$ , una disimilitud  $D$  y una preferencia  $P$  están ligadas por un **modelo de Minkowsky de potencia  $r$** , si existe un punto  $m$  tal que para todo  $i, j$ :

$$\begin{aligned} S_{ij} &= (D_{mi}^r + D_{mj}^r - D_{ij}^r)/2 \\ D_{ij}^r &= S_{ii} + S_{jj} - 2S_{ij} \\ D_{ij}^r &= P_{ij}^r + P_{ji}^r \end{aligned}$$

Para  $r = 0$  se llega a un análisis de naturaleza conjuntista, haciendo intervenir las nociones de diferencia de conjuntos e intersección; para  $r = 1$  es el enfoque en norma  $L_1$ ; para  $r = 2$  se trata del enfoque en norma euclídea.

Uno puede estar tentado de representar las cantidades que intervienen sobre un mismo gráfico (ver Figura 1). Consideremos dos puntos  $A$  y  $B$  en el plano  $\mathbb{R}^2$  provisto de un sistema de ejes. Para un modelo de Minkowsky de potencia  $r$ , la distancia  $D_{AB}$  será  $[AB]_r$ , la preferencia de  $A$  sobre  $B$  será  $[AM]_r$ , la preferencia de  $B$  sobre  $A$  será  $[BM]_r$ . La similitud entre  $A$  y  $B$  será  $[OM]_1$  si  $r = 1$ , y  $\langle OA, OB \rangle$  si  $r = 2$ .

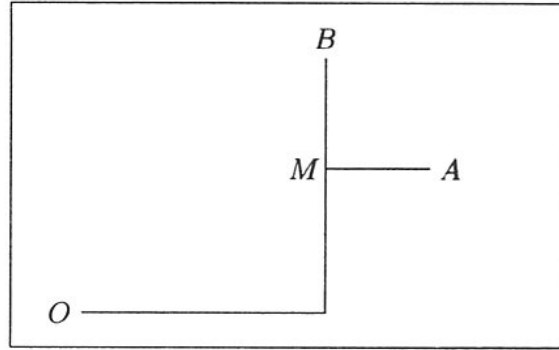


Figura 1: Representación de distancias, preferencias, similitudes y disimilitudes entre los puntos  $A$  y  $B$  sobre el plano  $\mathbb{R}^2$

La búsqueda de una representación como la anterior podrá obtenerse por medio de una adaptación de los métodos usuales de posicionamiento multidimensional (*multidimensional scaling*, como se conoce en inglés).

También se puede intentar un enfoque de clasificación definiendo una **ultrapreferencia** como una preferencia que verifica:

$$P_{AB} \leq \max\{P_{AB}, P_{BC}\}$$

para todo triplete de puntos  $A, B, C$ .

Toda ultrapreferencia es entonces equivalente a un árbol jerárquico no simétrico, sobre el cual la preferencia de  $A$  sobre  $B$  es la altura de nivel donde  $A$  va a unirse con  $B$  (ver Figura 2). Los algoritmos son análogos a los de clasificación jerárquica ascendente.

Otra posibilidad consiste en definir **preferencias aditivas**, como preferencias que verifican la desigualdad:

$$P_{AB} + P_{CD} \leq \max\{P_{AD} + P_{CB}, P_{AC} + P_{DB}\}$$



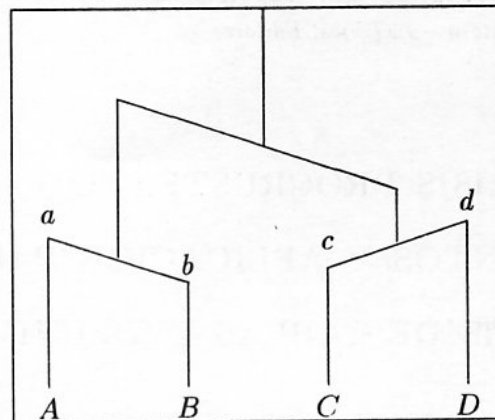


Figura 2: Representación de una ultrapreferencia por medio de un árbol jerárquico no simétrico

para todo  $A, B, C, D$ .

Una preferencia de este tipo es equivalente a un árbol aditivo, provisto de un origen, y sobre el cual la preferencia de  $A$  sobre  $B$  es la distancia de  $A$  al primer nodo, tomando en cuenta al origen, que reúne  $A$  y  $B$  (ver Figura 3).

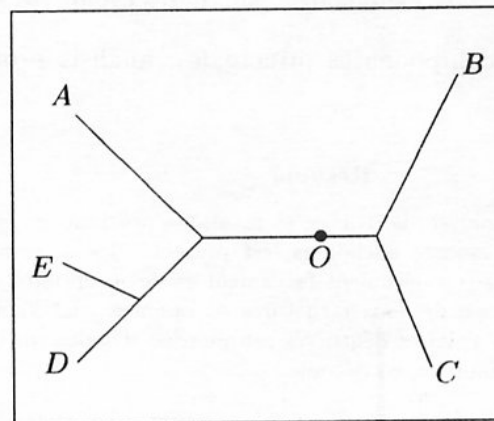


Figura 3: Representación de una preferencia aditiva por medio de un árbol aditivo

Los algoritmos son considerablemente más simples que los del árbol aditivo clásico, en la medida en que se conoce la distancia de cada punto al nodo más cercano.

La originalidad de este enfoque respecto al análisis ordinal tradicional, es que no busca una estructura de orden, optimizando un criterio, con un fin decisonal. Se puede tomar como un procedimiento descriptivo de las estructuras subyacentes a las preferencias, en una óptica de análisis de datos.

# ANÁLISIS PROCRUSTEANO: FUNDAMENTOS Y APLICACIÓN PARA UN ANÁLISIS DE TABLAS EVOLUTIVAS

BERNARD FICHET<sup>1</sup>

---

## Resumen

El análisis procrusteano, análisis de Tucker y rotaciones procrusteanas, de dos configuraciones de puntos en espacios euclídeos, se presenta aquí en términos de operadores. Soluciones y condiciones de unicidad se deducen fácilmente y propiedades de invarianza conducen al análisis procrusteano de dos estructuras de datos euclídeos. La extensión a los espacios provistos de productos escalares degenerados es abordada. Fundada en estos resultados, una metodología para datos evolutivos se deduce.

**Palabras claves:** análisis en componentes principales, análisis procrusteano, distancia euclídea.

## Résumé

L'analyse procrustéenne, analyse de Tucker et rotations procrustéennes, de deux configurations de points dans des espaces euclidiens, est présentée ici en termes d'opérateurs. Solutions et conditions d'unicité en découlent facilement et des propriétés d'invariance conduisent à l'analyse procrustéenne de deux structures de données euclidiennes. L'extension aux espaces munis de produits scalaires dégénérés est abordée. Fondée sur ces résultats, une méthodologie pour données évolutives en découle.

**Mots clés:** analyse en composantes principales, analyse procrustéenne, distance euclidienne.

El análisis inter-batería de Tucker concierne dos tablas de datos sobre los mismos individuos y variables diferentes. Es una extensión del análisis en componentes principales (A.C.P.). En términos geométricos, se consideran dos nubes de puntos en dos espacios euclídeos, cada uno completado de su centro de gravedad, de un centro de esfera circunscrita o de cualquier otro punto. Estas nubes son indexadas por el mismo conjunto. El análisis procrusteano de estas dos nubes es presentado independientemente de todo sistema de

---

<sup>1</sup>LABORATOIRE DE BIOMATHÉMATIQUES, FACULTÉ DE MÉDECINE, 27 BD. JEAN MOULIN, F-13385 MARSEILLE, FRANCIA

coordenadas, por el uso de formas bilineales no necesariamente simétricas y con la ayuda de la descomposición en valores singulares de operadores.

Algunos conceptos intrínsecos aparecen entonces: vectores y ejes procrusteanos, así como componentes procrusteanas de una nube respecto a otra. Otros invariantes salen simplemente a la luz: estabilidad de los ejes procrusteanos de una nube para toda isometría realizada sobre la segunda, y estabilidad de las componentes procrusteanas para toda transformación isométrica de cada nube. Se obtienen sistemas de componentes procrusteanas de dos estructuras de datos euclídeas completadas, independientes de las imágenes euclídeas. Las fórmulas de transición aseguran el ligamen entre estos dos sistemas.

El problema de la optimalidad de Tucker se escribe entonces fácilmente. Además, una mejor superposición de dos nubes se deduce directamente del conocimiento de los ejes procrusteanos. En particular, la unicidad es estudiada, distinguiendo unicidad de la transformación y unicidad de la rotación.

La extensión a los espacios provistos de productos escalares degenerados (formas bilineales semidefinidas positivas) se aborda enseguida. La introducción de ejes y componentes procrusteanas es más delicada, pues los operadores no son diagonalizables. Sin embargo, se les puede exhibir, lo que generaliza los resultados anteriores.

Se evoca enseguida una estrategia para la representación de estructuras de datos euclídeos que evolucionan en el tiempo. Superponiendo lo mejor posible la representación en el tiempo  $(t + 1)$  sobre la que se obtiene en el tiempo  $t$ , una estrategia como ésta ofrece trayectorias intra para los individuos. También es posible efectuar rotaciones de cada nube respecto a una nube tomada como referencia, o respecto a una nube promedio (compromiso). Diferentes escogencias se proponen para pesos que evolucionan en el tiempo o para parejas de estructuras en correspondencia y evolutivas. Así mismo, se puede hacer un A.C.P. de la figura global después de las rotaciones o efectuar las rotaciones después del A.C.P. de cada estructura. En este último caso, las fórmulas analíticas de las rotaciones planas se recuerdan. Todo esto representa un análisis fino.

Para un análisis global que ofrezca una representación de los diferentes tiempos, se propone la métrica procrustea entre distancias euclídeas. Con la ayuda de un contraejemplo, mostramos que ésta no es euclídea, lo que requiere del uso de una aproximación.

## GEOMETRICAL TOOLS IN CLASSIFICATION AND PATTERN RECOGNITION

JEAN-PAUL RASSON<sup>1</sup> - MARCEL RÉMON<sup>1</sup>

---

### Abstract

We investigate the solutions to the clustering and the discriminant analysis Problems when the points are supposed to be distributed according to Poisson Processes on convex supports. This leads to very intuitive criteria for homogeneous Poisson Processes based on the Lebesgue measures of convex hulls. For non homogeneous Poisson Processes, the Lebesgue measures have to be replaced by intensities integrated on convex hulls.

Similar geometrical tools, based on the Lebesgue measure, are used in the context of Pattern Recognition. First, a discriminant analysis algorithm is developed for estimating a convex domain when inside and outside points are available. Generalisation to non convex domains is explored.

### Resumen

Investigamos las soluciones a los problemas de clasificación automática y análisis discriminante cuando se supone que los puntos se distribuyen de acuerdo con un Proceso de Poisson sobre soportes convexos. Esto conduce a criterios muy intuitivos para Procesos de Poisson homogéneos basados en medidas de Lebesgue sobre envolventes convexas.

Herramientas geométricas similares, basadas en la medida de Lebesgue, son usadas en el contexto del Reconocimiento de Patrones. Primero, un algoritmo de análisis discriminante se desarrolla para estimar un dominio convexo cuando se pueden conocer los puntos internos y externos. Se explora la generalización a dominios no convexos.

## Introduction

This research has been initiated by the question of D.G. Kendall of how to estimate a bounded convex set observing only the realization of an homogeneous Poisson Process inside this convex set. The solution (Ripley and Rasson, 1977; Rasson, 1979) was an homothetic expansion of the convex hull of the sample from its centroid. The following question, raised by E. Diday, was naturally the problem of using these arguments in Clustering. This led us to the maximum likelihood estimation of the hypothesized support

---

<sup>1</sup>F.U.N.D.P., DÉPARTEMENT DE MATHÉMATIQUE, 8 REMPART DE LA VIERGE, B-5000 NAMUR, BÉLGICA; TEL.: + 32 81 724928 OR 724930; FAX: + 32 81 724914; E-MAIL: JPR@MATH.FUNDP.AC.BE OR MRE@MATH.FUNDP.AC.BE

for clustering, i.e. the union of  $K$  bounded convex sets. The solution was “find the partition of the points in  $K$  subgroups of such that the sum of the Lebesgue measures of their convex hulls is minimal” (Hardy and Rasson 1982).

Then came the question of the corresponding discriminant analysis raised by A.D. Gordon. The Bayesian solution for the same hypothesis (still for an homogeneous Poisson Process) was to affect the new point to the sample for which the Lebesgue measure added by convexity to its convex hull is minimal (Baufays and Rasson, 1984a, 1984b, 1985). But this could not classify the points belonging to more than one convex hull as this is often the case of pixels in images.

Then to deal with them, we moved to non-homogeneous independent Poisson Processes with convex supports. The main change in the solution was to replace the Lebesgue measure by the integrated intensity. But the same equation gave also the solution for points lying in the intersections of the supports.

## 1 The Stationary Poisson Process

### 1.1 The model

We consider that we deal with a clustering problem where the points we observe are generated by a Poisson Process and are distributed in  $D$  where  $D$  is the union of  $g$  disjoint domains ( $D_k$ ),  $1 \leq k \leq g$ . Our point of view will be that if we are able to find back these domains, making some inference about them, we will, in some sense, solve the clustering problem.

### 1.2 The maximum likelihood solution of the clustering problem

Let  $x$  denote the sample vector  $(x_1, \dots, x_n)$  with  $x_i \in \mathbb{R}^d, i = 1, \dots, n$ . The indicator function of a set  $A$  at the point  $y$  is defined by :  $\mathbb{1}_A(y) = 1$  if  $y \in A$  and 0 otherwise.

So, since with our hypothesis, the points will be independently and uniformly distributed on  $D$ , the likelihood function takes the form

$$F_D(x) = \frac{1}{(m(D))^n} \cdot \prod_{i=1}^n \mathbb{1}_D(x_i)$$

where  $m(D)$ , the Lebesgue measure of  $D$ , is the sum of the measures of the  $g$  subsets  $D_k (1 \leq k \leq g)$ . The domain  $D$ , parameter of infinite dimension, for which the likelihood is maximal, is, among all those which contain all the points, the one whose Lebesgue measure is minimal.

If we do not impose more conditions on the subsets, we can easily find  $g$  sets  $D_k$  which contain all the points and are such that the sum of their measures is zero. Thus there are many trivial solutions to the problem. Nevertheless, we can easily see that the problem of estimating a domain is not well-posed and that the weakest assumption that makes the domain  $D$  estimable is the convexity of the  $D_k$  (Baufays and Rasson, 1984).

With a partition of the set of points into  $g$  sub-domains having disjoint convex hulls, we can associate a whole class of estimators; indeed we only have to find  $g$  disjoint convex

sets, each of them containing one of them. For each partition, the likelihood has a local maximum: the convex hulls of the  $g$  subsets. The global maximum will be attained with the partition for which the sum of the Lebesgue measures of the convex hulls of the  $g$  subgroups is minimal. This is the solution we seek.

Practically, if the basic space is  $\mathbb{R}$ , we look for the  $g$  disjoint intervals containing all the points such that the sum of their lengths is minimal. In  $\mathbb{R}^2$  (or  $\mathbb{R}^3$ ), we try to find the  $g$  groups of points such that the sum of the areas (volumes) of their disjoint convex hulls is minimal.

### 1.3 The statistical model and the rule for the associated discriminant analysis

The conditional distribution for the  $k$ -th population is assumed to be uniform in a convex compact domain  $D_k$ , and the a priori probability  $p_k$  that an individual belongs to population  $k$  is proportional to the Lebesgue measure of  $D_k$ . The convex domains  $D_k$  are assumed disjoint. The density of population  $k$ ,  $f_k(x)$ , and the unconditional density  $f(x)$ , are respectively equal to:

$$f_k(x) = \frac{1}{m(D_k)} \cdot \mathbb{1}_{D_k(x)}$$

$$f(x) = \frac{1}{m(D)} \cdot \sum_{k=1}^g \mathbb{1}_{D_k(x)}$$

The decision rule is the Bayesian one, with the unknown parameters - the convex sets  $D_k$  - replaced by their maximum likelihood estimations.

Let  $X_k$  be the labeled sample of population  $k$ ,  $H(X_k)$  be its convex hull, and  $x$  be the individual to be assigned to one of the  $g$  populations. If  $x$  is allocated to the  $k$ -th group, the estimates of the domains  $D_k$  are equal to:

$$\widehat{D}_j = \begin{cases} H(X_j) & \text{if } j \neq k \\ H(X_j \cup \{x\}) & \text{if } j = k \end{cases}$$

The maximum likelihood estimate of  $p_k f_k(x)$  is therefore:

$$\widehat{p}_k f_k(x) = \frac{\mathbb{1}_{\widehat{D}_k}(x)}{\sum_{j=1}^g m(H(X_j)) + S_k(x)}$$

with  $S_k(x) = m(H(X_k \cup \{x\})) - m(H(X_k))$ .

As the convex sets are assumed disjoint,  $x$  can be allocated to the  $k$ -th group, if and only if  $H(X_k \cup \{x\})$  and  $H(X_j)$  are disjoint for all  $j \neq k$ . Otherwise, we define arbitrarily:  $S_k(x) = +\infty$ .

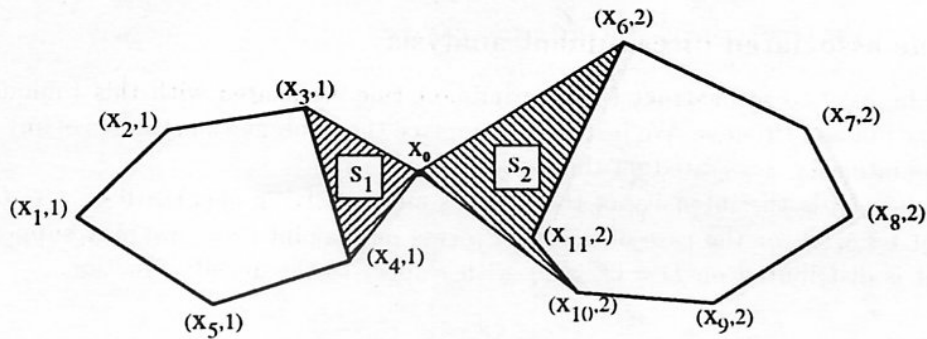


Figure 1.

The allocation rule is then :

*assign  $x$  to the  $k$ -th population if and only if  $S_k(x) < S_j(x), k \neq j$ .*

## 2 The Non-stationary Poisson Process

In order to avoid an heuristical part in the supervised classification procedure, we generalize now the basic hypothesis, namely the Homogeneous Poisson Process, one into the hypothesis of Non Homogeneous Poisson Process.

### 2.1 The Maximum Likelihood solution for the clustering problem

We shall consider therefore that we deal with a clustering problem where the points we observe are generated by a Non Homogeneous Poisson Process with intensity  $q(\cdot)$  (satisfying  $q(x) > 0$ ) and are observed only in  $D$ , where  $D$  is the union of  $g$  disjoint convex domains.

$$D = \bigcup_{k=1}^g D_k, \quad D_k \text{ convex disjoint.}$$

Our point of view will be again to try to find the maximum likelihood estimation of the  $D_k$ . If  $x$  denote the sample vector  $(x_1, \dots, x_n)$  with  $x_i \in \mathbb{R}^d, i = 1, \dots, n$  its likelihood will be

$$F_D(x) = \frac{1}{(m(D))^n} \cdot \prod_{i=1}^n \mathbb{1}_D(x_i) \cdot q(x_i)$$

where  $m(D) = \int_D q(x) dx$  and  $q(\cdot)$  is the intensity of the process.

Thus, if the intensity is known (or, maybe, has been estimated), the maximum likelihood clustering solution, since the maximum likelihood estimation of a convex domain based on a sample of points inside this domain is still the convex hull of this sample, is then:

*Find the  $g$  groups of points for which the sum of the intensities integrated on their convex hulls is minimal.*

## 2.2 The associated discriminant analysis

It would be easy to reconstruct the discriminant rule associated with this (unique) Non Stationary Poisson Process. We just have to replace the Lebesgue's measures of any convex set by the intensity integrated on this convex set.

Thus if  $q_k(\cdot)$  is the intensity of the Process and satisfying  $q_k(x) > 0 \Leftrightarrow x \in D_k$  (e.g.  $q_k(\cdot) = q(\cdot) \mathbb{1}_{D_k}(\cdot)$  for the case of unique Process on disjoint sets), we may suppose that any point is distributed on  $D = \cup_{k=1}^g D_k$  with respect to the density function

$$f(x) = \sum_{k=1}^g p_k \cdot f_k(x)$$

$$\text{where } f_k(x) = \frac{q_k(x)}{\int_{D_k} q_k(y) dy} \quad \text{and} \quad p_k = \frac{\int_{D_k} q_k(y) dy}{\sum_{k=1}^g \int_{D_k} q_k(y) dy}$$

As usual, the possibly convex supports  $D_k$  are estimated by the convex hulls  $H(X_k)$  of the training set points.

If we denote  $S = \sum_{k=1}^g \int_{H(X_k)} q_k(y) dy$  and  $S_k(x) = \int_{H(X_k \cup \{x\}) \setminus H(X_k)} q_k(y) dy$  then the Bayesian classification rule becomes:

*assign the new point  $x$  to the class  $k$  such that  $\widehat{p}_k f_k(x) = q_k(x)/(S + S_k(x))$  is maximal.*

When the local intensities  $q_k(x)$  do not depend on  $k$ , this rule simply consists in assigning  $x$  to the class  $k$  such that the added intensity  $S_k(x)$  is minimal. Thus, in this case and when the convex hulls are disjoint, we still keep the convex admissibility property i.e. "if the point  $x$  belongs to the convex hull of only one class,  $x$  is assigned to this class".

## 3 Pattern Recognition applications

### 3.1 The inside/outside problem : presentation

Suppose that  $X$  is a Poisson point process within a fixed finite window  $F \subset \mathbb{R}^d$ . In  $F$ , we have a compact convex domain  $D$ . We suppose that the Poisson process  $X$  is homogeneous on  $F$ , with density  $\lambda$ . We observe a fixed number  $t \geq 1$  of realizations of  $X$  in  $F$ , from which  $n$  turn out to be inside the domain  $D$  and  $m$  outside of  $D$  ( $t = n + m$ ). We want to estimate the unknown convex domain  $D$ . This problem is indeed the third problem of Grenander (1973).

The solution we propose to the problem is the use of the discriminant rule we have just described in cluster analysis. The situation here is quite similar as we have two disjoint domains  $D$  and its complementary  $\bar{D} = F \setminus D$ . The main difference lies in the non convexity of  $\bar{D}$ .

### 3.2 The discriminant analysis rule and the inside/outside problem

The same idea is applied to the inside/outside problem, with  $D$  and  $\bar{D}$  playing the role of  $C_1$  and  $C_2$ . The major difference here is that  $\bar{D}$  is no longer a convex set.



Let us note  $(x, y) = \{x_1, y_1, \dots, x_{n+m}, y_{n+m}\}$  the realizations of the homogeneous Poisson process  $X$  and the labeling variable  $Y : y_i = 1$  if  $x_i \in D$  and  $y_i = 2$  otherwise. Let  $y_i = 1$  for  $i = 1, \dots, n$  and  $y_i = 2$  for  $i = n+1, \dots, n+m$ , without restricting generality.

Conditionally on  $n$  and  $m$  fixed, the likelihood function for  $(x, y)$  is

$$\begin{aligned} L_D(x, y) &= \left( \frac{1}{m(D)^n} \prod_{i=1}^n \mathbb{1}_{[x_i \in D]} \right) \left( \frac{1}{m(\bar{D})^m} \prod_{i=n+1}^{n+m} \mathbb{1}_{[x_i \in \bar{D}]} \right) \\ &= \left( \frac{1}{m(D)} \right)^n \left( \frac{1}{m(\bar{D})} \right)^m \mathbb{1}_{[H(x_1, \dots, x_n) \subseteq D]} \mathbb{1}_{[J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n) \subseteq \bar{D}]} \end{aligned}$$

where  $J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n)$  is the "shadow" statistic defined in Hatchel, Meilijson and Nadas (1981). It is defined as:

$$J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n) = \bigcup_{i: y_i=2} \{x_i + \lambda(x_i - b) \in \mathbb{R}^d | \lambda \geq 0, b \in H(x_1, \dots, x_n)\}.$$

$(H(\cdot), J(\cdot | \cdot))$  is a minimal sufficient statistic for the estimation of  $D$ . See Figure 2.

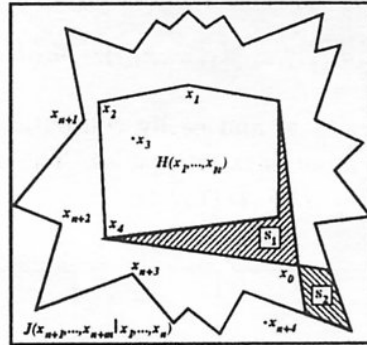


Figure 2.

It is known that  $J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n)$  has similar properties as the convex hull statistic  $H(x_1, \dots, x_n)$ . It is a consistent estimate of  $\bar{D}$ . It is robust with respect to small changes in the location of the data points. It satisfies the equivariance requirement and underestimates the Lebesgue measure of  $\bar{D}$  in the same way as  $m(H(x_1, \dots, x_n))$  does for  $m(D)$ , ie

$$E[m(J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n))] = \left( 1 - \frac{E[V_{m+1}]}{m+1} \right) m(\bar{D})$$

with  $E[V_{m+1}]$  being the expected number of extreme points of  $J(\cdot | \cdot)$  for  $m+1$  observations in  $\bar{D}$ . See Ripley and Rasson (1977) and Rémon (1993).

The use of  $H(x_1, \dots, x_n)$  and  $J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n)$  as estimates of the two unknown domains [here  $D$  and  $\bar{D}$ ] in the criterion proposed by Baufays and Rasson is the key idea of our discriminant algorithm.

One gets then the following boundary of the regions allocating to  $D$  and  $\hat{D}$ . This is the set of points  $x_0$  such that:

$$p_D \widehat{f}_D(x_0) = p_{\hat{D}} \widehat{f}_{\hat{D}}(x_0)$$

i.e.

$$\begin{aligned} & \frac{1}{m(H(x_1, \dots, x_n, x_0)) + m(J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n))} \\ &= \frac{1}{m(H(x_1, \dots, x_n)) + m(J(x_{n+1}, \dots, x_{n+m}, x_0 | x_1, \dots, x_n))} \end{aligned}$$

i.e.

$$S_1(x_0) = S_2(x_0)$$

where

$$S_1(x_0) \equiv m(H(x_1, \dots, x_n, x_0)) - m(H(x_1, \dots, x_n))$$

and

$$S_2(x_0) \equiv m(J(x_{n+1}, \dots, x_{n+m}, x_0 | x_1, \dots, x_n)) - m(J(x_{n+1}, \dots, x_{n+m} | x_1, \dots, x_n)).$$

This boundary gives us a practical and easily computable estimate  $\hat{D}$  for the unknown domain  $D$ . See Figure 3 for large data set results. The symmetric difference between  $D$  and  $\hat{D}$  is noted by  $D\Delta\hat{D} = D \cup \hat{D} \setminus D \cap \hat{D}$ .

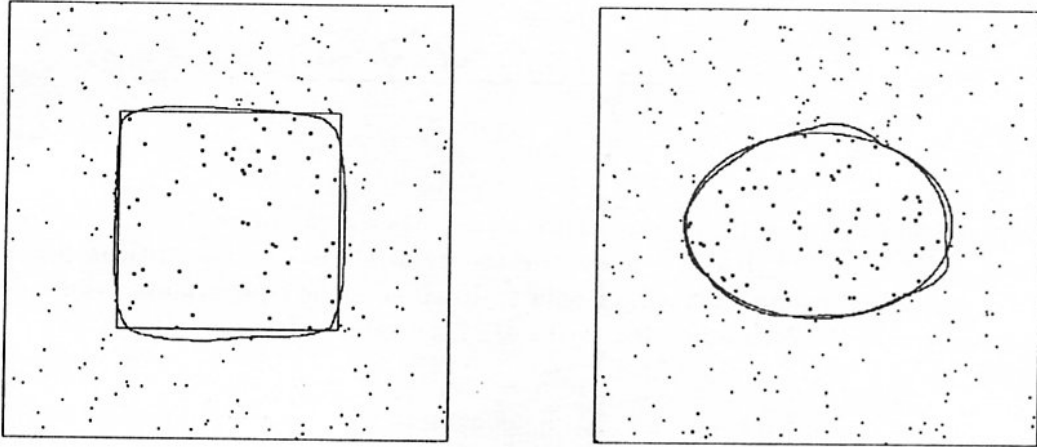


Figure 3a: Quadratic  $D$  with  $m(D) = 0.25$  :  $t = 250$  observations with  $n = 59$  from  $D$  yield  $\hat{D}$  with  $m(\hat{D}) = 0.26$  and  $m(D\Delta\hat{D}) = 0.018$ .

Figure 3b: Ellipsoidal  $D$  with  $m(D) = 0.20$  :  $t = 300$  observations with  $n = 68$  from  $D$  yield  $\hat{D}$  with  $m(\hat{D}) = 0.20$  and  $m(D\Delta\hat{D}) = 0.011$ .

### 3.3 Properties of $\hat{D}$

The estimator  $\hat{D}$  yields a consistent estimation of  $D$ , as it is bound by two consistent estimators  $H(\cdot)$  and  $J(\cdot, \cdot)$  of  $D$ . It has a piecewise continuous boundary. Unfortunately it happens not to be a convex set. On the other hand, this last feature turns out to be an advantage when a similar reasoning is applied to the estimation of non-convex domains.

This estimator  $\hat{D}$  is robust with respect to small changes in the location of data points, as it is based only on the shapes of  $H(\cdot)$  and  $J(\cdot, \cdot)$ , which are robust in this sense. Such a property is rare in spatial statistics. For instance, the estimator  $\tilde{D}$  proposed by Moore et al. (1988) can be very sensitive to small change in the location of data points.

Let us note that the time required for the computing of  $\hat{D}$  does not depend on the number of points, excepted for the computation of the convex hull and shadow statistic. The amount of required cpu-time is only a function of the precision asked for the estimator  $\hat{D}$ .

### Conclusions and future researches

Our estimate for  $D$ , based on a well known discriminant analysis criterion, seems to be a powerful tool for pattern recognition. Moreover, it is quite straightforward to generalize it to a non-homogeneous Poisson process.

This research is currently working on recognition of non convex domains. The first results seems very encouraging as shown by the estimation of some letter A. See figure 4 where our algorithm is compared to a discriminant rule based on the distance to the nearest neighbour.

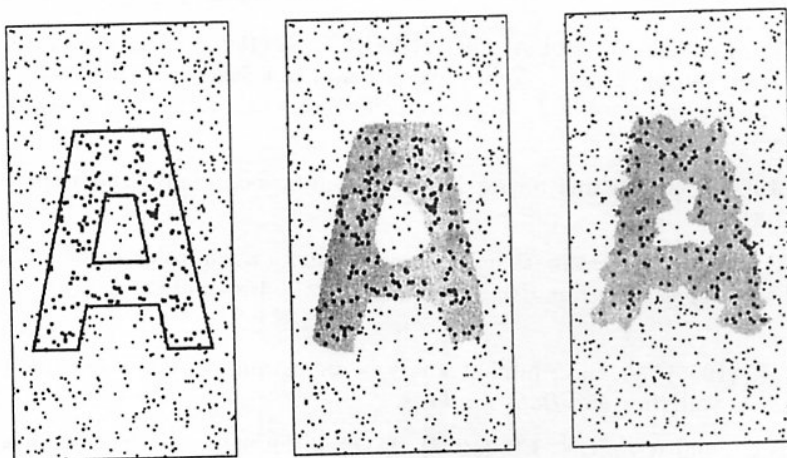


Figure 4a: A non-convex body to be estimated : the letter A.

Figure 4b: Our discriminant analysis algorithm.

Figure 4c: The estimation based on the distance to the nearest neighbour.

## References

- [1] BAUFAYS, P., RASSON, J.-P. (1984) “Une nouvelle règle de classement, utilisant l’enveloppe convexe et la mesure de Lebesgue”, *Statistique et Analyse des Données* **2**: 31–47.
- [2] BAUFAYS, P., RASSON, J.P. (1984) “Propriétés théoriques et pratiques et applications d’une nouvelle règle de classement”, *Statistique et Analyse des Données* **9**(3): 1–10.
- [3] BAUFAYS, P., and RASSON, J.P. (1985) “A new geometric discriminant rule”, *Computational Statistics Quarterly* **2**(1): 15–30.
- [4] DEGYTAR, Y.U., FINKELSH TEIN, M.Y. (1974) “Classification Algorithms Based on Construction of Convex Hulls of Sets”, *Engineering Cybernetics* **12**: 150–154.
- [5] DUDA, R.O., HART, P.E. (1973) *Pattern Recognition and Scene Analysis*. Wiley, Chichester.
- [6] EFRON, B. (1965) “The Convex Hull of a Random Set of Points”, *Biometrika* **52**: 331–453.
- [7] FISHER, L., VAN NESS, J.W. (1971) “Admissible Clustering Procedures”, *Biometrika* **58**: 91–104.
- [8] FUKUNAGA, K. (1972) *Introduction to Statistical Pattern Recognition*. Academic Press, New York.
- [9] GRENANDER, U. (1973) “Statistical geometry: a tool for pattern analysis”, *Bulletin of the American Mathematical Society* **79**: 829–856.
- [10] HAND, D.J. (1981) *Discrimination and Classification*. Wiley, Chichester.
- [11] HARDY, A., RASSON, J.-P. (1982) “Une Nouvelle Approche des Problèmes de Classification Automatique”, *Statistique et Analyse des Données* **7**: 41–56.
- [12] HARTIGAN, J.A. (1975) *Clustering Algorithms*. Wiley, Chichester.
- [13] MAC LACHLAN, G.J. (1992) *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York.
- [14] MOORE, M., LEMAY, Y. and ARCHAMBAULT, S. (1988) “Algorithms to reconstruct a convex set from sample points. Computing Science and Statistics”, Proceedings of the *20th Symposium on the Interface*, E.J. Wegman, D.T. Gantz and J.J. Miller (eds.), ASA, Virginia: 553–558.
- [15] RASSON, J.P. (1979) “Estimation de domaines convexes du plan”, *Statistique et Analyse des Données* **1**: 31–46.
- [16] REMON, M. (1994) “The estimation of a convex domain when inside and outside observations are available”, *Supplemento ai Rendiconti del Circolo Matematico di Palermo, serie II*, **35**: 227–235.
- [17] REMON, M. (1996) “A Discriminant analysis algorithm for the inside/outside problem”, *Computational Statistics and Data Analysis*.
- [18] RIPLEY, B.D., and RASSON, J.P. (1977) “Finding the edge of a Poisson forest”, *Journal of Applied Probability*, **14**: 483–491.
- [19] TOUSSAINT, G.T. (1980) “Pattern Recognition and Geometrical Complexity”, *Proc. Fifth Int. Conf. Pattern Recognition*, IEEE: 1324–1347.

# *Minicursos*

## MÉTODOS DE PUNTOS INTERIORES EN OPTIMIZACIÓN LINEAL

RAMIRO JOSÉ CÁCERES ESPINOZA<sup>1</sup>

---

### Objetivos

El objetivo del minicurso es introducir al estudiante a las diferentes ideas usadas en el diseño de métodos de puntos interiores para programación lineal.

### Público

El minicurso está dirigido fundamentalmente a estudiantes de licenciatura en matemáticas que hayan recibido una introducción a la programación lineal y sus métodos, y a estudiantes de posgrado o profesores interesados.

### Contenidos

En una parte inicial, se estudiarán las ideas básicas y características de métodos de puntos interiores tales como los de Khachiyan, Renegar, Vaidya, Karmarkar, Gonzaga.

El curso se puede extender a un curso de 15 horas, en el que en una segunda parte se puede mostrar una generalización del método de Vaidya al caso de problemas convexos en una clase particular, así como experiencias en la implementación computacional del algoritmo correspondiente.

### Bibliografía

- [1] Bazarra, M.S. (1990) *Linear Programming and Network Flows*. John Wiley & Sons, New York.
- [2] Cáceres, R. (1991) *Ueber die Zentren-Methoden für eine Klasse Konvexer Optimierungsaufgaben*. Dissertation, Humboldt Universität zu Berlin.

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA, UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA, LEÓN

- [3] Jarre, F. (1989) *The Method of Analytic Centers for Smooth Convex Programs*. Dissertation, Universität Wuerzburg.
- [4] Karmarkar, N. (1984) "A new polynomial-time algorithm for linear programming", *Combinatorica* 4.
- [5] Renegar, J. (1988) "A polynomial-time algorithm based on Newton's method for linear programming", *Mathematical Programming* 40
- [6] Vaidya, P. M. (1987) "An algorithm for linear programming which requires  $\mathcal{O}(((m+n)n^2 + (m+n)^{1.5}n)L)$  arithmetic operations", *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, New York, May 25-27.

## REDES NEURONALES: EXPERIMENTOS CON MATHEMATICA<sup>©</sup>

ILEANA CASTILLO<sup>1</sup> - JOSÉ LUIS ESPINOZA<sup>1</sup>  
GEOVANNI FIGUEROA<sup>1</sup> - WALTER MORA<sup>1</sup>

---

### 1 Introducción

En la experimentación con redes neurales, ya sea para familiarizarse con las características de éstas o para experimentar con nuevas arquitecturas. Se tienen varias alternativas: una es comprar algún software disponible comercialmente (opción de una flexibilidad limitada); otra es programar, digamos en C o Pascal. Esta segunda posibilidad ofrece gran flexibilidad, pero requiere gran experiencia de programación y/o invertir mucho tiempo de trabajo. En este minicurso exploraremos una posibilidad intermedia: aprender y experimentar con redes neurales usando MATHEMATICA. El programa se presta para que, con una pequeña introducción, podamos usarlo de manera efectiva. MATHEMATICA es primariamente un interpretador, pero es muy conveniente para el trabajo con redes pequeñas, que es lo que se maneja en el aprendizaje y la experimentación.

En este minicurso, después de una breve descripción del paradigma, iremos introduciendo los comandos de MATHEMATICA necesarios para hacer los cálculos que nos permitan familiarizarnos con las redes neurales que se presentan.

### 2 Redes Neurales

Una red neural es un paradigma computacional inspirado en el paralelismo del cerebro. En la figura que sigue se ilustra un modelo de red neural. Cada *nodo*, *unidad*, *elemento de procesamiento* o *neurona individual* es representada por un círculo. Las flechas señalan la dirección del flujo de datos de una unidad a la siguiente. Cada unidad tiene una sola salida que puede ser transmitida a varias unidades.

El nodo más simple hace una suma ponderada de  $N$  inputs  $x_1, x_2, \dots, x_N$  y pasa su resultado por una función de *activación*. El nodo es caracterizado por un umbral  $\theta$  y por el tipo de función de activación.

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE COSTA RICA, CARTAGO, COSTA RICA



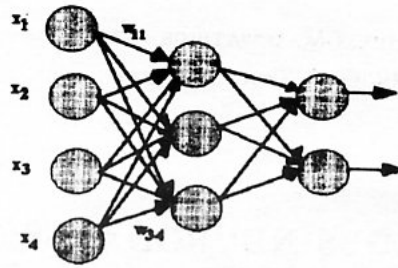


Figura 1: Ejemplo de red neural

Los modelos de redes neurales son especificados por la arquitectura de la red, las características de los nodos y las reglas de entrenamiento o aprendizaje (supervisado o insupervisado). Estas reglas especifican una manera de inicializar los pesos de la red y cómo éstos serán adaptados en el proceso de entrenamiento (sin embargo hay redes, como la de Hopfield y la de Hamming, que son generalmente usadas con pesos fijos y lo que se adapta es la salida de los nodos).

### 3 Redes neurales y MATHEMATICA

En MATHEMATICA todo es una expresión. Si uno evalúa cualquier tira de caracteres (usando la tecla `Insert` o `Shift-Enter`) entonces MATHEMATICA ejecuta la expresión si se trata de un comando, si no, devuelve la tira igual. En este programa no se definen las variables ni los tipos de manera explícita. Implícitamente, al usarse un comando, ya el tipo está definido por la naturaleza del comando. La idea es ir aprendiendo a dominar el ambiente del programa a través de los ejemplos.

#### 3.1 Cálculos en la red

Para los ejemplos que siguen usaremos la red de la Figura 1. En la capa de entrada, la red lee un patrón  $X=(x_1, x_2, \dots, x_N)$ . En la siguiente capa, la  $i$ -ésima unidad recibe el *input*:

$$net_i = \sum_{j=1}^n x_j w_{ij}$$

$w_{ij}$  es el peso de la conexión que va de la unidad  $j$  a la unidad  $i$ .

Para hacer este cálculo, supongamos que  $X$  es, por ejemplo, binario. También inicialicemos los pesos  $w_{ij}$  de cada capa como números aleatorios entre  $-1$  y  $1$ .

```
x1j=Table[Random[Integer,{0,1}],{4}]          (* inputs          *)
{1, 0, 1, 1}
```

```
W1=Partition[Table[Random[Real,{-1,1}],{12}],4] (* capa de pesos 1*)
W2=Partition[Table[Random[Real,{-1,1}],{6}],3]  (* capa de pesos 2*)
```

```
{ { 0.958118, -0.505832, 0.579135, 0.616119},
  {-0.188478, -0.154346, 0.494552, -0.084310},
  {-0.398497, -0.480943, -0.383355, 0.582014} }
```

```
{ {-0.156972, -0.580933, 0.835743},
  {0.850794, -0.670648, -0.355626} }
```

En MATHEMATICA, los comandos siempre empiezan con mayúsculas, sus argumentos están entre '[ ]' y si son varios se separan por comas. El comando `Table` genera una *lista* del tamaño especificado, en este caso 12 y 6 respectivamente. El comando que se aplica para generar la lista se pone en el lugar del primer argumento, en este caso es `Random`. La función del comando `Partition`, es hacer los grupos de pesos que usa cada neurona. En la segunda capa, cada neurona recibe 4 conexiones y en la tercera capa, cada neurona recibe 3 conexiones. Para tener acceso a cada grupo de conexiones, digamos el grupo *k*, se indica la sublista como `W1[[k]]`. Ahora vamos a calcular los *inputs* de la capa 2 y los *inputs* de la capa 3 así como la salida de la neurona. Para esto suponemos que la función de activación es la identidad para las neuronas de la capa 1 y 2. Para las neuronas de la capa 3, tomaremos como función de activación, la función  $f(net_i) = \frac{1}{1 + e^{-r net_i}}$  con  $r = 0.5$ .

```
net1=x1j.wij1[[1]]; (* funcion de activacion = identidad *)
net2=x1j.wij1[[2]];
net3=x1j.wij1[[3]];

x2j={net1,net2,net3} (* salidas de capa 2 *)
{2.15337, 0.221763, -0.199838}

sigm[neti_]:=1/(1+E^(-r neti)); r=0.5;

net4=sigm[ x2j.wij2[[1]] ]; (*funcion de activacion=sigm *)
net5=sigm[ x2j.wij2[[2]] ];

salida={net4,net5}
{0.346636, 0.852509} (* salida de la neurona *)
```

Las listas son consideradas como *vectores columna*. El punto y coma se usa tanto para evitar la salida de la ejecución de un comando (como aquí) como para separar varias opciones de un argumento.

Observe la forma de definir la función `sigm(neti)` (de variable `neti` y parámetro `r`), usando `neti_` para la variable (esto sólo se hace en la definición). El símbolo `:=` es usado para definir la función pero *sin calcularla*. Esto es útil en contextos donde las funciones que definimos deben esperar el valor de un parámetro (en este caso `r`). `'='` se usa para definir y *calcular* de una vez la función.

Para graficar la función se usa el comando `Plot`.

```
Clear[r]; r=1;
sig1=Plot[sigm[neti],{neti,-10,10}]
```

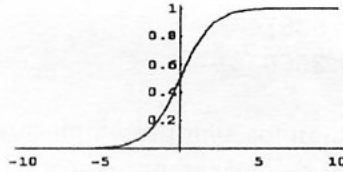


Figura 2: Sigm(x) con  $r = 1$

## 4 Método de retropropagación del gradiente

El método de retropropagación del gradiente es uno de los más importantes algoritmos para entrenar (adaptar los pesos) de una red neuronal multicapa, con al menos una capa oculta y sin ciclos. Consiste en una generalización de la regla delta (o de descenso del gradiente) que usualmente se define en una red de sólo dos capas.

### 4.1 Método de descenso del gradiente (regla delta)

En este método la regla de aprendizaje, entendida como la adaptación sucesiva de los pesos conforme se va entrenando la red con los patrones de entrenamiento, sigue la siguiente regla: Dado un patrón de entrada  $x_p$ , y la salida  $s_p$ , con una salida deseada  $d_p$ , se trata de minimizar el error cuadrático específico al evaluar este patrón  $x_p$ :

Minimizar  $E_p = e_p^2 = (d_p - s_p)^2$  donde  $s_p = \sum_{i=1}^n w_i x_{ip}$ ; por lo tanto:

$$E_p = (d_p - \sum_{i=1}^n w_i x_{ip})^2$$

Dado que  $E_p$  depende de los pesos  $w_i$ , se calcula el gradiente:

$$\frac{\delta E_p}{\delta w_i} = -2(d_p - \sum_{i=1}^n w_i(x_i)_p) = -2e_p(x_i)_p$$

De esta manera, en el paso  $t+1$  se ajustan los pesos en la dirección opuesta al gradiente. En lo sucesivo  $\eta$  es la tasa de aprendizaje y usualmente es más pequeña que uno:

$$w_i(t+1) = w_i(t) + \eta e(x_i)_p \text{ para } i = 1, \dots, n$$

### 4.2 Generalización de la regla delta: método de retropropagación del gradiente

Se trata ahora de extender la regla Delta para entrenar una red neuronal que tenga al menos una capa oculta. En lo que sigue, supondremos que hay una sola capa oculta, pero los cálculos se pueden extender fácilmente a varias capas ocultas.

En una red que sigue como regla de aprendizaje la retropropagación del gradiente, la primera capa de entrada contiene un número  $N$  de unidades que constituye la dimensión en que se representan los patrones de entrenamiento; denotamos un patrón  $x_p = (x_{p1}, x_{p2}, \dots, x_{pN})$ . Conectando la capa de entrada con la oculta hay una matriz de pesos  $w^h = (w_{ji}^h)_{i=1, \dots, N; j=1, \dots, L}$ , donde  $L$  es el número de unidades en la capa oculta,  $w_{ji}^h$  es el peso desde la unidad  $i$  de entrada hasta la unidad oculta  $j$ . Similarmente, hay una matriz de pesos entre la capa oculta y la de salida,  $w^s = (w_{kj}^s)_{j=1, \dots, L; k=1, \dots, M}$ , donde  $M$  es el número de unidades en la capa de salida.

En las unidades de la capa oculta se calcula  $net_{pj}^h = \sum_{i=1}^N w_{ji}^h x_{pi} + \theta_j^h$ . En las unidades de la capa de salida se calcula,  $net_{pk}^s = \sum_{j=1}^L w_{kj}^s i_{pj} + \theta_k^s$ . Aquí,  $i_{pj} = f(net_{pj}^h)$ , con  $f(x) = \frac{1}{1+e^{-x}}$  es una función sigmoide; los términos  $\theta^s$  son opcionales y corresponden a posibles unidades de sesgo que pueden agregarse en la capa de entrada y la oculta.

Formalmente, el entrenamiento parte de un conjunto de pares de vectores de la forma  $(x_p, d_p)$ , donde  $x_p \in \mathbb{R}^N$ ,  $d_p \in \mathbb{R}^M$ ,  $x_p$  es el patrón en sí,  $d_p$  es la salida deseada. Si denotamos  $s_{pk} = f(net_{pk}^s)$  las coordenadas de la salida calculada por el algoritmo, lo deseable es que todas las diferencias entre éstas y la salida deseada tiendan a cero conforme la red se vaya entrenando. La manera usual de plantearlo es minimizando el error  $E_p = \frac{1}{2} \sum_{k=1}^M \delta_{pk}^2$ , con  $\delta_{pk} = (d_{pk} - s_{pk})$ .

Una vez más, se calcula el gradiente de  $E_p$  respecto a los pesos  $w_{kj}^s$  y respecto a los pesos  $w_{ji}^h$ . En [3] se concluye que:

$$\frac{\delta E_p}{\delta w_{kj}^s} = -\delta_{pk} s_{pk} (1 - s_{pk}) i_{pj}$$

$$\frac{\delta E_p}{\delta w_{ji}^h} = -i_{pj} (1 - i_{pj}) x_{pi} \sum_{k=1}^M \delta_{pk} s_{pk} (1 - s_{pk}) w_{kj}^s$$

En seguida, se modifican los pesos de salida en la dirección opuesta al gradiente:

$$w_{kj}^s(t+1) = w_{kj}^s(t) + \eta \delta_{pk}^s i_{pj} \quad \text{donde: } \delta_{pk}^s = \delta_{pk} s_{pk} (1 - s_{pk}).$$

Una vez hecha esta modificación de pesos de la capa de salida, se hace la modificación de los de la capa oculta y debido a que en esta última el gradiente depende de los pesos modificados anteriormente, aquí es en donde se da la propagación hacia atrás:

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \eta \delta_{pj}^h x_{pi}$$

$$\text{donde: } \delta_{pj}^h = i_{pj} (1 - i_{pj}) \sum_{k=1}^M \delta_{pk}^s w_{kj}^s$$

### 4.3 Reconocimiento de letras

A continuación se presenta un ejemplo de red que será implementada en la sección siguiente, junto con el algoritmo de retropropagación del gradiente. En la figura 3 aparece la arquitectura de red para este problema. La intención es que al final del entrenamiento, la red sea capaz de reconocer las letras T y C dispuestas hacia arriba, hacia abajo, a la

izquierda o a la derecha. En total son 8 patrones de entrenamiento (que están en el archivo IoPairsTC) y se usa una red con nueve unidades de entrada correspondientes a las celdas de la cuadrícula, tres unidades en la capa oculta y una sola unidad en la capa de salida. La figura 4 ilustra el comportamiento del error  $E_p$  conforme se hacen las iteraciones; para este ejemplo en particular se observa que tal error tiende a cero, por lo que la red se entrena bien.

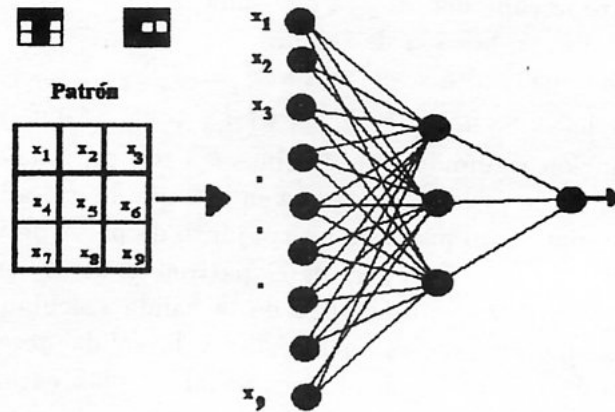


Figura 3: Red del problema T-C

#### 4.4 Implementación

```
(* Modulo Principal *)
bnpStandard[inNumber_,hidNumber_,outNumber_,ioPairs_,
eta_,numIters_] :=
Module[{errors,hidWts,outWts,ioP,inputs,
outDesired,hidOuts,outputs,outErrors,outDelta,
hidDelta},
hidWts = Table[Table[Random[Real,{-0.1,0.1}],{inNumber}],{hidNumber}];
outWts = Table[Table[Random[Real,{-0.1,0.1}],{hidNumber}],{outNumber}];
errors = Table[
(* seleccionar un patr'on *)
ioP = ioPairs[[Random[Integer,{1,Length[ioPairs]}]]];
inputs = ioP[[1]];
outDesired = ioP[[2]];
(* paso hacia adelante *)
hidOuts = sigmoid[hidWts.inputs];
outputs = sigmoid[outWts.hidOuts];
(* determinar errores y deltas *)
outErrors = outDesired-outputs;
outDelta = outErrors (outputs (1-outputs));
hidDelta = (hidOuts (1-hidOuts)) Transpose[outWts].outDelta;
(* actualizar pesos *)
outWts += eta Outer[Times,outDelta,hidOuts];
hidWts += eta Outer[Times,hidDelta,inputs];
(* agregar a la tabla el error cuadr'atico *)
outErrors.outErrors,{numIters}];
(* fin de tabla *)
Return[{hidWts,outWts,errors}]; ];

outs={0,0,0}; (* abrir outs para valores de retorno *)
```

```
outs = bnpStandard[9,3,1,ioPairsTC,1.3,500];
ListPlot[outs[[3]],PlotJoined->True]; (* Gr\afica de los errores *)
```

Patr\on	SalidaRed	d_k	Error
1	{0.135899}	{0.1}	{-0.0358992}
2	{0.869001}	{0.9}	{0.0309986}
3	{0.134872}	{0.1}	{-0.034872}
4	{0.858871}	{0.9}	{0.041129}
5	{0.134404}	{0.1}	{-0.0344043}
6	{0.868927}	{0.9}	{0.0310727}
7	{0.128799}	{0.1}	{-0.0287993}
8	{0.848687}	{0.9}	{0.0513133}

Error medio = 0.0013461186

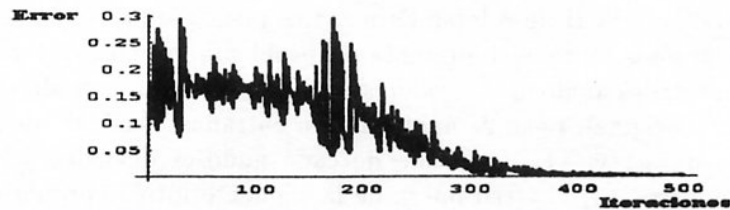


Figura 4: Gráfica de errores

## 5 Clasificación de patrones

Se presenta la red discreta de Hopfield, la cual será usada para la clasificación de patrones. Se describe el propósito y el diseño de esta red. El objetivo de la clasificación de patrones es asignar a los patrones de entrada una de las  $M$  clases disponibles. En lo que sigue vamos a suponer que los patrones de entrada son vectores estáticos  $x$  con  $N$  entradas reales denotadas por  $x_i$ ,  $i = 1, 2, \dots, N$ . Estas entradas representan las medidas de algunos atributos útiles para distinguir entre clases. Los patrones de entrada pueden ser vistos como puntos de un espacio multidimensional definido por las medidas de los atributos. El propósito de un clasificador de patrones es dividir este espacio multidimensional en regiones de decisión que indiquen a cuál clase pertenece cualquier patrón de entrada.

## 6 La red de Hopfield

Una de las más populares arquitecturas de red neuronal es la red de Hopfield, desarrollada por J.J. Hopfield en Caltech. Esta es una red bastante simple, como se muestra en la figura 5, consta de una única capa, en la cual cada unidad se conecta con todas las otras.

Esta red es clasificada como una **memoria asociativa**. Este tipo de memoria parece ser una de las funciones primordiales del cerebro. Usualmente asociamos la cara de un amigo con su nombre, o su nombre con su número telefónico.

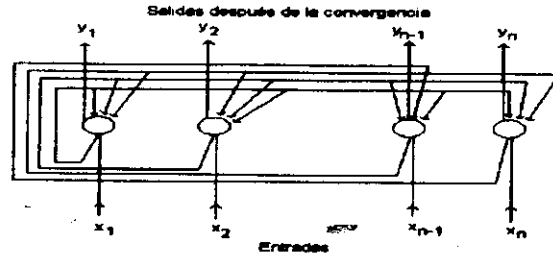


Figura 5: Arquitectura de una red de Hopfield.

La red de Hopfield es más apropiada para cuando los vectores de entrada tienen una representación binaria exacta, como en las imágenes en blanco y negro donde los elementos de entrada son píxeles, o texto ASCII y cada elemento de entrada puede representar caracteres en el código ASCII de 8 bits. Esta red es menos apropiada cuando los valores de entrada son continuos, porque se presenta un problema de representación al tratar de convertir estas cantidades analógicas a valores binarios. La red mostrada en la figura 5, es análoga a la versión original, tiene  $N$  unidades con entradas y salidas binarias, las cuales toman los valores de  $+1$  y  $-1$ . La salida de cada nodo es devuelta a los otros nodos ponderada con los pesos  $w_{ij}$ . La red opera de la siguiente forma: primero, los pesos son calculados usando los patrones de aprendizaje para cada clase. Luego un patrón cuya clase no es conocida se introduce en la red y esta itera actualizándola. Se considera que la red converge cuando las salidas no sufren cambios en las iteraciones sucesivas. El patrón especificado por los nodos de salida después de la convergencia es la salida de la red.

Si la entrada a la red ( $net_i$ ) es negativa el valor de salida es  $-1$ ; si la entrada a la red es positiva, el valor de salida es  $+1$ ; y si la entrada a la red es cero, el valor de salida no cambia.

Llamemos a la función de salida con  $\psi$ , la cual se define como sigue:

```
psi[inValue_,netIn_] := If[netIn > 0, 1,
  If[netIn < 0, -1, inValue]
```

Es conveniente, para simplificar las cosas, definir la función  $\phi$ , la cual toma como argumentos los vectores de entrada ( $x_i$ ) y las entradas de cada unidad y los mapea en la función  $\psi$ .

```
phi[inVector_List,netInVector_List] :=
  MapThread[psi[# ,#2]&,{inVector,netInVector}]
```

El vector de entrada a cada unidad ( $net_i$ ) se calcula de la forma usual como el producto punto del vector de entrada  $x_i$  y el vector de pesos  $w_i$ . Lo que resta es determinar el vector de pesos apropiado para los ejemplares que deseamos almacenar en la red.

Para determinar los pesos vamos a usar el método de aprendizaje conocido como la **regla de Hebb**. Esto se expresa matemáticamente como:

$$\Delta w_{ij} \propto x_i x_j$$

donde  $x_i$  y  $x_j$  son las salidas de las neuronas  $i$  y  $j$ .

Si hay más de un vector en el conjunto de aprendizaje<sup>2</sup>, podemos sumar la contribución de cada peso de los vectores individuales  $x_i$  para obtener la matriz de pesos siguiente:

$$w = \sum_{i=1}^P x_i x_i^t, \quad (1)$$

donde  $P$  es el número de vectores del conjunto de aprendizaje.

Asociado a la red de Hopfield existe una cantidad conocida como la **función de energía**, dada por:

$$E = -\frac{1}{2} x^t w x, \quad (2)$$

donde el vector  $x$  es el vector de entrada a la red. Esta ecuación puede escribirse como:

$$E = -\frac{1}{2} \sum_{i,j=1}^n x_i w_{ij} x_j, \quad (3)$$

donde  $n$  el número de neuronas en la red.

La función de energía para un vector de entrada muy próximo a uno de los patrones de aprendizaje es menos negativa que la energía de los patrones de aprendizaje. Sin embargo, conforme la red procesa el vector de entrada, debemos esperar que la energía del vector de salida evolucione hacia un estado de mínimo local del cual no salga, es decir, la red tratará de buscar un punto fijo.

Observe que una vez que hemos calculado el vector de salida y su energía tenemos dos formas de propagarlo a través de la red. La primera es calcular las entradas a cada neurona ( $net_i$ ) en paralelo y entonces calcular el nuevo vector de salida y la nueva energía. Este método se conoce como **actualización sincrónica**. El segundo método, llamado **actualización asincrónica**, calcula la nueva salida para una neurona a la vez, esta neurona generalmente es escogida al azar, entonces propaga esta nueva salida hacia las otras neuronas antes de escoger una nueva neurona. Aunque el método asincrónico parece ser el más indicado de como opera realmente nuestro cerebro usaremos el método sincrónico porque resulta más simple de programar.

El código para la red discreta de Hopfield aparece abajo. Primero se construye la matriz de pesos a partir del conjunto de aprendizaje esto lo realiza la función `makeHopfieldWts`. Después de calculada esta matriz se ejecuta la función `discreteHopfield`, hasta que el vector de salida no cambie.

```
psi[inValue_,netIn_] :=
  If [netIn >0,1,If[netIn <0,-1,inValue]];
phi[inVector_List,netInVector_List] :=
MapThread[psi[#,#2]&, {inVector,netInVector}];
energyHop[x_,w_] := -0.5 x.w.x;
makeHopfieldWts[traininPats_,printWts_:True] :=
Module[{wtVector},
  wtVector =
  Apply[Plus,Map[Outer[Times,#,#]&,trainPats]]];
```

<sup>2</sup>Al conjunto  $\{(x_1, d_1), (x_2, d_2), \dots, (x_P, d_P)\}$  de ejemplares, se le conoce como conjunto de aprendizaje.



```

    If[printWts,
      Print[];
      Print["La matriz de pesos es:"];
      Print[];
      Print[MatrixForm[wtVector]];
      Print[],Continue
    ]; (* fin del if *)
    Return[wtVector];
  ]; (* fin del Module *)
discreteHopfield[wtVector_,inVector_,printAll_:True]:=
Module[{done,energy,newEnergy,netInput,
  newInput,output},
  done=False;
  newInput=inVector;
  energy=energyHop[inVector,wtVector];
  If[printAll,
    Print[];
    Print["Vector de entrada = ",inVector];
    Print[];
    Print["Energ\{'\{i}a = ",energy];
    Print[], Continue
  ]; (* fin del if *)
  While[!done,
    netInput=wtVector . newInput;
    output=phi[newInput,netInput];
    newEnergy=energyHop[output,wtVector];
    If[printAll,
      Print[];
      Print["Vector de Salida = ", output];
      Print[];
      Print["Energ\{'\{i}a = ",newEnergy];
      Print[], Continue
    ]; (* fin del if *)
    If[energy==newEnergy,
      done = True,
      energy=newEnergy;
      newinput=output,
      Continue
    ]; (* fin del if *)
  ]; (* fin del while *)
  If[!printAll,
    Print[];
    Print["Vector de salida = ",output];
    Print[];
    Print["Energia = ", newEnergy];
    Print[];
  ]; (* fin del if *)
]; (* fin del module *)

```

El comportamiento de la red de Hopfield se ilustra en la figura 5. Una red de Hopfield con 100 unidades y por lo tanto 10.000 pesos fue entrenada para reconocer cuatro patrones, a saber:  $\{C, E, F, H\}$ . Cada una de estas letras en blanco y negro están formadas por 100 pixeles. Los elementos de entrada de la red toman el valor 1 para el negro y  $-1$  para el blanco. En el ejemplo presentado la letra "E" fue alterada al azar invirtiendo cada bit independientemente de  $+1$  a  $-1$  y viceversa con una probabilidad de 0.25. Este patrón fue introducido en la red y en la primera iteración se logra reconocer que el patrón introducido es la letra "E".

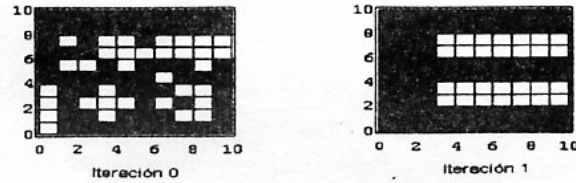


Figura 6: Comportamiento de la red de Hopfield

## 7 Optimización y Satisfacción de Restricciones

En esta sección se estudia el problema del agente viajero que es el clásico ejemplo del tipo de problema que tiene muchas soluciones, pero que una de ellas es considerada la *mejor*. Dada una lista de ciudades y un costo conocido de viajar de una ciudad a otra, el *problema del agente viajero* consiste en encontrar la ruta más eficiente de tal manera que todas las ciudades sean visitadas exactamente una vez, y que el costo total sea mínimo.

Las restricciones impuestas en el problema se dividen en *fuertes* y *débiles*. La restricción de que cada ciudad sea visitada exactamente una vez es fuerte, es decir, cualquier solución al problema debe satisfacerla. La restricción de que el costo sea mínimo es débil porque no todas las posibles soluciones la satisfacen. Seguidamente se analizan algunos detalles del problema y luego se aplica redes neurales para resolverlo.

### 7.1 El problema del agente viajero

Si dada una lista de ciudades uno escogiera aleatoriamente una ruta que pase por cada ciudad aleatoriamente, es muy probable que no sea la más eficiente en términos de costo. Una manera de encontrar la ruta con el mínimo costo es *enumeración*, es decir, encontrar todas las posibles soluciones y los costos respectivos. Sin embargo, esta técnica no es eficiente, pues en un problema con  $n$  ciudades hay  $n!$  rutas posibles. De estas  $n!$  rutas no todas son distintas. Por ejemplo, en un problema con 5 ciudades se considera que las rutas  $1 - 2 - 3 - 4 - 5$ ,  $2 - 3 - 4 - 5 - 1$  y  $5 - 4 - 3 - 2 - 1$  son iguales, pues no importa dónde se empieza, ni la dirección. Así que el número total de rutas distintas es  $n!/2n$ . Para el programa en MATHEMATICA definimos,  $\text{numTours}[n\_]:=n!/2n$ .

En un problema con pocas ciudades uno podría usar enumeración. Por ejemplo, si  $n = 5$ ,  $\text{numTours}[5] = 12$ . Pero si se dobla el número de ciudades, el número de rutas aumenta considerablemente,  $\text{numTours}[10] = 181.440$ . Para analizar el problema del agente viajero generalmente se representan matricialmente los costos de viajar de una ciudad a otra. Como la dirección no importa,  $c_{ij} = c_{ji}$ ,  $\forall i, j = 1, \dots, n$ . Además,  $c_{ii} = 0$ ,  $\forall i = 1, \dots, n$ . Por lo tanto, la matriz de costos es una matriz simétrica con ceros en la diagonal. En MATHEMATICA se generan los costos aleatoriamente para formar una matriz triangular inferior. Luego se transpone para formar una matriz triangular superior. Finalmente se suman las dos matrices triangulares. Para el caso  $n = 5$ , se hace lo siguiente.

```
costs= Table[If[i<=j,0,Random[Integer,{ 1,10}]],{ i,5 },{ j,5}]
costs=costs+Transpose[costs]
```

Para calcular el costo de una ruta en particular, se suman los elementos apropiados. Por ejemplo,

$$t1 = \text{costs}[[1,2]] + \text{costs}[[2,3]] + \text{costs}[[3,4]] + \text{costs}[[4,5]] + \text{costs}[[5,1]]$$

Como la selección aleatoria de una ruta no garantiza que sea la óptima y la enumeración exhaustiva puede ser abrumadora, en problemas como el del agente viajero generalmente uno se conforma con un heurístico que le permita encontrar rápidamente una buena solución. Las Redes neurales proporcionan un heurístico eficiente (con relación a enumeración exhaustiva), pero que no siempre es la mejor. En la siguiente sección se describe cómo se modela tal problema mediante redes neurales.

## 7.2 Redes neurales y el problema del agente viajero

En esta sección se verá cómo se aplica una red de Hopfield a la solución de problemas con restricciones en general y al problema del agente viajero (PAV) en particular. Primero se modifica la red de Hopfield un poco, de tal manera que los valores en las unidades de salida sean funciones continuas del *input* neto y no valores binarios.

### 7.2.1 La Red Continua de Hopfield

Se define  $u_i$  como el *input* neto del  $i$ -ésimo elemento de procesamiento. Recordamos que

$$u_i = \text{net}_i = \sum_{j=1}^n x_j w_{ij}$$

donde  $n$  es el número de unidades conectadas a la unidad  $i$ ,  $x_j$  es la salida de la unidad  $j$  y  $w_{ij}$  es el peso de la conexión de  $j$  a  $i$ . La función de salida utilizada es

$$g(\lambda, u) = 0,5(1 + \tanh[\lambda u]),$$

donde  $\lambda$  es una constante llamada el *parámetro de ganancia*. En particular,  $g[0,5, u]$  es idéntica a la función sigmoide. Denotamos la salida de la unidad  $i$  como  $v_i = g[\lambda, u_i]$ ,  $0 \leq v_i \leq 1$ .

### 7.2.2 Calculando pesos e inputs externos para el PAV

Si el parámetro de ganancia es suficientemente alto, la función de energía se puede escribir como

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n T_{ij} v_i v_j - \sum_{i=1}^n I_i v_i, \quad (4)$$

donde  $T_{ij}$  son los pesos y los  $I_i$  son las señales externas. Debe tenerse presente que una solución al PAV minimizará el valor de la función de energía.

En la red de Hopfield para el PAV cada unidad representa una hipótesis de que visitamos una ciudad particular en un punto particular de la secuencia de la ruta. Por ejemplo,

si tenemos que visitar cinco ciudades, las cuales denotamos con las letras mayúsculas  $A, B, C, D, E$ , la ruta  $B - A - E - C - D$  estaría representada por

$$\overbrace{01000}^A \quad \overbrace{10000}^B \quad \overbrace{00010}^C \quad \overbrace{00001}^D \quad \overbrace{00100}^E$$

o en notación matricial

$$\begin{array}{ccccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{C} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{D} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{E}. \end{array}$$

Supongamos por un momento que la única restricción del problema es que  $n$  ciudades sean visitadas de un total de  $N$ . Esta restricción puede ser representada con la siguiente ecuación:

$$\sum_{i=1}^N v_i = n, \quad (5)$$

donde  $n$  es el número de ciudades y  $N = n^2$  es el número de unidades en la red. La función de energía

$$E = (n - \sum_{i=1}^N v_i)^2 \quad (6)$$

es un mínimo cuando  $n$  de las unidades tengan salida 1. Además si añadimos el término

$$\sum_{i=1}^N v_i(1 - v_i) \quad (7)$$

entonces la función de energía será un mínimo para  $v_i \in \{0, 1\}$ . La nueva función de energía es entonces

$$E = (n - \sum_{i=1}^N v_i)^2 + \sum_{i=1}^N v_i(1 - v_i). \quad (8)$$

Desarrollando el lado derecho de (8) e ignorando el término  $n^2$ , la función de energía puede escribirse como

$$E = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N (-2)v_i v_j - \sum_{i=1}^N v_i(2n - 1) \quad (9)$$

la cual es idéntica a la ecuación (4) tomando

$$T_{ij} = \begin{cases} -2 & i \neq j \\ 0 & \text{otro caso} \end{cases}$$

$$I_i = 2n - 1$$

para todo  $i$ .

Lo anterior resuelve el problema de escoger  $n$ -de- $N$ . Sin embargo, el PAV es más complicado. En un PAV con cinco ciudades, un simple problema  $5 - de - 25$  convergería a una solución, pero ésta no garantiza que la restricción de visitar cada ciudad exactamente una vez sea satisfecha. En un problema de 2 ciudades uno añadiría la restricción de que una ciudad sea escogida entre 1 y 2, y la otra entre 3 y 4. De manera que la restricción fuerte se puede descomponer en problemas  $1 - de - 2$ .

## Bibliografía

- [1] Cohen, M.; Grossberg, S. (1983) "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks", *IEEE Trans. Syst. Man. Cybern.*, **13**: 815–826.
- [2] Duda, R.; Hart, P. (1973) *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York.
- [3] Freeman, J. (1994) *Simulating Neural Networks with Mathematica*. Addison-Wesley, Reading Mass.
- [4] Freeman, J. (1992) "Neural Networks With Mathematica." AI EXPERT.
- [5] Gray, J. (1994) *Mastering Mathematica*. Academic Press, New York.
- [6] Hopfield, J.J.; Tank, D. (1986) "Computing with neural circuits: a model", *Science*, **233**: 625–633.
- [7] Hopfield, J.J. (1984) "Neurons with graded response have collective computational properties like those of two-state neurons", *Proc. Natl. Acad. Sci., USA*, **81**: 3088–3092.
- [8] Hopfield, J.J. (1982) "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci., USA*, **79**: 2554–2558.
- [9] Kaufman, L.; Rousseeuw, P. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, New York.
- [10] Khanna, T. (1994) *Foundations of Neural Networks*. Addison-Wesley, Reading Mass.
- [11] Kohonen, T. (1984) *Self-organization and Associative Memory*, Springer-Verlag, New York.
- [12] Lechevallier, Y. (1995) "Construcción eficaz de una red neuronal a partir de un árbol de decisión", en *Memorias VII-VIII Simposios Métodos Matemáticos Aplicados a las Ciencias*, W. Castillo & J. Trejos (eds.), Editorial de la Universidad de Costa Rica, pp. 53–74.
- [13] Lippmann, R. (1987) "An Introduction to Computing with Neural Nets", IEEE ASSP.
- [14] Rosenblatt, R. (1959) *Principles of Neurodynamics*. Spartan Books, New York.
- [15] Winston, H. (1993) *Artificial Intelligence*. 3rd edition. Addison-Wesley, Reading Mass.

## PROGRAMACIÓN ESTOCÁSTICA EN OPTIMIZACIÓN

SERGIO DE LOS COBOS<sup>1</sup> - BLANCA ROSA PÉREZ<sup>1</sup> - MIGUEL ÁNGEL GUTIÉRREZ<sup>2</sup>

---

### Resumen

Hoy en día, el reto para varias disciplinas es proponer técnicas eficientes de solución, que por un lado garanticen soluciones "buenas", y por otro lado que sean rápidos y fáciles de implementar.

El propósito de este trabajo, es presentar las generalidades de algunos de los métodos heurísticos y de programación estocástica para problemas de optimización combinatoria, consideradas en el Estado del Arte como "sumamente satisfactorios".

Una característica importante de los métodos presentados es que proporcionan marcos generales de solución a una amplia variedad de problemas, no sólo de tipo combinatorio.

**Palabras clave:** Programación estocástica, búsqueda tabú, recocido simulado, algoritmos genéticos.

## 1 Introducción

En situaciones prácticas, uno de los problemas computacionales es la *explosión combinatoria*. La explosión combinatoria se encuentra en aquellas situaciones donde las elecciones están compuestas secuencialmente, permitiendo una vasta cantidad de posibilidades.

Algunos problemas clásicos donde la explosión combinatoria prevalece, muestran que un intento por generar todas las alternativas relevantes por computadora no es una tarea factible. Así por ejemplo, en el problema del agente viajero, en el cual se tiene que salir de una ciudad y regresar a la misma después de haber visitado (con costo de viaje mínimo) todas las demás ciudades, si se tienen  $n$  ciudades en total que recorrer entonces, dada una ciudad inicial existen  $(n - 1)!$  soluciones factibles, y si una computadora que pudiera ser programada para examinar soluciones a razón de un billón de soluciones por segundo; la computadora terminaría su tarea, para  $n = 25$  ciudades (que es un problema pequeño para muchos casos prácticos) en alrededor de 19,674 años.

---

<sup>1</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA, DEPARTAMENTO DE MATEMÁTICAS, AV. PURÍSIMA Y MICHOACÁN, DEL. IZTAPALAPA, C.P. 09340, MÉXICO D.F.; E-MAIL: COBOS@XANUM.UAM.MX.

<sup>2</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - AZCAPOTZALCO, DEPARTAMENTO DE SISTEMAS, SAN PABLO 180, DEL. AZCAPOTZALCO, C.P. 02200, MÉXICO D.F.

Aunque para algunos problemas de optimización como son los problemas de programación lineal, los problemas del transporte, los problemas de asignación, existen algoritmos rápidos y eficientes para muchos otros problemas de optimización combinatoria no los hay, como son los casos del problema de asignación cuadrático, del problema de localización de plantas y del problema del agente viajero, entre otros.

El principal problema de algunos algoritmos heurísticos es su inhabilidad para enfrentarse con puntos de optimalidad local, o en otras palabras, su inhabilidad de continuar la búsqueda hacia el óptimo global después de que un óptimo local se ha alcanzado.

Lo anterior ha propiciado que el enfoque de la *inteligencia artificial* haya revivido como solución de problemas difíciles. Recientemente varias aproximaciones han surgido del manejo de problemas de decisión complejos, como son: *algoritmos genéticos*, *redes neuronales*, *recocido simulado*, *búsqueda tabú*, *análisis de objetivos* y *búsqueda dispersa* entre otros, algunos de éstos, conocidos como métodos de *programación estocástica*

La búsqueda tabú, junto con la técnica del recocido simulado y los algoritmos genéticos, han sido singularmente calificados por el Committee on Next Decade of Operations Research (Condor(1988)) como “extremadamente promisorios” para el tratamiento futuro de aplicaciones prácticas. Cabe mencionar que estas técnicas son las que se reportan en el Estado del Arte y se utilizan con mayor frecuencia para abordar problemas “difíciles”, y su potencial reportado es “ilimitado”.

## 2 Programación combinatoria

Los problemas de optimización se dividen naturalmente en dos categorías: problemas de optimización con variables continuas y problemas de optimización con variables discretas. A estos últimos se les llama problemas de *optimización combinatoria*. En problemas de tipo continuo, se busca la solución sobre un conjunto con ciertas propiedades de continuidad y generalmente de convexidad; en los problemas de optimización combinatoria se busca la solución sobre un conjunto finito o infinito numerable. Estas dos clases de problemas tienen diferentes retos y los métodos para resolverlos son distintos.

### 2.1 Planteamiento del problema combinatorio

El problema de optimización combinatoria tiene la siguiente forma:

$$(P) \text{ Minimizar } C(x) : x \in X \subset R^n.$$

La función objetivo puede ser lineal o no lineal, y la condición  $x \in X$  se asume restringida a componentes específicos de  $x$  con valores discretos, las restricciones pueden ser incluso de tipo lógico.

**Definición 2.1** *Se define un movimiento  $s$  como una función definida sobre un subconjunto  $X(s)$  de  $X$  de la siguiente forma:*

$$s : X(s) \longrightarrow X.$$

Es decir, se considera que un *movimiento* es una transición de una solución factible a otra solución factible (transformada), el cual se puede describir mediante un conjunto de uno o varios *atributos*. Este mecanismo lo utilizan diferentes métodos, en general es utilizado por los métodos de optimización clásica como son los de programación lineal, programación entera, programación no lineal, etc.

Una **instancia** de un problema de optimización combinatoria puede formalizarse como una pareja  $(S, f)$  donde  $S$  denota el conjunto finito de todas las soluciones posibles y  $f$  la **función de costo**, mapeo definido por

$$f : S \rightarrow R. \quad (1)$$

En el caso de minimización, el problema es encontrar  $i_{opt} \in S$  que satisfaga

$$\forall i \in S, f(i_{opt}) \leq f(i), \quad (2)$$

en el caso de maximización, la  $i_{opt}$  que satisfaga

$$\forall i \in S, f(i_{opt}) \geq f(i). \quad (3)$$

A la solución  $i_{opt}$  se le llama una **solución globalmente óptima** y  $f_{opt} = f(i_{opt})$  denota el costo óptimo, mientras que  $S_{opt}$  denota el conjunto de soluciones óptimas.

De lo anterior, se puede decir que, un problema de **optimización combinatoria** es un conjunto  $I$  de instancias de un problema de optimización combinatoria.

**Definición 2.2** Sea  $(S, f)$  una instancia de un problema de optimización combinatoria. Entonces una **estructura de vecindades** es un mapeo

$$\mathcal{N} : S \rightarrow 2^S, \quad (4)$$

que define para cada solución  $i \in S$  un conjunto  $S_i \subset S$  de soluciones que son “cercanas” a  $i$  en algún sentido. El conjunto  $S_i$  se llama **vecindad** de la solución  $i$  y cada  $j \in S_i$  se llama un **vecino** de  $i$ . Además se supone que  $j \in S_i \Leftrightarrow i \in S_j$ .

**Definición 2.3** Sea  $(S, f)$  una instancia de un problema de optimización combinatoria y  $\mathcal{N}$  una estructura de vecindades. Entonces un **mecanismo de generación** es un medio para seleccionar una solución  $j$  de la vecindad  $S_i$  de la solución  $i$ .

**Definición 2.4** Si  $(S, f)$  denota una instancia de un problema de optimización combinatoria y sea  $\mathcal{N}$  una estructura de vecindades, entonces  $\hat{i} \in S$  se llama una **solución óptima local** o simplemente un **óptimo local** con respecto a  $\mathcal{N}$  si  $\hat{i}$  es mejor que o igual a, todas sus soluciones vecinas con respecto al costo. Específicamente, en el caso de minimización,  $\hat{i}$  se llama **solución mínima local** o simplemente un **mínimo local** si

$$\forall j \in S_i, f(\hat{i}) \leq f(j), \quad (5)$$

y en el caso de maximización,  $\hat{i}$  se llama una **solución máxima local** o simplemente un **máximo local** si

$$\forall j \in S_i, f(\hat{i}) \geq f(j). \quad (6)$$



**Definición 2.5** Si  $(\mathcal{S}, f)$  denota una instancia de un problema de optimización combinatoria y sea  $\mathcal{N}$  una estructura de vecindades. Entonces  $\mathcal{N}$  se llama **exacta** si para cada  $\hat{i} \in \mathcal{S}$  que es localmente óptimo con respecto a  $\mathcal{N}$ ,  $\hat{i}$  también es globalmente óptimo.

Por lo que un problema de *optimización combinatoria* es un conjunto digamos  $I$  de instancias de un problema de optimización combinatoria, de manera informal, una instancia está dada por los “datos de entrada” y la información suficiente para obtener una solución mientras que un problema es una colección de instancias del mismo tipo.

## 2.2 Problemas P y NP

Un aspecto importante de los problemas de optimización en general, es el *esfuerzo computacional*. Por ejemplo, el esfuerzo computacional para resolver el problema de asignación por medio del método Húngaro está en relación polinomial respecto al crecimiento del tamaño del problema, en cambio, para el problema del agente viajero, el esfuerzo computacional está en relación exponencial.

Aquellos problemas para los que se conoce un algoritmo con esfuerzo computacional de tipo polinomial, se dice que pertenecen a la clase  $P$ , en cambio, aquellos problemas de esfuerzo computacional no polinomial, por ejemplo de tipo exponencial, se dice que están en la clase  $NP$  (non-deterministic polynomial), como se observó anteriormente.

Un concepto importante es el de *reducibilidad o transformabilidad*, para el cual, suponga que se tiene un problema  $p_1$  con un algoritmo de solución conocido digamos  $a$ . Si se puede transformar cada instancia de otro problema  $p_2$  dentro de una instancia de  $p_1$  en tiempo polinomial, entonces se puede utilizar el algoritmo  $a$  para resolver a este último problema. Por lo que, podemos decir que  $p_1$  es al menos tan *difícil* de resolver como  $p_2$ .

Si cada problema NP es polinomialmente reducible a un problema  $p$ , entonces se dice que el problema  $p$  es *NP-difícil* (NP-hard). Si además, el problema  $p$  es un problema NP, entonces se dice que  $p$  es un problema *NP-completo* (NP-complete). Esto es muy importante puesto que nos lleva a un problema abierto en la actualidad: ¿ $P=NP?$ , puesto que, si se puede encontrar un algoritmo polinomial para un problema NP-completo, entonces se habrá encontrado un algoritmo polinomial para todos los problemas NP.

Para los problemas de optimización en la clase NP-completa, a la fecha, no se conocen algoritmos que los resuelvan en tiempo polinomial. Es por ello que se han desarrollado algoritmos de tipo heurístico para resolver instancias grandes de problemas que pertenecen a esta clase. Estos algoritmos, no necesariamente encuentran la mejor solución al problema, pero en general se pueden obtener “buenas” soluciones cuando se aplican, aquí se entiende “buena” solución en términos de cercanía al valor óptimo. La ventaja principal de este tipo de algoritmos es que son rápidos y corren en un tiempo polinomial.

## 2.3 Optimización, satisfacibilidad y semioptimización

Por ejemplo, en el problema del agente viajero se requiere de un trabajo de *optimización*: encontrar la trayectoria o ruta que sea tan barata como cualquier otra trayectoria o ruta factible. En otros problemas de optimización, ahora bien, el objetivo no es sólo el exhibir un objeto formal que satisfaga un conjunto de criterios establecidos sino también averiguar

que cualidades no marcadas posee en el espacio de candidatos. Ejemplo de este estilo es el problema de las N-reinas (véase por ejemplo Laguna (1993)) donde se requiere de la *satisfacibilidad* de que las reinas no se maten entre sí.

La mayoría de los problemas pueden poseer ambos tipos de trabajos. Por ejemplo, el problema de las N-reinas se puede declarar como sigue: encontrar el emplazamiento más barato de las N reinas en el tablero de ajedrez donde cada reina capturada introduce un costo de una unidad. Puesto de esta manera, el problema llega a ser muy difícil ya que no se conoce *a priori* cuando una solución de costo cero existe.

Siempre la diferencia en complejidad entre un trabajo de optimización y su contraparte de satisfacibilidad es substancial. El problema del agente viajero es un ejemplo de tal situación: encontrar una ruta a través de un conjunto de ciudades es trivial en el caso de que existan todas los posibles caminos entre cualesquiera dos ciudades, pero encontrar una ruta optimal es un problema NP-completo. En tales casos, uno se ve forzado a relajar los requerimientos de optimalidad y moderar para encontrar una "buena" solución utilizando sólo un razonable esfuerzo de búsqueda. Se dice que se tiene un problema de *semioptimización* (suboptimización) si hay un criterio de aceptación que tolera una vecindad alrededor de la solución optimal.

La mayoría de los problemas combinatorios prácticos son del tipo de semioptimización, por lo que se requiere de un balance razonable entre la calidad de la solución encontrada y el costo de búsqueda de tal solución. El problema básico al manejar un trabajo de semioptimización es el de idear algoritmos que garanticen cotas tanto en el esfuerzo de búsqueda como en su extensión para el cual el objetivo de optimización está comprometido. Un trabajo aún más ambicioso es el de equipar a tales algoritmos con un conjunto de parámetros ajustables que el usuario pueda cambiar para negociar entre la calidad de la solución y el monto de esfuerzo requerido.

La característica sobresaliente de los métodos de programación estocástica presentadas en este trabajo es su aplicación general y la habilidad para obtener soluciones arbitrariamente cercanas a la óptima. Sin embargo, el obtener soluciones de alta calidad puede requerir de mucho esfuerzo computacional. Una comparación importante que puede hacerse con respecto a si una técnica es general o no es la que se presenta con respecto a la programación lineal y la programación dinámica. La programación lineal es un modelo muy específico para solucionar problemas en donde la función objetivo y las restricciones deben ser lineales; si el problema no cae dentro de este esquema restringido, la programación lineal no puede aplicarse. Por otro lado, la programación dinámica se puede usar para resolver muchos problemas de optimización y su aplicabilidad depende de la habilidad que se tenga para definir las etapas del problema, las ecuaciones recursivas, las variables de estado y las variables de decisión para un problema específico. En este sentido, las técnicas aquí presentadas resultan ser análogas a la programación dinámica. El uso y la calidad de sus resultados dependen del arte y la habilidad con que se definan sus diferentes componentes como son: las evaluaciones de la función objetivo de la nueva solución con respecto a la anterior, la estructura de vecindades y el grado de refinamiento de la implantación en la computadora.

### 3 Búsqueda Tabú

La Búsqueda Tabú (BT) es un procedimiento heurístico introducido y desarrollado en su forma actual por Fred Glover (1989) y (1990a).

“La filosofía de la BT es la de manejar y explotar una colección de principios para resolver problemas de manera inteligente. Uno de los elementos fundamentales de la BT es el uso de la memoria flexible, desde el punto de vista de la BT, la memoria flexible envuelve el proceso dual de crear y explotar estructuras para tomar ventaja mediante combinar actividades de adquisición, evaluación y mejoramiento de la información de manera histórica” (Glover y Laguna (1993)).

Para un ejemplo numérico completo donde se ejemplifican diferentes aspectos de la BT, ver de los Cobos(1994). En términos generales, el método BT puede esbozarse consistente en:

1. Se desea moverse paso a paso desde una solución factible inicial de un problema de optimización combinatoria hacia una solución que proporcione el valor mínimo de la función objetivo  $C$ . Para esto, se puede representar a cada solución por medio de un punto  $s$  (en algún espacio) y se define una vecindad  $N(s)$  de cada punto  $s$ .

2. El paso básico del procedimiento consiste en empezar desde un punto factible  $s$  y generar un conjunto de soluciones en  $N(s)$ ; entonces se escoge al mejor vecino generado  $s^*$  y se posiciona en ese nuevo punto ya sea que  $C(s^*)$  tenga o no mejor valor que  $C(s)$ .

3. La característica importante de la búsqueda tabú es precisamente la construcción de una lista tabú  $T$  de movimientos: aquellos movimientos que no son permitidos (movimientos tabú) en la iteración presente. La razón de esta lista es la de excluir los movimientos que nos pueden regresar a algún punto de una iteración anterior. Además, las condiciones tabú tienen la meta de prevenir ciclos e inducir la exploración de nuevas regiones.

4. Las restricciones tabú no son inviolables bajo toda circunstancia. Cuando un movimiento tabú proporciona una solución mejor que cualquier otra encontrada, su clasificación tabú puede eliminarse. La condición que permite dicha eliminación se llama *criterio de aspiración*. Las restricciones tabú, permiten que un movimiento se observe como admisible si no se aplican, mientras que el criterio de aspiración permite que un movimiento se observe como admisible si se satisface.

En Glover y Laguna (1993), se indica que las aspiraciones son de dos clases: *aspiraciones de movimiento* y *aspiraciones de atributo*. Una aspiración de movimiento, cuando se satisface, revoca la clasificación tabú del movimiento. Una aspiración de atributo, cuando se satisface revoca el estatus tabú del atributo. En éste último caso el movimiento puede o no cambiar su clasificación tabú, dependiendo de sí la restricción tabú puede activarse por más de un atributo.

Los valores de movimiento por lo general proporcionan una base fundamental para evaluar la calidad de un movimiento. El mejor movimiento se selecciona del conjunto de

movimientos *admisibles*. Un movimiento es admisible si es no tabú o si su estatus tabú puede eliminarse por medio del criterio de aspiración. El mejor movimiento entonces se realiza y la estructura de datos tabú se actualiza.

La memoria de término corto de la BT constituye una forma de exploración que busca realizar el mejor movimiento posible, para satisfacer ciertas restricciones.

Además se requiere de una estructura de datos para guardar el seguimiento de los movimientos que son clasificados como tabú y para liberar aquellos movimientos de su condición tabú cuando su pertenencia a la memoria de término corto expire. El acompañamiento de la memoria basada en la pertenencia junto con la memoria basada en la frecuencia adicionan el efecto de que se puede estipular una historia selectiva de los estados encontrados durante la búsqueda, y reemplazando la vecindad actual por una vecindad modificada que depende de este proceso histórico.

5. Ahora bien, conforme la búsqueda progresa, la forma de la evaluación empleada por la búsqueda tabú llega a ser más adaptativa, incorporando referencias concernientes para la *intensificación* y la *diversificación* regional de búsqueda. Cabe aclarar que en las estrategias basadas en consideraciones de término corto la clasificación tabú sirve para identificar elementos de la vecindad del movimiento actual, mientras que en las estrategias de término intermedio y largo pueden no contener soluciones en esta vecindad, sino que por lo general consisten de seleccionar soluciones *élites* (óptimos locales de alta calidad) encontrados en varios puntos en el proceso de solución.

La longitud de la lista tabú es un parámetro, si es demasiado pequeño el ciclado ocurrirá, pero si es demasiado grande, restringirá bastante la búsqueda para poder saltar “valles profundos” (i.e. el mejor mínimo local, ver de los Cobos y Col.(1996)) del espacio de valores de la función objetivo.

La matriz de frecuencias es la que lleva la “historia” del procedimiento y es la que se utiliza para la formación de la función de memoria de término largo, la cual permite la diversificación de la búsqueda, es decir, es posible el dirigir la búsqueda “más cercana” ó “más alejada” de las regiones exploradas.

En el método de BT, por ejemplo, el “mejor” movimiento que se realiza en cada iteración se especifica como el movimiento admisible con el menor valor objetivo. Ahora bien, esta estrategia no garantiza que el movimiento seleccionado permita la búsqueda en la dirección de la solución optimal, por lo que, se requiere de técnicas que nos permitan integrar las estrategias de intensificación y diversificación de manera efectiva, basándose sobre las funciones de memoria de término intermedio y largo de la BT.

Si en alguna iteración ya no existen puntos admisibles, se tendrá que utilizar ahora las funciones de memoria de término intermedio (intensificación) y de término largo (diversificación).

La fase de intensificación proporciona una forma simple para enfocar la búsqueda al rededor de la mejor solución (o conjunto de soluciones élites) hasta el momento.

Una aproximación que está cercanamente unida a los orígenes de la BT y que proporciona un interjuego efectivo entre la intensificación y la diversificación es la *estrategia de oscilación*.

La estrategia de oscilación opera mediante el moverse hasta alcanzar una frontera, representada por la factibilidad o un estado de construcción que normalmente puede representarse por un punto donde el método puede parar. En vez de parar, ahora bien, la definición de vecindad se extiende o el criterio de evaluación para seleccionar movimientos se modifica, para permitir que se pueda cruzar esa frontera. La aproximación entonces procede para una profundidad específica más allá de la frontera y entonces se regresa. En este punto la frontera de nuevo se aproxima y se cruza, esta vez desde la dirección opuesta, procediendo a un nuevo punto en turno. El proceso de aproximar y cruzar repetidamente la frontera desde diferentes direcciones crea una forma de oscilación que es la que le da el nombre a la estrategia. El control sobre esta oscilación se establece mediante el generar evaluaciones y reglas de movimientos modificadas, dependiendo de la región en la cual se está actualmente navegando y de la dirección de la búsqueda.

## 4 Recocido Simulado

### 4.1 El proceso de recocido de un sólido

El algoritmo de recocido simulado está basado en una analogía entre la simulación de recocido de sólidos y la problemática de resolver problemas de optimización combinatoria de gran escala. Por esta razón el algoritmo se conoce como **recocido simulado**. Recocido denota un proceso de calentamiento de un sólido a una temperatura en la que sus granos deformados recrystalizan para producir nuevos granos. La temperatura de recocido o de recrystalización, depende del tipo de material, del grado de deformación del mismo, además de su uso futuro. Seguida a la fase de calentamiento, viene un proceso de enfriamiento en donde la temperatura se baja poco a poco. De esta manera, cada vez que se baja la temperatura, las partículas se reacomodan en estados de más baja energía hasta que se obtiene un sólido con sus partículas acomodadas conforme a una estructura de cristal. Si se comienza con un valor máximo de la temperatura, en la fase de enfriamiento del proceso de recocido, para cada valor de la temperatura  $T$  debe permitirse que se alcance su equilibrio térmico. Sin embargo, si el proceso de enfriamiento es demasiado rápido y no se alcance en cada etapa el equilibrio térmico, el sólido congelará en un estado cuya estructura será amorfa en lugar de la estructura cristalina de más baja energía. La estructura amorfa está caracterizada por una imperfecta cristalización del sólido.

El equilibrio térmico está caracterizado por la distribución de Boltzmann (Toda y col.(1983)). De acuerdo a esta distribución, la probabilidad que el sólido esté en un estado  $i$  con energía  $E_i$  a la temperatura  $T$ , viene dada por

$$P_T\{X = i\} = \frac{1}{Z(T)} \exp\left(\frac{-E_i}{k_B T}\right), \quad (7)$$

donde  $X$  es una variable aleatoria que denota el estado actual del sólido.  $Z(T)$  es una constante de normalización llamada la **función partición**, que esta definida como

$$Z(T) = \sum_j \exp\left(\frac{-E_j}{k_B T}\right), \quad (8)$$

donde la sumatoria se extiende sobre todos los posibles estados y  $k_B$  es una constante física conocida como la constante de Boltzmann. El factor  $\exp\left(\frac{-E_i}{k_B T}\right)$  se conoce como el factor de Boltzmann. La igualdad (7) es una función de densidad de probabilidad.

Por lo dicho anteriormente, el proceso de recocido consta de dos pasos fundamentales que son:

- Incrementar la temperatura del baño térmico a un valor máximo.
- Decrementar **cuidadosamente** la temperatura del baño térmico hasta que las partículas se reacomoden por sí mismas en un estado de mínima energía, denominado el estado fundamental del sólido.

Dado un estado  $i$  del sólido con energía  $E_i$ , se genera un estado subsecuente  $j$  aplicando un mecanismo de perturbación que transforma el estado actual en el siguiente estado por medio de una pequeña distorsión, por ejemplo, por el desplazamiento de una partícula. La energía del siguiente estado es  $E_j$ . Si la **diferencia de energía**,  $E_j - E_i$ , es menor o igual a cero, el estado  $j$  se acepta como el estado actual. Si la diferencia de energía es mayor que cero, el estado  $j$  se acepta con una probabilidad que esta dada por

$$\exp\left(\frac{E_i - E_j}{k_B T}\right), \quad (9)$$

donde  $T$  denota la temperatura del baño térmico y  $k_B$  es la constante de Boltzmann. La regla de decisión descrita arriba se conoce como el **criterio de Metropolis** y al algoritmo se le conoce como **algoritmo de Metropolis** (Metropolis y col.(1953)).

## 4.2 El algoritmo de recocido simulado

La simulación del proceso de recocido puede usarse para describir un proceso de generación de sucesiones de soluciones de un problema de optimización combinatoria en donde se vaya obteniendo, conforme el proceso avanza, mejores soluciones al mismo. Para este propósito, se puede observar una analogía entre el sistema físico y un problema de optimización combinatoria en donde cada solución del problema de optimización combinatoria puede verse como un estado del sólido y los valores de la función objetivo para cada solución del problema de optimización combinatoria como los niveles de energía del sólido. En resumen, se puede pensar en las siguientes equivalencias.

- Las soluciones de un problema de optimización combinatoria son equivalentes a los estados de un sistema físico.
- El costo de una solución es equivalente a la energía de un estado.

Además, se introduce un parámetro que juega el papel equivalente de la temperatura. Este parámetro se llama el **parámetro de control**. El algoritmo de recocido simulado puede verse como una iteración del algoritmo de Metropolis, evaluado en valores decrecientes del parámetro de control.

Como en los algoritmos de búsqueda local se supone la existencia de una estructura de vecindades y un mecanismo de generación. Se introducen las siguientes definiciones.

**Definición 4.1** Sea  $(S, f)$  denote una instancia de un problema de optimización combinatoria y denote por  $i$  y  $j$  dos soluciones con costo  $f(i)$  y  $f(j)$ , respectivamente. Entonces el criterio de aceptación determina si  $j$  se acepta de  $i$  a partir de aplicar la siguiente probabilidad de aceptación:

$$P_c\{\text{aceptar } j\} = \begin{cases} 1 & \text{si } f(j) \leq f(i) \\ \exp\left(\frac{f(i)-f(j)}{c}\right) & \text{si } f(j) > f(i), \end{cases} \quad (10)$$

donde  $c \in R^+$  denota el parámetro de control.

Claramente, el mecanismo de generación corresponde al mecanismo de perturbación en el algoritmo de Metropolis, mientras que el criterio de aceptación corresponde al criterio de Metropolis.

**Definición 4.2** Una transición es una acción combinada que transforma la solución actual en una subsecuente. Esta acción consiste de los siguientes dos pasos: (i) aplicación del mecanismo de generación, (ii) aplicación del criterio de aceptación.

Sea  $c_k$  denote el valor del parámetro de control y  $L_k$  el número de transiciones generadas en la  $k$ -ésima iteración del algoritmo de Metropolis. Entonces el algoritmo de recocido simulado puede describirse en pseudo-código como se muestra en la Figura 4.1

```

Comienza
  INICIALIZA ( $i_{inicial}, c_0, L_0$ )
   $k := 0$ 
   $i := i_{inicial}$ 
Repite
  para  $l := 1$  a  $L_k$  haz
    Comienza
      GENERA ( $j$  de  $S_i$ )
      si  $f(j) \leq f(i)$  entonces  $i := j$ 
      sino
        si  $\exp\left(\frac{f(i)-f(j)}{c_k}\right) >$  numero aleatorio en  $[0, 1)$  entonces  $i := j$ 
    finpara
     $k := k + 1$ 
    CALCULA-LONGITUD ( $L_k$ )
    CALCULA-CONTROL ( $c_k$ )
hasta criterio de paro
fin

```

Figura 4.1 Algoritmo de Recocido Simulado en pseudo-código

### 4.3 Aspectos generales

En las aplicaciones del algoritmo de recocido simulado comúnmente se desea implementarlo de manera que la sucesión de soluciones estén generadas a partir de valores decrecientes del parámetro de control. Las soluciones se generan continuamente tratando de transformar la solución actual en una subsecuente por medio de aplicar los mecanismos de generación y el criterio de aceptación. Las aplicaciones del algoritmo de recocido simulado requieren de la especificación de los siguientes puntos: (i) Una representación concisa del problema, (ii) un mecanismo de transición y (iii) un programa de enfriamiento. Cada uno de estos puntos se pueden enumerar con más detalle.

1. Una descripción concisa de la **representación del problema** consiste de una representación del espacio de soluciones y una expresión de la función de costo. La función de costo debe escogerse de manera que represente la efectividad de las soluciones con respecto al objetivo de optimización.
2. La generación de ensayos para transformar la solución actual en una subsecuente consiste de tres pasos. Primero, se debe generar una nueva solución aplicando un mecanismo de generación. Enseguida se debe calcular la diferencia de costo de las dos soluciones, por último, se hace una decisión de aceptar o no, la nueva solución. La evaluación de la nueva solución es lo que más tiempo consume en el algoritmo de recocido simulado y por lo tanto debe hacerse lo más eficientemente posible. El cálculo de la diferencia de costo se hace tomando en cuenta las diferencias entre ambas soluciones. Para muchas aplicaciones éste es el camino más rápido para calcular las diferencias en el costo. La decisión de aceptar la nueva solución se basa en un criterio de aceptación (vea definición 4.1).
3. Ejecutar el proceso de recocido, requiere de la especificación de los parámetros que determinan el programa de enfriamiento. Estos parámetros son el valor inicial del parámetro de control, una función que especifique el decremento del parámetro de control, la longitud de cada bloque donde permanece constante el parámetro de control y el criterio de paro.

En Aarts y Korst(1989), se muestra que si se selecciona el mecanismo de generación, el criterio de aceptación como en la definición 4.1, una estructura de vecindades y un programa de enfriamiento adecuado, el algoritmo de recocido simulado **converge con probabilidad 1** al conjunto de soluciones óptimas.

## 5 Algoritmos genéticos

Al reflexionar sobre la efectividad de la estrategia de la evolución biológica, parece obvio que las reglas de la evolución biológica son el resultado de un proceso de evolución en sí mismas. Suponga una población de organismos existentes con reglas hereditarias escasamente modificadas con la norma existente. Si esas modificaciones ayudan a la población para adaptarse más rápidamente a su medio ambiente particular, entonces esta población



tiene mayor oportunidad de sobrevivir en el futuro que una población con menos reglas hereditarias efectivas. Por tanto, la evolución biológica, da en sí misma un modelo optimal de operación. Esto origina la idea que los conceptos estrategicos de la evolución biológica pueden utilizarse con éxito para crear algoritmos, modelos y máquinas “inteligentes”.

El término técnico de *estrategia de evolución* para un programa, tiene el objetivo de copiar las reglas de la evolución biológica en la forma más exacta. Así de lejos, la imitación de la evolución de un modo simplificado ha probado ser muy exitoso. Con el tiempo se ha llegado a describir diferentes estilos de operaciones de evolución.

Una idea general de los algoritmos genéticos (AG), es que se pueden ver como una búsqueda aleatoria inteligente, en el sentido de obtener un muestreo aleatorio del espacio de soluciones.

Los algoritmos genéticos fueron desarrollados por Holland en la Universidad de Michigan por los 60's. El nombre de algoritmos genéticos se origina de la analogía entre la estructura genética de los cromosomas y de la representación de las soluciones de problemas por medio de anillos o vectores.

## 5.1 Componentes básicos de los AG

Los algoritmos genéticos están basados en la noción de la propagación de nuevas soluciones a partir de *soluciones padres*, empleando mecanismos modelados de aquellos que se aplican en genética. La mejor descendencia de las soluciones padres se retiene para una nueva generación, por lo que al proceder en forma de evolución se fomenta la *supervivencia de los más aptos*. Como la calidad de los más aptos (mejor descendencia) eventualmente construye el más alto nivel compatible con el medio ambiente (el problema es gobernado por restricciones), la mejor sobre todas las descendencias se registra y es el candidato propuesto por el método para una solución optimal.

Un algoritmo genético básico tiene tres operadores: *reproducción*, *cruzamiento* y *mutación*. La reproducción es un apareamiento aleatorio de individuos (soluciones muestra) de una población para crear una o más descendencias. El cruzamiento define el resultado como un cambio de gene (plan reproductivo), cuyo valor específico es conocido como *alelo*, *i.e.*, los aleles pueden concebirse como instancias en el sentido de sistemas expertos. El cambio de genes (tipo de información y sus atributos) siguen las reglas posicionales tradicionalmente modelada después de la reproducción biológica. Deben modificarse y particularizarse a diferentes tipos de problemas combinatorios, ahora bien, para que tenga sentido se debe permitir ciertas relaciones de restricciones para proporcionar una oportunidad al progénito de realizar un mejoramiento sobre sus padres.

Finalmente, una mutación es simplemente el introducir un elemento aleatorio, muchas veces usado para enmendar el resultado de un gene cambiado cuando el resultado no se encuentra exitoso bajo las restricciones apropiadas. A este respecto, la mutación está más fuertemente sesgada para ayudar en la aplicación de los algoritmos genéticos que en la genética biológica. La mutación diversifica el espacio de búsqueda y protege de la pérdida de material genético que puede darse en la reproducción y el cruzamiento.

En términos generales podemos considerar que los AG son técnicas de búsqueda aleatoria que imitan a los procesos observados en la evolución natural. Combinan la super-

vivencia de los más aptos (o mejores) entre estructuras de anillos (soluciones) con un estructurado cambio de información aleatoria.

Los algoritmos genéticos difieren de las técnicas de optimización tradicionales en muchos aspectos. Ellos trabajan con una codificación de las variables (típicamente como anillos o vectores) más que con las variables en sí mismas, y utilizan las reglas de transición probabilística para moverse de una población de soluciones a otras, más que de una solución sencilla a otra. La más interesante e importante característica de los AG es que ellos utilizan tan solo evaluaciones de la función objetivo. Esto es, los AG no utilizan ninguna información sobre diferenciabilidad, convexidad o alguna otra característica auxiliar. Esto hace que los AG sean fáciles de utilizar y de implementar en gran variedad de problemas de optimización.

Existen muchas variaciones de los algoritmos genéticos debido al uso de diferentes operadores de reproducción, cruzamiento y mutación. En general, un algoritmo genético para resolver un problema combinatorio debe contener cinco componentes principales (Davis y Steenstrup(1987)):

1. Una representación cromosómica de las configuraciones del espacio de búsqueda del problema.
2. Una forma de generar la población inicial de soluciones.
3. Una función de evaluación que juega el papel del medio ambiente, calificando las configuraciones en términos de su "valor de ajuste".
4. Una descripción precisa de las operaciones genéticas que alteran la composición de los "hijos" durante la reproducción.
5. Los valores de los parámetros que el algoritmo genético utiliza (tamaño de la población, probabilidades de aplicar las operaciones genéticas, etc.)

Holland y Golberg, desarrollan el concepto de *esquema* o patrón de similaridad para analizar el problema de convergencia para algoritmos que trabajan con representaciones de cadenas de bits. Un esquema es un patrón en el que ciertas posiciones del cromosoma son fijas. Por ejemplo, el esquema  $1*0**1$ , comprende a todos los cromosomas cuyo primer allele es 1, tercer allele es 0 y su sexto allele es 1. Note que \* representa posiciones en que no estamos interesados y que pueden tomar cualquier valor del alfabeto genético utilizado. A través de la noción de esquemas el teorema fundamental de los algoritmos genéticos, proporciona una cota inferior a la evaluación para un tipo de esquema en una generación, por lo que, se pueden mantener "buenos" esquemas después de un cierto número de generaciones. Esta cota depende del número de posiciones fijas del esquema, de la distancia entre las posiciones fijas y del criterio de evaluación de la generación anterior.

## Bibliografía

- [1] Aarts, E. y Korst, J.,(1989) *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, New York.

- [2] Cerny V. (1985) “Thermodynamical Approach to the Traveling Salesman Problem: An efficient simulation algorithm”, *Journal of Optimization Theory and Applications* **45**: 41–52.
- [3] Colin R.R. (Ed.) (1993) *Modern Heuristic Techniques for Combinatorial Problems*. John and Wiley Sons, New York.
- [4] Committee on the Next Decade of Operations Research (Condor 1988) “Operations Research: The Next Decade,” *Ops. Res.* **36**.
- [5] Davis L. and Streenstrup, M. (1987) “Genetic Algorithms and Simulated Annealing”, L. Davis (ed.), Pitman, London.
- [6] de los Cobos Silva S. G.; Pérez Salvador B.R. y Gutierrez Andrade. M.A. (1995) “Lineamientos de implantación de la búsqueda tabú para los problemas de calendari-zación”, IX Simposio *Métodos Matemáticas Aplicados a las Ciencias*, J. Trejos (ed.) (Turrialba, Costa Rica), pp. 23–31.
- [7] de los Cobos Silva S. G., Gutierrez Andrade. M.A. y Pérez Salvador B.R., (1996) “Un ejemplo de cómo saltar valles profundos en problemas combinatorios.”, *Revista Contactos*, No. 15, pp. 36–41.
- [8] Espinoza J. L., (1995) “Algoritmos Genéticos y Optimización de Funciones”, Memorias IX *Simposio Métodos Matemáticas Aplicados a las Ciencias*, J. Trejos (ed.) (Turrialba, Costa Rica) pp. 33–42.
- [9] Fox B. (1993) “Integrating and accelerating tabu search, simulated annealing, and genetic algorithms”, *Annals of Ops. Res.* **41**: 47–67.
- [10] Glover F. (1989) “Tabu search, Part I”, *ORSA Journal on Computing* **1**(3): 190–206.
- [11] Glover F. (1990) “Tabu search: a tutorial”, *Interfaces* **20**(4): 74–94.
- [12] Glover F. (1990a) “Tabu search, Part II”, *ORSA Journal on Computing* **2**(1): 4–31.
- [13] Glover F. (1993) “Genetic algorithms and scatter search: unsuspected potentials”, Technical report (August), School of Busines, University of Colorado at Boulder.
- [14] Glover F. and Laguna M. (1993) “Tabu Search”, in: *Modern Heuristic Techniques for Combinatorial Problems*, Colin R. Reeves (ed.), 70–150, Blackwell Scientific Publications, Oxford.
- [15] Goldberg D. E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Mass.
- [16] Gutierrez Andrade M. A.; de los Cobos S. S. G.; Pérez S. B. R. (1995) “El problema de localización de plantas: Análisis de técnicas de solución”, Memorias IX *Simposio Métodos Matemáticas Aplicados a las Ciencias*, J. Trejos (ed.) (Turrialba, Costa Rica) pp. 57–66.

- [17] Gutierrez Andrade M. A., de los Cobos S. S. G., Pérez S. B. R. (1995) "Búsqueda tabú: un procedimiento heurístico para solucionar problemas de optimización combinatoria", Rev. *En LINEA*, (aceptado).
- [18] Gutierrez Andrade M. A., de los Cobos S. S. G., Pérez S. B. R. (1996) "*Optimización con recocido simulado para el problema del conjunto independiente*", Rev. *En LINEA*, por publicarse.
- [19] Holland J. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.
- [20] Laguna M. and Glover F. (1993) "Integrating Target Analysis and Tabu Search for Improved Scheduling Systems", *Expert Systems with Application* **6**, 287-297.
- [21] Metropolis, N., Rosenbluth M., Rosenbluth A., Teller A., and Teller E. (1953) "Equations of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, **21**, pp. 1087-1092.
- [22] Murillo, A.(1996) "Particionamiento usando búsqueda tabú", IV *Encuentro Centroamericano de Investigación Matemática*, Antigua, Guatemala.
- [23] Piza, E. y Trejos, J. (1995) "Particionamiento usando sobrecalentamiento simulado y algoritmos genéticos", *Memorias IX Simposio Métodos Matemáticas Aplicados a las Ciencias*, J. Trejos (ed.) (Turrialba, Costa Rica) pp. 121-132.
- [24] Piza, E. y Trejos, J.(1996), "Clasificación automática: particionamiento mediante sobrecalentamiento simulado, IV *Encuentro Centroamericano de Investigación Matemática*, Antigua, Guatemala.
- [25] Rechenberg I. (1964) "Cybernetic solution path of an experimental problem", Royal Aircraft Establishment, Library Translation 1122, Farnborough.
- [26] Rechenberg I. (1989) *Artificial Evolution and Artificial Intelligence, Evolution in Medicine; Learning; Principles and Techniques*, Chapman and Hall, London.
- [27] Trejos Z. J., Piza V. E. y Figueroa M. G. (1996) *Clasificación automática mediante un algoritmo genético: resultados numéricos*, IV *Encuentro Centroamericano de Investigación Matemática*, Antigua, Guatemala.
- [28] Toda, M., Kubo, R. and Saito, M. (1983) *Statistical Physics*, Springer-Verlang.

# ANÁLISIS TIEMPO-FRECUENCIA DE SEÑALES CARDIOVASCULARES BAJO DIFERENTES CONDICIONES HEMODINÁMICAS

RAÚL F. JIMÉNEZ<sup>1</sup> - BRUNO GÜNTHER<sup>2</sup> - ERIC ZABALA<sup>3</sup>

---

## Resumen

Se utiliza el análisis tiempo-frecuencia, dado por la transformada en ondelettes continua (CWT), para estudiar las presiones arteriales bajo diferentes condiciones hemodinámicas, obtenidas de perros anestesiados mediante catéteres electrónicos. Este nuevo método matemático para analizar señales entrega dos representaciones bidimensionales: el módulo y la fase en plano tiempo-frecuencia y una representación tridimensional en el espacio tiempo-escala-frecuencia. El presente trabajo tiene como principal meta el estudio de los cambios de la frecuencia cardíaca bajo diferentes posiciones del animal (45°); de horizontal a cabeza arriba, de horizontal a cabeza abajo, oclusión manual de ambas arterias carótidas. Además, se estudia la variabilidad de la frecuencia cardíaca con un sencillo algoritmo ideado por nosotros, y el algoritmo de la CWT se implementó en un PC usando el software Matlab.

## 1 Introducción

Hoy en día es práctica común someter los datos numéricos (señales) obtenidos experimentalmente al análisis de Fourier. Este método entrega un espectro en el dominio de las frecuencias, mientras que la CWT permite un análisis de la frecuencia en función del tiempo, es decir, los métodos de Fourier entregan figuras estáticas del espectro, mientras que la CWT entrega una representación dinámica. En trabajos anteriores [5, 6] hemos aplicado la transformada discreta en ondelettes (DWT) para analizar las pulsaciones de presión y de velocidad de la sangre en perros anestesiados con respiración espontánea; en [7] comparamos esos resultados con respiración artificial y apnea; en [9] los efectos de la hemorragia. El presente trabajo trata con los cambios en la frecuencia cardíaca producidos por cambios de posición del animal, usando la CWT sobre pulsaciones de presión

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD DE CONCEPCIÓN, CHILE, CASILLA 4009; E-MAIL: RJIMENEZ@GAUSS.CFM.UDEC.CL

<sup>2</sup>DEPARTAMENTO DE FISIOLÓGIA Y BIOFÍSICA, ESCUELA DE MEDICINA, UNIVERSIDAD DE CHILE, SANTIAGO, CHILE

<sup>3</sup>DEPARTAMENTO DE INGENIERÍA ELÉCTRICA, UNIVERSIDAD DE CONCEPCIÓN, CHILE

sanguínea. Finalmente, para investigar el control reflejo del sistema cardiovascular, se analizan los efectos de la oclusión manual de ambas arterias carótidas.

### 1.1 Reseña de la transformada continua en ondelettes (CWT)

La CWT entrega una representación visual en el plano tiempo-frecuencia de las señales, lo que permite estudiar el comportamiento en frecuencia en todo instante, analizar transitorios y detectar impulsos y discontinuidades. Este método ha sido aplicado en diversos campos (procesamiento de sonidos, imágenes, sismología, mecánica, fractales, ver [10]), sin embargo, en fisiología hay pocos trabajos ([3] y [12]).

### 1.2 Definiciones

La CWT de una señal real  $f(t)$  respecto de una **ondelette analizante**  $\psi(t)$  se define por

$$CWT(s, \tau) = \frac{1}{\sqrt{s}} \int_{\mathbb{R}} f(t) \psi^* \left( \frac{t - \tau}{d} \right) dt \quad s > 0, \tau \in \mathbb{R} \quad (1)$$

donde  $\psi^*(t)$  denota la compleja conjugada de  $\psi(t)$ . Esta transformada está definida en un semiplano abierto "tiempo-escala"  $H$ , pero es usual considerar todo el plano, tomando  $-\log s$  como eje vertical donde las bajas frecuencias se encuentra en la parte baja del plano, según se muestra en la Figura 1. Para detalles ver [4].

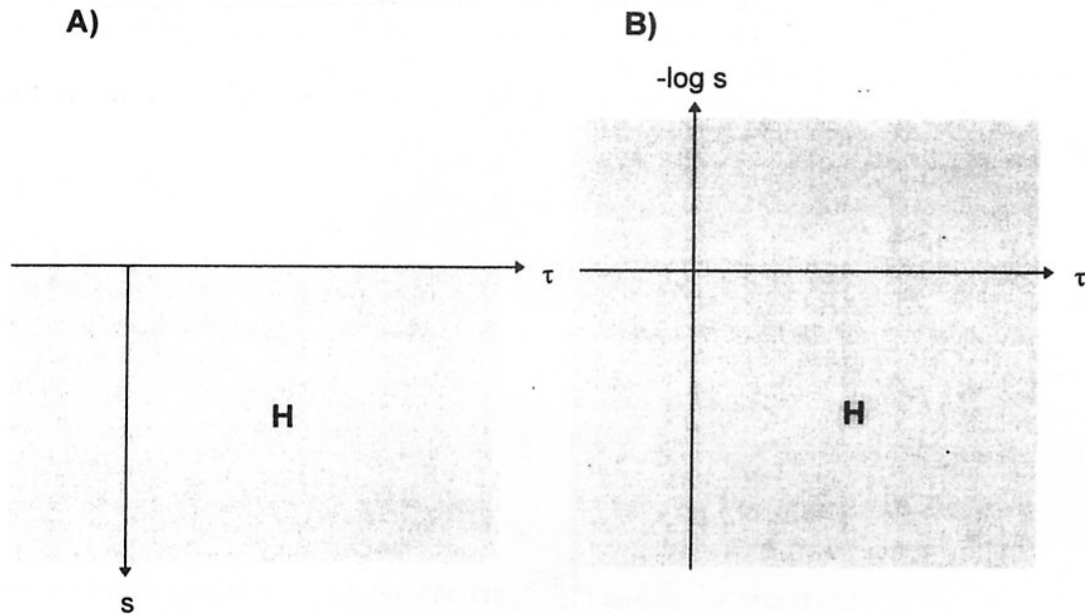


Figura 1: Semiplano  $H$  y plano  $H$  tiempo-escala

La ecuación (1) puede escribirse en términos de la transformada de Fourier  $f^*(t)$  y  $\psi^*(t)$  como

$$CWT(s, \tau) = \sqrt{s} \int f^*(\omega) \psi^*(s\omega) e^{j\tau\omega} d\omega, \quad (2)$$

donde hemos impuesto sobre  $\psi$  la **condición de admisibilidad**:

$$K_\psi = 2\pi \int |\psi^*(\omega)| d\omega |\omega|^{-1} < \infty. \quad (3)$$

Si  $\psi^*(\omega)$  es diferenciable, lo que suponemos, esto implica que  $\psi^*(0) = 0$ , ice  $\psi$  es media nula.

La motivación principal de la condición de admisibilidad es que implica la convergencia débil de

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle f, \psi_{s,\tau} \rangle \psi_{s,\tau} \frac{ds d\tau}{s^2} \quad (4)$$

donde  $\psi_{s,\tau}(t) = s^{-1/2} \psi((t - \tau)/s)$  y  $\langle \cdot, \cdot \rangle$  es el producto interno en  $L^2(\mathbb{R})$ , el espacio de las señales de energía finita.

Los coeficientes de ondelettes de  $f(t)$  son los productos interiores de  $L^2$  definidos por

$$\langle f, \psi_{s,\tau} \rangle = \int_{\mathbb{R}} f(t) \psi_{s,\tau}(t) dt, \quad s > 0, \tau \in \mathbb{R}. \quad (5)$$

Si en cierta zona de la señal hay bajas frecuencias, los coeficientes son grandes; si la señal es casi constante, los valores de los coeficientes son casi nulos. De esta manera la magnitud de los coeficientes de ondelettes indican el contenido espectral de la señal a una cierta escala  $s$ . Cuando  $s$  varía, los coeficientes cubren un amplio rango en frecuencia; a valores grandes de la escala  $s$  corresponden bajas frecuencias. Cambiando el parámetro (podemos mover el centro de localización en el eje del tiempo y cubrir así toda la señal (cada ondelette está centrada en el punto  $t = \tau$ ).

Este análisis es dependiente de la elección de la ondelette analizante; en este trabajo hemos elegido la ondelette de Morlet:

$$\psi(t) = \pi^{-1/4} e^{iw_0 t} e^{-t^2/2}, \quad \omega_0 \geq 5, \quad \text{donde } \psi^*(\omega) = \pi^{-1/4} e^{-(\omega - \omega_0)^2/2} \quad (6)$$

donde y en las escalas hemos usado un ancho de 8 octavas con 8 voces por octava, donde por voz entendemos la restricción de  $CWT(s_i, \tau)$  para valores discretos del parámetro  $s$ . De este modo el eje vertical está medido en octavas de suerte que los valores en la escala  $-s$ :  $-2, -1, 0, 1, 2, \dots$  corresponden los valores  $2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, \dots$  en Hz.

Para visualizar las prestaciones de esta transformada aplicada a una señal test, ver la Figura 2.; en B) presentamos el módulo de los coeficientes (notar que son números complejos) y en C) presentamos la fase de:

$$CWT(s, \tau) = |CWT(s, \tau)| e^{i\varphi(s, \tau)}, \quad 0 \leq \varphi(s, \tau) < 2\pi$$

de la señal test. En D) presentamos una representación 3D del módulo, que como veremos mas adelante, entrega información complementaria a la entregada por el módulo.

### 1.3 Implementacion del algoritmo

Los algoritmos para calcular la  $CWT$  requieren gran cantidad de memoria y rapidez; por ello es común el uso de estaciones de trabajo. Sin embargo, nosotros hemos logrado

implementar este algoritmo en un PC usando la *FFT* y el software MatLab. Incluso hemos mejorado las prestaciones de esta transformada eliminando el efecto de los bordes mediante un simple truco. Muy brevemente, diremos que aplicando traslaciones y dilataciones sobre la ondelette de Morlet, llegamos a la expresión

$$\psi(t, \tau, s) = e^{-j\omega_0(\frac{t-\tau}{s})} e^{-\frac{(\tau-t)^2}{2s^2}}$$

Reordenamos esta ecuación de manera que las funciones resultantes sean un producto de convolución en el tiempo:

$$CWT(s, \tau) = s^{-1/2} \int_{\mathbf{R}} f(t) e^{-j\omega \frac{\omega_0}{s}(\tau-t)} e^{-\frac{(\tau-t)^2}{2s^2}} dt$$

es decir

$$CWT(s, \tau) = s^{-1/2} \text{conv}(f(t), \psi(t, \tau, s)).$$

Entonces, discretizando y aplicando la transformada directa de Fourier a cada función y aplicando la operación inversa, obtenemos

Notar que el parámetro  $\tau$  está implícito en la convolución, por lo tanto la *CWT* sólo depende del parámetro de escala  $s$ , constituyéndose en el parámetro de la transformada y de ahí el nombre *continuo*.

La elección usual de la escala  $s$  es en potencias de 2, para el cual el rango en frecuencias está dividido en octavas. Además cada octava se divide en voces para dar **continuidad** a la representación gráfica.

Un gran número de octavas implica una mejor visualización pero, a veces los rangos de frecuencia de interés resultan muy pequeños. Para nuestras señales cardiovasculares es suficiente considerar 8 voces por octava.

## 2 Resultados

En la Figura 3 A) mostramos las pulsaciones de presión en condiciones normales, observándose cambios periódicos en la diástole y sístoles producidos por la respiración. En este caso la frecuencia media es de 138 mmHg. En la Figura 3 B) mostramos el módulo de la *CWT*, observándose que la mayor energía de la señal, es decir, una componente fundamental, (color rojo) se encuentra alrededor de  $2 \pm 0.5$  Hz; valor que corresponde a la frecuencia cardíaca del animal. En esta misma zona aparecen periódicamente manchas de color púrpura que corresponden a un fenómeno de **modulación en amplitud (AM)**, que ya fue descrito por nosotros en [6, 7], cuyo origen es aun desconocido y que no es detectable a simple vista ni con los métodos clásicos de Fourier. En la Fig. 3 C) presentamos la fase, donde los ciclos respiratorios están ubicados alrededor del valor  $2^{-3}$  octavas, correspondiente a 0.125 Hz. En el intervalo [3,4] octavas, es decir, en la zona de altas frecuencias, aparecen componentes de altas frecuencias de las pulsaciones arteriales, en particular, el cierre de las válvulas sigmoideas (incisura). En la Fig. 3 D) presentamos la representación en 3D del módulo, mirada desde la zona de bajas frecuencias y donde es posible observar con claridad el fenómeno de AM, como "cimas" de una cordillera.



En la Fig. 4 A) presentamos la respuesta cardiovascular a un cambio de posición de horizontal a cabeza arriba del animal, en un ángulo de  $45^{\circ}$ . Es notable observar los dramáticos cambios producidos por la autorregulación. En la Fig. 4 B) observamos: una pérdida de la energía al comenzar el tilting, los intentos de recuperación y un fenómeno de muy baja frecuencia cuyo origen desconocemos, por el momento. En la Fig. 5 A) presentamos la respuesta cardiovascular a un cambio de posición de horizontal a cabeza abajo del animal, en un ángulo de  $45^{\circ}$ . En la Figura 6 A) presentamos la autorregulación producida tras la oclusión de ambas arterias carótidas.

### Agradecimientos

Nuestra gratitud a la Srta. Sylvia Gutiérrez por su excelente asistencia técnica. Este trabajo fue apoyado por el P.I. 94.15.08-1 de la Universidad de Concepción, Dirección de Investigación.

### Bibliografía

- [1] FARKAS, G. A.; BAER, R. E.; ESTENNE, M.; DE TROYER, A. (1988) "Mechanical role of expiratory muscle during breathing in upright dogs", *J. Appl. Physiol.* **64**: 1060–1067.
- [2] GORINI, M.; ESTENNE, M. (1991) "Effect of head-up tilt on neural inspiration drive in the anesthetized dog", *Resp. Physiol.* **85**: 83–96.
- [3] GRAMATIKOV, B.; GEORGIEV, I. (1995) "Wavelets as alternative to short-time Fourier transform in signal-averaged electrocardiography", *Med. Bio. Eng. Comput.*, **33**(3): 482–487.
- [4] GROSSMANN, A.; KRONLAND-MARTINET, R.; MORLET, J. (1989) "Reading and understanding wavelet transforms", In *Wavelets: Time-Frequency Methods and Phase Space*, J.M. Combes, A. Grossmann & Ph. Tchamitchiam (eds.), Springer-Verlag, Berlin: 2–20.
- [5] GÜNTHER, B.; JIMÉNEZ, R. F.; VEGA, J. C. (1991) "Análisis por transformada en ondelettes de las presiones del pulso en el sistema cardiovascular", *Actas V Congreso Internacional de Biomatemática*, Instituto Profesional de Santiago, Santiago, Chile: 186–195.
- [6] GÜNTHER, B.; JIMÉNEZ, R. F.; PICARTE, C. (1993) "Wavelet analysis of arterial pressure and blood velocity pulsations in the aorta of anesthetized dogs", *Biol. Res.* **26**: 391–396.
- [7] GÜNTHER, B.; JIMÉNEZ, R. F. (1996) "Wavelet analysis of aortic pulsations in anesthetized dogs: Effects of spontaneous and positive-pressure respiration", *J. Biol. Sys.* **4**(1): 45–59.

- [8] JIMÉNEZ, R. F. (1991) "Detection of discontinuities by wavelet transforms", In: *Appl. Math. Engrg. Sci.*, C. Carasso, C. Conca, R. Correa & J.-P. Puel (eds.), Cépaduès, Toulouse: 277-304.
- [9] JIMÉNEZ, R. F. (1995) "Estudio de pulsaciones arteriales mediante la transformada en ondelettes", In: *Síndrome de Shock, Aspectos históricos, hemodinámicos, bioquímicos y clínicos*, Vivaldi & Hernández (eds.), Universidad de Concepción, Facultad de Ciencias Biológicas.
- [10] MEYER, Y.; ROQUES, S. (eds) (1993) *Progress in Wavelet Analysis and Applications*. Editions Frontières, France.
- [11] MORLET, D.; COUDERC, J. P.; TOUBOUL, P.; RUEL, P. (1995) "Wavelet analysis of high-resolution ECGs in post-infarction patients; role of the basic wavelet and of the analyzed lead", *Int. J. Biomed. Comput.* **39**(3): 311-325.

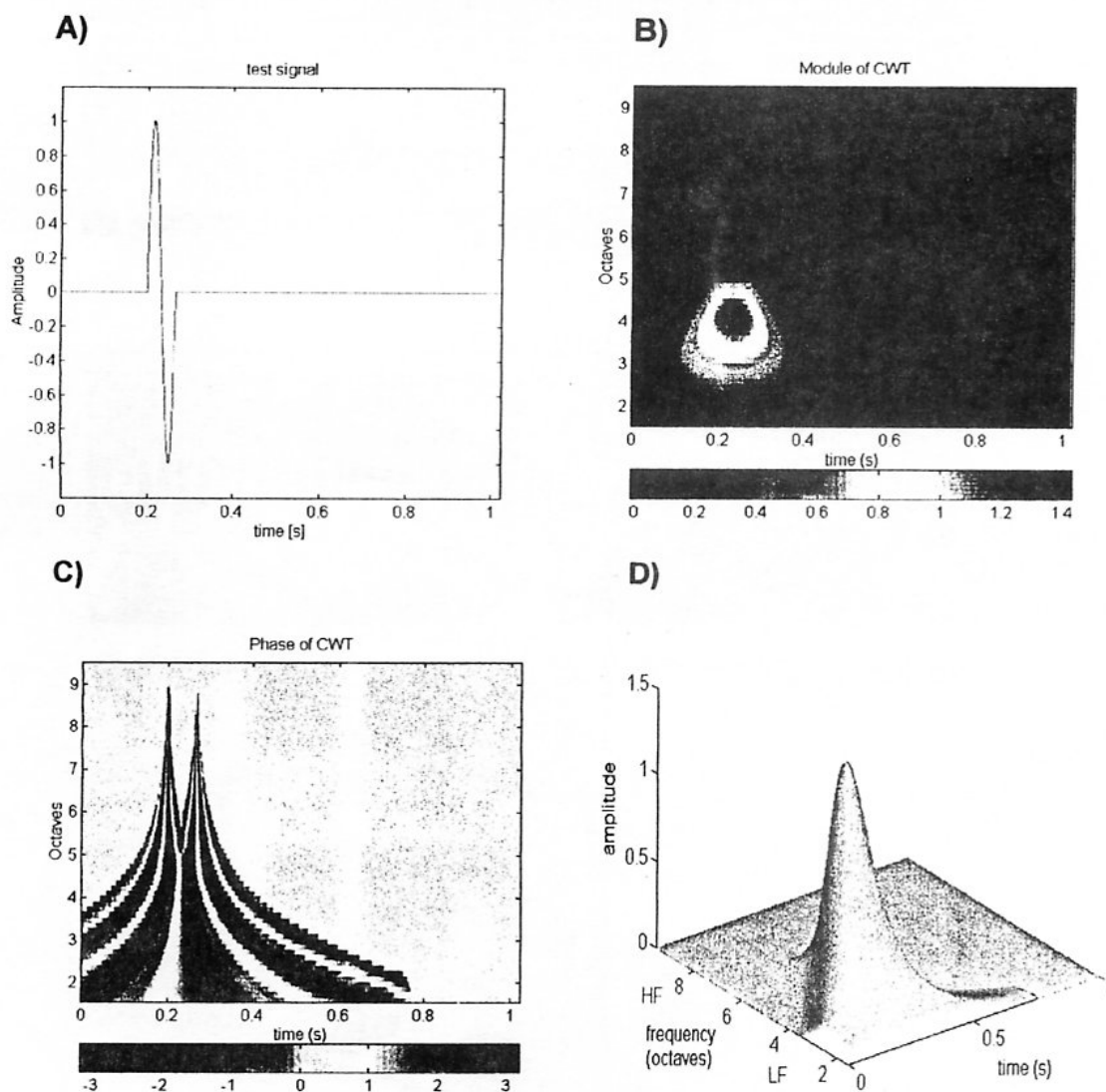


Figura 2. Análisis tiempo-frecuencia de una señal test usando la transformada continua de ondelettes.

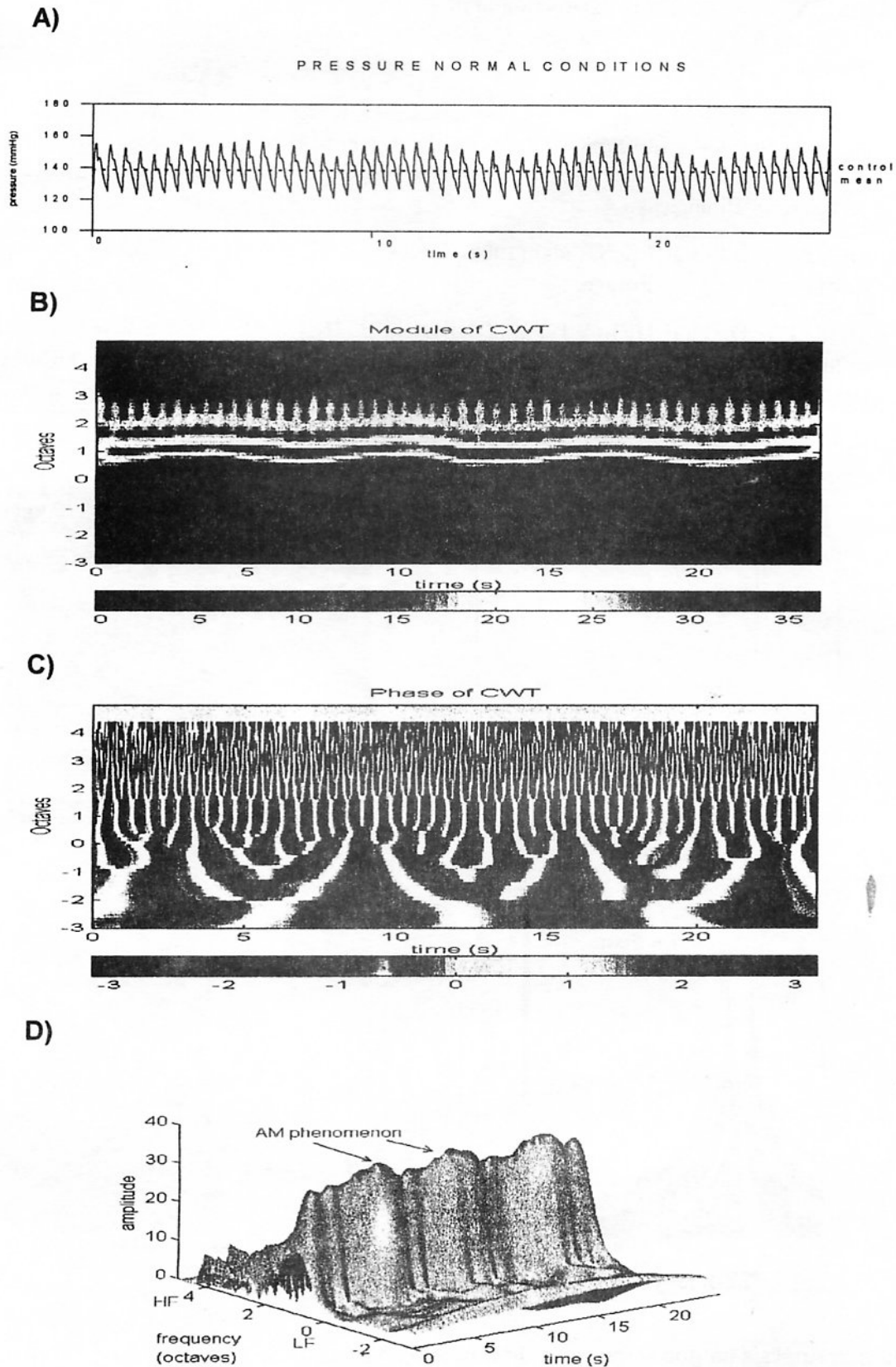
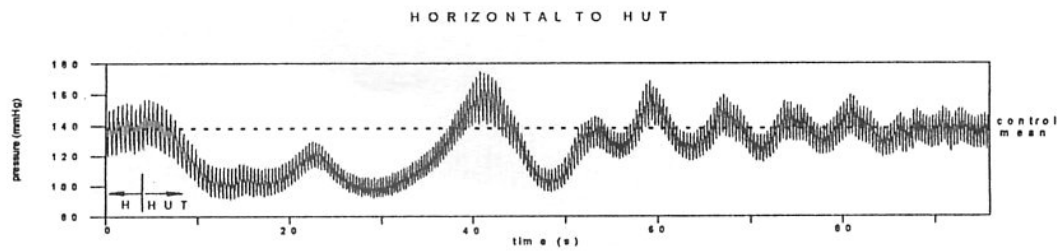
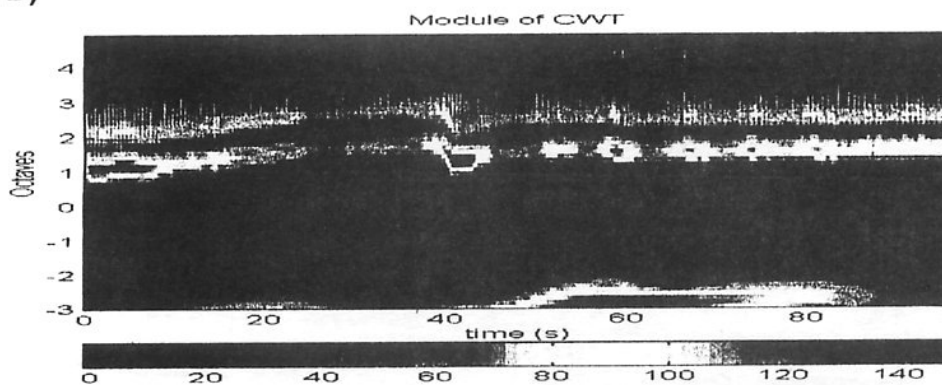


Figura 3. Análisis tiempo-frecuencia de pulsos de presión en condiciones normales.

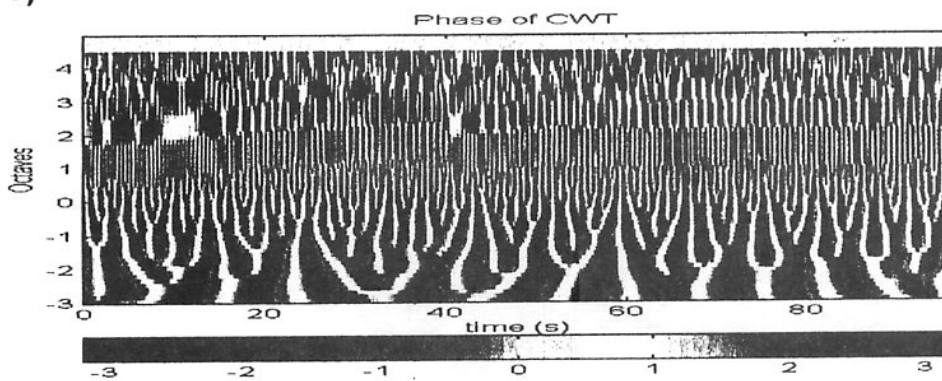
A)



B)



C)



D)

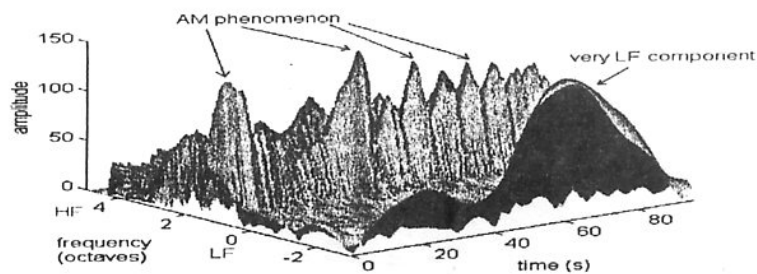


Figura 4. Análisis tiempo-frecuencia de señales de presión arterial durante el cambio de posición de horizontal (H) a cabeza arriba (HUT).

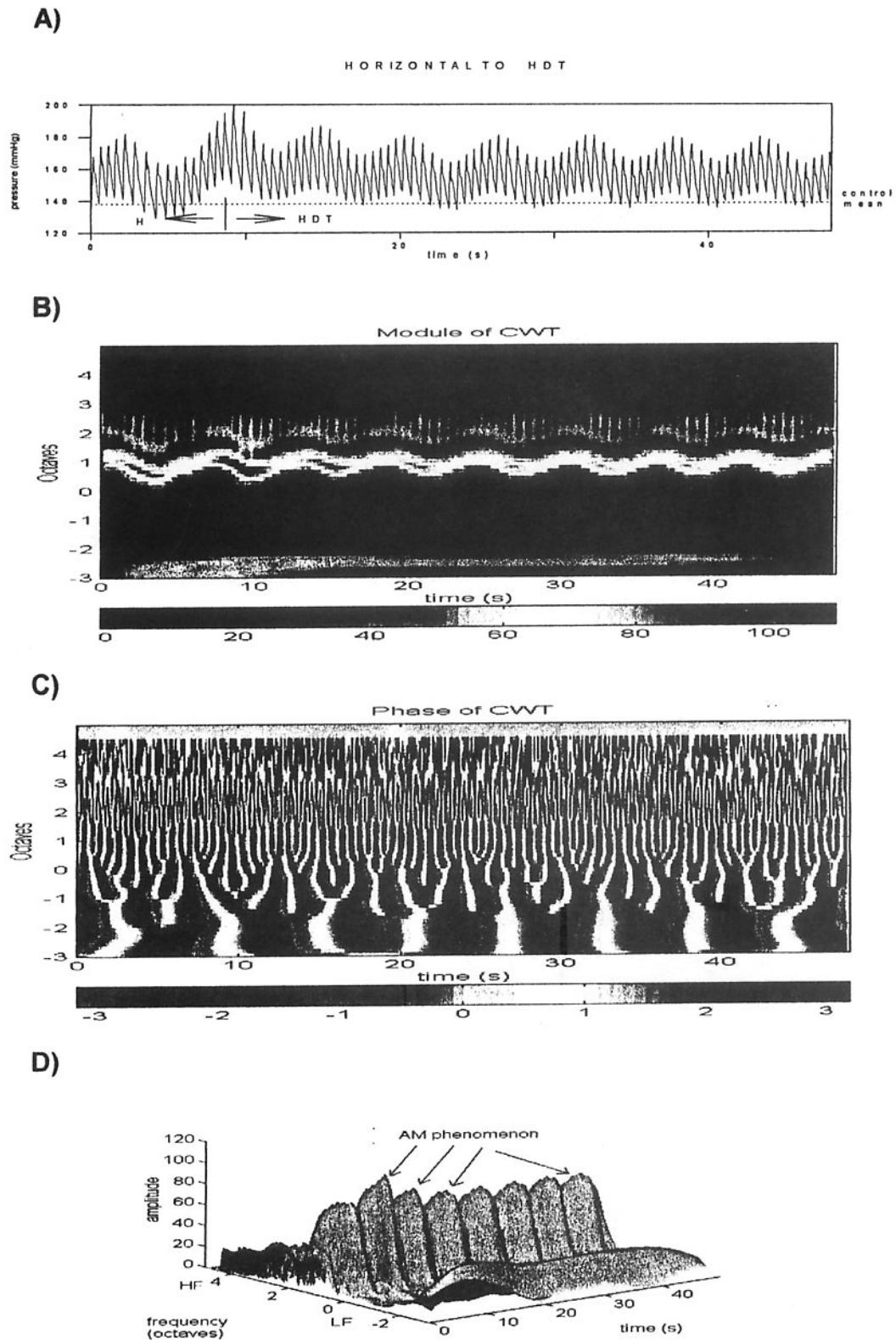


Figura 5. Análisi tiempo-frecuencia de pulsaciones arteriales por cambio de posición de horizontal a cabeza abajo en  $45^{\circ}$ .

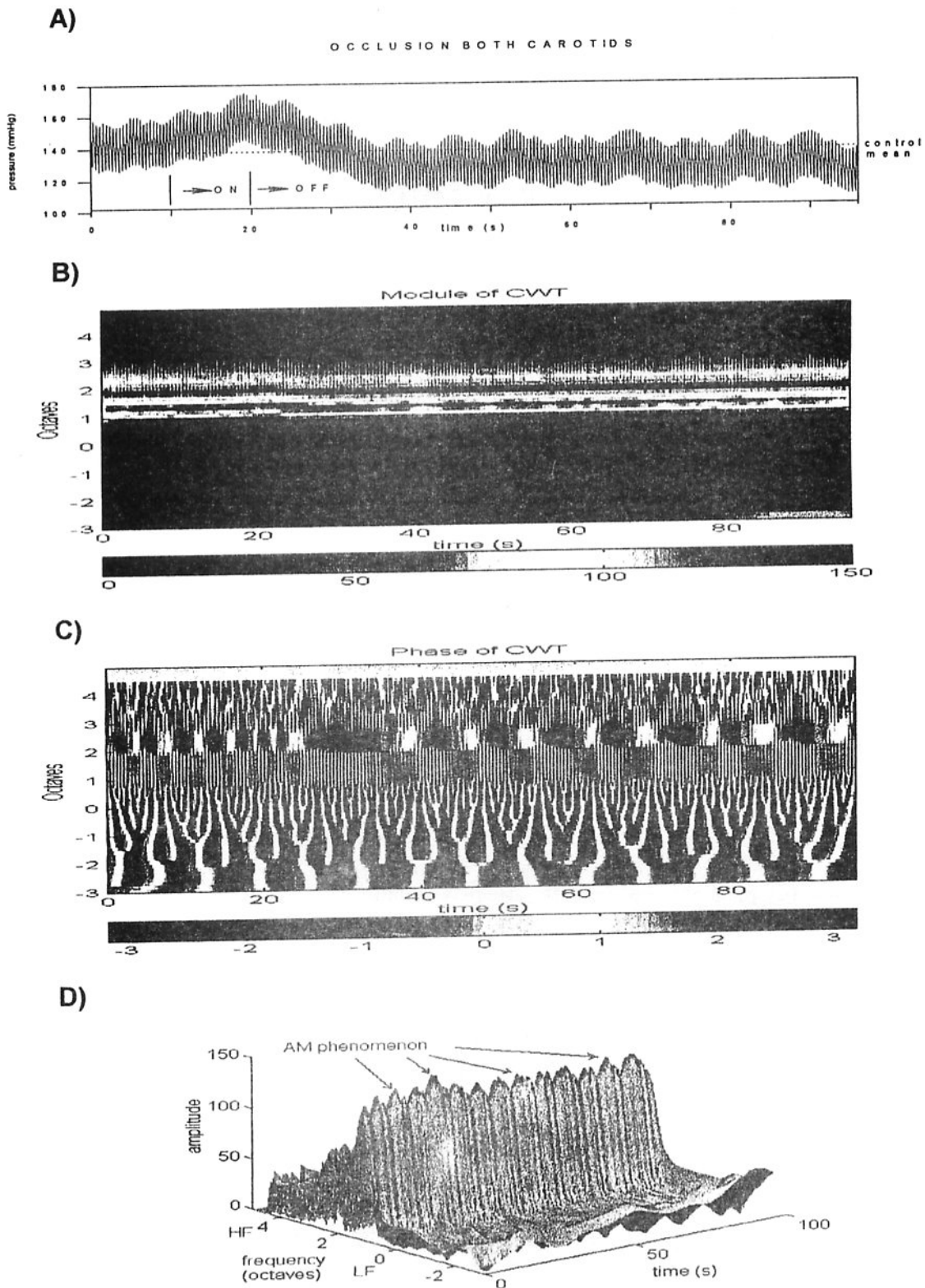


Figura 6. Análisis tiempo-frecuencia de pulsaciones de presión tras la oclusión manual de ambas arterias carótidas por 10 segundos.

## MÉTODOS NO PARAMÉTRICOS UNIVARIADOS Y MULTIVARIADOS

BELMER GARCIA NEGRILLO<sup>1</sup>

---

### Contenido

1. Análisis exploratorio y selección de la función score que da el test localmente más poderoso.
2. Pruebas de hipótesis para dos poblaciones.
3. Pruebas de hipótesis para dos o más poblaciones (one way, two way, factorial, bloques completos, bloques incompletos balanceados o no, split-plot, modelos jerárquicos, medidas repetidas, etc.).
4. Regresión no-paramétrica.
5. Modelo lineal general no-paramétrico.
6. Métodos multivariados completamente aleatorizados.
7. Métodos multivariados con bloques.

Se hará una presentación del software desarrollado.

---

<sup>1</sup>Av. ATILIO MARTINI 469, CIDADE UNIVERSITARIA, CEP. 13084-750 CAMPINAS SP. BRASIL; TEL.: +(55)(19)239 52 85; E-MAIL BELMER@IME.UNICAMP.BR

# IMPLEMENTACIÓN DEL ACP USANDO DISTANCIAS RELACIONALES EN PRESENCIA DE DOS GRUPOS DE VARIABLES

MARCO ALFARO<sup>1</sup> - MARIO VILLALOBOS<sup>1</sup>

---

## Resumen

En este artículo se implementa un algoritmo para realizar un ACP utilizando para ello métricas con efecto relacional. Hacemos además algunas pruebas en el caso de las métricas usuales aplicadas a las tablas de las notas escolares y de los peces de Amiard.

**Palabras Claves:** Análisis en Componentes Principales (ACP), métrica con efecto relacional, grupos de variables, datos estructurados.

## 1 Introducción

El Análisis en Componentes Principales (ACP) es una técnica de descripción estadística que trata de obtener, a partir de una tabla de datos definida por un conjunto de variables cuantitativas, un conjunto reducido de variables sintéticas, llamadas componentes principales, que son combinaciones lineales de las variables observadas y tales que son ortogonales y de varianza máxima.

Sea  $X$  la matriz  $n \times p$  que contiene la información de las  $p$  variables cuantitativas observadas sobre  $n$  individuos. Llamaremos  $E = \mathbb{R}^p$  el espacio de los individuos y  $F = \mathbb{R}^n$  el espacio de las variables, que en adelante supondremos centradas. Los individuos estarán dotados de pesos  $p_i > 0$  tales que  $\sum_{i=1}^n p_i = 1$ . Así, la matriz  $D = \text{diag}(p_i)$  define una métrica en  $F$ . Por la interpretación estadística de esta métrica, es la que usaremos en el espacio  $F$  [2, 8].

Si  $V = X^tDX$  es la matriz de varianzas-covarianzas, se sabe que la solución del ACP se obtiene al diagonalizar  $VM$ , donde  $M$  es la métrica (matriz simétrica definida positiva) que sirve para definir las distancias entre individuos.

Uno de los problemas abiertos en ACP es la escogencia de la métrica [1]. En efecto, dependiendo de la escogencia de ésta, pueden obtenerse resultados diferentes.

---

<sup>1</sup>ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA



Cuando se está en presencia de varios grupos de variables, se han propuesto las distancias relacionales [6], que tratan de tomar en cuenta la estructura definida por cada uno de los grupos.

Si  $M$  es la métrica del espacio de individuos, deseamos que ésta se vea complementada con la información de la estructura de las asociaciones observadas en el espacio de variables. Para simplificar la notación en este artículo, planteamos el caso de dos grupos de variables; sin embargo, todos los conceptos expuestos se pueden generalizar fácilmente para un número arbitrario de grupos de variables.

Proponemos, siguiendo la notación de A. Ibrahim y Y. Schektman [5], una implementación para realizar un ACP, que permite tomar en cuenta las estructuras de asociaciones entre los grupos de variables. Proponemos además un algoritmo para resolver el problema planteado.

## 2 Las distancias relacionales

Supóngase que se tienen dos grupos de variables:  $x^1(1), \dots, x^{p_1}(1)$  y  $x^1(2), \dots, x^{p_2}(2)$ , con  $p = p_1 + p_2$ . Entonces el  $i$ -ésimo individuo ( $i$ -ésima fila de  $X$ ) es  $x_i = x_i(1) + x_i(2)$ , donde  $x_i(1) = (x_i^1(1), \dots, x_i^{p_1}(1), 0, \dots, 0) \in \mathbb{R}^{p_1+p_2}$  y  $x_i(2) = (0, \dots, 0, x_i^1(2), \dots, x_i^{p_2}(2)) \in \mathbb{R}^{p_1+p_2}$ . De esta forma, el espacio de individuos se descompone como  $E = E_1 \oplus E_2$ , donde  $E_1 = \mathbb{R}^{p_1}$  y  $E_2 = \mathbb{R}^{p_2}$ . Por dualidad, el espacio de variables  $F$  se descompone como  $F = F_1 + F_2$ , donde  $F_h$  es el espacio generado por  $x^1(h), \dots, x^{p_h}(h)$ , con  $h = 1, 2$ .

Estas descomposiciones inducen la siguiente partición de la matriz  $M$  de la métrica y de la matriz de varianzas-covarianzas  $V$ :

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

donde  $M_{ii}$  es una métrica en  $E_i$ , con  $i = 1, 2$ , llamadas métricas intra, y

$$V_{11} = X_1^t D X_1, \quad V_{12} = X_1^t D X_2,$$

$$V_{22} = X_2^t D X_2, \quad V_{21} = V_{12}^t.$$

Aquí,  $X_1$  es la matriz  $n \times p_1$  cuyas filas son las  $x_i(1)$  y  $X_2$  es la matriz cuyas filas son las  $x_i(2)$ . Denotamos  $\mathcal{N}_h = \{x_1(h), \dots, x_n(h)\} \subseteq E_h$ , la nube de puntos descrita por las variables del grupo  $h$ .

Para simplificar, denotaremos además  $M_i = M_{ii}$  y  $V_i = V_{ii}$ . Sean  $\{c_1(h), \dots, c_{p_h}(h)\}$  y  $\{C^1(h), \dots, C^{p_h}(h)\}$  los vectores principales y las componentes principales del ACP de  $\mathcal{N}_h = (X_h, M_{hh}, D)$ .

**Definición:** Decimos que la métrica  $M$  tiene efecto relacional si

$$\cos_M(c_j(1), c_k(2)) = \rho(C^j(1), C^k(2)), \text{ para todo } j \in \{1, \dots, p_1\} \text{ y } k \in \{1, \dots, p_2\}.$$

donde  $\rho(\cdot, \cdot)$  es la correlación lineal entre las componentes  $C^j(1)$  y  $C^j(2)$ . Una distancia relacional tiene una expresión algebraica simple [6, 3], dada a continuación.

**Propiedad 1:**  $M$  tiene efecto relacional si y sólo si

$$M_{12} = M_1(V_1M_1)^{-1/2}V_{12}M_2(V_2M_2)^{-1/2}.$$

**Demostración:** Sean  $\{c_j(h)\}$  los ejes principales del ACP de  $(X_h, M_h, D)$ ,  $\{C^j(h) = X_hM_hc_j(h)\}$  las componentes principales, y  $\lambda_j = \lambda_j(h)$  los valores propios asociados. Entonces:

$$(V_2M_2)^{-1/2}c_j(2) = \frac{1}{\sqrt{\lambda_j}}c_j(2), \text{ y como } \sqrt{\lambda_j} = \|C^j(2)\|_D, \text{ se tiene que:}$$

$$(V_2M_2)^{-1/2}c_j(2) = \frac{c_j(2)}{\|C^j(2)\|_D}.$$

Luego,

$$X_2M_2(V_2M_2)^{-1/2}c_j(2) = \frac{X_2M_2c_j(2)}{\|C^j(2)\|_D} = \frac{C^j(2)}{\|C^j(2)\|_D}.$$

De la definición de métrica con efecto relacional, podemos escribir:

$$\begin{aligned} \langle c_j(1), M_{12}c_k(2) \rangle &= D \left[ \frac{C^j(1)}{\|C^j(1)\|_D}, \frac{C^k(2)}{\|C^k(2)\|_D} \right] \\ &= D[X_1M_1(V_1M_1)^{-1/2}c_j(1), X_2M_2(V_2M_2)^{-1/2}c_k(2)] \\ &= \langle M_1(V_1M_1)^{-1/2}c_j(1), V_{12}M_2(V_2M_2)^{-1/2}c_k(2) \rangle. \end{aligned}$$

donde  $\langle \cdot, \cdot \rangle$  es producto interno canónico y  $D[\cdot, \cdot]$  es el  $D$ -producto interno. Como además  $(V_iM_i)$  son  $M_i$ -simétricas [2], entonces, se obtiene finalmente que:

$$\forall j, k \quad \langle c_j(1), M_{12}c_k(2) \rangle = \langle c_j(1), M_1(V_1M_1)^{-1/2}V_{12}M_2(V_2M_2)^{-1/2}c_k(2) \rangle \quad \blacksquare$$

### 3 Algunas métricas con efecto relacional de uso frecuente

Consideramos aquí algunas métricas de uso frecuente:

$$M_i = I, \quad M_i = V, \quad M_i = V^{-1}, \quad M_i = D_{1/\sigma^2}$$

y estudiamos cómo se puede realizar el cálculo de las métricas con efecto relacional en tales casos.

Si denotamos por

$$S_i = M_i(V_i M_i)^{-1/2},$$

entonces podemos escribir, cuando  $M$  tiene efecto relacional:

$$M_{ij} = S_i V_{ij} S_j, \quad (1)$$

y tenemos así el siguiente resultado:

#### Propiedad 2:

- a) Si  $M_i = V_i$ , entonces  $S_i = I$ .
- b) Si  $M_i = V_i^{-1}$ , entonces  $S_i = V_i^{-1}$ .
- c) Si  $M_i = I$ , entonces  $S_i = V_i^{-1/2}$ .
- d) Si  $M_i = D_{1/\sigma^2}$ , entonces  $S_i = D_{1/\sigma^2}(V D_{1/\sigma^2})^{-1/2}$ .

En la parte **d)** de la propiedad anterior, se necesita hacer el cálculo de  $S_i = D_{1/\sigma^2}(V D_{1/\sigma^2})^{-1/2}$ , para hacer esto se utilizará la fórmula  $(V D_{1/\sigma^2})^{-1/2} = P^{-1} \text{diag}(1/\sqrt{\lambda_j}) P$  donde  $\lambda_j$  y  $P$  son los valores propios y la matriz de vectores propios de  $D_{1/\sigma^2}$  ortonormados, respectivamente.

Por la propiedad 3, el cálculo de valores y vectores propios de  $(V D_{1/\sigma^2})^{1/2}$  se reduce al cálculo de valores y vectores propios de la matriz de correlaciones, la cual es simétrica.

**Propiedad 3:** Los valores propios de la matriz  $V D_{1/\sigma^2}$  son los mismos que los de la matriz  $R = D_{1/\sigma} V D_{1/\sigma}$ , y los vectores propios de  $V D_{1/\sigma^2}$  son  $u_j = D_{\sigma} v_j$ , donde  $v_j$  es vector propio de  $R$ .

**Corolario:** Si  $C$  es la matriz de vectores propios de  $R$ , entonces  $P = D_{\sigma} C$  es la matriz de vectores propios de  $V D_{1/\sigma^2}$ .

## 4 Algoritmo

A continuación, presentamos el algoritmo pseudocódigo que fue aplicado a las tablas de datos. Los resultados aparecen en la sección siguiente.

1. Lectura de los datos
2. Centraje de las variables
3. Cálculo de la matriz de varianzas-covarianzas  $V$ .
4. Separación de la matriz  $V$  en las submatrices  $V_1, V_2, V_{12}$
5. Escogencia de la métrica  $M_1$  para el primer bloque de variables.
6. Cálculo de  $S_1$ , de acuerdo con la propiedad 2.
7. Escogencia de la métrica  $M_2$  para el segundo bloque de variables.
8. Cálculo de  $S_2$ , de acuerdo con la propiedad 2.
9. Cálculo de  $M_{12}$  según la ecuación (1).
10. Construcción de la métrica  $M$ , según

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

11. Realización del ACP con la métrica  $M$ 
  - 11-1. Diagonalización de  $VM$ .
  - 11-2. Cálculo de componentes principales.
12. Impresión de resultados.

## 5 Características de la implementación

El programa computacional está escrito en lenguaje *Pascal 7*, en el cual se utiliza el método de Jacobi para el cálculo de los valores propios y el método de Gauss para el cálculo de las matrices inversas.

El ACP se realiza utilizando los mismos algoritmos empleados por Bolaños en [1], genera una salida que puede ser un archivo de  $\text{\LaTeX}$ , ya sea por medio de gráficos o bien en modo texto, a elección del usuario. Este programa se encuentra a disposición de las personas interesadas.

## 6 Resultados numéricos

Aplicamos el algoritmo a las tablas de notas escolares y de los peces de Amiard. En el caso de la tabla de notas escolares se tienen siete variables observadas sobre diez estudiantes, las primeras cinco variables corresponden a las calificaciones, en una escala de 0 a 10, obtenidas por los estudiantes en Matemática, Ciencias, Español, Historia y Educación Física, y las otras dos variables corresponden a peso y estatura, estas últimas medidas en libras y centímetros, respectivamente.

La tabla de datos que se tiene es:

	$x^1$	$x^2$	$x^3$	$x^4$	$x^5$	$x^6$	$x^7$
	Matemática	Ciencias	Español	Historia	Ed. Física	Peso	Estatura
$x_1$	7.00	6.50	9.20	8.60	8.00	126.00	162.00
$x_2$	7.50	9.40	7.30	7.00	7.00	140.00	168.00
$x_3$	7.60	9.20	8.00	8.00	7.50	130.00	169.00
$x_4$	5.00	6.50	6.50	7.00	9.00	150.00	172.00
$x_5$	6.00	6.00	7.80	8.90	7.30	142.00	165.00
$x_6$	7.80	9.60	7.70	8.00	6.50	128.00	165.00
$x_7$	6.30	6.40	8.20	9.00	7.20	144.00	170.00
$x_8$	7.90	9.70	7.50	8.00	6.00	134.00	165.00
$x_9$	6.00	6.00	6.50	5.50	8.70	135.00	170.00
$x_{10}$	6.80	7.20	8.70	9.00	7.00	128.00	166.00

Como vemos, tenemos dos grupos de variables, bien diferentes por su naturaleza, que deberían tratarse por separado, por lo que se tiene un caso en el que puede aplicarse un ACP con una métrica con efecto relacional.

Con el interés de ilustrar las salidas del programa, se incluyen a continuación algunos de los resultados, con las escogencias de las métricas intra que se indican en cada caso.

De los resultados obtenidos, para los datos de las notas escolares, podemos observar que cuando  $M_2 = I$  (gráfico 1 y 3), el primer valor propio tiene un porcentaje de inercia superior al 99%, esto puede deberse a que en este caso se consideran las tallas y los pesos brutos, o sea valores más grandes que los de las notas (que están entre 0 y 10).

En el caso de la tabla de los peces de Amiard, se tienen 16 variables de las cuales las primeras nueve son medidas físicas, mientras que las restantes 7 son medidas de la cantidad de radioactividad, por lo que también se justifica el uso de las métricas con efecto relacional. Los resultados no se incluyen en este artículo.

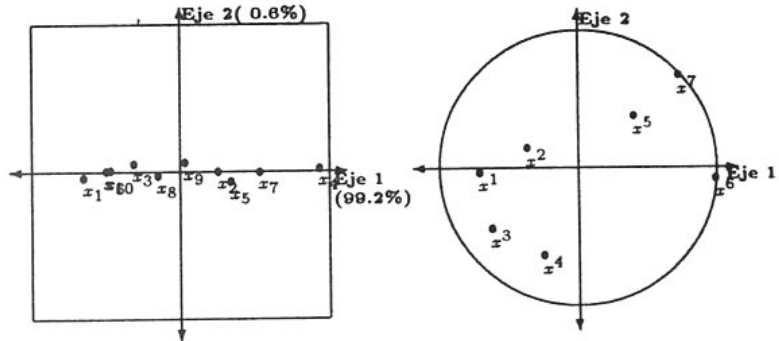
## 7 Perspectivas

A pesar de lo antes mencionado, por tratarse de un caso particular, no se puede concluir que exista un patrón. De hecho, en los cálculos realizados con la tabla de los peces de Amiard se observan comportamientos diferentes.

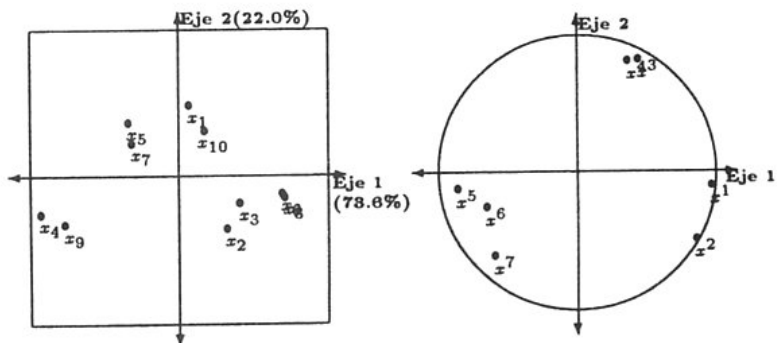
Los casos que hemos estudiado en el presente artículo se han realizado únicamente con datos en los que se observan dos grupos de variables, pero el método se puede aplicar, de

manera semejante, en el caso en que se tengan más de dos grupos de variables. Este será tema de futuras investigaciones, así como una comparación minuciosa entre los resultados obtenidos con estas métricas y las usuales.

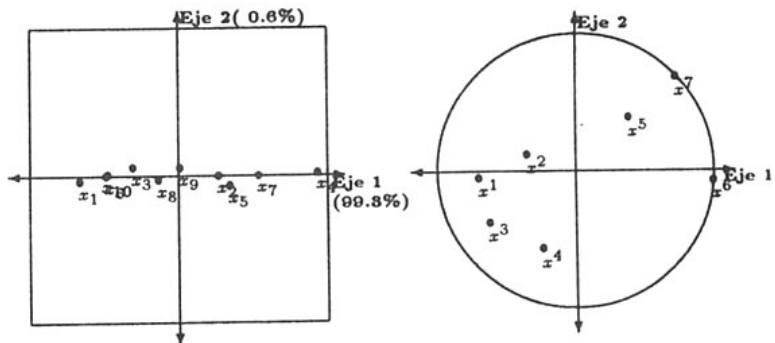
$$1) \quad M = \begin{bmatrix} I & M_{12} \\ M_{21} & I \end{bmatrix}$$



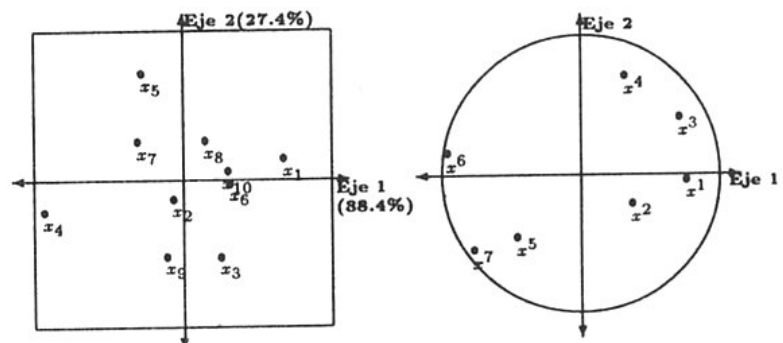
$$2) \quad M = \begin{bmatrix} I & M_{12} \\ M_{21} & V \end{bmatrix}$$



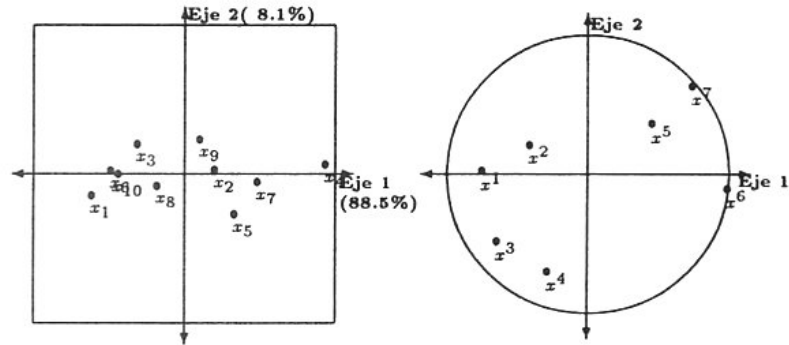
$$3) \quad M = \begin{bmatrix} D_{1/\sigma^2} & M_{12} \\ M_{21} & I \end{bmatrix}$$



$$4) \quad M = \begin{bmatrix} D_{1/\sigma^2} & M_{12} \\ M_{21} & D_{1/\sigma^2} \end{bmatrix}$$



$$5) \quad M = \begin{bmatrix} V^{-1} & M_{12} \\ M_{21} & V^{-1} \end{bmatrix}$$



## Bibliografía

- [1] Bolaños, F. (1996) *Contribución al estudio de las métricas en Análisis de Datos*. Tesis de Licenciatura. Universidad de Costa Rica, San Pedro.
- [2] Cailliez, F.; Pagès J.P. (1976) *Introduction à l'Analyse des Données*. Smash, Paris.
- [3] Croquette, A. (1980) *Quelques résultats synthétiques en analyse des données multidimensionnelles: optimalités et métriques à effets relationnels*. Thèse de 3<sup>o</sup> cycle, Université Paul Sabatier, Toulouse.
- [4] Diday E.; Lemaire J.; Pouget J.; Testu, F. (1982) *Eléments d'Analyse des Données*. Dunod, Paris.
- [5] Ibrahim, A.; Schektman Y. (1985) *Analyses en partitions principales, algorithmes et exemples*. Journées de la Société Francophone de Classification, La Grande-Motte, CNET: 61–89.
- [6] Schektman, Y. (1978) *Contribution à la mesure en facteurs dans les sciences expérimentales et à la mise en œuvre des calculs statistiques*. Thèse d'Etat, Université Paul Sabatier, Toulouse.
- [7] Trejos, J. (1994) *Contribution à l'Acquisition Automatique de Connaissances à Partir de Données Qualitatives*. Thèse de Doctorat, Université Paul Sabatier, Toulouse.
- [8] Trejos, J. (1996) *Introducción al Análisis de Datos*. Notas de curso, Universidad de Costa Rica-Universidad Nacional Autónoma de Nicaragua, León.

## POLINOMIOS Y SPLINE INTERPOLANTES EN $\mathbb{R}^n$

VERNOR ARGUEDAS<sup>1</sup> - ROBERTO MATA<sup>2</sup>

---

En este trabajo se analiza la forma de los polinomios y spline interpolantes en  $\mathbb{R}^n$  y se hace el estudio de algunas de sus propiedades.

Enunciamos y demostramos el siguiente teorema, el cual posiblemente sea muy conocido; sin embargo, no lo hemos visto escrito en ningún lado.

Sea  $f : [a, b] \rightarrow \mathbb{R}^n$ , con  $f = (f_1, \dots, f_n)^t$ , donde  $f \in C^{(n+1)}[a, b]$ . Es decir, todas las derivadas existen y son continuas, hasta la de orden  $(n + 1)$ .

Sea  $\{x_0 = a < x_1 < \dots < x_p = b\}$  una partición dada de  $[a, b]$  y sea  $L$  la función polinomial de interpolación de  $f$  con respecto a  $\{x_0, x_1, \dots, x_p\}$ . Entonces:

$$\|f(x) - L(x)\|_{\infty} \leq \frac{1}{(p+1)!} |(x-x_0)(x-x_1)\cdots(x-x_p)| \cdot \sup \|f^{(p+1)}(y)\|_{\infty}, \text{ con } y \in [a, b]$$

También describimos la forma de los polinomios interpolantes dobles en  $\mathbb{R}^n$ , mediante el siguiente teorema:

Sea  $f : [a, b] \rightarrow \mathbb{R}^n$ , con  $f = (f_1, \dots, f_n)^t$  y  $f \in C^{(2p+1)}[a, b]$ . Sea  $\{x_0 = a < x_1 < \dots < x_p = b\}$  una partición dada de  $[a, b]$ . Entonces, existe una única función polinomial  $H$ , de grado a lo sumo  $(2p + 1)$ , que coincide con  $f$  y  $f'$  en  $x_i$ , para  $i = 0, 1, \dots, p$ . Además, si  $f \in \text{cal}C^{(2p+2)}[a, b]$ , entonces:

$$\|f(x) - H(x)\|_{\infty} \leq \frac{1}{(2p+2p)!} (x-x_0)^2 \cdots (x-x_p)^2 \cdot \sup \|f^{(2p+2)}(y)\|_{\infty}, \text{ con } y \in [a, b].$$

Para cada  $i$ , la existencia y unicidad de la función polinomial  $H$ , se obtiene como consecuencia de la existencia y unicidad del polinomio de Hermite doble en cada una de las funciones componentes.

Finalmente, definimos los splines en  $\mathbb{R}^n$  y obtenemos el siguiente resultado:

Si  $f : [a, b] \rightarrow \mathbb{R}^n$ , con  $f = (f_1, \dots, f_n)^t$ , es dada, y cada  $f_i$  se aproxima en algún sentido con el spline  $s_i$ , entonces  $f$  se aproxima por el spline:  $(s_1, \dots, s_n)^t$ .

---

<sup>1</sup>ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

<sup>2</sup>SEDE DE GUANACASTE, UNIVERSIDAD DE COSTA RICA, LIBERIA, COSTA RICA



# APLICACIÓN DE LOS CONJUNTOS APROXIMADOS A LA SIMPLIFICACIÓN DE REGLAS EN SISTEMAS EXPERTOS Y CLASIFICACIÓN AUTOMÁTICA

ALCIDES ASTORGA<sup>1</sup>

---

## Resumen

En la construcción de sistemas automáticos un problema que se presenta es determinar, para problemas específicos, el menor conjunto de condiciones y reglas para la toma de decisiones. En este trabajo se presenta una metodología basada en los conjuntos aproximados, cuyo fundamento principal es el concepto de relación de equivalencia. Esta metodología se basa en la eliminación de condiciones superfluas de forma tal que el conjunto de reglas se mantenga consistente y la certeza de la decisión que se deba tomar no se vea afectada. También se analizan los resultados obtenidos, para un caso específico de datos.

**Palabras claves:** conjuntos aproximados, tablas de decisión, reglas de decisión, reducto, núcleo, algoritmos consistentes, atributos indispensables.

## 1 Introducción

Una tabla de decisión es un tipo especial de sistema de información, en la cual se especifica qué decisiones se deben tomar cuando algunas condiciones son satisfechas. Las tablas de decisión son formalismos de representación de conocimiento utilizados dentro del proceso de desarrollo de la base de conocimiento de los sistemas expertos y de la clasificación de datos.

La simplificación de tablas de decisión, tanto en número de reglas como en condiciones a verificar, es importante en muchas aplicaciones. En una tabla de decisión reducida, se pueden tomar las mismas decisiones que si se utilizaran todos los atributos, pero basándose en un número menor de condiciones. Este tipo de simplificación elimina el trabajo de ir chequeando condiciones innecesarias, o en algunas aplicaciones, realizar estudios costosos para arribar a conclusiones que eventualmente podrían ser obtenidas por medios más simples.

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE COSTA RICA, CARTAGO, COSTA RICA

El presente trabajo consta de dos partes. En la primera parte, se presenta aquellos fundamentos de la teoría de conjuntos aproximados -Rough sets - que son la base de los algoritmos que se implementarán. En la segunda parte, se comentan los resultados obtenidos en un caso específico. Para obtener esos resultados se programaron los correspondientes algoritmos en el lenguaje **TopSpeed<sup>©</sup>Modula**.

La teoría de los conjuntos aproximados fue establecida por el investigador polaco Zdzislaw Pawlak, a inicios de la década de los ochenta y se fundamenta en los conceptos de relaciones de equivalencia y particiones, los cuales, a criterio de Pawlak se pueden utilizar para describir el proceso de adquisición de conocimiento que siguen los humanos.

## 2 Sistemas de información y tablas de decisión

**Definición 2.1** Un *sistema de información* es un cuádruple  $S, S = (O, A, V, f)$  donde:

$O$  : es un conjunto no vacío de objetos.

$A$  : es un conjunto no vacío de atributos para los elementos de  $O$ .

$V$  : es un conjunto no vacío, el cual se define como  $V = \cup_{a \in A} V_a$  donde los elementos de  $V_a$  son los valores que puede tomar el atributo  $a$ .

$f$  : es una función tal que  $f : O \times A \rightarrow V$  donde  $f(o, a) \in V_a$  para cada  $o \in O$  y  $a \in A$ ,  $f$  recibe el nombre de *función de información* para  $o$ .

**Definición 2.2** Sean  $X \subset O$  y  $R$  una relación de equivalencia definida en  $O$ . Se dice que  $X$  es *R-definible* en  $O$  si  $X$  se puede expresar como la unión de algunas de las  $R$ -clases de equivalencia, en caso contrario, se dice que  $X$  es *R-inexacto* o *R-aproximado*.

**Definición 2.3** Sean  $S = (O, A, V, f)$  y  $P, Q \subset A$ . Los elementos de  $P$  serán llamados *atributos de condición* y los de  $Q$  *atributos de decisión*. Los sistemas donde se distinguen los atributos de condición y de decisión los llamaremos *tablas de decisión* y se denotarán por  $T = (O, A, P, Q)$ .

Para cada  $x \in O$  fijo, asociamos la función  $d_x(a) = f(x, a)$  para todo  $a \in P \cup Q$ , con  $P, Q \subset A$ . La función  $d_x$  será llamada una *regla de decisión* en  $T$ ,  $x$  será referida como una etiqueta para la regla de decisión.

**Notación:** Si  $d_x$  es una regla de decisión y  $N \subset A$ , entonces la restricción de  $d_x$  a  $N$  será denotada por  $d_x/N$ .

**Definición 2.4** Sea  $T = (O, A, P, Q)$ . La regla  $d_x$  es *consistente* en  $T$  si para todo  $y \neq x$ ,  $d_x/P = d_y/P$  implica que  $d_x/Q = d_y/Q$ , en cualquier otro caso la regla es *inconsistente*. Una tabla de decisión es *consistente* si todas las reglas de decisión son consistentes, de otro modo se dice que es *inconsistente*.

## 2.1 Algoritmos de decisión

Con el fin de simplificar la notación, las reglas de decisión las representaremos en la forma general  $\varphi \rightarrow_{PQ} \psi$ , donde se supone que  $\varphi$  y  $\psi$  son predicados con dominio  $P$  y  $Q$ , respectivamente, y nos referiremos a ellas como *reglas PQ de decisión*.

Cualquier conjunto finito de reglas, es referido como un *algoritmo de decisión* y será denotado por  $(P, Q)$ .

El algoritmo  $(P, Q)$  es *consistente* en  $T$  si y solo si todas sus reglas son consistentes en  $T$ , en caso contrario, el algoritmo se dice que es *inconsistente* en  $T$ . Este resultado se puede enunciar así:

**Teorema 1** Una regla  $\varphi \rightarrow_{PQ} \psi$  en un algoritmo  $(P, Q)$  es consistente, si y solo si para cualquier otra regla de decisión  $\varphi' \rightarrow_{PQ} \psi'$  en  $(P, Q)$ , se tiene que:

$$\varphi = \varphi' \rightarrow_{PQ} \psi = \psi'$$

En otras palabras, todas las decisiones en un algoritmo consistente quedan determinadas si se conocen los valores de los atributos de condición.

## 2.2 Simplificación de algoritmos consistentes

Sea  $(P, Q)$  un algoritmo consistente y sea  $a \in P$ .

1. Diremos que el atributo  $a$  *no es indispensable* en el algoritmo  $(P, Q)$  si y solo si el algoritmo  $(P - \{a\}, Q)$  es consistente, de otra forma se dice que el atributo  $a$  es indispensable en el algoritmo  $(P, Q)$ .
2. Si todos los atributos  $a$  en  $P$  son indispensables en el algoritmo  $(P, Q)$ , entonces el algoritmo  $(P, Q)$  será llamado *independiente*.
3. El conjunto de todos los atributos indispensables en un algoritmo  $(P, Q)$  será llamado el *núcleo* del algoritmo  $(P, Q)$  y será denotado  $\text{NUCLEO}(P, Q)$ .
4. El subconjunto de atributos  $R \subset P$  será llamado un *reducto* de  $P$  en el algoritmo  $(P, Q)$  si el algoritmo  $(R, Q)$  es independiente y consistente.
5. Si  $R$  es un reducto de  $P$  en el algoritmo  $(P, Q)$ , entonces el algoritmo  $(R, Q)$  se dice ser un *reducto del algoritmo*  $(P, Q)$ .

## 2.3 Simplificación de reglas

Sea  $\varphi \rightarrow_{PQ} \psi$  una  $PQ$  regla, y sea  $I = P - \{a\}$  con  $a$  un elemento arbitrario de  $A$ .

1. Diremos que el atributo  $a$  es *no indispensable en la regla*  $\varphi \rightarrow_{PQ} \psi$  si y solo si  $\varphi \rightarrow_{PQ} \psi$  es consistente en  $(P, Q)$  implica que  $\varphi \rightarrow_{IQ} \psi$  es consistente en  $(I, Q)$ . De cualquier otra forma se dice que  $a$  es indispensable en  $\varphi \rightarrow_{PQ} \psi$ .
2. Si todos los atributos  $a$  en  $P$  son indispensables en  $\varphi \rightarrow_{PQ} \psi$  entonces  $\varphi \rightarrow_{PQ} \psi$  será llamada *PQ independiente*.

3. El conjunto  $R$ ,  $R \subset P$ , será llamado un reducto de una regla  $\varphi \rightarrow_{PQ} \psi$  si  $\varphi \rightarrow_{PQ} \psi$  es  $RQ$  independiente y  $\varphi \rightarrow_{PQ} \psi$  consistente en  $(P, Q)$  implica que  $\varphi \rightarrow_{IQ} \psi$  es consistente en  $RQ$ .
4. Si  $R$  es un reducto de la  $PQ$  regla  $\varphi \rightarrow_{PQ} \psi$  entonces  $\varphi \rightarrow_{RQ} \psi$  se dice estar *reducida*.
5. El conjunto de todos los atributos indispensables en  $\varphi \rightarrow_{PQ} \psi$  será llamado el *núcleo* de  $\varphi \rightarrow \psi$  y se denotará por  $\text{NUCLEO}(\varphi \rightarrow_{PQ} \psi)$ .

**Teorema 2** Sea  $\varphi \rightarrow_{PQ} \psi$  una  $PQ$  regla, entonces se cumple que

$$\text{NUCLEO}(\varphi \rightarrow_{PQ} \psi) = \text{RED}(\varphi \rightarrow_{PQ} \psi)$$

donde  $\text{RED}(\varphi \rightarrow_{PQ} \psi)$  es el conjunto de todos los reductos de  $\varphi \rightarrow_{PQ} \psi$ .

### 3 Un caso de estudio

Como se mencionó anteriormente, en esta sección aparte de obtener un conjunto mínimo de atributos, así como de verificar con cuánta exactitud la información disponible puede ser usada para obtener la clasificación aportada por un experto, se buscará establecer un conjunto mínimo de reglas las cuales se pueden usar para ubicar objetos, en las diferentes clases propuestas por un experto.

El conjunto de datos que se estudiará en esta sección consiste de treinta animales los cuales han sido clasificados en cuatro clases por un grupo de zoólogos. La decisión para ubicar a los animales en una u otra clase se ha tomado con base en catorce atributos que han definido los expertos y los cuales han sido medidos u observados en los felinos. Para cada atributo se ha definido también un rango discreto de posibles valores, tal y como se indica a continuación:

- a *Aspecto del pelo*: 4 posibles valores
- b *Tipo de pelo*: 2 posibles valores
- c *Posee garras retráctiles*: 2 posibles valores
- d *Comportamiento de rapiña*: 3 posibles valores
- e *Forma de las orejas*: 2 posibles valores
- f *Presencia o no del hialoide*: 2 posibles valores
- g *Tamaño de la cruz*: 3 posibles valores
- h *Peso del animal*: 3 posibles valores
- i *Longitud del cuerpo*: 3 posibles valores
- j *Longitud de la cola con respecto al cuerpo*: 3 posibles valores
- k *Desarrollo de los colmillos*: 2 posibles valores
- l *Tamaño de la presa*: 3 posibles valores
- m *Trepador de árboles*: 2 posibles valores
- n *Caza corriendo o por sorpresa*: 2 posibles valores.

Si bien es cierto que el conjunto original de datos correspondía a la información de 30 animales, para efectos de este estudio lo hemos reducido a 25, con el fin de utilizar los

Tabla 1

Felino	Atributos														Clase
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	
1	1	1	2	1	1	2	3	3	3	2	2	1	1	2	1
2	3	1	2	3	1	2	3	3	3	2	2	1	1	1	1
3	2	1	2	3	1	2	3	3	2	2	2	2	2	1	1
4	2	2	2	1	1	2	2	2	2	3	2	2	2	1	1
5	2	1	1	1	1	1	3	2	2	3	1	2	1	2	4
6	1	1	2	2	1	1	2	3	2	3	2	2	2	1	3
7	4	1	2	3	1	2	2	2	2	3	2	3	2	1	2
8	2	1	2	1	2	1	2	2	2	1	1	3	2	2	3
9	2	1	2	2	1	1	2	2	2	2	1	3	2	1	3
10	2	2	2	2	2	1	2	2	2	1	2	2	2	1	3
11	1	1	2	2	2	1	2	2	1	1	1	3	2	2	3
12	2	1	2	2	1	1	1	1	2	2	1	3	1	1	3
13	1	1	2	2	1	1	1	2	2	3	1	3	2	1	3
14	1	2	2	3	2	1	1	2	1	2	1	3	2	1	3
15	2	1	2	3	1	1	1	1	1	2	1	3	2	1	3
16	1	2	2	2	1	1	1	1	1	2	1	3	1	1	3
17	1	1	2	2	2	1	1	1	1	1	1	3	2	1	3
18	2	1	2	3	1	1	1	1	1	2	1	3	2	1	3
19	2	1	2	2	1	1	1	1	1	2	1	3	2	1	3
20	1	2	2	3	1	1	1	1	1	1	1	3	2	1	3
21	2	1	2	2	1	1	1	1	1	1	1	3	2	2	3
22	1	2	2	3	1	1	1	1	1	1	1	3	2	1	3
23	2	1	2	3	1	1	1	1	1	2	1	3	2	1	3
24	1	1	2	3	1	1	1	1	1	2	1	3	2	1	3
25	2	2	2	3	1	1	1	1	2	2	1	2	2	1	3

cinco animales restantes para realizar un test rápido del conjunto de reglas de clasificación que se obtendrán al final. El conjunto de datos con los que se trabajará se indican en la Tabla 1.

### Paso 1: Eliminación de atributos de condición redundantes<sup>2</sup>

Con el fin de trabajar con un conjunto mínimo de condiciones y utilizando los resultados expuestos en [3], se elaboraron los algoritmos correspondientes, los cuales, como se mencionó anteriormente, fueron programados en **Modula**, obteniéndose el conjunto de atributos significativos que se indican en las Tablas 2 y 3. En este paso no se utilizan los atributos de decisión, los cuales aparecen en las Tablas 2 y 3 en la columna etiquetada con el nombre de Clase.

Para los siguientes pasos, tanto la Tabla 2, como la Tabla 3, deben ser visualizadas como un conjunto de reglas donde la columna de atributos constituyen las condiciones a verificar (atributos de condición) y las clases serán las decisiones que se deben tomar (atributos de decisión).

El siguiente paso consiste en obtener el menor conjunto de reglas que nos pueda servir como herramienta para ubicar cualquier felino dentro de las cuatro clases aportadas por el experto. Hay que tener presente que las reglas que se generen dependen de la información disponible en el sistema y del conocimiento que nos aportó el experto, es decir, de la

<sup>2</sup>Los fundamentos teóricos que sustentan los algoritmos que se utilizan en este paso son tratados por otros colegas en este Congreso, en su ponencia "Descripción de clases mediante conjuntos aproximados".

Tabla 2

Felino	Atributos						Clase
	a	b	d	g	i	j	
1	1	1	1	3	3	2	1
2	3	1	3	3	3	2	1
3	2	1	3	3	2	2	1
4	2	2	1	2	2	3	1
5	2	1	1	3	2	3	4
6	1	1	2	2	2	3	3
7	4	1	3	2	2	3	2
8	2	1	1	2	2	1	3
9	2	1	2	2	2	2	3
10	2	2	2	2	2	1	3
11	1	1	2	2	1	1	3
12	2	1	2	1	2	2	3

Tabla 3

Felino	Atributos						Clase
	a	b	d	g	i	j	
13	1	1	2	1	2	3	3
14	1	2	3	1	1	2	3
15	2	1	3	1	1	2	3
16	1	2	2	1	1	2	3
17	1	1	2	1	1	1	3
18	2	1	3	1	1	2	3
19	2	1	2	1	1	2	3
20	1	2	3	1	1	1	3
21	2	1	2	1	1	1	3
22	1	2	3	1	1	1	3
23	2	1	3	1	1	2	3
24	1	1	3	1	1	2	3
25	2	2	3	1	2	2	3

clasificación de cada animal en los clases indicadas.

**Paso 2: Consistencia del conjunto de reglas**

El conjunto de 25 reglas con las que se trabaja, es consistente. Para verificar esto, se programó el Teorema 1. Por lo que los atributos de decisión son totalmente dependientes de los atributos de condición, y por lo tanto, los atributos de condición son "suficientes" para tomar la decisión final y poder decir si un felino, incluido en la Tabla 1, pertenece o no, a las clases aportadas por el experto.

**Paso 3: Determinación de los atributos indispensables**

El objetivo en este paso es obtener la menor cantidad de atributos de condición que sean necesarios para poder ubicar los animales en las diferentes clases, para esto se procede a eliminar un solo atributo de las Tablas 2 y 3, usando otra vez el Teorema 1 se verifica si el conjunto de reglas resultante es consistente o no. Dentro de los cálculos hechos, se obtuvieron dos reductos, los cuales se muestran en la Tabla 4.

Debido a que el segundo conjunto de atributos contiene un menor número de elementos que el primero, el siguiente análisis se realizará considerando únicamente los atributos *d*, *g* y *j*, reduciéndose la tabla de información, a como se muestra en las Tablas 5 y 6.

Como se puede notar en las Tablas 5 y 6, las siguientes reglas son las mismas: la 2 y 3, la 10 y 11, la 12, 16 y 19, la 14, 15, 18, 23, 24 y 25, la 17 y 21, la 20 y 22.

Eliminando las reglas repetidas en las tablas anteriores, se obtienen las 14 reglas que se muestran en la Tabla 7.

**Paso 4: Obtención del núcleo de las reglas**

En esta parte se buscará el núcleo para cada una de las reglas de la Tabla 7, es decir, para cada regla se determinará el menor conjunto de atributos de condición que sean indispensables. En este caso, la eliminación de las condiciones se hace regla por regla, a diferencia del paso anterior, donde el atributo debía ser eliminado simultáneamente en todas las reglas, lo que se traducía en la eliminación de columnas.

El núcleo para el conjunto de reglas se muestra en la Tabla 8, los atributos que no son indispensables han sido sustituidos por el signo "-".

Tabla 1

Felino	Atributos														Clase
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	
1	1	1	2	1	1	2	3	3	3	2	2	1	1	2	1
2	3	1	2	3	1	2	3	3	3	2	2	1	1	1	1
3	2	1	2	3	1	2	3	3	2	2	2	2	2	2	1
4	2	2	2	1	1	2	2	2	2	3	2	2	2	1	1
5	2	1	1	1	1	1	3	2	2	3	1	2	1	2	4
6	1	1	2	2	1	1	2	3	2	3	2	2	2	1	3
7	4	1	2	3	1	2	2	2	2	3	2	3	2	1	2
8	2	1	2	1	2	1	2	2	2	1	1	3	2	2	3
9	2	1	2	2	1	1	2	2	2	2	1	3	2	1	3
10	2	2	2	2	2	1	2	2	2	1	2	2	2	1	3
11	1	1	2	2	2	1	2	2	1	1	1	3	2	2	3
12	2	1	2	2	1	1	1	1	2	2	1	3	1	1	3
13	1	1	2	2	1	1	1	2	2	3	1	3	2	1	3
14	1	2	2	3	2	1	1	2	1	2	1	3	2	1	3
15	2	1	2	3	1	1	1	1	1	2	1	3	2	1	3
16	1	2	2	2	1	1	1	1	1	2	1	3	1	1	3
17	1	1	2	2	2	1	1	1	1	1	1	3	2	1	3
18	2	1	2	3	1	1	1	1	1	2	1	3	2	1	3
19	2	1	2	2	1	1	1	1	1	2	1	3	2	1	3
20	1	2	2	3	1	1	1	1	1	1	1	3	2	1	3
21	2	1	2	2	1	1	1	1	1	1	1	3	2	2	3
22	1	2	2	3	1	1	1	1	1	1	1	3	2	1	3
23	2	1	2	3	1	1	1	1	1	2	1	3	2	1	3
24	1	1	2	3	1	1	1	1	1	2	1	3	2	1	3
25	2	2	2	3	1	1	1	1	2	2	1	2	2	1	3

cinco animales restantes para realizar un test rápido del conjunto de reglas de clasificación que se obtendrán al final. El conjunto de datos con los que se trabajará se indican en la Tabla 1.

### Paso 1: Eliminación de atributos de condición redundantes<sup>2</sup>

Con el fin de trabajar con un conjunto mínimo de condiciones y utilizando los resultados expuestos en [3], se elaboraron los algoritmos correspondientes, los cuales, como se mencionó anteriormente, fueron programados en **Modula**, obteniéndose el conjunto de atributos significativos que se indican en las Tablas 2 y 3. En este paso no se utilizan los atributos de decisión, los cuales aparecen en las Tablas 2 y 3 en la columna etiquetada con el nombre de Clase.

Para los siguientes pasos, tanto la Tabla 2, como la Tabla 3, deben ser visualizadas como un conjunto de reglas donde la columna de atributos constituyen las condiciones a verificar (atributos de condición) y las clases serán las decisiones que se deben tomar (atributos de decisión).

El siguiente paso consiste en obtener el menor conjunto de reglas que nos pueda servir como herramienta para ubicar cualquier felino dentro de las cuatro clases aportadas por el experto. Hay que tener presente que las reglas que se generen dependen de la información disponible en el sistema y del conocimiento que nos aportó el experto, es decir, de la

<sup>2</sup>Los fundamentos teóricos que sustentan los algoritmos que se utilizan en este paso son tratados por otros colegas en este Congreso, en su ponencia "Descripción de clases mediante conjuntos aproximados".

Tabla 4

Reductos					
Conjunto 1	a	b	d	g	
Conjunto 2			d		j

Tabla 5

Reglas	Atributos de condición			Atributos de decisión
	d	g	j	
1	1	3	2	1
2	3	3	2	1
3	3	3	2	1
4	1	2	3	1
5	1	3	3	4
6	2	2	3	3
7	3	2	3	2
8	1	2	1	3
9	2	2	2	3
10	2	2	1	3
11	2	2	1	3
12	2	1	2	3

Tabla 6

Reglas	Atributos de condición			Atributos de decisión
	d	g	j	
13	2	1	3	3
14	3	1	2	3
15	3	1	2	3
16	2	1	2	3
17	2	1	1	3
18	3	1	2	3
19	2	1	2	3
20	3	1	1	3
21	2	1	1	3
22	3	1	1	3
23	3	1	2	3
24	3	1	2	3
25	3	1	2	3

### Paso 5: Cálculo del reducto para el conjunto de reglas

Una vez que se ha determinado, para cada regla, el conjunto de atributos indispensables, es decir, el núcleo, el siguiente paso consiste en calcular los reductos, esto es, obtener para cada regla  $i$ , con  $1 \leq i \leq 14$ , aquellos conjuntos  $P$  tales que  $P \in \{d, g, j\}$  y para los cuales  $(P, \{d, g, j\})$  es consistente.

Para el cálculo de los correspondientes reductos se utilizaron los conceptos dados en la sección 2.3 (1 y 3) y el Teorema 2. Los resultados que se obtuvieron se muestran en la Tabla 9. Observe que cada regla de la Tabla 8 puede a su vez generar más de una regla reducida. Eliminando las reglas repetidas se obtiene la Tabla 10.

Las reglas anteriores se pueden representar en forma normalizada, aplicando reglas de factorización, sin embargo, para efectos de este trabajo representaremos las 11 reglas anteriores en forma separada y sin factorizar, como se indica en el paso 6.

### Paso 6: Escritura de las correspondientes reglas

#### Regla 1

Si (comportamiento de rapiña, 1) y (longitud de la cola con respecto al cuerpo, 2) **entonces** (tipo de felino, 1).

#### Regla 2

Si (tamaño de la cruz, 3) y (longitud de la cola con respecto al cuerpo, 2) **entonces** (tipo de felino, 1).

#### Regla 3

Si (comportamiento de rapiña, 3) y (tamaño de la cruz, 3) **entonces** (tipo de felino, 1).

#### Regla 4

Si (comportamiento de rapiña, 1) y (tamaño de la cruz, 2) y (longitud de la cola con respecto al cuerpo, 3) **entonces** (tipo de felino, 1).



Tabla 7

Reglas	Atributos de condición			Atributos de decisión
	<i>d</i>	<i>g</i>	<i>j</i>	
1	1	3	2	1
2	3	3	2	1
3	1	2	3	1
4	1	3	3	4
5	2	2	3	3
6	3	2	3	2
7	1	2	1	3
8	2	2	2	3
9	2	2	1	3
10	2	1	2	3
11	2	1	3	3
12	3	1	2	3
13	2	1	1	3
14	3	1	1	3

Tabla 8

Reglas	Atributos de condición			Atributos de decisión
	<i>d</i>	<i>g</i>	<i>j</i>	
1	-	-	2	1
2	-	3	-	1
3	1	2	3	1
4	-	3	3	4
5	2	-	-	3
6	3	-	-	2
7	-	-	1	3
8	-	-	-	3
9	-	-	-	3
10	-	-	-	3
11	-	-	-	3
12	-	1	-	3
13	-	-	-	3
14	-	-	-	3

**Regla 5**

Si (tamaño de la cruz, 3) y (longitud de la cola con respecto al cuerpo, 3) **entonces** (tipo de felino, 4).

**Regla 6**

Si (comportamiento de rapiña, 2) **entonces** (tipo de felino, 2).

**Regla 7**

Si (comportamiento de rapiña, 3) y (tamaño de la cruz, 2) **entonces** (tipo de felino, 2)

**Regla 8**

Si (comportamiento de rapiña, 3) y (longitud de la cola con respecto al cuerpo, 3) **entonces** (tipo de felino, 2)

**Regla 9**

Si (longitud de la cola con respecto al cuerpo, 1) **entonces** (tipo de felino, 3)

**Regla 10**

Si (tamaño de la cruz, 2) y (longitud de la cola con respecto al cuerpo, 2) **entonces** (tipo de felino, 3).

**Regla 11**

Si (tamaño de la cruz, 1) **entonces** (tipo de felino, 3).

**Paso 7: Test de verificación de las reglas**

Como se mencionó al inicio de la sección, el conjunto de datos original contenía la información correspondiente a 30 animales, aunque para efectos del estudio únicamente se trabajó con 25, con el fin de clasificar a los cinco animales restantes por medio de las reglas reducidas y comparar los resultados con los brindados por el experto. La información correspondiente se muestra en la Tabla 11.

Según lo que se obtuvo en los pasos anteriores los únicos atributos significativos para la clasificación son el *d*, *g* y *j*, por lo que simplificando la tabla se tiene la información que se muestra en la Tabla 12. Si al conjunto de datos anterior se le aplican las reglas obtenidas, el resultado es la clasificación que se muestra en la Tabla 13.

Tabla 9

Reglas	Atributos de condición			Atributos de decisión
	<i>d</i>	<i>g</i>	<i>j</i>	
1	1	-	2	1
	-	3	2	
2	3	3	-	1
	-	3	2	
3	1	2	3	4
4	-	3	3	3
5	2	-	-	2
6	3	2	-	3
	3	-	3	
7	-	-	1	3
8	2	-	-	3
	-	2	2	
9	2	-	-	3
	-	-	1	
10	2	-	-	3
	-	1	-	
11	2	-	-	3
	-	1	-	
12	-	1	-	3
13	2	-	-	3
	-	1	-	
	-	-	1	
14	-	-	1	3
	-	1	-	

Tabla 10

Reglas	Atributos de condición			Atributos de decisión
	<i>d</i>	<i>g</i>	<i>j</i>	
1	1	-	2	1
	-	3	2	
2	3	3	-	1
3	1	2	3	1
4	-	3	3	4
5	2	-	-	3
6	3	2	-	2
	3	-	3	3
7	-	-	1	3
8	-	2	2	3
9	-	1	-	3

Tabla 11

Felino	Atributo													Clase	
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>m</i>		<i>n</i>
1	2	1	2	3	1	2	3	3	2	1	2	1	2	1	1
2	1	1	2	3	1	1	1	1	1	1	1	3	2	1	3
3	3	1	2	3	1	1	1	1	1	2	1	3	2	2	3
4	1	1	2	3	1	1	1	1	1	2	1	3	2	1	3
5	4	1	2	3	1	1	1	1	1	3	1	3	2	1	3

Nótese que la ubicación de los animales 1 y 5, usando las reglas generadas, se presenta dudosa y no coincide totalmente con la decisión del experto. Las causas del por qué se presenta una situación como ésta pueden ser variadas, dentro de las que se pueden citar están:

- El conjunto de objetos era muy pequeño comparado con el número de atributos originales y los posibles valores que pueden tomar éstos, por cuanto, al ser 14 atributos, las combinaciones que se pueden hacer con éstos y sus posibles valores da un número muy alto, dejando por fuera demasiadas posibilidades y ser poco representativas las 25 muestras.
- El número de animales que se tenía por clase en la muestra no era muy representativo, pues únicamente se tenía información de un animal en la clase 4, y de otro en la clase

Tabla 12

Felino	Atributo			Clase
	<i>d</i>	<i>g</i>	<i>j</i>	
1	3	3	1	1
2	3	1	1	3
3	3	1	2	3
4	3	1	2	3
5	3	1	3	3

Tabla 13

Felino	Atributo			Clase
	<i>d</i>	<i>g</i>	<i>j</i>	
1	3	3	1	1 ó 3
2	3	1	1	3
3	3	1	2	3
4	3	1	2	3
5	3	1	3	2 ó 3

2, de 4 animales en la clase 1 y de 19 en la clase 3. Por lo que el sistema describía aceptablemente la clase 3, pero era muy pobre con respecto al resto de clases.

## 4 Conclusiones

Actualmente los conjuntos aproximados son utilizados en aplicaciones prácticas que involucren la construcción de tablas de decisión o bases de conocimiento en sistemas expertos, tal y como se menciona en [8]. Si bien es cierto esta es una teoría cuya aplicación es recomendable cuando la información disponible no es muy grande, para garantizar una efectiva implementación de los resultados es necesario proponer algoritmos muy eficientes, e incluso relacionarla con otras metodologías, tales como los algoritmos genéticos o redes neuronales. Queda pendiente, también, el aplicar esta metodología a casos concretos de nuestra realidad.

## Bibliografía

- [1] M. Montalbano (1974) *Decision Tables*. Science Research Associates Inc., USA.
- [2] A. Mrozek (1989) "Rough Set Dependency Analysis Among Attributes in Computer Implementation of Expert Inference Models", *Int. Journal of Man-Machine Studies*, Num. 30, pp. 243-259.
- [3] Z. Pawlak (1982) "Rough Sets", *International Journal of Computer and Information Sciences*, 11(9): 341-356.
- [4] Z. Pawlak (1984) "Rough Classification", *Int. J. Man-Machine Studies* 20: 469-483.
- [5] Z. Pawlak y E. Orłowska (1984) "Expressive Power of Knowledge Representation Systems", *Int. J. Man-Machine Studies* 20: 485-500.
- [6] Z. Pawlak (1986) "Rough Classification of Patients After Highly Selective Vagotomy for Duodenal Ulcer", *Int. Journal of Man-Machine Studies* 24: 413-433.
- [7] Z. Pawlak (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [8] Z. Pawlak (1995) "Rough Sets", *Communications of the ACM* 38(11): 89-96.
- [9] A. Szkladow y W. Ziarko (1993) "Rough Sets: Working with Imperfect Data", *AI Expert* 8(7): 36-41.

Tabla 12

Felino	Atributo			Clase
	<i>d</i>	<i>g</i>	<i>j</i>	
1	3	3	1	1
2	3	1	1	3
3	3	1	2	3
4	3	1	2	3
5	3	1	3	3

Tabla 13

Felino	Atributo			Clase
	<i>d</i>	<i>g</i>	<i>j</i>	
1	3	3	1	1 ó 3
2	3	1	1	3
3	3	1	2	3
4	3	1	2	3
5	3	1	3	2 ó 3

2, de 4 animales en la clase 1 y de 19 en la clase 3. Por lo que el sistema describía aceptablemente la clase 3, pero era muy pobre con respecto al resto de clases.

## 4 Conclusiones

Actualmente los conjuntos aproximados son utilizados en aplicaciones prácticas que involucren la construcción de tablas de decisión o bases de conocimiento en sistemas expertos, tal y como se menciona en [8]. Si bien es cierto esta es una teoría cuya aplicación es recomendable cuando la información disponible no es muy grande, para garantizar una efectiva implementación de los resultados es necesario proponer algoritmos muy eficientes, e incluso relacionarla con otras metodologías, tales como los algoritmos genéticos o redes neuronales. Queda pendiente, también, el aplicar esta metodología a casos concretos de nuestra realidad.

## Bibliografía

- [1] M. Montalbano (1974) *Decision Tables*. Science Research Associates Inc., USA.
- [2] A. Mrozek (1989) "Rough Set Dependency Analysis Among Attributes in Computer Implementation of Expert Inference Models", *Int. Journal of Man-Machine Studies*, Num. 30, pp. 243-259.
- [3] Z. Pawlak (1982) "Rough Sets", *International Journal of Computer and Information Sciences*, 11(9): 341-356.
- [4] Z. Pawlak (1984) "Rough Classification", *Int. J. Man-Machine Studies* 20: 469-483.
- [5] Z. Pawlak y E. Orłowska (1984) "Expressive Power of Knowledge Representation Systems", *Int. J. Man-Machine Studies* 20: 485-500.
- [6] Z. Pawlak (1986) "Rough Classification of Patients After Highly Selective Vagotomy for Duodenal Ulcer", *Int. Journal of Man-Machine Studies* 24: 413-433.
- [7] Z. Pawlak (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [8] Z. Pawlak (1995) "Rough Sets", *Communications of the ACM* 38(11): 89-96.
- [9] A. Szkladow y W. Ziarko (1993) "Rough Sets: Working with Imperfect Data", *AI Expert* 8(7): 36-41.

## ESTUDIO DE MÉTRICAS EN ANÁLISIS EN COMPONENTES PRINCIPALES MEDIANTE GENERADORES PRIMARIOS

FABRICIO BOLAÑOS<sup>1</sup>

Se aborda el problema de la escogencia de la métrica  $M$  en el espacio de individuos  $E = \mathbb{R}^p$  para el Análisis en Componentes Principales (A.C.P.) de una tabla centrada  $X_{n \times p}$ . Y. Schektman [3] propuso los generadores de métricas, entre ellos el generador primario, que se implementó en el caso del A.C.P. para hacer una exploración de la métrica  $M$ . Los generadores primarios de métricas se pueden definir por:  $M_r = w_1(r)\bar{M}_1 + w_2(r)\bar{M}_2$ , donde  $w_1(r) = \frac{m-r}{m} \in [0, 1]$ ,  $w_2(r) = 1 - w_1(r) \in [0, 1]$  y  $r \in \{0, \dots, m\}$ ,  $m \in \mathbb{N}$ .  $\bar{M}_1$  y  $\bar{M}_2$  pueden ser escogidos en el conjunto de métricas usuales  $\{V, V^{-1}, I, D_{1/\sigma^2}\}$ . Existen otros tipos de combinación [1] que aquí no presentamos. Citamos a continuación algunas propiedades que hemos obtenido en [1]

**Proposición 1 a)** Si  $\bar{M}_1, \bar{M}_2 \in \{V, V^{-1}, I_p\}$ , entonces los vectores propios de  $M_r$  son idénticos a los de  $V$  y los valores propios correspondientes se deducen de los de  $V$ .

b) Si  $\bar{M}_1 = D_{1/\sigma^2}$ ,  $\bar{M}_2 = V^{-1}$ , entonces los vectores propios de  $VM_r$  son iguales a los de  $VD_{1/\sigma^2}$ .

c) Si  $\lambda_1, \dots, \lambda_p$  son los valores propios de  $V$ , asociados a los vectores propios  $u_1, \dots, u_p$ , entonces tenemos:

1. Si  $\bar{M}_1 = V$ ,  $\bar{M}_2 = I$ , entonces los vectores propios de  $VM_r$  son iguales a los de  $V$  y los valores propios son:  $(\frac{m-r}{m})\lambda_i^2 + \frac{r}{m}\lambda_i$ .

2. Si  $\bar{M}_1 = I$ ,  $\bar{M}_2 = V^{-1}$  y  $r < m$ , entonces los vectores propios de  $VM_r$  son iguales a los de  $V$  y los valores propios son:  $(\frac{m-r}{m})\lambda_i + \frac{r}{m}$ .

d) Si  $\bar{M}_1 = V$ ,  $\bar{M}_2 = I$  entonces las componentes principales se pueden calcular por:  
 $C_\alpha^k = \sqrt{[(1-\alpha)\lambda_k + \alpha]} C_1^k$ .

e) Si  $\bar{M}_1 = I$ ,  $\bar{M}_2 = V^{-1}$  y  $r < m$ , entonces las componentes principales restantes se pueden calcular por:  $C_\alpha^k = \sqrt{[(1-\alpha) + \frac{\alpha}{\lambda_k}]} C_1^k$ .

<sup>1</sup> ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

- f) Si  $\bar{M}_1 = V$  y  $\bar{M}_2 = I$  entonces las correlaciones de las variables con las componentes principales de los respectivos A.C.P. son las mismas para todo  $r$ .
- g) Si  $\bar{M}_1 = I$ ,  $\bar{M}_2 = V^{-1}$  y si  $r < m$ , entonces las correlaciones de las variables con las componentes principales de los respectivos A.C.P. son las mismas que las anteriores.

La propiedad anterior también permite deducir propiedades interesantes para la inercia de la nube de puntos.

En [1] se pueden consultar los resultados obtenidos sobre varias tablas de datos.

## Bibliografía

- [1] Bolaños, F (1996) *Contribución al estudio de las métricas en Análisis de Datos*. Tesis de Licenciatura.
- [2] Cailliez, F.; Pagès, J.P. (1976) *Introduction à l'Analyse de Données*. SMASH, París.
- [3] Schektman, Y. (1983) *Quelques réflexions sur les métriques. Une approche originale pour la protection des méthodes relationnelles et l'étude de la régression multidimensionnelle*. Nota interna, Laboratorio MLAD, Universidad Paul Sabatier, Toulouse.
- [4] Trejos, J. (1994) *Génération de produits scalaires pour l'Analyse des Données*. Nota interna, PIMAD, Escuela de Matemática, Universidad de Costa Rica. 31pp.
- [5] Trejos, J.; Bolaños, F. (1995) "Generación de métricas útiles en análisis de datos," *IX Simposio Métodos Matemáticos Aplicados a las Ciencias*, J. Trejos (ed.), U.C.R.-I.T.C.R., Turrialba, pp 173-182.

# ESTUDIO EVOLUTIVO DE TABLAS DE CONTINGENCIA: APLICACIÓN A LOS EGRESOS HOSPITALARIOS DE MUJERES ENTRE 20 Y 44 AÑOS <sup>1</sup>

WILLIAM CASTILLO<sup>2</sup> - JORGE GONZÁLEZ<sup>2</sup>

---

## Resumen

En este artículo se presenta un resumen del Análisis en Componentes Principales Evolutivo mediante el cual se obtienen representaciones planas óptimas de las trayectorias. Se usa este método para estudiar la evolución de las enfermedades en mujeres de edad entre 20 y 44 años, egresadas del sistema hospitalario de Costa Rica, en el período 1980-90. En esta aplicación una trayectoria representa los cambios que una profesión experimenta durante el período considerado, en relación con las enfermedades.

**Palabras Claves:** ACP evolutivo, trayectoria, correlación longitudinal, tabla de contingencia.

## 1 Introducción

Nuestro objetivo es estudiar la evolución de un conjunto de  $K$  tablas de contingencia, de tamaño  $q \times p$  medidas anualmente y que denotamos  $N_k = (n_{ijk})$ , ( $k = 1, \dots, K$ ), donde  $n_{ijk}$  es el número de individuos que en el año  $k$  tienen al mismo tiempo la modalidad  $i$  en la primera variable y la modalidad  $j$  en la segunda variable.

En este trabajo analizaremos 11 tablas correspondientes a los años 1980 a 1990. Cada tabla es el cruce entre la variable "Profesión" con 9 modalidades y la variable "Enfermedad" con 14 modalidades, medidas en el conjunto de mujeres con edad entre 20 y 44 años atendidas durante este período en la CCSS. Usaremos la palabra profesión para referirnos a la primera variable y enfermedad para la segunda, aunque en algunos casos no sean ni profesiones ni enfermedades.

Es nuestro interés observar si alguna enfermedad tiene un comportamiento característico en una determinada categoría de empleo y si este comportamiento se modifica

---

<sup>1</sup>Esta ponencia es resultado de las investigaciones de los autores en el Programa de Investigación en Modelos y Análisis de Datos (PIMAD).

<sup>2</sup>PIMAD, ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

al pasar de un año a otro. También es interesante detectar si alguna profesión presenta tendencias preferentes por alguna enfermedad.

El procedimiento que utilizaremos se conoce en la literatura como "Estudio Evolutivo de Tablas de Contingencia" (EVTC) y propuesto por A. Carlier [2], el cual recurre a un particular Análisis en Componentes Principales (ACP) que busca reemplazar las " $p$ " categorías de la segunda variable por un número menor de nuevas variables llamadas factores, a través de los cuales observamos los cambios en las categorías de la primera variable ( trayectorias) en el tiempo.

Las características más sobresalientes de la evolución son detectadas usando un criterio de maximización que sólo toma en cuenta los cambios entre dos celdas consecutivas ( $n_{ijk} - n_{ij(k-1)}$ ). Para medir estos cambios es necesario adaptar los conceptos clásicos de inercia, varianza, Análisis en Componentes Principales y que serán denominados respectivamente como: evolución total (EVT), varianza longitudinal (varl) y Análisis en Componentes Principales Evolutivo (ACPE).

## 2 Definiciones básicas

Para analizar la distribución de la profesión  $i$  en las diferentes enfermedades, en el  $k$ -ésimo año, utilizamos el vector  $P_{ik} \in \mathbb{R}^p$ , definido por  $P_{ik} = (p_{ijk})_{j \in J} = \Pi_{ik} \delta_{ik}$ , donde  $\Pi_{ik} \in \mathbb{R}$  es un indicador del tamaño y  $\delta_{ik} = (\delta_{ijk})_{j \in J} \in \mathbb{R}^p$  es, en cada entrada, la fracción de la profesión  $i$  en cada enfermedad  $j$  en el año  $k$ .

Definimos:

$$\sum_{j=1}^p \sum_{k=1}^K n_{ijk} = n_{i..}, \quad \sum_{j=1}^p n_{ijk} = n_{i.k}, \quad \Pi_{ik} = \frac{K n_{i.k}}{n_{i..}}, \quad \delta_{ik} = \frac{1}{n_{i.k}} (n_{i1k}, \dots, n_{ipk}).$$

$\Pi_{ik}$  compara el tamaño de la categoría  $i$  en el año  $k$  con el promedio anual de esa categoría.

**Definición 2.1 (Trayectoria)** La trayectoria de la categoría  $i$  es el conjunto de  $K$  puntos consecutivos  $\{P_{ik} | k = 1, \dots, K\}$ .

## 3 Análisis en Componentes Principales Evolutivo

Por definición de  $P_{ik}$  las trayectorias se encuentran en el espacio  $\mathbb{R}^p$ , que llamamos usualmente espacio de los individuos. Nuestro interés es proyectarlas sobre un subespacio de menor dimensión usando una idea análoga al Análisis en Componentes Principales (ACP), que denominamos Análisis en Componentes Principales Evolutivo (ACPE).

El resultado final son unas representaciones planas de las trayectorias con mínima distorsión. Ello nos permite observar su comportamiento, el cual es explicado por algunas modalidades de la segunda variable.

### 3.1 Subespacio óptimo

Sea  $M = \text{diag}(b_1, \dots, b_p)$  una métrica diagonal en el espacio de los individuos  $\mathbb{R}^p$  y sea  $H$  un subespacio de  $\mathbb{R}^p$ .



**Definición 3.1 (Trayectoria proyectada)** Sea la trayectoria  $\{P_{ik}|k = 1, \dots, K\}$  en  $\mathbb{R}^p$ , llamamos trayectoria proyectada sobre  $H$  a  $\{P_H(P_{ik})|k = 1, \dots, K\}$  donde  $P_H$  es el operador de proyección  $M$ -ortogonal sobre  $H$ .

Para precisar cual es la condición de optimalidad para  $H$ , empecemos diciendo que lo que nos interesa es conservar la longitud de las trayectorias en  $H$  lo más cercana a sus longitudes en  $\mathbb{R}^p$ , por lo que nuestro interés está más enfatizado en las diferencias  $P_{ik} - P_{i(k-1)}$  que en los puntos  $P_{ik}$ . Necesitamos, por lo tanto, redefinir algunos conceptos clásicos de la estadística.

**Definición 3.2 (Nube  $P_I$ )** Denotamos con  $P_I = \{P_{ik} - P_{i(k-1)}|i = 1, \dots, q; k = 2, \dots, K\}$  la nube de  $(K-1)q$  puntos de  $\mathbb{R}^p$  y con  $D = \text{Diag}((D_s)$ , donde

$$D_s = \text{Diag}(a_s c_2, \dots, a_s c_K), \quad a_s > 0, \quad c_j > 0, \quad s = 1, \dots, q.$$

**Definición 3.3 (Evolución total de la Nube)** Denominamos Evolución Total de la nube  $P_I$  a su inercia.

$$EVT(P_I) = \sum_{i=1}^q \sum_{k=2}^K a_i c_k \|P_{ik} - P_{i(k-1)}\|_M^2 = \sum_{i=1}^q \sum_{j=1}^p \sum_{k=2}^K a_i b_j c_k (p_{ijk} - p_{ij(k-1)})^2$$

Como  $\|P_H(P_{ik} - P_{i(k-1)})\|_M \leq \|P_{ik} - P_{i(k-1)}\|_M$  podemos definir un subespacio óptimo como sigue.

**Definición 3.4 (Subespacio óptimo)** Definimos como subespacio óptimo, entre todos los subespacios de dimensión  $r$  al subespacio  $H_r$  que maximiza la inercia de la nube proyectada, esto es la cantidad:

$$EV_{H_r}(P_I) = \sum_{i=1}^q \sum_{k=2}^K a_i c_k \|P_H(P_{ik} - P_{i(k-1)})\|_M^2$$

Como sabemos, el subespacio  $H_r$  se obtiene al realizar el ACP  $(X, M, D)$  donde  $X$  es la matriz de individuos por variables cuyas filas son los puntos de la nube  $P_I$ . Luego  $H_r$  es generado por los vectores propios  $\{u_1, \dots, u_r\}$  asociados a los  $r$  primeros valores propios  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_r$  de la matriz  $VM$ , donde  $V = X^t D X$  es la matriz de varianza covarianza longitudinal, como veremos a continuación.

## 4 Varianza longitudinal, covarianza longitudinal

Notemos que la nube de puntos  $\{P_{ik}|k = 1, \dots, K\}$  define la matriz  $P$  de individuos  $(P_{ik})$  por variables  $(P^j)$  de tamaño  $Kq \times p$ :

$$P^t = (P_{11}^t, \dots, P_{1K}^t, \dots, P_{q1}^t, \dots, P_{qK}^t)$$

A su vez la nube  $P_I$  define la matriz de individuos  $(P_{ik} - P_{i(k-1)})$  por variables  $(\Delta P^j = (p_{ijk} - p_{ij(k-1)}) \in \mathbb{R}^{q(K-1)})$ :

$$X_{q(K-1) \times p}^t = \left( (P_{12} - P_{11})^t, (P_{13} - P_{12})^t, \dots, (P_{I2} - P_{I1})^t, \dots, (P_{qK} - P_{q(K-1)})^t \right)$$

Lo anterior nos sugiere las siguientes definiciones:

**Definición 4.1 (Covarianza longitudinal)** Sean las variables  $P^j$ ,  $P^{j'}$  en  $\mathbb{R}^{qK}$ . Se llama covarianza longitudinal entre  $P^j$  y  $P^{j'}$  a:

$$\text{covl}(P^j, P^{j'}) = \sum_{i=1}^q \sum_{k=2}^K a_i c_k (p_{ijk} - p_{ij(k-1)}) (p_{ij'k} - p_{ij'(k-1)}).$$

Análogamente la varianza longitudinal es

$$\text{varl}(P^j) = \sum_{i=1}^q \sum_{k=2}^K a_i c_k (p_{ijk} - p_{ij(k-1)})^2$$

Si  $\langle \cdot, \cdot \rangle_D$  denota el producto interno definido por la matriz de pesos  $D$ , tenemos:

$$\text{covl}(P^j, P^{j'}) = \langle \Delta P^j, \Delta P^{j'} \rangle_D.$$

Además si  $\text{varl}(P^j) = 0$ ,  $P^j$  se tiene que en cada individuo la variable  $j$  es constante en el tiempo ( $P^j$  es constante a trozos).

Diremos que el conjunto  $\{P^j\}_{j \in J}$  es  $l$ -ortonormal si el conjunto  $\{\Delta P^j\}_{j \in J}$  es  $D$ -ortonormal

## 5 Análisis de los individuos y variables

Del ACP de  $(X, M, D)$  tenemos:

$U = (u_1, \dots, u_p)$  matriz de vectores propios de  $VM$  asociados a los  $p$ -primeros valores propios,  $H_r$  el subespacio óptimo generado por los  $r$ -primeros vectores propios y  $C = (g^1, \dots, g^p)$  la matriz de componentes principales  $D$ -normalizadas,  $g^l \in \mathbb{R}^{q(K-1)}$ .

Para trazar las trayectorias sobre el subespacio óptimo  $H_r$  proyectamos (en suplementario) los puntos  $P_{ik}$ :  $P_H(P_{ik}) = \sum_{l=1}^r \sqrt{\lambda_l} f_{(ik)}^l u_l$ , donde  $\lambda_l$  es el valor propio  $l$ -ésimo y  $f_{(ik)}^l$  es un factor que calculamos a continuación.

Como  $P_H(P_{ik} - P_{i(k-1)}) = \sum_{l=1}^r \sqrt{\lambda_l} g_{ik}^l u_l$  se sigue que:  $f_{ik}^l - f_{i(k-1)}^l = g_{ik}^l$ . Al variar  $i = 1, \dots, q$ ,  $k = 2, \dots, K$  tenemos que  $f^l \in \mathbb{R}^{Kq}$  y  $\Delta f^l = g^l$ .

Si  $F = (f^1, \dots, f^p)$  es la matriz de tamaño  $qK \times p$  tenemos la relación matricial

$$PMUD \frac{1}{\sqrt{\lambda}} = F, \text{ donde } D \frac{1}{\sqrt{\lambda}} = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_p}}\right)$$

A los elementos del conjunto  $\{f^s\}$  los denominamos factores y ellos forman una base  $l$ -ortonormal para las variables  $p^j \in \mathbb{R}^{Kq}$ , puesto que  $\text{covl}(f^j, f^{j'}) = \langle g^j, g^{j'} \rangle_D$ .

## 6 Definición de las métricas

La escogencia de los parámetros  $a_i, c_k, b_j$  que definen las métricas  $M$  y  $D$  se hace de modo que se satisfagan las siguientes propiedades de estabilidad: al fusionar dos modalidades  $i$  y  $h$  tales que  $N_i = (n_{ijk})_{jk}$  y  $N_h = (n_{hjk})_{jk}$  sean proporcionales, la evolución total no cambia. Una condición análoga se exige para las modalidades de la otra variable. En el caso de la variable tiempo, la evolución total debe permanecer invariable cuando se realiza una nueva observación  $N_{k_0}$  entre los instantes  $k-1$  y  $k$ , tal que para algún  $\eta \in ]0, 1[$ ,  $N_{k_0} = N_{k-1} + \eta(N_k - N_{k+1})$ , donde  $N_k = (n_{ijk})_{ij}$ .

De las métricas propuestas por Carlier [2], se escogieron las siguientes porque se adecúan bien a los datos y los cálculos son más sencillos.

$$a_i = \frac{(n_{i..})^2}{K^2 \sum_j \sum_{k \geq 2} |n_{ijk} - n_{ij(k-1)}|}, i = 1, \dots, q.$$

$$b_j = \frac{1}{\sum_i \sum_{k \geq 2} |n_{ijk} - n_{ij(k-1)}|}, j = 1, \dots, p.$$

$$c_k = \frac{1}{\sum_i \sum_j |n_{ijk} - n_{ij(k-1)}|}, k = 1, \dots, K.$$

## 7 Estudio evolutivo de los egresos del sistema hospitalario estatal de Costa Rica

La metodología presentada se implementó en *Mathematica*<sup>3</sup>, lo cual permitió su aplicación al análisis de una sucesión de tablas de contingencia obtenidas cruzando las variables *profesión* y *enfermedad*, de las mujeres con edad entre 20 y 44 años, egresadas del sistema hospitalario nacional. Se consideró un período de 11 años desde 1980 hasta 1990 para un total de 11 tablas de datos<sup>4</sup>.

El análisis de los resultados del ACPE se hace en cuatro fases : descripción de los datos, el cálculo de los índices numéricos, la lectura de las trayectorias y las conclusiones.

### 7.1 Descripción de los datos

La variable *profesión* tiene las siguientes 11 modalidades (entre paréntesis aparece el número de las profesiones que se usaron en el estudio): Sin ocupación; profesionales y técnicos (1); directores, gerentes y administradores (2); empleados estatales (3); comerciantes vendedores (4); agricultores, ganaderos y trabajadores agrícolas (5); transportistas

<sup>3</sup>Los programas usados fueron elaborados por el profesor Carlos Arce S., de la Escuela de Matemática de la U.C.R.

<sup>4</sup>Los datos fueron aportados por la profesora Mayra Achío del Instituto de Investigaciones Sociales de la U.C.R.

(6); artesanos industriales textiles (7); artesanos industriales gráficos (8); estibadores, cargadores, bodegueros (9); servicios personales .

La variable *enfermedad* tiene las siguientes 16 modalidades (entre paréntesis aparece el número de las enfermedades que se usaron en el estudio): tumores benignos (1); psicosis (2); de las venas, vasos linfáticos y otras (3); asma (4); coleditiasis y colesistit (5); otras del aparato digestivo (6); aparato urinario (7); enfermedades de la mama (8); otras genitales femeninas (9); aborto (10); obstétricas directas; obstétricas indirectas (11); parto normal; piel y celular subcutáneo (12); otras osteomusculares (13); otras suplementarias (14).

Las modalidades sin ocupación (62%) y servicios personales (22%) representan juntas un 84% de la muestra, lo cual implicó que a nivel de las dos primeras direcciones del ACPE, estas modalidades se manifestaron con mucha fuerza opacando por completo las otras. Lo mismo ocurre con los diagnósticos *obstétricas directas* y *parto normal*. Al excluir del análisis estas modalidades nos quedamos con solo el 16% de la muestra, pero ello nos permitió observar algunas características a nivel de los grupos minoritarios de las mujeres jóvenes.

## 7.2 Índices numéricos : una ayuda para la interpretación

Los índices se refieren tanto a los individuos como a las variables. Los índices *globales* permiten seleccionar los elementos que entran en juego en el análisis de acuerdo con su aporte a la evolución total. Los índices de calidad sirven para juzgar qué tan buena es la aproximación de esos elementos en un subespacio de baja dimensión por proyección ortogonal. Expresando la evolución total en términos de las componentes principales y de los valores propios del ACPE, resulta natural definir los siguientes índices para los individuos (trayectorias) :

**Contribución del individuo  $i$  a EVT:**

$$ctr(i) = a_i \sum_{k=2}^K c_k \sum_{\alpha=1}^p \lambda_{\alpha} (g_{ik}^{\alpha})^2$$

donde  $EVT = \sum_{i=1}^q ctr(i)$ . Por lo que el porcentaje de evolución atribuible a  $i$  es :  $porc(i) = 100 \frac{ctr(i)}{\sum \lambda_{\alpha}}$ .

En el caso de los datos que analizamos, las profesiones juegan el papel de los individuos y los valores de  $porc(i)$  se reportan en la tabla que sigue :

$i$	1	2	3	4	5	6	7	8	9
$porc(i)$	24.6	5.3	22.4	11.7	6.0	1.5	15.8	6.6	6.0

**Calidad de la representación de una trayectoria :** tenemos que

$$ctr_r(i) = a_i \sum_{k=2}^K c_k \sum_{\alpha=1}^r \lambda_{\alpha} (g_{ik}^{\alpha})^2$$

es la contribución de la trayectoria  $i$  a la evolución proyectada sobre  $H_r$ . Por tanto la calidad de la representación de la trayectoria  $i$  se mide por medio de :  $porc_r(i) = 100 \frac{ctr_r(i)}{ctr(i)}$ .

Nuestro análisis se limita al primer plano principal con el 37% de la EVT. Para este plano los valores  $\text{porc}_2(i)$  son :

$i$	1	2	3	4	5	6	7	8	9
$\text{porc}_2(i)$	57.6	18.8	32.6	18.8	18.7	9.9	44.3	35.5	19.8

### 7.3 Índices de las variables

Similarmente se puede expresar la EVT en términos de los factores  $f^\alpha$  y de los valores propios del ACPE y definir los siguientes índices para las variables:

**Contribución de una variable  $P^j$  a EVT:**  $\text{ctrvar}(j) = b_j \text{var}_l(P^j) = b_j \sum_{\alpha=1}^r \lambda_\alpha u_{\alpha j}^2$ .

El porcentaje de la evolución atribuible a la variable es  $\% \text{ctrvar}(j) = 100 \frac{\text{ctrvar}(j)}{\text{EVT}}$ , donde  $\text{EVT} = \sum_{j=1}^p \text{ctrvar}(j)$ .

**Calidad de la representación de una variable en el subespacio  $F_r$ :**  $\text{calv}(j) = 100 \frac{\text{var}_l(\hat{P}^j)}{\text{var}_l(P^j)}$ . Es decir:  $\% \text{calv}(j) = 100 \frac{\sum_{\alpha=1}^r \lambda_\alpha u_{\alpha j}^2}{\sum_{\alpha=1}^p \lambda_\alpha u_{\alpha j}^2}$ .

Las variables (enfermedades) que muestran los mayores valores en los índices anteriores son las que explican el movimiento de las trayectorias. En nuestro caso son las que se indican en la tabla que sigue:

Enfermedad	$\% \text{ctrvar}(j)$	$\% \text{calv}(j)$
9	15.5	87
10	12.3	73
13	8.4	37
14	9.2	34

### 7.4 Interpretación de los factores

La asociación de las variables  $P^j$  con los factores  $f^\alpha$  se juzga por medio de la correlación longitudinal y el aporte de una variable a la evolución total "explicada" por un factor.

**Correlación de  $P^j$  con  $f^\alpha$ :** Como  $P^j$  se escribe en la forma  $P^j = \sum_{\alpha=1}^p \sqrt{\lambda_\alpha} u_{\alpha j} f^\alpha$  y los factores  $f^\alpha$  son  $l$ -ortonormados, se deduce que  $\langle P^j, f^\alpha \rangle_l = \langle \sum_{\alpha=1}^p \sqrt{\lambda_\alpha} u_{\alpha j} f^\alpha, f^\alpha \rangle_l = \sqrt{\lambda_\alpha} u_{\alpha j}$ . Por lo tanto la correlación entre  $P^j$  y  $f^\alpha$  es :

$$\text{corrl}(P^j, f^\alpha) = \frac{\langle P^j, f^\alpha \rangle_l}{\|P^j\|_l \|f^\alpha\|_l} = \frac{\sqrt{\lambda_\alpha} u_{\alpha j}}{\sqrt{\text{var}_l(P^j)}}$$

**Aporte de una variable al factor:** Sea  $\text{Pr}_{f^\alpha}(P^j) = \langle P^j, f^\alpha \rangle_l f^\alpha = \sqrt{\lambda_\alpha} u_{\alpha j} f^\alpha$ . Por lo tanto  $\text{var}_l(\text{Pr}_{f^\alpha}(P^j)) = \|\text{Pr}_{f^\alpha}(P^j)\|_l^2 = \lambda_\alpha u_{\alpha j}^2$ . Es decir  $\sum_{j=1}^p b_j \text{var}_l(\text{Pr}_{f^\alpha}(P^j)) = \lambda_\alpha \|u_\alpha\|_M^2 = \lambda_\alpha$ . Entonces el aporte de una variable a la evolución explicada por el factor  $f^\alpha$  se define por  $\% \text{apo}(P^j, f^\alpha) = 100 \frac{b_j \text{var}_l(\text{Pr}_{f^\alpha}(P^j))}{\lambda_\alpha} = 100 b_j u_{\alpha j}^2$ .

Cuanto mayor sea la magnitud de estos índices para unas variables y un factor, más las variables han determinado al factor y éste se interpreta como representativo de la variación de las variables. En el caso que nos ocupa las variables anteriores son las que tienen los valores más grandes en esos índices :

Enfermedad	$\text{corr}_l(P^j, f^1)$	$\% \text{apo}(P^j, f^1)$	$\text{corr}_l(P^j, f^2)$	$\% \text{apo}(P^j, f^2)$
9	-0.88	51	-0.57	23
10	-0.65	22	-0.56	30
13	-0.44	7	-0.41	11
14	-0.41	3	-0.40	6

### 7.5 Lectura del movimiento de las trayectorias

Hemos visto que  $P_{ik} = \Pi_{ik} \delta_{ik}$  donde  $\delta_{ik} = \left( \frac{n_{ijk}}{n_{i.k}} \right)_{j \in J}$  es la proporción en la distribución de la categoría  $i$  a través de las variables, en el año  $k$  y,  $\Pi_{ik} = \frac{K n_{i.k}}{n_{i..}} = \sum_{j=1}^p p_{ijk}$  es el tamaño o popularidad de la categoría  $i$  en el año  $k$  con respecto al promedio anual. Si  $P_{ik} \in \mathbb{R}^2$  se tienen las siguientes propiedades:

- Si una trayectoria  $i$  se “mueve” a lo largo de una línea que pasa por el origen, entonces la proporción  $\frac{\delta_{i2k}}{\delta_{i1k}} = \frac{n_{i2k}}{n_{i1k}}$  es independiente de  $k$ . En este caso el tamaño  $\Pi_{ik}$  crece si la trayectoria se aleja del origen y decrece si se acerca al origen.
- Si una trayectoria  $i$  se “mueve” paralelamente a la recta  $x + y = 1$ , entonces la proporción  $\frac{\delta_{i2k}}{\delta_{i1k}} = \frac{n_{i2k}}{n_{i1k}}$  cambia y el tamaño  $\Pi_{ik}$  se mantiene constante. Cuanto más se extienda la trayectoria más la proporción está cambiando.

Las propiedades anteriores se trasladan al espacio  $\mathbb{R}^p$  con tal que la recta  $x + y = 1$  sea sustituida por el hiperespacio  $H$  de ecuación:  $\sum_{j=1}^p x_j = 1$ . En efecto, si la distribución de la categoría  $i$ ,  $\delta_{ik} = \left( \frac{n_{ijk}}{n_{i.k}} \right)_j$  es constante a través de los años, entonces la trayectoria de la categoría  $i$  se mueve sobre una recta que pasa por el origen. Si por otra parte  $\Pi_{ik}$  no depende de  $k$ , la trayectoria se mueve paralelamente al hiperespacio  $H$ .

Aun cuando en proyección  $M$ -ortogonal sobre un subespacio principal, estas propiedades no se preservan, si un vector propio  $u_\alpha$  es paralelo a  $H$ , las trayectorias paralelas a  $u_\alpha$  son de tamaño constante en la medida que su movimiento es también paralelo a  $H$ . Entonces el factor  $f^\alpha$  se asocia con variaciones en la distribución  $\delta_{ik}$  año con año.

Un índice que nos permite juzgar el paralelismo de una trayectoria con un eje se construye así: el vector  $u_\alpha$  es paralelo a  $H$  si  $\sum_{j=1}^p u_{\alpha j} = 0$ . Es decir, si  $\langle u_\alpha, \beta \rangle_M = 0$ . Por otra parte,  $\cos(u_\alpha, \beta) = \frac{\langle u_\alpha, \beta \rangle_M}{\|u_\alpha\|_M \|\beta\|_M} = \langle u_\alpha, \beta \rangle_M$ . En consecuencia tenemos: si  $\cos^2(u_\alpha, \beta) \approx 0$ , es decir  $u_\alpha$  es casi  $M$  ortogonal a  $\beta$ , entonces el factor  $f^\alpha$  se asocia con variaciones en la distribución  $\delta_{ik}$  de la trayectoria  $i$  en los sectores donde ésta es aproximadamente paralela a  $u_\alpha$ .

Por otra parte, si  $\cos^2(u_\alpha, \beta) \approx 1$ , es decir  $u_\alpha$  es casi paralelo a  $\beta$ , entonces el factor  $f^\alpha$  se asocia con variaciones en el tamaño  $\Pi_{ik}$  en los sectores donde esto ocurre.

Para el primer eje se obtuvo  $\cos^2(u_1, \beta) = 0.726$  y para el segundo eje  $\cos^2(u_2, \beta) = 0.273$ . Por lo que se interpreta el primer eje como un indicador de la variación del tamaño  $\Pi_{ik}$  y el segundo como un indicador de los cambios estructurales  $\delta_{ik}$ .

## 8 Discusión de los resultados

Los gráficos que presentamos en el anexo corresponde a las cinco trayectorias con mayor aporte a EVT y mejor representadas en los dos primeros ejes. En estos gráficos cada vértice es un año y se indica el primero con 80. Con base en estos gráficos y los histogramas (no incluidos) de las enfermedades, se puede afirmar que:

- En el caso de las **profesionales y técnicos** se nota un aumento en los egresos del 80 al 81 (7% sobre la media del período) y del 81 al 82 (16%). Pero a partir del 88 desciende, alcanzando un 15% de disminución del 89 al 90.
- Hemos constatado a partir de los histogramas que hay una tendencia decreciente en el aborto alcanzando su mínimo en el año 87. A partir del año 82 hay una tendencia creciente en las enfermedades genitales hasta antes del 85. En el año 85 se alcanza el máximo en los diagnósticos de enfermedades osteomusculares. Correlativamente se tiene el mínimo para las enfermedades suplementarias y un descenso importante en el aborto.
- La trayectoria de las trabajadoras **Estatales** es similar a la anterior.
- **Comerciantes y vendedoras** no tiene variación en tamaño y su estructura varía principalmente a causa de las enfermedades osteomusculares.
- La trayectoria de **Artesanas y textileras** presenta solo variación en tamaño con tendencia creciente en el período. Este aumento se debe principalmente al aborto y a las enfermedades osteomusculares.
- La trayectoria de las trabajadoras de la **artesanía gráfica** indica una variación zigzagueante en tamaño y estructura. Este fenómeno se debe a que cada año crece el aborto y disminuyen las enfermedades genitales o viceversa.

Todas estas observaciones han sido verificadas con los datos.

## 9 Conclusiones

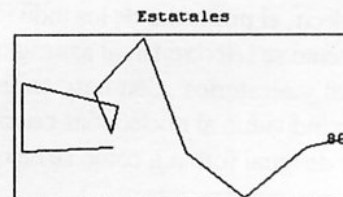
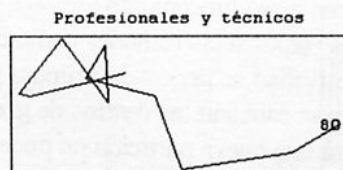
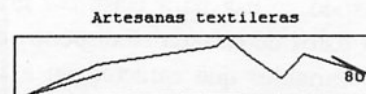
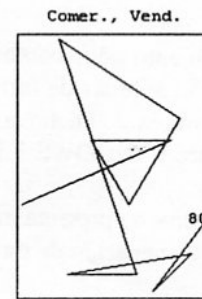
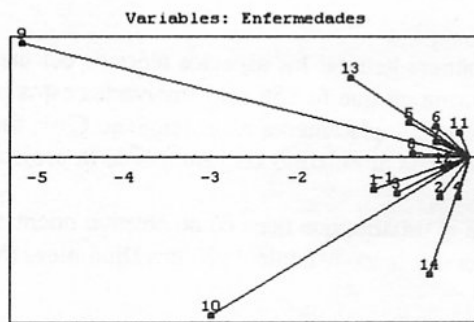
De acuerdo con los resultados obtenidos se puede afirmar que este método da cuenta de manera bastante satisfactoria de las variaciones en tamaño y en estructura de una sucesión de tablas de contingencias. Los gráficos han revelado claramente estas variaciones en las trayectorias. El gráfico de las variables ha servido para explicar dichas variaciones en términos de las cuatro enfermedades más importantes. Todas estas observaciones se confirman a nivel de los datos.

## Bibliografía

- [1] Cailleux, F. y Pagès, J.P. (1976) *Introduction a l'analyse des données*. SMASH, París.

- [2] Carlier, A. (1985) "Analyse des évolutions de tables de contingence, quelques aspects opérationnels. *Data Analysis and Informatics*, 4, E. Diday et al. (eds.), Amsterdam : North Holland, 421-428. París.
- [3] Carlier, A. y Ewing, G. (1991) *Contingency Table Analysis of Evolution with application to demographical data*. (Informe de investigación).
- [4] Castillo, W., González, J. (1994) "Análisis de tablas múltiples de datos," *Rev. de Matemática: Teoría y Aplicaciones*, 1(1): 47-55.
- [5] Escofier, B. y Pagès, J. (1988) *Analyses Factorielles Simples et Multiples*. Dunod, París.
- [6] Lavit, Ch. (1988) *Analyse Conjointe de Tableaux Quantitatifs*. Méthodes + Programmes. Masson, París.
- [7] Saporta, G. y Lavallard, F. (1996) *L'Analyse des Données Evolutives: Méthodes et Applications*. Editions Technip, París.

## Anexo





## SISTEMA DE GENERACIÓN DE PARTICIONES

BEN CHACÓN PEREIRA<sup>1</sup>

---

### Resumen

El proceso de particionamiento llamado Método de Nubes Dinámicas es combinado con el Análisis de Formas Fuertes en un paquete útil y de uso sencillo. Son presentadas las notas teóricas, los algoritmos implementados y una pequeña guía de uso del sistema ilustrada con gráficos.

## 1 Introducción

Este documento tiene como fin describir de manera general los aspectos teóricos del método de nubes dinámicas y el análisis de formas fuertes, y la forma en que fueron implementados estos procesos en el programa "parti.exe". Dicho sistema está construido completamente en el lenguaje C++, fue hecho para ejecutarse sobre WINDOWS 3.1 o superior, y aprovecha al máximo las ventajas de la programación orientada a objetos.

Posteriormente se presenta una guía rápida para el usuario, que tiene como objetivo orientarlo en el uso del Sistema de Generación de Particiones, también denominado Módulo de Nubes Dinámicas (MND).

## 2 El Método de Nubes Dinámicas

Este método se usa para clasificar grandes cantidades de datos homogéneos, es decir, del mismo tipo. Los datos de entrada se disponen dentro de una matriz, donde las columnas pueden ser vistas como las variables que caracterizan a los individuos y las filas como los individuos a ser clasificados. El método de nubes dinámicas utiliza esta matriz para crear una partición que dividirá el conjunto de individuos en subconjuntos disjuntos cuyos elementos serán muy semejantes, pero diferentes de los de los otros subconjuntos.

En el método de nubes dinámicas cada subconjunto de individuos es representado por su centro de gravedad, es decir, el promedio de los individuos que pertenecen a ese subconjunto.

Inicialmente se seleccionan al azar varios individuos, los cuales serán tomados como centros de gravedad iniciales y aleatorios. Con base en estos centros de gravedad se genera la primera partición, asignándole cada individuo al núcleo más cercano. Seguidamente se calculan los centros de gravedad de cada partición, y de igual forma a como se hizo antes, se generará una nueva partición un poco diferente

---

<sup>1</sup> PIMAD, ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

a la obtenida antes. Se continuará iterando sobre estas dos operaciones, aproximando poco a poco la partición final, hasta que se llegue al punto en que no haya cambio de una iteración a la otra. Aquí se detiene el algoritmo.

Como se dijo antes, el objetivo del método de nubes dinámicas, es que las clases obtenidas sean lo más homogéneas posibles, y que estén bastante separadas unas de otras. Para poder medir el grado en que fue logrado este objetivo es que se definen los siguientes elementos.

Existen dos tipos de distancias que son abordados en este documento.

- **Métrica Euclídea:** Es la distancia clásica en espacios vectoriales y en particular en  $\mathfrak{R}^n$ . Equivale al concepto natural de distancia entre objetos del mundo real. La fórmula para esta distancia entre dos individuos  $x=(x_1, x_2, \dots, x_m)$  y  $y=(y_1, y_2, \dots, y_m)$  es:

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$$

Al utilizar la distancia euclídea, a cada individuo le es asignado el peso  $1/n$ .

- **Distancia Chi cuadrado:** Es un tipo de distancia que basa su valor en la "forma" de las filas y no en su magnitud. Es útil para analizar tendencias en el conjunto de individuos. Esta distancia no solo se basa en las coordenadas de los individuos, sino también en valores de la tabla de datos de todos los individuos analizados, a saber:

$x_{.i}$  : Suma de los elementos de la  $i$ -ésima columna de la tabla.

$x_{i.}$  : Suma de los elementos de la  $i$ -ésima fila de la tabla.

$T$  : Suma de todos los elementos de la tabla.

$$d(x, y) = \sqrt{\sum_{i=1}^n \frac{x_{.i}}{T} \left( \frac{x_i}{x_{i.}} - \frac{y_i}{y_{i.}} \right)^2}$$

Cuando se utiliza este tipo de distancia, al  $i$ -ésimo individuo se le asigna el peso  $x_{i.}$ .

Si el peso del  $i$ -ésimo individuo se denotan  $p_i$ , existe una partición formada de  $k$  clases  $C_1, \dots, C_k$ , y al peso de la  $l$ -ésima clase se le denomina como  $q_l = \sum_{i \in C_l} p_i$ , entonces el centro de gravedad del

conjunto de individuos es  $g = \sum_{i=1}^n p_i x_i$ . El centro de gravedad de la  $l$ -ésima clase se denota por

$$g_l = \frac{1}{q_l} \sum_{i \in C_l} p_i x_i$$

La inercia total mide la dispersión de los elementos alrededor de su centro de gravedad, y está definida como  $I = \sum_{i=1}^n p_i d^2(x_i, g)$ . Para medir qué tan separadas están las clases entre sí se mide la dispersión de los centros de gravedad, alrededor del centro de gravedad global. A esto se le llama inercia

inter-clases, y se escribe  $I_{inter} = \sum_{l=1}^k q_l d^2(g_l, g)$ . Análogamente, podemos medir la dispersión de la  $l$ -

ésima clase por medio de su inercia interna,  $I_{int ra}(l) = \sum_{x_i \in C_l} p_i d^2(x_i, g_l)$ , y podemos medir la homogeneidad de la partición total, por medio de las dispersiones internas de cada clase, esto se llama inercia intra-classes, y se define como la suma de las inercias internas:  $I_{int ra} = \sum_{l=1}^k I_{int ra}(l)$ .

El fin último del método de nubes dinámicas consiste en minimizar la homogeneidad de la partición, es decir, que se quiere minimizar  $I_{int ra}$ , pero por la igualdad de Fisher ( $I = I_{int er} + I_{int ra}$ ), esto es equivalente a maximizar  $I_{int er}$ , ya que  $I$  es un valor fijo para cada tabla. Es por esto que  $I_{int er}$  es un buen índice de la calidad de una partición. Pero es mejor aún si normalizamos esta cantidad dividiendo por  $I$ . Así que el objetivo del método de nubes dinámicas es buscar una partición que maximice la cantidad  $\frac{I_{int er}}{I}$ .

### Algoritmo de nubes dinámicas

#### Etapa 0:

1. Escoger al azar  $k$  individuos diferentes  $g_1^{(0)}, g_2^{(0)}, \dots, g_k^{(0)}$  de entre las  $K$  filas de la tabla.
2. para  $i := 1$  hasta  $n$  hacer:

$$\text{asignar el individuo } i \text{ al centro } g_j^{(0)} \text{ tal que } d(x_i, g_j^{(0)}) = \min_{j=1, \dots, k} [d(x_i, g_j^{(0)})]$$

así se forman  $k$  clases  $C_1^{(0)}, C_2^{(0)}, \dots, C_k^{(0)}$

3. Calcular la inercia intra-classes  $I_{int ra}^{(0)} = \sum_{l=1}^k \sum_{x_i \in C_l^{(0)}} p_i d^2(x_i, g_l^{(0)})$

#### Etapa h:

1. Calcular los centros de gravedad  $g_i^{(h)}$  de las clases  $C_i^{(h-1)}$ :  $g_i^{(h)} = \frac{1}{q_i^{(h-1)}} \sum_{x_i \in C_i^{(h-1)}} x_i$

2. para  $i := 1$  hasta  $n$  hacer:

$$\text{asignar el individuo } i \text{ al centro } g_j^{(h)} \text{ tal que } d(x_i, g_j^{(h)}) = \min_{j=1, \dots, k} [d(x_i, g_j^{(h)})]$$

así se forman  $k$  clases  $C_1^{(h)}, C_2^{(h)}, \dots, C_k^{(h)}$

3. Calcular la inercia intra-classes  $I_{int ra}^{(h)} = \sum_{l=1}^k \sum_{x_i \in C_l^{(h)}} p_i d^2(x_i, g_l^{(h)})$

**hasta** : que se cumpla una de dos situaciones:

- Se alcanza el número máximo de iteraciones
- ningún individuo cambia de clase.

### 3 El Análisis de Formas Fuertes

En el método de nubes dinámicas, la primera selección de los centros de gravedad fue hecha de manera aleatoria, por lo que el resultado puede variar de una ejecución a otra. Por esto se recomienda ejecutar el algoritmo varias veces y luego seleccionar las mejores particiones. Con base en estas particiones se puede obtener una partición más fina, en donde cada parte agrupa a los individuos que quedaron en la misma clase en todas las particiones seleccionadas. Los nuevos grupos generados de esta manera reciben el nombre de formas fuertes.

En resumen, el análisis de formas fuertes consiste en realizar una intersección de particiones para generar una partición más fina.

El conjunto de  $r$  particiones se puede representar por medio de una matriz  $Y$  de  $n \times r$ , donde la  $i$ -ésima fila representa al  $i$ -ésimo individuo y la  $j$ -ésima columna a la  $j$ -ésima partición. De esta manera el valor  $y_{ij}$  indica el número de clase en que es clasificado el  $i$ -ésimo individuo según la  $j$ -ésima partición.

A partir de  $Y$  se desea construir la matriz  $F$  de  $d \times r$ , llamada matriz de formas fuertes, donde la  $l$ -ésima fila representa a la  $l$ -ésima forma fuerte y la  $j$ -ésima columna a la  $j$ -ésima partición. Es decir que  $f_{lj}$  indica el número de clase en la  $j$ -ésima partición a la que pertenecen los individuos de la  $l$ -ésima forma fuerte.

También es útil construir un vector  $Q$  de  $n$  entradas, tal que cada entrada hace referencia al número (entre 1 y  $d$ ) de forma fuerte a la que pertenece el individuo correspondiente.

#### Algoritmo

```

i := 1, d := 1
para j := 1 hasta r hacer:
     $f_{1j} := y_{1j}$  (la primer fila de  $F$  es la primera de  $Y$ )
para i := 2 hasta n hacer:
    ya := falso
    l := 1
    Mientras (ya = falso) y (l ≤ d) hacer
        j := 1
        Mientras ( $f_{ij} = y_{ij}$ ) y (j < r+1) hacer
            j := j+1
        Si ( $f_{ir} = y_{ir}$ ) y (j = r+1) entonces
            ya := verdadero
        Sino l := l+1
    Si (ya = verdadero) entonces (fue reconocida la forma fuerte de i)
         $q_i := l$ 
    Sino d := d+1 (fue encontrada una nueva forma fuerte)
        para j := 1 hasta r hacer:
             $f_{dj} := y_{ij}$ 
         $q_i := d$ 

```

**Nota:** Observe que el algoritmo anterior es lineal en  $n$  y  $r$ .


## 4 Uso del Sistema

A continuación se presentan los pasos básicos a seguir para realizar una ejecución del programa.

### 4.1 Selección del archivo de datos.

Al principio del proceso de generar una partición se deberá proporcionar al sistema los datos iniciales sobre los que se trabajará. Estos datos se encontrarán en un archivo de texto cuya extensión será **DAT**.

El archivo está organizado como una matriz con  $n$  filas y  $m$  columnas. La primer línea del archivo consiste del encabezado, en el cual consta de los números  $n$  y  $m$  separados por un espacio. Las restantes líneas contienen la información correspondiente a la matriz. Cada línea contiene los datos de un individuo, la cual está compuesta de  $m$  campos separados por blancos o tabulaciones.

Para abrir el archivo de datos debe presionar el botón correspondiente, el cual se muestra a continuación: . También puede realizar esta operación seleccionando la opción **Archivo | Abrir (datos)** del menú del sistema.

Seguidamente se observará una ventana de selección de archivos como la que se muestra en la siguiente figura. En ella debe elegir el archivo a leer. Recuerde que este debe cumplir con las convenciones de formato descritas anteriormente.


Seleccione el archivo a procesar y presione el botón de Aceptar para continuar.



Figura 1. Ventana de abrir archivo de datos.

### 4.2 Selección de la métrica.

Después de abrir el archivo de datos, el siguiente paso es seleccionar la métrica a utilizar en el resto del proceso. Para esto presione el botón que se muestra a continuación, y que se encuentra en la

barra de herramientas: . También puede elegir la métrica mediante la opción **Opciones | Métrica** del menú del sistema. Seguidamente aparecerá la ventana de selección de métrica.

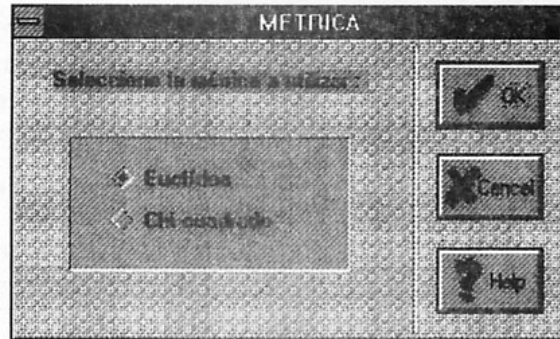



Figura 2. Ventana de selección de la métrica.

Existen dos posibles opciones para la distancia:

- Métrica Euclídea
- Distancia Chi cuadrado

### 4.3 Ejecución del Método Nubes Dinámicas.

Luego de elegir la métrica, debe proceder a ejecutar el Método de Nubes Dinámicas. Presione con el mouse el botón que se muestra seguidamente: . También puede acceder este proceso mediante la opción **Partición | Nubes Dinámicas** del menú del sistema.

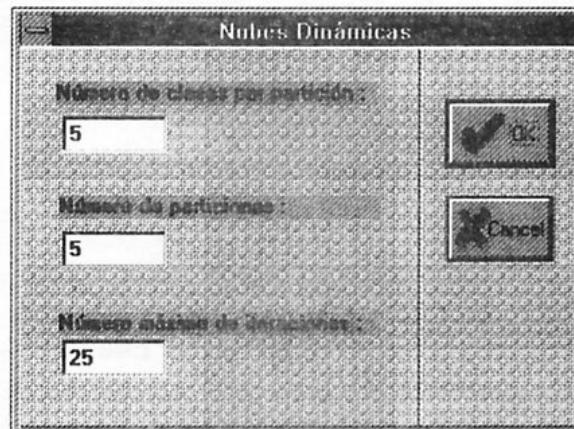


Figura 3. Ventana de Nubes Dinámicas.

En la ventana de Nubes Dinámicas encontrará tres campos que obligatoriamente deberá llenar para ejecutar el método:

**Número de clases por partición:** Aquí debe digitar el número de subconjuntos en que será clasificado el conjunto de curvas de carga en cada ejecución de Nubes Dinámicas. Debe ser un valor entre 1 y 16.

**Número de particiones:** El proceso de Nubes Dinámicas es ejecutado varias veces. Esto es necesario ya que en cada ejecución del método se elige una partición inicial al azar, sobre la que se trabaja. Con correr el proceso varias veces se pretende eliminar el efecto de aleatoriedad de esa partición inicial.


**Número máximo de iteraciones:** El proceso de Nubes Dinámicas es iterativo, y no es fácil predecir cuántas iteraciones realizará o cuánto durará. Es por esta razón que se acota el número de iteraciones de cada corrida con un valor máximo. Este ítem tiene un valor predeterminado de 25.

Una vez que el programa lee y acepta los datos anteriores, se activará un proceso que clasifica el conjunto de curvas de carga seleccionado anteriormente con base en la métrica elegida.

Cuando finalice el proceso, se presentará una tabla que muestra la inercia explicada de cada una de las particiones realizadas. La inercia explicada es un valor entre 0 y 1 que es calculada mediante la fórmula:  $\text{Inercia Inter} / \text{Inercia Total}$ . Este índice indica la calidad de la clasificación. Un valor cercano a 0 indica que los subconjuntos de individuos son muy dispersos y que la clasificación es de mala calidad, mientras que uno cercano a 1, que los individuos se reparten en grupos compactos y bien diferenciados, lo cual corresponde a una partición de buena calidad.

Usted debe interpretar esa información y tomar una decisión: ¿cuántas de las particiones recién creadas se deben tomar en cuenta en el Análisis de Formas Fuertes?. La respuesta puede variar según el número de particiones, la cantidad de clases por partición y los objetivos del análisis. Si no tiene idea o no está seguro puede ejecutar Formas Fuertes varias veces seguidas hasta quedar satisfecho con los resultados.

#### 4.4 Realización el Análisis de Formas Fuertes.

Elija la opción **Partición | Formas Fuertes** del menú del sistema o presione el botón que se muestra a continuación: .

En la ventana de Formas Fuertes debe digitar el número de particiones obtenidas de Nubes Dinámicas, que serán utilizadas en el proceso de Formas Fuertes. Este valor debería ser obtenido con base en la información de la tabla anterior. En el ejemplo de la siguiente figura fueron seleccionadas 4 particiones.

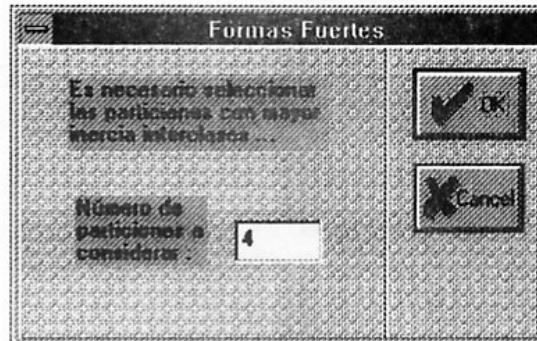



Figura 4. Ventana de Formas Fuertes.

El programa automáticamente tomará las mejores particiones, basándose en su inercia explicada, y con base en ellas obtendrá una partición final.

Seguidamente se presenta una tabla en que se especifica el número de clases y el porcentaje de individuos de la población total que pertenecen a cada clase. Si la partición final recién creada no le satisface, puede volver a correr Formas Fuertes con una cantidad mayor o menor de particiones procedentes de Nubes Dinámicas.

#### 4.5 Generación de los archivos de la partición generada.

Cuando Formas Fuertes genere una partición final adecuada, y se desee guardar estos datos en archivos, se debe elegir la opción **Archivo | Guardar (particiones)** del menú del sistema o presione con

el mouse el botón que se muestra seguidamente: .

A continuación podrá ver la ventana de guardar archivo de datos, como se muestra en la Figura 5. En esta debe seleccionar o digitar el nombre del archivo en que serán guardados los datos finales del proceso. La extensión del archivo debe ser **DAT**.

El sistema guardará cuatro archivos, con el mismo nombre, pero con diferente extensión:

- **<archivo>.DAT** es un archivo de texto que contiene una tabla con el mismo formato que la tabla que se leyó al iniciar el proceso, y que contiene la información de los centros de gravedad de las clases de la partición final.
- **<archivo>.CAR** es un archivo de texto con formato de vector, en que se indican las cardinalidades de las clases de la partición final. Este es el archivo de cardinalidades.
- **<archivo>.ENG** es un archivo de texto con formato de vector, en que se indica la energía de cada clase. La energía es lo que antes fue definido como  $x_i$ .
- **<archivo>.PAR** es el archivo de texto con formato de vector en que se guarda la partición final. Este corresponde al vector  $Q$  del algoritmo de formas fuertes.



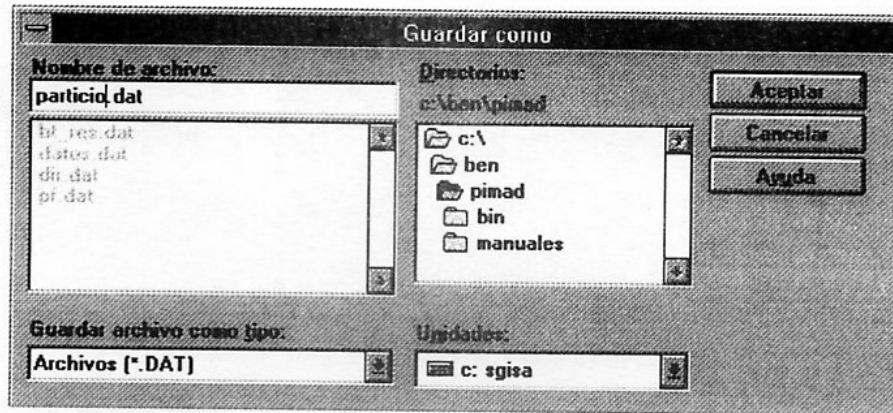




Figura 5. Ventana de guardar archivo de partición.

Presionando en cualquier momento el botón  podrá tener acceso al índice de ayuda del sistema. Desde ahí podrá navegar por las pantallas de ayuda e ir conociendo más a fondo las diferentes opciones del programa.

Finalmente puede salir del sistema, ya sea presionando con el mouse el botón , o por medio de la opción **Archivo | Salir** del menú del sistema.

## Bibliografía

- Benzécri, J. P. (1984) *L'Analyse des Données. Tome 1: La Taxonomie*. 4a edición. Dunod, Paris.
- Castillo, W. (1989) "Métodos y resultados en clasificación automática", *Revista Ciencia y Tecnología*, XIII (1-2): 105-116.
- Castillo, W.; Rodríguez, O.; Trejos J.; Chacón B. (1996) "Clasificación Jerárquica Ascendente según el método de Ward", documento del proyecto *Construcción de tipologías de curvas de carga del sistema eléctrico: metodología matemática y diseño de software*. Universidad de Costa Rica - Instituto Costarricense de Electricidad, San Pedro.
- Celeux, G. (1987) *Les formes fortes*. Support de cours, Ecole Modulad "Méthodes de Classification pour l'Analyse des Données", Strasbourg, pp. 49-56.
- Celeux, G.; Diday, E.; Govaert, G.; Lechevallier, Y.; Ralambondrainy, H. (1989) *Classification Automatique des Données. Environnement Statistique et Informatique*. Dunod-Informatique, Paris.
- Diday, E. et collaborateurs (1980) *Optimisation en Classification Automatique*. INRIA, Le Chesnay.
- Diday, E.; Lemaire, J.; Pouget, J.; Testu, F. (1982) *Eléments d'Analyse des Données*. Dunod, Paris.
- Espinoza, J.L.; Trejos J. (1989) "Clasificación por particiones", *Revista Ciencia y Tecnología*, XIII (1-2): 129-154.
- Lerman, I.C. (1981) *Classification et Analyse Ordinale des Données*. Dunod, Paris.
- Roux, M. (1985) *Algorithmes de Classification*. Masson, Paris.

# BÚSQUEDA TABÚ: METAPROCEDIMIENTO DE MEJORA PARA EL PROBLEMA DE INVENTARIOS MULTIPRODUCTO

SERGIO DE LOS COBOS<sup>1</sup> - BLANCA ROSA PÉREZ<sup>1</sup>  
MIGUEL ÁNGEL GUTIÉRREZ<sup>2</sup> - MANUEL ORDORICA<sup>3</sup>

---

## Resumen

En este trabajo se presenta una nueva propuesta de la Búsqueda Tabú como meta-procedimiento de mejora, para el problema de inventarios multiproducto con demanda determinística y costos de reaprovisionamiento conjunto, que, al aplicarse obtiene mejores resultados con respecto a métodos reportados como altamente eficientes. Se presenta la experiencia computacional.

**Palabras clave:** sistemas de inventario multiproducto con demanda determinística y costos de reaprovisionamiento conjunto, Búsqueda Tabú.

## 1 Introducción

En la actualidad son grandes los retos a los que se debe de enfrentar las empresas, entre los que destacan la competencia creciente, las dificultades de abastecimiento, la sobrea-bundancia de artículos y la escases de materia prima, por lo que es de importancia cada día mayor la adecuada administración económica de los inventarios.

Una forma de asegurar la continuidad de las operaciones es el almacenamiento, sin embargo, esto desencadena costos suplementarios, lo que por efecto ocasiona una reducción del margen de utilidad, por otro lado, la falta de inventarios adecuados puede propiciar un rompimiento en la continuidad de operación y altos costos de oportunidad por no poder satisfacer la demanda.

---

<sup>1</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA, DEPARTAMENTO DE MATEMÁTICAS, AV. PURÍSIMA Y MICHOACÁN, DEL. IZTAPALAPA, C.P. 09340, MÉXICO D.F.; E-MAIL: COBOS@XANUM.UAM.MX.

<sup>2</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - AZCAPOTZALCO, DEPARTAMENTO DE SISTEMAS, SAN PABLO 180, DEL. AZCAPOTZALCO, C.P. 02200, MÉXICO D.F.

<sup>3</sup>EL COLEGIO DE MÉXICO, CAMINO AL AJUSCO 20, DEL. ALVARO OBREGÓN, C.P. 10740, MÉXICO D.F., MÉXICO

En este trabajo se desarrolló para caso de los los modelos de inventario multiproducto con demanda determinística y con costos de reaprovisionamiento conjunto, un algoritmo de mejora original, que utiliza un esquema de la búsqueda tabú para este tipo de problemas, además proporciona la mayor eficiencia en cuanto a las soluciones obtenidas con respecto a los algoritmos reportados como de los más eficientes. Se reporta en la experiencia computacional, la eficiencia de este nuevo marco de solución propuesto como la mayor en todas las instancias consideradas, con respecto de los demás algoritmos probados.

El trabajo se desarrolla como sigue: en la Sección 1, se describe el modelo de inventarios multiproducto para el caso de demanda determinística, así como algunos de los algoritmos que resuelven el problema con eficiencia; en la Sección 2, se introduce a la técnica de la Búsqueda Tabú y se describen algunos métodos de manejo de la lista tabú de tipo dinámico, en la Sección 3 se propone a la Búsqueda Tabú como un metaprocedimiento de mejoramiento para el problema de inventarios propuesto. Finalmente en la Sección 4, se presenta la experiencia computacional.

## 2 Inventarios multiproducto: caso determinístico

El problema de inventarios multiproductos consiste en general de determinar cuánto y cuándo ordenar de cada artículo de los que se controlan, para que, la demanda sea satisfecha con costos mínimos. En este trabajo se supone que la entrega por parte de los proveedores es inmediata y se realiza en una sola entrega, no se permite déficit y los artículos nunca llegan a ser obsoletos una vez que se inventarean. Además se supone que diferentes artículos de un mismo proveedor se pueden entregar en un solo paquete y se realiza a intervalos regulares de tiempo, digamos  $t_1$ .

El problema para  $n$  artículos por tanto se traduce en determinar la frecuencia de los ciclos de abasto tanto individual como conjunto, por lo que se puede observar fácilmente que el cociente de la frecuencia individual del artículo  $j$  digamos  $t_j$  entre la frecuencia conjunta es un entero positivo  $\beta_j$  mayor o igual a uno, i.e.,  $t_j = \beta_j t_1$ ,  $j = 1, 2, \dots, n$

Los artículos tienen diferentes ciclos, pero el que marca las oportunidades para ordenar es el menor duración  $t_1$ . Se ordena tantas veces como lo requiera dicho artículo en una unidad de tiempo,  $1/t_j$ , y cada vez que se ordena se incluyen en la orden los artículos cuyo ciclo así lo indica.

Los costos que intervienen en el modelo son de: ordenación, adquisición e inventario. Un costo fijo  $K$  se carga cada vez que se realiza una orden. Un costo por línea  $k_j$  se produce cada vez que se carga el artículo  $j$  en la orden correspondiente.

El costo debido al inventario se carga sobre el inventario promedio durante un ciclo y está dado por:

$$\frac{1}{2} \sum_{j=1}^n h_j d_j t_j, \quad (1)$$

donde:  $h_j$ ,  $d_j$ , y  $t_j$  son el inventario promedio, la demanda promedio y el tiempo del ciclo del artículo  $j$ , respectivamente.

De lo anterior se tiene que el costo total por unidad de tiempo es:

$$CT = \frac{1}{t_1} \left( K + \sum_{j=1}^n \frac{k_j}{\beta_j} + \sum_{j=1}^n c_j d_j \right) + \frac{1}{2} t_1 \sum_{j=1}^n h_j d_j \beta_j, \quad (2)$$

el cual se desea minimizar. El costo total no es sólo función de  $t_1$ , sino también de las  $\beta_j$ , para  $j = 1, 2, \dots, n$  y donde las  $c_j$ ,  $j = 1, 2, \dots, n$  son los costos unitarios.

Narro y de los Cobos(1994) proponen aproximar los valores de  $\beta_j$   $j = 1, 2, \dots, n$  y después encontrar el valor de  $t_1$  que minimiza a  $CT(t_1)$ , de la siguiente manera:

El costo correspondiente a cada artículo ordenado se puede expresar como:  $\alpha_j K$ , con  $\alpha_j \geq 0$ ,  $\sum_{j=1}^n \alpha_j = 1$ . Si la política óptima para cada artículo  $j$ , considerado como independiente, aconseja ordenar  $x_j$  unidades en cierto instante, entonces se tiene que:

$$K \delta \left( \sum_{j=1}^n x_j \right) + \sum_{j=1}^n k_j \delta(x_j) \geq \sum_{j=1}^n (\alpha_j K + k_j) \delta(x_j), \quad (3)$$

donde  $\delta(x) = 1$  si  $x > 0$ , y cero de otra forma.

La política óptima es una política factible para el sistema de inventario del artículo  $j$  con costo por ordenar  $(\alpha_j K + k_j)$ . Si el costo mínimo correspondiente es  $(CT)_j(\alpha_j)$ , una cota inferior para el costo de la política óptima completa está dada por:  $\sum_{j=1}^n (CT)_j(\alpha_j)$ , para cualquier asignación de  $\alpha_i$   $i = 1, 2, \dots, n$ , por lo que la mejor cota inferior está dada por:

$$\text{Minimizar}_{\alpha_i} \sum_i (CT)_i(\alpha_i) \quad (4)$$

$$\text{s.a. } \alpha_i \geq 0, \quad i = 1, 2, \dots, n, \quad \sum_{i=1}^n \alpha_i = 1, \quad (5)$$

el cual es un problema de programación no lineal entero mixto. Ahora bien, las formas en que se calculan los ciclos y se aproximan los múltiplos son diversas. Si  $l$  es el real que resulta del cociente del tiempo de duración del ciclo correspondiente y el de menor longitud y  $\beta$  es el entero más próximo, entonces se debe satisfacer que:  $\beta(\beta + 1) < l^2 < (\beta + 1)l$ , en particular:

$$l_j'^2 = \frac{k_j}{d_j h_j} \max \left\{ \frac{d_i h_i}{K + k_i}, i = 1, 2, \dots, n \right\}, \quad (6)$$

$$\beta_j'^2 = \frac{1 + \sqrt{1 + 4l_j'^2}}{2}, \quad (7)$$

$$t_1' = \sqrt{\frac{2 \left( K + \sum_{i=1}^n \frac{k_i}{\beta_i'} \right)}{\sum_{i=1}^n h_i d_i \beta_i'}}, \quad (8)$$

$$l_j^2 = \frac{2(\alpha_j K + k_j)}{\frac{d_j h_j}{t_1'^2}}, \quad (9)$$

$$\beta_j^2 = \frac{1 + \sqrt{1 + 4l_j^2}}{2}, \quad (10)$$

por lo que se desea resolver:

$$\text{Minimizar } \frac{1}{t_1} \left( K + \sum_{j=1}^n \frac{k_j}{\beta_j} + \sum_{j=1}^n c_j d_j \right) + \frac{1}{2} t_1 \sum_{j=1}^n h_j d_j \beta_j \quad \text{s.a. } t_1 \geq 0. \quad (11)$$

### 3 Búsqueda Tabú

La Búsqueda Tabú (BT) es un procedimiento heurístico de “alto nivel” introducido y desarrollado en su forma actual por Fred Glover (1989) y (1990a), el cual se utiliza con gran éxito para resolver problemas de optimización cuya característica principal es la de “escapar” de la optimalidad local. Para una lista de aplicaciones véase por ejemplo Glover y Laguna (1993).

Para un ejemplo numérico completo donde se ejemplifican diferentes aspectos de la BT, ver de los Cobos(1994). En términos generales, el método BT puede esbozarse consistente en:

1. El paso básico del procedimiento consiste en empezar desde un punto factible  $s$  y generar un conjunto de soluciones en  $N(s)$ ; entonces se escoge al mejor vecino generado  $s^*$  y se posiciona en ese nuevo punto ya sea que  $C(s^*)$  tenga o no mejor valor que  $C(s)$ .

2. La característica importante de la búsqueda tabú es precisamente la construcción de una lista tabú  $T$  de movimientos: aquellos movimientos que no son permitidos (movimientos tabú) en la iteración presente. La razón de esta lista es la de excluir los movimientos que nos pueden regresar a algún punto de una iteración anterior.

Sin embargo, la eliminación de ciclos no es la última meta en el proceso de búsqueda. En algunos casos, una buena trayectoria de búsqueda resultará al visitar una solución encontrada antes. El objetivo de manera amplia es el de continuar estimulando el descubrimiento de nuevas soluciones de alta calidad.

3. Las restricciones tabú no son inviolables bajo toda circunstancia. Cuando un movimiento tabú proporciona una solución mejor que cualquier otra encontrada, su clasificación tabú puede eliminarse. La condición que permite dicha eliminación se llama *criterio de aspiración*.

4. Ahora bien, conforme la búsqueda progresa, la forma de la evaluación empleada por la búsqueda tabú llega a ser más adaptativa, incorporando referencias concernientes para la *intensificación* y la *diversificación* regional de búsqueda. Cabe aclarar que en las estrategias basadas en consideraciones de término corto la clasificación tabú sirve para identificar elementos de la vecindad del movimiento actual, mientras que en las estrategias de término intermedio y largo pueden no contener soluciones en esta vecindad, sino que por lo general consisten de seleccionar soluciones *élites* (óptimos locales de alta calidad) encontrados en varios puntos en el proceso de solución. Dichas soluciones élites se identifican como elementos de un conglomerado regional en las estrategias de intensificación regional, y como elementos de diferentes conglomerados en las estrategias de diversificación regional.

### 3.1 Manejo dinámico de la lista tabú

El *manejo de la lista tabú* significa la actualización de ésta, *i.e.*, la decisión de cuántos y cuáles movimientos se pondrán como tabú para una iteración de la búsqueda. La mayoría de las implantaciones de la búsqueda tabú utilizan un manejo de la lista tabú de tipo estático. De manera más precisa, los movimientos permanecen como tabú durante un número de iteraciones, por lo que la eficiencia del algoritmo depende de la elección de la duración del estatus tabú o, equivalentemente, de la longitud de esta lista.

Este manejo de la lista tabú de tipo dinámico permite el exámen más detallado de la región factible, por lo que es posible romper ciclos y diversificar la búsqueda.

Existen varios métodos de lista tabú dinámica, entre los que destacan: el método de la *sucesión (secuencia) de Cancelación* (CSM) y el método de *eliminación inversa* (REM) propuestos en Glover (1990a), los que a continuación se presentan.

#### 3.1.1 Método de la sucesión de cancelación

Si  $LT$  es la lista tabú que implícitamente define al conjunto de movimientos tabú. Como punto inicial, considérese a  $LT$  dada de la siguiente manera:  $LT = (e(1), e(2), \dots, e(q))$ . Supongamos que los elementos de  $LT$  son indexados por la iteración, identificando el punto en el cual fue adicionado a la lista, y  $q$  es el índice de la iteración actual. Correspondientemente se puede identificar a la lista de soluciones:  $(x(1), x(2), \dots, x(q))$  tal que, para cada  $i$ ,  $e(i)$  es el atributo asociado con el movimiento aplicado a  $x(i)$ .

Se asocia con  $LT$ , una *lista tabú activa*,  $LTA$ , la cual consiste sólo de los elementos de  $LT$  que no han sido cancelados, por lo que  $LTA$  es una subsucesión de  $LT$  que contiene el mismo último elemento y que se puede describir como:

$$LTA = (e(p), \dots, e(h), e(i), e(j), \dots, e(q), e(q+1)).$$

Si la adición de  $e(q+1)$  constituye la primera vez que algún elemento cancela a un elemento previo, entonces por la propiedad de suficiencia se tiene que ninguna de cualesquiera dos soluciones generadas alejadamente pueden ser la misma. Más aún, la solución  $x(i)$  no puede duplicarse por la solución de cualquier sucesión de movimientos futuros a menos de que cada elemento de la sucesión  $e(j), \dots, e(q)$  esté cancelado.

A esta sucesión que cae entre el cancelamiento del elemento  $e(q+1)$  y el elemento cancelado  $e(i)$  se le llama la *sucesión de cancelación* o *C-sucesión*. Dado que se previene la cancelación de  $e(q)$  por  $e(q+1)$ , se puede asegurar que el último elemento en cada *C-sucesión* sucesiva permanecerá sin cancelar, por lo que ninguna solución puede repetirse. Esto es una condición suficiente pero no necesaria para eliminar repetición de soluciones.

El segundo proceso dinámico para manejar listas tabú está basado en la generación de restricciones de movimientos implicadas de manera general por las *C-sucesiones* con cotas que se pueden tanto expandir como contraer. Lo apropiado de expandir las cotas de una *C-sucesión* proviene del hecho de que la adición de un nuevo elemento a la  $LTA$  proteje contra el regreso a una solución "guardada" por otras *C-sucesiones* y así pueda incluirse entre esas sucesiones. Existe un obstáculo para realizar esto dado que el nuevo elemento es no adyacente a los elementos previos de las *C-sucesiones* y por lo general no existe manera de rearrreglar a los elementos y a las cotas, tal que, esas *C-sucesiones* se

puedan definir por los tipos convenientes de estructuras de datos que se introducen para manejar a las *C-sucesiones* ordinarias.

El marco para sobrepasar este obstáculo opera mediante la eliminación sucesiva de los elementos de la lista *LT* por una secuencia inversa y no por una secuencia para mantener a *LTA* actualizada en cada iteración, por lo que el método se denomina de *eliminación inversa*. El estatus tabú asignado por REM representa un criterio necesario y suficiente para prevenir el visitar soluciones conocidas, como se explica en Glover (1990a) y en Dammeyer y Vo $\beta$  (1993).

### 3.1.2 Método de eliminación inversa

Sin un control adicional, un proceso de búsqueda puede visitar en la siguiente iteración una solución óptima local al moverse en una vecindad de ésta. Considere que los atributos de todos los movimientos se almacenan en una *lista de corridas*, *i.e.*, una lista que representa las trayectorias de las soluciones que se han encontrado. Por tanto, para evitar el ciclado, la lista de corridas se rastreará hacia atrás para determinar todos los movimientos que se considerarán tabú para evitar una solución ya explorada. Para este propósito una secuencia de cancelación residual (RCS) se contruye a pasos mediante una traza. (El rastreo hacia atrás de la lista de corridas se le llamará traza). En cada paso de una traza (paso de traza), exactamente un atributo se procesa y se empieza con el último (más reciente) hacia el primero.

Se inicia con una RCS vacía, sólo se adicionan aquellos atributos cuyo complemento no está en la sucesión, de otra forma, sus complementos en RCS se eliminan, por lo que en cada paso de traza se conocen qué atributos pueden invertirse para que la solución actual regrese a alguna ya examinada durante el proceso de búsqueda.

Si los atributos restantes en la RCS pueden invertirse mediante exactamente un movimiento, entonces ese movimiento es tabú en la siguiente iteración, puesto que de otra forma este movimiento producirá una solución anterior. La Tabla 1 proporciona un ejemplo para construir una sucesión de cancelación residual.

Como se había comentado, la asignación del estatus tabú mediante REM es un criterio necesario y suficiente para prevenir visitar soluciones conocidas.

Lista de Corridas: 1 2 6 4 1 5 2 3 2 5 (último mvto. 5) iteración 10				
Posición	Paso de Traza	RCS	long.	mvto. tabú
10	1	5	1	5
9	2	$\tilde{2}$ 5	2	-
8	3	3 $\tilde{2}$ 5	2	-
7	4	3 5	2	-
6	5	3	1	$\tilde{3}$
5	6	1 3	2	-
4	7	4 1 3	3	-
3	8	6 4 1 3	4	-
2	9	$\tilde{2}$ 6 4 1 3	5	-
1	10	$\tilde{2}$ 6 4 3	4	-

Tabla 1. Sucesión de Cancelación Residual (RCS).

Suponga que en la iteración  $i$  del procedimiento de búsqueda el  $k$ -ésimo paso de traza proporciona una RCS de longitud  $l$ . Entonces, en la iteración  $i + 1$  la longitud de la RCS en el paso de traza  $k + 1$  es igual a  $l - 1$  o a  $l + 1$ . Para mayores detalles véase a a Glover y a Dammeyer y Vo $\beta$ .

### 3.1.3 Adaptación para problemas enteros de variable acotada

En Glover (1990a), se puede observar que los atributos de movimiento que satisfacen las condiciones necesarias y suficientes descritas anteriormente, resultan un medio natural para crear un atributo para representar el incremento o decremento de cada variable entera para un valor específico. Pero esto produce  $2U(j)$  atributos para cada variable entera  $x_j$ , donde  $U(j)$  es la cota superior de la  $j$ -ésima variable, por lo que se puede requerir de arreglos de dimensión muy grande.

Glover propone una estructura de datos alternativa que sólo requiere de un espacio para los arreglos de  $3m$ , donde  $m$  es el número total de variables.

Lo anterior se basa en el hecho de que no es la sucesión de los elementos en la C-sucesión lo importante, sino su identificación. Entonces los arreglos *predecesor* y *sucesor* sirven sólo como un medio conveniente para poder trazar esta identidad.

Para trazar la información en el caso general de cota superior, se utiliza un arreglo *desviación* el cual identifica al vector diferencia  $x(i) - x(q)$ , i.e.,  $desviación(j) = x_j(i) - x_j(q)$ . Se inicia con todas las entradas del arreglo *desviación* iguales a cero. Cada  $e(h)$  encontrado durante la traza de  $e(q)$  a  $e(i)$  en la lista tabú registra la información de que una  $x_j$  particular fue incrementada o decrementada, y sobre la base de esta información el correspondiente valor de *desviación(j)* se cambia en dirección opuesta. Sea  $n$  el valor de la suma de los valores absolutos de las entradas del arreglo *desviación*. Entonces cuando  $n=1$ , el único elemento unido por los arreglos *predecesor* y *sucesor* es el de índice  $j$  tal que  $desviación(j) = 1$  ó  $-1$ , indicando que la  $x_j$  apropiada es tabú para incrementarse o decrementarse respectivamente.

## 4 Marco de mejoramiento por Búsqueda Tabú

Como se ha visto, el método de la BT empieza con una solución factible inicial y en el proceso de ejecución, el procedimiento actualiza a los arreglos y elementos de la función memoria. Entonces el proceso se repite hasta que el criterio de terminación se encuentra.

A fin de iniciar la búsqueda para mejorar los planes maestros de los pedidos a proveedores una orden inicial se requiere. Para utilizar a la búsqueda tabú como un meta-procedimiento de mejora, en lugar de utilizar valores fijos para las  $\beta_j$ ,  $j = 1, 2, \dots, n$ , se propone que cada  $\beta$  esté dentro de un rango de valores dado por:

$$\beta_j \in [1, 2\beta_j^*], \text{ donde } \beta_j^* = \left\lceil \sqrt{\frac{1 + \sqrt{1 + 4l_j^2}}{2}} \right\rceil \text{ para } j = 1, 2, \dots, n,$$

donde  $\lceil x \rceil$  es el entero mayor más cercano de  $x$ .



Por tanto, lo que se desea es resolver:

$$\text{Minimizar } \frac{1}{t_1} \left( K + \sum_{j=1}^n \frac{k_j}{\beta_j} + \sum_{j=1}^n c_j d_j \right) + \frac{1}{2} t_1 \sum_{j=1}^n h_j d_j \beta_j,$$

$$\text{s.a. } t_1 \geq 0, \quad \beta_j^* \in [1, 2\beta_j^*], \quad j = 1, 2, \dots, n,$$

el cual es un problema de programación no lineal entero mixto, pero como se observa la función objetivo resulta ser una función convexa en  $t_1$ , por lo que, una vez fijos los valores de  $\beta_j$   $j = 1, 2, \dots, n$ , el valor óptimo para  $t_1$  se encuentra fácilmente.

Consideraremos que una solución es de la forma:

$$\pi = (\pi(1), \pi(2), \dots, \pi(n)) \text{ donde } \pi(i) \in [1, 2\beta_i^*], \quad i = 1, 2, \dots, n,$$

entonces, definimos un movimiento de la forma  $\pi' = \pi \pm e_j$ ,  $j = 1, 2, \dots, n$ , tal que  $\pi'(i) \in [1, 2\beta_i^*]$ ,  $i = 1, 2, \dots, n$ . En una implantación sencilla de BT, la lista tabú se mantiene en forma implícita registrando sólo el índice  $j$  de la variable entera cuyo valor ha cambiado en la iteración, utilizando un manejo dinámico de la lista tabú.

## 5 Resultados computacionales

Esta adaptación probó su robustez al encontrar mejores soluciones respecto a las de algoritmos eficientes de la literatura, en particular se compararon con el algoritmo de Kaspi y Rosenblatt(1985).

n		KyR		BT	
		Costo	Tiempo	Costo	Tiempo
5	min.	4908	0.0297	4908	0.0297
	max.	6928	0.0183	6928	0.0183
	prom.	5805	0.0237	5805	0.0237
10	min.	8743	0.0216	8716	0.0213
	max.	10125	0.0174	10064	0.0171
	prom.	9632	0.0179	9602	0.0177
15	min.	7992	0.0201	7896	0.0207
	max.	11993	0.0144	11993	0.0144
	prom.	10448	0.0158	10446	0.0160
20	min.	12907	0.0134	12864	0.0129
	max.	15767	0.0100	15732	0.0099
	prom.	14650	0.0122	14609	0.0118
25	min.	13497	0.0188	13346	0.0172
	max.	17361	0.0131	17309	0.0127
	prom.	15035	0.0166	14958	0.0158
30	min.	18593	0.0137	18354	0.0123
	max.	21038	0.0128	20679	0.0106
	prom.	19947	0.0132	19615	0.0111

Tabla 2. Valores generados por 100 instancias aleatorias.

Para probar la robustez de la eficiencia del algoritmo propuesto en este trabajo se generaron 100 instancias aleatorias para cada uno de los métodos, con tamaños  $n = 5, 10, 15, 20, 25$  y  $30$ , y con los siguientes rangos de las variables:

$$2000 \leq d_j \leq 50000, d_j \text{ entero}, j = 1, 2, \dots, n,$$

$$1 \leq h_j \leq 8, j = 1, 2, \dots, n,$$

$$1 \leq k_j \leq 12, j = 1, 2, \dots, n,$$

los resultados se proporcionan en la Tabla 2.

La eficiencia se calculó tomando como base el valor  $z_{tabu}$ , esto es, el valor de la función objetivo obtenido por el algoritmo propuesto en este trabajo y se utilizó la ecuación:

$$eficiencia = 1 - \frac{z - z_{tabu}}{z_{tabu}}.$$

Los resultados se proporcionan en la Tabla 3. Se puede observar de la Tabla 3, que la eficiencia del algoritmo propuesto utilizando un marco de búsqueda tabú es la más elevada. El número de iteraciones bajo el marco de búsqueda tabú que se realizaron fueron de  $2n$  donde  $n$  es el número de artículos considerados. Cabe señalar que en el estudio de Kaspi y Rosenblatt compararon la efectividad de varios algoritmos generando 132000 ejemplos y cubriendo 132 clases diferentes.

n		KyR	BT
5	min.	1	1
	max.	1	1
	prom.	1	1
10	min.	0.996	1
	max.	0.993	1
	prom.	0.996	1
15	min.	0.999	1
	max.	1	1
	prom.	0.999	1
20	min.	0.996	1
	max.	0.997	1
	prom.	0.997	1
25	min.	0.988	1
	max.	0.996	1
	prom.	0.994	1
30	min.	0.986	1
	max.	0.982	1
	prom.	0.983	1

Tabla 3. Eficiencia Comparativa.

## 6 Conclusiones

Como se puede observar, la eficiencia del algoritmo propuesto, es mayor en todas las instancias respecto a algoritmos especializados. Aún con esquemas sencillos de la búsqueda

tabú, se obtienen excelentes resultados, pero aún existen por explorar muchas otras adaptaciones del mismo.

## Bibliografía

- [1] Dammeyer F. and Voß S. (1993) “Dynamic Tabu List Management Using the Reverse Elimination Method”, *Annal of Operations Researc* **41**: 31–46.
- [2] de los Cobos Silva S.(1994) *La Técnica de la Búsqueda Tabú y sus Aplicaciones*. (julio) Tesis Doctoral, Division de Estudios de Posgrado, U.N.A.M., México.
- [3] de los Cobos Silva S. G., Pérez Salvador B.R.,y Gutierrez Andrade. M.A. (1995) “Lineamientos de implantación de la búsqueda tabú para los problemas de calendari-zación”, *IX Simposio Métodos Matemáticos Aplicados a las Ciencias*, J. Trejos (ed.) (Turrialba, Costa Rica), pp. 23–31.
- [4] Glover F. (1989) “Tabu Search, Part I”, *ORSA Journal on Computing* **1**(3): 190–206.
- [5] Glover F. (1990a) “Tabu Search, Part II”, *ORSA Journal on Computing* **2**(1): 4–31.
- [6] Kaspi M. and Rodenblatt M.J. (1985) “The Effectiveness of Heuristic Algorithm for Multi-Item Inventory Systems with Joint Replenishment Costs”, *International Journal of Production Research* **23**: 109–116.
- [7] Murillo F. A.(1996) “Particionamiento usando Búsqueda Tabú”, *IV Encuentro Centroamericano de Investigación Matemática*, Antigua, Guatemala.
- [8] Narro Ramírez A.E. and De los Cobos Silva S.G.(1994) “Algoritms for Multi-item Inventory Systems”, *15th International Symposium on Mathematical Programming*, Ann Arbor, Michigan, U.S.A.

# CONTROLABILIDAD DIRECCIONAL PARA SISTEMAS NO LINEALES EN EL PLANO A TRAVÉS DE UN PROCESO DE LINEALIZACIÓN<sup>1</sup>

VÍCTOR DELGADO<sup>2</sup>

---

## Resumen

Para un sistema de control no lineal en el plano se plantea el problema de determinar condiciones que permitan la factibilidad de perturbar la dinámica a través de controles realimentados cuyas respectivas trayectorias se acerquen a un punto estacionario y, eventualmente, lo alcancen en tiempo finito. Tal proceso se realiza a través de una linealización del sistema.

**Palabras Claves:** Sistemas de control linealmente equivalentes; controlabilidad direccional, modelo poblacional.

## 1 Introducción

Para un sistema de control no lineal en el plano:

$$\begin{aligned}x' &= f(x, y) \\y' &= g(x, y) + u\end{aligned}\tag{1}$$

donde  $f$  y  $g$  son funciones de clase  $C^1$  en torno a un punto crítico de  $F(f, g)$ , se plantea el problema de determinar condiciones sobre  $f, g, u$  que permitan la factibilidad de una perturbación del sistema (1) a través de controles admisibles cuyas respectivas trayectorias se acerquen al punto crítico y, eventualmente, lo alcancen en tiempo finito.

El camino natural hacia tal objetivo es considerar una linealización del sistema (1) en torno al punto crítico. Si tal proceso origina un sistema lineal controlable, este puede transformarse en otro linealmente equivalente y para el cual sea posible determinar controles cuyas respectivas trayectorias se aproximen rectilíneamente a un objetivo prefijado (controlabilidad direccional).

---

<sup>1</sup>Proyecto S-96-08 Dirección de investigación UACH Valdivia, Chile

<sup>2</sup>INSTITUTO DE MATEMÁTICAS, UACH VALDIVIA, CHILE; E-MAIL: VDELGADO@VALDIVIA.UCA.UACH.CL

## 2 Resultados

Recientemente hemos obtenido un par de resultados cuya interpretación en el plano nos proporciona información útil a nuestro propósito. Específicamente:

1. Un sistema de control  $x' = A * x + B * u$  en  $\mathbb{R}^n$  es linealmente equivalente al sistema  $x' = Ax + Bu$  si existe una matriz real y no singular  $P$  tal que  $A* = PAP^{-1}$  y  $B* = PB$  ([5], pág. 83).

Si el rango de la matriz  $( B \ AB \ \dots \ A^{n-1}B )$  es  $n$ , entonces el sistema  $x' = Ax + Bu$  es controlable en el sentido que dada una condición inicial  $x_0$  y un objetivo de llegada  $x_1$ , existe un control admisible cuya respectiva trayectoria transfiere  $x_0$  a  $x_1$  en tiempo finito.

En lo que sigue  $A^t$  indica la matriz transpuesta de  $A$  y para una matriz cuadrada  $A = (a_{ij})$ , la notación  $a_{ij}^{(k)}$  denota los elementos de  $A^k$ , con  $A^0$  la matriz identidad. Además para naturales  $a$  y  $b$  menores o iguales a  $n$  se define:

$$\begin{aligned} a * b &= n \text{ si } a = b \\ a * b &= n - b \text{ si } a \neq b \end{aligned}$$

**Teorema 1.** Si para el sistema de control  $x' = Ax + bu$ , con  $A = (a_{ij})$  de orden  $n$  y  $b = (0 \ \dots \ 1)^t$ , el rango de la matriz  $P = (A^{n-1}b \ \dots \ Ab \ b)$  vale  $n$ , entonces es linealmente equivalente al siguiente:

$$\begin{aligned} x'_1 &= x_2 \\ &\vdots \\ x'_{n-1} &= x_n \\ x'_n &= a_n x_1 + \dots + a_1 x_n + u \end{aligned} \tag{2}$$

donde para cada  $p = 1, \dots, n$  se verifica:

$$a_p = (\det P)^{-1} \det(a_{ij}^{(p*j)}) \tag{3}$$

Este resultado está en proceso de arbitraje como se menciona en [2]. Al aplicar el teorema anterior en el plano se obtiene que el sistema:

$$\begin{aligned} x' &= ax + by \\ y' &= cx + dy + u \quad b \neq 0 \end{aligned}$$

es linealmente equivalente al sistema: En efecto, en este caso

$$\begin{aligned} A &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ P &= \begin{bmatrix} AB & B \end{bmatrix} = \begin{bmatrix} b & 0 \\ d & 1 \end{bmatrix}, \text{ donde } B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ A^2 &= \begin{bmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{bmatrix} \end{aligned}$$

En consecuencia de (3) se obtiene:

$$a_1 = \frac{1}{b} \det \begin{bmatrix} ab + bd & 0 \\ bc + d^2 & 1 \end{bmatrix} = a + d = \text{tr}A$$

$$a_2 = \frac{1}{b} \det \begin{bmatrix} b & ab + bd \\ d & bc + d^2 \end{bmatrix} = bc - ad = -\det A$$

2. Sea  $M$  una variedad lineal de codimensión  $k$  en  $\mathbb{R}^n$ ,  $x_0 = (x_{1_0}, \dots, x_{n_0})$  un elemento de  $M$  y  $x(t, u) = (x_1(t, u), \dots, x_n(t, u))$  con condición inicial  $x(0, u) = x_0$ , una solución del sistema (2).

Además  $\langle x, y \rangle$  denota el producto interno usual en  $\mathbb{R}^n$  y  $S^0$  el subespacio ortogonal al subespacio  $S$ .

**Teorema 2.** Sea  $S$  un subespacio lineal de codimension  $k$ ,  $1 \leq k < n$  y  $S^0$  el subespacio ortogonal generado por los vectores  $m_i$ ,  $i = 1, \dots, k$ , tales que sus últimas componentes no nulas son respectivamente  $m_{i,n_i} = -1$ . Si  $x(t, u)$  es una solución del sistema (1) y pertenece a  $M = x_0 + S$  para  $t$  no negativo, entonces el control  $u = u(t)$  satisface:

$$u = \langle (N^{n-n_i})^t m_i, Ax \rangle \quad (4)$$

para cada  $i = 1, \dots, k$ , donde  $A$  es la matriz acompañante del sistema (1) y  $N$  es una matriz nilpotente de índice  $n$ .

La demostración de este resultado se desarrolla con detalle en [1].

El problema de la permanencia de soluciones de un sistema de control en variedades lineales, aparece también en [4].

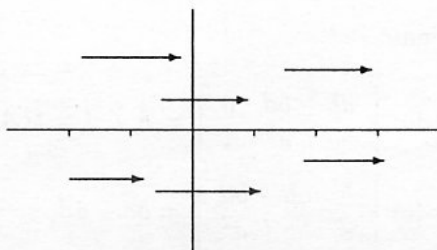
Para aplicar el teorema anterior en el plano se considera el sistema

$$\begin{aligned} x' &= y \\ y' &= ax + by + u \end{aligned}$$

y un vector  $\mathbf{m}$  que genera un subespacio ortogonal a una línea recta dada.

Se originan los siguientes casos:

- Si  $\mathbf{m} = (-1, 0)$  entonces de (4) se obtiene el control realimentado  $u = -ax - by$  y la solución trivial  $x(t) = x_0, y(t) = 0$ .
- Si  $\mathbf{m} = (0, -1)$  entonces de (4) se obtiene el control realimentado  $u = -ax - by$  y la solución  $x(t) = y_0 t + x_0, y(t) = y_0$ .



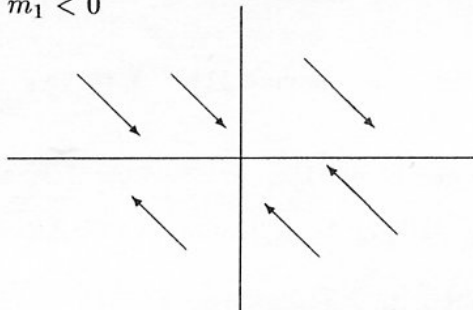
- Si  $\mathbf{m} = (m_1, -1)$  entonces de (4) se obtiene el control realimentado  $u = m_1 y - ax - by$  y se origina el sistema

$$\begin{aligned}x' &= y \\y' &= m_1 y\end{aligned}$$

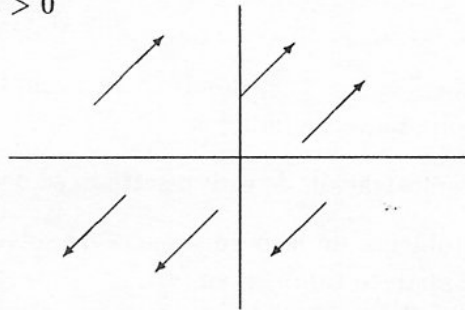
Luego

$$\begin{aligned}x(t) &= \frac{y_0}{m_1}(e^{m_1 t} - 1) + x_0 \\y(t) &= y_0 e^{m_1 t}\end{aligned}$$

$m_1 < 0$



$m_1 > 0$



En consecuencia si la linealización de (1) origina un sistema controlable, éste se transforma en otro equivalente de la forma:

$$\begin{aligned}x' &= y \\y' &= px + qy + u\end{aligned}\tag{5}$$

Al sistema (5) se le aplica la condición necesaria dada por (4) y se considera el conjunto de todas las condiciones iniciales que pueden ser transferidas en tiempo finito al punto crítico  $(x_0, y_0)$  a través de una línea recta  $M$ . Tal conjunto se denota por  $I(M)$ .

Para cada condición inicial  $(x_1, y_1)$  del sistema (1) que pertenezca al conjunto  $I(M)$  sea  $u(x_1, y_1)$  el control cuya respectiva trayectoria del sistema linealizado transfiere  $(x_1, y_1)$  al punto  $(x_0, y_0)$  en un tiempo finito  $t(x_1, y_1)$ .

El anterior control proporciona una trayectoria

$$(x(t, u(x_1, y_1)), y(t, u(x_1, y_1)))$$

del sistema (1) tal que para  $t = 0$  sale del punto  $(x_1, y_1)$  y para  $t = t(x_1, y_1)$  alcanza el punto  $(x_0^*, y_0^*)$ . El conjunto de tales puntos  $(x_1^*, y_1^*)$  proporciona valiosa información para el problema que estudiamos.

### 3 Aplicación a un modelo biomatemático

Se considera el modelo poblacional de Volterra-Lotka con control de captura sobre la población predadora ([3], pág. 61-80):

$$\begin{aligned}x' &= ax - by \\y' &= -cy + dxy + u; a, b, c, d > 0\end{aligned}\quad (6)$$

Al linealizar el sistema (6) en torno al punto estacionario  $(c/d, a/b, 0)$  se obtiene el siguiente sistema controlable:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 0 & -bc/d \\ ad/b & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \quad (7)$$

Usando el teorema 1 se transforma el sistema controlable (7) en otro linealmente equivalente:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -ac & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \quad (8)$$

La transformación lineal que realiza la equivalencia envía el punto  $(c/d, a/b)$  del sistema (7) al punto  $(-1/b, a/b)$  del sistema (8).

Se aplica ahora el teorema 2 para introducir en el sistema (8) el control realimentado  $u = acx + my$ , donde  $m$  es una constante negativa adecuada. Entonces se obtienen trayectorias que transfieren una condición inicial  $(x^*, y^*)$ ,  $x^* < -1/b$ ,  $y^* > a/b$ , al punto  $(-1/b, a/b)$  en tiempo finito. Esta propiedad se transfiere al sistema (7) a través de la inversa de la matriz de transformación y se obtienen trayectorias que salen de la condición inicial  $(x^*, y^*)$ ,  $x^* > c/d$ ,  $y^* > a/b$ , y alcanzan el punto  $(c/d, a/b)$  en el tiempo:

$$t = (x^* + 1/b) \log(a/by^*) / (y^* - a/b)$$

Al aplicar al sistema (6) el control que realiza la dinámica anterior se obtiene una trayectoria que se acerca al punto estacionario  $(c/d, a/b)$  del sistema poblacional de Volterra-Lotka.

### Bibliografía

- [1] Delgado, V. (1992) "Trayectorias de un sistema de control en  $\mathbb{R}^n$  sobre variedades lineales de codimensión 1 y  $n - 1$ ", *CUBO* 8: 71-74.
- [2] Delgado, V.; Alvarado, V. (1996) "Construcción explícita de un operador para sistemas de control linealmente equivalentes", (en proceso de arbitraje)
- [3] Beltrami, E. (1987) *Mathematics for Dynamic Modeling*. Academic Press, New York.
- [4] Chukwu, E. N. (1984) "Total Controllability to Affine Manifolds of Control Systems", *Journal of Optim.Th.and Appl.* 42(2): 181-199.
- [5] Lee, E. B.; Markus L. (1987) *Foundations of Optimal Control Theory*. John Wiley and Sons, New York.



## ANÁLISIS EN MODELOS DE REGRESIÓN CON DATOS CENSURADOS

JORGE DOMÍNGUEZ DOMÍNGUEZ<sup>1</sup>

---

### Resumen

Para estimar los parámetros en un modelo de regresión con datos censurados se han propuesto varios procedimientos, el objetivo en este trabajo es discutir y comparar las estimaciones que se obtienen de aplicar los métodos de máxima verosimilitud, el de mínimos cuadrados iterados propuesto por Schmee - Hahn y el esquema iterativo sugerido por Hamada - Wu. En el estudio se considera que la respuesta sigue una distribución de probabilidad Weibull. Se hace una extensión de estos procedimientos al caso de variabilidad no constante.

### 1 Introducción

Estimar el tiempo de falla de componentes mecánicos y eléctricos en la manufactura de un artículo, o el tiempo de degradación de un producto son objetivos relevantes dentro del contexto industrial. En ocasiones, es de interés estudiar el tiempo de mantenimiento de un equipo, es importante conocer el tiempo de cambio entre piezas en un equipo. Frecuentemente en estas situaciones existen variables tales como el esfuerzo, temperatura o carga a las que se someten los componentes, o productos y estas influyen en la determinación de la respuesta. Para evaluar el desempeño de un producto, existen otras variables de respuesta, como el número de kilómetros recorridos por un automóvil, número de veces que un sistema ha prestado un servicio. En el desarrollo de una estrategia experimental se consideran variables y valores de estas que tienen impacto en aumentar la durabilidad de un producto.

La existencia de covariables permiten considerar el estudio de modelos de regresión, con el tiempo de falla como variable de respuesta, los modelos son de utilidad para poder predecir la durabilidad de un equipo. El interés está en establecer la relación entre las variables  $T$ : tiempo de falla ( $T > 0$ ) y las covariables  $X$ . Para ello, frecuentemente se usa la transformación  $Y = \log T$  para representar el tiempo de falla y describir las distribuciones probabilísticas tanto en términos de  $Y$  como de  $T$ , usualmente la distribución  $Y$  es más simple. Así el modelo de regresión paramétrico es:

---

<sup>1</sup>CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS, GUANAJUATO, MÉXICO

$$Y = \log T = X\beta + \sigma_x \varepsilon \quad (1)$$

$\beta_{px1}$  y  $\sigma_x = \sigma(x)$  son los parámetros, en general, la distribución de probabilidad  $Y$  se representa por parámetros de forma, escala o posición, cabe notar que estos guardan una relación con los parámetros de la distribución  $T$ . El análisis de datos del tiempo de falla considerando el modelo 1 supone que la varianza es constante, esto es,  $\sigma(x) = cte$ . Esta suposición permite asegurar la estimación de los parámetros  $\beta, \sigma_x$  del modelo por máxima verosimilitud con mayor precisión.

En varias ocasiones no se satisface el supuesto de que  $\sigma(x) = cte$ , dada esta situación surge el interés por estudiar las propiedades estadísticas de los estimadores, en este trabajo también se considera el caso de que  $\sigma(x) = \exp(\alpha_0 + X\alpha_1)$ , entonces se dice que existe efecto de dispersión. El impacto sobre el efecto de dispersión se deberá principalmente al parámetro  $\alpha_1$ , sin embargo es importante estudiar los estimadores del parámetro  $\theta = (\beta_0, \beta_1, \alpha_0, \alpha_1)$  y la respuesta estimada  $\hat{Y}$ . En este planteamiento se considera que la covariable  $X$  tiene efecto tanto en la media como en la variabilidad.

Estudios iniciales cuando  $\sigma(x) \neq cte$ , los realizaron Bartlett & Kendall (1946), posteriormente Box & Draper (1959). Desde mediados de los 80 a la fecha se ha mostrado considerable interés en modelar la variabilidad y en estimar los efectos de dispersión. La mayoría de publicaciones que hacen alusión a este tema se refieren al análisis de efectos de dispersión, las investigaciones entorno a este tema, se pueden ubicar bajo el nombre de estimación de la función varianza.

Entre las aportaciones más recientes a la estructura de modelar la variación, se pueden citar Nair & Pregibon (1988) quienes consideran el análisis de efectos de dispersión en diseños  $2^k$ . En un desarrollo más general (McCullagh & Nelder, 1989) presentan la modelación conjunta de la media y dispersión. Davidian & Carroll (1987) discuten varios procedimientos para la estimación de la función varianza. Una descripción con más detalle de estos procedimientos la presentan (Carroll & Ruppert, 1988). Un estudio de efectos de dispersión en modelos lineales generalizados la reporta Smyth (1989). Sin embargo en ninguno de estos trabajos se considera el caso de datos censurados.

El objetivo de este trabajo es estimar los parámetros del modelo 1, por un procedimiento derivado del método de máxima verosimilitud y compararlo con los métodos propuestos por Schmees - Hahn y el de Hamada - Wu, en primera instancia se considerará el caso de  $\sigma(x) = \sigma$ , para ilustrar esta comparación se empleará un conjunto de datos reportados en la literatura estadística. Se describirá la estrategia para extender al estudio a modelos de regresión con variabilidad no constante usando la función  $\sigma(x) = \exp(X\alpha)$ , para tal fin se agregan otros dos procedimientos derivados del método de máxima verosimilitud

La estructura de los apartados de este trabajo es como sigue, en la siguiente parte se describirán las características de los datos. Los procedimientos para estimar los parámetros se muestran en 3, un ejemplo para comparar la estimación de los parámetros considerando los procedimientos propuestos se presentan en el apartado 4. Finalmente, en el último apartado se describen los resultados.

## 2 Característica de los datos

Dentro del marco de este tema, los métodos estándar para estimar el parámetro  $\theta$  se basa en procedimientos de máxima verosimilitud y en torno a este se derivan las propiedades estadísticas, cuyos resultados generalmente son asintóticos. Sin embargo, es de interés en el medio industrial considerar el comportamiento empírico de  $\alpha_1$ , en **muestras pequeñas**, esto abre la posibilidad de estudiar la estimación estadística de  $\theta$  considerando este aspecto.

**Muestra con censura**, una característica anexa al análisis de este tipo de datos surge porque algunas unidades pueden fallar antes de un período establecido o en una fase de experimentación. Esta circunstancia da lugar a un conjunto de datos incompletos para el análisis estadístico, esta situación es común en estudios de tiempo de falla y se denomina censura. La censura proporciona situaciones especiales e interesantes dentro del análisis estadístico, la censura aparece de distintas maneras, en Lawless (1982), pag. 31 ss., se realiza una discusión sobre estos conceptos con detalle.

El tipo de censura que se considera para los datos que se modelarán por la expresión 1, es el denominado censura tipo I, éste se define como sigue, considere un período de observación  $C$ , entonces una unidad observa un tiempo de falla  $T_i$ , si  $T_i$  es menor o igual que  $C$  entonces la unidad no se censura, en caso contrario la unidad se censura y el tiempo de observación es  $C$ . Así, si  $T_o$  es un tiempo de observación de una unidad, se tiene que  $T_o = \min(T_i, C)$ . En una situación como la planteada, indica que hay censura por la derecha, este trabajo se desarrolla considerando este tipo de censura. También, se puede establecer censura por la izquierda, sin embargo el análisis estadístico desarrollado para un caso sigue de modo similar para el otro. El número de unidades sin censura observada en la muestra es aleatorio.

**Distribuciones.** El interés principal en el análisis de datos de vida, es modelar mediante las distribuciones de probabilidad la variable aleatoria: tiempo de falla, los modelos paramétricos más comunes en la literatura son los de Weibull, Exponencial, Gama, Gumbell o Valores Extremos, log logística, log normal entre otros, en este trabajo se considerarán las distribuciones Weibull y Valores Extremos (DVE). Recientemente Hallinam (1993) realiza una revisión sobre la distribución Weibull, en la cual propone uniformizar la notación para varias formas de la distribución Weibull, ya que distintas expresiones han aparecido desde el planteamiento inicial de esta, dando lugar a confusiones entre los usuarios al aplicar la distribución en la modelación de problemas.

Una excelente presentación que describe las propiedades matemáticas y estadísticas en la distribución Weibull y DVE la presentan (Johnson y Kotz (1970), Cap. 20,21) respectivamente. Dentro del marco del estudio de las distribuciones para la variable de falla o tiempo de vida, Lawless (1982), Kalbfleisch & Prentice (1980) estudian con detalle los procedimientos de estimación y las propiedades estadísticas de los parámetros para ambas distribuciones. Otro libro que tiene presencia en las referencias lo escribieron Cox y Oakes (1984) el cual ofrece un panorama resumido de los conceptos estadísticos en confiabilidad. A nivel de aplicaciones podemos referirnos Crowder et. al. (1991), el cual hace una breve descripción del papel que desempeña la estadística en el análisis de datos en confiabilidad y luego se desarrollan intuitivamente estas ideas ilustrándolas con datos hipotéticos y reales. En la mayoría de estas citas presentan los conceptos de estadística

Bayesiana para analizar datos generados por la variable tiempo de falla.

Finalmente, aplicaciones a la ingeniería se pueden consultar en Nelson (1990), ahí presenta la teoría y práctica de pruebas de vida acelerada. Más aplicaciones y desarrollos de procedimientos en confiabilidad aparecen en Kececioglu (1991), O'Connor (1991).

Existe una estrecha relación entre las distribuciones Weibull y valores extremos (DVE). Esta se puede establecer mediante el siguiente procedimiento. La variable  $T$  sigue una distribución Weibull  $W(\lambda(x), \gamma(x))$  cuya función densidad es:

$$f(t) = \lambda(x)\gamma(x)(\lambda(x)t)^{\gamma(x)-1} \exp(-(\lambda(x)t)^{\gamma(x)}),$$

la función de confiabilidad  $S(t)$  para la distribución Weibull toma la forma:

$$S(t) = \exp(-(\lambda(x)t)^{\gamma(x)})$$

Si en lugar de trabajar con la variable  $T$ , se propone  $Y = \log T$ , entonces  $Y \sim DVE$ , por el método de transformación se sigue que la función de densidad de  $Y$  es:

$$f(Y) = \exp\left(\frac{Y - \mu(x)}{\sigma(x)} - \exp\left(\frac{Y - \mu(x)}{\sigma(x)}\right)\right) \quad (2)$$

con  $Y \in (-\infty, \infty)$ . La relación de los parámetros en ambas densidades está dada por  $\mu(x) = \log \lambda(x)$  y  $\sigma(x) = \gamma(x)^{-1}$ . Para expresarlo en la forma lineal como en  $\mu(x) = X\beta$ .  $S(t)$  para la DVE es:

$$S(Y) = \exp\left(-\exp\left(\frac{Y - \mu(x)}{\sigma(x)}\right)\right) \quad (3)$$

Es importante notar que la relación entre las variables  $T$  y  $Y$  en el modelo 1 siguen respectivamente una distribución Weibull y de DVE, esto indica que  $\varepsilon \sim DVE$ .

### 3 Procedimientos de estimación

Existen varios procedimientos planteados en la literatura estadística con el objeto de estimar los parámetros  $\theta = (\beta_0, \beta_1, \sigma(x)) = (\beta, \sigma(x))$  del modelo 1, en este trabajo se propone comparar el desempeño de tres procedimientos, el primero relacionado con la función de verosimilitud y los métodos iterativos propuestos por Schmee - Hahn y el de Hamada - Wu, en esta comparación se considera  $\sigma(x) = \sigma$ . Por otro lado, es de interés estudiar el caso cuando  $\sigma(x) = \exp(\alpha_0 + X\alpha_1)$ , esto da lugar a la variabilidad no constante en el modelo de regresión, para esta situación se considerarán solamente procedimientos relacionados con la función de verosimilitud, todos los procedimientos se describirán a continuación.

#### 3.1 Máxima verosimilitud

El primero de los procedimientos de estimación de máxima verosimilitud estima en forma conjunta los parámetros  $\theta$  y se le denotará por PMV, éste consiste en considerar la función densidad conjunta de la variable tiempo de falla  $t_i$  y de la indicadora  $\delta_i$ , es decir

$$\delta_i = \begin{cases} 1 & \text{si } t_i \leq C \text{ no hay censura} \\ 0 & \text{si } t_i > C \text{ si existe censura} \end{cases}$$

donde  $C$ : tiempo de censura,  $i = 1, \dots, n$ , así esta densidad es  $f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$ , entonces la función de verosimilitud es:

$$L_i(\beta, \sigma(x)) = f(t_i)^{\delta_i} S(t_i)^{1-\delta_i},$$

el logaritmo de esta expresión es:

$$l_i(\theta) = \log L_i(\beta, \alpha) = (\delta_i \log f(t_i) + (1 - \delta_i) \log S(t_i)) \quad (4)$$

La variable aleatoria  $\varepsilon$  en el modelo 1 sigue una distribución de valores extremos estándar. Donde  $f(\varepsilon)$  y  $S(\varepsilon)$  para esta distribución están dadas por las expresiones 2 y 3. La función de verosimilitud para esta situación es como sigue:

$$l(\theta) = \log \prod L_i(\beta, \sigma(x))$$

entonces:

$$l(\theta) = \log(\sigma(x)) + \sum_i^r \frac{Y_i - X_i \beta}{\sigma(x)} - \sum_i^n \exp \frac{Y_i - X_i \beta}{\sigma(x)}$$

los valores estimados del parámetro  $\theta$  se obtienen optimizando la función  $l(\theta)$ , por el método Newton- Raphson. En este procedimiento se obtiene la matriz de varianzas covarianzas que resulta de utilidad para hacer inferencias sobre los parámetros, (ver Lawless 1982, cap 6).

Con objeto de extender el estudio a efectos de dispersión se tiene que  $\sigma(x) = \exp(\alpha_0 + X\alpha_1)$  entonces el parámetro  $\theta = (\beta, \sigma(x)) = (\beta, \alpha)$ , para estimar éste se anexan dos procedimientos. El primero de ellos se denomina pseudoverosimilitud y se denotará por PPV, en éste, inicialmente se estima el parámetro  $\beta$  por máxima verosimilitud, el valor estimado de  $\beta$  se sustituye en la expresión 1, luego por máxima verosimilitud se estima el parámetro  $\alpha$ , el proceso continúa hasta encontrar una convergencia de ambos parámetros. El otro consiste en hacer una estimación ponderada del parámetro  $\beta$  por mínimos cuadrados, el siguiente paso es sustituir este valor estimado en la expresión 1 para estimar  $\alpha$  por máxima verosimilitud, el proceso termina hasta alcanzar la convergencia, éste se denota por PON.

### 3.2 Procedimiento de Schmee y Hahn (SH)

Schmee - Hahn (1979), sugieren el uso del método de mínimos cuadrados Iterados como una aproximación al análisis de datos censurados. El método opera de la siguiente manera:

- Primera etapa:

1. Se obtiene un estimador inicial de mínimos cuadrados para el modelo 1, se consideran los valores censurados como fallas.
2. Usando este ajuste, se calcula el tiempo de falla esperado de cada dato censurado.

- Segunda etapa:
  - 3) Estos valores esperados se usan en lugar de los tiempos de censura, se realiza un nuevo ajuste de mínimos cuadrados.
  - 4) Se repite el punto 2 de la primera etapa.
- Tercera etapa: El procedimiento continúa hasta que se obtiene una convergencia.

### 3.3 Procedimiento de Hamada y Wu (HW)

Una propuesta alternativa al método de Schmee-Hahn la propusieron Hamada-Wu (1991), el procedimiento es muy similar al anterior solo que en este caso se utiliza el método de máxima verosimilitud en lugar del de mínimos cuadrados, además de proponer un modelo inicial que se ajuste adecuadamente a los datos, el detalle del procedimiento es el siguiente:

- Primera etapa:
  1. Se obtiene un modelo inicial  $X_o\beta_o$ , tal que explique al conjunto de datos.
  2. Se estima por máxima verosimilitud los parámetros del modelo 1
  3. Usando este ajuste, se calcula el tiempo de falla esperado para cada dato censurado.
  4. Se repiten los pasos 2 y 3 hasta alcanzar la convergencia y obtener el modelo  $X_i\beta_i$ .

La expresión para obtener el valor del tiempo de falla esperado en los dos procedimientos anteriores esperado es la siguiente:

$$E(Y|Y > C) = X_i\beta_i + \sigma \frac{f(z)}{S(z)}$$

donde  $z = \frac{Y - X_i\beta_i}{\sigma}$ .

## 4 Aplicación de los procedimientos

La programación para estimar los parámetros del modelo siguiendo los diferentes procedimientos se realizó usando el sistema GAUSS versión 3.14, cuyo conjunto de procedimientos, funciones y generador de números aleatorias están altamente probados para asegurar eficiencia y exactitud en los resultados numéricos.

Ejemplo motorettes. En la tabla 1 se describen los resultados de una prueba de vida acelerada de aislante eléctrico en 40 motorettes, reportados originalmente por Crawford (1970). 10 motorettes se probaron en cuatro temperaturas. En total fallaron 17 motorettes, resultaron sin falla 23 de ellos.

La variable de respuesta es  $T$ : el tiempo (hrs.) de falla que se supone tiene una distribución Weibull,  $Y = \log(T)$  entonces  $Y$  se distribuye como una DVE. A la covariable temperatura se le aplicó la transformación de — Arrhenius,

$$X = \frac{1000}{273.2 - T}$$

150 °C	170 °C	190 °C	220 °C
	1764	408	408
	2774	408	408
	3444	1344	504
	3542	1344	504
	4860	1440	504
	4860		
	5196		
en 150 °C sin falla a 8064 horas			
en 170 °C 3 fallaron en 5448 horas			
en 190 °C 5 fallaron en 1680 horas			
en 220 °C 5 fallaron en 528 horas			

Tabla 1: Datos de 40 motoretas

con estas transformaciones se ajustó el modelo 1 usando los tres procedimientos, los valores estimados, se describen en la tabla 2, donde  $l(\theta)$  es el valor de la función de verosimilitud,  $Y_{130}$  es la proyección a 130°C, en el centro de la distribución, con lo cual se puede estimar la proyección del tiempo de falla en horas como se puede observar en el último renglón

	PMV	SH	HW
$\beta_0$	-13.423	-13.396	-10.520
$\beta_1$	9.755	9.680	8.385
$\sigma$	0.538	0.471	0.657
$l(\theta)$	-24.687		
$Y_{130}$	10.771	10.612	10.276
	47619	40619	29027

Tabla 2: Estimación de los parámetros

Haciendo un análisis descriptivo de la tabla 1 se puede observar que existe una variabilidad entre los datos dentro de cada nivel de temperatura, claro que ambas transformaciones ayudan a linealizar el problema y con ello estar próximo a un valor de  $\sigma(x) = cte$ , sin embargo, se puede extender el análisis de estos datos tomando en cuenta la presencia de variabilidad, se hizo considerando los tres procedimientos relacionados con la estimación de máxima verosimilitud, para ello solo se consideraron los datos de la temperatura mayor que 150°C, los resultados se muestran en la tabla 3. Análogamente a la descripción anterior se realiza la proyección en  $Y_{130}$ , con los tiempos de falla.

## 5 Análisis de Resultados

Se observa de la tabla 2 que las estimaciones de PMV y SH son muy similares, sin embargo esta diferencia es notoria en la proyección del tiempo falla, puesto que, PMV representa 7000 horas más de funcionamiento de los motores. En estos datos el procedi-

miento HW tiene un bajo desempeño, aunque ellos proponen tener un modelo adecuado desde el inicio, esto puede dar lugar a ese pequeño desajuste.

	PMV	PPV	PON
$\beta_0$	-12.852	-11.542	-11.166
$\beta_1$	9.465	8.889	8.544
$\alpha_0$	-11.325	-6.873	-10.073
$\alpha_1$	4.728	2.723	4.312
$l(\hat{\theta})$	-21.733	-22.387	-29.382
$Y_{130}$	10.662	10.5049	10.024
	41056	36465	22572

Tabla 3: Estimación de los parámetros en presencia de variación

El análisis de estos datos considerando que  $\sigma(x) = \exp(\alpha_0 + X\alpha_1)$ , se observa que los estimadores son muy similares, sin embargo en la proyección la diferencia se ve grande. El PMV muestra mayor tiempo de falla. No obstante que se ha trabajado con 30 datos los resultados son muy similares cuando se ha considerado que  $\sigma(x) = \exp(\alpha_0)$ . Sin embargo es de interés estadístico probar el impacto del efecto de dispersión, es decir cuando  $\alpha_1 = 0$ , mediante la razón de verosimilitud se puede probar la siguiente hipótesis  $H_0 : \theta = (\beta_0, \beta_1, \alpha_0, \alpha_1) = (\beta_0^o, \beta_1^o, \alpha_0^o, 0)$ . Entonces la razón de verosimilitud es

$$R(\theta) = -2 \log \frac{L(\tilde{\theta})}{L(\hat{\theta})} = 2(l(\hat{\theta}) - l(\tilde{\theta}))$$

donde  $l(\tilde{\theta})$  es la estimación de máxima verosimilitud cuando la hipótesis  $H_0$  es cierta,  $l(\hat{\theta})$  es la estimación cuando  $\alpha_1 \neq 0$ .  $R(\theta) = 3.166$ , donde  $l(\tilde{\theta}) = -22.953$  (con 30 datos), comparando la  $R(\theta)$  con una  $\chi^2$  con  $gl=3$ , no se rechaza la hipótesis  $H_0$ , no se puede afirmar de una variabilidad existente en los datos.

Conviene plantear varias situaciones mediante un estudio Monte Carlo para ver el desempeño de los procedimientos, y considerar los casos donde exista variabilidad.

## Bibliografía

- [1] Bartlett, M.S. & Kendall, D.G. (1946) "The statistical analysis of variance-heterogeneity and the logarithmic transformation", *Journal of the Royal Statistical Society*, Serie B, 8: 128 – 138.
- [2] Box, G.E.P. & Drapper, N.R. (1959) "A basis for the selection of a response surface design", *Journal American Statistical Association* 54: 622–654.
- [3] Carroll, R.L. & Ruppert, D. (1988) *Transformation and Weighting in Regression*. Chapman and Hall, New York.
- [4] Cox, D.R. & Oakes, D. (1984) *Analysis of Survival Data*. Chapman & Hall, New York.



- [5] Crowder, M.J.; Kimber, A.C.; Smith, R.L.; Sweeting, T.J. (1991) *Statistical of Reliability Data*. Chapman & Hall, New York.
- [6] Davidan, M. & Carroll, R.J. (1987) "Variance function estimation", *Journal American Statistical Association* **82**(400): 1079 – 1091.
- [7] Hallinan, A.J. (1993) "A review of the Weibull distribution" *Journal Quality Technology* **25**(2): 85–93.
- [8] Hamada M. & Wu, C.F.J. (1991) "Analysis of censored data from highly fractionated experiments", *Technometrics* **33**(1).
- [9] Johnson, N.L. & Kotz, S. (1970) *Distributions in Statistics. Continuous Univariate Distributions*, Vol. 1. John Wiley & Sons. New York.
- [10] Kalbfleisch, J.D. & Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.
- [11] Kececioglu, D. (1991) *Reliability Engineering Handbook*, Vol. I y II. ASQC, Milwaukee.
- [12] Lawless, J.F. (1982) *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York.
- [13] McCullagh, P. y Nelder, J.S. (1988) *Generalizad Linear Models*, 2nd. ed. Chapman and Hall, New York.
- [14] Nair, V.N., & Pregibon, D. (1988) "Analyzing dispersion effects from replicated factorial experiments", *Technometrics* **30**: 247–257.
- [15] Nelson, W. (1990) *Accelerated Testing-Statistical Models, Test Plans and Data Analyses*. John Wiley & Sons, New York.
- [16] Schmee, J. & Hahn, G. (1979) "A simple method for regression analysis with censored data", *Technometrics* **21**: 417–432.
- [17] Smyth, G.K. (1989) "Generalized linear models with varying dispersion", *Journal of the Royal Statistical Society, Serie B.* **51**(1): 47–60.
- [18] O'Connor, P. (1991) *Practical Reliability Engineering*. ASQC, Milwaukee

## INFORMATION ET VALEURS PRONOSTIQUES DE CERTAINES VARIABLES DANS LE CANCER DU SEIN

F. MAX DORVILLE<sup>1</sup> - JÉRÔME MANUCEAU<sup>1</sup>  
MARYLÈNE TROUPÉ<sup>1</sup> - JEAN VAILLANT<sup>1</sup>

---

### Résumé

Dans le cadre de l'analyse de durées de vie, la notion d'information mutuelle entre  $\sigma$ -algèbres est utilisée pour sélectionner un ensemble de covariables qualitatives ou quantitatives à fortes valeurs prédictives. Le taux d'incertitude expliqué est testé sur la base d'une décomposition similaire [7] à celle de l'analyse de variance. Les données manquantes sur certaines covariables ne conduisent pas à l'élimination des unités statistiques concernées. On évite ainsi la perte d'information qui en découlerait. Ces résultats sont appliqués à l'étude de la durée de vie de 1304 malades atteints du cancer du sein suivis sur une période de 10 ans. La technique utilisée permet de trouver un nombre limité de facteurs ayant une valeur prédictive intéressante.

**Mots-clés:** entropie,  $\sigma$ -algèbre, information, analyse de survie

### Resumen

En el marco del análisis de duración de vida, la noción de información mutua entre  $\sigma$ -álgebras utilizada para seleccionar un conjunto de covariables cualitativas o cuantitativas con fuertes valores predictivos. La tasa de incertidumbre explicada es probada sobre la base de una descomposición similar [7] a la del análisis de la varianza. Los datos faltantes sobre ciertas covariables no conducen a la eliminación de las unidades estadísticas involucradas. Se evita así la pérdida de información que se generaría. Estos resultados se aplican al estudio de la duración de vida de 1304 enfermos de cáncer del seno observados durante un período de 10 años. La técnica utilizada permite encontrar un número limitado de factores que tienen un valor predictivo interesante.

**Palabras claves:** entropía,  $\sigma$ álgebra, información, análisis de sobrevivencia.

## 1 Introduction

La notion d'entropie d'une loi de probabilité  $P$  sur un espace  $\Omega$  est un outil intéressant pour caractériser l'incertitude associée à  $P$  [8]. Quand  $\Omega$  est dénombrable, l'entropie de  $P$  notée  $H(P)$  est définie par :

---

<sup>1</sup>UNIVERSITÉ DES ANTILLES-GUYANE, UFR SCIENCES EXACTES ET NATURELLES-CAMPUS DE FOUILLOLE, DÉPARTEMENT DE MATHÉMATIQUES-INFORMATIQUE, 97159 POINTE-À-PITRE CEDEX, FRANCE

$$H(P) = - \sum_{\omega \in \Omega} P(\omega) \ln(P(\omega)) \quad (1)$$

avec la convention  $0 \ln 0 = 0$ , ce qui revient en fait à écrire

$$H(P) = - \sum_{\omega \in \Omega'} P(\omega) \ln(P(\omega)) \quad (2)$$

où  $\Omega' = \{\omega \in \Omega \mid P(\omega) > 0\}$ .

Pour une variable aléatoire  $X$  sur  $\Omega$  dénombrable, on définit alors son entropie notée  $H(X)$  par celle de sa loi image  $P_X$  :

$$H(X) = - \sum_{x \in X(\Omega)} P(X^{-1}(\{x\})) \ln(P(X^{-1}(\{x\}))). \quad (3)$$

$H(X)$  mesure l'incertitude sur la valeur de  $X(\omega)$  pour un choix (ou réalisation) d'un élément  $\omega$  de  $\Omega$  selon la loi  $P$ . Ainsi, si  $\mathcal{A} = (A_i)_{i \in I}$  est une partition de  $\Omega$ , on définit l'entropie de cette partition par  $H(\mathcal{A}) = H(X_{\mathcal{A}})$  où  $X_{\mathcal{A}}$  est la variable aléatoire définie de la façon suivante :

$$\begin{aligned} X_{\mathcal{A}} : \Omega &\longrightarrow \{0, 1\}^I \\ \omega &\longmapsto (I_{A_i}(\omega))_{i \in I} \end{aligned}$$

On a par conséquent,

$$H(X_{\mathcal{A}}) = - \sum_{i \in I} P(A_i) \ln(P(A_i)). \quad (4)$$

$H(X_{\mathcal{A}})$  mesure l'incertitude sur l'appartenance à l'une des parties  $A_i$  avant l'observation d'un élément  $\omega$  de  $\Omega$ .

Si on observe deux variables aléatoires  $X$  et  $Y$  sur  $\Omega$ , l'entropie de  $X$  conditionnellement à une valeur  $y$  prise par  $Y$  est l'entropie de la loi de  $X$  conditionnelle à l'événement  $Y = y$ , que l'on note  $H(X \mid Y = y)$ . On définit alors l'entropie de  $X$  conditionnelle à  $Y$  que l'on note  $H(X \mid Y)$  par

$$H(X \mid Y) = \sum_{y \in Y(\Omega)} P(Y = y) H(X \mid Y = y) \quad (5)$$

qui est l'espérance de la fonction aléatoire qui à  $y$  fait correspondre  $H(X \mid Y = y)$ .

Si on considère maintenant une loi de probabilité  $P$  sur  $\Omega$  cette fois non dénombrable, quelques précautions doivent être prises pour la généralisation formelle de la notion d'entropie vue précédemment. Ceci conduit à l'entropie dite généralisée  $H(P; \mu)$  relative à une mesure de référence  $\mu$  sur  $\Omega$  ([2],[9]) et à l'entropie d'une  $\sigma$ -algèbre  $\mathcal{A}$  de type séparable [1] que l'on notera par la suite  $H(\mathcal{A})$ . Le lien formel entre entropie de *Shannon* [10] et entropie de *Kullback* [4] est donné, par exemple, dans [9] ainsi que leurs propriétés respectives.

## 2 Information entre $\sigma$ -algèbres

Soient  $\mathcal{A}$  et  $\mathcal{B}$  deux  $\sigma$ -algèbres, on note  $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{A} \cup \mathcal{B})$  la  $\sigma$ -algèbre engendrée par  $\mathcal{A} \cup \mathcal{B}$ .

**Définition 1**  $\mathcal{A}$  et  $\mathcal{B}$  étant deux  $\sigma$ -algèbres de type séparable, l'entropie de  $\mathcal{A}$  conditionnelle à  $\mathcal{B}$  est définie par

$$H(\mathcal{A} | \mathcal{B}) = H(\mathcal{A} \vee \mathcal{B}) - H(\mathcal{B}). \quad (6)$$

Cette entropie représente l'incertitude restant sur  $\mathcal{A}$  quand on connaît  $\mathcal{B}$ .  
On rappelle les propriétés suivantes

### Propriété 1

1.  $\mathcal{A} \subset \mathcal{B} \Rightarrow H(\mathcal{A}) \leq H(\mathcal{B})$ .
2.  $H(\mathcal{A} \vee \mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B})$  et  $H(\mathcal{A} | \mathcal{B}) \leq H(\mathcal{A})$ .
3. Si  $\mathcal{A}$  et  $\mathcal{B}$  sont d'entropies finies, alors on a:  
 $H(\mathcal{A} \vee \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B})$  et  $H(\mathcal{A} | \mathcal{B}) = H(\mathcal{A})$  si et seulement si  $\mathcal{A}$  et  $\mathcal{B}$  sont indépendants. On écrit alors  $\mathcal{A} \perp \mathcal{B}$ .

Pour une bonne quantification  $I(\mathcal{A}, \mathcal{B})$  de l'information entre  $\sigma$ -algèbres  $\mathcal{A}$  et  $\mathcal{B}$ , il est nécessaire de poser en postulats quelques propriétés exigibles de  $I(\mathcal{A}, \mathcal{B})$ . Il nous paraît souhaitable que  $I(\mathcal{A}, \mathcal{B})$  vérifie les postulats suivants :

1.  $\mathcal{A} \perp \mathcal{B} \Leftrightarrow I(\mathcal{A}, \mathcal{B}) = 0$ ,
2.  $I(\mathcal{A}, \mathcal{B}) = I(\mathcal{B}, \mathcal{A})$ ,
3.  $\mathcal{A} \subset \mathcal{C} \Rightarrow I(\mathcal{A}, \mathcal{B}) \leq I(\mathcal{C}, \mathcal{B})$ .

En choisissant

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A} \vee \mathcal{B}), \quad (7)$$

tous ces postulats sont vérifiés et on a de plus,

$$I(\mathcal{A}, \mathcal{A}) = H(\mathcal{A}). \quad (8)$$

Quand on n'a pas de modèles paramétriques suffisamment raisonnables par rapport au phénomène observé (variable d'intérêt  $Y$  et  $k$  covariables  $X_1, \dots, X_k$ ), une démarche non paramétrique est envisageable à partir de l'information fournie par les tables de contingences associées à ces observations de façon analogue à celle utilisée par [6] et [7]. Nous utilisons ici un principe de maximum d'information par rapport à une  $\sigma$ -algèbre de référence [5], cette dernière correspondant dans le cas présent aux observations de la variable d'intérêt  $Y$ .

### 3 Méthode de sélection de covariables

La méthode employée est basée sur l'estimation des entropies conditionnelles et non conditionnelles de la variable  $Y$  et des covariables  $X_1, \dots, X_k$ . Ces estimations sont calculées à partir des fréquences empiriques fournies par des tables de contingences à plusieurs entrées. Pour les covariables continues, une discrétisation est donc nécessaire.

#### 3.1 Préparation des tables de contingences

Pour diminuer l'occurrence de faibles effectifs dans les tables de contingences, le nombre de modalités d'une covariable est parfois restreint en utilisant un principe de maximum d'entropie ci-après décrit.

##### Quelques rappels sur le principe de maximum d'entropie

Parmi les lois de probabilité définies sur un espace  $\Omega$  de cardinal  $m$ , la loi uniforme sur  $\Omega$  est celle qui maximise l'entropie, ce maximum valant  $\ln(m)$ .

Le principe de maximum d'entropie est applicable à une variable  $Z$  devant être quantitative ou qualitative ordonnée. Il consiste à réduire le nombre initial de modalités à une valeur  $k$  en constituant  $k$  classes de modalités consécutives d'effectifs le plus voisin possible.

Par exemple, pour une variable  $Z$  présentant 10 modalités d'effectifs représentés dans le tableau 1, une réduction à 3 modalités par le principe de maximum d'entropie se ferait comme suit:

Modalité initiale	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	$Z_7$	$Z_8$	$Z_9$	$Z_{10}$
Effectif	3	16	10	7	19	8	11	5	1	14
Nouvelle modalité	$Z'_1$			$Z'_2$			$Z'_3$			
Nouvel effectif	29			34			31			

Tabla 1: Exemple de regroupement de modalités par le principe de maximum d'entropie

#### 3.2 Méthode de sélection

##### 3.2.1 Information portée par une covariable

Soient  $Y$  la variable d'intérêt,  $X$  une covariable quelconque,  $(n_{ij})$  le tableau de contingence qui leur est associé, et  $p = (p_{ij})$  la matrice des espérances des fréquences relatives.

La première étape de la méthode de sélection consiste à quantifier le degré de liaison entre  $X$  et  $Y$ . Pour cela, on utilise un coefficient de corrélation basé sur l'entropie de *Shannon*. Celui-ci est défini comme suit:

**Définition 2** L'information relative de  $X$  sur  $Y$  a pour expression

$$I_r(X, Y) = \frac{I(X, Y)}{H(Y)} \quad (9)$$

En utilisant les égalités (3), (6) et (7) on a:

$$I_r(X, Y) = \frac{H(Y) - H(Y | X)}{H(Y)} \in [0, 1], \quad (10)$$

ce qui montre que  $I_r(X, Y)$  peut être interprété comme un pourcentage d'information de  $X$  sur  $Y$ .

En utilisant les égalités (3) et (10), on obtient alors la propriété suivante:

**Propriété 2** Une expression de l'information relative de  $X$  sur  $Y$  est:

$$I_r(X, Y) = \frac{\sum_{i,j} p_{ij} \ln\left(\frac{p_{ij}}{p_i \cdot p_j}\right)}{\sum_j p_j \ln(p_j)} \quad (11)$$

On en déduit la propriété suivante:

**Propriété 3** Un estimateur de l'information relative de  $X$  sur  $Y$  est:

$$T = \hat{I}_r(X, Y) = \frac{\sum_{i,j} \hat{p}_{ij} \ln\left(\frac{\hat{p}_{ij}}{\hat{p}_i \cdot \hat{p}_j}\right)}{\sum_j \hat{p}_j \ln(\hat{p}_j)} \quad (12)$$

où  $\hat{p}_{ij}$ ,  $\hat{p}_i$  et  $\hat{p}_j$  sont les notations classiques pour les fréquences relatives empiriques.

Cet estimateur est convergent et suit asymptotiquement la loi normale  $\mathcal{N}(I_r(X, Y), \text{var}(T))$

où  $\text{var}(T)$  est la variance asymptotique de  $T$ . Son expression est

$$\text{var}(T) = \frac{1}{n} \text{var} \left( \sum_{i,j} \frac{\partial I_r(X, Y)}{\partial p_{ij}}(p) \mathbb{I}_{ij} \right) \quad (13)$$

où  $n = \sum_{i,j} n_{ij}$  est le nombre d'unités statistiques considérées, et les  $\mathbb{I}_{ij}$  sont les variables indicatrices associées au tableau  $(n_{ij})$ .

En utilisant les égalités (11) et (13), on obtient la propriété suivante:

**Propriété 4** Une expression de la variance asymptotique de l'estimateur  $T$  de l'information relative de  $X$  sur  $Y$  est:

$$\text{var}(T) = \frac{1}{n} \left[ \frac{1}{H^2(Y)} \sum_{i,j} p_{ij} \left[ \ln\left(\frac{p_{ij}}{p_i \cdot p_j}\right) + T \ln(p_j) \right]^2 \right] \quad (14)$$

dont un estimateur noté  $\widehat{\text{var}}(T)$  est obtenu en remplaçant  $p_{ij}$ ,  $p_i$  et  $p_j$  par leurs estimations empiriques  $\hat{p}_{ij}$ ,  $\hat{p}_i$  et  $\hat{p}_j$ .

### 3.2.2 Information portée par plusieurs covariables

De manière analogue au paragraphe précédent, on peut calculer l'information relative combinée de plusieurs covariables  $X_1, \dots, X_k$  sur  $Y$ . L'évaluation de la variance d'information permet alors d'établir des régions de confiance et d'éliminer les covariables les moins pertinentes.

**Vocabulaire:** Lorsque  $Y$  est la variable de survie,  $I_r(X, Y)$  est appelé valeur pronostique de  $X$  et noté  $I_r(X)$ .

### 3.2.3 Méthode pratique de sélection

On sélectionne progressivement les différentes covariables à l'aide d'un processus itératif. Notons  $E$  l'ensemble constitué de toutes les covariables  $X_1, \dots, X_k$ .

La première étape permet de construire l'ensemble  $E_1$  défini par

$$E_1 = \left\{ X_l \in E \ / \ \hat{I}_r(X_l, Y) + u_\alpha \sqrt{\widehat{\text{var}}(\hat{I}_r(X_l, Y))} \geq p_1 \right\}$$

et à une étape  $h$  où  $h \geq 2$ , on a

$$E_h = \left\{ (X_l, x) \in E \times E_{h-1} \ / \ \hat{I}_r((X_l, x), Y) + u_\alpha \sqrt{\widehat{\text{var}}(\hat{I}_r((X_l, x), Y))} \geq p_h \right\}$$

où  $p_1, p_h$  et  $\alpha$  sont des seuils fixés par l'utilisateur, et  $u_\alpha$  est le quantile de la loi normale vérifiant  $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$ ,  $\Phi$  étant la fonction de répartition de la loi centrée réduite.

#### Remarques

1. La suite de seuils  $(p_h)_{h \in \mathbf{N}^*}$  doit être strictement croissante car le rajout d'une covariable conduit à l'augmentation de l'information relative.
2. Le processus de sélection décrit présente l'avantage de toujours permettre, à une étape donnée, l'intégration d'une covariable non sélectionnée à une étape antérieure.

Dans la pratique, les critères d'arrêt sont les suivants:

- i)* un nombre maximum d'itérations est fixé par l'utilisateur,
- ii)* un seuil maximum au delà duquel l'utilisateur considère que l'information relative est suffisante est fixé,
- iii)* toute combinaison de *i)* et *ii)*,
- iv)* il n'a pas été possible de rajouter des covariables.

## 4 Hiérarchisation des modalités d'une covariable

Dans le cas où  $Y$  est la variable de survie, on appellera taux de survie au temps  $t$  le quotient du nombre de malades vivants par le nombre de malades non censurés. La courbe de survie est la courbe des taux de survie aux différents temps. Le taux de guérison d'une sous-population est le taux de survie à la fin de l'expérimentation.

Chaque covariable divise la population de malades en une partition correspondant aux différentes modalités (*cf. exemples présentés en fig. 3 et 4*). Ces dernières seront ordonnées en fonction de leur taux de guérison.

Si deux modalités ont des courbes de survie très proches, on pourra les regrouper en une modalité unique.

## 5 Application au cancer du sein

### 5.1 Présentation des données

Les données proviennent d'une enquête réalisée à Marseille par le professeur J.-M. SPITALIER et le docteur D. HANS. Elles concernent 1304 patients suivis sur une période de 10 ans. Les variables observées sur chaque patient, au nombre de 45, sont quantitatives et qualitatives. Le traitement appliqué aux différents malades dépendait des valeurs prises par ces variables.

Un prétraitement a permis de ne retenir qu'une quinzaine d'entre elles à savoir : 1-âge, 2-classe clinique, 3-classe thermographique, 4-classe sénographique, 5-classe échographique, 6-PEV clinique, 7-diamètre clinique, 8-allure clinique, 9-côté, 10-histologie, 11-histologie N, 12-nombre de ganglions envahis, 13-récepteurs œustradiol, 14-récepteurs progestérone, 15-stade UICC.

### 5.2 Sélection des covariables

Les calculs sont effectués à l'aide de programmes écrits en APL ou en Turbo Pascal. La première étape permet de sélectionner 5 covariables qui sont: la classe thermographique, le nombre de ganglions envahis, l'histologie, l'histologie N et le stade UICC avec pour seuils:  $p_1 = 5\%$  et  $\alpha = 5\%$  (*cf. Fig. 1*).

La seconde étape confirme ce choix avec pour seuils  $p_2 = 15\%$  et  $\alpha = 5\%$  (*cf. Fig. 2*). Mais à la troisième étape, la meilleure combinaison de 3 covariables est l'âge, l'histologie et le nombre de ganglions envahis: cette combinaison donne environ 40% d'information relative au décès.

### 5.3 Hiérarchisation des modalités d'une covariable

Sachant qu'une covariable a une bonne valeur pronostique, il est intéressant de pouvoir affecter aux différentes valeurs possibles de cette covariable une courbe de survie (*cf. Fig. 3 et 4*). Ces courbes représentent la survie conditionnelle à la valeur prise par cette covariable chez un malade et sont des outils de pronostic médical.



L'étude séparée des 4 covariables âge, histologie, nombre de ganglions envahis et stade UICC indique les modalités ayant les meilleures espérances de survie. Ainsi, pour la covariable nombre de ganglions envahis observés par histologie (*cf. Fig. 3*), 0, 1 ou 2 ganglions envahis donnent des courbes de survie voisines et au dessus de la survie moyenne alors que la modalité "3 à 4 ganglions envahis" est nettement moins bonne. La modalité "plus de 4 ganglions envahis" l'est encore moins.

L'âge pris isolément apporte peu d'information sur la survie d'où l'enchevêtrement des courbes de survie associées à ses modalités (*cf. Fig. 4*).

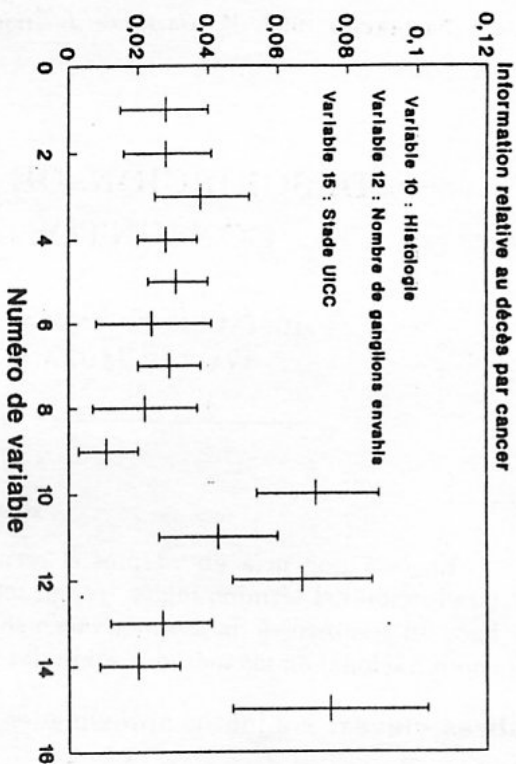
## 6 Conclusions

Nous avons présenté une méthode asymptotique qui ne s'applique que dans le cas d'un nombre important d'unités statistiques. Elle permet la sélection de covariables apportant le plus d'information sur la variable d'intérêt. Son application au cas de malades atteints du cancer du sein a permis de mettre en évidence quatre covariables principales. Les courbes de survie tracées pour chacune des modalités d'une covariable renseignent de manière plus précise sur la valeur pronostique.

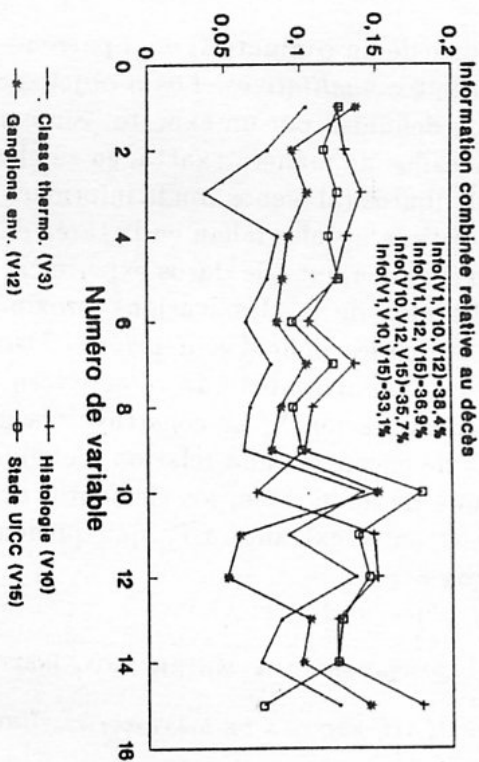
## Bibliografía

- [1] ARCONTE, A., KHATTABI, MANUCEAU, J., MARTIAS, C. (1995) Mutual Information between  $\sigma$ -algèbras, *Prépublication UAG*, 95/07.
- [2] DALEY & VERES-JONES (1988) *An introduction to the theory of point processes*. Springer-Verlag, New-York.
- [3] EL HASNAOUI, A. (1993) *Le concept du gain d'information: une nouvelle approche en épidémiologie quantitative*. Thèse de doct. - Univ. de Montpellier,
- [4] KULLBACK, S. (1959) *Information Theory and Statistics*. John Wiley & Sons, New York.
- [5] MANUCEAU, J., TROUPÉ, M., VAILLANT, J., "Information between  $\sigma$ -algebras: applications to selection of pronostic variables", *to appear*.
- [6] RAO, C.R. (1982) "Diversity: Its measurement, decomposition, apportionment and analysis", *Sankhya Series A*, 44: 1-22.
- [7] RAO, C.R. (1986) "Generalization of ANOVA through entropy and cross entropy functions", *Probability theory and mathematical statistics* 2: 477-494, Sciences Press, Utrecht.
- [8] RÉNYI, A. (1961) On measures of entropy and information. *Proc. of the fourth Berkeley Symposium on Mathematical Statistics & Probability* 1: 547-561.
- [9] ROBERT, C. (1990) "An entropy concentration theorem: applications in artificial intelligence and descriptive statistics", *J. Appl. Prob.* 27: 303-313.
- [10] SHANNON, C.E. (1948) "A mathematical theory of communication", *Bell System Technical Journal* 27: 379-423 & 623-656.

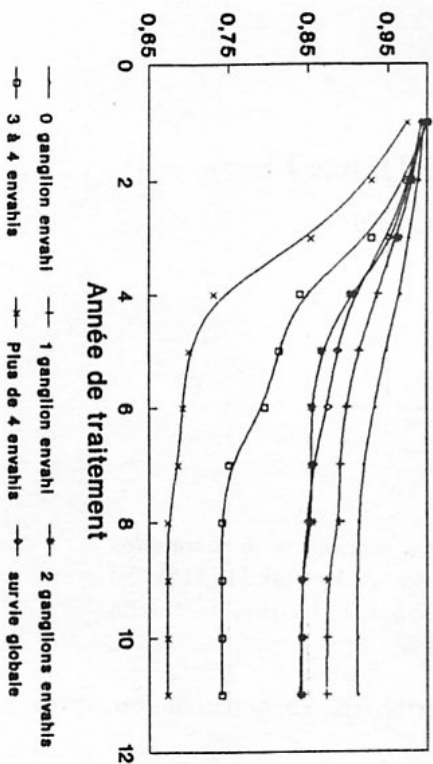
# 1. Données de Survie



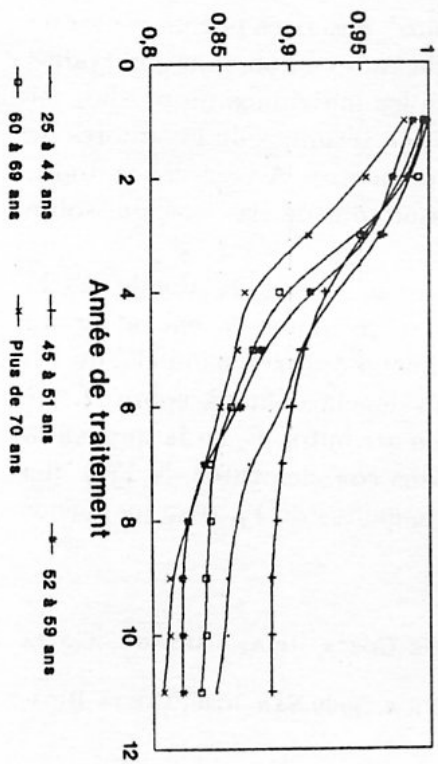
# 2. Données de Survie



# 3. Courbes de survie par classe de V12



# 4. Courbes de survie par classe d'âge



## DESCRIPCIÓN DE CLASES MEDIANTE CONJUNTOS APROXIMADOS

JOSÉ LUIS ESPINOZA<sup>1</sup> - GEOVANNI FIGUEROA<sup>1</sup>  
WALTER MORA<sup>1</sup> - JAVIER TREJOS<sup>1,2</sup>

---

### Resumen

En esta ponencia abordamos el estudio de los llamados *conjuntos aproximados* (traducción del término inglés "rough sets"), introducidos por Z. Pawlak [10, 15]. Se hace un resumen de la teoría y luego se comentan aplicaciones de la implementación computacional de los métodos sobre datos reales y simulados.

**Palabras claves:** conjuntos aproximados, reducción de atributos, caracterización aproximada.

### 1 Introducción

Se parte de un conjunto  $\Omega$  de  $n$  patrones, objetos o 'individuos', descritos por un vector de  $p$  atributos *cualitativos*. Los  $n$  objetos están divididos en clases, que han sido observadas o bien definidas por un experto. Sin embargo, posiblemente los individuos no puedan ser clasificados de manera exacta, en su clase asignada a priori, en términos de los valores de sus atributos (al menos con la información disponible en el momento). A veces los métodos estadísticos usuales fallan en la tarea de producir buenos resultados de clasificación, sobre todo si el conjunto de datos experimentales es muy pequeño.

La idea de la clasificación aproximada es la siguiente: sea  $\Omega = \cup Y_i$  donde las  $Y_i$  son las clases definidas *a priori*. Dada la clase  $Y_j$ , lo que se quiere es encontrar un conjunto de atributos que caractericen totalmente a una 'buena aproximación'  $Y'_j$  de  $Y_j$ . La aproximación  $Y'_j$  se construye vía una relación de equivalencia definida sobre  $\Omega$ . Se trata de encontrar una relación, definida por un conjunto de atributos  $P$ , en la que, en la medida de lo posible, los elementos de  $Y_j$  estén relacionados con elementos de  $Y_j$  y que los elementos extraños a  $Y_j$ , que queden relacionados con elementos de  $Y_j$ , sean los menos posibles.

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE COSTA RICA, CARTAGO, COSTA RICA

<sup>2</sup>PIMAD, ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

La construcción de una aproximación de  $Y_j$  permite también eliminar objetos y atributos redundantes en la caracterización de este conjunto. La reducción de atributos permite crear una dependencia entre un conjunto mínimo de atributos y un conjunto de 'decisiones' (variables explicativas y variables a explicar), lo que se constituye en un 'modelo' para  $Y_j$ .

La presentación que sigue está basada en los desarrollos teóricos de Z. Pawlak y sus colaboradores [8], [9], [10], [11], [12], [13], [15].

La implementación computacional, así como una descripción de los datos de los ejemplos se pueden ver en [3].

## 1.1 Conceptos básicos de conjuntos aproximados

Un conjunto aproximado es un modelo matemático, el cual usaremos para trabajar con clasificación aproximada. Existe una teoría determinística de conjuntos aproximados y una teoría probabilística. En ambos modelos, sin embargo, el objetivo central es caracterizar un concepto  $Y_j$  en términos de los conceptos (elementales)  $P_1, P_2, \dots, P_s$  que constituyen una partición de  $\Omega$  inducida por un subconjunto de atributos  $P$ .

### 1.1.1 Relación de indiscernibilidad

Sea  $\Omega$  el conjunto de  $n$  objetos o patrones representado en la tabla de datos dada,  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ .

Sea  $Y = \{Y_1, Y_2, \dots, Y_k\}$  una partición de  $\Omega$ , en  $k$  clases, observadas o definidas por un experto. Cada  $Y_j$  es un concepto. Por ejemplo, si  $\Omega$  es un conjunto de pacientes,  $Y_1$  podría denotar el concepto 'pacientes con recuperación satisfactoria', etc.

Sea  $X$  el conjunto de  $p$  atributos *cualitativos* definido en la tabla de datos:  $X = \{x^1, x^2, \dots, x^p\}$ ;  $x^j(\omega_h)$  es el valor del atributo  $x^j$  en el individuo  $\omega_h$ .

Dado  $P \subseteq X$ ,  $P \neq \emptyset$ , se define la relación  $\tilde{P}$  de la siguiente manera:

$$\omega_i \tilde{P} \omega_h \iff \forall x^j \in P; x^j(\omega_i) = x^j(\omega_h).$$

La relación  $\tilde{P}$  es una relación de equivalencia. La partición que induce la relación  $\tilde{P}$  sobre  $\Omega$  se denotará  $P^*$ . Las clases de equivalencia se denotarán con corchetes, v.gr.  $[\omega_i]_{\tilde{P}}$  es la clase del objeto  $\omega_i$ .

### 1.1.2 Aproximación de conjuntos

Dado un conjunto de atributos  $P \subseteq X$ ,  $P$  induce una aproximación de  $Y_j \in Y$ . En realidad, la aproximación consta de dos conjuntos, una aproximación inferior, denotada  $\underline{P}(Y_j)$ , y una aproximación superior denotada  $\overline{P}(Y_j)$ .  $\underline{P}(Y_j)$  es el conjunto de objetos de  $\Omega$  que pueden ser clasificados con certeza como pertenecientes a  $Y_j$ , usando el conjunto de atributos  $P$  y  $\overline{P}(Y_j)$  es el conjunto de objetos de  $\Omega$  que *posiblemente* pueden ser clasificados como pertenecientes a  $Y_j$ , usando el conjunto de atributos  $P$ .

Al conjunto  $Fr_P(Y_j) = \overline{P}(Y_j) - \underline{P}(Y_j)$  se le llama la  $P$ -frontera de  $Y_j$ . También es llamada *región de duda*.

Formalmente,  $\forall Y_j \in Y$ :

$$\underline{P}(Y_j) = \{\omega \in \Omega \text{ tal que } [\omega]_{\tilde{P}} \subseteq Y_j\},$$

$$\overline{P}(Y_j) = \{\omega \in \Omega \text{ tal que } [\omega]_{\tilde{P}} \cap Y_j \neq \emptyset\}.$$

Claramente  $\underline{P}(Y_j) \subseteq \overline{P}(Y_j)$ . Si la partición es muy 'gruesa' puede pasar que  $\underline{P}(Y_j) = \emptyset$ . En la figuras 1, se ilustran los conceptos de aproximación superior e inferior.

*Ejemplo 1.:* Consideremos la siguiente tabla de datos dada en la tabla 1.

Tabla 1.

$\Omega$	Atributos			
	$x^1$	$x^2$	$x^3$	$x^4$
$\omega_1$	0	1	1	0
$\omega_2$	2	1	0	1
$\omega_3$	0	0	0	1
$\omega_4$	1	0	2	0
$\omega_5$	0	1	1	0
$\omega_6$	2	1	0	1
$\omega_7$	0	1	1	0
$\omega_8$	1	0	2	1
$\omega_9$	0	0	0	1

Tabla 1: Tabla de datos

Si  $P = \{x^2, x^3, x^4\}$  entonces  $P^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_6\}, \{\omega_3, \omega_9\}, \{\omega_4\}, \{\omega_8\} \}$   
 Sea  $Y_1 = \{\omega_1, \omega_3, \omega_7\}$ , entonces:

- $\underline{P}(Y_1) = \emptyset$ ,
- $\overline{P}(Y_1) = [\omega_1] \cup [\omega_3] = \{ \omega_1, \omega_5, \omega_7, \omega_3, \omega_9 \}$ .

Como se ve, respecto a  $P$ , los elementos de  $Y_1$  no tienen certeza de ser bien clasificados: la relación  $\tilde{P}$  relaciona 'muchos' elementos de  $Y_1$  con elementos extraños a  $Y_1$ .

Si consideramos  $Y_2 = \{ \omega_2, \omega_3, \omega_4, \omega_6 \}$  entonces:

- $\underline{P}(Y_2) = \{\omega_2, \omega_6\} \cup \{\omega_4\}$ ,
- $\overline{P}(Y_2) = \{\omega_2, \omega_6\} \cup \{\omega_4\} \cup \{\omega_3, \omega_9\}$ .

Como se ve, la relación  $\tilde{P}$  caracteriza muy bien los elementos de  $Y_2$  pues "casi siempre" relaciona elementos de  $Y_2$  con elementos de  $Y_2$ .  $\square$

### 1.1.3 Calidad de la aproximación

La calidad de la aproximación del conjunto  $Y_j$  respecto al conjunto de atributos  $P$  se define como:  $\mu_P(Y_j) = \frac{\text{card } \underline{P}(Y_j)}{\text{card } \overline{P}(Y_j)}$ . Claramente  $0 \leq \mu_P(Y_j) \leq 1$ .  $\mu_P(Y_j)$  expresa la proporción entre los objetos correctamente clasificados por  $P$  y todos los objetos que tienen alguna posibilidad de ser clasificados en  $Y_j$  vía  $P$ .

Del ejemplo 1 de la sección anterior, tendríamos que  $\mu_P(Y_1) = \frac{0}{5} = 0$  y  $\mu_P(Y_2) = \frac{3}{5} = 0.6$ .

En el enfoque probabilístico, tratado más adelante, se usa una función de entropía para medir la caracterización del concepto  $Y_1$  por la partición  $P^*$ .

## 1.2 Caracterización aproximada

Sea  $Y = \{Y_1, Y_2, \dots, Y_k\}$  una clasificación (disjunta) de  $\Omega$ , en  $k$  clases. La  $P$ -aproximación inferior y superior de  $Y$  se definen como  $\underline{P}(Y) = \{\underline{P}(Y_1), \underline{P}(Y_2), \dots, \underline{P}(Y_k)\}$  y  $\overline{P}(Y) = \{\overline{P}(Y_1), \overline{P}(Y_2), \dots, \overline{P}(Y_k)\}$ , respectivamente.

La calidad de la aproximación de la clasificación  $Y$ , por el conjunto de atributos  $P$ , está dada por:

$$\mu_P(Y) = \frac{\sum_{i=1}^k \text{card } \underline{P}(Y_i)}{\text{card } \Omega},$$

que es una proporción entre los objetos correctamente clasificados, vía  $P$ , y todos los objetos de  $\Omega$ .

Pawlak también propone como índice la exactitud de la aproximación de la clasificación  $Y$ , por el conjunto de atributos  $P$ , mediante:

$$\gamma_P(Y) = \frac{\sum_{i=1}^k \text{card } \underline{P}(Y_i)}{\sum_{i=1}^k \text{card } \overline{P}(Y_i)}$$

### 1.2.1 Reducción de atributos

Decimos que un atributo  $x^j$  de  $P$ , es redundante en  $P$  si  $P$  y  $P' = P - \{x^j\}$  producen la misma partición de  $\Omega$ . En este caso,  $\text{card}(P^*) = \text{card}(P'^*)$  y, por supuesto,  $\mu_P(Y) = \mu_{P'}(Y)$ .

Los conjuntos de atributos que no tienen atributos redundantes se llaman *independientes*, i. e.  $P$  es independiente si y sólo si para todo  $P' \subset P$ , se tiene que  $\text{card} P'^* < \text{card} P^*$ . En otro caso, el conjunto se llama *dependiente*. Los subconjuntos de los conjuntos independientes producen particiones más 'gruesas', pues hacen fusiones de clases enteras (no mezcla partes de clases).

*Ejemplo 2.* Respecto a la Tabla 1, el conjunto  $P = \{x^2, x^3, x^4\}$  es independiente. Un cálculo rápido deja ver que:

- $P^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_6\}, \{\omega_3, \omega_9\}, \{\omega_4\}, \{\omega_8\} \}$  ,
- $\{x^2, x^3\}^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_6\}, \{\omega_3, \omega_9\}, \{\omega_4, \omega_8\} \}$  ,
- $\{x^2, x^4\}^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_6\}, \{\omega_3, \omega_8, \omega_9\}, \{\omega_4\} \}$  ,
- $\{x^3, x^4\}^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_3, \omega_6, \omega_9\}, \{\omega_8\}, \{\omega_4\} \}$  ,
- $\{x^2\}^* = \{ \{\omega_1, \omega_2, \omega_5, \omega_6, \omega_7\}, \{\omega_3, \omega_4, \omega_8, \omega_9\} \}$  ,
- $\{x^3\}^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_3, \omega_6, \omega_9\}, \{\omega_4, \omega_8\} \}$  ,
- $\{x^4\}^* = \{ \{\omega_1, \omega_4, \omega_5, \omega_7\}, \{\omega_2, \omega_3, \omega_6, \omega_9, \omega_8\} \}$  .

En cambio, el conjunto  $P_1 = \{x^1, x^2, x^4\}$  es dependiente pues:

$$P_1^* = \{x^1, x^2\}^* = \{ \{\omega_1, \omega_5, \omega_7\}, \{\omega_2, \omega_6\}, \{\omega_3, \omega_9\}, \{\omega_4\}, \{\omega_8\} \}$$

Obsérvese que  $\{x^2, x^3, x^4\}$  y  $\{x^1, x^2\}$  producen la misma partición de  $\Omega$ , pero ambos son independientes  $\square$

Sean  $P, R \subseteq X$ , se dice que el conjunto de atributos  $R$  *depende* del conjunto de atributos  $P$  si se da la inclusión de relaciones  $\tilde{P} \subseteq \tilde{R}$ , es decir, si  $\tilde{R}$  es más 'gruesa' que  $\tilde{P}$ . Esto también se denota  $P \rightarrow R$ . En los conjuntos independientes, los atributos no dependen entre sí, es decir,  $\forall i, j : \{x^i\} \not\rightarrow \{x^j\}$ .

$P' \subset P$  es una *reducción* del conjunto de atributos  $P$  si es un conjunto *maximal* en  $P$ , es decir, si  $P'$  es independiente y no está contenido en otro independiente (de  $P$ ). Por supuesto, el conjunto de atributos  $X$ , definido por la tabla de datos, puede tener más de una reducción.

En la práctica, nos interesa lo que se denomina el *conjunto mínimo*: el 'más pequeño' conjunto independiente de  $X$  que asegura la misma calidad de la clasificación  $Y$  hecha por las reducciones de  $X$ . Igual que las reducciones, pueden haber varios conjuntos minimales. En el ejemplo 2,  $\{x^2, x^3, x^4\}$  y  $\{x^1, x^2\}$  son reducciones, pero este último es un conjunto minimal.

Toda la presentación anterior corresponde al enfoque llamado determinístico de los conjuntos aproximados. En \*\* se puede encontrar un tratamiento para tablas de datos incompletos. Para este caso, con la información disponible, un experto completa los datos de tal manera que un individuo  $x_i$  puede estar representado por uno o más juegos de vectores  $(a_1, \dots, a_N)$ , de acuerdo a las posibles situaciones que se puedan presentar a la luz de la 'poca' información disponible. En este contexto las 'clases' se traslapan, pero esto no afecta las definiciones de aproximación superior e inferior.

## 2 Resultados numéricos

Se ha aplicado la implementación computacional a varios juegos de datos [3].

1. Dos conjuntos de datos simulados [17].
2. Datos zoológicos [1].
3. Datos de los almacenes [4].
4. Datos de los peces de Amiard [2].

En particular

- a. Datos de cáncer de pecho: son datos provenientes de M. Zwitter & M Soklic<sup>3</sup>. La tabla contiene observaciones sobre 9 atributos explicativos sobre 286 individuos que han tenido cáncer del pechi divididos en 2 clases: con cáncer recurrente (85 pacientes) y no recurrente (201 pacientes). Los atributos explicativos son: edad (6 modalidades), edad de la menopausia (3 mods.), tamaño del tumor (11 mods.), *inv. nodes* (7 mods.), *node caps* (2 mods.), grado de malignidad (3 mods.), pecho (2 mods.), localización (5 mods.) y radiación (2 mods.). Nosotros seleccionamos al azar una muestra de 70 pacientes para crear el archivo de aprendizaje, dejando el resto de los pacientes para el archivo de test.
- b. Datos de los pintores [6]: en 1708, Roger de Piles, miembro de la Academia Francesa de Pintura y Escultura, publicó la obra "Balance de los Pintores", que, entre otras críticas sobre la pintura de su época, contiene el juicio del experto sobre 56 pintores según cuatro criterios: composición, dibujo, color y expresión, dado en forma de nota en la escala de 0 (mínimo) a 20 (máximo). Los pintores están divididos en 8 clases, según la escuela a la que pertenecen: Renacimiento, Manierista, Seicento, Veneciano, Lombardo, XVI, XVII, Francés. En vista del carácter cuantitativo de las variables explicativas, nosotros las codificamos en 3 clases cada una, usando un método objetivo óptimo [5], usando el algoritmo de Fisher de particionamiento con una sola variable.

En la tabla 2 presentamos los principales resultados obtenidos, con las seis tablas de datos. La tabla contiene la información del número de objetos, número de atributos explicativos, número de clases a explicar, número de reducciones obtenidas, número de atributos por reducción, número de clases de objetos para la primera reducción, exactitud de la aproximación para la primera reducción y calidad de la aproximación para la primera reducción.

En la tabla 2 se puede apreciar que hay dos tipos de resultados. Por una parte, las tablas con bastantes atributos, para las que se reducen los atributos mediante la metodología estudiada y que forman bastantes clases; esto es normal ya que entre más variables explicativas menos chance tienen dos individuos de ser "indiscernibles". En tales casos se

<sup>3</sup>University Medical Centre, Institute of Oncology, Ljubljana, Eslovenia.



Tabla de datos	Características de las tablas			Características de las reducciones			Resultados para la 1ª reduc.	
	# objetos	# atrib.	# clases	# reduc.	# atrib por reduc.	# clases para 1ª reduc.	exactitud de la aprox.	calidad de la aprox.
simulados 1	100	3	2	1	2	4	0.25	0.40
simulados 2	20	5	2	1	4	6	0.667	0.80
zoológicos	101	16	7	5	11	59	1.0	1.0
felinos	30	14	4	6	6 ó 8	26	1.0	1.0
vagotomía	77	11	4	2	9	74	0.949	0.974
cáncer de pecho	70	9	2	3	7	70	1.0	1.0
almacenes	100	4	2	1	4	15	0.25	0.4
pintores	54	4	8	1	4	28	0.147	0.314
Amiard	23	16	3	25	4,5 ó 6	23	1.0	1.0

Tabla 2: Características de las tablas de datos analizadas y resultados obtenidos para la primera reducción de cada ejemplo, mediante conjuntos aproximados

obtiene buena calidad en la aproximación (incluso, en 4 casos la aproximación es exacta y en los datos de vagotomía es casi exacta). Por otra parte, para los conjuntos con cuatro o menos atributos los resultados son muy pobres. En efecto, entre menos atributos se tengan menos chance habrá de encontrar reducciones (en los datos de almacenes y pintores, no se encontró ni siquiera una reducción), lo que significa que todos los atributos explicativos son significativos. Quedan clases muy numerosas (por simple combinatoria) y por lo tanto la calidad de la aproximación debe ser pobre. Nosotros pensamos que quizás los modelos probabilísticos de conjuntos aproximados puedan servir para mejorar la caracterización de conjuntos de datos con estas características.

Queremos destacar que los resultados que nosotros obtenemos sobre los datos zoológicos y de felinos concuerdan con los resultados obtenidos por otros investigadores [1, 17]. Los resultados son también muy buenos para la segunda tabla de datos simulados y los datos de vagotomía, lo cual también concuerda con otros resultados [17, 13]. Finalmente, los resultados obtenidos para los datos de cáncer de pecho hacen pensar la utilidad de este método para la caracterización de pacientes con esta enfermedad. Una descripción más detallada de estos resultados y sus interpretaciones será próximamente abordada.

Presentamos a continuación un ejemplo de resultados obtenidos, tal como se obtienen en el programa (excepto la formación de clases). En particular, se puede apreciar la exactitud de la aproximación para cada una de las clases a explicar. La formación de las clases para las reducciones se puede apreciar en [3]

## 2.1 Datos de cáncer de pecho

\*\*\*\*\* RESUMEN DE RESULTADOS \*\*\*\*\*

NOMBRE DEL ARCHIVO DE DATOS: brcanc2.dat

NUMERO DE INDIVIDUOS 70

NUMERO DE ATRIBUTOS 9

HAY 3 REDUCCIONES

1 1 1 1 0 1 1 1 0  
 1 1 1 0 1 1 1 1 0  
 1 1 1 0 0 1 1 1 1

CADA REDUCCION TIENE 70 CLASES

EXACTITUD DE LA APROXIMACION PARA CADA CLASE

	Clase #	P_Inf	P_Sup	Exactitud
==>	[1] 48	4.8000000000E+01	4.8000000000E+01	1.0000000000E+00
==>	[2] 22	2.2000000000E+01	2.2000000000E+01	1.0000000000E+00

EXACTITUD DE LA APROXIMACION ==> 1.0000000000E+00

CALIDAD DE LA APROXIMACION ==> 1.0000000000E+00

## 2.2 Datos de los pintores

\*\*\*\*\* RESUMEN DE RESULTADOS \*\*\*\*\*

NOMBRE DEL ARCHIVO DE DATOS: pintor2.dat

NUMERO DE INDIVIDUOS 54

NUMERO DE ATRIBUTOS 4

HAY 1 REDUCCIONES

1 1 1 1

CADA REDUCCION TIENE 28 CLASES

EXACTITUD DE LA APROXIMACION PARA CADA CLASE

	Clase #	P_Inf	P_Sup	Exactitud
==>	[1] 10	4.0000000000E+00	1.8000000000E+01	2.2222222222E-01
==>	[2] 7	2.0000000000E+00	1.3000000000E+01	1.5384615385E-01
==>	[3] 6	0.0000000000E+00	2.1000000000E+01	0.0000000000E+00
==>	[4] 9	4.0000000000E+00	1.4000000000E+01	2.8571428571E-01
==>	[5] 7	2.0000000000E+00	2.0000000000E+01	1.0000000000E-01
==>	[6] 4	3.0000000000E+00	6.0000000000E+00	5.0000000000E-01
==>	[7] 7	2.0000000000E+00	1.4000000000E+01	1.4285714286E-01
==>	[8] 4	0.0000000000E+00	9.0000000000E+00	0.0000000000E+00

EXACTITUD DE LA APROXIMACION ==> 1.4782608696E-01

CALIDAD DE LA APROXIMACION ==> 3.1481481481E-01

### 3 Conclusiones y perspectivas

De acuerdo con los resultados obtenidos hasta el momento, podemos decir que la teoría de conjuntos aproximados es útil para caracterizar una partición de un conjunto de objetos mediante la determinación de un subconjunto de atributos que son necesarios, entre todos los atributos observados sobre los objetos. Destacamos que dentro de nuestro trabajo, desarrollamos una metodología objetiva para codificar una tabla con variables cuantitativas; ésto nos permite tratar tablas con variables cuantitativas a través de una codificación.

Quedan algunos problemas abiertos y que aún no han sido abordados, según la bibliografía consultada. En primer lugar está el tratamiento de datos cuantitativos; además de la codificación optimal que usamos, nosotros creemos que quizás podría pensarse en considerar 'clases' de objetos en el sentido de que  $\omega_i, \omega_h$  están relacionados según un conjunto de variables cuantitativas  $P$  si  $\forall x^j \in P : |x^j(\omega_i) - x^j(\omega_h)| \leq \epsilon_j$ , donde  $\epsilon_j$  es un rango de tolerancia que tendría asociado la variable  $x^j$ , tal que se considere los valores como 'parecidos' (este valor de  $\epsilon_j$  podría calcularse a partir de una medida de dispersión, tal como la desviación estándar, la desviación cuartil o la amplitud). En segundo lugar, está la definición de una red neuronal a partir de las caracterizaciones de las clases obtenidas a través de los conjuntos aproximados, siguiendo la idea de Sethi y desarrollada por Lechevallier [7], que construye una red neuronal a partir de un árbol de decisión.

### Bibliografía

- [1] Albert, O.; Brito, P.; Bertrand, P.; Diday, E.; Jacques, G. (1991) *Logiciel de Classification Pyramidale*. ILOG, Rapport de fin d'étude.
- [2] Cailliez, F.; Pagès, J.-P. (1976) *Introduction à l'Analyse des Données*. SMASH, Paris.
- [3] Espinoza, J.L.; Figueroa, G.; Mora W.; Trejos, J. (1996) "Estudio de problemas de programación matemática mediante técnicas estocásticas y neuronales", Primer informe de investigación. Enero, 1996. Instituto Tecnológico de Costa Rica.
- [4] Celeux, G.; Nakache, J.-P. (1994) *Analyse Discriminante sur Variables Qualitatives*. Polytechnica, Paris.
- [5] Espinoza, J.L. (1995) "Partición óptima: el algoritmo de Fisher", en Memorias del Seminario de Maestría *Métodos de Clasificación y Discriminación*, Universidad de Costa Rica-Instituto Tecnológico de Costa Rica, noviembre 1995.
- [6] Jambu, M. (1978) *Classification Automatique pour l'Analyse des Données*. Dunod, Paris.
- [7] Lechevallier, Y. (1995) "Construcción eficaz de una red neuronal a partir de un árbol de decisión", en Memorias VII-VIII Simposios *Métodos Matemáticos Aplicados a las Ciencias*, W. Castillo & J. Trejos (eds.), Editorial de la Universidad de Costa Rica, pp. 53-74.

- [8] Marek, W.; Pawlak, Z. (1976) "Mathematical storage and retrieval systems: mathematical foundations", *Theoretical Computer Science* **1**, pp. 331-354.
- [9] Mrózek, A. (1990) "Rough sets and dependency analysis among attributes in computer implementations of expert's inference models", en B. Gaines & J. Boose (eds.) *Machine Learning and Uncertain Reasoning*, Academic Press (Series: Knowledge-Based Systems), **3**, pp. 243-259.
- [10] Pawlak, Z. (1982) "Rough sets", *Int. J. of Computer and Information Sciences*, **11**(5), pp. 341-356.
- [11] Pawlak, Z. (1984) "Rough sets and decision tables", LNCS 84, pp. 187-196.
- [12] Pawlak, Z. (1984) "On learning - A rough sets approach", LNCS 84, pp. 197-227.
- [13] Pawlak, Z.; Slowinski, K.; Slowinski, R. (1986) "Rough classification of patients after highly selective vagotomy for duodenal ulcer", *Int. J. Man-Machine Studies*, **24**, 413-433.
- [14] Pawlak, Z.; Wong, S.K.M.; Ziarko, W. (1990) "Rough sets: probabilistic versus deterministic approach", en B. Gaines & J. Boose (eds.) *Machine Learning and Uncertain Reasoning*, Academic Press (Series: Knowledge-Based Systems), **3**, pp. 227-241.
- [15] Pawlak, Z. (1991) *Rough Sets. Theoretical Aspects of Reasoning About Data*. Kluwer Acad. Publ., Dordrecht.
- [16] Slowinski, R.; Stefanowski, J. (1989) "Rough classification in incomplete information systems", *Mathl. Comput. Modelling*, Vol. **12**(10-11), pp.1347-1257.
- [17] Trejos, J. (1994) *Contribution à l'Acquisition Automatique de Connaissances à Partir de Données Qualitatives*. Thèse de Doctorat, Université Paul Sabatier, Toulouse.

## SOBRE MULTIRRESOLUCIONES, ELEMENTOS FINITOS Y UN PROBLEMA DE DINÁMICA DE FLUIDOS<sup>1</sup>

MIJAIL GUILLEMARD<sup>2</sup>

### Resumen

Se presenta un problema ilustrativo de dinámica de fluidos y algunas bases para resolverlo numéricamente usando elementos finitos. Un método alternativo se prepara con ideas de multirresoluciones y ondeletas.

### 1 Introducción

Este artículo es el primero de una posible serie sobre las aplicaciones de multirresoluciones y ondeletas a la solución numérica de un problema de dinámica de fluidos. La meta es comparar, con un ejemplo, los diferentes métodos de solución basados en elementos finitos y multirresoluciones. Estas notas describen el problema, notaciones y una conexión entre las multirresoluciones y elementos finitos. La motivación del problema está en el estudio práctico del efecto que tiene un emisario submarino en el ambiente.

Un emisario submarino consiste en una serie de tuberías que descargan desperdicios orgánicos en el fondo marino. Al proponer la construcción de este tipo de estructuras es necesario estudiar el efecto que estas tienen en el ambiente. Un primer paso es simular el comportamiento de las descargas, para esto hay que describir el movimiento que produce la descarga de un fluido en otro. La formación del flujo de aguas negras en el agua de mar se le llama *plume*, es una forma cilíndrica que aumenta de diámetro con el tiempo.

Las ecuaciones en derivadas parciales que describen el movimiento del plume se basan en los principios de conservación. Hay varios métodos que se usan para resolver estas ecuaciones, algunos se basan en diferencias finitas y elementos finitos. En años recientes han surgido nuevas herramientas basadas en ideas de ondeletas y multirresoluciones para la resolución numérica de las ecuaciones en derivadas parciales.

En la sección 1 se revisan algunos principios de conservación y las ecuaciones que se derivan. En la sección 2 se describen rápidamente, para referencia futura, un par de modelos ya establecidos. La sección 3 presentan definiciones básicas y una conexión entre elementos finitos y multirresoluciones.

<sup>1</sup>Este trabajo fue posible gracias al apoyo ofrecido por el Programa de Estudios Ambientales de la Escuela de Biología de la Universidad de Costa Rica.

<sup>2</sup>ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA; MGUILLEM@CARIARI.UCR.AC.CR

## 2 Principios de conservación

Hay varios principios de conservación que se usan para describir las ecuaciones de un fluido, algunos de estos son los principios de conservación de masa, momento y concentración. Antes de verlos se describe algo de notación.

Siguiendo [1] y [2], se define a  $\Omega$  como una región abierta de  $\mathbb{R}^n$  ( $n$  usualmente es 2 ó 3) con frontera localmente lipschitziana, i.e., todo punto de la frontera tiene una vecindad que es localmente una función lipschitziana. La razón de esta definición es permitir flexibilidad para incluir regiones con esquinas, por ejemplo un fluido en un cubo. La función escalar  $\rho(t, x)$  es la densidad del punto  $x \in \mathbb{R}^n$  en el momento  $t$ ;  $\phi(t, x)$  la función de flujo del fluido: la posición de la partícula  $x$  en el momento  $t$  se denota  $\phi_t(x) = \phi(t, x)$ , el camino que sigue una partícula  $x$  es  $\sigma(t) = \phi_t(x)$ . Se denota el vector velocidad por  $v(t, x) = \frac{d\phi}{dt}$ , la concentración por  $c(t, x)$ , la presión por  $p(t, x)$  y un campo vectorial que representa las fuerzas exteriores por  $f(t, x)$ .

El concepto de derivada material de una función escalar  $\psi(t, x)$  aparece a menudo, esta es la derivada usual de  $\psi(t, \phi(t, x))$  en función de  $t$ :  $D\psi/Dt = \partial\psi/\partial t + v \cdot \nabla\psi$ .

### 2.1 Conservación de masa

El principio de conservación de masa dice que el cambio en la masa de una región arbitraria  $W \subseteq \Omega$  depende de la masa que entra por la frontera de  $W$  y que sale de  $\partial W$ . Otra forma de entenderlo es decir que la masa de la región  $W$  permanece constante mientras se mueve, o sea,

$$\frac{d}{dt} \int_{\phi_t(W)} \rho(t, x) dx = 0. \tag{1}$$

El axioma anterior es equivalente a las siguientes condiciones:

$$\frac{D\rho}{Dt} + \rho \operatorname{div} v = 0 \tag{2}$$

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \tag{3}$$

$$\frac{d}{dt} \int_W \rho dx = - \int_{\partial W} \rho v \cdot n ds \tag{4}$$

La equivalencia entre 2 y el principio de conservación de masa 1 se da por un teorema llamado el *teorema de transporte* [1]. Pasar de 2 a 3 es usar la definición de derivada material, la condición 4 es el primer axioma que se menciona y se conecta con los demás al usar Stokes:  $\int_{\partial W} \rho v \cdot n ds = \int_W \nabla \cdot (\rho v) dx$ , y tomar una sucesión de dominios  $W_i$  que convergen a un punto. A las ecuaciones 2 y 3 se les llama *ecuaciones de continuidad*.

### 2.2 Conservación del momento

Un principio análogo a la ley de Newton se aplica en el caso de los fluidos: la variación de la cantidad de momento depende de las fuerzas exteriores que se aplican al objeto y de las tensiones internas. Este es el *principio de tensión de Cauchy*: sea  $\tau(t, x, n)$  un campo

vectorial con  $n$  un vector unitario (las tensiones internas actuando en la superficie del elemento de control), sea  $v(t, x)$  un campo vectorial,  $\rho(t, x)$  una función positiva, y  $f(t, x)$  un campo vectorial; este sistema  $(v, \rho, f, \tau)$  cumple el principio de tensión de Cauchy si:

$$\frac{d}{dt} \int_{\phi_t(W)} \rho v dx = \int_{\phi_t(W)} \rho f dx + \int_{S_t} \tau ds, \quad (5)$$

donde  $S_t$  es la frontera de  $\phi_t(W)$ . La ecuación que se deduce de este principio es conocida como la *forma local del balance del momento, primera ley de Cauchy* o la *ecuación de movimiento*:

$$\rho \frac{Dv}{Dt} = \rho f + \nabla T, \quad (6)$$

donde  $T$  es un tensor llamado *tensor de tensión*, que proviene del campo vectorial  $\tau$ :  $\tau(t, x, n) = T(t, x)n$ , y  $Dv/Dt$  es la derivada material de  $v(t, x)$ .

### 2.3 Fluidos ideales

Un *fluido ideal* es un *fluido incompresible y perfecto*. En un fluido perfecto las tensiones tangenciales en la superficie no existen, la ecuación se simplifica al suponer el vector  $T \cdot n$  paralelo a  $n$  para todo vector unitario  $n$ . Así, el tensor  $T$  puede describirse con ayuda una función escalar  $p(t, x)$  llamada presión del fluido. Un fluido incompresible es un fluido en donde el volumen de una región arbitraria de  $\Omega$  no cambia al trasladarse con el tiempo, un cálculo muestra que esto es equivalente a que  $\text{div } v = 0$ . Las ecuaciones de un fluido ideal resultan ser:

$$\begin{aligned} \rho \frac{Dv}{Dt} &= \rho f - \nabla p && \text{ecuación de movimiento} \\ \frac{D\rho}{Dt} &= 0 && \text{ecuación de continuidad} \\ \text{div } v &= 0 && \text{incompresibilidad} \end{aligned} \quad (7)$$

Algunas veces la ecuación de movimiento aparece en otra forma equivalente, que puede deducirse de la definición de derivada material  $Dv/Dt$  (nótese que  $v$  es un campo vectorial y no una función escalar):

$$\rho \left( \frac{\partial v}{\partial t} + (v \cdot \nabla)v \right) = \rho f - \nabla p, \quad (8)$$

Un cálculo,  $(v \cdot \nabla)v = \frac{1}{2} \nabla v^2 - v \times (\nabla \times v)$ , permite escribir la ecuación de movimiento de otra forma familiar en la literatura:

$$\rho \left[ \frac{\partial v}{\partial t} + \frac{1}{2} \nabla v^2 - v \times (\nabla \times v) \right] = \rho f - \nabla p. \quad (9)$$

### 2.4 Fluidos viscosos, ecuación de Navier–Stokes

En general los fluidos no son perfectos. Hay tensiones tangenciales de entre las tensiones internas, tal cual el movimiento de una sopa espesa con otra mas liviana. Estas tensiones

producen nuevos términos en la ecuación de movimiento que aparecen al suponer que el tensor  $T$  depende de variables como la velocidad, la divergencia de la velocidad, la densidad y otros.

Para un fluido compresible aparecen dos constantes  $\lambda$  y  $\mu$ . La constante  $\mu$  se le llama usualmente *viscosidad dinámica*, la constante  $\lambda$  está sujeta a debates, usualmente es nula, pero no lo es en el caso de velocidades grandes. Las ecuaciones de movimiento que resultan se llaman *ecuaciones de Navier-Stokes*. Para fluidos compresibles se tiene:

$$\rho \frac{Dv}{Dt} = \rho f - \nabla p + (\lambda + \mu)\nabla(\text{div } v) + \mu\Delta v \quad (10)$$

y en el caso de un fluido incompresible:

$$\frac{Dv}{Dt} = f - \frac{1}{\rho}\nabla p + \nu\Delta v, \quad (11)$$

con  $\nu = \mu/\rho$  es la *viscosidad cinética*.

## 2.5 Número de Reynolds y Froude, ecuaciones de Stokes

El concepto de *similitud dinámica* sirve para comparar sistemas que tienen los mismos comportamientos aunque tengan dimensiones físicas diferentes, es de utilidad para hacer experimentos a escala. Se ve en la siguiente sección un ejemplo de lo anterior con el modelo RSB. Para comparar los sistemas se usan varios valores que caracterizan las ecuaciones de movimiento, como el *número de Reynolds* y el *número de Froude*. Los problemas de dinámica de fluidos se separan en dos familias, una en las que el número de Reynolds es pequeño y otros en los que no, la primera es más sencilla que la segunda. En el caso de números de Reynolds pequeños la ecuación de movimiento se simplifica al eliminar los términos  $(v \cdot \nabla)v$  y  $\partial v/\partial t$  de la ecuación de Navier-Stokes, así se obtiene las *ecuaciones de Stokes*:

$$-\nu\Delta v + \nabla p = f \quad (12)$$

$$\nabla v = 0 \quad (13)$$

## 2.6 Conservación de la concentración

El caso que nos ocupa necesita que se estudie la mezcla de dos fluidos en una región  $\Omega$ , se requiere algún principio que describa el comportamiento de la concentración de las especies que están en el ambiente. Simplificamos las cosas al suponer que el cambio de la concentración sólo se produce por el movimiento mecánico de los fluidos.

La idea que se usa es que si sólo hay difusión por movimiento de los fluidos, la variación con respecto al tiempo de la concentración es nula: esto se describe por la ecuación  $Dc/Dt = 0$  [3]. Como esto es una derivada material se obtiene:  $\partial c/\partial t + v \cdot \nabla c = 0$ . Con la ecuación de continuidad se obtiene:

$$\frac{\partial}{\partial t}(\rho c) + v \cdot \nabla(\rho c) + \rho c \nabla \cdot v = 0. \quad (14)$$



### 3 Modelos UM y RSB

Para referencia y comparaciones futuras se mencionan dos de los muchos modelos distribuidos por la agencia de protección ambiental de los Estados Unidos para el estudio de descargas de aguas negras en el océano. El primero se basa en ecuaciones simplificadas que se deducen de los principios de conservación. El segundo es un modelo experimental basado en los números de Froude.

#### 3.1 Modelo UM

Al integrar las ecuaciones 3, 7, 14 sobre un elemento de control —vea [4] y [5]— y con un sistema de coordenadas adecuado, se obtiene un sistema de ecuaciones diferenciales ordinarias con las variables masa  $m$ , velocidad de la corriente  $U$ , velocidad del plume  $V$ , densidad del plume  $\rho$ , densidad del ambiente  $\rho_a$ , vector de gravedad  $g$ , radio del plume  $b$ , espesor del elemento del plume  $h$ :

$$\begin{aligned} \frac{dm}{dt} &= -\rho A_p U + \rho A_t \alpha |V| && \text{Conservación de masa,} \\ \frac{d(mV)}{dt} &= U \frac{dm}{dt} - m \frac{\rho_a - \rho}{\rho} g && \text{Conservación del momento,} \\ \frac{d(m\rho)}{dt} &= \rho_a \frac{dm}{dt} && \text{Conservación de energía.} \end{aligned} \quad (15)$$

La ecuación de conservación de masa incluye el término extra  $\rho A_t \alpha |V|$ . Este término proviene de la *aspiración inducida de Taylor*. La constante  $\alpha$  se le llama a veces *constante de aspiración*. El valor  $A_t$  es el área de la sección del plume en contacto con el ambiente, el área de aspiración de Taylor. Su valor es  $A_t = 2\pi b h$ . El valor del coeficiente de área  $A_p$  es

$$A_p = \pi b \Delta b + 2bh - \frac{\pi}{2} b^2 \frac{\partial \psi}{\partial s} h. \quad (16)$$

Para completar este sistema otras ecuaciones se ocupan, entre ellas una ecuación de estado para el ambiente y para el plume:

$$\begin{aligned} m &= \rho \pi b^2 h && \text{Cálculo para la masa.} \\ \rho &= \rho(T, c) && \text{Ecuación de estado.} \end{aligned} \quad (17)$$

El sistema consiste en tomar diferencias finitas discretizando la variable de tiempo. Entonces las expresiones con derivadas pasan a ser expresiones discretas y se obtiene un sistema recurrente fácil de programar.

#### 3.2 Modelo RSB

Roberts, Snyder y Baumgartner produjeron modelos en laboratorios que simulan la situación del emisario submarino a escala. Los datos que resultan se agrupan en una serie de fórmulas y así se produce el modelo RSB. El modelo se basa en el número de Froude, se calcula el número de Froude de la situación real para predecir con los datos experimentos su evolución. Detalles de los cálculos se pueden encontrar en [6], [7] y [8].

## 4 Elementos finitos y multirresoluciones

### 4.1 Problemas variacionales y ecuaciones diferenciales

El problema de encontrar soluciones de una ecuación diferencial es, en algunas aplicaciones (como en problemas estáticos), una consecuencia de un principio físico descrito en forma variacional. La equivalencia entre la ecuación diferencial y el problema variacional es básica para escoger el algoritmo que lo resuelve. Por ejemplo, a una ecuación lineal  $Lu = f$  se le asocia una funcional  $I(v) = (Lv, v) - 2(f, v)$ , con  $(\cdot, \cdot)$  el producto interno del espacio en que se trabaja. Buscar un mínimo para la funcional  $I$  equivale a resolver la ecuación  $Lu = f$ , es decir invertir el operador  $L$ . La prueba de esto no es difícil suponiendo que  $L$  sea un operador simétrico y positiva. A la ecuación  $Lu = f$  se le llama la *ecuación de Euler* de  $I$ . Otro problema asociado a la ecuación  $Lu = f$  es encontrar  $u$  tal que  $(Lu, v) = (f, v)$  para todo  $v$ , esto es lo que se llama la *forma débil* o la *forma de Galerkin* de la solución del problema.

### 4.2 Elementos finitos

Antes de ver algo sobre elementos finitos se da una lista de algunos espacios de funciones importantes:

$$\begin{aligned} \mathcal{D}(\Omega) &:= \{ f \in C^\infty(\Omega) : \text{sop } f \text{ compacto} \}, \\ W^{m,p}(\Omega) &:= \{ f \in L^p(\Omega) : \partial^\alpha f \in L^p(\Omega), |\alpha| \leq m \}, \\ H^m(\Omega) &:= W^{m,2}(\Omega) \\ W_0^{m,p}(\Omega) &:= \text{Clausura de } \mathcal{D}(\Omega) \text{ en } W^{m,p}(\Omega). \end{aligned}$$

Análogamente, se definen  $\mathcal{D}(\Omega, \mathbb{R}^n)$ ,  $W^{m,p}(\Omega, \mathbb{R}^n)$ , etc., como espacios de funciones de valores vectoriales con componentes en los espacio de  $\mathbb{R}$ . Entonces se define:

$$\begin{aligned} \mathcal{V}(\Omega) &:= \{ u \in \mathcal{D}(\Omega, \mathbb{R}^n) : \text{div } u = 0 \}, \\ H &:= \text{Clausura de } \mathcal{V}(\Omega, \mathbb{R}^n) \text{ en } L^2(\Omega), \\ V &:= \text{Clausura de } \mathcal{V}(\Omega, \mathbb{R}^n) \text{ en } H_0^1(\Omega, \mathbb{R}^n). \end{aligned}$$

Al espacio  $W^{m,p}(\Omega)$  se le llama *espacio de Sobolev*, con  $m \in \mathbb{N}$  y  $1 \leq p \leq \infty$ , es un espacio de Banach con la norma  $(\|u\|_{W^{m,p}(\Omega)})^p = \sum_{|\alpha| \leq m} \|\partial^\alpha u\|_{L^p(\Omega)}^p$ . Aquí  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ ,  $|\alpha| = \alpha_1 + \dots + \alpha_n$  y  $\partial^\alpha u = \partial^{|\alpha|} u / \partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}$ . Al tomar  $p = 2$ ,  $H^m(\Omega)$  es un espacio de Hilbert con el producto escalar definido por  $((u, v))_{H^m(\Omega)} = \sum_{|\alpha| \leq m} (\partial^\alpha u, \partial^\alpha v)$ . Se define a  $\Omega$  como en la Sección 1, aunque nos limitamos ahora al caso  $n = 2$ .

La idea de los elementos finitos es aproximar las soluciones de una ecuación en derivadas parciales por medio de descomposiciones del espacio en donde están [2], [9], [10]. Si el problema es resolver una ecuación  $Lx = y$  con  $y \in K$  dado,  $x \in K$  y  $L: K \rightarrow H$ . Se "aproximan" los espacios  $K$  y  $H$  con ciertos espacios  $K_n$  y  $H_n$  donde el problema  $Lx = y$  sea mas fácil. Un primer concepto que se necesita es el de aproximaciones externas e internas:

Una aproximación externa de un espacio normado  $W$  consta de un espacio normado  $F$ , un isomorfismo  $w: W \rightarrow F$  y una tripleta  $\{W_h, p_h, r_h\}_{h \in \mathcal{H}}$ , con  $W_h$  un espacio normado,  $p_h: W_h \rightarrow F$  una función lineal continua y  $r_h: W \rightarrow W_h$  una función. En el caso de  $F = W$  y  $w$  la identidad tenemos una aproximación interna de  $W$ .

A los operadores  $p_h$  y  $r_h$  se les llama, respectivamente, operadores de prolongación (o extensión) y restricción. Se necesita ahora, algún concepto de error al hacer la aproximación:

**Definición 1** Para  $h \in \mathcal{H}$ ,  $u \in W$  y  $u_h \in W_h$  se dice que:  $\|wu - p_h u_h\|_F$  es el error entre  $u$  y  $u_h$ ,  $\|u_h - r_h u\|_{W_h}$  es el error discreto entre  $u$  y  $u_h$ ,  $\|wu - p_h r_h u\|_F$  es el error de truncaje de  $u$ .

Las ideas de estabilidad y convergencia para estos operadores son:

**Definición 2** El operador de extensión se dice estable si la norma dada por  $\|p_h\| = \sup_{u_h \in W_h} \|p_h u_h\|_F$  con  $\|u_h\|_{W_h} = 1$ , se puede acotar superiormente independientemente de  $h$ .

**Definición 3** Se tiene una aproximación externa convergente si se cumple lo siguiente: Para todo  $u \in W$  se tiene  $\lim_{h \rightarrow 0} p_h r_h u = wu$  en la topología fuerte de  $F$ . Para toda sucesión  $u_h$  de elementos de  $W_h$ ,  $h \rightarrow 0$ , tal que  $p_h u_h$  converge a algún elemento  $\phi$  en la topología débil de  $F$ , tenemos  $\phi = wu$  para algún  $u \in W$ .

### Ejemplo 1: una aproximación interna de $H_0^1(\Omega)$

Antes de ver la aproximación, se define para un  $h \geq 0$  una triangulación admisible de  $\Omega$  al conjunto  $\mathcal{S}_h$ , que es una familia de triángulos  $s$  (en el caso general para  $n \in \mathbb{N}$  se tiene  $n$ -símplices), que cumplen lo siguiente:

1.  $\Omega(h) = \bigcup_{s \in \mathcal{S}_h} s \subset \Omega$ ,
2. Si  $s$  y  $s' \in \mathcal{S}_h \Rightarrow \dot{s} \cap \dot{s}' = \emptyset$ , donde  $\dot{s}$  es el interior de  $s$  y además  $s \cap s'$  es o bien vacío, o bien una cara de ambos triángulos.

Para cada triangulación  $\mathcal{S}_h$  se le asocian varios números,  $\rho(h) = \sup_{s \in \mathcal{S}_h} \rho_s$ ,  $\rho'(h) = \inf_{s \in \mathcal{S}_h} \rho'_s$ ,  $\sigma(h) = \sup_{s \in \mathcal{S}_h} (\rho_s / \rho'_s)$ , donde  $\rho_s$  es el diámetro de la bola más pequeña que contiene a  $s$  y  $\rho'_s$  es el diámetro de la bola más grande que está contenido en  $s$ .

Una triangulación es regular si  $\sigma(h)$  permanece acotado mientras  $\rho(h) \rightarrow 0$ , y si  $\Omega(h)$  converge a  $\Omega$  si para cada conjunto compacto  $K \subset \Omega$ , existe un  $\delta(K) > 0$  tal que  $\rho(h) \leq \delta(K) \Rightarrow K \subset \Omega(h)$ . Se denota a  $\mathcal{H}_\Omega$  al conjunto de triangulaciones regulares de  $\Omega$ .

Para definir la aproximación interna de  $H_0^1(\Omega)$  se da primero una triangulación admisible  $\mathcal{S}_h$  de  $\Omega$ , se define el espacio  $W_h$  como el espacio de funciones continuas con soporte  $\Omega(h)$  y cuyas componentes son polinomios de grado menor o igual a 2 en cada triángulo  $s \in \mathcal{S}_h$ . Su producto escalar está inducido por el de  $H_0^1(\Omega)$ .

Para definir una base de  $W_h$  se recuerda que todo polinomio de orden menor o igual que 2 está únicamente determinado por sus valores en los vértices  $A_i$  y puntos medios  $A_{ij}$

( $i \neq j$ ) de las aristas de cada triángulo  $s$ . La unión de estos puntos se denota  $\mathcal{U}_h$ , y los puntos de  $\mathcal{U}_h$  que están en el interior de  $\Omega(h)$  se denota  $\dot{\mathcal{U}}_h$ . La propiedad anterior implica que existe una única función  $u_h \in W_h$  que toma valores dados en los puntos  $M \in \dot{\mathcal{U}}_h$ . Construyendo con esto funciones  $w_{hM}$  tal que  $w_{hM}(M) = 1$ ,  $w_{hM}(P) = 0$ , para todo  $P \in \dot{\mathcal{U}}_h$ ,  $P \neq M$ , ( $M \in \dot{\mathcal{U}}_h$ ), se obtiene una base de  $W_h$ , es decir  $u_h = \sum_{M \in \dot{\mathcal{U}}_h} u_h(M) w_{hM}$  para toda función  $u_h \in W_h$ .

Se define el operador de extensión por la identidad:  $p_h u_h = u_h$ , para todo  $u_h \in W_h$  y el operador de restricción para una función  $u \in \mathcal{D}(\Omega)$  por  $r_h u(M) = u(M)$ , para todo  $M \in \dot{\mathcal{U}}_h$ . Finalmente se tiene la siguiente propiedad: la aproximación interna de  $H_0^1(\Omega)$  definida en lo anterior es estable y convergente si  $h$  está en una triangulación regular de  $\Omega$ . Para probar esto se acota la expresión  $\|p_h r_h u - u\|$  y se obtiene para una triangulación regular:  $\|p_h r_h u - u\| \leq c(u) \rho^2(h) \sigma(h) \leq c(u) a \rho^2(h)$ , para un par de constantes  $a$  y  $c(u)$  que dependen de la condición de la triangulación regular y de la función  $u$ .

### Ejemplo 2: aproximación externa de $V$

Para este caso se define primero el espacio  $F$  como  $H_0^1(\Omega)$  y  $w$  es la identidad, el espacio  $V_h$  es un subespacio de  $W_h$  del ejemplo anterior pero con funciones con componentes que son polinomios de grado a lo más 2, tal que en cada triángulo  $s$  de  $\mathcal{S}_h$  se tiene  $\int_s \nabla u_h = 0$ . Esto último es la condición discreta de  $\nabla u = 0$ . El operador de extensión es, como antes, la identidad, ya que  $V_h \subset H_0^1(\Omega)$ . El operador de restricción es más delicado por las condiciones extras: es  $r_h u = u_h = u_h^1 + u_h^2$ , para  $u \in \mathcal{V}$ , con  $u_h^1$  definido por  $u_h^1(M) = u(M)$  para todo  $M \in \dot{\mathcal{U}}_h$ , y  $u_h^2$  es una corrección para que  $r_h u$  esté en  $V_h$ . Se puede probar [2] que con estas últimas definiciones esta aproximación externa de  $V$  es estable y convergente si  $h$  está en una triangulación regular  $\mathcal{H}_\Omega$  de  $\Omega$ .

### 4.3 El problema de Stokes

Volviendo al caso de las ecuaciones de Stokes 12, 13, se describe el problema en forma variacional para después aplicar las aproximaciones de los elementos finitos. Primero note que la ecuación  $-\nu \Delta u + \nabla p = f$  para  $f \in L^2(\Omega)$  se puede escribir al tomar productos escalares con una función  $v \in \mathcal{V}$  y se obtiene  $(-\nu \Delta u + \nabla p, v) = (f, v)$ . Al integrar por partes el término  $(-\Delta u, v)$ , se obtiene  $\sum_i (\nabla u_i, \nabla v_i) = ((u, v))$ , y la parte  $(\nabla p, v)$  pasa a ser  $-(p, \nabla v) = 0$ , así se obtiene:  $\nu((u, v)) = (f, v)$  para  $v \in \mathcal{V}$ . Esta última ecuación es la que se maneja con las aproximaciones por elementos finitos, al utilizar el procedimiento del ejemplo anterior y demostrar que en efecto este produce una sucesión de valores  $u_n$  que convergen a la solución  $u$  de  $\nu((u, v)) = (f, v)$  para  $v \in \mathcal{V}$ , a medida que  $\rho(h) \rightarrow 0$ .

### 4.4 Multirresoluciones

Una multirresolución es una de las herramientas principales en la construcción de las ondeletas [11, 12]. La idea es aproximar un espacio, en este caso  $L^2(\mathbb{R}^n)$ , por subespacios crecientes con propiedades especiales:

**Definición 4** Una multirresolución de  $L^2(\mathbb{R}^n)$  es una sucesión creciente  $V_j$ ,  $j \in \mathbb{Z}$ , de subespacios lineales cerrados de  $L^2(\mathbb{R}^n)$  con las siguientes propiedades:

1.  $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ ,  $\bigcup_{j \in \mathbb{Z}} V_j$  es denso en  $L^2(\mathbb{R}^n)$  para todo  $j \in \mathbb{Z}$ .
2.  $f(x) \in V_j \iff f(2x) \in V_{j+1}$ ,  $\forall j \in \mathbb{Z}$ ;
3.  $f(x) \in V_j \iff f(x - 2^{-j}k) \in V_j$ ,  $\forall k \in \mathbb{Z}^n$ ;
4. existe una función  $g(x) \in V_0$  tal que la sucesión  $\{g(x - k) : k \in \mathbb{Z}^n\}$  es una base de Riesz de  $V_0$ .

Una base de Riesz de un espacio de Hilbert  $H$  es una sucesión  $\{e_k\}_{k \in \mathbb{N}}$  que genera un subespacio denso en  $H$ , tal que hay dos constantes  $A > B > 0$  con  $B\|a\|_{l^2(\mathbb{R})} \leq \|\sum a_k e_k\| \leq A\|a\|_{l^2(\mathbb{R})}$ , para todo  $a \in l^2(\mathbb{R})$ . Estas condiciones son equivalentes a tener un isomorfismo  $T: l^2(\mathbb{N}) \rightarrow H$ . Una multirresolución es  $r$ -regular ( $r \in \mathbb{N}$ ) si la función  $g$  cumple  $|\partial^\alpha g(x)| \leq C_m(1+|x|)^{-m}$ , para todo entero  $m$  y todo multi-índice  $\alpha = (\alpha_1, \dots, \alpha_n)$  con  $|\alpha| \leq r$ .

Hay varios usos para las multirresoluciones, se mencionan dos que interesan en este contexto. Una multirresolución puede servir para construir una base ortogonal de algún espacio (como  $L^2(\mathbb{R}^n)$ ), que se compone de los dilatados y los trasladados de alguna función o familia de funciones  $\psi_1, \dots, \psi_n$ , es decir las funciones  $\Psi_i^{a,b}(x) = \psi_i(ax + b)$  forman una base del espacio (la base de ondeletas). Aquí  $a$  es un factor de escala que puede ser una matriz o un escalar (como  $2^j$ ) y  $b$  un factor de traslación.

Las multirresoluciones también pueden usarse para aproximar funciones de un espacio, la naturaleza de la definición es para eso: los espacios  $V_j$  son los espacios en donde quedan las aproximaciones. Esto ya hace pensar en la analogía entre este concepto y el de elementos finitos. Antes de ver una propiedad que permite ligar las nociones, hay unas propiedades básicas que se necesitan:

**Teorema 1** Sea  $\{V_j : j \in \mathbb{Z}\}$  una multirresolución de  $L^2(\mathbb{R}^n)$ . Entonces hay dos constantes  $c_1 \geq c_2 > 0$  tal que para casi todo  $\xi \in \mathbb{R}^n$  se tiene  $c_1^2 \leq \sum_k |\hat{g}(\xi + 2k\pi)|^2 \leq c_2^2$ , con  $k \in \mathbb{Z}^n$ . Con esto, si además  $\phi \in L^2(\mathbb{R}^n)$  se define por

$$\hat{g}(\xi) = \hat{\phi}(\xi) \left( \sum_{k \in \mathbb{Z}^n} |\hat{g}(\xi + 2k\pi)|^2 \right)^{1/2}, \quad (18)$$

entonces  $\{\phi(x - k) : k \in \mathbb{Z}^n\}$  es una base ortonormal de  $V_0$ .

Otra propiedad que cabe mencionar es que  $(\sum_k |\hat{g}(\xi + 2k\pi)|^2)^{1/2}$  es una función  $C^\infty$  cuando la función  $g$  es  $r$ -regular.

Estas propiedades, entre otras cosas, preparan el camino para fabricar una base de ondeletas del espacio  $L^2(\mathbb{R}^n)$ , pero además permiten construir el puente entre las multirresoluciones y los elementos finitos.

### 4.5 Conexión entre multirresoluciones y elementos finitos

Como se vio antes las ideas de “extensión” y “restricción” son básicas para los elementos finitos. Las propiedades de estos operadores que ahora describimos, conectan y permiten una primera comparación entre las multirresoluciones y elementos finitos.

Tomamos en lo que sigue la región  $\Omega = \mathbb{R}^n$ , aproximándolo por el retículo  $\Gamma = h\mathbb{Z}^n$ , con  $h \geq 0$ . Se puede dar una construcción general de los operadores de extensión y restricción con dos funciones  $\lambda$  y  $\mu$  con las siguientes propiedades:

1. Sea  $\lambda \in L^\infty(\mathbb{R}^n)$  con soporte compacto y  $\int \lambda(x) dx = 1$ ;
2. Sea  $\mu \in H^1(\mathbb{R}^n)$  con soporte compacto,  $\int \mu dx = 1$  y  $\partial^\alpha \hat{\mu}(2k\pi) = 0$  cuando  $k \in \mathbb{Z}^n$ ,  $k \neq 0$  y  $0 \leq |\alpha| \leq r$ ;
3.  $\mu$  y  $\lambda$  se relacionan por  $\int \lambda(x)\mu(y)(x-y)^\alpha dx dy = 0$  con  $1 \leq |\alpha| \leq r$ .

Ahora se define la extensión y la restricción por:

$$(r_h u)(hk) = h^{-n} \int \lambda(h^{-1}x - k)u(x) dx \tag{19}$$

$$(p_h u_h)(x) = \sum_{k \in \mathbb{Z}^n} u_h(hk)\mu(h^{-1}x - k) \tag{20}$$

con  $u \in L^2(\mathbb{R}^n)$  y  $k \in \mathbb{Z}^n$  y  $u_h \in l^2(\Gamma)$ .

J. P. Aubin demuestra en [13] que las condiciones anteriores implican que si  $0 \leq s \leq r$  y  $u \in H^s(\mathbb{R}^n)$ , entonces  $p_h r_h(u)$  converge a  $u$  en la norma de  $H^s(\mathbb{R}^n)$  mientras que  $h$  tiende a cero. Finalmente, la conexión de estos operadores con las multirresoluciones  $r$  regulares, es que la función  $g$  que genera la base de Riesz cumple la propiedad de la función  $\mu$  cuando  $g$  es de soporte compacto, se obtiene con esto una nueva visión de los elementos finitos pero con toda la maquinaria extra de las ondeletas que se puede aprovechar. Es posible probar que la función  $\phi$  cumple la propiedad  $\partial^\alpha \hat{\phi}(2k\pi) = 0$  para todo  $k \in \mathbb{Z}^n \setminus \{0\}$  y  $|\alpha| \leq r$ . Al derivar la relación 18 entre  $\phi$  y  $g$ , se obtiene  $\partial^\alpha \hat{g}(2k\pi) = 0$ , es decir, la función  $g$  cumple la condición necesaria para generar una extensión en el sentido de 19. El retículo, que se puede usar, tomando  $h = 2^{-j}$ , es  $\Gamma_j = 2^{-j}\mathbb{Z}^n$ . La extensión  $P_j: l^2(\Gamma_j) \rightarrow L^2(\mathbb{R}^n)$  es dado por  $\alpha(\gamma) \mapsto f(x) = \sum_{\gamma \in \Gamma_j} \alpha(\gamma)g(2^j(x - \gamma))$ .

## 5 Conclusiones

Las últimas secciones hablan solo de una de las técnicas con las que se aproximan soluciones de ecuaciones diferenciales. Los elementos finitos y diferencias finitas resultan ser dos métodos con muy buenos resultados para estudiar casos en que la geometría del problema es complicada (una región  $\Omega$  complicada). Los métodos espectrales dan resultados mas precisos en regiones sencillas. Los métodos mixtos buscan aprovechar y juntar todas estas diferentes técnicas para mejorar las aproximaciones. Las ondeletas, por su naturaleza, son una buena herramienta para aplicarlas a estos métodos mixtos.

En trabajos futuros se prepara lo necesario para dar comparaciones numéricas entre los métodos que se mencionaron en estas notas y ver el camino para dichos métodos mixtos.

## Bibliografía

- [1] T. J. R. Hughes y J. E. Marsden (1976) *A Short Course in Fluid Mechanics*. Publish or Perish. Wilmington, DL.
- [2] R. Temam (1984) *Navier-Stokes equations*. Third edition, Elsevier Science, Amsterdam.
- [3] L. D. Landau y E. M. Lifschitz (1982) *Fluid Mechanics*. Course of Theoretical Physics, Vol. 6, Pergamon Press, London.
- [4] Michael Schatzmann (1979) "A integral model of plume rise", *Atmospheric Environment* **13**: 721-731.
- [5] Walter E. Frick (1984) "Non-empirical closure of the plume equations", *Atmospheric Environment* **18**: 653-662.
- [6] P. J. W. Roberts, W. H. Snyder y D. J. Baumgartner (1989) "Ocean Outfalls I: Submerged wastefield formation", *ASCE Journal of Hydraulic Engineering* **115**: 1-25.
- [7] P. J. W. Roberts, W. H. Snyder y D. J. Baumgartner (1989) "Ocean Outfalls II: spatial evolution of submerged wastefield", *ASCE Journal of Hydraulic Engineering* **115**: 26-48.
- [8] P. J. W. Roberts, W. H. Snyder y D. J. Baumgartner (1989) "Ocean Outfalls III: effect on diffuser design on submerged wastefield", *ASCE Journal of Hydraulic Engineering* **115**: 49-70.
- [9] A.J. Baker (1983) *Finite Element Computational Fluid Mechanics*. Series in Computational Methods in Mechanics and Thermal Sciences, World Scientific, Singapore.
- [10] Mario A. Villalobos A. (1991) *Algunos aspectos analíticos funcionales en la teoría de elementos finitos aplicado fundamentalmente a las ecuaciones del tipo de Navier-Stokes*. Tesis de licenciatura. Escuela de Matemática, UCR.
- [11] Y. Meyer (1992) *Wavelets and Operators*. Cambridge University Press, Cambridge.
- [12] Ingrid Daubechies (1992) *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61, SIAM, Philadelphia, PA.
- [13] J. P. Aubin (1980) *Approximation of elliptic boundary-value problems*. Krieger Publ. Co., Huntington, NY.

## MULTIVARIATE ELLIPTICALLY COUNTEROURED LINEAR MODELS

A. K GUPTA<sup>1</sup>

---

In this article the main results of the Multivariate elliptically countered linear models are considered. Multivariate elliptically countered linear models arise in many areas of application. See the reference given in this paper for some of these applications. Note that applications of Multivariate elliptically countered linear models are not exhausted by applications in physics, multivariate statistical analysis, and the theory of nonordered structures. Now, theory of random matrices is also used in the theory of stability of solutions of stochastic systems, in linear stochastic programming, in molecular chemistry and in the theory of experiment design.

This paper is designed for statisticians, mathematicians, physicists, scientists, and engineers of different specialties, who use random matrices and theoretical-probability methods in their work.

---

<sup>1</sup>BOWLING GREEN STATE UNIVERSITY, OHIO, USA



## OPTIMIZACIÓN CON RECOCIDO SIMULADO PARA EL PROBLEMA DE ASIGNACIÓN CUADRÁTICA Y COMPARACIONES CON OTRAS TÉCNICAS DE SOLUCIÓN

MIGUEL ANGEL GUTIÉRREZ ANDRADE<sup>1</sup> - SERGIO DE LOS COBOS SILVA<sup>2</sup>  
BLANCA ROSA PÉREZ SALVADOR<sup>2</sup>

---

### Resumen

El objetivo de este artículo es presentar la bondad de la técnica de recocido simulado para el problema de asignación cuadrático, así como los lineamientos a seguir para implementarlo en la computadora. Para tal propósito, se presentan tres métodos alternativos de solución del problema: un método exacto, un método heurístico y un método de búsqueda local. El método exacto y el heurístico que se seleccionaron, se reportan en la literatura como los más eficientes, vea [BUR90]. El análisis comparativo de los cuatro algoritmos se hizo en cuanto a tiempo de ejecución, así como a su cercanía con el valor óptimo de la función objetivo, cuando se conoce tal solución. Adicionalmente se prueba para las instancias de Nugent. Para cada algoritmo se reporta la experiencia computacional y al final se hace un análisis comparativo de los cuatro métodos.

**Palabras clave:** recocido simulado, algoritmos heurísticos, problema de asignación cuadrático, optimización matemática, simulación matemática.

### 1 Descripción

Una variante importante del problema de localización clásico, es cuando el costo asociado por colocar una planta en un determinado lugar depende no únicamente de las distancias entre plantas y de las demandas respectivas, sino también de la interacción con otras plantas.

Una formulación matemática de esta problemática fue inicialmente planteada por Kopmans y Beckmann ([KOO57]) y consiste en minimizar

---

<sup>1</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - AZCAPOTZALCO, DEPARTAMENTO DE SISTEMAS, AVENIDA SAN PABLO 180, COL. REYNOSA, C.P. 02200, MÉXICO D.F.; E-MAIL: GAMA@HP9000A1.UAM.MX

<sup>2</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA

$$f(\pi) = \sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{\pi(i)\pi(k)} \quad (1)$$

donde  $\pi$  es una permutación del conjunto  $N = \{1, 2, \dots, n\}$  y  $a_{ik}$ ,  $b_{ik}$  para  $i, k = 1, \dots, n$ , son números reales. Equivalentemente, se desea encontrar una permutación  $\pi$  del conjunto  $N$  que minimice el valor en (1.1).

Una interpretación del problema planteado consiste en suponer que  $n$  módulos se han colocado en un tablero y se han conectado por medio de "alambres". Los módulos deben colocarse de manera que la longitud total del alambre que los conecta sea mínima.

Denote la distancia entre las posiciones  $i$  y  $k$  en el tablero por  $a_{ik}$ . Sea  $b_{jl}$  el número de conexiones entre el módulo  $j$  y el módulo  $l$ . Entonces una permutación  $\pi$  del conjunto  $N = \{1, 2, \dots, n\}$  asigna una posición en el tablero a todo módulo. Si  $j$  se sitúa en la posición  $i$  y  $l$  está situado en la posición  $k$ , el producto  $a_{ik} b_{\pi(i)\pi(k)}$  es la longitud del alambre entre los módulos  $j = \pi(i)$  y  $l = \pi(k)$ . Por lo tanto la longitud total del alambreado es

$$\sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{\pi(i)\pi(k)},$$

y el problema se reduce a

$$\underset{\pi}{\text{minimizar}} \quad z = \sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{\pi(i)\pi(k)}$$

El problema anterior, se conoce comúnmente en la literatura como *el problema de asignación cuadrático* (QAP). Una formulación similar del QAP es aplicable para el diseño de paneles de control y teclados de máquinas. Existen muchas otras aplicaciones del QAP por ejemplo: problemas de ubicación de edificios, planeación de hospitales; en deportes, en problemas de calendarización; en sistemas de información, un orden óptimo para interrelacionar datos en una cinta magnética; en química, para analizar reacciones químicas para componentes orgánicas; en balancear turbinas para autos de carreras; en arqueología, modelos para ordenar datos arqueológicos.

## 2 Algoritmo Exacto

Se conoce que el problema de asignación cuadrático es NP-duro (vea [BUR90]). Por lo tanto, hasta ahora, únicamente se conocen algoritmos de enumeración implícita que resuelven el problema de manera óptima. En esta sección se describe un algoritmo para encontrar las soluciones óptimas que según experiencia computacional de Burkard (vea [BUR90]) es de los algoritmos más eficientes que se conocen para resolver el problema de manera exacta. Este método se conoce como el método de la perturbación.

### 2.1 Descripción del algoritmo

El algoritmo de ramificación y acotamiento se puede describir como sigue: Sea  $M \subseteq N$  donde  $N = \{1, 2, \dots, n\}$ . Denote por  $S_n$  el conjunto de todas las permutaciones  $\pi$  de  $N$ .

La restricción de  $\pi$  a  $M$  se llama *permutación parcial*  $(\pi, M)$ . Además sea  $\phi_M = \{\pi \in S_n | \pi_M = (\pi, M)\}$  denote el conjunto de todas las permutaciones que son extensiones de la permutación parcial  $(\pi, M)$ . Se define  $\pi(M) = \{\pi(i) | i \in M\}$ .

Sea una instancia de un problema de asignación cuadrático con matrices  $A$  y  $B$  y una permutación parcial  $(\pi, M)$  dada. Una permutación  $\pi \in \phi_M$  tiene la función objetivo siguiente:

$$z(\pi) = \sum_{i,p \in M} a_{ip} b_{\pi(i)\pi(p)} + \sum_{i \in M, p \notin M} [a_{ip} b_{\pi(i)\pi(p)} + a_{pi} b_{\pi(p)\pi(i)}] + \sum_{i,p \notin M} a_{ip} b_{\pi(i)\pi(p)}. \quad (2)$$

la primera suma en (2.1) puede calcularse de manera exacta conociendo la permutación parcial  $(\pi, M)$ . Para la segunda y tercera sumatoria en (2.1) se puede encontrar una cota inferior resolviendo un problema de asignación lineal con una matriz de costos  $C = (c_{ij})$  ( $i \notin M, j \notin \pi(M)$ ). Sin embargo la solución óptima de este problema de asignación lineal da sugerencias de como debe hacerse una asignación simple ( $i \rightarrow j$ ) en el paso siguiente. Las coeficientes  $c_{ij}$  se escogen como cotas inferiores para la contribución de la asignación simple  $i$  a  $j$  en el valor de la función objetivo  $z(\pi)$ . Estas cotas inferiores  $c_{ij}$  pueden determinarse de la siguiente manera. En principio

$$a_{ii} b_{jj} + \sum_{p \in M} (a_{ip} b_{j\pi(p)} + a_{pi} b_{\pi(p)j})$$

es la contribución a  $c_{ij}$ , que depende de la permutación parcial  $(\pi, M)$ . Este valor puede mejorarse por un paso de reducción.

Reducción de la matriz A:

$$\begin{aligned} \bar{a}_i &= \min_{p \notin M, p \neq i} a_{ip} & (i \notin M); \\ \bar{a}_{ip} &= a_{ip} - \bar{a}_i & (i, p \notin M; i \neq p); \\ \bar{a}_p &= \min_{i \notin M, i \neq p} \bar{a}_{ip} & (p \notin M); \\ a_{ip} &= \bar{a}_{ip} - \bar{a}_p & (i, p \notin M; i \neq p). \end{aligned}$$

Es fácil mostrar que

$$\bar{a}_i \sum_{q \notin \pi(M), q \neq j} b_{jq} + \bar{a}_i \sum_{q \notin \pi(M), q \neq j} b_{qj} \quad (i \notin M, j \notin \pi(M))$$

es la contribución adicional que aparece en  $c_{ij}$ , si la asignación simple ( $i \rightarrow j$ ) se hace. En forma análoga la matriz  $B$  puede reducirse y esto permite un mejoramiento de la cota. La cota puede aún mejorarse. La idea es que se puede obtener una cota inferior para todos los productos escalares  $\sum_{i=1}^n a_i b_{\pi(i)}$  de dos vectores  $\mathbf{a}$  y  $\mathbf{b}$  no negativos, si el orden de los elementos del vector  $\mathbf{a}$  se toma en forma creciente y el de los elementos del vector  $\mathbf{b}$  se toma en forma decreciente y formamos el producto escalar  $\langle \mathbf{a}, \mathbf{b} \rangle$  de estos dos vectores

ordenados. Si se aplica lo anterior para reducir los elementos  $a_{ip}$  ( $i, p \notin M; i \neq p$ ) y  $b_{jq}$  ( $j, q \notin \pi(M); j \neq q$ ) y se definen los vectores:

$$\begin{aligned} a_i &= (a_{ip} | p \notin M, p \neq i) & i \notin M, \\ b_j &= (b_{jq} | q \notin \pi(M), q \neq j) & j \notin \pi(M), \end{aligned}$$

entonces  $\langle a, b \rangle$  es una contribución adicional a la cota  $c_{ij}$  (llamada la "cota de Gilmore"). Ahora se resuelve un problema de asignación lineal con matriz de costos  $C = (c_{ij})$  ( $i \notin M, j \notin \pi(M)$ ). La experiencia computacional (vea [BUR90]) muestra que estos problemas deben resolverse en forma óptima. Sea  $z_L$  el valor de la función objetivo y  $(\bar{c}_{ij})$  los costos reducidos en la solución óptima del problema de asignación lineal con coeficientes de costo  $c_{ij}$ . Entonces una cota inferior para el valor de la función objetivo en (2.1) de las permutaciones  $\pi \in \phi_M$  es

$$z(\phi_M) = \sum_{i,p \in M} a_{ip} b_{\pi(i)\pi(p)} + z_L.$$

La ramificación en el método de la perturbación viene dada de la siguiente forma: Se inicializa con la permutación parcial, en la que  $M$  es el conjunto vacío. Sea  $z$  una cota superior para el valor óptimo de la función objetivo. Si  $z(\phi_M) < z$  se escoge una pareja de índices  $(i, j)$   $i \notin M, j \notin \pi(M)$  por las reglas de costos alternativos, reemplace  $M$  por  $M \cup \{i\}$  y extienda la permutación parcial por la asignación simple ( $i \rightarrow j$ ). Para esta nueva permutación parcial  $(\bar{\pi}, \bar{M})$  se calcula una nueva cota  $z(\phi_{\bar{M}})$ . Los coeficientes  $c_{ij}$  nuevos pueden obtenerse a partir de la solución óptima del último problema de asignación lineal. Si  $|M| = n - 2$ , el valor de las dos permutaciones correspondientes se calcula de manera exacta y eventualmente la cota  $z$  se mejora. Si  $z(\phi_{\bar{M}}) \geq z$ , se regresa a la última permutación parcial encontrada  $(\pi, M)$  con  $z(\phi_{\bar{M}}) < z$ . La asignación simple ( $i \rightarrow j$ ), seleccionada en la última transición de  $(\pi, M)$  a su sucesor, se elimina y la correspondiente entrada  $(i, j)$  de la matriz  $\bar{C} = (\bar{c}_{ij})$  se bloquea. Si el elemento bloqueado es el primer elemento bloqueado en la matriz  $\bar{C} = (\bar{c}_{ij})$  correspondiente a  $(\pi, M)$ , una nueva asignación simple se escoge en el mismo renglón o columna. Si el renglón  $i$  (o columna  $j$ ) contiene más elementos bloqueados, la nueva asignación simple se toma sobre este renglón índice (columna índice). Si esto no es posible se regresa a la permutación parcial predecesora de la permutación parcial  $(\pi, M)$ . La solución óptima tiene el valor  $z$  cuando el conjunto  $M$  se hace vacío por segunda ocasión.

Las reglas de costos alternativos usados son las siguientes. Sea  $\pi^*$  una solución óptima del problema de asignación lineal con costos  $c_{ij}$  y sean  $(\bar{c}_{ij})$  denoten los costos reducidos en el tableau óptimo. Calcule para todo  $(i, \pi^*(i))$  las expresiones

$$\alpha_i = \min_{j, j \neq \pi^*(i)} \bar{c}_{ij}; \quad \beta_i = \min_{p, p \neq i} \bar{c}_{p\pi^*(i)};$$

el valor  $\alpha_i + \beta_i$  constituye una cota inferior para el incremento de la cota  $z_L$ , si la asignación simple ( $i \rightarrow \pi^*(i)$ ) no se realiza. Se puede escoger la pareja  $(i, \pi^*(i))$  para la siguiente

asignación simple que tiene costos alternativos maximal. Por lo tanto se determina el índice  $i$  que resuelve el problema:

$$\max_{i \notin M} (\alpha_i + \beta_i)$$

y a partir de aquí se define la nueva permutación parcial  $(\bar{\pi}, \bar{M})$ :

$$\bar{M} = M \cup \{i\}$$

$$\bar{\pi}(i) = \pi^*(i)$$

$$\bar{\pi}(p) = \pi(p) \quad \text{para toda } p \in M.$$

## 2.2 Experiencia computacional

La prueba del algoritmo se llevó a cabo generando instancias del problema de asignación cuadrático de manera aleatoria. Para la matriz  $A$  se generaron números aleatorios entre 0 - 100 y para la matriz  $B$  se generaron números aleatorios entre 0 - 50. Para cada valor de  $n = 5, 6, 7, 8, 12$  se generaron 100 instancias. La Tabla 1 reporta el tiempo promedio. Cabe mencionar que para  $n = 13$  el programa tardó más de 6 horas en una instancia generada aleatoriamente, por lo que se decidió cortar la tabla hasta el valor de  $n = 12$ .

n	ALGORITMO EXACTO
	tiempo (seg.)
5	0.08
6	0.25
7	1.03
8	3.98
10	73.45
12	1713.99

Tabla 1.

n	ALGORITMO EXACTO
	tiempo (seg.)
5	0.06
6	0.22
7	0.61
8	2.70
12	520.31

Tabla 2. Problemas de Nugent.

Adicionalmente, para cada uno de los valores anteriores de  $n$  se corrieron los problemas de prueba de Nugent (vea [NUG68]). Los resultados de estas ejecuciones aparecen en la Tabla 2.

## 3 Algoritmo heurístico

Como ya se ha mencionado, el problema de asignación cuadrático es NP-duro y desde el punto de vista práctico, es imposible resolver instancias "grandes" del problema. Existen instancias con  $n = 20$  para las que hasta la fecha no se conoce la solución óptima (vea [NUG68] y [BUR90]), por lo que cuando se requiere resolver problemas de mayor tamaño, se debe recurrir a algoritmos heurísticos que no necesariamente obtienen la solución óptima del problema, pero dan una "buena" solución. Burkard y Stratmann ([BUR78] y [BUR90])

proponen un algoritmo para encontrar soluciones subóptimas al problema que denominan el método de incrementar grado de libertad. Este algoritmo se describe a continuación.

### 3.1 Descripción del algoritmo

Usando la notación de la sección anterior, se comienza con el conjunto vacío  $M$  y se escoge de manera arbitraria un índice  $i \notin M$  y se asigna este índice  $i$  a un índice  $j \in \{1, 2, \dots, n\}$ , tal que  $z(j)$  sea minimal. Para  $j \notin \pi(M)$  se define  $z(j)$  por

$$z(j) = a_{ii}b_{jj} + \sum_{p \in M} [a_{ip}b_{j\pi(p)} + a_{pi}b_{\pi(p)j}] + \sum_{p,q \in M} a_{pq}b_{\pi(p)\pi(q)} \quad (3)$$

y para  $j = \pi(i_0)$ ,  $i_0 \in M$ ;  $k \notin \pi(M)$  se define  $z(j)$  por

$$\begin{aligned} z(j) &= a_{ii}b_{jj} + a_{i_0i_0}b_{kk} + a_{ii_0}b_{jk} + a_{i_0i}b_{kj} \\ &+ \sum_{p \in M \setminus \{i_0\}} [a_{ip}b_{j\pi(p)} + a_{pi}b_{\pi(p)j}] + a_{i_0p}b_{k\pi(p)} + a_{pi_0}b_{\pi(p)k} \\ &+ \sum_{p,q \notin M \setminus \{i_0\}} a_{pq}b_{\pi(p)\pi(q)}, \end{aligned} \quad (4)$$

se define  $m = |M| + 1$ . Para determinar para  $i \notin M$  un índice  $j \in \{1, \dots, n\}$  con valor minimal  $z(j)$ , se tienen que evaluar  $m(n - m + 1)$  sumas de la forma (3.1) o (3.2). Como  $m(n - m + 1)$  se incrementa con  $m = 1, 2, \dots, \lfloor \frac{n}{2} \rfloor$  este método se llama el método de incrementar grados de libertad. La permutación final que se obtiene por este método depende de la sucesión de índices  $i$  que se escojan. Por lo tanto, si se realiza este procedimiento con sucesiones diferentes de índices  $i$  (es decir, diferentes permutaciones de la sucesión básica  $i = 1, 2, \dots, n$ ), se obtendrán diferentes soluciones. Se puede tomar ventaja de esta observación y ejecutar el método para selecciones diferentes de los índices libres  $i \notin M$ . En el programa, la subrutina de asignación se repite REP veces comenzando con permutaciones generadas aleatoriamente para los índices  $i$ .

### 3.2 Experiencia computacional

Se utilizaron las mismas 100 instancias generadas para la prueba del algoritmo exacto para  $n = 5, 6, 7, 8, 10, 12$  y se generaron aleatoriamente 100 instancias para cada valor de  $n = 15, 20, 25, 30$  con las mismas características, ya descritas en el algoritmo exacto, para las matrices  $A$  y  $B$ . La Tabla 3 reporta el tiempo promedio.

ALGORITMO HEURÍSTICO	
n	tiempo (seg.)
5	0.08
6	0.16
7	1.28
8	3.44
10	1.01
12	2.03
15	4.75
20	14.55
25	35.02
30	71.80

Tabla 3.

ALGORITMO HEURÍSTICO			
n	Mejor Solución Conocida	Valor	Tiempo (seg.)
5	50 (óptima)	50	0.11
6	86 (óptima)	86	0.17
7	148 (óptima)	148	0.28
8	214 (óptima)	226	0.44
12	578 (óptima)	612	2.03
15	1150 (óptima)	1202	4.78
20	2570	2774	14.56
30	6124	6630	71.84

Tabla 4. Problemas de Nugent.

Para  $n = 5, 6, 7, 8, 12, 15, 20, 30$  se corrieron los problemas de prueba de Nugent (vea [NUG68]). Los resultados de estas ejecuciones aparecen en la Tabla 4.

## 4 Algoritmo de recocido simulado

El algoritmo de recocido simulado puede usarse para atacar este problema y obtener soluciones que no necesariamente son óptimas pero se puede mostrar un comportamiento muy favorable del método con respecto a instancias que se conoce la solución, así como para aquellas instancias que no se conoce la solución óptima pero se conoce una mejor solución. Para aplicar el algoritmo de recocido simulado se necesitan definir ciertos puntos que a continuación se describen.

### 4.1 Descripción del algoritmo

Como ya se ha mencionado, el espacio de soluciones consiste de todas las permutaciones de  $n$  elementos. Para cada permutación  $\pi$ , se define la función de costo como

$$f(\pi) = \sum_{i=1}^n \sum_{p=1}^n a_{ip} b_{\pi(i)\pi(p)}. \quad (5)$$

La generación de nuevas soluciones se hace a partir de tomar permutaciones que difieran de la actual en únicamente dos cambios. Se comienza con una permutación  $\pi$  y se seleccionan aleatoriamente dos índices  $i$  y  $k$  y a partir de estos índices se define la nueva permutación  $\bar{\pi}$  por

$$\bar{\pi}(i) \equiv \pi(k), \quad \bar{\pi}(k) \equiv \pi(i) \quad (6)$$

$$\bar{\pi}(l) \equiv \pi(l), \quad l = 1, 2, \dots, n, \quad l \neq i, k.$$

La diferencia en la función de costo entre la solución actual y la nueva solución generada a partir del cambio  $i, k$  puede calcularse de la siguiente expresión:

$$\Delta f = \sum_{l=1, l \neq i, k} [(a_{il} - a_{kl})(b_{\pi(k)\pi(l)} - b_{\pi(i)\pi(l)}) + (a_{li} - a_{lk})(b_{\pi(l)\pi(k)} - b_{\pi(l)\pi(i)})] \quad (7)$$

$$+ (a_{ii} - a_{kk})(b_{\pi(k)\pi(k)} - b_{\pi(i)\pi(i)})$$

$$+ (a_{ik} - a_{ki})(b_{\pi(k)\pi(i)} - b_{\pi(i)\pi(k)})$$

donde  $\Delta f$  denota la diferencia en el costo entre la solución propuesta menos la solución actual al hacer el intercambio. Si  $\Delta f < 0$ , entonces  $\bar{\pi}$  tiene un valor en la función objetivo menor que  $\pi$ . Usando estas observaciones se puede establecer el algoritmo de recocido simulado.

El Paso 1 del algoritmo se hace tomando una permutación aleatoria, del valor objetivo de esta permutación aleatoria se toma  $c_0$  igual al 20% del valor objetivo de esta permutación aleatoria y se inicia un ciclo de  $r = 2n$  veces el Paso 6, contabilizando el porcentaje de soluciones más malas que la actual aceptadas. Si este porcentaje es mayor o igual al 80%, entonces se toma el valor  $c_0$  si no es el caso, se incrementa un 10% y se repite el proceso hasta cumplir la condición de aceptación del 80% de soluciones malas. A partir de aquí, se procede con el algoritmo, cada vez que se ejecuta el Paso 7, se decrementa el valor del parámetro de control con  $c \leftarrow 0.9 \times c$  y  $r \leftarrow 1.05 \times r$ . El proceso continúa hasta encontrar un valor de  $c$  en el que, dentro del Paso 6, no se acepte ninguna nueva solución.

### Algoritmo de Recocido Simulado para el Problema de Asignación Cuadrático

**Propósito:** Obtener una permutación  $\pi$  que minimice el valor de la función en (1.1).

#### Descripción:

**Paso 1:** Tome  $c$  un valor grande apropiado.

**Paso 2:** Obtenga una permutación aleatoria  $\pi$  del conjunto  $\{1, 2, \dots, n\}$ .  
Obtenga el valor correspondiente de la función objetivo  $f(\pi)$ .

**Paso 3:** Asigne el valor verdadero a la variable lógica CHANGE.

**Paso 4:** Si la variable CHANGE tiene valor falso, entonces pare.

**Paso 5:** Haga CHANGE igual a falso.

**Paso 6:** Repita los puntos 1-9  $r$  veces.

1.- Tome una permutación arbitraria de dos cambios  $\bar{\pi}$

2.- Evalúe el valor  $\Delta f$  de la expresión (20)

3.- Si  $\Delta f < 0$  entonces vaya a 7.

4.- Sea  $P(\Delta f) = \exp(-\Delta f/c)$ .

5.- Genere un número aleatorio  $x$ , uniformemente distribuido en  $(0, 1)$ .

6.- Si  $x \geq P(\Delta f)$  entonces vaya a 9.

7.- Acepte la nueva permutación  $\pi \leftarrow \bar{\pi}$ ,  $f(\pi) \leftarrow f(\pi) + \Delta f$ .

8.- Si  $\Delta f \neq 0$  entonces CHANGE  $\leftarrow$  verdadero.



9.- Fin del paso 6.

**Paso 7** Reduzca  $c$  e incremente  $r$ .

**Paso 8** Vaya al paso 4.

## 4.2 Experiencia computacional

Se utilizaron las mismas 100 instancias generadas de manera aleatoria para  $n = 5, 6, 7, 8, 10, 12, 15, 20, 25, 30$  que se usaron para el algoritmo heurístico, con las características ya descritas. La Tabla 5 reporta el tiempo promedio. La Tabla 6 reporta los resultados para las instancias de prueba de Nugent.

## 5 Algoritmo de búsqueda local

Uno de los puntos importantes del algoritmo de recocido simulado es el hecho de que cada vez que se ejecuta el **Paso 6** del mismo, no solamente acepta permutaciones  $\pi$  para las que se tenga un mejor valor objetivo, sino que también se aceptan con cierta probabilidad permutaciones cuyo valor objetivo es más malo. Con el objeto de poder comparar los resultados del algoritmo de recocido simulado con el algoritmo de búsqueda local correspondiente, también se implementó este último en la computadora.

n	ALGORITMO REC. SIM.
	tiempo (seg.)
5	1.73
6	2.25
7	2.98
8	3.79
10	6.40
12	12.73
15	15.21
20	20.04
25	28.62
30	43.65

Tabla 5.

ALGORITMO RECOCIDO SIMULADO			
n	Mejor Solución	Valor	Tiempo
	Conocida		(seg.)
5	50 (óptima)	50	0.27
6	86 (óptima)	86	1.27
7	148 (óptima)	148	1.58
8	214 (óptima)	214	1.76
12	578 (óptima)	578	2.86
15	1150 (óptima)	1150	14.17
20	2570	2570	36.80
30	6124	6150	76.13

Tabla 6. Problemas de Nugent.

n	ALGORITMO BUS. LOC.	
	tiempo (seg.)	
5		0.13
6		0.21
7		0.29
8		0.43
10		0.79
12		1.07
15		2.40
20		5.64
25		11.11
30		20.96

Tabla 7.

ALGORITMO BÚSQUEDA LOCAL			
n	Mejor Solución	Valor	Tiempo
	Conocida		(seg.)
5	50 (óptima)	50	0.11
6	86 (óptima)	86	0.22
7	148 (óptima)	148	0.28
8	214 (óptima)	214	0.33
12	578 (óptima)	592	1.32
15	1150 (óptima)	1178	2.41
20	2570	2644	5.27
30	6124	6354	15.81

Tabla 8. Problemas de Nugent.

### 5.1 Experiencia computacional

Se utilizaron las mismas 100 instancias generadas de manera aleatoria para  $n = 5, 6, 7, 8, 10, 12, 15, 20, 25, 30$  que se usaron para los algoritmos anteriores. La Tabla 7 reporta el tiempo promedio empleado por el algoritmo. La Tabla 8 reporta los resultados para las instancias de prueba de Nugent.

## 6 Análisis comparativo de los métodos

Como ya se mencionó, se generaron en forma aleatoria instancias del problema de asignación cuadrático. Para cada una de estas instancias generadas, se corrieron los cuatro algoritmos y se calculó el tiempo promedio de solución. Los resultados aparecen en las Tablas 1, 3, 5 y 7 respectivamente y aparecen para los cuatro algoritmos en la Tabla 9.

Cabe mencionar que el tiempo de solución del algoritmo de recocido simulado es muy variable, ya que depende en gran medida del valor inicial del parámetro de control  $c$ , del número  $r$  de veces que se ejecuta el paso 6 (del algoritmo) y de la forma en que se reduzca el parámetro  $c$  y se incremente el parámetro  $r$ . Para el análisis realizado en este documento, se tomó un valor inicial de  $c$  de tal manera que la tasa de aceptación de las permutaciones propuestas fuera mayor o igual a 0.8 y los decrementos fueran del 0.1 (paso 7 del algoritmo). El número  $r$  inicial que se tomó en cada caso fue  $r = 2 \times n$  y los incrementos fueron de 0.05 (paso 7 del algoritmo). También se calculó la eficiencia de cada uno de los algoritmos por medio de la ecuación (6.1) La ecuación (6.1) se usó para los valores de  $n = 5, 6, 7, 8, 10, 12$ . Para los valores mayores de  $n$ , como no se puede calcular la solución óptima, la eficiencia se calculó tomando como base el valor  $z_{rec}$ , esto es, el valor objetivo que da el algoritmo de recocido simulado y se usó la ecuación equivalente:

$$\text{eficiencia} = 1 - \frac{z - z_{rec}}{z_{rec}} \quad (8)$$

Los cálculos que aparecen en la Tabla 9 son con base en los promedios obtenidos en las 100 instancias resueltas para cada valor de  $n$  de los valores en la función objetivo de la solución correspondiente. En la Tabla 9, puede observarse que las eficiencias del algoritmo de recocido simulado son las más elevadas y los tiempos de ejecución del algoritmo de recocido simulado, crece de manera razonable; incluso, para instancias pequeñas se obtienen tiempos de ejecución más grandes que para el algoritmo heurístico. Esta situación cambia para valores de  $n$  grandes.

Otra característica importante del algoritmo de recocido simulado es que las variaciones en las soluciones obtenidas al resolver una misma instancia con soluciones iniciales diferentes, es mucho menor que las variaciones en las soluciones que dan los otros dos algoritmos, esto es, la solución del algoritmo de recocido simulado es menos dependiente de la solución inicial proporcionada. En cambio para el algoritmo heurístico y más aún, para el algoritmo de búsqueda local, las soluciones dependen en gran medida de la solución con que se inicialice. Este comportamiento se puede observar en la Tabla 10 en donde se muestran los resultados al ejecutar cada uno de los algoritmos 100 veces con soluciones iniciales generadas de manera aleatoria para cada uno de los problemas de Nugent.

n	ALGORITMO							
	EXACTO		HEURISTICO		REC. SIM.		BUS. LOC.	
	eficien.	tiempo (seg.)	eficien.	tiempo (seg.)	eficien.	tiempo (seg.)	eficien.	tiempo (seg.)
5	1.00	0.08	0.9891	0.08	0.9999	1.73	0.9747	0.13
6	1.00	0.25	0.9803	0.16	0.9982	2.25	0.9666	0.21
7	1.00	1.03	0.9767	0.28	0.9936	2.98	0.9676	0.29
8	1.00	3.98	0.9737	0.44	0.9899	3.79	0.9630	0.43
10	1.00	73.45	0.9684	1.01	0.9905	6.40	0.9629	0.79
12	1.00	1713.99	0.9674	2.03	0.9940	12.73	0.9512	1.07
15	—	—	0.9693	4.75	1.0000	15.21	0.9640	2.40
20	—	—	0.9512	14.55	1.0000	20.04	0.9518	5.64
25	—	—	0.9527	35.02	1.0000	28.62	0.9551	11.11
30	—	—	0.9543	71.80	1.0000	43.65	0.9580	20.96

Tabla 9. Comparación de los algoritmos para resolver el QAP con respecto al tiempo de ejecución y a su eficiencia.

n	HEURÍSTICO			REC. SIM.			BUS. LOC.		
	Mejor	Media	Peor	Mejor	Media	Peor	Mejor	Media	Peor
5	50	51.6	58	50	50.4	58	50	55.2	62
6	86	88.5	94	86	86.3	92	86	90.2	94
7	148	153.9	164	148	148.7	156	148	154.5	172
8	214	228	252	214	215.3	222	214	224.5	248
12	578	620.3	662	578	600.2	640	578	610.8	650
15	1188	1231	1294	1150	1172	1214	1150	1204	1290
20	2692	2749	2794	2570	2657	2728	2604	2680	2742
30	6488	6636	6766	6124	6333	6442	6242	6364	6496

Tabla 10. Comparación entre los algoritmos heurístico, recocido simulado y búsqueda local con respecto a la dependencia de la solución inicial

Finalmente se hará mención que todos los programas computacionales y las ejecuciones de los mismos, se realizaron en una computadora 486 de 33 mhz.

## 7 Conclusiones

Las pruebas computacionales con el problema de asignación cuadrático, revelan un excelente comportamiento del algoritmo de recocido simulado. Las soluciones subóptimas difieren en 1-2% de la mejor solución conocida. Además si se ejecuta el procedimiento varias veces con diferentes valores iniciales, se obtienen en todos los casos la solución óptima o las mejores soluciones que se conocen (según sea el caso). Por otra parte, la implementación del algoritmo en la computadora, no requiere de gran capacidad de memoria y permite un tiempo razonable para obtener buenas soluciones en las computadoras personales.

## Bibliografía

- [AAR89] Aarts, E. y Korst, J. (1989) *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, Chichester.
- [BUR78] Burkard, R.E. y Stratmann, K. (1978) "Numerical investigations on quadratic assignment problems", *Naval Research Logistics Quarterly* 25: 129-148.
- [BUR84] Burkard, R.E. y Rendl, F. (1984) "A thermodynamically motivated simulation procedure for combinatorial optimization problems", *European Journal of Operational Research* 17: 169-174.
- [BUR90] Burkard, R.E. (1990) "Locations with spatial interactions: the quadratic assignment problem", *Discrete Location Theory*, Mirchandani & Francis (eds.), John Wiley & Sons.
- [KOO57] Koopmans, T.C. y Beckmann, M.J. (1957) "Assignment problems and the location of economic activities", *Econometrica* 25: 53-76.
- [LAA88] Laarhoven P.J.M. van y Aarts E.H.L. (1988) *Simulated Annealing: Theory and Applications*. Reidel.
- [NUG68] Nugent, C.E., Vollmann, T.E. y Ruml, J. (1968) "An experimental comparison of techniques for assignment of facilities to locations", *Operations Research* 16: 150-173.

OPTIMIZACIÓN CON RECOCIDO SIMULADO  
PARA  
EL PROBLEMA DE ASIGNACIÓN CUADRÁTICA Y  
COMPARACIONES CON OTRAS TÉCNICAS DE  
SOLUCIÓN

MIGUEL ANGEL GUTIÉRREZ ANDRADE<sup>1</sup> – SERGIO DE LOS COBOS SILVA<sup>2</sup>  
BLANCA ROSA PÉREZ SALVADOR<sup>2</sup>

---

**Resumen**

El objetivo de este artículo es presentar la bondad de la técnica de recocido simulado para el problema de asignación cuadrático, así como los lineamientos a seguir para implementarlo en la computadora. Para tal propósito, se presentan tres métodos alternativos de solución del problema: un método exacto, un método heurístico y un método de búsqueda local. El método exacto y el heurístico que se seleccionaron, se reportan en la literatura como los más eficientes, vea [BUR90]. El análisis comparativo de los cuatro algoritmos se hizo en cuanto a tiempo de ejecución, así como a su cercanía con el valor óptimo de la función objetivo, cuando se conoce tal solución. Adicionalmente se prueba para las instancias de Nugent. Para cada algoritmo se reporta la experiencia computacional y al final se hace un análisis comparativo de los cuatro métodos.

**Palabras clave:** recocido simulado, algoritmos heurísticos, problema de asignación cuadrático, optimización matemática, simulación matemática.

---

<sup>1</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - AZCAPOTZALCO, DEPARTAMENTO DE SISTEMAS, AVENIDA SAN PABLO 180, COL. REYNOSA, C.P. 02200, MÉXICO D.F.; E-MAIL: GAMA@HP9000A1.UAM.MX

<sup>2</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA - IZTAPALAPA

## 1 Descripción

Una variante importante del problema de localización clásico, es cuando el costo asociado por colocar una planta en un determinado lugar depende no únicamente de las distancias entre plantas y de las demandas respectivas, sino también de la interacción con otras plantas.

Una formulación matemática de esta problemática fue inicialmente planteada por Koopmans y Beckmann ([KOO57]) y consiste en minimizar

$$f(\pi) = \sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{\pi(i)\pi(k)} \quad (1)$$

donde  $\pi$  es una permutación del conjunto  $N = \{1, 2, \dots, n\}$  y  $a_{ik}$ ,  $b_{ik}$  para  $i, k = 1, \dots, n$ , son números reales. Equivalentemente, se desea encontrar una permutación  $\pi$  del conjunto  $N$  que minimice el valor en (1.1).

Una interpretación del problema planteado consiste en suponer que  $n$  módulos se han colocado en un tablero y se han conectado por medio de “alambres”. Los módulos deben colocarse de manera que la longitud total del alambre que los conecta sea mínima.

Denote la distancia entre las posiciones  $i$  y  $k$  en el tablero por  $a_{ik}$ . Sea  $b_{jl}$  el número de conexiones entre el módulo  $j$  y el módulo  $l$ . Entonces una permutación  $\pi$  del conjunto  $N = \{1, 2, \dots, n\}$  asigna una posición en el tablero a todo módulo. Si  $j$  se sitúa en la posición  $i$  y  $l$  está situado en la posición  $k$ , el producto  $a_{ik} b_{\pi(i)\pi(k)}$  es la longitud del alambre entre los módulos  $j = \pi(i)$  y  $l = \pi(k)$ . Por lo tanto la longitud total del alambrado es

$$\sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{\pi(i)\pi(k)},$$

y el problema se reduce a

$$\underset{\pi}{\text{minimizar}} \quad z = \sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{\pi(i)\pi(k)}$$

El problema anterior, se conoce comúnmente en la literatura como *el problema de asignación cuadrático* (QAP). Una formulación similar del QAP es aplicable para el diseño de paneles de control y teclados de máquinas. Existen muchas otras aplicaciones del QAP por ejemplo: problemas de ubicación de edificios, planeación de hospitales; en deportes, en problemas de calendarización; en sistemas de información, un orden óptimo para interrelacionar datos en una cinta magnética; en química, para analizar reacciones

químicas para componentes orgánicas; en balancear turbinas para autos de carreras; en arqueología, modelos para ordenar datos arqueológicos.

## 2 Algoritmo Exacto

Se conoce que el problema de asignación cuadrático es NP-duro (vea [BUR90]). Por lo tanto, hasta ahora, únicamente se conocen algoritmos de enumeración implícita que resuelven el problema de manera óptima. En esta sección se describe un algoritmo para encontrar las soluciones óptimas que según experiencia computacional de Burkard (vea [BUR90]) es de los algoritmos más eficientes que se conocen para resolver el problema de manera exacta. Este método se conoce como el método de la perturbación.

### 2.1 Descripción del algoritmo

El algoritmo de ramificación y acotamiento se puede describir como sigue: Sea  $M \subseteq N$  donde  $N = \{1, 2, \dots, n\}$ . Denote por  $S_n$  el conjunto de todas las permutaciones  $\pi$  de  $N$ . La restricción de  $\pi$  a  $M$  se llama *permutación parcial*  $(\pi, M)$ . Además sea  $\phi_M = \{\pi \in S_n \mid \pi_M = (\pi, M)\}$  denote el conjunto de todas las permutaciones que son extensiones de la permutación parcial  $(\pi, M)$ . Se define  $\pi(M) = \{\pi(i) \mid i \in M\}$ .

Sea una instancia de un problema de asignación cuadrático con matrices  $A$  y  $B$  y una permutación parcial  $(\pi, M)$  dada. Una permutación  $\pi \in \phi_M$  tiene la función objetivo siguiente:

$$z(\pi) = \sum_{i,p \in M} a_{ip} b_{\pi(i)\pi(p)} + \sum_{i \in M, p \notin M} [a_{ip} b_{\pi(i)\pi(p)} + a_{pi} b_{\pi(p)\pi(i)}] + \sum_{i,p \notin M} a_{ip} b_{\pi(i)\pi(p)}. \quad (2)$$

la primera suma en (2.1) puede calcularse de manera exacta conociendo la permutación parcial  $(\pi, M)$ . Para la segunda y tercera sumatoria en (2.1) se puede encontrar una cota inferior resolviendo un problema de asignación lineal con una matriz de costos.  $C = (c_{ij})$  ( $i \notin M, j \notin \pi(M)$ ). Sin embargo la solución óptima de este problema de asignación lineal da sugerencias de como debe hacerse una asignación simple ( $i \rightarrow j$ ) en el paso siguiente. Los coeficientes  $c_{ij}$  se escogen como cotas inferiores para la contribución de la asignación simple  $i$  a  $j$  en el valor de la función objetivo  $z(\pi)$ . Estas cotas inferiores  $c_{ij}$  pueden determinarse de la siguiente manera. En principio

$$a_{ii}b_{jj} + \sum_{p \in M} (a_{ip}b_{j\pi(p)} + a_{pi}b_{\pi(p)j})$$

es la contribución a  $c_{ij}$ , que depende de la permutación parcial  $(\pi, M)$ . Este valor puede mejorarse por un paso de reducción.

Reducción de la matriz A:

$$\begin{aligned} \bar{a}_i &= \min_{p \notin M, p \neq i} a_{ip} & (i \notin M); \\ \bar{a}_{ip} &= a_{ip} - \bar{a}_i & (i, p \notin M; i \neq p); \\ \tilde{a}_p &= \min_{i \notin M, i \neq p} \bar{a}_{ip} & (p \notin M); \\ a_{ip} &= \bar{a}_{ip} - \tilde{a}_p & (i, p \notin M; i \neq p). \end{aligned}$$

Es fácil mostrar que

$$\bar{a}_i \sum_{q \notin \pi(M), q \neq j} b_{jq} + \tilde{a}_i \sum_{q \notin \pi(M), q \neq j} b_{qj} \quad (i \notin M, j \notin \pi(M))$$

es la contribución adicional que aparece en  $c_{ij}$ , si la asignación simple  $(i \rightarrow j)$  se hace. En forma análoga la matriz  $B$  puede reducirse y esto permite un mejoramiento de la cota. La cota puede aún mejorarse. La idea es que se puede obtener una cota inferior para todos los productos escalares  $\sum_{i=1}^n a_i b_{\pi(i)}$  de dos vectores  $\mathbf{a}$  y  $\mathbf{b}$  no negativos, si el orden de los elementos del vector  $\mathbf{a}$  se toma en forma creciente y el de los elementos del vector  $\mathbf{b}$  se toma en forma decreciente y formamos el producto escalar  $\langle a, b \rangle$  de estos dos vectores ordenados. Si se aplica lo anterior para reducir los elementos  $a_{ip}$  ( $i, p \notin M; i \neq p$ ) y  $b_{jq}$  ( $j, q \notin \pi(M); j \neq q$ ) y se definen los vectores:

$$\begin{aligned} a_i &= (a_{ip} | p \notin M, p \neq i) & i \notin M, \\ b_j &= (b_{jq} | q \notin \pi(M), q \neq j) & j \notin \pi(M), \end{aligned}$$

entonces  $\langle a, b \rangle$  es una contribución adicional a la cota  $c_{ij}$  (llamada la “cota de Gilmore”). Ahora se resuelve un problema de asignación lineal con matriz de costos  $C = (c_{ij})$  ( $i \notin M, j \notin \pi(M)$ ). La experiencia computacional (vea [BUR90]) muestra que estos problemas deben resolverse en forma óptima. Sea  $z_L$  el valor de la función objetivo y  $(\bar{c}_{ij})$  los costos reducidos en la



solución óptima del problema de asignación lineal con coeficientes de costo  $c_{ij}$ . Entonces una cota inferior para el valor de la función objetivo en (2.1) de las permutaciones  $\pi \in \phi_M$  es

$$z(\phi_M) = \sum_{i,p \in M} a_{ip} b_{\pi(i)\pi(p)} + z_L.$$

La ramificación en el método de la perturbación viene dada de la siguiente forma: Se inicializa con la permutación parcial, en la que  $M$  es el conjunto vacío. Sea  $z$  una cota superior para el valor óptimo de la función objetivo. Si  $z(\phi_M) < z$  se escoge una pareja de índices  $(i, j)$   $i \notin M$ ,  $j \notin \pi(M)$  por las reglas de costos alternativos, reemplace  $M$  por  $M \cup \{i\}$  y extienda la permutación parcial por la asignación simple  $(i \rightarrow j)$ . Para esta nueva permutación parcial  $(\bar{\pi}, \bar{M})$  se calcula una nueva cota  $z(\phi_{\bar{M}})$ . Los coeficientes  $c_{ij}$  nuevos pueden obtenerse a partir de la solución óptima del último problema de asignación lineal. Si  $|M| = n - 2$ , el valor de las dos permutaciones correspondientes se calcula de manera exacta y eventualmente la cota  $z$  se mejora. Si  $z(\phi_{\bar{M}}) \geq z$ , se regresa a la última permutación parcial encontrada  $(\pi, M)$  con  $z(\phi_{\bar{M}}) < z$ . La asignación simple  $(i \rightarrow j)$ , seleccionada en la última transición de  $(\pi, M)$  a su sucesor, se elimina y la correspondiente entrada  $(i, j)$  de la matriz  $\bar{C} = (\bar{c}_{ij})$  se bloquea. Si el elemento bloqueado es el primer elemento bloqueado en la matriz  $\bar{C} = (\bar{c}_{ij})$  correspondiente a  $(\pi, M)$ , una nueva asignación simple se escoge en el mismo renglón o columna. Si el renglón  $i$  (o columna  $j$ ) contiene más elementos bloqueados, la nueva asignación simple se toma sobre este renglón índice (columna índice). Si esto no es posible se regresa a la permutación parcial predecesora de la permutación parcial  $(\pi, M)$ . La solución óptima tiene el valor  $z$  cuando el conjunto  $M$  se hace vacío por segunda ocasión.

Las reglas de costos alternativas usados son las siguientes. Sea  $\pi^*$  una solución óptima del problema de asignación lineal con costos  $c_{ij}$  y sean  $(\bar{c}_{ij})$  denoten los costos reducidos en el tableau óptimo. Calcule para todo  $(i, \pi^*(i))$  las expresiones

$$\alpha_i = \min_{j, j \neq \pi^*(i)} \bar{c}_{ij}; \quad \beta_i = \min_{p, p \neq i} \bar{c}_{p\pi^*(i)};$$

el valor  $\alpha_i + \beta_i$  constituye una cota inferior para el incremento de la cota  $z_L$ , si la asignación simple  $(i \rightarrow \pi^*(i))$  no se realiza. Se puede escoger la pareja  $(i, \pi^*(i))$  para la siguiente asignación simple que tiene costos alternativos maximal. Por lo tanto se determina el índice  $i$  que resuelve el problema:

$$\max_{i \notin M} (\alpha_i + \beta_i)$$

y a partir de aquí se define la nueva permutación parcial  $(\bar{\pi}, \bar{M})$ :

$$\begin{aligned} \bar{M} &= M \cup \{i\} \\ \bar{\pi}(i) &= \pi^*(i) \\ \bar{\pi}(p) &= \pi(p) \quad \text{para toda } p \in M. \end{aligned}$$

## 2.2 Experiencia computacional

La prueba del algoritmo se llevó a cabo generando instancias del problema de asignación cuadrático de manera aleatoria. Para la matriz  $A$  se generaron números aleatorios entre 0 - 100 y para la matriz  $B$  se generaron números aleatorios entre 0 - 50. Para cada valor de  $n = 5, 6, 7, 8, 12$  se generaron 100 instancias. La Tabla 1 reporta el tiempo promedio. Cabe mencionar que para  $n = 13$  el programa tardó más de 6 horas en una instancia generada aleatoriamente, por lo que se decidió cortar la tabla hasta el valor de  $n = 12$ .

ALGORITMO EXACTO	
n	tiempo (seg.)
5	0.08
6	0.25
7	1.03
8	3.98
10	73.45
12	1713.99

Tabla 1.

ALGORITMO EXACTO	
n	tiempo (seg.)
5	0.06
6	0.22
7	0.61
8	2.70
12	520.31

Tabla 2. Problemas de Nugent.

Adicionalmente, para cada uno de los valores anteriores de  $n$  se corrieron los problemas de prueba de Nugent (vea [NUG68]). Los resultados de estas ejecuciones aparecen en la Tabla 2.

## 3 Algoritmo heurístico

Como ya se ha mencionado, el problema de asignación cuadrático es NP-duro y desde el punto de vista práctico, es imposible resolver instancias “grandes”

del problema. Existen instancias con  $n = 20$  para las que hasta la fecha no se conoce la solución óptima (vea [NUG68] y [BUR90]), por lo que cuando se requiere resolver problemas de mayor tamaño, se debe recurrir a algoritmos heurísticos que no necesariamente obtienen la solución óptima del problema, pero dan una “buena” solución. Burkard y Stratmann ([BUR78] y [BUR90]) proponen un algoritmo para encontrar soluciones subóptimas al problema que denominan el método de incrementar grado de libertad. Este algoritmo se describe a continuación.

### 3.1 Descripción del algoritmo

Usando la notación de la sección anterior, se comienza con el conjunto vacío  $M$  y se escoge de manera arbitraria un índice  $i \notin M$  y se asigna este índice  $i$  a un índice  $j \in \{1, 2, \dots, n\}$ , tal que  $z(j)$  sea minimal. Para  $j \notin \pi(M)$  se define  $z(j)$  por

$$z(j) = a_{ii}b_{jj} + \sum_{p \in M} [a_{ip}b_{j\pi(p)} + a_{pi}b_{\pi(p)j}] + \sum_{p,q \in M} a_{pq}b_{\pi(p)\pi(q)} \quad (3)$$

y para  $j = \pi(i_0)$ ,  $i_0 \in M$ ;  $k \notin \pi(M)$  se define  $z(j)$  por

$$\begin{aligned} z(j) &= a_{ii}b_{jj} + a_{i_0i_0}b_{kk} + a_{i_0i}b_{jk} + a_{i_0i}b_{kj} \\ &+ \sum_{p \in M \setminus \{i_0\}} [a_{ip}b_{j\pi(p)} + a_{pi}b_{\pi(p)j} + a_{i_0p}b_{k\pi(p)} + a_{pi_0}b_{\pi(p)k}] \\ &+ \sum_{p,q \notin M \setminus \{i_0\}} a_{pq}b_{\pi(p)\pi(q)}, \end{aligned} \quad (4)$$

se define  $m = |M| + 1$ . Para determinar para  $i \notin M$  un índice  $j \in \{1, \dots, n\}$  con valor minimal  $z(j)$ , se tienen que evaluar  $m(n - m + 1)$  sumas de la forma (3.1) o (3.2). Como  $m(n - m + 1)$  se incrementa con  $m = 1, 2, \dots, \lfloor \frac{n}{2} \rfloor$  este método se llama el método de incrementar grados de libertad. La permutación final que se obtiene por este método depende de la sucesión de índices  $i$  que se escojan. Por lo tanto, si se realiza este procedimiento con sucesiones diferentes de índices  $i$  (es decir, diferentes permutaciones de la sucesión básica  $i = 1, 2, \dots, n$ ), se obtendrán diferentes soluciones. Se puede tomar ventaja de esta observación y ejecutar el método para selecciones diferentes de los índices libres  $i \notin M$ . En el programa, la subrutina de asignación se repite REP veces comenzando con permutaciones generadas aleatoriamente para los índices  $i$ .

### 3.2 Experiencia computacional

Se utilizaron las mismas 100 instancias generadas para la prueba del algoritmo exacto para  $n = 5, 6, 7, 8, 10, 12$  y se generaron aleatoriamente 100 instancias para cada valor de  $n = 15, 20, 25, 30$  con las mismas características, ya descritas en el algoritmo exacto, para las matrices  $A$  y  $B$ . La Tabla 3 reporta el tiempo promedio.

n	ALGORITMO HEURÍSTICO
	tiempo (seg.)
5	0.08
6	0.16
7	1.28
8	3.44
10	1.01
12	2.03
15	4.75
20	14.55
25	35.02
30	71.80

Tabla 3.

ALGORITMO HEURÍSTICO			
n	Mejor Solución	Valor	Tiempo
	Conocida		(seg.)
5	50 (óptima)	50	0.11
6	86 (óptima)	86	0.17
7	148 (óptima)	148	0.28
8	214 (óptima)	226	0.44
12	578 (óptima)	612	2.03
15	1150 (óptima)	1202	4.78
20	2570	2774	14.56
30	6124	6630	71.84

Tabla 4. Problemas de Nugent.

Para  $n = 5, 6, 7, 8, 12, 15, 20, 30$  se corrieron los problemas de prueba de Nugent (vea [NUG68]). Los resultados de estas ejecuciones aparecen en la Tabla 4.

## 4 Algoritmo de recocido simulado

El algoritmo de recocido simulado puede usarse para atacar este problema y obtener soluciones que no necesariamente son óptimas pero se puede mostrar un comportamiento muy favorable del método con respecto a instancias que se conoce la solución, así como para aquellas instancias que no se conoce la solución óptima pero se conoce una mejor solución. Para aplicar el algoritmo de recocido simulado se necesitan definir ciertos puntos que a continuación se describen.

#### 4.1 Descripción del algoritmo

Como ya se ha mencionado, el espacio de soluciones consiste de todas las permutaciones de  $n$  elementos. Para cada permutación  $\pi$ , se define la función de costo como

$$f(\pi) = \sum_{i=1}^n \sum_{p=1}^n a_{ip} b_{\pi(i)\pi(p)}. \quad (5)$$

La generación de nuevas soluciones se hace a partir de tomar permutaciones que difieran de la actual en únicamente dos cambios. Se comienza con una permutación  $\pi$  y se seleccionan aleatoriamente dos índices  $i$  y  $k$  y a partir de estos índices se define la nueva permutación  $\bar{\pi}$  por

$$\bar{\pi}(i) \equiv \pi(k), \quad \bar{\pi}(k) \equiv \pi(i) \quad (6)$$

$$\bar{\pi}(l) \equiv \pi(l), \quad l = 1, 2, \dots, n, \quad l \neq i, k.$$

La diferencia en la función de costo entre la solución actual y la nueva solución generada a partir del cambio  $i, k$  puede calcularse de la siguiente expresión:

$$\begin{aligned} \Delta f = \sum_{l=1, l \neq i, k}^n [(a_{il} - a_{kl})(b_{\pi(k)\pi(l)} - b_{\pi(i)\pi(l)}) + (a_{li} - a_{lk})(b_{\pi(l)\pi(k)} - b_{\pi(l)\pi(i)})] \\ + (a_{ii} - a_{kk})(b_{\pi(k)\pi(k)} - b_{\pi(i)\pi(i)}) \\ + (a_{ik} - a_{ki})(b_{\pi(k)\pi(i)} - b_{\pi(i)\pi(k)}), \end{aligned} \quad (7)$$

donde  $\Delta f$  denota la diferencia en el costo entre la solución propuesta menos la solución actual al hacer el intercambio. Si  $\Delta f < 0$ , entonces  $\bar{\pi}$  tiene un valor en la función objetivo menor que  $\pi$ . Usando estas observaciones se puede establecer el algoritmo de recocido simulado.

El Paso 1 del algoritmo se hace tomando una permutación aleatoria, del valor objetivo de esta permutación aleatoria se toma  $c_0$  igual al 20% del valor objetivo de esta permutación aleatoria y se inicia un ciclo de  $r = 2n$  veces el Paso 6, contabilizando el porcentaje de soluciones más malas que la actual aceptadas. Si este porcentaje es mayor o igual al 80%, entonces se toma el valor  $c_0$  si no es el caso, se incrementa un 10% y se repite el proceso hasta cumplir la condición de aceptación del 80% de soluciones malas. A partir

de aquí, se procede con el algoritmo, cada vez que se ejecuta el Paso 7, se decrementa el valor del parámetro de control con  $c \leftarrow 0.9 \times c$  y  $r \leftarrow 1.05 \times r$ . El proceso continúa hasta encontrar un valor de  $c$  en el que, dentro del Paso 6, no se acepte ninguna nueva solución.

### Algoritmo de Recocido Simulado para el Problema de Asignación Cuadrático

**Propósito:** Obtener una permutación  $\pi$  que minimice el valor de la función en (1.1).

**Descripción:**

**Paso 1:** Tome  $c$  un valor grande apropiado.

**Paso 2:** Obtenga una permutación aleatoria  $\pi$  del conjunto  $\{1, 2, \dots, n\}$ .

Obtenga el valor correspondiente de la función objetivo  $f(\pi)$ .

**Paso 3:** Asigne el valor verdadero a la variable lógica CHANGE.

**Paso 4:** Si la variable CHANGE tiene valor falso, entonces pare.

**Paso 5:** Haga CHANGE igual a falso.

**Paso 6:** Repita los puntos 1-9  $r$  veces.

1.- Tome una permutación arbitraria de dos cambios  $\bar{\pi}$

2.- Evalúe el valor  $\Delta f$  de la expresión (20)

3.- Si  $\Delta f < 0$  entonces vaya a 7.

4.- Sea  $P(\Delta f) = \exp(-\Delta f/c)$ .

5.- Genere un número aleatorio  $x$ , uniformemente distribuido en  $(0, 1)$ .

6.- Si  $x \geq P(\Delta f)$  entonces vaya a 9.

7.- Acepte la nueva permutación  $\pi \leftarrow \bar{\pi}$ ,  $f(\pi) \leftarrow f(\pi) + \Delta f$ .

8.- Si  $\Delta f \neq 0$  entonces CHANGE  $\leftarrow$  verdadero.

9.- Fin del paso 6.

**Paso 7** Reduzca  $c$  e incremente  $r$ .

**Paso 8** Vaya al paso 4.

## 4.2 Experiencia computacional

Se utilizaron las mismas 100 instancias generadas de manera aleatoria para  $n = 5, 6, 7, 8, 10, 12, 15, 20, 25, 30$  que se usaron para el algoritmo heurístico, con las características ya descritas. La Tabla 5 reporta el tiempo promedio. La Tabla 6 reporta los resultados para las instancias de prueba de Nugent.

## 5 Algoritmo de búsqueda local

Uno de los puntos importantes del algoritmo de recocido simulado es el hecho de que cada vez que se ejecuta el **Paso 6** del mismo, no solamente acepta permutaciones  $\pi$  para las que se tenga un mejor valor objetivo, sino que también se aceptan con cierta probabilidad permutaciones cuyo valor objetivo es más malo. Con el objeto de poder comparar los resultados del algoritmo de recocido simulado con el algoritmo de búsqueda local correspondiente, también se implementó este último en la computadora.

n	ALGORITMO REC. SIM.
	tiempo (seg.)
5	1.73
6	2.25
7	2.98
8	3.79
10	6.40
12	12.73
15	15.21
20	20.04
25	28.62
30	43.65

Tabla 5.

ALGORITMO RECOCIDO SIMULADO			
n	Mejor Solución	Valor	Tiempo
	Conocida		(seg.)
5	50 (óptima)	50	0.27
6	86 (óptima)	86	1.27
7	148 (óptima)	148	1.58
8	214 (óptima)	214	1.76
12	578 (óptima)	578	2.86
15	1150 (óptima)	1150	14.17
20	2570	2570	36.80
30	6124	6150	76.13

Tabla 6. Problemas de Nugent.

n	ALGORITMO BUS. LOC.
	tiempo (seg.)
5	0.13
6	0.21
7	0.29
8	0.43
10	0.79
12	1.07
15	2.40
20	5.64
25	11.11
30	20.96

ALGORITMO BÚSQUEDA LOCAL			
n	Mejor Solución	Valor	Tiempo
	Conocida		(seg.)
5	50 (óptima)	50	0.11
6	86 (óptima)	86	0.22
7	148 (óptima)	148	0.28
8	214 (óptima)	214	0.33
12	578 (óptima)	592	1.32
15	1150 (óptima)	1178	2.41
20	2570	2644	5.27
30	6124	6354	15.81

Tabla 7.

Tabla 8. Problemas de Nugent.

### 5.1 Experiencia computacional

Se utilizaron las mismas 100 instancias generadas de manera aleatoria para  $n = 5, 6, 7, 8, 10, 12, 15, 20, 25, 30$  que se usaron para los algoritmos anteriores. La Tabla 7 reporta el tiempo promedio empleado por el algoritmo. La Tabla 8 reporta los resultados para las instancias de prueba de Nugent.

## 6 Análisis comparativo de los métodos

Como ya se mencionó, se generaron en forma aleatoria instancias del problema de asignación cuadrático. Para cada una de estas instancias generadas, se corrieron los cuatro algoritmos y se calculó el tiempo promedio de solución. Los resultados aparecen en las Tablas 1, 3, 5 y 7 respectivamente y aparecen para los cuatro algoritmos en la Tabla 9.

Cabe mencionar que el tiempo de solución del algoritmo de recocido simulado es muy variable, ya que depende en gran medida del valor inicial del parámetro de control  $c$ , del número  $r$  de veces que se ejecuta el paso 6 (del algoritmo) y de la forma en que se reduzca el parámetro  $c$  y se incremente el parámetro  $r$ . Para el análisis realizado en este documento, se tomó un valor inicial de  $c$  de tal manera que la tasa de aceptación de las permutaciones propuestas fuera mayor o igual a 0.8 y los decrementos fueran del 0.1 (paso 7 del algoritmo). El número  $r$  inicial que se tomó en cada caso fue  $r = 2 \times n$  y los incrementos fueron de 0.05 (paso 7 del algoritmo). También se calculó la eficiencia de cada uno de los algoritmos por medio de la ecuación (6.1) La ecuación (6.1) se usó para los valores de  $n = 5, 6, 7, 8, 10, 12$ . Para los valores mayores de  $n$ , como no se puede calcular la solución óptima, la eficiencia se calculó tomando como base el valor  $z_{rec}$ , esto es, el valor objetivo que da el algoritmo de recocido simulado y se usó la ecuación equivalente:

$$\text{eficiencia} = 1 - \frac{z - z_{rec}}{z_{rec}} \quad (8)$$

Los cálculos que aparecen en la Tabla 9 son con base en los promedios obtenidos en las 100 instancias resueltas para cada valor de  $n$  de los valores en la función objetivo de la solución correspondiente. En la Tabla 9, puede observarse que las eficiencias del algoritmo de recocido simulado son las más



elevadas y los tiempos de ejecución del algoritmo de recocido simulado, crece de manera razonable; incluso, para instancias pequeñas se obtienen tiempos de ejecución más grandes que para el algoritmo heurístico. Esta situación cambia para valores de  $n$  grandes.

Otra característica importante del algoritmo de recocido simulado es que las variaciones en las soluciones obtenidas al resolver una misma instancia con soluciones iniciales diferentes, es mucho menor que las variaciones en las soluciones que dan los otros dos algoritmos, esto es, la solución del algoritmo de recocido simulado es menos dependiente de la solución inicial proporcionada. En cambio para el algoritmo heurístico y más aún, para el algoritmo de búsqueda local, las soluciones dependen en gran medida de la solución con que se inicialice. Este comportamiento se puede observar en la Tabla 10 en donde se muestran los resultados al ejecutar cada uno de los algoritmos 100 veces con soluciones iniciales generadas de manera aleatoria para cada uno de los problemas de Nugent.

n	ALGORITMO							
	EXACTO		HEURISTICO		REC. SIM.		BUS. LOC.	
	eficien.	tiempo (seg.)	eficien.	tiempo (seg.)	eficien.	tiempo (seg.)	eficien.	tiempo (seg.)
5	1.00	0.08	0.9891	0.08	0.9999	1.73	0.9747	0.13
6	1.00	0.25	0.9803	0.16	0.9982	2.25	0.9666	0.21
7	1.00	1.03	0.9767	0.28	0.9936	2.98	0.9676	0.29
8	1.00	3.98	0.9737	0.44	0.9899	3.79	0.9630	0.43
10	1.00	73.45	0.9684	1.01	0.9905	6.40	0.9629	0.79
12	1.00	1713.99	0.9674	2.03	0.9940	12.73	0.9512	1.07
15	—	—	0.9693	4.75	1.0000	15.21	0.9640	2.40
20	—	—	0.9512	14.55	1.0000	20.04	0.9518	5.64
25	—	—	0.9527	35.02	1.0000	28.62	0.9551	11.11
30	—	—	0.9543	71.80	1.0000	43.65	0.9580	20.96

Tabla 9. Comparación de los algoritmos para resolver el QAP con respecto al tiempo de ejecución y a su eficiencia.

n	HEURÍSTICO			REC. SIM.			BUS. LOC.		
	Mejor	Media	Peor	Mejor	Media	Peor	Mejor	Media	Peor
5	50	51.6	58	50	50.4	58	50	55.2	62
6	86	88.5	94	86	86.3	92	86	90.2	94
7	148	153.9	164	148	148.7	156	148	154.5	172
8	214	228	252	214	215.3	222	214	224.5	248
12	578	620.3	662	578	600.2	640	578	610.8	650
15	1188	1231	1294	1150	1172	1214	1150	1204	1290
20	2692	2749	2794	2570	2657	2728	2604	2680	2742
30	6488	6636	6766	6124	6333	6442	6242	6364	6496

Tabla 10. Comparación entre los algoritmos heurístico, recocido simulado

y búsqueda local con respecto a la dependencia de la solución inicial

Finalmente se hará mención que todos los programas computacionales y las ejecuciones de los mismos, se realizaron en una computadora 486 de 33 mhz.

## 7 Conclusiones

Las pruebas computacionales con el problema de asignación cuadrático, revelan un excelente comportamiento del algoritmo de recocido simulado. Las soluciones subóptimas difieren en 1-2% de la mejor solución conocida. Además si se ejecuta el procedimiento varias veces con diferentes valores iniciales, se obtienen en todos los casos la solución óptima o las mejores soluciones que se conocen (según sea el caso). Por otra parte, la implementación del algoritmo en la computadora, no requiere de gran capacidad de memoria y permite un tiempo razonable para obtener buenas soluciones en las computadoras personales.

## Referencias

- [AAR89] Aarts, E. y Korst, J. (1989) *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, Chichester.
- [BUR78] Burkard, R.E. y Stratmann, K. (1978) "Numerical investigations on quadratic assignment problems", *Naval Research Logistics Quarterly* **25**: 129–148.

- [BUR84] Burkard, R.E. y Rendl, F. (1984) “A thermodynamically motivated simulation procedure for combinatorial optimization problems”, *European Journal of Operational Research* **17**: 169–174.
- [BUR90] Burkard, R.E. (1990) “Locations with spatial interactions: the quadratic assignment problem”, *Discrete Location Theory*, Mirchandani & Francis (eds.), John Wiley & Sons.
- [KOO57] Koopmans, T.C. y Beckmann, M.J. (1957) “Assignment problems and the location of economic activities”, *Econometrica* **25**: 53–76.
- [LAA88] Laarhoven P.J.M. van y Aarts E.H.L. (1988) *Simulated Annealing: Theory and Applications*. Reidel.
- [NUG68] Nugent, C.E., Vollmann, T.E. y Ruml, J. (1968) “An experimental comparison of techniques for assignment of facilities to locations”, *Operations Research* **16**: 150–173.

## ESTUDIO DE ALGUNOS ASPECTOS RELATIVOS A LAS PARTÍCULAS SUSPENDIDAS EN LA CIUDAD DE MÉXICO

B. HERNANDEZ<sup>1</sup> - F. ALDAPE<sup>1</sup> - R. HERNANDEZ<sup>1</sup>

---

A partir de muestras de partícula suspendida total en el noroeste de la Ciudad de México utilizando la técnica PIXE (Proton Induced X-Ray Emission), fueron identificados y cuantificados diez elementos (S, K, Ca, Ti, V, Mn, Fe, Cu, Zn, Pb). Las concentraciones permitieron apreciar los niveles de contaminación y la determinación de los factores de enriquecimiento permitió separar la componente antropogénica de la natural. El estudio se complementa con un análisis estadístico multivariado de las mediciones realizadas y algunas variables de carácter meteorológico. Se presentan algunas conclusiones y recomendaciones.

---

<sup>1</sup>ININ, México

## PAQUETE DIDÁCTICO DE OPTIMIZACIÓN EN REDES

RAFAEL LÓPEZ BRACHO<sup>1</sup> - M. PAULA ORTUÑO SÁNCHEZ<sup>1</sup>  
FELIPE CARRILLO ROMERO<sup>1</sup> - M. TERESA RODRÍGUEZ MARTÍNEZ<sup>1</sup>

---

### Resumen

El paquete computacional "*Optimización en Redes para Windows*", versión 2.0 usa programación orientada a objetos para resolver problemas de Flujo máximo en una red, utilizando el algoritmo de Ford y Fulkerson, con la rutina de Edmonds y Karp para la obtención del camino aumentante en cada iteración del algoritmo, de Ruta más corta entre 2 vértices de una red utilizando, a elección del usuario, el algoritmo de Bellman o el de Dijkstra y de Árbol de Expansión de peso mínimo, utilizando el algoritmo de Kruskal.

### Presentación

El paquete permite optimizar redes de hasta 125 vértices y 7750 arcos; esta versión está orientada a fines didácticos, por lo que todas las redes a optimizar deben introducirse directamente en la pantalla en cada interacción con el sistema, esto lógicamente crea una dificultad de saturación de la pantalla en redes muy grandes (se estima que con más de 10 vértices), la cual es una limitación física de visualización, pero no de ejecución del programa.

El paquete aprovecha los códigos definidos para representar al objeto gráfica y el funcionamiento de las estructuras de datos así desarrolladas, a través de la programación orientada a objetos.

Esta versión del paquete servirá como apoyo didáctico del curso Investigación de Operaciones I, ofrecido a estudiantes de Ingeniería de la Universidad Autónoma Metropolitana Azcapotzalco.

Para cada problema, se va observando en la pantalla la construcción de la red que lo modela, lo cual permite explicar las características del modelo.

Durante la aplicación, para el caso del problema de Flujo Máximo se puede mostrar en la pantalla, si se desea, la construcción paso a paso de las trayectorias aumentantes.

En todos los problemas se muestra al final en la pantalla la solución obtenida.

A continuación se muestran los problemas abordados y los algoritmos utilizados.

## 2 Problema de determinar el flujo máximo en una red

Considérese una red de tubería (drenaje, cables eléctricos, etc). Teniendo cada una, una capacidad (límite superior de la cantidad de flujo que por ella puede pasar por unidad de tiempo). Se busca determinar la cantidad máxima de flujo que se puede enviar desde un nodo origen o fuente  $F$  a un nodo destino o sumidero  $S$ , sabiendo que el flujo es conservativo (la cantidad de flujo que entra en un nodo intermedio es igual a la cantidad de flujo que sale del mismo).<sup>1</sup>

### 2.1 Formulación

Sea  $(X,U)$  una gráfica dirigida representando la red de tubería del problema y una aplicación:

$$c: U \rightarrow \mathbb{R}$$

que a cada arco  $u \in U$  hace corresponder su capacidad  $c(u)$ . Por otra parte, sea  $u_r = (SF) \in U$  un arco especial "arco de regreso" con capacidad infinita.

El problema de flujo máximo de  $F$  a  $S$  en la red  $R=(X,U,c)$  consiste en buscar un vector  $f \in \mathbb{R}^m$  ( $m=|U|$ ) tal que:

- i)  $f$  sea un flujo en la gráfica  $G=(X,U)$
- ii)  $0 \leq f(u) \leq c(u); u \in U$
- iii)  $f(u_r)$  sea máximo bajo las condiciones anteriores.

Este problema puede formularse como el programa lineal:

$$\max Z = f(u_r)$$

s.a

$$Ef = 0$$

$$0 \leq f \leq c$$

donde  $E$  representa la matriz de incidencia de los arcos de la gráfica  $(X,U)$ , esto es:

Para la gráfica  $G=(X,U)$  con  $m$  vértices y  $n$  arcos, la matriz  $E$  es una matriz  $m \times n$  donde cada renglón está asociado con un vértice, y cada columna está asociada con un arco, el elemento de la  $i$ -ésima línea y  $j$ -ésima columna es:

$$e_{ij} = \begin{cases} +1 & \text{si el nodo } i \text{ es la extremidad inicial del arco } u_j \\ -1 & \text{si el nodo } i \text{ es la extremidad final del arco } u_j \\ 0 & \text{en otro caso} \end{cases}$$

La  $i$ -ésima restricción del sistema  $Ef = 0$  representa entonces la conservación del flujo en el nodo  $i$ .

El teorema de Flujo Máximo y Corte Mínimo, probado por Ford y Fulkerson en 1956, es fundamental para el estudio del problema de Flujo Máximo y del algoritmo de solución que ellos mismos desarrollaron en 1957. Este teorema establece lo siguiente:

<sup>1</sup> UNIVERSIDAD AUTONOMA METROPOLITANA, DPTO. DE SISTEMAS, AV. SAN PABLO 180  
COL. REYNOSA TAMAULIPAS, DEL. AZCAPOTZALCO, MEXICO D. F. CP 02200  
E-MAIL : rlb@hp9000a1.uam.mx

**Teorema de Flujo Máximo y Corte Mínimo.** Si un vector de flujo factible existe en una red de flujo  $R=(X,U,c)$ , el valor del flujo factible máximo es igual a la capacidad mínima de los cortes separando S de F.

El algoritmo de Ford y Fulkerson utiliza el Teorema anterior como criterio de detención. A partir de un flujo factible inicial, construye sucesivamente trayectorias uniendo S a F a través de las cuales se puede incrementar el flujo, por lo que son llamadas trayectorias aumentantes; continúa este proceso hasta que ya no es posible construir alguna de esas trayectorias. Al terminar la aplicación del algoritmo se tendrá el Flujo Máximo de F a S y se exhibirá igualmente el corte separando S de F, de capacidad mínima.

Jack Edmonds y Richard M. Karp, en su artículo "Theoretical Improvements in Algorithm Efficiency for Network Flow Problems", publicado en J. ACM 19, (1972), demostraron que si se usa una "búsqueda a lo ancho" (breadth first search), en el algoritmo del marcaje y siempre se escoge el camino aumentante más corto, entonces el algoritmo terminará en a lo más  $O(|X|^2|U|)$  aun si se permiten capacidades irracionales.

La descripción del Algoritmo y del proceso de obtención de la trayectoria aumentante propuesto por Edmonds y Karp se muestran a continuación.

## 2.2 Algoritmo de Resolución

### 2.2.1 Algoritmo de Ford y Fulkerson

1. Construir un flujo factible  $f$  inicial, puede ser 0 para todos los arcos.
2. Encontrar una trayectoria aumentante  $P$  de F a S.
3. Si no existe, entonces terminar, el flujo actual es óptimo; de otra manera  
-Calcular la holgura  $c_{ij}-f_{ij}$  de cada arco hacia adelante en  $P$  e identificar por  $\varepsilon_1$  la mínima de ellas, esto es:

$$\varepsilon_1 = \min \{ c_{ij}-f_{ij} \mid (i,j) \text{ es un arco hacia adelante en } P \}$$

-Identificar por  $\varepsilon_2$  el mínimo de los flujos de los arcos hacia atrás en  $P$ , esto es:

$$\varepsilon_2 = \min \{ f_{ij} \mid (i,j) \text{ es un arco hacia atrás en } P \}$$

Y calcular la holgura mínima  $\varepsilon$  por :

$$\varepsilon = \min \{ \varepsilon_1, \varepsilon_2 \}$$

Refinar ahora el flujo agregando  $\varepsilon$  al flujo de todos los arcos hacia adelante en  $P$  y restando  $\varepsilon$  al de los arcos hacia atrás en  $P$ . Esto es:

El nuevo flujo  $\hat{f}_{ij}$  se define por:

$$\hat{f}_{ij} = \begin{cases} f_{ij} + \varepsilon & \text{si } (i,j) \text{ es un arco hacia adelante en } P \\ f_{ij} - \varepsilon & \text{si } (i,j) \text{ es un arco hacia atrás en } P \\ f_{ij} & \text{si } (i,j) \text{ no pertenece a } P \end{cases}$$

4. Considérese el flujo  $\hat{f}$  obtenido en 3 y regresar al paso 2.

### 2.2.2 Método de Edmonds y Karp

*Etapa 1. Construcción de un árbol con raíz F.*

Sea  $f=(f_{ij})$  un flujo factible. Etiquete F con (0,\*), F está ahora etiquetado pero no examinado. La lista es {F}.

*Etapa 2. Selección de un nodo de la lista para examinarlo.*

Si la lista es vacía, pasar a la Etapa 5, si no, seleccionar el primer nodo de la lista y examinarlo.

*Etapa 3. Examen del nodo.*

Sea  $i$  el nodo a ser examinado.

a) Etiquetado hacia adelante. Identifique todos los nodos no etiquetados  $j$  que satisfagan  $(i,j) \in U$  y  $f_{ij} < c_{ij}$  etiquete cada uno de ellos con  $(i,+)$  y colóquelos al final de la lista.

b) Etiquetado en reversa. Identifique todos los nodos no etiquetados  $j$  que satisfagan  $(j,i) \in U$  y  $f_{ji} > 0$ , etiquete cada uno de ellos con  $(i,-)$  y colóquelos al final de la lista.

Borre el nodo  $i$  de la lista. Si  $s$  está ya etiquetado, se tiene ya un camino aumentante, pasar a la Etapa 4. Si  $s$  no está aún etiquetado pasar a la Etapa 2.

*Etapa 4. Flujo Aumentante.*

Empezando con  $S$ , usar las etiquetas de los vértices para construir el camino aumentante  $P$  de  $F$  a  $S$ . Terminar.

*Etapa 5. Terminación.*

El vector de flujo actual  $f$  es un vector de flujo máximo. Sean  $V$  el conjunto de nodos en el árbol y  $V'$  su complemento; éstos definen el corte  $(V, V')$  de capacidad mínima separando  $S$  de  $F$  en la red. Terminar.

*Observación:* Si a cada nodo etiquetado  $i$  se le agrega la etiqueta  $d_i$  la cual indica la profundidad del nodo  $i$  al ser etiquetado, esto es, la longitud del camino simple de  $F$  a  $i$  obtenido, dado que cada nodo es etiquetado después de que su padre lo ha sido, la política "primer nodo etiquetado, primer nodo examinado", garantiza que el nodo seleccionado para ser examinado siempre será un nodo con profundidad mínima desde  $F$  y le transmitirá ésta incrementada en 1 unidad a sus nodos hijos.

## 3 Problema de determinar la ruta más corta entre dos vértices de una gráfica

Este problema puede considerarse el más sencillo de los problemas de itinerarios. Consiste en determinar la trayectoria que se debe de seguir para ir de un vértice  $\gamma$  a un vértice  $\eta$  de una gráfica, en la que se asigna una distancia a cada una de las aristas, de manera que la distancia total recorrida sea mínima. Lógicamente los dos vértices a unir deben pertenecer a la misma componente conexa. En el problema puede considerarse que la gráfica es o no dirigida,



pero en el primer caso, la dirección de las aristas originará una condición más fuerte de **conectividad**, en el sentido de que debe existir al menos un camino que una los dos vértices, y que al recorrer cada arista lo haga en el sentido dado por su dirección. Este tipo de problema se puede presentar en una ciudad a algún vehículo que debe desplazarse de un sitio a otro a través de la traza de calles y avenidas de la misma, respetando su sentido de circulación. Lo mismo ocurre en la búsqueda de un itinerario a través de la red carretera de una región, en cuyo caso la gráfica se puede considerar no dirigida. Situaciones semejantes se presentan a través de una red fluvial, una red de oleoductos o gasoductos a través de los cuales se envía alguno de estos fluidos, redes de agua potable, redes eléctricas, etc.

### 3.1 Formulación

Sean  $G=(X,U)$  una gráfica dirigida y una aplicación  $d:U \rightarrow \mathbb{R}$  que a cada arco  $u \in U$  hace corresponder su distancia o longitud  $d(u)$ .

El problema de encontrar la trayectoria más corta entre 2 vértices  $\gamma$  y  $\eta$  de  $X$  es equivalente al problema de enviar una unidad de flujo del vértice  $\gamma$  al vértice  $\eta$  a un mínimo costo, donde  $d(u)$  es considerado el costo unitario del arco  $u$ .

Este problema se modela por el programa lineal:

$$\min Z = \sum_{u \in U} d(u)f(u)$$

s. a

$$\sum_{\{u \in U | I(u)=x\}} f(u) - \sum_{\{u \in U | T(u)=x\}} f(u) = \begin{cases} -1 & \text{si } x = \gamma \\ 0 & \forall x \in X - \{\gamma, \eta\} \\ 1 & \text{si } x = \eta \end{cases}$$

$$f(u) \geq 0 \quad \forall u \in U$$

donde  $f(u)$  representa el flujo del arco  $u$

$I(u)$  representa al vértice extremidad inicial del arco  $u$

$T(u)$  representa al vértice extremidad final del arco  $u$

y las restricciones representan las ecuaciones de conservación del flujo en los vértices.

Dado que la matriz de restricciones de un problema de flujo es totalmente unimodular, y las variables básicas de cada solución básica del programa lineal tienen asociadas las aristas de un árbol de expansión de la gráfica, efectivamente, para encontrar la trayectoria más corta de  $\gamma$  a  $\eta$  en la red, el programa lineal construye las trayectorias más cortas de  $\gamma$  a los restantes vértices de la gráfica. Las variables básicas que pertenecen a la trayectoria uniendo a  $\gamma$  con  $\eta$  tendrán valor 1 para su flujo, y las restantes variables básicas tendrán valor 0. Los algoritmos que se han programado tienen enfoque glotón, obtienen las arborescencias de rutas más cortas y son más eficientes que el Algoritmo Simplex en el sentido de que son deterministas polinomiales.

### 3.2 Algoritmo de Dijkstra

Este algoritmo, propuesto por E. W. Dijkstra en 1959 [DIJ 59], construye el árbol de trayectorias más cortas a partir del vértice  $\gamma$ , agregando en cada iteración el vértice aún no incluido en el árbol que sea más cercano a éste, junto con la arista que da la distancia considerada. Este algoritmo está basado en el resultado siguiente:

**Teorema.** Sean  $G=(X,U)$  una gráfica dirigida y  $d:U \rightarrow \mathbb{R}$  una aplicación que asigna a cada arco  $u$  una distancia  $d(u) \geq 0$ . Sea  $T$  un árbol de  $G$  definido por un conjunto de vértices  $V \subset X$ , que constituye una arborescencia de trayectorias más cortas a partir de una raíz  $\gamma$ . Para cada vértice  $i \in V$ , sea  $\pi(i)$  la longitud de la trayectoria más corta de  $\gamma$  a  $i$  en  $G$ . Sean  $p \in V$ ,  $q \in V' = X - V$  satisfaciendo:

$(p,q) \in U$ , y

$$\pi(p) + d(p,q) = \min \{ \pi(i) + d(i,j) : i \in V, j \in V', (i,j) \in U \}.$$

Entonces, si se agrega el arco  $(p,q)$  a  $T$  y se define  $\pi(q) = \pi(p) + d(p,q)$ , entonces se extiende  $T$  a  $T'$ , la cual es la arborescencia de trayectorias más cortas desde  $\gamma$ , definida en los vértices de  $V \cup \{q\}$ .

Comenzando con el vértice raíz, el resultado del teorema anterior puede ser usado para construir en  $n-1$  iteraciones la arborescencia buscada. Dijkstra observó que aplicar directamente el resultado anterior, llevaría a considerar cada arco en diferentes iteraciones. En su algoritmo él evita esta situación, cada arco es analizado una sola vez y el esfuerzo computacional se reduce a un orden de  $O(|U|)$ .

#### Algoritmo.

*Etapa 0 (Etapa Inicial).*

Etiquetar la raíz  $\gamma$ , con la marca permanente  $(0, \emptyset)$ . Etiquetar cada vértice  $j \neq \gamma$  en forma temporal con la marca  $(d(\gamma,j), \gamma)$  si existe el arco  $(\gamma,j)$  y  $(\infty, \emptyset)$  si no existe el arco  $(\gamma,j)$ .

Tomar  $S := \{ \gamma \}$ ,  $T := X - \{ \gamma \}$ .

*Etapa 1. (Designación de Etiquetas Permanentes).*

Seleccionar  $k \in T$  tal que  $\pi(k) = \min \{ \pi(j) \mid j \in T \}$

Hacer permanente la etiqueta temporal del vértice  $k$ .

Tomar  $S := S \cup \{ k \}$ ,  $T := T - \{ k \}$

Si  $T = \emptyset$ , alto; se han completado los cálculos. Si no, pasar a la etapa 2.

*Etapa 2 (Revisión de Etiquetas Provisionales).*

Para toda  $j \in T$ , hacer  $\pi(j) = \min \{ \pi(j), \pi(k) + d(k,j) \}$ , donde  $d(k,j) = \infty$  si no existe el arco  $(k,j)$ .

Si  $\pi(j) = \pi(k) + d(k,j)$ , entonces cambiar la etiqueta del vértice  $j$  por la etiqueta temporal  $(\pi(j), k)$ . Sino, dejarla igual.

Pasar a la etapa 1.

Donde  $\pi(j)$  es la distancia más corta desde el vértice raíz  $\gamma$  al vértice  $j$ .  
 $S$  es el conjunto de vértices para los cuales ya se ha calculado la distancia más corta que los separa de  $\gamma$ .  
 $T$  es el conjunto de vértices para los cuales aún no se ha calculado la distancia más corta que los separa de  $\gamma$ .

Al finalizar, para cada vértice  $j$  la etiqueta permanente  $(\pi(j),k)$  indica en sus componentes: la distancia más corta que lo separa del vértice  $\gamma$  y el vértice predecesor en la trayectoria más corta, respectivamente. Las segundas componentes de las etiquetas permanentes permiten reconstruir la trayectoria que une a  $\gamma$  con  $j$ .

La única limitante del algoritmo de Dijkstra, además de que el vértice  $\gamma$  debe ser raíz, es que todas las distancias deben ser no negativas.

*Observación.* El algoritmo de Dijkstra proporciona una numeración de los vértices de una gráfica en orden ascendente de la distancia que los separa de la raíz.

### 3.3 Algoritmo de Bellman

Partiendo del problema de obtención de un flujo unitario de costo mínimo en una red, en su problema dual, cada variable  $\pi(x)$  está asociada a un vértice  $x$  de la gráfica y es interpretada como la longitud de la trayectoria más corta de la fuente  $\gamma$  al vértice  $x$ . Estas variables deben satisfacer las condiciones siguientes:

$$\pi(\gamma)=0$$

$$\pi(j)=\min\{\pi(i)+d(i,j):(i,j)\in U\} \quad \forall j\in X, j\neq\gamma$$

Estas condiciones son conocidas como ecuaciones de Bellman-Ford y proporcionan condiciones necesarias para las longitudes de las trayectorias más cortas de los vértices de una gráfica al vértice raíz  $\gamma$ . Inversamente, si para cada vértice  $x\in X$  se satisfacen las ecuaciones de Bellman-Ford y existe una trayectoria en la gráfica a cada vértice  $x\in X$  con longitud  $\pi(x)$ , entonces estas trayectorias son de longitud mínima.

Bellman [BEL 58] aprovechó estas propiedades para idear el algoritmo que lleva su nombre, el cual puede aplicarse en cualquier gráfica dirigida con la condición de que no contenga circuitos, y a diferencia del algoritmo de Dijkstra, puede aceptar longitudes negativas en los arcos. El algoritmo es un procedimiento recursivo que parte de aplicar el principio de optimalidad de la Programación Dinámica al problema. Los vértices son etiquetados uno a uno, uno por cada iteración, de manera que se aprovecha la ausencia de circuitos, para dar un orden a los vértices que asegura el cumplimiento de las ecuaciones de Bellman-Ford. Todos los arcos son revisados en una sola ocasión, por lo que el número de operaciones está en función lineal del número de ellos ( $|U|=m$ ), obteniendo una complejidad computacional para este algoritmo  $O(m)$ .

#### Algoritmo.

*Etapa 0 (Etapa Inicial).*

Tomar  $\pi(\gamma)=0$ ,  $S=\{\gamma\}$

*Etapa 1. (Búsqueda de Vértice a Etiquetar).*

Buscar un vértice fuera de  $S$  teniendo todos sus predecesores en  $S$ :

- Si no existe ese vértice, terminar. Se tiene ya la arborescencia o  $\gamma$  no es raíz de la gráfica.

- Si existe ese vértice  $y$ , pasar a la etapa 2.

*Etapa 2. (Etiquetación del vértice seleccionado).*

$$\text{Tomar } \pi(y) = \min_{\{u \in U \mid T(u) = y\}} \{\pi(I(u)) + d(u)\}$$

$$S := S \cup \{y\}$$

Marcar el arco  $u$ .

Regresar a la etapa 1.

Donde  $\pi(y)$  indica la longitud de la trayectoria más corta uniendo  $y$  a  $\gamma$

$I(u)$  representa la extremidad inicial del arco  $u$

$T(u)$  representa la extremidad final del arco  $u$

$S \subseteq X$  es el subconjunto de vértices para los cuales ya se calculó la distancia más corta desde la raíz  $\gamma$ .

$$T = X - S.$$

## 4 Problema del árbol de expansión de peso mínimo

El problema de encontrar el árbol de expansión de peso mínimo, es uno de los problemas más sencillos y bellos de optimización en redes, y es de utilidad en una gran variedad de problemas.

Los datos de entrada en este problema son una gráfica conexa  $G=(X,U)$  y una aplicación  $c:U \rightarrow \mathbb{R}$  que asigna a cada arista un peso. El peso total del árbol se obtiene por la suma del peso de sus aristas. El peso de las aristas puede representar una longitud, un costo, una capacidad, etc. según el problema de que se trate. El objetivo es el diseño de una red que conecte todos los vértices usando la mínima longitud de cable o de otro recurso. Este problema a diferencia del de ruta más corta, no privilegia a algún vértice (usuario) en particular, sino que busca la posible conexión de todos ellos a menor costo. No es importante si la gráfica es dirigida o no, porque lo importante es la conexidad de todos los vértices.

Existen varios algoritmos eficientes para resolver este problema, los más conocidos son el de Kruskal y el de Prim. En este paquete hemos programado el algoritmo de Kruskal, el cual, después de algunas propiedades será expuesto.

### 4.1 Algunos resultados teóricos

A continuación se enuncian algunos resultados que permiten garantizar la eficiencia en la aplicación del algoritmo de Kruskal.

*Definición.* Dados un árbol de expansión  $T$  de una gráfica  $G=(X,U)$  y una arista  $u$  no perteneciente a  $T$ . Cuando  $u$  es agregado a  $T$ , el ciclo formado es llamado Ciclo Fundamental de  $u$  con respecto a  $T$ .

*Definición.* Dados un árbol de expansión  $T$  de una gráfica  $G=(X,U)$  y una arista  $u$  perteneciente a  $T$ . Cuando  $u$  es retirado de  $T$ , el corte formado es llamado Corte Fundamental de  $u$  con respecto a  $T$ . (ver la definición de corte en 2.2).

*Teorema.* Sea  $T$  un árbol de expansión en  $G$ . Entonces  $T$  es un árbol de expansión de peso mínimo si, y sólo si, cada arista en  $T$  es una arista de peso mínimo en su corte fundamental con respecto a  $T$ .

*Teorema.* Sea  $T$  un árbol de expansión en  $G$ . Entonces  $T$  es un árbol de expansión de peso mínimo si, y sólo si, cada arista que no pertenece a  $T$  es una arista de peso máximo en su ciclo fundamental asociado con  $T$ .

El algoritmo ideado por Kruskal en 1956 [KRU 56] parte de una lista de las aristas ordenada en base a su peso, de menor a mayor y del bosque trivial formado por los vértices de  $G$  y  $m$  aristas. En cada iteración, se toma el primer elemento de la lista y se agrega al bosque si no forma un ciclo, y se elimina en caso contrario. Este proceso se continúa hasta constituir el árbol  $T$ , el cual, lógicamente satisface los dos teoremas anteriores. Cada arista se habrá revisado a lo más una vez y el algoritmo es entonces de orden  $O(m \log n)$  tomando en cuenta que los algoritmos para ordenar de menor a mayor son de orden  $O(m \log m)$  y  $m < n^2$ . ( $m=|U|$ ,  $n=|X|$ ).

## 2 Algoritmo de Kruskal

### *Etapa 0 (Inicialización).*

Ordenar las aristas de  $G$  de acuerdo a su peso, de menor a mayor e integrarlas en una lista  $l$ . Comenzar con el bosque trivial  $(\{1\}, \emptyset), (\{2\}, \emptyset), \dots, (\{n\}, \emptyset)$ .

### *Etapa 1 (Etapa General).*

Seleccionar la arista de menor peso en  $l$ . Sea  $u=(i,j)$ .

- Si  $i, j$  pertenecen ambos a la misma componente conexa del bosque, entonces borre la arista  $u$  y repita la etapa.
- Si  $i, j$  pertenecen a diferentes componentes conexas del bosque, agregue la arista  $u$  al bosque para unir las dos componentes que conecta en una sola. Si sólo queda una componente conexa, se tiene el árbol de expansión de peso mínimo buscado, termine. Si no, realice una nueva etapa.

## Descripción técnica del paquete

El paquete fue desarrollado utilizando BorlandC++<sup>®</sup>, versión 3.1, por ser un sistema de desarrollo profesional para crear aplicaciones DOS y Windows, el cual proporciona un soporte completo de C++, entre ellos:

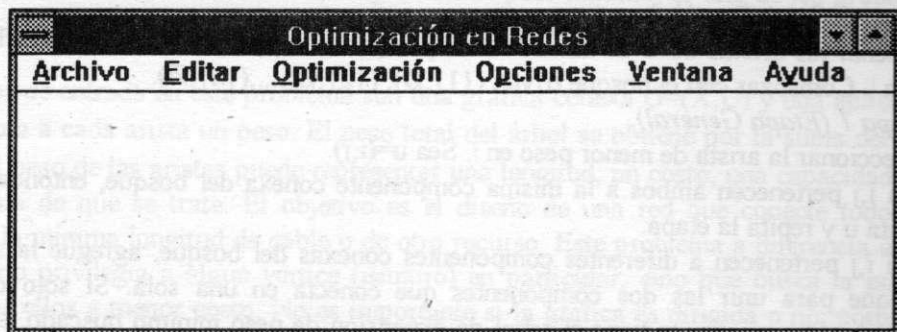
- AT&T C++ 3.0
- Esquemas para clases genéricas (Templates)
- Librerías de clases contenedoras para listas, arreglos, conjuntos, entre otras.
- Múltiple herencia.
- Librería de clases estándar.
- Clases bases virtuales y clases abstractas
- Funciones virtuales, sobrecarga de funciones y sobrecarga de operadores.

Para ejecutar el programa **Redw** de manera óptima se requiere que la computadora tenga las siguientes características:

- Procesador 80286, o superior
- Monitor VGA a color
- Sistema operativo MS-DOS 6.0 o superior y Microsoft Windows 3.0 o superior
- 4 MB de memoria RAM mínimo
- Mouse
- Unidad de disco flexible de 3.5" de alta densidad
- 1 MB de espacio en disco duro para instalación
- Impresora de matriz de puntos compatible con gráficas en IBM® o Epson® o impresora Láser compatible con Windows (si se desean imprimir las gráficas)
- Coprocesador matemático (incrementa la rapidez de ejecución)

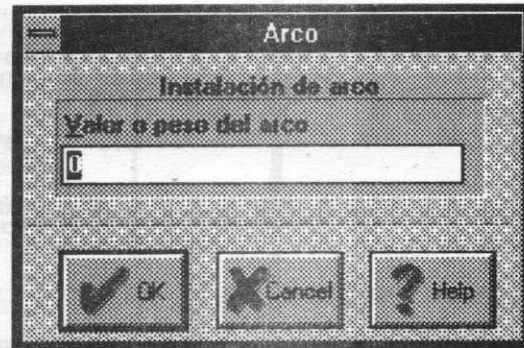
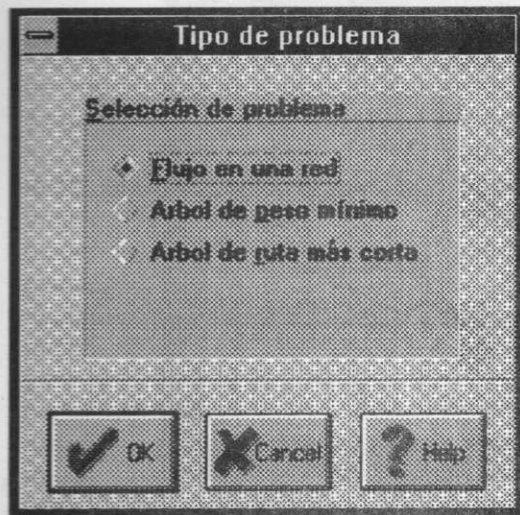
### 5.1 Ejecución del programa

La ventana de el paquete computacional "*Optimización en Redes para Windows*" versión 2.0 es la que se muestra a continuación:

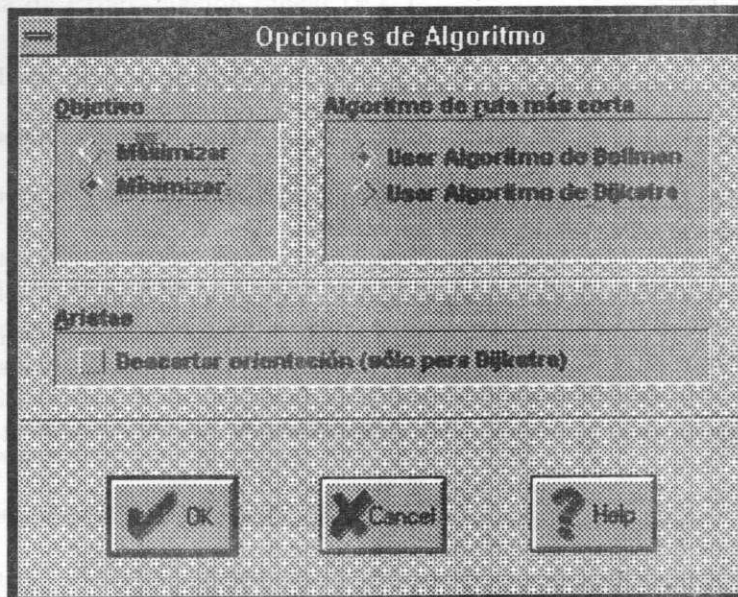


Puede observarse que en la parte superior de esta ventana, aparecen algunas de las ya conocidas opciones para los usuarios del ambiente Windows tales como: **A**rchivo con su respectivo menú que incluye *Nuevo, Abrir, Guardar, Guardar como, Imprimir, Especificar impresora* y *Salir*; **V**entana con las opciones *Cascada, Mosaico, Ordenar iconos* y *Cerrar todo* y **A**yuda que cuenta con un *Índice* y un manual en línea de *Como hacer*.

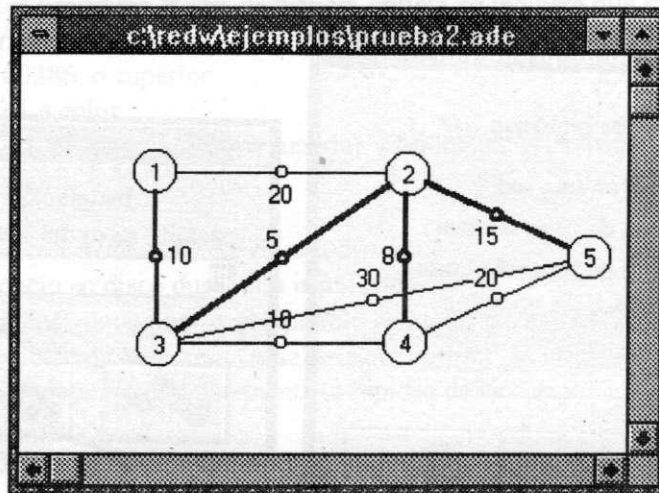
Las opciones **E**ditar, **O**ptimización y **O**pciones son propias del paquete computacional. La alternativa **E**ditar permite crear una nueva gráfica de trabajo o modificar alguna que haya sido creada anteriormente y salvada por este mismo programa. El menú desplegado en esta opción permite *Insertar Nodo|Arista, Borrar Nodo|Arista, Alterar aristas* y *Mover*. Algunos de los cuadros de diálogo que interactúan con el usuario en estas operaciones son en el que se selecciona el tipo de problema a resolver y en el que se solicita la capacidad de una arista al insertarla o alterarla. Estas ventanas se presentan a continuación:



Si ya se eligió alguna gráfica o ya se tiene editada alguna, se puede seleccionar **Opciones**, teniendo dos alternativas: *Algoritmo* y *Etiquetas Visibles*. *Etiquetas visibles* hace que los valores de aristas sean visualizados. *Algoritmo* se aplica al problema de ruta más corta (más larga) y árbol de peso mínimo (máximo), y se utiliza para selección de alternativas, el cuadro de diálogo correspondiente es el siguiente :



Una vez elegidas las opciones adecuadas, al seleccionar **Optimización** se puede *escoger Maximizar Flujo* o *Crear Árbol* de acuerdo al tipo de problema y el programa lo resuelve y presenta la solución en una ventana de trabajo como la que se presenta en la figura siguiente :



En esta ventana de trabajo se representan en la gráfica con líneas gruesas a los arcos saturados en el problema de Flujo, el árbol de peso mínimo (ó máximo), o el árbol de rutas más cortas o más largas. Por ejemplo, en la figura anterior se eligió *Crear árbol*, para un problema de árbol de peso mínimo y la solución está dada por el conjunto de aristas  $\{(1,3), (2,3), (2,5), (2,4)\}$ , o sea las aristas que están en color más oscuro.

## 6 Bibliografía

- [BEL 58] Bellman, R. E., *On a Routing Problem*, Quart. Appl. Math., Vol. 16, 1958, 87-90.
- [BOO 91] Booch, G., *Object Oriented Design*, The Benjamin/Cummings Publishing Company Inc., Redwood City, CA, 1991.
- [DIJ 59] Dijkstra, E. W., *A Note on Two Problems in Connection with Graphs*, Numerische Mathematik, Vol. 1, 1959, Pags. 269-271.
- [EDM 72] Edmonds, J. y R. M. Karp, *Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems*, Journal of ACM, Vol. 19, 1972, Pags. 248-264.
- [FAI 92] Faison, T., *Borland C++ 3.1 Object Oriented Programming*, Sams Publishing, 1992.
- [FOR 56] Ford, L. R. y D. R. Fulkerson, *Maximum Flow Through a Network*, Canadian Journal of Mathematics, Vol. 8, 1956, Pags. 399-404.
- [FOR 62] Ford, L. R. y D. R. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, NJ, 1962.
- [HEK 92] Hekmatpour, S., *C++ Guía para Programadores en C*, Prentice Hall Hispanoamericana, S. A., México D. F., 1992.
- [JOH 66] Johnson, E., *Networks and Basic Solutions*, Operations Research, Vol. 14, 1966, Pags. 619-623.
- [KRV 56] Kruskal, J. B., *On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem*, Proceedings of the AMS, Vol. 7, 1956, Pags. 48-50.
- [LAW 76] Lawler, E. L., *Combinatorial Optimizations: Networks and Matroids*, Holt, Rinehart and Winston, New York, 1976.
- [MUR 92] Murty, K. G., *Network Programming*, Prentice Hall Inc., Englewood Cliffs, N.J., 1992.
- [PAP 82] Papadimitriou, C. H. and K. Steiglitz, *Combinatorial Optimization: Algorithm and Complexity*, Prentice Hall Inc., Englewood Cliffs, N.J., 1982.
- [SAK 77] Sakarovich, M., *Optimisation dans les Réseaux*, ENSIMAG, Grenoble, 1977.
- [SHE 92] Sheldon, T., *Windows 3 Manual de Referencia*, Mc.GrawHill, México, 1992.
- [ZAD 72] Zadeh, N., *Theoretical Efficiency of Edmonds-Karp Algorithm for Computing Maximal Flows*, Journal of ACM, Vol. 19, 1972, Pags. 184-192.



## UN ENFOQUE GEOMÉTRICO PARA EL ANÁLISIS DE SENSIBILIDAD EN PROGRAMACIÓN LINEAL

MARIO MARÍN<sup>1</sup>

### Resumen

Este artículo brinda un punto de vista sobre el análisis de sensibilidad. El trabajo presenta un enfoque geométrico sobre el análisis de sensibilidad para problemas de programación lineal, que explota las propiedades de los conjuntos convexos en conjugación con las ideas básicas del método simplex y la información disponible en los tableaux mismos.

**Palabras clave:** Sensibilidad, Programación Lineal, Invarianza de óptimo.

### 1 Introducción

Dado el problema de Programación Lineal (P.L.)

$$\begin{aligned} \text{Minimizar } z &= cx \\ \text{Sujeto a} \\ Ax &\leq b \\ x_1, \dots, x_n &\geq 0 \end{aligned} \tag{1}$$

donde  $A$  es una matriz de rango  $m$ ,  $x, c \in \mathbb{R}^n$  y  $b \in \mathbb{R}^m$ .

Si  $x^*$  es solución factible óptima, se plantea la pregunta qué tan sensible al cambio es  $x^*$  ante perturbaciones en cualquiera de los componentes del problema (1). A saber, cambios en el vector de costos, vector del lado derecho o los coeficientes de la matriz tecnológica.

Desde el inicio mismo de la programación lineal, se ha reconocido la importancia de estudiar y resolver este problema. Muchos esfuerzos, desde los más diferentes puntos de vista, se han realizado para lograr resolverlo de manera exacta y lo más simple posible. Este artículo da una alternativa para enfrentarlo y podría resumirse en los siguientes términos:

Dado un vértice  $x^*$  de la región factible del problema (1), se determina un conjunto de vectores en  $\mathbb{R}^n$  que al ser sustituidos por  $c$  en el (P.L) mantienen a  $x^*$  como solución óptima.

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE COSTA RICA, CARTAGO, COSTA RICA

Tal conjunto se identificará como p-cono de invarianza de óptimo asociado al vértice  $x^*$ . Se analiza cómo usar este enfoque en el análisis de sensibilidad para coeficientes de la función objetivo y se plantean algunas posibilidades de uso en otros análisis de sensibilidad.

Se divide la presentación en las siguientes secciones: 2 presenta una descripción del problema y algunas convenciones que se usan; 3 y 4 los resultados básicos; 5 algunas consideraciones sobre cómo usar los resultados y algunos comentarios adicionales; 6 dos ejemplos y 7 presenta una serie de comentarios a modo de conclusiones.

## 2 Descripción del problema

Al analizar la figura 1, observamos que para cualquier vector  $c$  en la región que se indica como el p-cono, el valor óptimo se alcanza en el vértice  $x^*$ . También es cierto que este p-cono está generado por los vectores normales a los hiperplanos  $H_1$  y  $H_2$ .

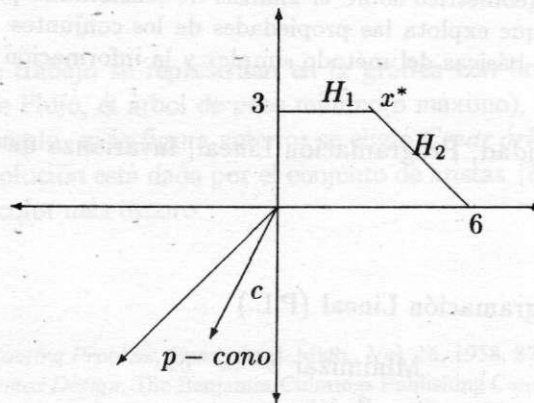


Figura 1: Cono de Invarianza

El presente trabajo da una caracterización de este p-cono en el caso general y ofrece algunas opciones sobre cómo usar esta caracterización en el análisis de sensibilidad. El interés primordial va dirigido hacia el estudio de algunos casos particulares del problema:

$$\begin{aligned} \text{Minimizar } z &= \sum_{i=1}^n (c_i + \delta_i v_i) x_i \\ \text{Sujeto a } & \\ & Ax \leq b \\ & x_1, \dots, x_n \geq 0 \end{aligned} \quad (2)$$

con solución  $x^*$  donde  $\delta_i$  es un parámetro y  $v_i$  es un valor conocido. Diferentes elecciones del vector  $v$  permitirán enfoques específicos del análisis de sensibilidad. En particular, si se tiene  $v_i = 0 \forall i \neq k$ , y  $v_k = 1$ , se estará analizando la sensibilidad para el coeficiente

$c_k$ , aún más se puede determinar un intervalo en el cual puede variar  $c_k$  sin que  $x^*$  cambie su condición de óptimo. Ese intervalo se llamará el rango de tolerancia para  $c_k$ .

Cada restricción de la forma  $x_i \geq 0$  se asociará con un semiespacio canónico de la forma  $-x_i \leq 0$ .

### 3 La región factible y el problema PL

Si  $H$  es un hiperplano de  $\mathbf{R}^n$  entonces se puede caracterizar por medio del siguiente lema:

**Lema 1** *Si  $H$  es un hiperplano en  $\mathbf{R}^n$  existe un par  $(\omega, p)$  donde  $w \in S^{n-1}$  y  $p \in \mathbf{R}^+$  tal que*

$$H = \{x \in \mathbf{R}^n / \bar{x} \cdot \bar{w} = p\} \quad (3)$$

donde  $S^{n-1}$  es la esfera de radio 1 en  $\mathbf{R}^n$ .

el vector  $\omega$  se dice normal al plano, y  $p$  es la distancia del hiperplano al origen. Además el par  $(\omega, p)$  no es único.

**Definición 1** *Dado un hiperplano  $H$  se divide el espacio en dos semiespacios cerrados.*

$$H^- = \{x \in \mathbf{R}^n / \bar{x} \cdot \bar{w} \leq p\} \quad H^+ = \{x \in \mathbf{R}^n / \bar{x} \cdot \bar{w} \geq p\} \quad (4)$$

**Definición 2** *Se llamará conjunto poliédrico a la intersección de un número finito de semiespacios cerrados.*

**Teorema 1** *En un conjunto poliédrico acotado  $S$  cada  $x \in S$  puede representarse como una combinación lineal de los puntos extremos de  $S$ , que son a la vez vértices.*

**Teorema 2** *Si  $S$  es conjunto poliédrico de  $\mathbf{R}^n$  que no contiene líneas y  $f$  es cualquier funcional lineal de  $\mathbf{R}^n$  en  $\mathbf{R}^n$ , acotado en  $S$ , entonces  $f$  alcanza sus valores extremos en un vértice de  $S$ .*

Como el conjunto de restricciones de un P.L. tiene la forma

$$\bar{x} \cdot \bar{w}_i \leq b_i \quad \forall i = 1 \dots m \quad (5)$$

la región de factibilidad del P.L. es un conjunto poliédrico. Además por las restricciones de no negatividad no puede contener ninguna línea.

Como una aplicación inmediata de la observación y el teorema anterior se obtiene el siguiente teorema:

**Teorema 3** *Si el P.L. tiene región de factibilidad  $R$  no vacía y la función objetivo es acotada sobre  $R$  entonces alcanza su valor óptimo en un vértice de  $R$ .*

## 4 El p-cono de invariabilidad de óptimo asociado a una solución básica factible

Olvidando un poco la formalidad, se puede decir que la esencia del método Simplex es:

- Seleccionar un extremo  $x$  de la región de factibilidad.
- Considerando todas las aristas que llegan a  $x$ , si la función objetivo no decrece a lo largo de ninguna de ellas, entonces el  $x$  seleccionado es óptimo.
- Caso contrario, se traslada el punto  $x$  a través de una de esas aristas que genera descenso en la función objetivo, para llegar a una nueva solución factible y continuar iterando.

Se puede establecer la siguiente sucesión de resultados que conducirán a la caracterización del p-cono de invariabilidad de óptimo asociado a la solución básica factible  $x^*$ .

**Teorema 4** *Cada solución básica factible  $x^*$ , vértice de la región, es la intersección de  $n$  hiperplanos. Cada arista que sale de un vértice es la intersección de  $n - 1$  de los  $n$  hiperplanos que se intersecan en el vértice.*

*Nótese que aquí se incluyen los hiperplanos canónicos,  $x_i = 0$  para  $i = 1 \dots n$ .*

**Demostración :** Como  $x^*$  es solución básica factible, después de reordenar columnas, si fuera necesario, se tendría:

$$x^* = (x_{B_1}, \dots, x_{B_k}, x_{N_1}, \dots, x_{N_{n-k}})$$

$$A = [B, N]$$

$$Bx_B = b \tag{6}$$

$$Nx_N = 0 \tag{7}$$

De (6)  $x^*$  satisface las ecuaciones de  $k$  hiperplanos de  $Ax = b$  de (7)  $x^*$  satisface las ecuaciones de  $n - k$  hiperplanos canónicos.

Además si  $L$  es una arista que sale de  $x^*$  debe satisfacer al menos  $n - 1$  hiperplanos. Como  $x^* \in L$  entonces se concluye el resultado.

**Definición 3** *Si  $x^*$  solución óptima, se tiene que  $x^*$  es la intersección de los hiperplanos  $H_1, \dots, H_n$ . Estos hiperplanos están caracterizados por los pares:*

$$(\omega_1, p_1), \dots, (\omega_n, p_n)$$

*respectivamente. Se define el p-cono de invariabilidad de óptimo asociado a  $x^*$  por*

$$I = \left\{ c = \sum_{i=1}^n -\alpha_i \omega_i \quad \text{con } \alpha_i \geq 0 \quad \forall i \right\}$$

**Teorema 5** Si  $x^*$  es una solución básica factible óptima al problema 1.1, entonces  $x^*$  continúa como solución básica factible óptima si el valor del vector de costos  $c$  se sustituye por cualquier valor  $c'$  en el  $p$ -cono de invariabilidad de óptimo asociado a  $x^*$

**Demostración:** Si  $x$  está en cualquier arista debe satisfacer  $n - 1$  hiperplanos. Sin perder generalidad los  $n - 1$  primeros, como

$$c'x^* = \sum_{i=1}^n -\alpha_i \omega_i x^* = \sum_{i=1}^n -\alpha_i p_i$$

mientras que

$$c'x = \sum_{i=1}^n -\alpha_i \omega_i x = \sum_{i=1}^{n-1} -\alpha_i p_i - \alpha_n \omega_n x$$

como  $x$  debe estar en la región factible entonces  $\omega_n x \leq p_n$  lo cual implica que:  $c'x^* \leq c'x \quad \forall x$  en una arista que llegue al vértice  $x^*$  y así  $x^*$  es óptimo. Como una aplicación de este teorema se obtiene el siguiente teorema:

**Teorema 6** Si  $c = \sum_{i=1}^n -\alpha_i \omega_i$  con  $\alpha_i = 0$  para algún  $i$  entonces  $c$  tiene tolerancia cero para la variable básica  $i$ .

## 5 Consideraciones sobre los cálculos por realizar

De la solución básica factible óptima  $x^*$  obtenida y de los tableaux inicial y final, se puede obtener la información necesaria para hacer este análisis. La aplicación más básica sería, dado un vector perturbado  $c'$  verificar si está en el  $p$ -cono de invariabilidad de óptimo para  $x^*$ , y esto se logra resolviendo el sistema

$$\sum_{i=1}^n -\alpha_i \omega_i = c'$$

y verificando que cada  $\alpha_i \geq 0$ .

**Teorema 7** La matriz asociada al sistema:

$$-\alpha_1 \omega_1 - \alpha_2 \omega_2 - \dots - \alpha_n \omega_n = c'$$

tiene la forma:

$$A = \begin{bmatrix} I_{k,k} & D_{k,n-k} \\ 0_{n-k,k} & C_{n-k,n-k} \end{bmatrix},$$

donde

$$[-\omega_1^t, \dots, -\omega_{n-k}^t] = \begin{bmatrix} D_{k,n-k} \\ C_{n-k,n-k} \end{bmatrix}$$

para los  $n-k$  hiperplanos no canónicos.

**Teorema 8** Si la matriz  $A$  tiene la forma

$$A = \begin{bmatrix} I_{k,k} & D_{k,n-k} \\ 0_{n-k,k} & C_{n-k,n-k} \end{bmatrix}$$

y  $C_{n-k,n-k}$  es invertible. Entonces

$$A^{-1} = \begin{bmatrix} I_{k,k} & -D_{k,n-k}C_{n-k,n-k}^{-1} \\ 0_{n-k,k} & C_{n-k,n-k}^{-1} \end{bmatrix}$$

La demostración del teorema 8 es producto de la posibilidad de reordenar las filas y las columnas de un sistema de ecuaciones. El teorema 8 se demuestra verificando el producto. Finalmente la matriz  $C_{mm}^{-1}$  se puede obtener del tableau óptimo como la matriz resultado de multiplicar por  $-1$  la transpuesta de la matriz básica inversa.

## 6 Ejemplos

Considere el siguiente ejemplo, adaptado de Dantzig (1963)

$$\begin{aligned} &\text{Minimizar} && -12x_1 - 20x_2 - 18x_3 - 40x_4 \\ &\text{Sujeto a} && \\ &4x_1 + 9x_2 + 7x_3 + 10x_4 &\leq & 6000 \\ &x_1 + x_2 + 3x_3 + 40x_4 &\leq & 4000 \\ &x_1, x_2, x_3, x_4 &\geq & 0 \end{aligned}$$

con tableaux inicial y final dados como sigue

		$z$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	LD
0	$z$	-1	-12	-20	-18	-40	0	0	0
1	$x_5$	0	4	9	7	10	1	0	6000
2	$x_6$	0	1	1	3	40	0	1	4000

		$z$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	LD
0	$z$	-1	0	$\frac{20}{3}$	$\frac{10}{3}$	0	$\frac{44}{15}$	$\frac{4}{15}$	18667
1	$x_1$	0	1	$\frac{7}{3}$	$\frac{5}{3}$	0	$\frac{4}{15}$	$-\frac{1}{15}$	$\frac{4000}{3}$
2	$x_4$	0	0	$-\frac{1}{30}$	$-\frac{1}{30}$	1	$-\frac{1}{150}$	$\frac{2}{175}$	$\frac{1000}{15}$

La solución obtenida es  $x_1 = \frac{4000}{3}$ ,  $x_2 = 0$ ,  $x_3 = 0$  y  $x_4 = \frac{1000}{15}$  de donde el cono de invariabilidad  $I$  para este vértice está generado por los vectores:

$$(-4, -9, -7, -10), (0, 1, 0, 0), (0, 0, 1, 0), (-1, -1, -3, -40)$$

y para verificar si un valor  $c'$  está en este cono, bastaría con resolver el sistema siguiente y comprobar que su solución es positiva:

$$\alpha_1(-4, -9, -7, -10) + \alpha_2(0, 1, 0, 0) + \alpha_3(0, 0, 1, 0) + \alpha_4(-1, -1, -3, -40) = c' \quad (8)$$

después de intercambios de filas 1 con 2 y 2 con 3 la matriz asociada a este sistema tiene la forma

$$A = \begin{bmatrix} 1 & 0 & -9 & -1 \\ 0 & 1 & -7 & -3 \\ 0 & 0 & -4 & -1 \\ 0 & 0 & -10 & -40 \end{bmatrix}$$

que tiene por inversa

$$A^{-1} = \begin{bmatrix} 1 & 0 & -\frac{7}{3} & \frac{1}{30} \\ 0 & -1 & -\frac{1}{3} & -\frac{1}{30} \\ 0 & 0 & -\frac{4}{15} & \frac{1}{150} \\ 0 & 0 & \frac{1}{15} & \frac{2}{75} \end{bmatrix}$$

Si en el ejemplo anterior se estudia la sensibilidad de  $c_1$ , entonces debería analizarse en qué casos el sistema (8) tiene solución positiva si

$$c' = (-20, -18, -12 + \lambda, -40).$$

La solución tiene la forma:

$$A^{-1}c' = \begin{bmatrix} -\frac{7}{3}\lambda + \frac{20}{3} \\ -\frac{5}{3}\lambda + \frac{10}{3} \\ -\frac{4}{15}\lambda + \frac{44}{15} \\ \frac{1}{15}\lambda + \frac{4}{15} \end{bmatrix}$$

lo cual da un intervalo de variación  $\lambda \in [-4, 2]$  y un rango de  $[-16, -10]$  para  $c_1$ .

Un análisis similar permite obtener las tolerancias para los demás coeficientes.

$$\left[-\frac{20}{3}, \infty\right[ \text{ para } c_3$$

$$\left[-\frac{10}{3}, \infty\right[ \text{ para } c_2$$

$$[-240, -10] \text{ para } c_4.$$

Como ejemplo adicional considérese el siguiente caso adaptado de Hillier-Lieberman (1982) [4].

La *Wyndor Glass Co.* es un productor de vidrio de alta calidad, incluyendo ventanas y puertas de vidrio. Tiene tres plantas. Los marcos de aluminio y la herrería se hacen en la planta 1; los marcos de madera se fabrican en la planta 2 y la planta 3 se usa para producir el vidrio y montar los productos. La empresa tiene cierta capacidad de producción libre y se han solicitado dos nuevos productos. Uno de estos (producto 1) es una puerta de vidrio (de 8 por 8 pies) con marco de aluminio. El otro (producto 2) es una ventana grande (de 4 por 6 pies) con marco de madera y de guillotina.

Se ha determinado que se podría vender, de cualquiera de los dos productos, la cantidad que pudiera producirse con la capacidad disponible. Sin embargo, como los dos productos competirán por la misma capacidad de producción en la planta 3, no se ve claro qué mezcla

de productos será la más ventajosa. Además la Empresa está interesada en conocer la sensibilidad de la decisión tomada ante variaciones en la utilidad de los productos

En la siguiente tabla se resume la información con respecto a las capacidades disponibles por planta, las capacidades de esa planta requeridas por unidad producida de cada producto y la utilidad unitaria de cada producto.

	Cap requerida por unidad Prod 2	Cap. requerida por unidad Prod 1	Capacidad disponible
planta 1	1	0	4
Planta 2	0	2	12
Planta 3	3	2	18

Considerando que la utilidad por unidad del producto 1 es de 3 y por el producto 2 es de 5 el problema se formula como el P.L. siguiente:

$$\text{Minimizar } -3x_1 - 5x_2$$

Sujeto a:

$$x_1 \leq 4$$

$$2x_2 \leq 12$$

$$3x_1 + 2x_2 \leq 18$$

$$x_1, x_2 \geq 0$$

(9)

Donde  $x_1$  y  $x_2$  representan las cantidades de producto 1 y producto 2 que se producen.

Los tableaux inicial y final están dados como sigue:

Ec N.	Var. Bas.	z	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	LD
0	z	-1	-3	-5	0	0	0	0
1	$x_3$	0	1	0	1	0	0	4
2	$x_4$	0	0	2	0	1	0	12
3	$x_5$	0	3	2	0	0	1	18

Ec N.	Var. Bas.	z	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	LD
0	z	-1	0	0	0	$\frac{3}{2}$	1	36
1	$x_3$	0	0	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	2
2	$x_2$	0	0	1	0	$\frac{1}{2}$	0	6
3	$x_1$	0	1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	2

La solución obtenida es  $x_1 = 2$ , y  $x_2 = 6$ .

Además el cono de invariabilidad  $I$  para este vértice está generado por los vectores:

$$(0, -2), (-3, -2)$$

Para verificar si un vector de costos  $c'$  está en  $I$  el sistema

$$\lambda(0, -2) + \beta(-3, -2) = c',$$



debe tener solución positiva. De los tableaux se obtiene que las matrices  $A$  y  $A^{-1}$  asociadas a este sistema son:

$$A = \begin{bmatrix} 0 & -3 \\ -2 & -2 \end{bmatrix}$$

que tiene por inversa

$$A^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{2} \\ -\frac{1}{3} & 0 \end{bmatrix}$$

Si se quiere estudiar la sensibilidad de  $c_1$ , entonces debería analizarse en qué casos el sistema anterior tiene solución positiva si  $c' = (-3 + \lambda, -5)$ .

La solución tiene la forma:

$$A^{-1}c' = \begin{bmatrix} -1 + \frac{1}{3}\lambda + \frac{5}{2} \\ 1 - \frac{1}{3}\lambda \end{bmatrix}$$

lo cual da un intervalo de variación  $\lambda \in [-\frac{9}{2}, 3]$  y el rango de tolerancia  $[-\frac{15}{2}, 0]$  para  $c_1$ .

Un análisis similar permite obtener el rango de tolerancia  $]-\infty, 2]$  para  $c_2$

## 7 Conclusiones

Se logra caracterizar de una manera, ante todo fácil de llevar a la práctica, una metodología para hacer análisis de sensibilidad en el problema de programación lineal perturbado en uno o varios de los coeficientes de la función objetivo. Toda la información requerida se puede obtener de los tableaux inicial y óptimo del método simplex.

Una ventaja del método radica en que no requiere de cálculos complicados y sería viable de programar. Además, sólo una vez debe obtenerse la matriz de sensibilidad de costos, y cualquier análisis para un nuevo vector  $c'$  solamente requiere del producto de la matriz por éste.

Otra posibilidad, como se demostró en el ejemplo, es que permite realizar análisis de rango de tolerancia para cualquier costo de la función objetivo. Cabe aquí destacar que el análisis de sensibilidad en bloque, como se propone en [Wendal 1985], puede resultar muy atractivo pero también puede quedarse corto si se piensa que las perturbaciones de varios costos simultáneamente están estrechamente ligadas por la matriz de sensibilidad.

Este enfoque puede usarse para caracterizar vectores de costo sensibles, que corresponden con vectores en las faces del  $p$ -cono o vectores más sensibles aún, si quedan sobre aristas de éste. De igual manera puede explotarse el conocimiento de estas regiones para determinar qué costos deben de mantenerse muy estables para tener un óptimo dado o bien, qué modificación de costos producirá un óptimo específico de los posibles en la región.

Una posibilidad que queda abierta es el estudio de variaciones simultáneas de componentes, cómo alteran la región factible y cuándo se mantendrá la condición óptimo de un vértice dado.

## Bibliografía

- [1] Bazaraa, M. S. y Jarvis J. J. (1977) *Linear Programming and Network Flows*. Wiley & Sons, New York.
- [2] Dantzig, G. B. (1963) *Linear Programming and Extensions*. Princeton University Press, Princeton.
- [3] Gal, T. y Nedona, J. (1972) "Multiparametric linear programming", *Management Sci.* **18**: 406-422.
- [4] Hillier, L. y Lieberman, G. (1980) *Introducción a la Investigación de Operaciones*. McGraw-Hill, México.
- [5] Lay S.R. (1982) *Convex Sets and Their Applications*. John Wiley & Sons, New York.
- [6] Ravi, N. y Wendell, R. (1991) "A tolerance approach to sensitivity analysis of matrix coefficients in linear programming", *Management Sci.* **6**: 333-347.
- [7] Steuer, R. (1981) "Algorithms for linear programming problems with interval objective functions coefficients", *Mathematics of Operations Research* **3**: 333-347.
- [8] Wendell, R. (1985) "A tolerance approach to sensitivity analysis in linear programming", *Management Sci* **31**: 564-567.

## COMPARACIÓN DE ÍNDICES EN LA CLASIFICACIÓN JERÁRQUICA ASCENDENTE

ALEX MURILLO<sup>1</sup>

### Resumen

Este artículo se dedicará a presentar un resumen de algunos índices utilizados para determinar los niveles significativos de los árboles jerárquicos de clasificación, asociándole a estos niveles las mejores particiones de una clasificación jerárquica automática ascendente, cuyas clases tengan los individuos más próximos y las clases sean lo más separadas entre sí. Luego se dará una comparación en algunas tablas de datos reales y simulados.

**Palabras claves:** clasificación, clasificación automática, índice, árbol jerárquico, nivel significativo, partición.

### 1 Introducción

El proceso de clasificación permite sintetizar la información y dar una visión global de un problema en estudio; además, brinda una representación esquemática simple, a partir de una tabla rectangular de datos.

El resultado que aporta una clasificación tiene una utilidad de significado en la información y el objetivo más simple de la clasificación consiste en repartir los objetos a clasificar, en clases, lo más homogéneas posibles, donde cada una sea bien diferente de la otra, tal que objetos de la misma clase sean más parecidos entre ellos, a los objetos pertenecientes a clases diferentes, es decir, se desea procesar simultáneamente las semejanzas y las diferencias entre los objetos.

Las jerarquías, con la comodidad de sus interpretaciones visuales, constituyen desde hace mucho tiempo, una forma de clasificación muy popular y son frecuentemente utilizadas con el interés de tener clases bien significativas. En este artículo, se seguirán usando jerarquías binarias.

Entre los métodos de clasificación tenemos los agregativos o ascendentes y los divisivos o descendentes, los cuales no obtienen en general los mismos resultados.

Los algoritmos ascendentes, o también llamados de aglomeraciones sucesivas, son los más antiguos, los más comunes y los más corrientemente utilizados. Entre sus ventajas está la rapidez, y generalmente dan buenos resultados.

<sup>1</sup>PIMAD, ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

En el algoritmo general de clasificación jerárquica ascendente (CJA) se supone que existen disimilitudes, entre todo par de objetos; se procede a reunir los dos objetos más próximos y se recalculan las distancias entre la nueva clase y los restantes, se reitera hasta que todos los objetos se aglomeren en una sola clase; obteniendo un árbol jerárquico, donde los nodos representan la fusión sucesiva y la altura el valor de la distancia entre los dos objetos o clases fusionados, conocida como índice de la jerarquía. La dificultad reside en la elección de una fórmula para recalculan las distancias entre las fusiones que reflejen fielmente los objetos. Si se tienen  $n$  objetos, el algoritmo CJA sólo necesitan comparar  $\frac{(n-k)(n-k-1)}{2}$  fusiones, donde  $n - k$  es el número de clases en la iteración  $k$ , con  $k = 1, \dots, n - 1$ .

## 2 Índices que determinan los niveles significativos

Se presenta un resumen de los índices más utilizados en la bibliografía, con el fin de determinar los niveles significativos de los árboles jerárquicos binarios, que ayudarán a la escogencia de las mejores particiones en un sentido que será definido.

### 2.1 Índice de variables

Se basa en la utilización de los conjuntos difusos generados como técnica de cálculo e interpretación para hacer un conteo de variables [11, 12].

Se considerará una tabla de datos de tamaño  $n \times p$ , donde  $n$  es el número de individuos a clasificar del conjunto  $\Omega$  y  $p$  es el número de variables cuantitativas positivas que fueron medidas sobre los individuos de  $\Omega$ . Así,  $x_i^j$  es el valor de la variable  $v_j$  medida sobre el individuo  $x_i \in \Omega$ , luego se trasladan al intervalo  $[0, 1]$ , por  $y_i^j = \frac{x_i^j}{s_j}$ , donde  $s_j = \sum_{i=1}^n x_i^j$ .

Se aplica la primera iteración del algoritmo CJA a la tabla de datos trasladada y se consideran los  $p$  conjuntos difusos en el universo  $\Omega^{(0)} = \Omega$ :

$$\tilde{v}_j = \left\{ (x_i, y_i^j) : x_i \in \Omega, \text{ y su respectivo } y_i^j \right\}, \text{ con } j = 1, \dots, p.$$

Sea  $\alpha$  el *parámetro de significancia* en  $[0, 1]$ , el cual representa el nivel de exigencia mínimo que el individuo  $x_i \in \Omega$  debe satisfacer en la variable  $v_j$  para ser tomado en cuenta en el conjunto de nivel  $\alpha$ , esto es:

$$(v_j)_\alpha^0 = \left\{ x_i \in \Omega : y_i^j \geq \alpha \right\}, \text{ con } j = 1, \dots, p.$$

Así, una variable  $v_j$  es bien significativa para representar separabilidad entre grupos, con respecto a esta variable, si  $|(v_j)_\alpha^0| \geq \beta n \geq 1$ , siendo  $\beta \in [0, 1]$  un porcentaje dado en forma empírica, llamado *parámetro de representatividad*.

Ahora, se calcula un porcentaje de las variables bien significativas, denominado *índice de variables*, en notación se tiene:

$$\mathcal{V}_0(\alpha) = \frac{1}{p} \left| \left\{ v_j : |(v_j)_\alpha^0| \geq \beta n \geq 1 \right\} \right|.$$

En la siguiente iteración del algoritmo CJA, el universo cambia a  $\Omega^{(1)} = \{\{x_1\}, \dots, \{x_i, x_j\}, \dots, \{x_n\}\}$ , y generalizando a la iteración  $k$ , el universo es la partición  $\Omega^{(k)}$ . Entonces es importante tomar en cuenta los conjuntos difusos generados en el universo  $\Omega^{(k)} \subset \mathcal{P}(\Omega)$  por el conjunto difuso  $\tilde{v}_j$ . La función de pertenencia, en la iteración  $k$ , es la suma acotada de los valores de los individuos en la misma clase  $h$ , esto es

$$\tilde{v}_j = \left\{ \left( h, \mu_{\tilde{v}_j} (h) \right) : h \in \Omega^{(k)} \right\}, \text{ en donde } \mu_{\tilde{v}_j} (h) = \min \left\{ 1, \sum_{x_i \in h} y_i^j \right\}.$$

Por lo tanto, el índice de variables a considerar en la iteración  $k = 0, \dots, n - 1$  es:

$$\mathcal{IV}_k(\alpha) = \frac{1}{p} \left| \left\{ v_j : |(v_j)_\alpha^k| \geq \beta(n - k) \geq 1 \right\} \right|, \text{ si } (v_j)_\alpha^k = \left\{ h \in \Omega^{(k)} : \mu_{\tilde{v}_j} (h) \geq \alpha \right\}.$$

Considere un valor del parámetro de significancia  $\alpha$  fijo pero arbitrario en el intervalo  $[0, 1]$ , se define como salto del índice de variables en  $\alpha$  al valor  $\Delta \mathcal{IV}_k(\alpha) = \mathcal{IV}_k(\alpha) - \mathcal{IV}_{k-1}(\alpha)$ . Si este salto es significativamente grande en la iteración  $k$ , indica que la cantidad de variables, que determinan la separabilidad de las clases, aumentó drásticamente.

Considere una escala de parámetros de significancia de la siguiente forma:  $\alpha_{\min} < \alpha_{\min} + e < \alpha_{\min} + 2e < \dots < \alpha_{\max}$ , con  $\alpha_{\min}, \alpha_{\max} \in [0, 1]$ , con  $e$  un valor de espaciado fijo, dado en forma empírica.

Se observa que si la variable  $v_j$  representa separabilidad, a un nivel  $\alpha$  de la escala, pero disminuye su separabilidad al siguiente nivel de la escala, se interpreta como el mayor nivel para el cual la variable representa separabilidad. Generalizando esta idea, se obtiene que si  $\mathcal{IV}_k(\alpha) > \mathcal{IV}_k(\alpha + e)$ , algunas variables están en su mayor nivel de separabilidad.

Como se desea encontrar la iteración  $k$  del algoritmo CJA en la cual la partición correspondiente  $\Omega^{(k)}$  sea de clases bien separadas entre sí, lo que se obtiene encontrando el mayor salto  $\Delta \mathcal{IV}_k(\alpha)$  para el cual las variables están en el mayor nivel que representa separabilidad, y esto se logra para:

$$\max_{k=0, \dots, n-1} \max_{\alpha \in J} \Delta \mathcal{IV}_k(\alpha), \text{ en donde } J = \{\alpha_{\min}, \alpha_{\min} + e, \alpha_{\min} + 2e, \dots, \alpha_{\max}\}.$$

Tomando en cuenta que puede haber dos iteraciones diferentes de salto máximo iguales, en tal caso se escogería la que tenga mayor parámetro de significancia. Sin embargo, este puede no ser único, en este caso se le deja al investigador analizar cuál partición es la más conveniente para los requerimientos del estudio que está efectuando.

Por lo tanto, la iteración obtenida anteriormente busca maximizar la separabilidad entre las clases de la partición.

Más detalles sobre la definición de este índice se encuentran en [11].

## 2.2 Índice de individuos

Este índice utiliza conjuntos difusos para el cálculo e interpretación y se basa en un conteo de individuos [11, 12].

Se parte de la misma tabla de datos  $n \times p$  antes mencionada y se aplicará el algoritmo CJA, pero ahora se consideran los conjuntos difusos:

$$\tilde{x}_i = \left\{ (v_j, y_i^j) : v_j \text{ es una variable, con } j = 1, \dots, p \right\}, \text{ donde } i = 1, \dots, n.$$

El conjunto de nivel  $\alpha$  correspondiente al conjunto difuso  $\tilde{x}_i$  es

$$(x_i)_\alpha = \left\{ v_j : y_i^j \geq \alpha, j = 1, \dots, p \right\}, \text{ con } i = 1, \dots, n.$$

Sea  $\Omega^{(k)} = \{P_1, \dots, P_{n-k}\}$  la partición de  $\Omega$  en la iteración  $k$ , con  $k = 0, \dots, n-1$ .

El índice de individuos, en la primera iteración es

$$\mathcal{I}_0(\alpha) = \frac{1}{n} \left| \left\{ x_i \in \Omega^{(0)} : |(x_i)_\alpha| \geq \beta p \geq 1 \right\} \right|.$$

Se denotará con  $\tilde{P}_l$ , con  $l = 1, \dots, n-k$  el conjunto difuso que se obtiene al sumar todos los conjuntos difusos  $\tilde{x}_i$  tal que  $x_i \in P_l$ , en notación simbólica se tiene:

$$\tilde{P}_l = \sum_{x_i \in P_l} \tilde{x}_i = \left\{ (v_j, \mu_{\tilde{P}_l}(v_j)) : v_j \text{ es variable, con } j = 1, \dots, p \right\},$$

en donde  $\mu_{\tilde{P}_l}(v_j) = \sum_{x_i \in P_l} y_i^j$ , con  $l = 1, \dots, n-k$ . Además, se cumple  $\mu_{\tilde{P}_l}(v_j) \leq 1$ .

Por lo tanto, en cada iteración  $k$  del algoritmo CJA, hay  $n-k$  conjuntos difusos  $\tilde{P}_l$ , es por esta razón que en el cálculo del índice de individuos, cuando se generaliza a todas las iteraciones, se debe ponderar por la cardinalidad de  $P_l$ , para que siempre refleje un porcentaje del total de los individuos de  $\Omega$ , entonces el índice sería:

$$\mathcal{I}_k(\alpha) = \frac{1}{n} \sum_{P_l \in Q_k(\alpha)} |P_l|, \text{ en donde } Q_k(\alpha) = \left\{ P_l \in \Omega^{(k)} : |(P_l)_\alpha| \geq \beta p \geq 1 \right\}.$$

El índice de individuos lo que obtiene son las clases que sobrepasan el parámetro de significancia mínimo en un valor del parámetro de representatividad aceptable, lo que se interpreta como las clases homogéneas.

Por lo tanto, si se maximiza el salto  $\Delta \mathcal{I}_k(\alpha)$  se encuentra la iteración  $k$  en la cual las clases son lo más homogéneas entre sí, en notación se tiene

$$\max_{k=0, \dots, n-1} \max_{\alpha \in J} \Delta \mathcal{I}_k(\alpha), \text{ en donde } J = \{\alpha_{\min}, \alpha_{\min} + e, \alpha_{\min} + 2e, \dots, \alpha_{\max}\} \subseteq [0, 1].$$

Se recomienda que ante diferentes iteraciones con salto máximo igual se escoja la iteración que tenga mayor nivel de exigencia  $\alpha$ , sin embargo, ésta puede no ser única; en este caso, se le deja al experto decidir cuál es la partición más conveniente.

Por lo tanto, este índice encuentra una o varias iteraciones en la que los grupos son lo más homogéneos posibles.

Más detalles sobre la definición de este índice se encuentran en [11].

### 2.3 Índice del codo

El índice del codo es muy popular, pero nosotros desconocemos donde se originó. Consiste en encontrar el máximo cambio de inercia intra-clase de una iteración a la siguiente.

Sean  $\Omega, h_{n-2}, \dots, h_1$  los niveles o nodos de la jerarquía indexada y  $w_{n-1}, w_{n-2}, \dots, w_1$  las inercias intra-clases en cada nivel. Sea  $\Delta w_k = w_k - w_{k-1}$  el cambio de inercia del nivel  $h_{k-1}$  al nivel  $h_k$  con  $k = 2, \dots, n-1$ . El método del codo consiste en encontrar la iteración  $k$  tal que

$$\max_{k=2, \dots, n-1} \frac{\Delta w_k}{\Delta w_{k-1}}$$

La analogía gráfica, consiste en encontrar el lugar donde el codo del gráfico  $(h_k, \Delta w_k)$  es el más pronunciado, con igual espaciado en el eje de las abscisas.

### 2.4 Índice de Diday

Con el fin de detectar una partición bien significativa, de la jerarquía indexada binaria, Diday [2] sugiere utilizar un corte simple más o menos arbitrario.

Considere la curva formada por los puntos  $(h_k, w_k)$  donde  $k = 1, \dots, n-1$ , siendo  $h_k$  los elementos de la jerarquía binaria y  $w_k$  los índices correspondientes; con los  $h_k$  distribuidos uniformemente sobre el eje de las abscisas y el orden de su construcción es dado por el algoritmo CJA. Si la curva crece rápidamente en  $h_k$ , se retiene la partición, desgraciadamente, esta partición es difícilmente justificable, en sentido de un criterio de optimización que sea matemáticamente bien definido.

Diday lo que desea encontrar es el máximo crecimiento de la curva, esto es análogo a maximizar pendientes y como los  $h_k$  están igualmente espaciados, corresponde a

$$\max_{k=2, \dots, n-1} \Delta w_k, \text{ donde } \Delta w_k = w_k - w_{k-1}.$$

Este índice es usado por muchos investigadores. Sin embargo, Everitt lo cataloga como un método informal [3].

### 2.5 Índice de Jambu

Jambu utiliza varios índices para que un programa extraiga automáticamente la clase de individuos con información significativa, se escogió uno de ellos para nuestro análisis comparativo. Sea  $c_k$  la contribución absoluta de la inercia intra-clase en el nivel  $k$  con  $k = 1, \dots, n-1$ , además, sea  $I(\Omega)$  la inercia total de  $\Omega$ .

Se elige como nivel de corte, el primer nivel número  $N$  tal que

$$\sum_{k=1}^N \frac{c_k}{I(\Omega)} \geq \gamma,$$

donde  $\gamma$  es un valor fijo, que representa un porcentaje de varianza acumulada a partir del nivel más alto de la clasificación.

Estos índices permiten eliminar del estudio, los primeros niveles de la clasificación, pues el espíritu del método de clasificación jerárquica establece que las clases más altas son de varianza más fuerte y las clases más bajas son de varianza más débil [5].

## 2.6 Índice de Lerman

Para cortar el árbol jerárquico en uno o en varios niveles que correspondan a las mejores particiones, tal que la partición seleccionada maximice el número de veces en que ocurra lo siguiente: todo par de individuos de una misma clase de la partición escogida, son más próximos que cualquier otro par de individuos en clases distintas.

Sobre la base de este principio, Lerman [6, 7] propuso un índice para identificar las mejores particiones de una jerarquía binaria dada. Este índice fue estudiado y aplicado en [1]. Si la proximidad entre los individuos es dada por una matriz, cuyas entradas debajo de la diagonal son todas diferentes, entonces se construye un índice  $S(k)$ , llamado estadística global, de la forma siguiente: para cada partición  $P_k$  de la jerarquía, sea

$$S(k) = \frac{z_k - rs/2}{\sqrt{rs(f+1)/12}},$$

en donde

- $f = \frac{n^2 - n}{2}$ .
- $s$  es el número de pares de individuos que están en clases distintas de  $P_k$ .
- $r$  es el número de pares de individuos que están en la misma clase de todas las clases  $P_k$ . Observe que  $f = r + s$ .
- $z_k$  es el número de veces que para  $P_k$  se cumple lo siguiente: sean  $a, b, c, d$  cuatro individuos distintos del conjunto a clasificar, entonces la proximidad entre  $a$  y  $b$  es menor que la proximidad entre  $c$  y  $d$  si y sólo si  $a$  y  $b$  están en distintas clases mientras que  $c$  y  $d$  están en la misma clase.

Normalmente ocurre que los valores  $S(1), \dots, S(n-1)$  muestran una tendencia global creciente, exhibiendo máximos locales; las particiones que corresponden a los picos más pronunciados son las mejores. A estas particiones se les llama niveles significativos o particiones significativas. Así, el analista escogerá una partición, teniendo en cuenta este criterio y el número de clases.

Empíricamente se ha observado que cuando se produce un crecimiento *acelerado* de los valores  $S(k)$  hasta un nivel y luego *desacelerado*, la clase formada en ese nivel puede, con frecuencia, ser objeto de alguna interpretación. Por lo tanto, conviene calcular las primeras diferencias de la estadística global  $D(k) = S(k) - S(k-1)$  y ubicar sus máximos locales. Los nodos correspondientes se denominan nodos significativos o clases significativas y la función  $D(k)$  estadística local.



### 2.7 Índice de Mojena

Mojena sugiere un procedimiento basado en el tamaño relativo de la diferencia de niveles en el árbol. En detalle, la propuesta es seleccionar la partición correspondiente al primer nivel  $k$  del árbol que satisfaga

$$w_k > \bar{w} + t\sigma_w, \text{ con } k = 1, \dots, n - 1,$$

donde  $\bar{w}$  y  $\sigma_w$  es el promedio y la desviación estándar respectivamente de las inercias intra-clase  $w_1, \dots, w_{n-1}$ . Además,  $t$  es una constante, que Mojena sugiere que el valor esté en el rango de 2.75 a 3.5 para que dé buenos resultados [9, 3].

### 2.8 Índice de Milligan y Cooper

Milligan y Cooper reportan detalladamente, en una investigación [8] acerca de índices, que el índice de Mojena está dentro de los más satisfactorios. Además, ellos sugieren que el valor de  $t$  para la regla de Mojena debe ser 1.25.

## 3 Comparación de los índices en los ejemplos numéricos

En la tabla 1 se hace una comparación entre los órdenes de ejecución de los índices analizados en este trabajo, con la siguiente notación. Sea  $n$  el número de individuos a clasificar y sea  $p$  el número de variables observadas sobre los individuos de  $\Omega$ , generalmente se tiene que  $p \leq n$  y el valor  $c$  se puede considerar constante, pues es el número de escalones de  $\alpha$ .

Índice	Orden
Índice de variables	$O(n^3 + cn^2p)$
Índice de individuos	$O(n^3 + cn^2p)$
Índice del codo	$O(n^3 + n^2p)$
Índice de Diday	$O(n^3 + n^2p)$
Índice de Jambu	$O(n^3 + n^2p)$
Índice de Mojena	$O(n^3 + n^2p)$
Índice de Milligan y Cooper	$O(n^3 + n^2p)$
Índice de Lerman	$O(n^5 + n^2p)$

Tabla 1: Orden de ejecución de los índices

En la tabla 2 se encuentran los resultados numéricos, donde los renglones son los diferentes índices que se compararon y las columnas son los seis ejemplos que se analizaron. En los lugares que hay un guión, significa que el índice correspondiente no recomienda ninguna partición.

En los índices de variables y de individuos se utilizó la escala de valores 10, 20, 30, 40, 50, 60, 70, 80, 90, del parámetro de significancia  $\alpha$ , y el parámetro de representatividad entre 0.2 y 0.4.

El ejemplo de ocho puntos en el plano  $XY$ , se utilizó en [11], con el fin de analizar el comportamiento de los índices mencionados en este trabajo, en un ejemplo tal que se conozca a priori la clasificación más satisfactoria; los ocho puntos que se utilizaron están graficados en la figura 1.

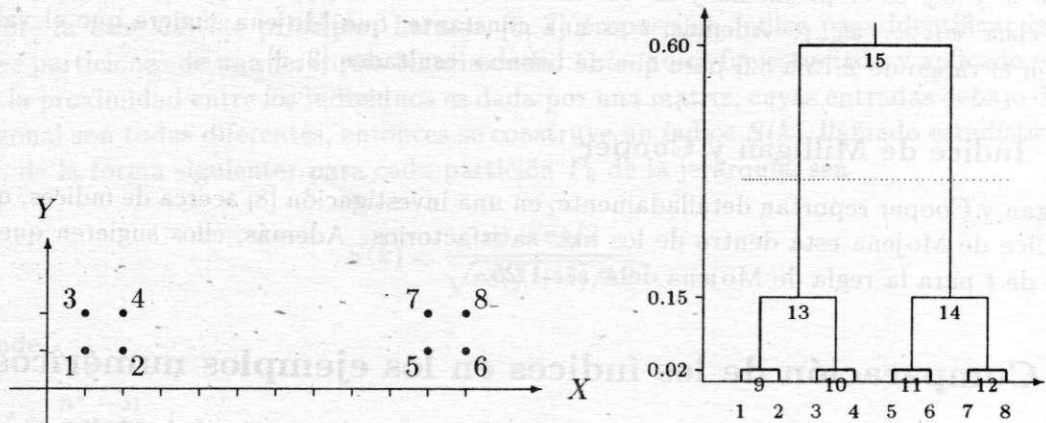


Figura 1: Ocho puntos en el plano  $XY$  y el árbol jerárquico

El árbol correspondiente a la jerarquía obtenida de CJA con la disimilitud euclídea clásica y la agregación del incremento de la inercia de Ward, se observa en la figura 1. La línea horizontal que corta el árbol jerárquico corresponde a la partición sugerida. Es bastante claro, de la tabla 2, pues la mayoría de índices afirma que el árbol hay que partirlo antes del nodo 15, asociándole la partición  $\{1, 2, 3, 4\}$ ,  $\{5, 6, 7, 8\}$  de  $\Omega$ .

En los siguientes ejemplos por efectos de espacio no se presentarán los árboles jerárquicos, pero todos los detalles de estos resultados se encuentran en [11].

Los datos del ejemplo PSYSOC se tomaron de [14] y se refiere a que, las causas de muerte son características del estado de salud mental de la sociedad, en diferentes países occidentales. El objetivo es establecer una clasificación de países en función de las tasas de mortalidad, calculadas por 100000 habitantes. Las variables son cuantitativas. Los índices de la tabla 2 sugieren que el árbol hay que partirlo antes del nodo 37 o el 36.

Los datos de las notas francesas han sido utilizados en diversos estudios [10, 15]. Estos datos corresponden a las notas escolares obtenidas por unos estudiantes franceses. El objetivo es hacer una clasificación de estudiantes de acuerdo con su rendimiento en cinco materias, teniendo en cuenta que la escala francesa es de 0 a 20. Los índices de la tabla 2 sugieren que el árbol hay que partirlo antes del nodo 17 o el 16.

En 1708, un eminente miembro de la Academia Francesa de Pintura y Escultura, Roger de Piles, publica la obra *Balance des Peintres*, que entre otras críticas sobre la pintura de su época, contiene para 56 pintores, una serie de juicios desde el punto de vista estético, para ello tomaron cuatro criterios considerados como esenciales, estos son: la composición,

Índices	8 Punt.	PSYSOC	Notas F.	Pintores	Peces Ami.	Sociomat.
Variables	15	36	16	46	45	45
Individuos	15	37	16	47	44	47
Codo	15	22	14	47	45	39
Diday	15	37	17	47	44	47
Jambu R1	15	37	17	46	44	46
Mojena	-	37	-	47	45	47
Millig. Coop.	15	36	17	47	44	46
Lerman	14	35	15	38	43	41

Tabla 2: Resultados numéricos para los índices antes mencionados

el diseño, el color y la expresión. Estos juicios personales de Piles se ubicaron en la escala de 0 a 20. De los 56 pintores, Eduardo Piza [13] seleccionó los 24 pintores más renombrados hoy en día. Además, se han utilizado en diversos estudios del Análisis de Datos [15, 10, 4]. Los índices de la tabla 2 sugieren que el árbol hay que partirlo antes del nodo 47 o el 46.

Los peces de Amiard se pueden ver como un ejemplo pedagógico del Análisis de Datos [4]. Este ejemplo considera un estudio del metabolismo de radio-contaminación, entre los peces. Se desea responder a dos interrogantes: ¿la contaminación de peces está asociada a la duración de contacto con la radioactividad? y ¿la medida de un pez está relacionada con su radio-contaminación? En un experimento, 24 peces se repartieron en tres acuarios radio-contaminados de forma idéntica, pero con diferencias en la duración de contacto. Los peces de 1 al 8 están en el acuario 1, los peces de 9 al 17 están en el acuario 2 y los peces de 18 al 24 están en el acuario 3. El pez número 17 se murió durante el experimento. Se les midieron 16 características cuantitativas. Además, este ejemplo ha sido usado en el análisis de diversos métodos del Análisis de Datos [15, 10]. Los índices de la tabla 2 sugieren que el árbol hay que partirlo antes del nodo 44 o el 45.

A los 24 alumnos de una clase de tercer grado, le atribuyen una nota de 0 a 20 a cada uno de sus compañeros y se registra en la llamada sociomatrix de Thomas. La nota que un alumno atribuye a otro mide la "afinidad" que siente para él. El objetivo del análisis es dominar el balance de las afinidades. Este ejemplo es utilizado por Jambu [4, 10, 15]. Los índices de la tabla 2 sugieren que el árbol hay que partirlo antes del nodo 47.

## 4 Conclusiones

En el análisis de la complejidad efectuado a los índices que determinan el corte del árbol jerárquico binario, se obtiene que los índices de variables y de individuos tienen un tiempo de ejecución bastante aceptable.

Por los resultados obtenidos en la tabla 2, hay indicios que los índices de variables y de individuos se comportan bien por coincidir con los índices más utilizados por los expertos para estos fines, además, de contar con una justificación matemática más explícita. La

importancia de esta coincidencia, es porque las clasificaciones obtenidas tienen importantes interpretaciones, lo que respalda que se trata de buenas decisiones.

## Bibliografía

- [1] Castillo, W. y Arce, C. "Clasificación con índices probabilísticos: una aplicación a las encuestas de opinión pública". VII y VIII Simposio, en Métodos Matemáticos Aplicados a las Ciencias, San José (Costa Rica), febrero 1994.
- [2] Diday, E., Lemaire, J., Pouget, J. et Testu, F. *Éléments d'analyse de données*. Dunod, Paris, 1982.
- [3] Everitt, B. *Cluster analysis*. Edward Arnold, New York, 1993.
- [4] Jambu, M. *Classification automatique pour l'analyse des données*. Dunod, Paris, 1978.
- [5] Jambu, M. *Exploration informatique et statistique des données*. Dunod, Paris, 1989.
- [6] Lerman, I. "Foundations of the likelihood linkage analysis (LLA) classification method". *Applied Stochastic Models and Data Analysis*, 7:63-76, 1991.
- [7] Lerman, I. et Ghazzali, N. "Quoi retenir d'un arbre de classification? un essai en quantification d'image numérisée". IRISA Rennes, Publication Interne No. 568, 1990.
- [8] Milligan, G. and Cooper, M. "An examination of procedures for determining the number of clusters in a data set". *Psychometrika*, 50:159-179, 1985.
- [9] Mojena, R. and Wishart, D. "Stopping rules for Ward's clustering method". Meeting, in Compstat (Toulouse, IASC): Physica-Verlag, Vienna, 1980.
- [10] Murillo, A. "Particionamiento usando búsqueda tabú". Simposio, en Memorias del IV Encuentro Centroamericano de Investigadores en Matemática, Antigua Guatemala, enero 1996.
- [11] Murillo, A. *Proposición de un índice para la interpretación de árboles ascendentes de clasificación basado en conjuntos difusos*. Tesis para optar al grado de Magister Scientiae en Computación, I.T.C.R., Cartago, Costa Rica, 1996.
- [12] Murillo, A. y Trejos, J. "Proposición de índices difusos en la clasificación jerárquica ascendente". *Revista de Matemática: Teoría y Aplicaciones*, 3(2), 1997.
- [13] Piza, E. "La clasificación automática jerárquica aglomerativa". *Revista Ciencias Económicas*, 7(1):95-111, 1987.
- [14] Roux, M. *Algorithmes de classification*. Masson, Paris, 1985.
- [15] Trejos, J. "Propiedades y aplicaciones del número equivalente en análisis de datos". Simposio, en Memorias del IV Encuentro Centroamericano de Investigadores en Matemática, Antigua Guatemala, enero 1996.

VALORES EXTREMOS DE FORMAS CUADRÁTICAS:  
UNA PROPIEDAD ÓPTIMA DE LAS  
COMPONENTES PRINCIPALES

MANUEL MURILLO<sup>1</sup> - MIGUEL ALPÍZAR<sup>2</sup>

---

Este trabajo presenta una pequeña reseña sobre el desarrollo histórico de las componentes principales. Algunos resultados algebraicos de Rao sobre los valores y vectores propios asociados a un par de matrices, maximización de la traza y minimización de la norma, además de los aportes de Darroch, Okamoto y Kanazawa. Se deducen algunas propiedades y se muestra una propiedad óptima para las componentes principales.

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE COSTA RICA, CARTAGO, COSTA RICA

<sup>2</sup>ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

## MÉTODOS NO PARAMÉTRICOS: CONSIDERACIONES SOBRE VENTAJAS Y DESVENTAJAS RESPECTO A LOS MÉTODOS PARAMÉTRICOS

GRACIELA NAVARRO<sup>1</sup>

---

Cuando se realizan pruebas de hipótesis, análisis de regresión y correlación, estimación de parámetros, etc., la mayoría de las técnicas estadísticas que se utilizan en nuestro país son denominadas clásicas o paramétricas, e inclusive la bibliografía en una gran parte versa sobre este tipo de técnicas paramétricas. Sin embargo, en años recientes se ha tomado conciencia de la existencia de otro tipo de procedimientos estadísticos denominados "no paramétricos", los cuales han sido aplicados en otros países en problemas industriales, científicos, de administración y en otras áreas del conocimiento.

El objetivo de esta ponencia es presenar el concepto, los antecedentes históricos de los métodos no paramétricos y de distribución libre, además de ciertos criterios de comparación que validan la utilización de los métodos no paramétricos en la investigación.

Según Bradley [1], "una prueba no paramétrica es una que no plantea hipótesis alguna sobre el valor de un parámetro, y una prueba de distribución libre no plantea supuestos sobre la forma precisa de la población muestreada". Las definiciones no son mutuamente excluyentes, y una prueba puede ser de ambos tipos.

A fin de estar completamente claros sobre lo que se desea decir con distribución libre, es necesario distinguir entre tres tipos de distribuciones: a) la de la población muestreada, b) la de la característica realmente usada por la prueba, y c) la del estadístico de la prueba. La distribución de la cual están "libres" las pruebas corresponde al tipo a), y la "libertad" de que disfrutan es usualmente relativa.

### Bibliografía

- [1] Bradley, J. (1968) *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, NJ.
- [2] Marascuilo, L.; McSweeney, M. (1980) *Non-Parametric and Distribution-Free Methods of Social Sciences*. Brooks Cole Publ. Co., USA.
- [3] Randles, H. R.; Wolfe, D. A. (1984) *Introduction to the Theory of Nonparametrics Statistics*. John Wiley & Sons, New York.

---

<sup>1</sup> DEPARTAMENTO DE MATEMÁTICA, UNIVERSIDAD LATINA DE COSTA RICA, LOURDES DE MONTES DE OCA, COSTA RICA

# SIMULAÇÃO NUMÉRICA DA MODELAGEM DOS POLUENTES AEROSSÓIS, SEGUNDA PARTE: UMA PRIMEIRA ABORDAGEM

SONIA ELENA PALOMINO CASTRO<sup>1</sup>

---

## Abstract

In this paper we intend to model, to numerically approximate and to simulate the concentration of air pollutants in a given domain. The evolution model we use accounts for diffusion and advection phenomena, that is to say the transportation by predominant wind currents is included in the model. Approximation techniques are those of Finite Elements for space variable and Crank Nickolson for time.

All of the results were applied on a particular case that happened in a brasilian city of São Paulo.

## Resumen

En este artículo, tratamos de modelar, aproximar numéricamente y simular la concentración de contaminantes aéreos en un dominio dado. La evolución del modelo que usamos considera fenómenos de difusión, es decir el transporte por medio de corrientes de aire predominantes se incluye en el modelo. Las técnicas de aproximación son las de Elementos Finitos para el espacio de variables y de Crank Nicholson para el tiempo.

Todos los resultados se aplicaron en el caso particular de la ciudad brasileña de São Paulo.

## Introdução

No VI Congresso Internacional de Biomatemática (1993) apresentou-se o trabalho intitulado: "*Modelaje y Simulación de Poluentes Aerosoles en el Aire*", cujo conteúdo foram: o tratamento teórico do modelo matemático via uma equação de evolução e a discretização do problema dos poluentes aerossóis que se dispersam numa região industrializada. Neste trabalho apresentamos os resultados do modelo aplicado a um problema de poluição no

---

<sup>1</sup>UNIVERSIDADE FEDERAL DE SANTA CATARINA, DPTO. DE MATEMÁTICA, CAMPUS UNIVERSITARIO, TRINDADE, CEP 88040 - 900, FLORIANÓPOLIS, SANTA CATARINA - BRASIL; e-mail: palomino@mtm.ufsc.br

ar que aconteceu em Paulínia, uma cidade brasileira localizada a 20 Km. da Universidade Estadual de Campinas, UNICAMP - São Paulo.

Numa pesquisa (vide [5]) a poluição atmosférica, expressão concreta das transformações ambientais ocasionada em geral pelas usinas e centros industriais em regiões como o Sudeste (e no Brasil em geral), aumenta os ataques de asma, provoca enfisema e bronquite e aumenta os casos de pessoas com problemas respiratórios e de ordem cardíaca; isto acontece, por exemplo, nas cidades de Perus, Cubatão, Pólo Petroquímico de Camaçari, Bahia; onde se impõe há muito tempo um quadro de saúde nos seus moradores dos mais deprimentes. Outra pesquisa, [4], da mesma autora analisa as características epidemiológicas que o Município de Paulínia apresenta de forma histórica no período de 1975-1989 constatando que, embora o curto espaço de tempo, as alterações significativas nesse período confirmam os padrões de industrialização de regiões de países onde normalmente os padrões epidemiológicos estão associados a "modernização" e a industrialização. É de se destacar que não é apenas que a saúde da população é afetada, também a Biota toda é submetida a esses riscos de contaminação.

Só conhecendo os valores da concentração de poluentes emitidos pelas fontes poluidoras no ambiente e comparando-os com os padrões permitidos que poderia-se inferir o nível de prejuízo ocasionado. É por tal motivo que o objetivo do trabalho é o de apresentar um modelo para descrever a evolução da concentração de certos poluentes do ar em termos dos fenômenos de difusão e de advecção - além de apresentar alguns métodos de aproximação numérica, servindo assim como ponto de partida para trabalhos posteriores.

Estabeleceremos em primeira instância o modelo matemático dado por uma equação diferencial parcial, esta equação que descreve o fenômeno evolutivo de difusão-advecção será submetida a discretização cujos resultados serão finalmente analisados.

## 1 Abordagem do problema: formulação clássica

Nesta abordagem trata-se numericamente o problema considerando condições de fronteira mistas na região, garantem-se a existência e a unicidade da solução na formulação variacional, [8], e faz-se uma primeira redução do problema ao caso bidimensional, [1].

Consideremos o domínio espacial como um paralelepípedo ( $\Omega$ ) com fronteiras: as faces laterais ( $\partial\Omega_0$ ) onde supomos uma concentração nula de poluentes e as tampas superior e inferior ( $\partial\Omega_1$ ) onde supomos um fluxo nulo, visualizamos esta situação na figura 1.

Seguindo os modelos estabelecidos por [6] e [7], entre outros, podemos formular matematicamente o modelo considerando uma poluição inicial constante  $u_0$  no instante  $t = 0$ .

A equação junto às condições iniciais e de fronteira definem a seguinte formulação clássica:

$$\begin{cases} \frac{\partial u}{\partial t} - \operatorname{div}(\alpha \nabla u) + \operatorname{div}(Vu) + \sigma u = f, \\ (x, t) \in \Omega \times [0, T], \text{ com} \\ u(x, 0) = u_0 \text{ em } \Omega, \\ u(x, t) = 0 \text{ em } \partial\Omega_0, \text{ e} \\ \frac{\partial u}{\partial n}(x, t) = 0 \text{ em } \partial\Omega_1. \end{cases} \quad (1)$$



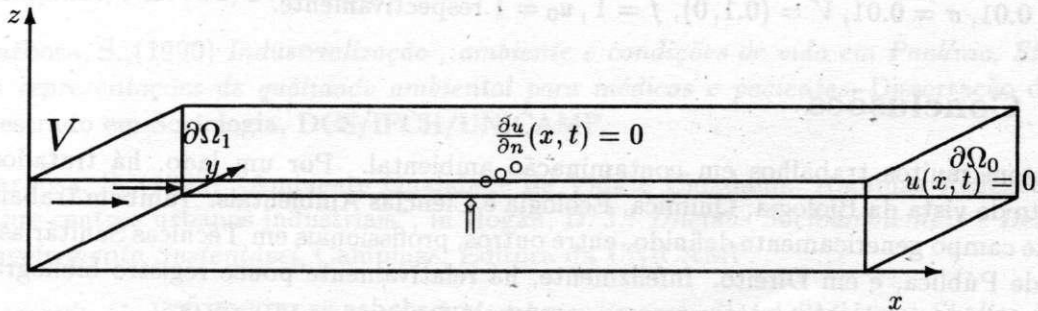


Figura 1: Domínio do Problema numa Primeira Abordagem

Onde  $u = u(x, t)$  é a concentração dos poluentes em  $\Omega \times [0, T]$ , assim  $\frac{\partial u}{\partial t}$  é a taxa de variação dessa concentração respeito da variável tempo.

Os outros termos são descritos por:

- $div(\alpha \nabla u)$ : este termo representa a difusão propriamente dita, onde  $\alpha = \alpha(u, x, t)$  é o coeficiente de difusão, que em geral pode depender do próprio poluente, da posição e inclusive do tempo;
- $div(Vu)$ : representa o transporte do poluente seguindo a direção dada pelo vetor velocidade  $V = (V_1, V_2, V_3)$  das correntes de vento predominantes;
- $\sigma u$ : com  $\sigma$  como coeficiente de decaimento, este termo descreve a parte do poluente que se decompõe interagindo com o meio ambiente e assume outras formas;
- $f = f(x, u, t)$ : identifica a fonte alimentadora ou poluidora localizada em  $x$  que em geral pode depender do tempo e do poluente.

## 2 Experiências computacionais e simulações

Nesta abordagem se utilizou um código computacional em FORTRAN 77, todos os cálculos se realizaram inicialmente em microcomputadores IBM PC AT 486 e pelas suas limitações e para obter melhores resultados passamos a usar as estações de trabalho em sistema UNIX. Como foi mencionado na introdução deste trabalho, o passo do modelo contínuo ao modelo discreto foi descrito em [1]. Nesta fase de discretização as funções  $\{\phi_i(x, y)\}$  representam as funções de base do espaço de discretização espacial dos elementos finitos.

As figuras exibidas mostram a evolução da concentração dos poluentes em 10 e 50 passos no tempo ( $NPT = 10$  e  $NPT = 50$  respectivamente) e com o intuito de descrever a dispersão dos poluentes pela influência do vento numa determinada região é que os resultados podem ser comparados com aqueles onde a influência do vento é nula. Para

as simulações teóricas foram usados dois conjuntos diferentes de parâmetros <sup>2</sup> indicados nas figuras, na primeira  $\alpha = 0.005$ ,  $\sigma = 0.01$ ,  $V = (0.1, 0)$ ,  $f = 1$ ,  $u_0 = 1$  e na última  $\alpha = 0.01$ ,  $\sigma = 0.01$ ,  $V = (0.1, 0)$ ,  $f = 1$ ,  $u_0 = 1$  respectivamente. <sup>3</sup>

### 3 Conclusões

Existem muitos trabalhos em contaminação ambiental. Por um lado, há tratados do ponto de vista da Biologia, Química, Ecologia e Ciências Ambientais. Também trabalham neste campo genericamente definido, entre outros, profissionais em Técnicas Sanitárias, em Saúde Pública, e em Direito. Infelizmente, há relativamente pouco registro bibliográfico de participação de Matemáticos neste campo - guardadas as proporções.

É assim que na bibliografia disponível praticamente não têm sido encontrados trabalhos que enfoquem o problema de evolução da poluição atmosférica usando a linguagem das Equações Diferenciais Parciais e ainda menos na obtenção de soluções aproximadas usando seja técnicas de Diferenças Finitas ou ainda de Elementos Finitos (FEM). O Modelo da Pluma de Gauss, como método analítico, apenas ilustra o problema em certos casos estacionários.

Nas soluções aproximadas do problema se observa uma concentração maior na zona de incidência da fonte poluidora, onde a presença do termo advectivo nesta abordagem é notável segundo a ilustração dos gráficos pelo deslocamento desses poluentes no decorrer do tempo. Isto significa que pode dar-se o caso que em cidades ou bairros vizinhos o prejuízo seja bem maior do que na região onde a fonte está localizada, dependendo, obviamente, da magnitude e direção do vento e da quantidade e tipo de poluente emitido. Na primeira abordagem, em que o problema foi reduzido apenas às variáveis horizontais, optamos pela utilização de um código numérico com Elementos Finitos de 2ª ordem (FEM/2ª). A possibilidade de recorrer à rede SUN em ambiente UNIX foi importante nesta escolha.

Além do desenvolvimento de melhores técnicas numéricas, trabalhos futuros também irão fazer uso de contatos com o NEPAM (Núcleo de Estudos de Pesquisas Ambientais) para o uso de parâmetros em estudos reais - visando efetiva contribuição em situações de estudos, de avaliação e de decisão sobre centros poluidores para o justo e imprescindível equilíbrio, garantindo uma melhor qualidade de vida e repetindo a natureza no real interesse e bem estar tanto da coletividade presente, como das gerações futuras.

### Bibliografia

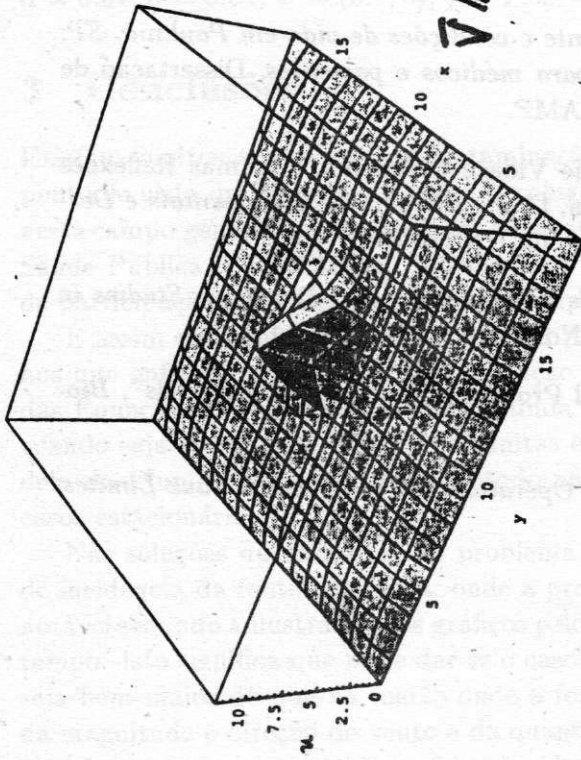
- [1] Palomino Cástro, S.; Meyer, João F. C. (1993) "Modelaje y simulación de poluentes aerosoles en el aire", *VI Congreso Internacional de Biomatemática*, São José, Costa Rica.
- [2] Palomino Castro, S. E. (1993) *Modelagem Matemática e Aproximação Numérica do Estudo de Poluentes no Ar*, Tese de Mestrado, UNICAMP, Campinas, S.P.

<sup>2</sup> conservando uma coerência nas unidades físicas apresentamos apenas as grandezas desses parâmetros

<sup>3</sup> sem perda de generalidade foi considerado que o vento aponta no sentido positivo do eixo X

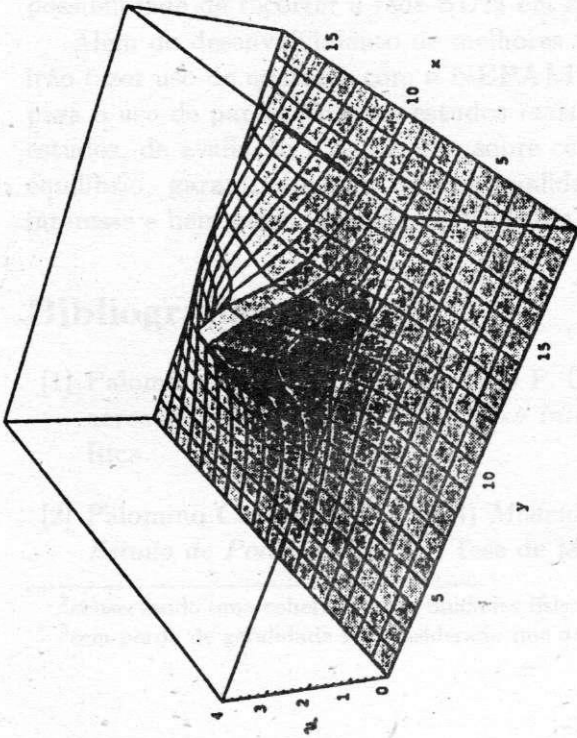
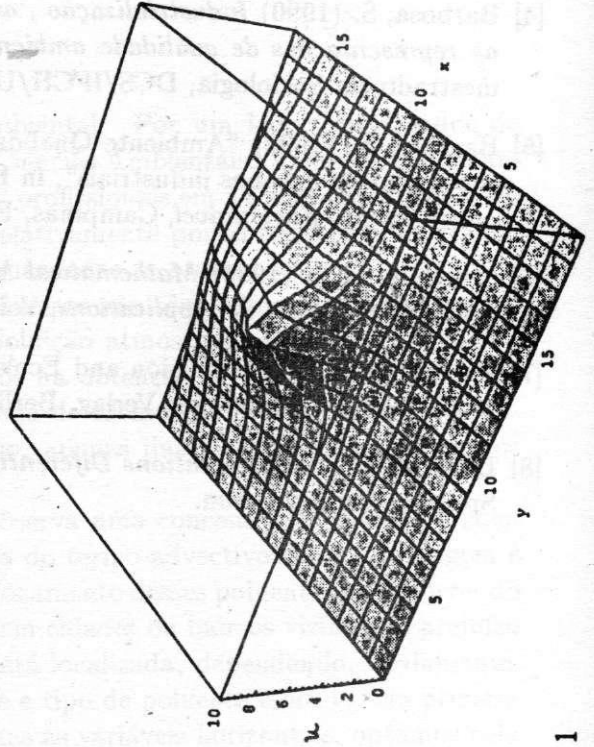
- [3] Sevá, O., Ferreira, L. C. (1989) *O Projeto da Termoelétrica em Paulínia, SP. A questão energética e a depredação Sócio Ambiental*, Coletânea de Ensaio e Estudos, Campinas, NEPAM/UNICAMP.
- [4] Barbosa, S. (1990) *Industrialização, ambiente e condições de vida em Paulínia, SP: as representações de qualidade ambiental para médicos e pacientes*, Dissertação de mestrado em Sociologia, DCS/IFCH/UNICAMP.
- [5] Barbosa, S. (1992) "Ambiente Qualidade de Vida e Cidadania. Algumas Reflexões sobre centros urbanos industriais", in Hogan, D. J.: *Dilemas Sócioambientais e Desenvolvimento Sustentável*, Campinas, Editora da UNICAMP.
- [6] Marchuk, G. I. (1986) *Mathematical Models in Environmental Problems, Studies in Mathematical and its Applications*, Vol 16, North Holland, Amsterdam.
- [7] Okubo, A. (1980) "Diffusion and Ecological Problems: Mathematical Models", *Bi-mathematics* 10, Springer-Verlag, Berlin.
- [8] Lions, J. L. (1961) *Equations Différentielles Opérationnelles et Problèmes aux Limites*, Springer-Verlag, Berlin.

(SUREDat6: NPT=10, V<sub>1</sub>= 0.1)



$\bar{V} \parallel (1100)$

(SUREDat6: NPT=50, V<sub>1</sub>= 0.1)



(SUREDat9: NPT=50, V<sub>1</sub>= 0.1)

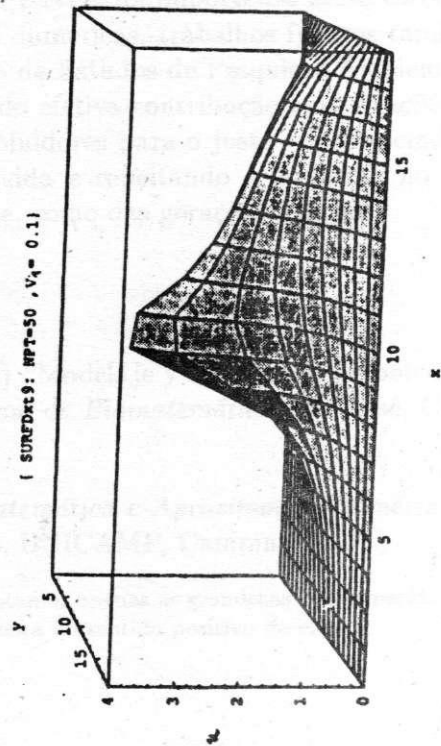


Figura 1.  $\alpha = 0.01, \sigma = 0.01, V_1 = 0.1, V_2 = 0, f = 1 e u_0 = 1$

Figura 2.  $\alpha = 0.005, \sigma = 0.01, V_1 = 0.1, V_2 = 0, f = 1 e u_0 = 1$

# SIMULAÇÃO NUMÉRICA DA MODELAGEM DOS POLUENTES AEROSSÓIS, SEGUNDA PARTE: UMA PRIMEIRA ABORDAGEM

SONIA ELENA PALOMINO CASTRO<sup>1</sup>

---

## Abstract

In this paper we intend to model, to numerically approximate and to simulate the concentration of air pollutants in a given domain. The evolution model we use accounts for diffusion and advection phenomena, that is to say the transportation by predominant wind currents is included in the model. Approximation techniques are those of Finite Elements for space variable and Crank Nickolson for time.

All of the results were applied on a particular case that happened in a brasilian city of São Paulo.

## Resumen

En este artículo, tratamos de modelar, aproximar numéricamente y simular la concentración de contaminantes aéreos en un dominio dado. La evolución del modelo que usamos considera fenómenos de difusión, es decir el transporte por medio de corrientes de aire predominantes se incluye en el modelo. Las técnicas de aproximación son las de Elementos Finitos para el espacio de variables y de Crank Nicholson para el tiempo.

Todos los resultados se aplicaron en el caso particular de la ciudad brasileña de São Paulo.

## Introdução

No VI Congresso Internacional de Biomatemática (1993) apresentou-se o trabalho intitulado: "*Modelaje y Simulación de Poluentes Aerosoles en el Aire*", cujo conteúdo foram: o tratamento teórico do modelo matemático via uma equação de evolução e a discretização do problema dos poluentes aerossóis que se dispersam numa região industrializada. Neste trabalho apresentamos os resultados do modelo aplicado a um problema de poluição no

---

<sup>1</sup>UNIVERSIDADE FEDERAL DE SANTA CATARINA, DPTO. DE MATEMÁTICA, CAMPUS UNIVERSITARIO, TRINDADE, CEP 88040 - 900, FLORIANÓPOLIS, SANTA CATARINA - BRASIL; e-mail: palomino@mtm.ufsc.br

ar que aconteceu em Paulínia, uma cidade brasileira localizada a 20 Km. da Universidade Estadual de Campinas, UNICAMP - São Paulo.

Numa pesquisa (vide [5]) a poluição atmosférica, expressão concreta das transformações ambientais ocasionada em geral pelas usinas e centros industriais em regiões como o Sudeste (e no Brasil em geral), aumenta os ataques de asma, provoca enfisema e bronquite e aumenta os casos de pessoas com problemas respiratórios e de ordem cardíaca; isto acontece, por exemplo, nas cidades de Perus, Cubatão, Pólo Petroquímico de Camaçari, Bahia; onde se impõe há muito tempo um quadro de saúde nos seus moradores dos mais deprimentes. Outra pesquisa, [4], da mesma autora analisa as características epidemiológicas que o Município de Paulínia apresenta de forma histórica no período de 1975-1989 constatando que, embora o curto espaço de tempo, as alterações significativas nesse período confirmam os padrões de industrialização de regiões de países onde normalmente os padrões epidemiológicos estão associados a "modernização" e a industrialização. É de se destacar que não é apenas que a saúde da população é afetada, também a Biota toda é submetida a esses riscos de contaminação.

Só conhecendo os valores da concentração de poluentes emitidos pelas fontes poluidoras no ambiente e comparando-os com os padrões permitidos que poderia-se inferir o nível de prejuízo ocasionado. É por tal motivo que o objetivo do trabalho é o de apresentar um modelo para descrever a evolução da concentração de certos poluentes do ar em termos dos fenômenos de difusão e de advecção - além de apresentar alguns métodos de aproximação numérica, servindo assim como ponto de partida para trabalhos posteriores.

Estabeleceremos em primeira instância o modelo matemático dado por uma equação diferencial parcial, esta equação que descreve o fenômeno evolutivo de difusão-advecção será submetida a discretização cujos resultados serão finalmente analisados.

## 1 Abordagem do problema: formulação clássica

Nesta abordagem trata-se numericamente o problema considerando condições de fronteira mistas na região, garantem-se a existência e a unicidade da solução na formulação variacional, [8], e faz-se uma primeira redução do problema ao caso bidimensional, [1].

Consideremos o domínio espacial como um paralelepípedo ( $\Omega$ ) com fronteiras: as faces laterais ( $\partial\Omega_0$ ) onde supomos uma concentração nula de poluentes e as tampas superior e inferior ( $\partial\Omega_1$ ) onde supomos um fluxo nulo, visualizamos esta situação na figura 1.

Seguindo os modelos estabelecidos por [6] e [7], entre outros, podemos formular matematicamente o modelo considerando uma poluição inicial constante  $u_0$  no instante  $t = 0$ .

A equação junto às condições iniciais e de fronteira definem a seguinte formulação clássica:

$$\begin{cases} \frac{\partial u}{\partial t} - \operatorname{div}(\alpha \nabla u) + \operatorname{div}(Vu) + \sigma u = f, \\ (x, t) \in \Omega \times [0, T], \text{ com} \\ u(x, 0) = u_0 \text{ em } \Omega, \\ u(x, t) = 0 \text{ em } \partial\Omega_0, \text{ e} \\ \frac{\partial u}{\partial n}(x, t) = 0 \text{ em } \partial\Omega_1. \end{cases} \quad (1)$$

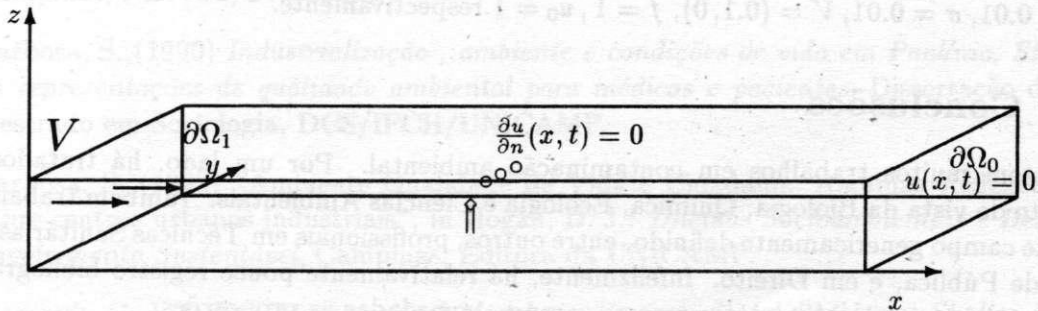


Figura 1: Domínio do Problema numa Primeira Abordagem

Onde  $u = u(x, t)$  é a concentração dos poluentes em  $\Omega \times [0, T]$ , assim  $\frac{\partial u}{\partial t}$  é a taxa de variação dessa concentração respeito da variável tempo.

Os outros termos são descritos por:

- $div(\alpha \nabla u)$ : este termo representa a difusão propriamente dita, onde  $\alpha = \alpha(u, x, t)$  é o coeficiente de difusão, que em geral pode depender do próprio poluente, da posição e inclusive do tempo;
- $div(Vu)$ : representa o transporte do poluente seguindo a direção dada pelo vetor velocidade  $V = (V_1, V_2, V_3)$  das correntes de vento predominantes;
- $\sigma u$ : com  $\sigma$  como coeficiente de decaimento, este termo descreve a parte do poluente que se decompõe interagindo com o meio ambiente e assume outras formas;
- $f = f(x, u, t)$ : identifica a fonte alimentadora ou poluidora localizada em  $x$  que em geral pode depender do tempo e do poluente.

## 2 Experiências computacionais e simulações

Nesta abordagem se utilizou um código computacional em FORTRAN 77, todos os cálculos se realizaram inicialmente em microcomputadores IBM PC AT 486 e pelas suas limitações e para obter melhores resultados passamos a usar as estações de trabalho em sistema UNIX. Como foi mencionado na introdução deste trabalho, o passo do modelo contínuo ao modelo discreto foi descrito em [1]. Nesta fase de discretização as funções  $\{\phi_i(x, y)\}$  representam as funções de base do espaço de discretização espacial dos elementos finitos.

As figuras exibidas mostram a evolução da concentração dos poluentes em 10 e 50 passos no tempo ( $NPT = 10$  e  $NPT = 50$  respectivamente) e com o intuito de descrever a dispersão dos poluentes pela influência do vento numa determinada região é que os resultados podem ser comparados com aqueles onde a influência do vento é nula. Para

as simulações teóricas foram usados dois conjuntos diferentes de parâmetros <sup>2</sup> indicados nas figuras, na primeira  $\alpha = 0.005$ ,  $\sigma = 0.01$ ,  $V = (0.1, 0)$ ,  $f = 1$ ,  $u_0 = 1$  e na última  $\alpha = 0.01$ ,  $\sigma = 0.01$ ,  $V = (0.1, 0)$ ,  $f = 1$ ,  $u_0 = 1$  respectivamente. <sup>3</sup>

### 3 Conclusões

Existem muitos trabalhos em contaminação ambiental. Por um lado, há tratados do ponto de vista da Biologia, Química, Ecologia e Ciências Ambientais. Também trabalham neste campo genericamente definido, entre outros, profissionais em Técnicas Sanitárias, em Saúde Pública, e em Direito. Infelizmente, há relativamente pouco registro bibliográfico de participação de Matemáticos neste campo - guardadas as proporções.

É assim que na bibliografia disponível praticamente não têm sido encontrados trabalhos que enfoquem o problema de evolução da poluição atmosférica usando a linguagem das Equações Diferenciais Parciais e ainda menos na obtenção de soluções aproximadas usando seja técnicas de Diferenças Finitas ou ainda de Elementos Finitos (FEM). O Modelo da Pluma de Gauss, como método analítico, apenas ilustra o problema em certos casos estacionários.

Nas soluções aproximadas do problema se observa uma concentração maior na zona de incidência da fonte poluidora, onde a presença do termo advectivo nesta abordagem é notável segundo a ilustração dos gráficos pelo deslocamento desses poluentes no decorrer do tempo. Isto significa que pode dar-se o caso que em cidades ou bairros vizinhos o prejuízo seja bem maior do que na região onde a fonte está localizada, dependendo, obviamente, da magnitude e direção do vento e da quantidade e tipo de poluente emitido. Na primeira abordagem, em que o problema foi reduzido apenas às variáveis horizontais, optamos pela utilização de um código numérico com Elementos Finitos de 2ª ordem (FEM/2ª). A possibilidade de recorrer à rede SUN em ambiente UNIX foi importante nesta escolha.

Além do desenvolvimento de melhores técnicas numéricas, trabalhos futuros também irão fazer uso de contatos com o NEPAM (Núcleo de Estudos de Pesquisas Ambientais) para o uso de parâmetros em estudos reais - visando efetiva contribuição em situações de estudos, de avaliação e de decisão sobre centros poluidores para o justo e imprescindível equilíbrio, garantindo uma melhor qualidade de vida e repetindo a natureza no real interesse e bem estar tanto da coletividade presente, como das gerações futuras.

### Bibliografia

- [1] Palomino Cástro, S.; Meyer, João F. C. (1993) "Modelaje y simulación de poluentes aerosoles en el aire", *VI Congreso Internacional de Biomatemática*, São José, Costa Rica.
- [2] Palomino Castro, S. E. (1993) *Modelagem Matemática e Aproximação Numérica do Estudo de Poluentes no Ar*, Tese de Mestrado, UNICAMP, Campinas, S.P.

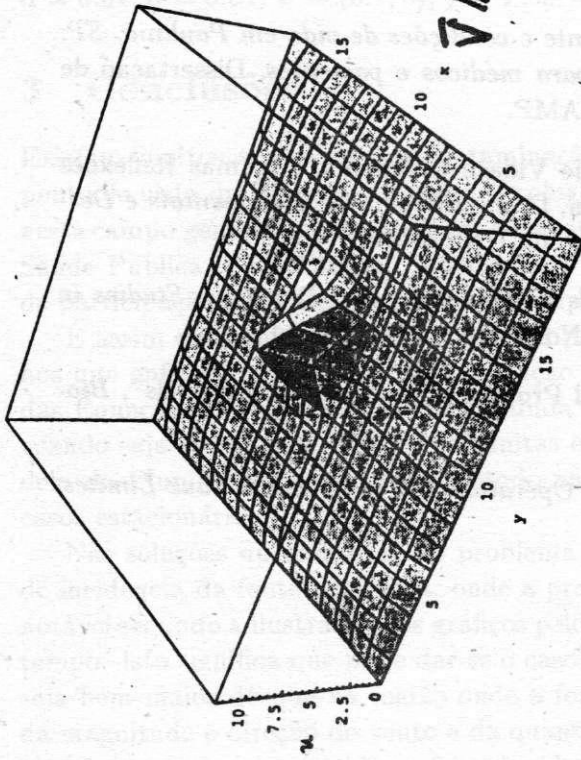
<sup>2</sup> conservando uma coerência nas unidades físicas apresentamos apenas as grandezas desses parâmetros

<sup>3</sup> sem perda de generalidade foi considerado que o vento aponta no sentido positivo do eixo X

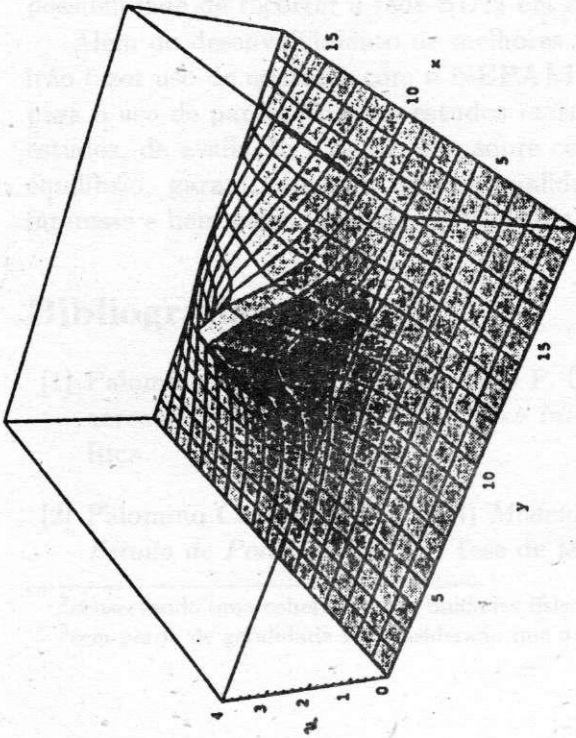
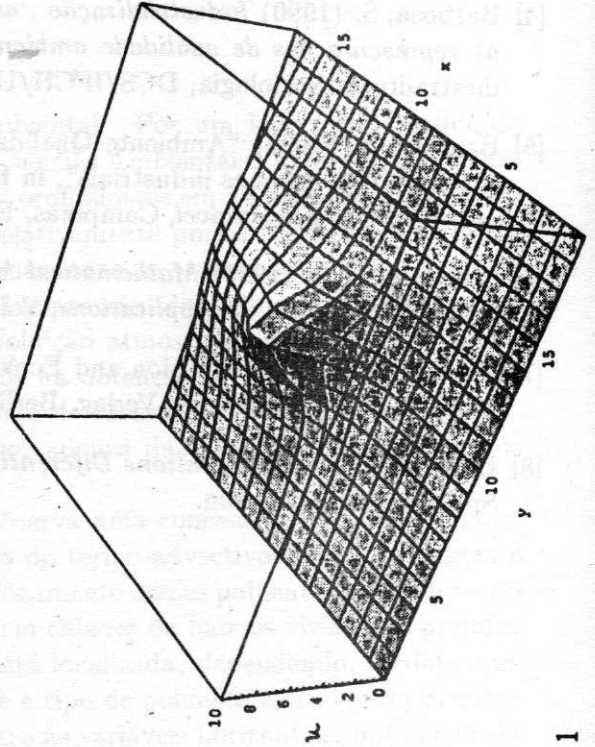


- [3] Sevá, O., Ferreira, L. C. (1989) *O Projeto da Termoelétrica em Paulínia, SP. A questão energética e a depredação Sócio Ambiental*, Coletânea de Ensaios e Estudos, Campinas, NEPAM/UNICAMP.
- [4] Barbosa, S. (1990) *Industrialização, ambiente e condições de vida em Paulínia, SP: as representações de qualidade ambiental para médicos e pacientes*, Dissertação de mestrado em Sociologia, DCS/IFCH/UNICAMP.
- [5] Barbosa, S. (1992) "Ambiente Qualidade de Vida e Cidadania. Algumas Reflexões sobre centros urbanos industriais", in Hogan, D. J.: *Dilemas Sócioambientais e Desenvolvimento Sustentável*, Campinas, Editora da UNICAMP.
- [6] Marchuk, G. I. (1986) *Mathematical Models in Environmental Problems, Studies in Mathematical and its Applications*, Vol 16, North Holland, Amsterdam.
- [7] Okubo, A. (1980) "Diffusion and Ecological Problems: Mathematical Models", *Bi-mathematics* 10, Springer-Verlag, Berlin.
- [8] Lions, J. L. (1961) *Equations Différentielles Opérationnelles et Problèmes aux Limites*, Springer-Verlag, Berlin.

(SUREDat6: NPT=10, V<sub>1</sub>= 0.1)



(SUREDat6: NPT=50, V<sub>1</sub>= 0.1)



(SUREDat9: NPT=50, V<sub>1</sub>= 0.1)

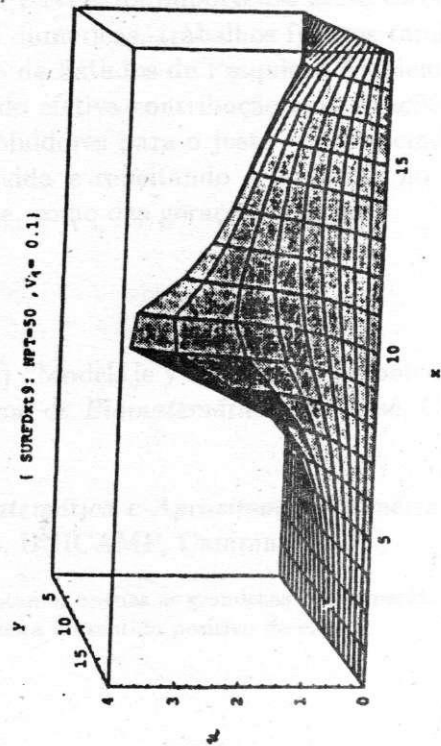


Figura 1.  $\alpha = 0.01, \sigma = 0.01, V_1 = 0.1, V_2 = 0, f = 1 e u_0 = 1$

Figura 2.  $\alpha = 0.005, \sigma = 0.01, V_1 = 0.1, V_2 = 0, f = 1 e u_0 = 1$

## OPERACIONALIZACIÓN DEL CONCEPTO DE DESARROLLO SOSTENIBLE A TRAVÉS DE MODELOS BIOECONÓMICOS

JOSÉ F. PASTRANA<sup>1</sup>

---

En este artículo se presenta primero el concepto de desarrollo sostenible de acuerdo con la FAO, para luego reelaborarlo en términos de un modelo bioeconómico, obteniendo finalmente una formulación operacional de ese concepto.

En él se presentan también los conceptos de sistema, modelo, caos, componentes económico y biológico y se discute la inhabilidad del sistema económico de libre empresa, en su forma corriente, para procurar el desarrollo sostenible. Además, se hace referencia a varios aspectos matemáticos relacionados con la solución del sistema de ecuaciones diferenciales del modelo bioeconómico que se ofrece como base para el desarrollo del concepto.

**Palabras clave:** sistema, modelo, caos, desarrollo sostenible.

---

<sup>1</sup>ESCUELA DE ESTADÍSTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

# REGRESIÓN LINEAL Y ANÁLISIS DE LA VARIANZA USANDO EL PAQUETE ESTADÍSTICO LINEAL VERSIÓN 2.3

BLANCA ROSA PÉREZ SALVADOR<sup>1</sup> - SERGIO G. DE LOS COBOS SILVA<sup>1</sup>  
MIGUEL ANGEL GUTIÉRREZ ANDRADE<sup>2</sup>

---

## Resumen

En el minicurso de regresión lineal y análisis de varianza, usando el paquete de estadísticas *Lineal*, se pretende dar una introducción teórica a estas dos importantes técnicas estadísticas, establecer los supuestos en los que se sustentan, indicar el tipo de datos a los que se les puede aplicar y posteriormente, realizar algunos ejercicios prácticos utilizando el paquete de estadísticas ya mencionado, que es una herramienta sencilla y eficaz para realizar los cálculos del análisis de regresión y análisis de varianza.

## 1 Introducción

El modelo de regresión lineal considera que existe una relación de asociación entre una variable respuesta  $Y$  y una o más variables explicativas  $X_i$ ;  $i = 1, 2, \dots, m$ , dada por el modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

donde  $\varepsilon$  es el error de observación, una variable aleatoria con media cero y varianza  $\sigma^2$ . Se supone que los errores son independientes y no correlacionados con las variables explicativas.

Las inferencias que se hacen al modelo de regresión lineal son dos:

1. Estimación de los parámetros del modelo  $\beta_i$ ,  $i = 0, 1, \dots, m$  y  $\sigma^2$  y
2. Prueba de hipótesis sobre los mismos parámetros.

---

<sup>1</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA, UNIDAD IZTAPALAPA, DEPTO. DE MATEMÁTICAS, AV. MICHOACÁN Y CALLE LA PURÍSIMA, COL. VICENTINA, CP 09340 MÉXICO D. F., MÉXICO

<sup>2</sup>UNIVERSIDAD AUTÓNOMA METROPOLITANA, UNIDAD ATZCAPOTZALCO, DEPTO DE SISTEMAS, AV. SAN PABLO 180, COL. REINOSA TAMAHULIPAS, CP 02200 MÉXICO D. F., MÉXICO

Por ejemplo, se puede estimar una relación funcional entre el ingreso de una persona y su nivel de escolaridad.

El modelo de análisis de varianza, supone la aplicación de dos o más tratamientos a un conjunto de muestras independientes de la población. Los tratamientos son  $T_1, T_2, \dots, T_k$ , al aplicarse los tratamientos se tiene interés en observar la variable respuesta  $Y$ , se supone que la media de la variable respuesta al aplicarse el tratamiento  $T_i$  es  $\mu_i$ . La inferencia que se hace con el análisis de varianza es la prueba de hipótesis de diferencia de medias de los tratamientos.

Por ejemplo, considere que se está experimentando en una estación agrícola, y se aplica una diferente concentración de fertilizantes a  $k$  parcelas sembradas de maíz, se quiere medir el rendimiento del cultivo por hectárea.  $Y_{ij}$  es el rendimiento de la hectárea  $i$  que se le aplicó la concentración  $j$  de fertilizante. El análisis de varianza se utiliza, en este caso, para determinar si las distintas concentraciones de fertilizante dan un rendimiento por hectárea semejante.

El paquete de estadísticas *Lineal* fue desarrollado para ser usado en computadoras personales de recursos mínimos, pero con resultados precisos, fácil manejo y rapidez de cálculos. Efectúa los cálculos necesarios para una regresión lineal simple, la regresión lineal múltiple y el análisis de varianza.

En el minicurso se pretende enseñar el cuadro teórico de las dos técnicas estadísticas y la interpretación de los resultados obtenidos a través del paquete.

## 2 Presentación de *Lineal*

El Paquete estadístico *Lineal*, permite tener los cálculos del **análisis de regresión lineal**, simple y múltiple, y el **análisis de varianza** con uno y dos criterios de clasificación.

Los resultados producidos por el paquete son precisos y pueden ser comparados con los que generan programas de mucha capacidad y tradición como el SPSS o el SAS.

### 2.1 Capacidad de lectura y escritura

El paquete toma los datos de un archivo en disco, con extensión ".dat"; y deposita los resultados en un archivo en disco con el mismo nombre y extensión ".sal", o bien, los imprime en papel a través de una impresora, de matriz de puntos o laser, conectada a la computadora. Al mismo tiempo, los resultados van apareciendo en la pantalla.

Si los resultados se escribieron en un archivo en disco y se quiere tenerlos impresos en papel, puede hacerlo utilizando la siguiente instrucción desde el drive en donde esta el archivo con los resultados:

```
copy nombre-del-archivo.sal prn
```

### 2.2 Requerimientos del sistema y limitaciones del programa

**Requerimientos mínimos.** El programa puede ser ejecutado eficientemente en una maquina 286 o superior, con al menos un drive disponible (inclusive de  $5\frac{1}{4}$ , doble densidad). El programa aumenta considerablemente la velocidad si se corre desde el disco duro.

El análisis de regresión puede hacerse hasta con 10 variables de 250 observaciones cada una. El análisis de varianza puede hacerse hasta con 250 datos, clasificados con uno o con dos criterios, sin considerar efecto de iteración.

El programa, de esta manera, fue desarrollado para **ser usado por estudiantes**, que por lo general **tienen acceso a computadoras con poca capacidad y sin disco duro**.

### 3 Contenido

La primera pantalla de *Lineal*, contiene una presentación y un menú con las tres opciones siguientes:

- **INFORMACIÓN**; que al seleccionarse despliega una pantalla de información acerca de lo que *Lineal* es.
- **ESTADÍSTICAS LINEALES**; que al seleccionarse lo lleva propiamente al programa *Lineal*; él cual se inicia presentando una pantalla con el submenú que se muestra en la figura 1.
- **SALIDA** que al seleccionarse le permite salir del programa *Lineal*.

**Figura 1** Pantalla de la opción **ESTADÍSTICAS LINEALES**.

arcHivo	Datos	Calcula	Regresión	Anova	Salir
[F1] Ayudas	Elije una opción			[ALT-X] Salir	

#### 3.1 Contenido de ESTADÍSTICAS LINEALES

Este es un cuadro sinóptico del contenido de los menús que conforman la opción ESTADÍSTICAS LINEALES.

- ESTADÍSTICAS LINEALES.

– arcHivo

- \* Crear un archivo
- \* Elegir un archivo

- \* Salir
- Datos
  - \* Listar datos
  - \* Modificar una observación
  - \* Eliminar una observación
  - \* Aumentar datos
  - \* A o M una Variable
  - \* Salir
- Calcula
  - \* Producto
  - \* Cociente
  - \* Logaritmo Natural
  - \* Logaritmo base 10
  - \* Raíz cuadrada
- Regresión
  - \* Estadísticas
  - \* Gráfica
- Anova
  - \* Estadísticas

Se pretende que los estudiantes se familiaricen con estas opciones y que puedan interpretar los resultados.

## ANÁLISIS DE REGRESIÓN DE GINI

BLANCA ROSA PÉREZ<sup>1</sup> - SERGIO DE LOS COBOS<sup>1</sup> - MIGUEL ANGEL GUTIÉRREZ<sup>2</sup>

**Palabras clave:** diferencia media, programación lineal, estimación robusta.

El análisis de regresión de Gini, es una alternativa al método de mínimos cuadrados, para estimar los parámetros del modelo de regresión lineal. Los estimadores obtenidos con la regresión de Gini son estimadores robustos, por lo que fallas en los supuestos de linealidad, no correlacionalidad, independencia y normalidad no les afecta de manera considerable. En el trabajo se presentan los estimadores por el método de Gini, y algunas maneras de probar los supuestos del método de mínimos cuadrados.

Se define la desviación media de Gini de una v.a.  $X$  como  $MD(X) = E(|X - M|)$ , donde  $M$  es la mediana de  $X$ . Se pueden estimar dos rectas utilizando la desviación media de Gini: (a) la recta que minimiza la expresión  $f(\alpha, \beta) = \sum |\varepsilon_i| = \sum_{i=1}^n |y_i - \alpha - \beta x_i|$ ; y (b) la que minimiza  $g(\beta) = \sum_{1 \leq i < j \leq n} |\varepsilon_i - \varepsilon_j| = \sum_{1 \leq i < j \leq n} |y_i - y_j - \beta(x_i - x_j)|$ .

Para minimizar estas dos funciones se tienen las siguientes propiedades: [P1] Las dos funciones  $f(\alpha, \beta)$  y  $g(\beta)$ , son convexas. [P2] El problema de minimizar  $g(\beta) = \sum_{i < j} |\varepsilon_i - \varepsilon_j| = \sum |y_i - y_j - \beta(x_i - x_j)|$  es equivalente al problema de programación lineal: minimizar  $H(\beta, \gamma) = \gamma$  sujeto a las  $2^{n(n-1)/2}$  restricciones  $\sum_{i < j} \pm(\varepsilon_i - \varepsilon_j) \leq \gamma$ . [P3] El problema de minimizar  $f(\alpha, \beta) = \sum |\varepsilon_i| = \sum |y_i - \alpha - \beta x_i|$  es equivalente al problema de programación lineal: minimizar  $H(\alpha, \beta, \gamma) = \gamma$  sujeto a las  $2^n$  restricciones  $\sum_{i < j} \pm \varepsilon_i \leq \gamma$ . [P4] Los vértices de  $g(\beta)$  coinciden con la solución de las ecuaciones  $\varepsilon_i - \varepsilon_j = 0$ , para todo  $i \neq j$ . Esto implica que  $\varepsilon_i - \varepsilon_j = y_i - y_j - \beta(x_i - x_j) = 0$  y que  $\beta_{ij} = \frac{y_i - y_j}{x_i - x_j}$  los vértices de la gráfica de la función coinciden con la pendiente entre dos puntos muestrales. [P5] Los vértices de  $f(\alpha, \beta)$  coinciden con la solución de dos sistemas de ecuaciones, cuya solución es:  $\beta_{ij} = \frac{y_i - y_j}{x_i - x_j} \alpha_{ij} = y_i - \beta x_i = y_j - \beta x_j$ .

Si la solución encontrada por minimizar la función  $f$  ó  $g$  es única e igual a  $\hat{\beta} = \frac{y_{i_0} - y_{j_0}}{x_{i_0} - x_{j_0}}$ , se puede ver como un promedio ponderado  $\sum m_{ij} w_{ij}$  con  $w_{i_0, j_0} = 1$  y  $w_{i, j} = 0$ ,  $\forall i, j \neq i_0, j_0$ . Si la solución no es única, (son dos o más vertices los que minimizan la función objetivo,  $V$  es el conjunto de estos vértices),  $\hat{\beta}$  se toma como el promedio ponderado  $\hat{\beta} = \sum m_{ij} w_{ij}$  con  $w_{ij} = \frac{|x_i - x_j|}{\sum |x_k - x_l|}$  si  $(x_i, x_j)$  y  $(x_k, x_l)$  se relacionan con los elementos de  $V$ , y  $w_{ij} = 0$  en otro caso. De este modo, sólo cuentan las pendientes que minimizan la desviación media de Gini.

La desviación media de Gini también puede usarse para el caso de regresión lineal múltiple.

### Referencias

- [1] Olkin I. and Yitzhaki S. (1992) "Gini regression analysis", *Int. Stat. Rev.* **60**(2): 185-192.
- [2] Pérez, B. R.; de los Cobos, S.; Gutiérrez, M. A. (1995) "Análisis de regresión de Gini", X Foro Nacional de Estadística y II Congreso Iberoamericano de Estadística, México, pp. 182-187.

<sup>1</sup> UNIVERSIDAD AUTÓNOMA METROPOLITANA, UNIDAD IZTAPALAPA, DEPTO. DE MATEMÁTICAS, AV. MICHOACÁN Y CALLE LA PURÍSIMA, COL. VICENTINA, CP 09340 MÉXICO D. F., MÉXICO

<sup>2</sup> U. A. M.- ATZCAPOTZALCO, DEPTO DE SISTEMAS, MÉXICO D. F. MÉXICO



## ANÁLISIS DE ESTRUCTURAS DE OPINIÓN PÚBLICA<sup>1</sup>

JORGE POLTRONIERI<sup>2</sup>

Este trabajo se presentan los resultados de los análisis hechos a través de nueve años de investigación de las estructuras de la opinión pública en Costa Rica. La interpretación y los comentarios se hacen desde una doble perspectiva:

- una perspectiva microscópica del sico-sociólogo, en que basa en los mecanismos que permiten el ajuste entre el individuo y su medio ambiente, de cómo se construyen las representaciones y de cómo se pasa de las representaciones sociales a las representaciones individuales y recíprocamente. Aquí se procede principalmente por diferencias insistiendo en la diversidad de opiniones en los diferentes grupos sociales.
- una perspectiva macroscópica del politólogo, donde la atención gira en torno al actor y sobre las relaciones que tiene con las opiniones. Aquí el porcentaje toma toda su importancia. Producto por agregación de las opiniones individuales, es una fuente de legitimidad que puede jugar un papel importancia en la correlación de fuerzas.

En el estudio de valoración, actitudes y sus estructuras se evidencia que es necesario tomar en consideración los siguientes aspectos:

- Escoger una metodología de estudio que esté acorde con las siguientes interrogantes: ¿Cómo hacer de manera que el sistema de implementado no sea complicado sin dejar de ser eficaz? ¿Cómo definir el contenido y la forma del cuestionario? ¿Cuándo y a qué ritmo deberá éste ser administrado?
- Definir el conjunto de tratamientos estadísticos a realizarse; algunos tratamientos pueden ser previstos de antemano.

El Análisis en Componentes Principales toma toda su importancia como metodología para enfocar el problema.

El espectro de los problemas que es posible plantear es bastante variado, por lo que la organización del sistema debe permitir manipular y tratar los datos en forma versátil.

Lo considerado anteriormente nos permite hablar de una *economía de roles*. Es el triple juego de las interacciones siguientes:

- Entre actores sociales: el debate.
- Entre actores y el público: el juego de la oferta y de la demanda de roles.
- Entre individuos sociales: la diferenciación que lleva a las sociedades democráticas modernas hacia nuevos equilibrios, bajo el impulso de los conflictos, donde el debate entre instituciones y las opiniones quedan en correspondencia con sus diversidades.

<sup>1</sup>El desarrollo completo del trabajo aparece en: Poltronieri, J. (1996) "Análisis de las estructuras de la opinión pública", *Revista de Matemática: Teoría y Aplicaciones* 3(2).

<sup>2</sup>PIMAD, ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

## PIMAD 2.1: LOS ALGORITMOS

OLDEMAR RODRÍGUEZ ROJAS<sup>1</sup>

---

### Resumen

En este artículo se presentan en pseudocódigo los principales algoritmos que están implementados en cada uno de los módulos del programa PIMAD 2.1, esto con el fin de que los usuarios del programa puedan estar seguros de que los cálculos que el programa realiza se adaptan realmente a su problema.

### 1. Introducción

El PIMAD 2.1 ha sido implementado para efectuar análisis de datos siguiendo las metodologías francesas. Además ha sido diseñado ejecutarlo bajo Windows 3.1, Windows for Workgroups 3.11 y WindowsNT. Además funciona bastante bien bajo el sistema operativo Windows95. Para una guía de su uso consulte [Rodríguez96]

Esta versión de programa consta de 4 módulos, el primero permite realizar análisis univariados y bivariados de datos, el segundo está diseñado para el análisis en componentes principales, el tercero permite efectuar análisis de tablas múltiples de datos utilizando el método Statis y el cuarto módulo permite realizar clasificación jerárquica. En las siguientes secciones se describen con detalle los algoritmos más importantes implementados en cada uno de estos módulos.

### 2. El módulo de análisis en componentes principales

#### 2.1 El algoritmo

En esta sección se presenta el algoritmo del Análisis en Componentes Principales (ACP) utilizado por el *PIMAD 2.1*. Se supondrá que en el espacio de los individuos se usa la métrica:

---

<sup>1</sup> Escuela de Matemática, Universidad de Costa Rica.

$$M = D_{1/S^2} = \begin{pmatrix} 1/S_1^2 & 0 & \dots & 0 \\ 0 & 1/S_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/S_p^2 \end{pmatrix},$$

y que sobre el espacio de los caracteres se usa la métrica  $D_{1/n}$  dada por la matriz:

$$D_{1/n} = \begin{pmatrix} 1/n & 0 & \dots & 0 \\ 0 & 1/n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/n \end{pmatrix}.$$

### Algoritmo:

**Paso 1:** Se centra la tabla de datos  $X$ , restando a cada entrada el promedio de la columna respectiva, se le llama de nuevo la tabla  $X$ .

**Paso 2:** Se calcula la matriz de correlación  $R$ .

**Paso 3:** Se calculan los valores y vectores propios (ortonormales) de  $R$ ,  $v_1, v_2, \dots, v_p$  y  $\lambda_1, \lambda_2, \dots, \lambda_p$  respectivamente.

**Paso 4:** Se calculan las componentes principales mediante la fórmula:

$$C^i = XMv_i \text{ con } i = 1, 2, \dots, p.$$

**Paso 5:** Se calculan las coordenadas de los caracteres, para graficarlos en los círculos de correlación. Se obtiene una matriz de correlaciones entre los caracteres originales y los componentes principales.

**Paso 6:** Se generan los planos principales y los círculos de correlación.

## 2.2 ¿Cómo ejecutar este algoritmo?

Con el menú ACP-Paso-a-Paso, que se presenta en la Figura 1, se puede realizar un Análisis en Componentes Principales (ACP) siguiendo, paso por paso, el algoritmo que presentado en la sección 2.1.

El paso 1 del algoritmo (centrar la tabla de datos) es ejecutado en forma automática por el programa cuando el archivo es cargado en memoria. Con la opción ACP-Paso-a-Paso|Calcular la Matriz Correlación se puede ejecutar el paso 2 del algoritmo, es decir, se calcula la matriz de correlación, esta matriz es guardada en un archivo tipo ASCII denominado MCORRELA.TXT. Para ejecutar el paso 3 del algoritmo (calcular los vectores

y valores propios) ejecute la opción **ACP-Paso-a-Paso|Calcular Vectores y Valores Propios**, los vectores propios son almacenados en el archivo VECTORP.TXT y los valores propios en el archivo VALORP.TXT. Al paso 4 del algoritmo (calcular las componentes principales) corresponde la opción **ACP-Paso-a-Paso|Calcular Componentes Principales**, las componentes principales son guardadas en forma columnar en el archivo COMPONE.TXT. El paso 5 del algoritmo puede ser ejecutado con la opción **ACP-Paso-a-Paso|Calcular Principales Correlaciones**, las principales correlaciones son almacenadas en el archivo PRINCORR.TXT en forma columnar. Para generar los gráficos del plano principal y del círculo de correlación (paso 6 del algoritmo) están las opciones **ACP-Paso-a-Paso|Graficar el Plano Principal**, y **ACP-Paso-a-Paso|Graficar el Círculo de Correlación** respectivamente.

ACP-Paso-a-Paso
Calcular los Vectores y Valores Propios
Calcular Componentes Principales
Cargar Componentes Principales
Graficar el Plano Principal...
Generar Archivo LATEX del Plano Principal
Calcular Principales Correlaciones
Cargar Principales Correlaciones
Graficar el Círculo de Correlación
Generar Archivo LATEX del Círculo de Correlación

FIGURA 1. Menú **ACP-Paso-a-Paso**.

Con la opción **ACP-Paso-a-Paso|Cargar Componentes Principales** se pueden cargar directamente en memoria las componentes principales para generar el plano principal, ya sea en formato Windows o LATEX. El archivo con las componentes principales debe tener el mismo formato que el especificado en la sección 3.1 de [Rodríguez96] para la tabla de datos. Además con la opción **ACP-Paso-a-Paso|Cargar Principales Correlaciones** se pueden cargar directamente en memoria las principales correlaciones para generar el círculo de correlación, también en formato Windows o LATEX. El archivo con las principales correlaciones también debe tener el mismo formato que el especificado para la tabla de datos.

Como ya se mencionó, con este menú se pueden generar gráficos en formato LATEX, tanto para el plano principal como para el círculo de correlación con las opciones **Generar Archivo LATEX del Plano Principal** y **Generar Archivo LATEX del Círculo de Correlación** respectivamente. Estos gráficos son guardados en un archivo extensión .TEX, el nombre es seleccionado por el usuario. Para imprimir el gráfico de este tipo debe compilarse con algún compilador LATEX estándar.

### 3. El módulo para el método Statis

#### 3.1 El algoritmo

En esta sección se presenta un algoritmo para el método Statis que utiliza el PIMAD 2.1. Este algoritmo es presentado con mayor detalle en [González-Rodríguez 95].

El algoritmo lo dividimos en: Entrada de datos, Intraestructura e Interestructura. En la métrica  $D$  consideramos todos los pesos iguales y  $M_k$  es la matriz identidad.

**Entrada:** Se reciben  $r$  tablas (matrices) de datos  $X_k$  individuo $\times$ variables de tamaño  $n \times p_k$ , centradas respecto a  $D$ , donde  $n$  es el número de individuos, el mismo en las  $r$  tablas,  $p_k$  es el número de variables en la  $k$ -ésima medición,  $D = \text{diag}(1/n, 1/n, \dots, 1/n)$  es la matriz de pesos, invariable para las  $r$  tablas.

**Interestructura:** (pasos 1-4). Como los individuos en las diferentes mediciones han sido los mismos, comparamos su distribución espacial a través de las matrices  $W_i = X_i X_i^t$ . Para ello usamos la métrica definida por el producto interno:

$$\langle W_i, W_j \rangle_\phi = \text{traza}(W_i W_j) = \langle \vec{W}_i, \vec{W}_j \rangle_I$$

donde  $\vec{W}_i \in \mathbb{R}^{n^2}$  es un vector formado por todas las filas de la matriz  $W_i$  para  $i = 1, 2, \dots, r$ .

**Paso 1:** Calcular las matrices  $W_i$  de tamaño  $n \times n$  dadas por  $W_i = X_i X_i^t$  para  $i = 1, 2, \dots, r$

**Paso 2:** Generar la matriz

$$X = \left[ \vec{W}_1 | \vec{W}_2 | \dots | \vec{W}_r \right]_{n^2 \times r}$$

**Paso 3:** Efectuar el Análisis en Componentes Principales del triplete  $(X, D_{\frac{1}{\sigma^2}}, \frac{1}{n^2} J_{n^2})$ . (Se utiliza el algoritmo presentado en la sección 4). Luego se guarda el primer vector propio de este ACP, que denotamos por  $u = (u_1, \dots, u_r)$ , y el primer valor propio  $\lambda_1$  para usarlos en el compromiso.

**Paso 4:** Graficar el círculo de correlaciones (en éste aparecerán representados los vectores  $\vec{W}_i$ ).

**Intraestructura:** (pasos 5-13). El estudio evolutivo de individuos y variables lo realizamos con la factorización del compromiso  $\sum_{k=1}^r \beta_k W_k$  obtenida del ACP de  $(\bar{X}, I_l, D)$ , donde  $l = \sum_{k=1}^r p_k$ ,  $\beta_k$  está definido en el paso 5 y  $\bar{X}$  se define como sigue:

$$\tilde{X} = \left[ \sqrt{\beta_1} X_1 \mid \sqrt{\beta_2} X_2 \mid \dots \mid \sqrt{\beta_r} X_r \right]_{n \times \sum_{k=1}^r p_k}$$

Dado que el operador  $WD$  del ACP de  $(\tilde{X}, I_l, D)$  es igual al compromiso, se obtiene una representación simultánea de las  $\sum_{k=1}^r p_k$  variables en la base  $D$ -ortonormal de los vectores propios de  $WD$ , que denominamos ejes del compromiso.

Pero el ACP anterior no resuelve el problema de la representación simultánea de los individuos, para ello definimos las coordenadas de los individuos de la  $k$ -ésima tabla como la  $D$ -proyección ortogonal de las filas de la matriz  $W_k$  sobre los ejes del compromiso. Para una justificación de esta definición ver [Lavit 88].

Una trayectoria es la representación de un mismo individuo a través de las diferentes tablas, en los ejes del compromiso.

#### **Evolución de las variables: (pasos 5-10)**

**Paso 5:** Calcular  $\beta = (\beta_1, \beta_2, \dots, \beta_r) = \frac{1}{\sqrt{\lambda_1}} u$ .

**Paso 6:** Calcular la matriz por bloques  $\tilde{X}$  como sigue:

$$\tilde{X} = \left[ \sqrt{\beta_1} X_1 \mid \sqrt{\beta_2} X_2 \mid \dots \mid \sqrt{\beta_r} X_r \right]_{n \times \sum_{k=1}^r p_k}$$

**Paso 7:** Efectuar el ACP  $(\tilde{X}, I_l, D)$  (no reducido).

**Paso 8:** Guardar la matriz de componentes principales, denotada por:

$$C = [C_1, C_2, \dots, C_l] \text{ donde } l = \sum_{k=1}^r p_k.$$

**Paso 9:** Graficar el círculo de correlación de este ACP, en el cual se podrá estudiar la evolución de las variables.

**Paso 10:** Graficar el plano principal de este ACP, en el que aparecen representados los  $n$  individuos promedio, ver definición en [Lavit 88].

#### **Evolución de los individuos: (pasos 11-13)**

**Paso 11:** Calcular la matriz IND definida por bloques como sigue:

$$\text{IND} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_r \end{bmatrix}_{(r) \times n}$$

**Paso 12:** Calcular las coordenadas de los individuos a través de las diferentes tablas, denotada por  $\Omega$ , mediante el siguiente producto matricial:

$$\Omega = \text{IND} \times C.$$

**Paso 13:** Graficar el plano principal. En éste se podrá estudiar la evolución de los individuos.

### 3.2 ¿Cómo ejecutar este algoritmo?

Mediante el menú **InterEstructura** y el menú **IntraEstructura** se puede realizar un análisis de tablas múltiples siguiendo el algoritmo de la sección anterior paso por paso.

El menú **InterEstructura** se presenta en la Figura 2. La primera opción de este menú **Calcula Wi-Genera X** corresponde a los pasos 1, 2 y 3 del algoritmo. La opción **Círculo de Correlación de los Wi** corresponde al paso 4 del algoritmo.

El menú **IntraEstructura** se presenta en la Figura 3. Mediante este menú se puede analizar la evolución de las variables y de los individuos.

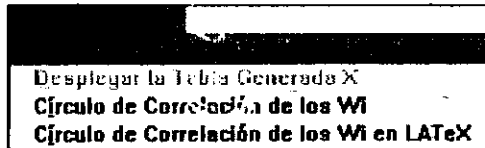


FIGURA 2. Menú *InterEstructura*.

El submenú **Trayectoria de las Variables** corresponde a los pasos 5 al 10 del algoritmo de la sección anterior. Así con la opción **Trayectoria de las Variables|Calcular Beta** se puede ejecutar el paso 5 del algoritmo. Para ejecutar el paso 6 del algoritmo ejecute la opción **Trayectoria de las Variables|Calcular  $\sim X$** . Para ejecutar el paso 7 del algoritmo, es decir efectuar el  $ACP(\bar{X}, I, D)$  no reducido, de deben ejecutar las opciones **Matriz de Correlaciones**, **Vectores Propios**, **Principales Correlaciones**, y **Componentes Principales** del menú **Trayectoria de las Variables**. El paso 8 del algoritmo (guardar la matriz de componentes principales) es ejecutado en forma automática por el programa. El paso 9 del algoritmo corresponde a la opción **Trayectoria de las Variables|Círculo de Correlación** y el paso 10 corresponde a la opción **Trayectoria de las Variables|Plano Principal-Individuos Promedio**.

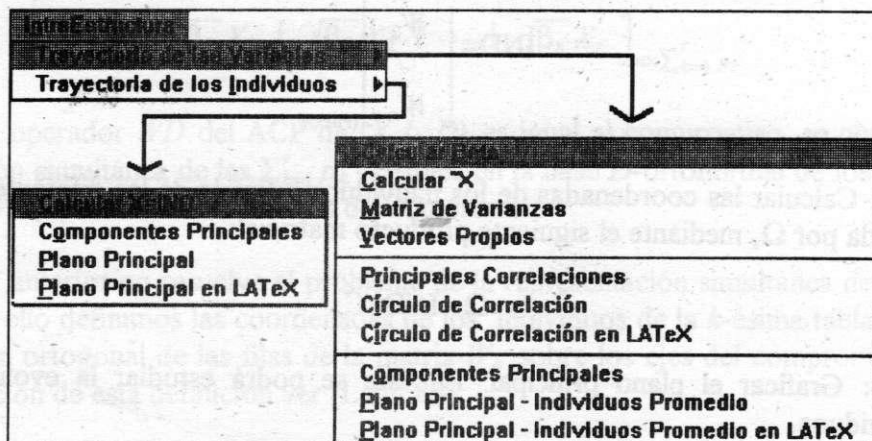


FIGURA 3. Menú IntraEstructura.

El submenú **Trayectoria de los Individuos** corresponde a los pasos 11 al 13 del mismo algoritmo. De modo tal que para calcular la matriz IND definida en el paso 11 del algoritmo ejecute la opción **Calcular X=IND**, el paso 12 del algoritmo corresponde al opción **Componentes Principales**, y el paso 13 corresponde a la opción **Plano Principal** de este menú.

## 4. El módulo de clasificación jerárquica

### 4.1 El algoritmo utilizado por este módulo

Los métodos de clasificación jerárquica ascendente tienen como propósito central construir un conjunto de particiones, llamado jerarquía binaria, por un procedimiento que se puede resumir en cuatro pasos:

1. Se inicia con la partición cuyas clases se reducen a un elemento.
2. Si  $P$  es la partición construida en el paso  $k$ , entonces se localizan dos clases de  $P$  que sean las más próximas. Se trata de una optimización local en  $P$  que se hace en cada paso del algoritmo.
3. Se fusionan las dos clases encontradas anteriormente y se define a partir de  $P$ , una nueva partición, suprimiendo las dos clases fusionadas y agregando la nueva clase.
4. Repetir los pasos 2. y 3. hasta que la partición resultante tenga solo una clase.

Los métodos de clasificación jerárquica requieren de un índice para medir la proximidad entre individuos, llamado disimilitud, y de otro para medir la proximidad entre grupos de individuos, llamado índice de agregación. Se usarán la distancia euclídea clásica y la distancia  $\chi^2$  como índice de disimilitud y el conocido índice de agregación de Ward como índice de agregación, cuya definición es:



$$\delta_w(C_i, C_j) = \frac{p(C_i)p(C_j)}{p(C_i)+p(C_j)} d^2(g_i, g_j)$$

donde  $p(C_i) = \sum_{x \in C_i} p_x$  y  $g_i = \sum_{x \in C_i} \frac{p_x}{p(C_i)} x$  son, respectivamente, el peso y el centro de gravedad de la clase  $C_i$ .

Seguidamente se presenta en pseudocódigo el algoritmo de clasificación jerárquica utilizado por el PIMAD 2.1. [Castillo-Rodríguez-Trejos-Chacón 95]

#### Entrada:

- La tabla de datos  $X$ , de modo tal que en las filas están los individuos y en las columnas las variables.
- La distancia  $d$  escogida.
- la disimilitud  $\delta$  escogida.

#### Salidas:

- Una partición del conjunto de filas de  $X$ .
- Representación gráfica de la jerarquía binaria.

#### Algoritmo:

##### Paso 0: Inicialización

1. Poner  $k = 0$ .
2. Definir  $P_0 = \{C_1, \dots, C_m\}$ .
3. Definir la jerarquía inicial:  $H_0 = P_0$ .
4. Calcular la matriz inicial de agregaciones:  $\delta_w(C_h, C_i)$  para  $h < i$ .

##### Paso 1: Formación de una nueva clase

1. Se determinan las dos clases de  $P_k$  más cercanos en el sentido de la agregación  $\delta_w$ . Es decir, si  $C_s$  y  $C_q$  son estas dos clases, entonces  $\delta_w(C_s, C_q) = \min\{\delta_w(X, Y) \mid X, Y \in P_k\}$ . Si hay varias clases que tienen esta propiedad, se escogen dos arbitrariamente. Por ejemplo, las dos que primero fueron encontradas.
2. Se fusionan las clases  $C_s$  y  $C_q$ , como sigue:  $C = C_s \cup C_q$ .

##### Paso 2: Actualización

1. Sea  $k \leftarrow k + 1$ .
2. Sea  $P_k \leftarrow [P_{k-1} \cup \{C\}] - \{C_s, C_q\}$ .
3. Sea  $H_k \leftarrow H_{k-1} \cup \{C\}$ .
4. Calcular la matriz de agregaciones sobre las clases de  $P_k$ .

**Paso 3: Test**

1. Si  $k < n - 2$  regresar al Paso 1.
2. Si  $k = n - 2$ , hacer la última fusión y terminar.

**4.2 ¿Cómo ejecutar este algoritmo?**

El algoritmo de la sección anterior se puede ejecutar paso por paso mediante el menú **Clasificación**, este menú se presenta en la Figura 4.

El paso 0 del algoritmo presentado en la sección anterior se ejecuta en forma automática por el programa. Para realizar el paso 1 del algoritmo ejecute la opción **Generar la Jerarquía**, cada nueva partición generada en el paso  $k$  es grabada en forma automática por el programa en el archivo ASCII denominado PARTI.K, donde  $K=k$  (el  $K$  se refiere a la extensión del archivo). El paso 2 del algoritmo se ejecuta también con la opción **Generar la Jerarquía**, la matriz de agregaciones en el paso  $k$  es grabada también en forma automática, en un archivo cuyo nombre es MATRIZT.K, donde  $K=k$ .

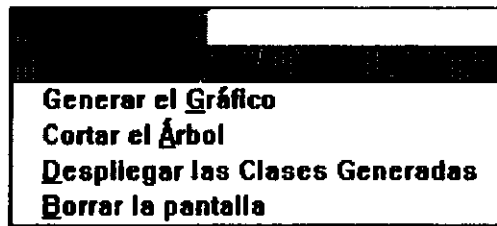


FIGURA 4. Menú Clasificación.

**5. Referencias**

- [Castillo 89] Castillo W. "Métodos y resultados en clasificación automática". Revista Ciencia y tecnología, vol. XIII, No. 1-2, pp. 105-116, 1989.
- [Castillo-Rodríguez-Trejos-Chacón 95] Castillo W. Rodríguez O. Trejos J. Chacón B. "Construcción de tipologías de consumidores del sistema eléctrico". Memorias del IV Encuentro Centroamericano de Investigadores en Matemática, Guatemala 1995.
- [González-Rodríguez 95] González J. Rodríguez O. "Algoritmo e implementación del método Stasis". Memorias del IX simposio de métodos matemáticos aplicados a las ciencias, 1995.
- [Lavit 88] Lavit Ch. "Analyse conjointe de tableaux quantitatifs". Ed. Mason, Paris.
- [Murillo 96] Murillo A. "Proposición de un índice para la interpretación de árboles ascendentes de clasificación basados en conjuntos difusos". Tesis para optar al grado de Magister Scientiae en Computación, Instituto Tecnológico de Costa Rica, 1996.

- [Piza 87] Piza E. "*La clasificación automática jerárquica aglomerativa*". Revista Ciencias económicas, vol. VII, No. 1, pp. 95-111, 1987.
- [Rodríguez 93] Rodríguez O. "*Análisis orientado a objetos: Una aplicación al análisis de datos*". Tesis para optar al grado de Magister Scientiae en Computación, Instituto Tecnológico de Costa Rica, 1993.
- [Rodríguez 96] Rodríguez O. "*PIMAD 2.1 Guía al usuario*". Memorias del X simposio de métodos aplicados a la ciencias, 1996.
- [Trejos96] Trejos J. "*Introducción al análisis de datos*". Escuela de matemática, Universidad de Costa Rica, 1996.

## PIMAD 2.1: GUÍA AL USUARIO

OLDEMAR RODRÍGUEZ ROJAS<sup>1</sup>

---

### Resumen

En este artículo se presenta una guía breve para el usuario del PIMAD 2.1. Paralelo a este artículo en [Rodríguez96] se pueden consultar los principales algoritmos que el están implementados en cada uno de los módulos del programa.

### 1. Introducción

El PIMAD 2.1 es un sistema desarrollado por el autor del presente trabajo, con el propósito de realizar análisis datos siguiendo las metodologías francesas. El desarrollo de este sistema ha tomado casi tres años, la impleme no reducido, de deben ejecutar las opciones **Matriz de Correlaciones**, **Vectores Propios**, **Principales Correlaciones**, y **Componentes Principales** del menú **Trayectoria de las Variables**. El paso 8 del algoritmo (guardar la matriz de componentes principales) es ejecutado en forma automática por el programa. El paso 9 del algoritmo corresponde a la opción **Trayectoria de las Variables|Círculo de Correlación** y el paso 10 corresponde a la opción **Trayectoria de las Variables|Plano Principal-Individuos Promedio**.

ntación del mismo se realizó utilizando el lenguaje de programación C++, en la versión 4.5 de la empresa Borland. Además el sistema ha sido diseñado y programado por completo utilizando las técnicas de la Orientación a Objetos. El PIMAD 2.1 ha sido diseñado para ser ejecutado bajo Windows 3.1, Windows for Workgroups 3.11 y WindowsNT. Además funciona bastante bien bajo el sistema operativo Windows95.

Esta versión de programa consta de 4 módulos, el primero permite realizar análisis univariados y bivariados de datos, el segundo está diseñado para el análisis en componentes principales, el tercero permite efectuar análisis de tablas múltiples de datos utilizando el método Statis y el cuarto módulo permite realizar clasificación jerárquica. En las siguientes secciones se describe con detalle la forma en que se debe utilizar cada uno de estos módulos, además se explica como se instala el programa.

### 2. ¿Cómo se instala el Programa PIMAD 2.1?

En la versión 2.1 los archivos del sistema vienen empaquetados en dos discos de instalación, por esta razón para instalar los diferentes módulos del programa se debe ejecutar el

---

<sup>1</sup> Escuela de Matemática, Universidad de Costa Rica.

archivo INSTALAR.BAT, éste se encuentra en el disquete de instalación número 1. Este programa creará un directorio denominado C:\TEMPORAL en el cual copiarán y descomprimirán todos los archivos necesarios para instalar el sistema PIMAD 2.1.

En el directorio C:\TEMPORAL quedará ubicado el archivo INSTALL.EXE, ejecute este programa desde Windows y siga las instrucciones que el instalador le proporciona. Al finalizar el proceso de instalación el directorio C:\TEMPORAL puede ser borrado.

El proceso BATCH que está en el archivo INSTALAR.BAT fue diseñado para funcionar desde una unidad de 1.44 ubicada en A:. Si su unidad de 1.44 está ubicada en B: edite con cualquier editor que permita archivos ASCII, el archivo INSTALAR.BAT y cambie la línea 4 por la siguiente:

B:

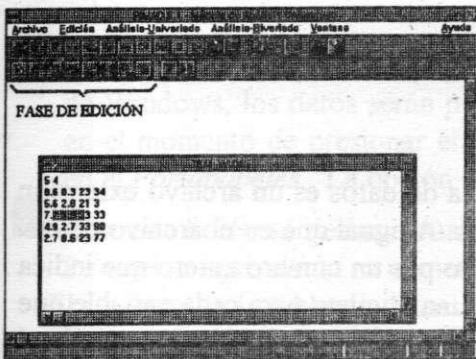


Una vez instalado el sistema se creará la ventana que de la Figura 1. En esta ventana aparece un icono para cada uno de los módulos del sistema.

FIGURA 1. Ventana de la aplicación

### 3. El módulo de análisis univariado y bivariado

Para ejecutar el módulo de análisis univariado y bivariado presione doble clic sobre el icono anterior. Cuando el programa se ejecuta se presenta la pantalla que aparece en la Figura 2. Como puede verse en ésta, el programa puede ejecutarse mediante el menú principal o bien mediante la barra de botones



("tool bar"). Se recomienda usar el programa mediante la barra de botones pues es mucho más rápido. Si se desea conocer la función de cada botón, basta deslizar el cursor del mouse sobre éste y un rótulo que aparece en la parte inferior de la pantalla le dará una explicación breve, para una explicación más amplia se puede consultar la ayuda del sistema. Como puede verse en la Figura 2 este modulo, al igual que los demás que forman parte de PIMAD 2.1, tiene un grupo de botones que permiten editar la tabla de datos y los archivos de etiquetas.

FIGURA 2. Ventana principal del módulo

#### 3.1 El editor de datos (Fase de edición de datos)

El sistema PIMAD 2.1 funcionan con base a tres archivos tipo ASCII, uno para la tabla de datos extensión TXT, otro para las etiquetas de las variables extensión ETV, y el tercero para las etiquetas de los individuos extensión ETI.

El formato de tabla de datos es el siguiente: en la primera fila deben aparecer dos números enteros, el primero indica el número de individuos de la tabla de datos (número de

El formato de tabla de datos es el siguiente: en la primera fila deben aparecer dos números enteros, el primero indica el número de individuos de la tabla de datos (número de filas de la matriz) y el segundo indica el número de variables de la tabla de datos (número de columnas de la matriz). Luego deben aparecer los datos de cada individuo, uno en una fila separados por blancos. Por ejemplo, la siguiente tabla de datos tiene cinco individuos (en este caso estudiantes) y cuatro variables (en este caso las notas de los estudiantes en matemática, ciencias, español e inglés). Esta tabla se puede almacenar en un archivo cuyo nombre podría ser EJEMPLO.TXT.

5	4		
90	92	70	65
95	92	80	70
60	60	90	92
50	50	75	75
60	65	80	88

Para esta tabla de datos se deben crear dos archivos de etiquetas. Uno para los individuos extensión ETI. En la primera fila de este archivo debe aparecer un número entero que indica el número de individuos que hay en la tabla de datos y en las siguientes filas una etiqueta para cada individuo. Por ejemplo, el archivo EJEMPLO.ETI corresponde al archivo de etiquetas de individuos de la tabla de datos presentada arriba, y su contenido es el siguiente:

```
5
Ana
Pedro
Luis
Damaris
Alfredo
```

El segundo archivo de etiquetas asociado a una tabla de datos es un archivo extensión ETV el cual corresponde a las etiquetas de las variables. Al igual que en el archivo de etiquetas de individuos, este archivo debe estar encabezado por un número entero que indica el número de variables presentes en la tabla de datos y una etiqueta para cada variable que este presente en la tabla de datos. Por ejemplo, el archivo EJEMPLO.ETV corresponde al archivo de etiquetas de variables de la tabla de datos presentada arriba, y su contenido es el siguiente:

```
4
Matemática
Ciencias
Español
Inglés
```

Con el propósito de facilitar la creación y edición de la tabla de datos y de los archivos de etiquetas, el PIMAD 2.1 proporciona procedimientos para editar archivos tipo ASCII. Para esto se proporcionan una serie de botones que se explican a continuación:

- Con este botón se puede crear un nuevo archivo tipo ASCII, la opción equivalente en el menú es **Archivo|Crear Nueva Tabla de Datos para Edición...**
- Con este botón se puede cargar, en una ventana de edición, un archivo ASCII existente para modificarlo, la opción equivalente en el menú es **Archivo|Abrir una Nueva Tabla de Datos para Edición...**
- Con este botón se puede guardar la tabla que se está editando, el botón estará apagado si no hay ninguna tabla cargada en la ventana de edición. La opción equivalente en el menú es **Archivo|Guardar la Tabla.**
- Este botón permite cortar o eliminar una porción de texto o de datos en archivo, el texto debe estar previamente seleccionado con el mouse. Los datos eliminados se almacenan temporalmente en el *Portapapeles* de Windows, por lo que pueden ser pegados en otra parte del mismo archivo, en otro archivo o incluso en otra aplicación de Windows. Es importante hacer notar que una acción realizada con este botón puede ser anulada con el botón ■. La opción equivalente en el menú es **Edición|Cortar.** Este botón estará apagado si no existe ninguna porción del archivo seleccionado. No debe confundirse este botón con el utilizado para cortar el árbol binario ■ en el módulo de clasificación jerárquica.
- Este botón se utiliza para enviar una porción del archivo de datos al *Portapapeles* de Windows, el texto debe estar previamente seleccionado con el mouse. La opción equivalente en el menú es **Edición|Copiar.** Estos datos pueden ser pegados en otra parte del mismo archivo, en otro archivo o incluso en otra aplicación de Windows. Este botón también estará apagado si no existe ninguna porción del archivo seleccionado.
- Con este botón se pueden pegar los datos que están almacenados en el *Portapapeles* de Windows, los datos serán pegados en la posición en donde se encuentre el cursor en el momento de presionar el botón. Este botón estará apagado si no existen datos en el *Portapapeles.* La opción equivalente en el menú es **Edición|Pegar.**
- Con este botón se deshace cualquier operación realizada con el botón de Cortar ■. La opción equivalente en el menú es **Edición|Deshacer.**

El formato de la tabla de datos y de los archivos de etiquetas es el mismo para todos los módulos de PIMAD 2.1. Además todos los módulos poseen un editor de datos idéntico al explicado en esta sección.


### 3.2 Cargar los datos

El módulo de análisis univariado y bivariado recibe dos archivos, un archivo extensión TXT con la tabla de datos y otro extensión ETV con las etiquetas de las variables. El formato de estos archivos es como se explicó en la sección 3.1.

Para cargar la tabla de datos en memoria presione el botón ■, o equivalentemente seleccione la opción **Archivo|Cargar Tabla de Datos en Memoria.** Una vez ejecutada esta

operación el sistema le presenta una caja de diálogo mediante la cual se puede seleccionar la tabla deseada, posteriormente le presentará otra caja de diálogo para que seleccione el archivo de etiquetas para las variables. El sistema le informará si los datos fueron cargados satisfactoriamente.

### 3.3 Seleccionar las variables para el análisis












Para seleccionar las variables a las cuales se les desea hacer algún tipo de análisis con este módulo dé clic sobre el botón , o en forma equivalente seleccione la opción de menú **Análisis-Univariado|Seleccionar las Variables**.

Para seleccionar una variable presione doble clic sobre ésta. Para realizar análisis univariado se pueden seleccionar tantas variables como se deseen, y luego se pueden efectuar los cálculos como se explicará en la siguiente sección. Para realizar análisis bivariado se deben seleccionar solamente dos variables, en caso contrario el sistema le enviará un mensaje de error.

### 3.4 Cálculos del análisis univariado

Es posible realizar los cálculos del análisis univariado mediante el menú principal o mediante la barra de botones o herramientas. La barra de herramientas es una fila de botones en la parte superior de la ventana principal la cual contiene los principales comandos de la aplicación. Presionar clic en uno de los botones es en realidad una forma rápida de seleccionar una alternativa del menú. Los Botones en la barra de herramientas están activos o inactivos de acuerdo con el estado de la aplicación.

En la siguiente tabla se explican cada uno los botones para análisis univariado y su opción equivalente en el menú. La definición estadística de cada una de estas medidas de tendencia central y de cada uno de los gráficos se pueden consultar en [Trejos96].

Botón	Acción	Equivalencia en el menú
	Calcula la media aritmética.	Análisis-Univariado Media Aritmética.
	Calcula la mediana.	Análisis-Univariado Mediana.
	Calcula la media de los extremos.	Análisis-Univariado Media de los Extremos
	Calcula la varianza.	Análisis-Univariado Varianza.
	Calcula la desviación estándar.	Análisis-Univariado Desviación Estándar
	Calcula el coeficiente de variación.	Análisis-Univariado Coeficiente de variación
	Calcula la desviación media.	Análisis-Univariado Desviación Media
	Calcula la desviación cuartil.	Análisis-Univariado Desviación Cuartil
	Calcula la extensión.	Análisis-Univariado Extensión
	Produce un histograma.	Análisis-Univariado Histograma
	Produce las cajas de dispersión.	Análisis-Univariado Cajas de Dispersión.



### 3.5 Cálculos del análisis bivariado

También es posible realizar los cálculos del análisis bivariado mediante el menú principal o mediante la barra de botones o herramientas.

En la siguiente tabla se explican cada uno los botones para análisis bivariado y su opción equivalente en el menú. La definición estadística de cada una de los índices y gráficos del análisis bivariado que se pueden calcular con PIMAD 2.1 están definidos en [Trejos96].

Botón	Acción	Equivalencia en el menú
	Produce el diagrama de dispersión.	Análisis-Bivariado D <i>Diagrama de Dispersión.</i>
	Calcula la covarianza.	Análisis-Bivariado C <i>ovarianza.</i>
	Calcula la correlación.	Análisis-Bivariado C <i>orrelación.</i>

### 3.6 Otros botones

El módulo de análisis univariado tiene cuatro botones más. Estos cuatro botones se encuentran también en los demás módulos de PIMAD 2.1 con la misma función.

Este botón se utiliza para imprimir el gráfico que esté desplegado en pantalla, esto incluye todos los cálculos del análisis univariado y los del análisis bivariado, este botón es equivalente a la opción **A**rchivo|**I**mprimir Gráfico. No se debe confundir este botón con el botón , pues este último se utiliza únicamente para imprimir los archivos tipo ASCII que están activos en una ventana de edición, es decir la tabla de datos o los archivos de etiquetas.

Este botón se utiliza para borrar la información que está desplegada en la pantalla, es decir, cualquiera de los cálculos del análisis univariado y los del análisis bivariado. Es equivalente a la opción **A**nálisis-**U**nivariado|**B**orra Gráfico.

Con este botón se puede invocar la ayuda del sistema, en la cual se encuentra explicado con detalle el funcionamiento completo del sistema. Es equivalente a la opción **A**yuda|**C**ontenidos.

Con este botón se puede salir del módulo. Es equivalente a la opción **A**rchivo|**S**alir o a presionar las teclas ALT-F4 simultáneamente.




## 4. El módulo de análisis en componentes principales


### 4.1 Cargar los datos

El módulo de análisis en componentes principales utiliza tres archivos, un archivo extensión TXT con la tabla de datos, un archivo extensión ETI con la etiquetas de los

individuos y un archivo extensión ETV con las etiquetas de las variables. El formato de estos archivos es el mismo que en los otros módulos, tal como se explicó en la sección 3.

Para cargar la tabla de datos y los archivos de etiquetas presione el botón , o equivalentemente seleccione la opción **A**rchivo|**C**argar Tabla de Datos en Memoria. Una vez ejecutada esta operación el sistema le presenta una caja de diálogo mediante la cual se puede seleccionar la tabla deseada, luego presentará otra caja de diálogo para seleccionar el archivo de etiquetas para los individuos y finalmente presentará otra caja de diálogo para escoger el archivo de etiquetas para las variables. El sistema le informará si los datos fueron cargados satisfactoriamente.

#### 4.2 El plano principal y el círculo de correlación

Tanto el plano principal como el círculo de correlación pueden generarse desde la barra de herramientas.  Presionando este botón se producen todos los cálculos necesarios para generar el plano principal, como son la matriz de correlación, los vectores y valores propios, las componentes principales. Mientras estos cálculos se están produciendo el sistema presenta una barra de estado (un medidor conocido en Windows como "gauge") que informa sobre el porcentaje de avance en la ejecución del algoritmo. Luego el programa presenta la caja de diálogo de la Figura 3, en ésta se puede escoger si se desean graficar todos los individuos o más bien seleccionar algunos, si se escoge seleccionar individuos entonces el sistema presenta una caja de diálogo en la que se pueden seleccionar los individuos. En la caja de diálogo de la Figura 3 también se pueden seleccionar los ejes en que se producirá el plano, así como el tamaño del gráfico (pequeño, mediano o grande). Este botón es equivalente a la opción de menú **A**CP-Directo|**P**lano Principal.

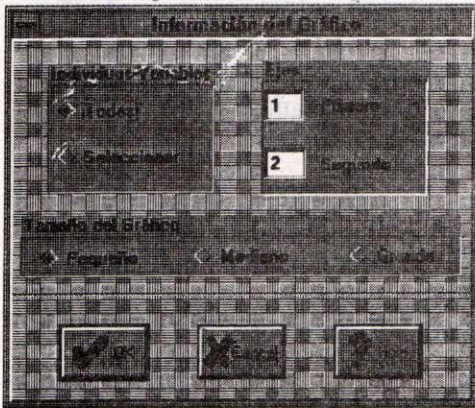




FIGURA 3. Información sobre los gráficos.

 Presionando este botón se producen todos los cálculos necesarios para generar el círculo de correlación, como son la matriz de correlación, los vectores y valores propios, y las principales correlaciones. Al igual que cuando se grafica un plano principal, el sistema presenta una barra de estado, luego presenta la caja de diálogo de la Figura 3, para finalmente presentar el círculo de correlación en pantalla. Este botón es equivalente a la opción de menú **A**CP-Directo|**C**írculo de Correlación.


Una vez que el gráfico está en pantalla se puede cambiar el tamaño de éste presionando el botón . Al presionar este botón el tamaño del gráfico varía de pequeño a mediano, de mediano a grande, de grande a pequeño y así sucesivamente. Este botón es equivalente a la opción de menú **A**CP-Directo|**M**odificar el Tamaño del Gráfico.

Con el menú **ACP-Paso-a-Paso** se puede realizar un análisis en componentes principales (ACP) siguiendo, paso por paso, el algoritmo presentado en [Rodríguez96]. Los resultados intermedios, como son la matriz de correlación, valores y vectores propios, componentes principales y las principales correlaciones, son guardados en archivos tipo ASCII, con lo que podrán ser utilizados con otros fines después a la ejecución del programa.

## 5. El módulo de Stasis





### 5.1 Cargar los datos

El módulo para análisis de tablas múltiples mediante el método Stasis necesita para funcionar una secuencia de archivos para las tablas de datos que contemple el estudio, estos archivos deben tener extensión TXT, uno para cada tabla de datos, además otro archivo debe tener extensión ETI para las etiquetas de los individuos y otro archivo debe tener extensión ETV para las etiquetas de las variables. El formato de estos archivos es el mismo que en los otros módulos, tal como se explicó en la sección 3.

Para cargar las tablas de datos y los archivos de etiquetas presione el botón , o equivalentemente seleccione la opción **Archivo|Cargar Tabla de Datos en Memoria**. Una vez ejecutada esta operación el sistema le presentará una secuencia de cajas de diálogo mediante la cuales se pueden seleccionar las tablas de datos deseadas. Para seleccionar una tabla escoja el nombre del archivo y luego presione el botón *OK* y así sucesivamente seleccione todas las tablas deseadas. Después de seleccionar la última tabla presione el botón *Cancel*, luego el sistema presentará una caja de diálogo para seleccionar el archivo de etiquetas para los individuos y finalmente presentará otra caja de diálogo para que seleccione el archivo de etiquetas de las variables. El sistema le informará si los datos fueron cargados satisfactoriamente.

### 5.2 El menú Stasis-Directo

Mediante este menú, o bien mediante la barra de botones se pueden generar los gráficos finales del método Stasis sin necesidad de conocer los cálculos intermedios que realiza el algoritmo. En la siguiente tabla se explican éstas opciones.

Botón	Acción	Equivalencia en el menú
	Círculo de correlación de los $W_i$ .	Statis-Directo Círculo de Correlación de $W_i$ .
	Círculo de las trayectorias de la var.	Statis-Directo Círculo Trayectoria Variables
	Plano Principal de los ind. promedio.	Statis-Directo Plano Individuos Promedio.
	Plano de la trayectoria de los ind.	Statis-Directo Círculo Trayectoria de los Ind.


Mediante los menús InterEstructura y IntraEstructura se puede realizar un análisis de tablas múltiples siguiendo el algoritmo presentado en [Rodríguez96] paso por paso.

## 6. El módulo de clasificación jerárquica




### 6.1 Cargar los datos

El módulo de clasificación jerárquica recibe un archivo extensión TXT con la tabla de datos y otro extensión ETI con la etiquetas de los individuos. El formato de estos archivos es el mismo que en los otros módulos, tal como se explicó en la sección 3.


Para cargar la tabla de datos en memoria presione el botón , o equivalentemente seleccione la opción **A**rchivo|**C**argar Tabla de Datos en Memoria. Una vez ejecutada esta operación el sistema le presenta una caja de diálogo mediante la cual se puede seleccionar la tabla deseada y posteriormente le presentará otra caja de diálogo para que seleccione el archivo de etiquetas para los individuos. El sistema le informará si los datos fueron cargados satisfactoriamente.


### 6.2 Realizando los cálculos

El primer paso en una clasificación jerárquica es seleccionar el índice de disimilitud, no obstante seleccionar el índice de disimilitud es opcional, pues si no se selecciona ninguno, el sistema usará por defecto la *distancia euclídea* como índice de disimilitud.

Para escoger un índice de disimilitud presione el botón , o equivalentemente escoja la opción **O**pciones|**C**ambiar **D**isimilitud del menú principal. El sistema permite como índices de disimilitud distancia euclídea y la distancia  $\chi^2$ . La definición de estos dos índices se puede consultar en [Piza87].


El segundo paso importante es seleccionar el índice de agregación. En el PIMAD 2.1 seleccionar el índice de agregación es también opcional, pues si no se selecciona ninguno, el sistema usará por defecto índice de agregación de Ward, para la definición de éste índice consulte [Castillo89].

Para escoger un índice de agregación presione el botón , o equivalentemente escoja la opción **O**pciones|**C**ambiar **A**gregación del menú principal. Luego aparecerá la caja de diálogo mediante la cual se puede escoger el índice de agregación deseado. La definición de cada uno de los índices de agregación que utiliza éste módulo de PIMAD 2.1, que aparecen en ésta caja de diálogo, están definidos con detalle en [Piza87].


El siguiente paso en la ejecución del programa es generar la jerarquía binaria, para esto presione el botón , o en forma equivalente escoja la opción **C**lasificación|**G**enerar **J**erarquía... en el menú principal. Al escoger esta opción el programa presenta una barra de estado ("gauge") que le informará sobre progreso que está teniendo la ejecución del algoritmo.


El objetivo de este proceso es generar una serie de archivos y datos que son necesarios para graficar el árbol binario o jerarquía binaria.

Una de las partes más importantes en la ejecución de la clasificación jerárquica es generar el gráfico del árbol, pues este es el que permitirá la escogencia de la partición deseada.

Para graficar el árbol presione el botón , o en forma equivalente seleccione **Clasificación|Generar Gráfico** en el menú principal. En forma casi inmediata el sistema desplegará el árbol en la pantalla.

Una vez que el árbol está desplegado en la pantalla, el proceso siguiente es escoger la partición deseada, para esto se debe "cortar" el árbol en el nivel adecuado, para más detalles consultar [Murillo96].

Para cortar el árbol presione clic sobre el botón , o en forma equivalente seleccione la opción **Clasificación|Cortar el Árbol**. De forma inmediata el cursor se transforma en una "flecha-tijera", luego presione clic en el nivel del árbol en donde desea que se produzca el corte, seguidamente el sistema trazará una línea que indica donde se cortó el árbol y en la parte superior del gráfico indicará cuántas clases tiene la partición seleccionada. Este proceso puede repetirse tantas veces como sea necesario hasta encontrar la partición deseada.

Luego de seleccionar la partición deseada, es de gran importancia conocer qué "individuos" pertenecen a cada clase, para esto presione clic sobre el botón , o en forma equivalente seleccione la opción **Clasificación|Desplegar las Clases Generadas**. Esta opción mostrará la información en la que se especifica a que clase pertenece cada uno de los individuos.

## 7. Referencias

- [Castillo89] Castillo W. "Métodos y resultados en clasificación automática". Revista Ciencia y tecnología, vol. XIII, No. 1-2, pp. 105-116, 1989.
- [Castillo-Rodríguez-Trejos-Chacón95] Castillo W. Rodríguez O. Trejos J. Chacón B. "Construcción de tipologías de consumidores del sistema eléctrico". Memorias del IV Encuentro Centroamericano de Investigadores en Matemática, Guatemala 1995.
- [González-Rodríguez95] González J. Rodríguez O. "Algoritmo e implementación del método Stasis". Memorias del IX simposio de métodos matemáticos aplicados a las ciencias, 1995.
- [Lavit88] Lavit Ch. "Analyse conjointe de tableaux quantitatif". Ed. Mason, Paris.
- [Murillo96] Murillo A. "Proposición de un índice para la interpretación de árboles ascendentes de clasificación basados en conjuntos difusos". Tesis para optar al grado de Magister Scientiae en Computación, Instituto Tecnológico de Costa Rica, 1996.
- [Piza87] Piza E. "La clasificación automática jerárquica aglomerativa". Revista Ciencias económicas, vol. VII, No. 1, pp. 95-111, 19987.
- [Rodríguez93] Rodríguez O. "Análisis orientado a objetos: Una aplicación al análisis de datos". Tesis para optar al grado de Magister Scientiae en Computación, Instituto Tecnológico de Costa Rica, 1993.
- [Rodríguez96] Rodríguez O. "PIMAD 2.1: Los algoritmos". Memorias del X simposio de métodos aplicados a la ciencias, 1996.
- [Trejos96] Trejos J. "Introducción al análisis de datos". Escuela de matemática, Universidad de Costa Rica, 1996.

## LA ECUACIÓN DE EULER PARA PROBLEMAS DE CÁLCULO VARIACIONAL CON RETARDO

LORENA SALAZAR<sup>1</sup>

### Resumen

En este artículo se presenta una generalización de la ecuación de Euler para el problema de cálculo variacional con retardo

$$J[X] = \int_0^T F(t, X(t), X(t-\tau), \dot{X}(t)) dt$$

y se analiza cuando esta vía clásica nos puede llevar a la solución del problema de minimización en el conjunto de funciones admisibles

$$U = \{X : [-\tau, T] \rightarrow \mathbb{R} : X, \dot{X} \text{ son continuas, } X(t) = \phi(t) \text{ en } [-\tau, 0] \text{ y } X(T) = B\}.$$

### 1 Introducción

Cuando se considera el problema más simple de cálculo de variaciones, como es el de minimizar el funcional

$$J[X] = \int_a^b F(t, X, \dot{X}) dt$$

definido en el conjunto de funciones  $X(t)$  las cuales tienen primeras derivadas continuas en  $[a, b]$  y satisfacen  $X(a) = A, X(b) = B$ , es bien sabido que una condición necesaria para que  $J[X]$  alcance un extremo en  $X$ , es que  $X(t)$  satisfaga la ecuación de Euler

$$\frac{d}{dt} F_{\dot{X}} = F_X \quad (1)$$

En este artículo se presenta un análogo de la ecuación (1) para el funcional

$$J[X] = \int_0^T F[t, X(t), X(t-\tau), \dot{X}(t)] dt, \quad (2)$$

para el que existe un retardo en el tiempo para la variable  $X$ , donde  $X(t) \in U$ , el conjunto de las funciones admisibles definido por

$$U = \{X : [-\tau, T] \rightarrow \mathbb{R} : X, \dot{X} \text{ son continuas, } X(t) = \phi(t) \text{ en } [-\tau, 0] \text{ y } X(T) = B\}$$

<sup>1</sup>ESCUELA DE MATEMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA

Para efectos de existencia de soluciones para el problema de minimizar (2) asumimos además que la función

$$F : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$$

es continua y posee derivadas parciales continuas.

## 2 Generalización de la ecuación de Euler

Antes de derivar la ecuación diferencial de Euler asociada a nuestro funcional con retardo  $J[X]$ , enunciaremos algunas definiciones y resultados de la teoría clásica del cálculo variacional que usamos en la prueba de nuestro resultado. Vamos a denotar por

$$\nabla J[X; V] = J[X + V] - J[X]$$

el incremento de  $J[X]$  correspondiente al incremento  $V(t)$  de la variable independiente  $X(t)$ , donde  $V$  es una función admisible, y por  $\delta J[V]$  la variación ó diferencial de  $J[X]$ . Así

$$\nabla J[X; V] = \delta J[V] + \epsilon \|V\|$$

donde  $\delta J[V]$  es un funcional lineal y  $\epsilon \rightarrow 0$  cuando  $\|V\| \rightarrow 0$ .

**Resultado 1** *Una condición necesaria para que un funcional diferenciable  $J[X]$  alcance un extremo en  $X$  es que su variación  $\delta J$  se anule en  $X$ .*

El siguiente resultado es conocido como la derivada de Gâteaux, y señala otra forma de expresar la variación de un funcional.

**Resultado 2** *Para un funcional diferenciable  $J[X]$  se tiene que*

$$\delta J = \frac{\partial}{\partial \epsilon} (J[X + \epsilon V])|_{\epsilon=0}$$

Efectivamente, por un lado

$$\begin{aligned} \nabla J[X; V] &= \int_a^b (F[t, X + V, \dot{X} + \dot{V}] - F[t, X, \dot{X}]) dt \\ &= \int_a^b (F_X[t, X, \dot{X}]V + F_{\dot{X}}[t, X, \dot{X}]\dot{V}) dt + \dots \end{aligned}$$

por el Teorema de Taylor. Aquí los puntos suspensivos denotan las derivadas parciales de mayor orden relativos a  $V$  y  $\dot{V}$ . Por lo tanto

$$\delta J = \int_a^b (F_X[t, X, \dot{X}]V + F_{\dot{X}}[t, X, \dot{X}]\dot{V}) dt$$

Por otro lado

$$\begin{aligned} \frac{\partial}{\partial \epsilon} (J[X + \epsilon V])|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon} \left( \int_a^b F[t, X + \epsilon V, \dot{X} + \epsilon \dot{V}] dt \right) \Big|_{\epsilon=0} \\ &= \int_a^b (F_X[t, X, \dot{X}]V + F_{\dot{X}}[t, X, \dot{X}]\dot{V}) dt = \delta J \end{aligned}$$

El siguiente lemma es conocido como el lema de Du Bois-Raymond, y será usado en la prueba del teorema a probar.

**Lema 1** Si  $\alpha(t)$  es continua en  $[a, b]$  y si

$$\int_a^b [\alpha(t)\dot{V}(t)] dt = 0$$

para toda función  $V(t)$  en  $D_1(a, b)$ <sup>2</sup> tal que  $V(a) = V(b) = 0$ , entonces  $\alpha(t) = C$ , donde  $C$  es una constante.

Finalmente enunciamos el teorema que generaliza la ecuación de Euler para el caso en que hay retardo en el tiempo.

**Teorema 1** Sea  $J[X]$  un funcional de la forma

$$J[X] = \int_0^T F[t, X(t), X(t-\tau), \dot{X}(t)] dt, \quad (3)$$

definido en  $U$ <sup>3</sup>. Entonces una condición necesaria para que  $J[X]$  alcance un extremo en  $X$ , es que  $X(t)$  satisfaga la ecuación de Euler

$$\frac{d}{dt} (F_{\dot{X}}[t, X(t), X(t-\tau), \dot{X}(t)]) = \begin{cases} F_X[t, X(t), X(t-\tau), \dot{X}(t)] + F_X[t+\tau, X(t+\tau), X(t), \dot{X}(t+\tau)] \\ \text{en } [0, T-\tau] \\ F_X[t, X(t), X(t-\tau), \dot{X}(t)], \text{ en } [T-\tau, T] \end{cases}$$

**Prueba:**

Introducimos los siguientes cambios de variable para simplificar la prueba:

$$Y(t) = X(t-\tau), \quad Z(t) = X(t+\tau) \quad (4)$$

Entonces

$$F[t, X(t), X(t-\tau), \dot{X}(t)] = F[t, X(t), Y(t), \dot{X}(t)]$$

Ahora consideramos la dirección admisible  $V$  (i.e.  $X + \epsilon V \in U$  para cualquier  $\epsilon > 0$ ). Usando resultado 2, se tiene que

$$\begin{aligned} \delta J &= \frac{\partial}{\partial \epsilon} \left( \int_0^T F[t, X(t) + \epsilon V(t), X(t-\tau) + \epsilon V(t-\tau), \dot{X}(t) + \epsilon \dot{V}(t)] dt \right) \Big|_{\epsilon=0} \\ &= \int_0^T F_X[t, X(t), Y(t), \dot{X}(t)] V(t) dt + \int_0^T F_Y[t, X(t), Y(t), \dot{X}(t)] V(t-\tau) dt \\ &\quad + \int_0^T F_{\dot{X}}[t, X(t), Y(t), \dot{X}(t)] \dot{V}(t) dt \end{aligned}$$

<sup>2</sup>  $D_1(a, b) = \{V : [a, b] \rightarrow \mathbb{R} : V, \dot{V} \text{ son continuas}\}$

<sup>3</sup>  $U = \{X : [-\tau, T] \rightarrow \mathbb{R} : X, \dot{X} \text{ son continuas}, X(t) = \phi(t) \text{ en } [-\tau, 0] \text{ y } X(T) = B\}$ .



Ahora con un cambio de variable y usando (4) se obtiene

$$\begin{aligned} & \int_0^T F_Y[t, X(t), Y(t), \dot{X}(t)]V(t - \tau) dt \\ &= \int_{-\tau}^{T-\tau} F_Y[t + \tau, X(t + \tau), Y(t + \tau), \dot{X}(t + \tau)]V(t) dt \\ &= \int_{-\tau}^{T-\tau} F_Y[t + \tau, Z(t), X(t), \dot{Z}(t)]V(t) dt. \end{aligned}$$

Dado que  $V$  es una dirección admisible, se tiene que

$$X + \epsilon V \in U \Leftrightarrow (X + \epsilon V)(t) = \phi(t) \text{ en } [-\tau, 0] \Leftrightarrow V(t) = 0 \text{ en } [-\tau, 0]$$

así

$$\int_{-\tau}^{T-\tau} F_Y[t + \tau, Z(t), X(t), \dot{Z}(t)]V(t) dt = \int_0^{T-\tau} F_Y[t + \tau, Z(t), X(t), \dot{Z}(t)]V(t) dt.$$

Por lo tanto

$$\begin{aligned} \delta J &= \int_0^T F_X[t, X(t), Y(t), \dot{X}(t)]\dot{V}(t) dt + \int_0^{T-\tau} F_Y[t + \tau, Z(t), X(t), \dot{Z}(t)]V(t) dt \\ &+ \int_0^T F_{\dot{X}}[t, X(t), Y(t), \dot{X}(t)]\dot{V}(t) dt \\ &= \int_0^{T-\tau} \{ F_X[t, X(t), Y(t), \dot{X}(t)] + F_Y[t + \tau, Z(t), X(t), \dot{Z}(t)] \} V(t) dt \\ &+ \int_{T-\tau}^T F_X[t, X(t), Y(t), \dot{X}(t)]V(t) dt + \int_0^T F_{\dot{X}}[t, X(t), Y(t), \dot{X}(t)]\dot{V}(t) dt. \end{aligned}$$

Definiendo

$$\Psi(t) = F_{\dot{X}}[t, X(t), Y(t), \dot{X}(t)] \text{ en } [0, T], \text{ y}$$

$$\Phi(t) = \begin{cases} F_X[t, X(t), Y(t), \dot{X}(t)] + F_Y[t + \tau, Z(t), X(t), \dot{Z}(t)] & \text{en } [0, T - \tau] \\ F_X[t, X(t), Y(t), \dot{X}(t)] & \text{en } [T - \tau, T], \end{cases}$$

y usando resultado 1, obtenemos

$$\begin{aligned} \delta J &= \int_0^{T-\tau} \Phi(t)V(t) dt + \int_{T-\tau}^T \Phi(t)V(t) dt + \int_0^T \Psi(t)\dot{V}(t) dt \\ &= \int_0^T \Phi(t)V(t) dt + \int_0^T \Psi(t)\dot{V}(t) dt = 0. \end{aligned}$$

Ahora integrando la primera integral de  $\delta J$  por partes, con

$$G(t) = \int_0^t \phi(s) ds \tag{5}$$

obtenemos

$$\int_0^T \Phi(t)V(t) dt = V(t)G(t)|_0^T - \int_0^T G(t)\dot{V}(t) dt = - \int_0^T G(t)\dot{V}(t) dt$$

dado que  $V(T) = 0$  y  $V(0) = 0$ . Por lo tanto

$$\begin{aligned} \delta J = 0 &\Leftrightarrow \int_0^T \Phi(t)V(t) dt + \int_0^T \Psi(t)\dot{V}(t) dt = 0 \\ &\Leftrightarrow - \int_0^T G(t)\dot{V}(t) dt + \int_0^T \Psi(t)\dot{V}(t) dt = 0 \\ &\Leftrightarrow \int_0^T [\Psi(t) - G(t)]\dot{V}(t) dt = 0 \end{aligned}$$

Como lo anterior debe ser satisfecho para cualquier direcci3n admisible  $V$ , se obtiene por el lema de Du Bois-Raymond que,

$$\Psi(t) - G(t) = C,$$

para alguna constante  $C$ . Entonces por (5) se concluye que

$$C + F_{\dot{X}}[t, X(t), Y(t), \dot{X}(t)] = \int_0^t \Phi(s) ds, \quad t \in [0, T]$$

Finalmente reemplazando  $\Phi(t)$  obtenemos

$$F_{\dot{X}}[t, X(t), Y(t), \dot{X}(t)] = \begin{cases} C + \int_0^t (F_X[s, X(s), Y(s), \dot{X}(s)] + F_Y[s + \tau, Z(s), X(s), \dot{Z}(s)]) ds, \\ \text{en } [0, T - \tau] \\ \int_0^t F_X[s, X(s), Y(s), \dot{X}(s)] ds + C, \quad \text{en } [T - \tau, T] \end{cases}$$

y cambiando a las variables originales:

$$\frac{d}{dt} (F_{\dot{X}}[t, X(t), X(t - \tau), \dot{X}(t)]) = \begin{cases} F_X[t, X(t), X(t - \tau), \dot{X}(t)] + F_{X(t-\tau)}[t + \tau, X(t + \tau), X(t), \dot{X}(t + \tau)] \\ \text{en } [0, T - \tau] \\ F_X[t, X(t), X(t - \tau), \dot{X}(t)], \quad \text{en } [T - \tau, T] \end{cases} \quad (6)$$

queda probado el Teorema 1. Para simplificar la notación, denotemos por

$$F_X(t) = F_X[t, X(t), X(t - \tau), \dot{X}(t)]$$

$$F_{\dot{X}}(t) = F_X[t, X(t), X(t - \tau), \dot{X}(t)]$$

$$F_{X(t-\tau)}(t) = F_{X(t-\tau)}[t, X(t), X(t - \tau), \dot{X}(t)]$$

$$F_{X(t-\tau)}(t + \tau) = F_{X(t-\tau)}[t + \tau, X(t + \tau), X(t), \dot{X}(t + \tau)]$$

de modo que (6) se reduce a

$$\frac{d}{dt}[F_{\dot{X}}(t)] = \begin{cases} F_X(t) + F_{X(t-\tau)}(t + \tau) & \text{en } [0, T - \tau] \\ F_X(t), & \text{en } [T - \tau, T] \end{cases} \quad (7)$$

### 3 Casos especiales de la ecuación de Euler

En esta sección analizamos algunos casos especiales de (6) en que alguna de las variables esta ausente y el caso de integrandos particulares en que esta ecuación podría ser resuelta.

**Caso 0:** El integrando no depende de la variable  $X(t - \tau)$ , es decir no existe retardo en nuestro funcional. En este caso

$$J[X] = \int_0^T F[t, X(t), \dot{X}(t)] dt, \quad (8)$$

es el funcional más simple dentro de la teoría de cálculo de variaciones, para el cual la ecuación de Euler asociada es

$$\frac{d}{dt} F_{\dot{X}} = F_X \quad (9)$$

como se había señalado en (1) al inicio de este artículo.

**Caso 1:** El integrando no depende de la variable  $X(t)$ . En este caso

$$J[X] = \int_0^T F[t, X(t - \tau), \dot{X}(t)] dt, \quad (10)$$

y la ecuación de Euler asociada a (10) se reduce a

$$\frac{d}{dt}[F_{\dot{X}}(t)] = \begin{cases} F_{X(t-\tau)}(t + \tau) & \text{en } [0, T - \tau] \\ 0 & \text{en } [T - \tau, T] \end{cases}$$

la cual es de nuevo una ecuación diferencial con retraso y adelanto a la vez.

**Caso 2:** El integrando no depende de la variable  $\dot{X}(t)$ . En este caso

$$J[X] = \int_0^T F[t, X(t), X(t - \tau)] dt, \quad (11)$$

y la ecuación de Euler asociada a (11) se reduce a

$$0 = \begin{cases} F_X(t) + F_{X(t-\tau)}(t + \tau) & \text{en } [0, T - \tau] \\ F_X(t) & \text{en } [T - \tau, T] \end{cases}$$

**Caso 3:** El integrando es de la forma

$$J[X] = \int_0^T M[t, X(t), \dot{X}(t)] + N[t, X(t - \tau)] dt, \quad (12)$$

y la ecuación de Euler asociada a (12) se reduce a

$$\frac{d}{dt}[M_{\dot{X}}] = \begin{cases} M_{X(t)}(t) + N_{X(t-\tau)}(t + \tau) & \text{en } [0, T - \tau] \\ M_{X(t)}(t) & \text{en } [T - \tau, T] \end{cases}$$

la cual es una ecuación diferencial donde no hay retraso, solo adelanto, y por lo tanto podría ser posible encontrar una solución.

Finalmente, se presenta a continuación un ejemplo ilustrativo que muestra tanto la utilidad del resultado anterior, así como también sus grandes limitaciones, en la mayoría de los casos.

**Ejemplo 1** *Minimice el funcional*

$$J[X] = \int_0^2 [\dot{X}(t)^2 + X(t)^3 + X(t-1)X(t)\dot{X}(t)] dt$$

$$X(t) = 0 \quad \text{si } t \in [-1, 0]; \quad X(2) = 0$$

La ecuación de Euler para este funcional es:

$$\frac{d}{dt}[2\dot{X}(t) + X(t-1)X(t)] = \begin{cases} 3X(t)^2 + X(t-1)\dot{X}(t) + X(t+1)\dot{X}(t+1) \\ \text{si } 0 \leq t \leq 1, \\ 3X(t)^2 + X(t-1)\dot{X}(t) \quad \text{si } 1 \leq t \leq 2 \end{cases}$$

la cual es una ecuación diferencial que presenta tanto un atraso como un adelanto, lo cual dificulta enormemente el estudio de estos casos. Sin embargo en este caso particular, se puede notar que la ecuación de Euler es satisfecha por la trayectoria  $X_0(t) \equiv 0$ . Usando expansiones de la serie de Fourier coseno para hallar una aproximación a este problema, como sugiere [7], y el paquete de optimización llamado GAMS (General Algebraic Modeling

t	$X_n(t)$	$X(t)$
0.00000	0.00000	0.00000
0.20000	0.00000	0.00000
0.40000	0.00000	0.00000
0.60000	0.00000	0.00000
0.80000	0.00000	0.00000
1.00000	0.00000	0.00000
1.20000	0.00000	0.00000
1.40000	0.00000	0.00000
1.60000	0.00000	0.00000
1.80000	0.00000	0.00000
2.00000	0.00000	0.00000

Tabla 1: GAMS: Ejemplo 1 (n=25)

System), usado en [6], el problema original se transforma en el siguiente problema de aproximación:

$$\begin{aligned}
 J[X_n] = & \int_0^1 \left[ \sum_{k=1}^n -(k\pi/T) a_k \text{sen}(k\pi/T) \right]^2 + \left[ \frac{a_0}{2} + \sum_{k=1}^n \cos\left(\frac{kt\pi}{T}\right) \right]^3 dt \\
 & + \int_1^2 \left[ \sum_{k=1}^n \frac{-k\pi}{T} a_k \text{sen}\left(\frac{kt\pi}{T}\right) \right]^2 + \left[ \frac{a_0}{2} + \sum_{k=1}^n \cos\left(\frac{mt\pi}{T}\right) \right]^3 \\
 & + \left[ \frac{a_0}{2} + \sum_{k=1}^n \cos\left(\frac{k(t-1)\pi}{T}\right) \right] \cdot \left[ \frac{a_0}{2} + \sum_{k=1}^n \cos\left(\frac{kt\pi}{T}\right) \right] \\
 & \cdot \left[ \sum_{k=1}^n -(k\pi/T) a_k \text{sen}(k\pi/T) \right] dt
 \end{aligned}$$

con las condiciones

$$\frac{a_0}{2} + \sum_{k=1}^n a_k = 0; \quad \frac{a_0}{2} + \sum_{k=1}^n (-1)^k a_k = 0$$

Los resultados muestran que efectivamente la solución a este problema es  $X = 0$ .

#### 4 Conclusión

Se ha presentado en este artículo una generalización de la Ecuación de Euler para funcionales que tienen un retraso en la variable  $X(t)$ , resultando en general una ecuación que incluye no solo un retraso sino que también se genera un adelanto, dificultando su solución, y por lo tanto dificultando esta vía clásica de atacar un problema variacional de este tipo. Solo para casos especiales de integrandos se podría obtener una solución sin

mucha complicación, siempre y cuando la ecuación de Euler asociada incluya solo uno de los dos, adelanto ó retraso. En [6] se presenta una vía alternativa usando series de Fourier para implementar una especie de método directo para atacar estos problemas, sin tratar de resolver la ecuación de Euler.

## Bibliografía

- [1] Akhiezer, N. (1962) *The Calculus of Variations*. Blaisdell Publishing Company, New York.
- [2] Gelfand, I. M. and Fomin, S. V. (1963) *Calculus of Variations*. Prentice-Hall Inc, New Jersey.
- [3] Hughes, D. (1968) "Variational and optimal control problems with delayed argument", *Journal of Optimization Theory and Applications* 2(1): 1-15.
- [4] Krasnov, Makarenko y Kiseliiov (1974) *Cálculo Variacional*. Stanford University, Dover Publications, New York.
- [5] Rosenblueth, J.F (1988) "Systems with time delay in the calculus of variations: a variational approach", *IMA Journal of Mathematical Control and Information* 5: 125-145.
- [6] Salazar, L. (1994) *Using GAMS to solve Optimization Problems with time Delay*. Tesis de Maestría, The University of Toledo, Ohio, Estados Unidos.
- [7] Salazar, L. (1995) "Series de Fourier aplicadas a problemas de cálculo de variaciones con retardo", *Revista de Matemática: Teoría y Aplicaciones* 2(1): 57-68.
- [8] Weinstock, R. (1974) *Calculus of Variations with Applications to Physics and Engineering*. Stanford University, Dover Publications, New York.

# ENFOQUE COMPARTIMENTAL PARA CONTAR PARTICIONES GENERALIZADAS DE NÚMEROS ENTEROS POSITIVOS

OSVALDO SKLIAR<sup>1</sup> - GUILLERMO OVIEDO<sup>1</sup> - VÍCTOR MEDINA<sup>1</sup>

## Resumen

Se caracterizan, para números enteros positivos, las nociones de particiones generalizadas, tanto restringidas como no restringidas. Se presenta un enfoque compartimental a partir del que se obtienen fórmulas para computar los números de particiones existentes -de los dos tipos mencionados- de cualquier número entero positivo. Con la terminología utilizada, las particiones clásicas se obtienen como casos particulares de las generalizadas restringidas igualando el número de compartimentos considerado al número entero positivo cuyo número de particiones se busca computar. Se aportan, en forma de dos figuras y una tabla, algunos cálculos efectuados mediante el uso de un programa de computadora en C++, basado en las fórmulas mencionadas.

Se resuelve, utilizando el mismo enfoque compartimental al que se hizo referencia, un problema estrechamente vinculado a los previamente considerados: el de determinar los números de distribuciones -tanto restringidas como no restringidas- de  $N$  partículas distinguibles en  $M$  compartimentos.

**Palabras clave:** particiones de números enteros, enfoque compartimental.

## 1 Introducción

Se entiende por partición de un entero positivo  $N$ , en el sentido clásico, a un arreglo  $(X_1, X_2, \dots, X_N)$  de enteros que satisfacen:

$$X_1 + X_2 + \dots + X_N = N, \quad (1)$$

$$X_1 \geq X_2 \geq \dots \geq X_N \geq 0. \quad (2)$$

Las particiones de  $N = 5$ , por ejemplo, son:  $(5, 0, 0, 0, 0)$ ,  $(4, 1, 0, 0, 0)$ ,  $(3, 2, 0, 0, 0)$ ,  $(3, 1, 1, 0, 0)$ ,  $(2, 2, 1, 0, 0)$ ,  $(2, 1, 1, 1, 0)$  y  $(1, 1, 1, 1, 1)$ . Puede afirmarse, entonces, que existen 7 particiones del número 5.

<sup>1</sup>UNIVERSIDAD NACIONAL, HEREDIA, COSTA RICA

La notación usual para el número de particiones de un entero positivo  $N$  es  $p(N)$ . Esta función se denomina en la literatura especializada *función de partición* y ha sido estudiada en profundidad. ([1], [2] pp.29-43, [3], pp.166-193, [4] pp.216-227) y [5]. En este trabajo, sin embargo, se utilizará para dicha función la siguiente notación:  $n_p(N)$ . La razón de este cambio irrelevante es la siguiente: se desea que el lector, al encontrarse con el símbolo de la función en cuestión, evoque de inmediato la expresión "número de particiones de  $N$ ". (Como es obvio, "n" trae a la mente la palabra "número", el subíndice "p" recuerda la palabra "particiones" y el símbolo "N" especifica el entero positivo a cuyo número de particiones se hace referencia.) Al utilizar la notación que se acaba de introducir, puede expresarse, p. ej.,  $n_p(5) = 7$ .

En este trabajo se generalizará, de diversas maneras, la clásica noción de partición de un número entero y se especificará una forma de contar el número de cada uno de los tipos de particiones a que se hará referencia.

Puede pensarse, p. ej., en cambiar la ecuación (1) por la siguiente:

$$X_1 + X_2 + \dots + X_M = N$$

$N$  -la suma- puede ser diferente del número entero positivo  $M$  -el número de sumandos-. (Sobre estas particiones puede consultarse [1], [2] y [4]). Desde luego, si se hace  $M = N$ , se obtiene (1).

Puede caracterizarse, pues, la noción de una "partición generalizada" de  $N^2$  como un arreglo  $(X_1, X_2, \dots, X_M)$  de enteros que satisfacen:

$$X_1 + X_2 + \dots + X_M = N \quad (3)$$

$$X_1 \geq X_2 \geq \dots \geq X_M \geq 0 \quad (4)$$

Sean, por ejemplo,  $N = 5$  y  $M = 3$ . ¿Cuáles son las particiones generalizadas correspondientes? Estas:  $(5, 0, 0)$ ,  $(4, 1, 0)$ ,  $(3, 2, 0)$ ,  $(3, 1, 1)$  y  $(2, 2, 1)$ .

El número de particiones generalizadas de  $N$  será una función que dependerá no sólo de  $N$  sino también de  $M$ . Para hacer referencia a dicho número se utilizará la siguiente notación:  $n_{pg}(N, M)$ .

Para el ejemplo considerado ( $N = 5$  y  $M = 3$ ) se tiene, pues,  $n_{pg}(5, 3) = 5$ . Si  $N = 4$  y  $M = 2$  -para brindar otro ejemplo- las correspondientes particiones generalizadas son  $(4, 0)$ ,  $(3, 1)$  y  $(2, 2)$ . De modo que  $n_{pg}(4, 2) = 3$ .

El principal objetivo de este trabajo es brindar un enfoque, de tipo compartimental, que posibilite contar fácilmente el tipo especificado de particiones generalizadas de números enteros positivos. Recuérdese, sin embargo, que si se hace  $N = M$  se tiene el caso de la clásica partición de un número entero positivo. De modo que el enfoque presentado también es válido para el caso usual. Expresado de otra forma:  $n_{pg}(N, N) = n_p(N)$ .

Un objetivo adicional es considerar otros tipos de particiones generalizadas de números enteros positivos y especificar -utilizando el mencionado enfoque compartimental- las correspondientes formas de contarlas.

<sup>2</sup>También llamada "partición de  $N$  en  $M$  partes" ([3] p.29).



## 2 El enfoque compartimental que se propone

Considérese nuevamente la siguiente ecuación:

$$X_1 + X_2 + \dots + X_M = N$$

A cada uno de los sumandos  $X_i$  se le hará corresponder un compartimento  $C_i$ . (Puede pensarse en los compartimentos como recipientes en los cuales es factible colocar ciertos elementos como, por ejemplo, partículas.) A cada una de las unidades simples se le hará corresponder una partícula. Admítase que las partículas son indistinguibles entre ellas. Para una partición cualquiera,  $X_i$  será el número de partículas contenidas en  $C_i$ .  $N$  se interpretará como el número total de partículas distribuidas entre los  $M$  compartimentos considerados.

Como previamente se especificó, una de las particiones generalizadas para  $N = 5$  y  $M = 3$  es la siguiente:  $(3, 2, 0)$ . En la **Figura 1** se esquematiza la forma de interpretar, mediante el enfoque compartimental, dicha partición.



Representación compartimental de la partición  $(3, 2, 0)$ .

**Figura 1**

En el primer compartimento ( $C_1$ ) se han incluido 3 partículas (simbolizadas mediante 3 "puntos"). El segundo compartimento ( $C_2$ ) contiene 2 partículas. El tercer compartimento no contiene partículas. (O, en otras palabras,  $C_3$  contiene 0 partículas.)

En la **Figura 2** se ha representado, adoptando el enfoque compartimental, una de las particiones en el sentido clásico —o sea, tales que  $N = M$ — del número 5. Se trata de ésta:  $(2, 2, 1, 0, 0)$ .



Representación compartimental de la partición  $(2, 2, 1, 0, 0)$ .

**Figura 2**

### 3 Una forma de contar las particiones generalizadas previamente caracterizadas

Se definirá, ante todo, para cada número real  $x$ , la función  $[x]^*$ .

$$[x]^* = \begin{cases} x & \text{si } x \text{ es un número entero} \\ [x] + 1 & \text{en otro caso} \end{cases}$$

siendo  $[x]$  la "parte entera" de  $x$ .

Dados los números reales  $x_1$  y  $x_2$ , la expresión  $\min\{x_1, x_2\}$  hará referencia al mínimo del conjunto formado por esos números.

A continuación se introducirá la expresión que se propone para contar el número de particiones generalizadas previamente consideradas. (Para las cuales, recuérdese debe respetarse la condición (4).)

$$\begin{aligned} n_{pg}(N, M) = & \sum_{x_1 = \left[\frac{N}{M}\right]^*}^N \min\{x_1, N-x_1\} \sum_{x_2 = \left[\frac{N-x_1}{M-1}\right]^*}^{\min\{x_2, N-x_1-x_2\}} \min\{x_2, N-x_1-x_2\} \dots \\ & \sum_{x_{M-1} = \left[\frac{N-x_1-x_2-\dots-x_{M-2}}{2}\right]^*}^{\min\{x_{M-2}, N-x_1-x_2-\dots-x_{M-2}\}} 1 \end{aligned} \quad (5)$$

Tres ejemplos de la precedente ecuación son los siguientes:

#### Ejemplo 1

$$n_{pg}(10, 3) = \sum_{x_1=4}^{10} \min\{x_1, 10-x_1\} \sum_{x_2=\left[\frac{10-x_1}{2}\right]^*} 1$$

#### Ejemplo 2

$$\begin{aligned} n_{pg}(137, 48) = & \sum_{x_1=3}^{137} \min\{x_1, 137-x_1\} \sum_{x_2=\left[\frac{137-x_1}{47}\right]^*}^{\min\{x_2, 137-x_1-x_2\}} \sum_{x_3=\left[\frac{137-x_1-x_2}{46}\right]^*} \dots \end{aligned}$$

$$\sum_{x_{47} = \left[ \frac{137 - x_1 - x_2 - \dots - x_{46}}{2} \right]^*}^{\min\{x_{46}, 137 - x_1 - x_2 - \dots - x_{46}\}} 1$$

**Ejemplo 3**

$$n_{pg}(5, 5) = n_p(5) =$$

$$\sum_{x_1=1}^5 \sum_{x_2 = \left[ \frac{5-x_1}{4} \right]^*}^{\min\{x_1, 5-x_1\}} \sum_{x_3 = \left[ \frac{5-x_1-x_2}{3} \right]^*}^{\min\{x_2, 5-x_1-x_2\}} \sum_{x_4 = \left[ \frac{5-x_1-x_2-x_3}{2} \right]^*}^{\min\{x_3, 5-x_1-x_2-x_3\}} 1$$

Serán justificadas, a continuación, las expresiones dadas para los límites inferiores y superiores de las sumatorias de la fórmula (5).

**1. Justificación de las expresiones dadas para los límites inferiores de las mencionadas sumatorias.**

Hay dos casos a considerar: (I) el caso en que  $N$  es divisible por  $M$  y (II) el caso en que  $N$  no es divisible por  $M$ .

**Caso (I)** Considérese, en primer lugar, que  $N$  es divisible por  $M$ . Es fácil demostrar, entonces, que el valor mínimo que puede asumir  $x_1$  es, precisamente,  $\frac{N}{M}$ , que para el caso que se está analizando coincide con  $\left[ \frac{N}{M} \right]^*$ . En efecto; si asumiese un valor menor, o sea que  $x_1 \leq \frac{N}{M} - 1$ , entonces por la condición restrictiva (4),  $x_i \leq \frac{N}{M} - 1$  para  $1 \leq i \leq M$  y por lo tanto  $X_1 + X_2 + \dots + X_M \leq M \left( \frac{N}{M} - 1 \right) = N - M < N$ , no cumpliéndose la condición restrictiva (3). Puesto que este resultado es absurdo, debe rechazarse la posibilidad de que  $x_1$  pueda tomar valores menores que  $\frac{N}{M}$ . En cambio, sí resulta admisible que  $x_1$  asuma el valor mínimo  $\frac{N}{M}$ . Por la restricción (4) cada uno de los restantes compartimentos podrá contener, a lo sumo,  $\frac{N}{M}$  partículas. Por supuesto, si  $x_1 = \frac{N}{M}$ , la condición (3) implica que cada uno de los restantes  $M - 1$  compartimentos deberá contener también, exactamente  $\frac{N}{M}$  partículas.

**Caso (II)** Considérese ahora el caso en el cual  $N$  no es divisible por  $M$ . En este caso  $\frac{N}{M} = c + d$ , donde  $c = \left[ \frac{N}{M} \right]$  y  $d$  (la "parte fraccionaria" de  $N/M$ )

es un número para el cual se cumple  $1 > d > 0$ . En este caso, el menor valor que puede asumir  $x_1$  es  $\left[\frac{N}{M}\right]^* = c + 1$ . En efecto; si asumiese un valor menor, entonces  $x_1 \leq c$  y por la condición restrictiva (4) ninguno de los compartimentos restantes podría tener un número de partículas superior a  $c$ ; por lo tanto  $X_1 + X_2 + \dots + X_M \leq M \cdot c < M \cdot c + M \cdot d = N$ . Pero esto último es absurdo, pues para que  $(X_1, X_2, \dots, X_M)$  sea una partición generalizada de  $N$  tiene que cumplirse la condición (3). Por lo tanto, debe rechazarse la posibilidad de que  $X_1$  asuma como valor mínimo  $c$  u otro menor y debe fijarse dicho valor mínimo para  $X_1$  así:  $X_1 = c + 1 = \left[\frac{N}{M}\right]^*$ .

Una vez comprendidas las razones por las cuales se fija  $X_1 = \left[\frac{N}{M}\right]^*$  como límite inferior de la primera sumatoria de la fórmula (5) —ordenando las sumatorias de izquierda a derecha— resulta obvio por qué se fijan como se ha hecho los límites inferiores de las otras sumatorias. Considérese, p. ej., la segunda sumatoria. Una vez que se asigna un número  $X_1$  de partículas al compartimento  $C_1$ , quedan disponibles  $N - X_1$  partículas para distribuir entre los restantes  $M - 1$  compartimentos, respetando las condiciones:

$$\begin{aligned} X_2 + X_3 + \dots + X_M &= N - x_1 \\ X_2 \geq X_3 &\geq \dots \geq X_M \geq 0 \end{aligned}$$

Si se aplica a este nuevo problema el razonamiento hecho anteriormente, la primera sumatoria debe tener como límite inferior  $X_2 = \left[\frac{N - x_1}{M - 1}\right]^*$ . Pero, en realidad, esta primera sumatoria del nuevo problema es la segunda sumatoria del problema original. Y el mismo tipo de razonamiento puede aplicarse sucesivamente a los límites inferiores de las restantes sumatorias de la fórmula (5). ¿Por qué la fórmula (5) incluye  $M - 1$  sumatorias? Porque se tomó  $M$  como igual al número de compartimentos. Fijados los valores  $X_1, X_2, \dots, X_{M-1}$  de partículas contenidas, respectivamente, en los compartimentos  $C_1, C_2, \dots, C_{M-1}$ , el número de partículas contenidas en el compartimento  $C_M$  resulta determinado sin ambigüedad. Dicho número es  $X_M = N - x_1 - x_2 - \dots - x_{M-1}$ .

## 2. Justificación de las expresiones dadas para los límites superiores de las mencionadas sumatorias.

No hay obstáculo para que un número de partículas superior al límite inferior de la primera sumatoria se encuentre contenido en  $C_1$ . De hecho la totalidad de las partículas ( $N$ ) pueden estar contenidas en  $C_1$ . Por ello el límite superior de la primera sumatoria es igual a  $N$ .

Por (3) y (4) el número de partículas contenidas en  $C_2$  no puede ser mayor que  $x_1$  y tampoco puede ser mayor que  $N - x_1$ , que es el número de partículas disponibles para ser distribuidas entre los compartimentos que quedan una vez que se excluye

el  $C_1$ . Por consiguiente, el límite superior de la segunda sumatoria de la fórmula (5) debe ser  $\min\{x_1, N - x_1\}$ . De manera análoga resulta sencillo fijar los límites superiores de las restantes sumatorias de la fórmula (5).

### 4 Algunos resultados computacionales

A continuación se aportan, en forma de dos figuras y una tabla, algunos resultados computacionales obtenidos mediante un programa para computadora elaborado en C++, basado en las fórmulas previamente especificadas.

Por una razón que resultará justificada al leerse el apartado 5 de este trabajo, las que hasta aquí se llamaron "particiones generalizadas", son denominadas -en las mencionadas figuras y en la tabla- "particiones generalizadas restringidas".

Números de Particiones Generalizadas Restringidas							
$N$	10	20	30	40	50	60	70
$M$							
10	42	530	3590	16928	62740	195491	533975
30	42	627	5604	37241	202139	943442	3910071
50	42	627	5604	37338	204226	966370	4085881
70	42	627	5604	37338	204226	966467	4087968
90	42	627	5604	37338	204226	966467	4087968
100	42	627	5604	37338	204226	966467	4087968

### 5 Particiones generalizadas no restringidas de números enteros positivos

¿Qué sucede si se cumple la ecuación (3) pero se elimina la restricción (4)? Ello da lugar a otro tipo de particiones generalizadas de  $N$  que podemos calificar de no restringidas (pgr). Para ellas se obtiene:

$$n_{pgr}(N, M) = \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} \sum_{x_3=0}^{N-x_1-x_2} \dots \sum_{x_{M-1}=0}^{N-x_1-x_2-\dots-x_{M-2}} 1$$

### 6 Distribuciones en $M$ compartimentos de $N$ partículas distinguibles entre ellas

Resulta de interés considerar la posibilidad de que las partículas sean distinguibles entre ellas. Hasta ahora se había considerado dichas partículas como indistinguibles entre ellas. Así, p. ej., si dada la partición esquematizada en la **Figura 2**, se toma una de las partículas de  $C_2$  y se la coloca en  $C_3$  y también se traslada la partícula de  $C_3$  a  $C_2$ , se obtiene la

misma partición. Ello no ocurriría si las partículas pudiesen distinguirse entre ellas. Desde luego, si éste es el caso, el enfoque utilizado ya no resulta válido para estudiar particiones de números enteros. Puede, en cambio, en la situación descripta, hacerse referencia a distribuciones en  $M$  compartimentos de  $N$  elementos -p. ej., partículas- distinguibles entre ellos.

Las distribuciones que se están considerando serán calificadas de restringidas (dr) si se acepta como válida la condición restrictiva (4) o bien no restringidas (dnr) si no se impone dicha condición restrictiva.

Mediante la fórmula (6) será expresado el número de distribuciones del tipo que se considera si es válida la condición restrictiva (4). Y mediante la fórmula (7) se expresará el número de distribuciones resultante si no se impone esa condición restrictiva.

$$\begin{aligned}
 n_{dr}(N, M) = & \sum_{x_1=\lfloor \frac{N}{M} \rfloor}^N \sum_{x_2=\lfloor \frac{N-x_1}{M-1} \rfloor}^{\min\{x_1, N-x_1\}} \sum_{x_3=\lfloor \frac{N-x_1-x_2}{M-2} \rfloor}^{\min\{x_2, N-x_1-x_2\}} \dots \\
 & \sum_{x_{M-1}=\lfloor \frac{N-x_1-x_2-\dots-x_{M-2}}{2} \rfloor}^{\min\{x_{M-2}, N-x_1-x_2-\dots-x_{M-2}\}} \binom{N}{x_1} \binom{N-x_1}{x_2} \dots \\
 & \binom{N-x_1-\dots-x_{M-2}}{x_{M-1}} \quad (6)
 \end{aligned}$$

$$\begin{aligned}
 n_{dnr}(N, M) = & \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} \dots \sum_{x_{M-1}=0}^{N-x_1-x_2-\dots-x_{M-2}} \binom{N}{x_1} \binom{N-x_1}{x_2} \dots \\
 & \binom{N-x_1-\dots-x_{M-2}}{x_{M-1}} \quad (7)
 \end{aligned}$$

Mediante el coeficiente binomial  $\binom{N}{x_1}$  se expresa, pues, el número de maneras diferentes de seleccionar  $x_1$  partículas -las que ocupan  $C_1$ - del total de  $N$ ; mediante el coeficiente binomial  $\binom{N-x_1}{x_2}$  se expresa el número de maneras diferentes de seleccionar  $x_2$  partículas -las que ocupan  $C_2$ - del número de partículas disponibles ( $N-x_1$ ) una vez que del total de ellas ( $N$ ) se resta  $x_1$ , etc.

Nótese que  $n_{dnr}$  también puede expresarse así:  $n_{dnr} = M^N$ . En efecto; para cada una de las  $N$  partículas hay  $M$  posibilidades: las de ubicarse en uno cualquiera de los  $M$  compartimentos. De donde se deduce la siguiente interesante igualdad del cálculo combinatorio:

$$M^N = \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} \cdots \sum_{x_{M-1}=0}^{N-x_1-x_2-\cdots-x_{M-2}} \binom{N}{x_1} \binom{N-x_1}{x_2} \cdots \binom{N-x_1-\cdots-x_{M-2}}{x_{M-1}}$$

## 7 Conclusiones

Se obtuvo una fórmula —la (5)— para calcular, de manera exacta, el número  $n_{pg}(N, M)$  de particiones generalizadas del entero  $N > 0$ . Si  $N = M$  se obtiene el número  $n_p$  de particiones (clásicas) de dicho entero  $N$ . (En la literatura especializada esta última función aritmética es usualmente denominada  $p(N)$ .) También se aportan expresiones para resolver otros problemas afines al que acaba de especificarse.

Posiblemente el rasgo de mayor interés de este trabajo es el siguiente: las diversas fórmulas presentadas se obtuvieron mediante consideraciones matemáticas sencillas a partir de un mismo enfoque compartimental.

A pesar de la simplicidad de los recursos matemáticos utilizados —que no van más allá del cálculo combinatorio elemental— los resultados obtenidos permitirán abordar —como se mostrará en ulteriores trabajos— problemas de gran interés en mecánica estadística, como los concernientes al análisis detallado de las fluctuaciones.

## Bibliografía

- [1] Andrews, George E. (1976) *The Theory of Partitions*, Addison-Wesley Publ. Co., Reading, Massachusetts.
- [2] Chandrasekharan, K. (1970) *Arithmetical Functions*. Springer-Verlag, New York.
- [3] Hall, Marshall Jr. (1967) *Combinatorial Theory*. Blaisdell Publ. Co., Waltham, Massachusetts.
- [4] Rademacher, Hans (1973) *Topics in Analytic Number Theory*. Springer-Verlag, New York.
- [5] Boza, J.; Murillo, A. (1995) "Una nueva manera de contar particiones", *IX Simposio Métodos Matemáticos Aplicados a las Ciencias*, Javier Trejos (ed.), Universidad de Costa Rica-Instituto Tecnológico de Costa Rica, Turrialba.

## NUEVO ENFOQUE COMPARTIMENTAL PARA LA SOLUCIÓN DEL PROBLEMA DEL ISOMORFISMO DE GRAFOS

OSVALDO SKLIAR<sup>1</sup> - GUILLERMO OVIEDO<sup>1</sup> - VÍCTOR MEDINA<sup>1</sup>

---

Previamente se presentó un enfoque compartimental para determinar si dos grafos cualesquiera son o no isomorfos.<sup>2</sup> Dicho enfoque utiliza la simulación de procesos de difusión entre compartimentos asociados a los vértices de ambos grafos. En este trabajo se presenta un nuevo enfoque que sustituye la simulación de procesos de difusión por la de procesos de "antidifusión". Estos tienen la ventaja de explotar mejor que los previamente mencionados cualquier asimetría de los grafos considerados. Aunque, en esta oportunidad, solo se trata el caso de los grafos dirigidos, resulta factible generalizar el nuevo enfoque -como se hará en ulteriores trabajos- para cualquier tipo de grafos.

**Palabras clave:** isomorfismo de grafos, enfoque compartimental, antidifusión.

---

<sup>1</sup>UNIVERSIDAD NACIONAL, HEREDIA, COSTA RICA

<sup>2</sup>Skliar O., T. Láscaris Comneno y V. Medina. 1994. *Enfoque compartimental para la solución del problema del isomorfismo de grafos*. Editor: G. Mora. Memorias II Encuentro Centroamericano de Investigadores en Matemática. I Parte. San Ramón. pp. 113-137.



## ESTRATEGIAS COMPUTACIONALES APLICADAS A LA EVALUACIÓN ANIMAL

YASNAY TEIJEIRO<sup>1</sup> - REINALDO LÓPEZ<sup>1</sup>

---

En el mejoramiento animal, la selección de los animales de forma eficiente garantiza el mejoramiento genético de las nuevas generaciones. Es por ello que se investigan los modelos matemáticos y las estrategias de computación que permiten resolver el problema de manera más eficiente. Es conocido que los problemas de selección, estimación del valor genético, estimación de los componentes de varianza, etc generan sistemas de ecuaciones enormes, tamaño que se incrementa cuando se añaden las relaciones entre los animales. Por todo esto se realizó un estudio de los modelos matemáticos y estrategias computacionales más usadas en los últimos años para encontrar las que ofrezcan un uso óptimo de la memoria y una máxima velocidad de convergencia.

Se presentarán una serie de modelos matemáticos, en los que los rasgos genotípicos tienen que ser evaluados a través de mediciones u observaciones fenotípicas en los animales y éstas están influenciadas por los efectos ambientales y genéticos. Lo que nos interesa es estimar la parte genotípica de las observaciones. Entre los modelos que se presentarán están el modelo semental, el modelo semental más abuelo materno y el modelo animal. También se verán métodos y algoritmos para la estimación de los parámetros, así como algunas consideraciones para disminuir los requerimientos de memoria y para acelerar la velocidad de convergencia.

---

<sup>1</sup>DEPARTAMENTO DE MATEMÁTICA, UNIVERSIDAD LATINA DE COSTA RICA, LOURDES DE MONTES DE OCA, COSTA RICA

## ANÁLISIS DISCRIMINANTE APLICADO A UN ESTUDIO TAXONÓMICO DE SCIRPUS SECCIÓN MONOCEPHALES

MARTA TORRES DE PLAZA<sup>1</sup> - LIDIA BENÍTEZ DE PARRA<sup>1</sup>  
GRACIELA PONESSA DE MERCADO<sup>2</sup>

---

### Introducción

**Scirpus** Linné, es un género cosmopolita de 250 especies a nivel mundial (Koydma, 1958). **Scirpus**, en Argentina, está representado por 18 especies y en Tucumán por 10, de las cuales 4 pertenecen a la sección **Monocephales**: **S.acaulis** Phil., **S.chubutensis** C.B.Clarke, **S.desertícola** Phil. y **S.macrolepis** Phil.

En Pflanzenfamilien (Pax, 1888) presenta al género dividido en 2 subgéneros teniendo en cuenta la presencia o ausencia de las flores con setas hipóginas; **Eu.Scirpus** e **Isolepis**, las especies citadas para Tucumán están incluidas en el primer grupo.

El género está organizado en 17 secciones y 7 grupos (Koyama, 1958). El grupo 5, **Scirpus** Sección **Micranthi**, al cual incorporamos las especies en estudio, se caracteriza por presentar plantas pequeñas de hábitat gramíneo, inflorescencia en cabezuela simple, con flores perfectas.

El objetivo del presente trabajo es el estudio taxonómico de la sección **Monocephales**. Se trata de delimitar las 4 especies, realizando la tipificación y estudios de morfología externa e interna a fin de describirlas, ilustrarlas y proporcionar claves para su posterior reconocimiento. Esto se debe a que **Scirpus** es un género heterogéneo del cual se segregaron recientemente **Schoenoplectus**, **Bolboschoenus** e **Isolepis** teniendo en consideración el tipo de embrión y posición de la inflorescencia respecto al tallo.

---

<sup>1</sup>UNIVERSIDAD NACIONAL DE TUCUMÁN, FACULTAD DE CIENCIAS NATURALES E INSTITUTO MIGUEL LILLO, CÁTEDRA DE MATEMÁTICA, MIGUEL LILLO 205, (4000) SAN MIGUEL DE TUCUMÁN, REPÚBLICA ARGENTINA

<sup>2</sup>FUNDACIÓN MIGUEL LILLO, MIGUEL LILLO 205, (4000) SAN MIGUEL DE TUCUMÁN, REPÚBLICA ARGENTINA

## 1 Materiales y métodos

Para la revisión taxonómica se han seguido los métodos clásicos consignándose para las descripciones rasgos de morfología externa e interna a nivel de microscopio óptico (M.O.), microscopio electrónico de barrido (M.E.B.) y de transmisión (M.E.T.).

Se consultaron materiales depositados en herbarios del país: Instituto de Botánica Darwinión (SI), Museo de Ciencias Naturales de La Plata (LP) y Fundación Miguel Lillo (Lil.).

Se presenta una descripción del género y de la sección. El estudio anatómico para M.O. se realizó en material fresco y de herbario, fijados en F.A.A. (formol alcohol - ácido acético), se realizaron cortes a mano alzada y a través del proceso de inclusión en parafina (Johansen, 1940) y doble coloración safranina-fast green (Diseño de Stritmatter, 1979).

Los esquemas se realizaron con el auxilio de una cámara clara y se los dibujó respetando la simbología de Metcalfe (1972).

Las muestras para el estudio de ultraestructura se procesaron en el Laboratorio de Microscopía Electrónica del Noroeste Argentino (LAMENOA).

El material se fijó en glutaraldehído con buffer fosfato. Se deshidrató para su posterior inclusión en resina epoxi en el caso de ser observados al M.E.T. Los aquenios que fueron procesados para ser observados al M.E.B. obviando punto crítico, se adhirieron a una cinta engomada y fueron metalizados con oro paladio en un evaporador al vacío (Schuyler, 1971).

El análisis discriminante (AD) se realizó en base al estudio de doce caracteres, escogiendo al azar quince ejemplares de cada especie. Los caracteres o variables consideradas fueron: número de hojas (NOH), porte de la planta (PORTE), longitud de la lámina foliar (LAMFLONG), latitud lámina foliar (LAMFLAT), longitud de vaina (VAILONG), latitud de vaina (VAILAT), longitud de gluma (GLULONG), latitud de gluma (GLULAT), longitud de aquenio (AQLONG), latitud de aquenio (AQLAT), longitud de estilo (ESTLONG), longitud estomática (ESTOMA). Se aplicó el AD usando el paquete estadístico "BMD Biomedical Computer Programs" (BMDP-Universidad de California) y del mismo programa 7M correspondiente al "Stepwise Discriminant Analysis", a fin de observar cuales de ellas poseían valor discriminatorio desde el punto de vista estadístico.

Con los datos de las muestras se comprobó que las variables respondían a una distribución aproximadamente normal, de acuerdo con el test de asimetría y curtosis. Además, se analizó la matriz de covarianza para cada muestra, resultando las mismas aproximadamente homogéneas.

## 2 Análisis discriminante

De las doce variables medidas, cinco fueron seleccionadas por su eficiencia discriminatoria, elegida por el método "Stepwise" (por pasos), siendo estas según el orden dado por el programa: PORTE, LAMFLONG, ESTILO, VAILONG, y NOH.

El orden establecido se debe a que en primer lugar realiza un ANOVA con un criterio de clasificación, de donde se obtienen valores para el estadístico F que determinarán de acuerdo a su valor el ingreso de la primera variable (PORTE) a la función discriminante,

o función de clasificación (combinación lineal de las variables que mejor caracterizan las diferencias de las especies).

En los pasos siguientes, el valor de  $F$  es computado a partir del análisis de Varianzas y Covarianzas, determinando éste el ingreso de las siguientes variables.

Se plantea como hipótesis nula, que las especies no presentan diferencias entre ellas. Para aceptar o rechazar ésta hipótesis se recurre al estadístico  $U$  (Lambda de Wilks) y en base a éste se analiza el estadístico  $V = m * \ln U$ , que se distribuye como  $\chi^2$  con  $pq$  gl. ( $p = n^\circ$  de variables seleccionadas y  $q = n^\circ$  de muestras - 1). El estadístico  $U$  dió igual a 0.0174350 con 5.3 y 53 gl.,  $m = 54.5$  por lo que al analizar el estadístico  $V = -54.5 * \ln(0.0174350) = 220.6855$  frente al valor de tabla de  $\chi^2$  15 gl = 32.8 se concluye: rechazar la hipótesis nula a un nivel de significación  $\alpha = 0.005$ , es decir que las diferencias entre las especies son existen.

Las funciones de clasificación obtenidas son:

Variable	Grupo			
	SA	SD	SM	SCh
NOH	0.2572	0.7314	0.2252	0.4548
PORTE	1.9544	3.6910	3.5381	7.8746
LAMFLONG	1.1485	0.3128	1.1798	-0.3023
VAI LONG	6.2090	4.4411	4.1360	2.9314
ESTILO	11.2604	4.0204	7.1596	3.1624
Constante	-63.3238	-33.6083	-40.3358	-45.6454

Luego, mediante la función de clasificación, cada caso es asignado a un grupo para el cual el valor de la probabilidad posterior es máximo. El análisis fue satisfactorio para las cuatro especies con un porcentaje de 100%, 86.7%, 86.7%, 100% respectivamente, de clasificados correctamente, no habiendo contado con una muestra de entrenamiento se valida haciendo uso del método "Jackknife".

La figura 1 muestra lo dicho anteriormente, donde el eje horizontal es la dirección donde las especies exhiben la máxima separación (variable canónica 1) y el eje vertical muestra la máxima separación en una dirección ortogonal al eje horizontal (variable canónica 2). Los valores sobre las líneas indican la distancia de Mahalanobis entre los centros.

### Estudio agregando una nueva variable

Se analizó el comportamiento de las especies *S. deserticola*, *S. macrolepis* y *S. chubutensis* agregando una nueva variable: escape (SCAPO). No se tuvo en cuenta la cuarta especie por poseer un brevísimo escape, despreciable desde el punto de vista de la cuantificación. Con este análisis se confirma la diferencia entre especies, rechazando la hipótesis nula a un nivel de significación de 0.005 (Gráfico N° 2.)

El estadístico  $U$  dió igual a 0.0174350 con 5.3 y 53 gl,  $m = 54.5$ , por lo que al analizar el estadístico  $V = -54.5 * \ln(0.0174350) = 220.6855$  frente al valor de tabla de  $\chi^2$  15 gl = 32.8 se concluye rechazar la hipótesis nula a un nivel de significación  $\alpha = 0.005$ , es decir que las diferencias entre las especies existen.

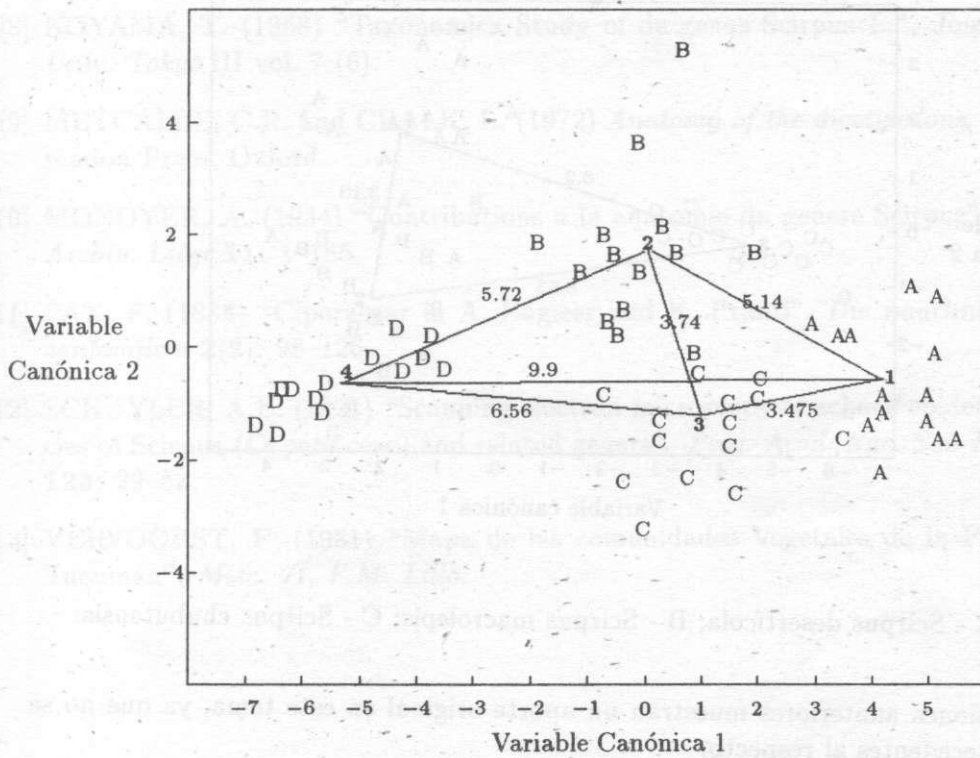


Figura 1: Plano principal del primer Análisis Discriminante; A - *Scirpus Acaulis*; B - *Scirpus desertícola*; C - *Scirpus macrolepis*; D - *Scirpus chubutensis*.

Luego, mediante la función de clasificación cada caso es asignado dentro de un grupo para el cual el valor de la probabilidad posterior es máximo. El análisis fue satisfactorio para las tres especies con un porcentaje de 100%, 86.7% y 100% respectivamente, de clasificados correctamente.

En la figura 2 se observa, al igual que en la figura 1, las diferentes especies en un espacio canónico bidimensional.

### 3 Discusión y conclusiones

- A través del AD se pudo establecer que las Especies son diferentes y que los caracteres seleccionados con mayor poder discriminante, son los que morfológicamente tienen peso taxonómico a nivel de Especie.
- El AD ha permitido ver cuales son los caracteres de valor diagnóstico que se pueden usar para realizar la clave de reconocimiento de las especies.
- La naturaleza de las papilas de las ramas estigmáticas queda desechado como carácter de valor diagnóstico debido al resultado dado por el AD.

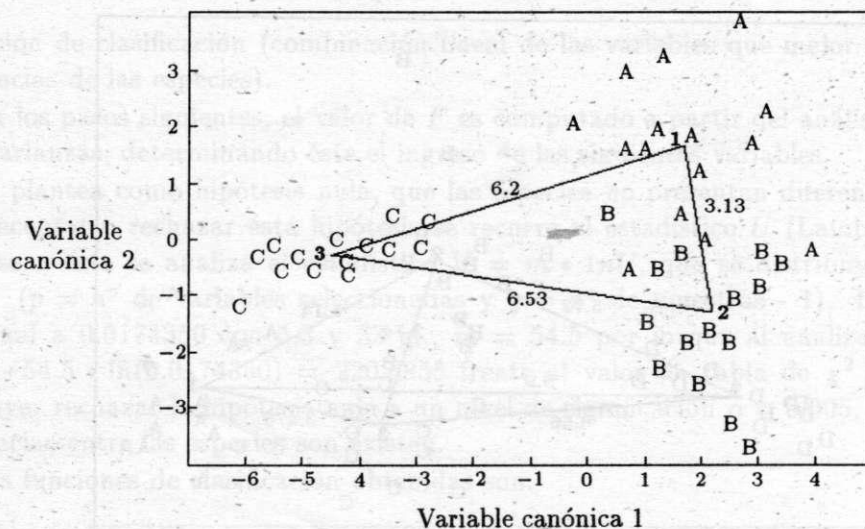


Figura 2: A - *Scirpus desertícola*; B - *Scirpus macrolepis*; C - *Scirpus chubutensis*.

Las conclusiones anteriores muestran un aporte original en este tema, ya que no se encuentran antecedentes al respecto.

Se pudo establecer que el AD puede ser usado como herramienta de trabajo para dilucidar problemas de naturaleza taxonómica.

## Bibliografía

- [1] BARROS, E. (1943) *Genera et Species Plantarum Argentinarum IV Pars prima Cyperaceae Scripoideae*.
- [2] BEETLE, A.A. (1940) "Studies in the genus *Scirpus* L.", *American Journ. Bot.* **27**:63-64.
- [3] BLASER (1941) *Amer. Journ. Bot.* **28**: 548-551.
- [4] CLARKE, C. (1908) "New genera and species of Cyperaceae" *Kew. Bull. Add. Ser.* **8**: 1-196.
- [5] DIZEO, G. de STRITTMATER (1979) "Modificación de una coloración safraninafast green", *Boletín de la Sociedad Argentina de Botánica XVIII* (3-4): 121-122.
- [6] JOHANSEN, D.A. (1940) *Plant Microtechnique*. New York.
- [7] KENDALL, M. (1980) *Multivariate Analysis*, 2da. Edic. Charles Griffin and Company L.T.D. London and High Wycombe.

- [8] KOYAMA, T. (1958) "Taxonomica Study of de genus *Scirpus* L.", *Jour. Fac. Sci. Univ. Tokyo* III vol. 7 (6).
- [9] METCALFE, C.R. and CHALK, L. (1972) *Anatomy of the dicotyledons*, Vol. I. Clarendon Press, Oxford.
- [10] MONOYER, A. (1934) "Contributions a la anatomie du genere *Scirpus*", *Inst. Bot. Archiv. Liège* 11: 1-185.
- [11] PAX, F. (1888) "Ciperacear in A. Engleer and K. Prantl", *Die naurlinchen Pflanzenfamilien* 2(2): 98-126.
- [12] SCHUYLER, A.E. (1971) "Scanning electron microscopy of achene epidermis in species of *Scirpus* (Cyperaceae) and related genera", *Proc. Acad. Nat. Sci. Philadelphia* 123: 29-52.
- [13] VERVOORST, F. (1981) "Mapa de las comunidades Vegetales de la Provincia de Tucumán", *Misc. 71, F.M. Lillo*.

## ALGORITMO DE LA DISTRIBUCIÓN NORMAL

JUAN JOSÉ VARGAS<sup>1</sup>

---

A partir del desarrollo en series de Taylor aplicado a la función de distribución normal estándar, se propone un algoritmo para calcular esta función, el cual converge en relativamente pocas iteraciones y no requiere del uso de otras funciones adicionales. La precisión de los cálculos se controla mediante un parámetro dado en el algoritmo. Se presenta una tabla con algunos valores calculados para precisiones de 5 y 10 dígitos significativos, y se comparan las iteraciones requeridas en cada caso.

El algoritmo es el siguiente, siendo  $x$  el punto sobre el que quiere evaluarse la función y  $s$  la salida:

$s := 0$

$t := x$

$n := 1$

**while**  $|t| \geq \epsilon$  **do**

$s := s + t$

$t := \frac{t * (-x) * x * n}{(n + 1) * (n + 2)}$

$n := n + 2$

**endwhile**

$s := s / \sqrt{\pi} + 0.5$

---

<sup>1</sup>ESCUELA DE CIENCIAS DE LA COMPUTACIÓN E INFORMÁTICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA



# FACTORES DE DIFERENCIACIÓN DE LA MORTALIDAD EN LA PROVINCIA DE BUENOS AIRES EN LA DÉCADA DEL 80

GUILLERMO VELÁZQUEZ<sup>1</sup> - HERNÁN OTERO<sup>1</sup>  
CLAUDIA MARINELLI<sup>2</sup> - SEBASTIÁN TORCIDA<sup>2</sup>

---

## Resumen

A partir de las tablas de mortalidad correspondientes a cada partido de la provincia de Buenos Aires se representa la distribución de los principales indicadores de la mortalidad (tasa de mortalidad infantil y esperanza de vida al nacimiento). En un segundo paso, se realiza el análisis espacial de dicha distribución (pautas de contigüidad, agrupamientos regionales, etc) y se agrupan los partidos en unidades de análisis mayores: zonas sanitarias. Finalmente se relaciona la mortalidad bonaerense con variables vinculadas a diversas dimensiones del desarrollo: niveles de instrucción, socioeconómico y sanitario, para concluir estableciendo un balance sobre la validez explicativa de estos factores en los niveles de mortalidad observados. Desde el punto de vista estadístico, las etapas mencionadas implicaron la selección de las variables y su posterior transformación, sumado a un análisis gráfico-descriptivo a través de las técnicas de componentes principales y cluster.

## 1 Introducción

En el contexto de la producción existente sobre mortalidad en los países de la región, el presente trabajo busca satisfacer dos objetivos básicos: 1) la medición de los niveles de mortalidad de la provincia de Buenos Aires a principios de la década de 1980 a partir de un nivel de agregación espacial intermedio (zonas sanitarias) y 2) la elaboración de un marco explicativo que nos permita avanzar en la comprensión de las diferencias observadas.

Es importante destacar que la utilización del espacio como variable analítica permite acceder no sólo al conocimiento de las diferencias espaciales sino también al de las inequidades sociales y económicas asociadas a cada región. Concebir al espacio como una de

---

<sup>1</sup>PROPIEP, FACULTAD DE CIENCIAS HUMANAS, UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, TANDIL, REPÚBLICA ARGENTINA

<sup>2</sup>NUCOMPA, FACULTAD DE CIENCIAS EXACTAS, UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, REPÚBLICA ARGENTINA

las dimensiones básicas del proceso de diferenciación y estructuración social y económica de un país no sólo resulta válido en términos teóricos sino que constituye una vía de acceso obligada al estudio de las diferencias, dadas las falencias en los datos sociales que caracterizan a las estadísticas oficiales.

En la investigación social no existe una metodología experimental estricta; la producción de conocimientos se desarrolla siguiendo una espiral inductiva-hipotética-deductiva que conduce a explorar tanto como a experimentar, cuando esto es posible. Los pasos aquí desarrollados, que combinan el Análisis de Componentes Principales y el Análisis de Cluster, responden a esa exigencia metodológica, al permitirnos analizar la información proporcionada por datos multidimensionales mediante una aproximación gráfica que respeta las características originales de éstos.

## 2 Fuentes

El presente estudio se basa en tablas de mortalidad de la provincia de Buenos Aires correspondientes al período 1980-1982<sup>3</sup> (Otero, Velázquez: 1995), a partir de las cuales se obtiene la Tasa de Mortalidad Infantil (TMI) y la expectativa de vida al nacer ( $E(o)$ ) de las 11 zonas sanitarias en que se divide la Provincia de Buenos Aires.

Asimismo, para el análisis de las variables utilizadas se ha recurrido a información de diversos organismos públicos: INDEC (Censo Nacional de Población 1980), Ministerio de Salud de la PBA (Estadísticas Vitales), Ministerio de Economía de la PBA (Serie de indicadores socio-económicos).

## 3 Metodología

### 3.1 Problemas de las fuentes de información

Los principales problemas para la confección de las tablas se refieren a la calidad y cobertura de la información disponible. En lo concerniente a la información censal, la clasificación según edades no suministra información desagregada por intervalos anuales para la población menor de cinco años, lo que dificulta el cálculo de las tasas de mortalidad. En su reemplazo, hemos utilizado la media de los nacimientos correspondientes a los años 1976-1983, solución que permite evitar, por otra parte, el problema bastante frecuente en la información censal del subregistro de niños.

En el caso de las actas de defunciones, se presentan problemas en lo relativo a la adscripción espacial de la defunción ya que las actas presentan tres lugares posibles para la confección de las tablas: el lugar de nacimiento (provincia para los nativos, país para los extranjeros), el departamento de ocurrencia de la defunción y el departamento de residencia habitual. La diferencia entre estos lugares puede ser importante en función

---

<sup>3</sup>Para evitar la influencia de variaciones aleatorias, producidas por la escasa población de algunos partidos o por la existencia de años de mortalidad excepcional, se ha trabajado con el promedio de las defunciones correspondientes a los años 1980, 1981 y 1982. No se tomó el promedio pericenso (1979-1981), como hubiera sido más conveniente, porque en el año 79 cambiaron los criterios de recolección de la información en la provincia.

de las migraciones internas cuyo nivel y destino es diferencial según los distintos partidos. Como la movilidad de la población hace además que una proporción importante de muertes se produzca fuera del lugar de nacimiento, el partido de residencia habitual fue preferido a los otros dos indicadores.

Con el fin de garantizar el mayor grado de comparabilidad con las tablas elaboradas para el total del país y jurisdicciones por J. Somoza y M. Müller (INDEC, 1988) se han seguido, en la mayor parte de los casos, criterios similares a los establecidos por dichos autores. Así, por ejemplo, en lo concerniente al nivel de cobertura de las fuentes se ha seguido el criterio según el cual los registros pueden considerarse completos en la provincia de Buenos Aires, lo que implica no corregir los datos de base. Para evitar que eventuales variaciones aleatorias producidas por el número de efectivos en juego en cada zona se trasladaran a las restantes funciones de la tabla, se procedió previamente a ajustar mediante un método manual o gráfico la función de cocientes de mortalidad; al igual que en las tablas nacionales del INDEC (1988) hemos procedido a ajustes de tipo "conservador" y sólo para aquellos casos en los que la función  $q(x,n)$  presentaba un perfil manifiestamente atípico. Para la conversión de las tasas centrales de mortalidad en probabilidades para cada intervalo de edad se siguió el método de Reed y Merrell (1939)<sup>4</sup>.

## 3.2 Análisis estadístico

### 3.2.1 Diferencias espaciales de la mortalidad

Como es sabido existe una alta correlación entre la expectativa de vida al nacer y la mortalidad infantil, tanto porque esta última interviene con un alto peso relativo en el cálculo de la primera, como porque las condiciones socio-económicas afectan de modo análogo las probabilidades de morir en el primer tramo de vida y en los intervalos posteriores. El Gráfico 1 (ver Anexo) permite apreciar la fuerte asociación lineal entre ambas variables (algo esperable dado el nivel de agregación de los datos): como puede observarse, ambos indicadores varían inversamente.

Dicho gráfico induce a pensar en una tipología de la mortalidad de la provincia compuesta por tres niveles: 1) alto (zonas III, IX, X y XI), 2) medio (I, II, IV, VI y VIII) y 3) bajo (zonas V y VII), y caracterizada por una alta contigüidad espacial (ver Mapa, en Anexo), resultando evidente la menor mortalidad de los partidos del norte y del oeste del conurbano bonaerense y las peores condiciones socio-sanitarias de la Pampa Deprimida. Debe destacarse que, a pesar de tratarse de un contexto habitualmente visto como homogéneo en relación al conjunto del país, las diferencias observadas son bastante significativas: 9,5 por mil entre los valores extremos de las tasas de mortalidad infantil (36,9 por mil en la zona IX contra 27,4 en la zona VII) y 3,3 años en la expectativa de vida (69,68 en la zona XI contra 72,99 en la zona VII).

A fin de evitar una imagen distorsionada del fenómeno, debe tenerse en cuenta la distribución espacial de la población en la provincia: las regiones de mortalidad baja

<sup>4</sup>Como es sabido, la función de probabilidad propuesta por estos autores constituye el mejor grado de ajuste de la asociación existente entre la tasa y la probabilidad de mortalidad de un intervalo de edades determinado.

(zonas V y VII) y media (zona VI) representan por si mismas el 67,4% de la población total de la provincia en 1980, mientras que un significativo 32,6% de la población total del primer estado argentino se hallaba hacia 1980, en términos relativos, en niveles de mortalidad alta.

Al igual que lo observado para el total del país (Mychaszula y Acosta, 1990) se registra una sobremortalidad infantil masculina, que varía entre 12,6 defunciones por mil en la zona sanitaria XI y 6,0 en la zona II. La diferencia de expectativa al nacer según sexos varía entre 6,8 años en la zona sanitaria VI y 10,2 en la zona I.

### 3.2.2 Factores de diferenciación de zonas sanitarias según variables de desarrollo

Con el objeto de describir el comportamiento de cada zona sanitaria y su relación con las restantes, se seleccionaron las siguientes variables explicativas: defunciones infantiles (NMI), población (Pobla), producto bruto interno (PBI), camas de hospital (Camas), población con necesidades básicas insatisfechas (PNBI), población sin instrucción (Ps/I), población con estudios secundarios o terciarios (Pc/SyT), población urbana (PUr), depósitos en bancos privados (Dep) y expectativa de vida al nacer (Eo) (ver "Tabla de datos ...", en Anexo).

Dado que el rango de variación de las variables elegidas era significativamente heterogéneo, se elaboró en primera instancia la correspondiente matriz de rangos de Spearman (ver Anexo), estableciendo un ranking de las zonas con respecto a cada variable. En la misma es posible observar que la posición de una determinada zona relativa a la variable población prácticamente determina su posición respecto del resto de las variables.

Se decidió entonces eliminar del análisis dicha variable y expresar las variables originales restantes como proporciones, de manera que las nuevas variables permanecieran sensibles a las diferencias iniciales entre las zonas, pero teniendo rangos de variación comparables. Las variables transformadas resultaron:

- Indicadores de mortalidad
  - TMI: Tasa de mortalidad infantil (por mil)
  - E(o): Expectativa de vida al nacer
- Distribución espacial
  - PPUr: proporción de población urbana
- Infraestructura hospitalaria del sector público
  - PCamas: Proporción de camas de hospital
- Indicadores de desarrollo económico y social
  - Riqueza
    - \* PPBI: Producto bruto interno per cápita

- \* PDep: Depósitos bancarios privados per cápita
- Infraestructura y calidad de vida
  - \* PPNBI: Proporción de población con Necesidades Básicas Insatisfechas
- Indicadores del nivel de instrucción
  - \* PPs/I: Proporción de personas de cinco años y más que no asiste a la escuela
  - \* PPc/SyT: Proporción de población que tiene estudios secundarios o terciarios

Para representar la distribución de las zonas se realizó un Análisis de Componentes Principales (ACP) con las 9 variables transformadas. Los primeros dos autovalores de la matriz de correlación resultaron  $\lambda_1 = 6,08$ ;  $\lambda_2 = 2,26$  y el porcentaje de la variación total explicado por las dos primeras componentes fue del 81,57%. Se calcularon, para cada uno de los indicadores, los coeficientes de correlación con las dos primeras componentes principales, los que permitieron obtener el gráfico de los h-plots (Gráfico 3). A partir del mismo puede apreciarse que:

- las longitudes de los vectores, en su mayoría cercanas a uno, indican un buen porcentaje de reconstrucción de las variables en el plano de las dos primeras componentes;
- se observa la presencia de correlaciones contrarias a las esperadas, como en el caso de las variables E(o) y PPNBI, en donde la correlación es provocada por tres observaciones influyentes (Gráfico 2), mientras que en el caso TMI y PPBI no debe perderse de vista que el porcentaje de reconstrucción de TMI es inferior al 80%;
- la proyección de los vectores sobre cada eje (componente) permite interpretar a la primera componente (eje horizontal) como "Nivel socio-económico" y a la segunda componente (eje vertical) como "Nivel de instrucción".

Posteriormente, al representar las zonas en el plano de las dos primeras componentes (Gráfico 4), la proximidad de algunas de ellas condujo a pensar en agrupamientos distantes pero con alta cohesión interna. Para corroborar dicha hipótesis se llevó a cabo un Análisis de Cluster con ligamiento completo, utilizando la matriz de distancias euclídeas (debido a la homogeneidad de las variables transformadas); y la lectura del dendograma (ver Anexo) confirmó la presencia de tres grupos: zonas I y VIII (de comportamiento promedio, por encontrarse próximas al origen de coordenadas); zonas V, VI y VII por un lado y zonas III, IV, IX y X por otro (en donde el nivel socio-económico influye decisivamente, al hallarse alejadas de la segunda componente); en tanto, la zona II y la zona XI presentan un comportamiento aislado, a pesar de que en ambas el nivel de instrucción tiene una gran injerencia (pero con distinto sentido en cada caso).

## 4 Discusión

1. El factor que mide la oferta del sistema de salud tiene correlación positiva con las

variables de mortalidad<sup>5</sup>, opuesta a lo esperado. Este resultado es paradójico sólo en apariencia, ya que los datos disponibles se refieren solamente al sistema oficial de salud, que tiene mayor peso relativo en las zonas de mayor riesgo social. Por otra parte, debe considerarse que la población de las zonas V, VI y VII tiene acceso al sistema sanitario de la Capital Federal y que, en consecuencia, los indicadores de dichas zonas subregistran la oferta real del sistema de salud. Debe recalcar que no se trata aquí del clásico problema del lugar de ocurrencia de las defunciones ya que, como se mencionó anteriormente, los niveles de mortalidad fueron calculados a partir del domicilio de residencia habitual de las personas. En futuros trabajos debería hacerse un análisis con mayor cantidad de indicadores de infraestructura sanitaria, lo cual podría cambiar el sentido de la correlación aquí obtenida.

2. La urbanización presenta correlación muy alta con la mortalidad (la menor mortalidad ocurre en áreas urbanas, Gráfico 3). Como hemos visto en el punto 3.1 la cobertura de los datos es completa, por lo que no existe aquí un problema de subregistro de defunciones; por otra parte, en caso de un eventual subregistro, es de esperar que el mismo fuese mayor en las zonas rurales, lo que reforzaría los resultados obtenidos. A pesar de la mayor proporción de población con NBI existente en las zonas urbanas, éstas cuentan con ventajas significativas en relación a la zona rural que le permiten compensar los efectos negativos de la estructura social, al menos en lo relativo a la mortalidad. Entre tales ventajas, la mayor accesibilidad al sistema de salud público (instituciones hospitalarias y disponibilidad de medicamentos), a la información en general y la existencia de rasgos propios de la cultura urbana ocupan un lugar destacado.
3. Los indicadores de riqueza (depósitos bancarios y PBI per cápita) muestran mayor mortalidad en regiones de mayor riqueza, correlación opuesta a la esperable en teoría. A las críticas habitualmente formuladas contra los indicadores per cápita (desconocimiento de la distribución, grave por ejemplo en los depósitos bancarios de regiones rurales de mayor polarización social), debe sumarse el hecho de que el PBI mide la riqueza generada y no la ganancia percibida.
4. La correlación negativa entre mortalidad y proporción de población con NBI constituye la segunda paradoja del estudio. Una posible vía de explicación de este fenómeno se encuentra en la naturaleza misma del indicador, cuya composición (especialmente en aspectos tales como ocupación y vivienda) tiende a subestimar el número de pobres rurales<sup>6</sup>. Si se quitasen las zonas más urbanizadas (V, VI y VII) la correlación sobre los datos originales sería nula.
5. En todos los casos anteriores, las correlaciones de las variables exógenas con cualquiera de los indicadores de mortalidad (TMI y E(0)) explican lo mismo a pesar de

<sup>5</sup>En todas las conclusiones siguientes alta mortalidad implica alta TMI y baja e(0)

<sup>6</sup>Un rancho rural, por ejemplo, no es considerado vivienda precaria. Análogamente, la desocupación es básicamente un problema urbano. Por otra parte, como ha mostrado Susana Torrado (1993:p. 214) a partir de datos del censo 80, cualquiera sea el estrato social considerado, "el nivel de vida es menor cuanto más pequeño es el tamaño del asentamiento de residencia".

la menor reconstrucción de TMI. Por otra parte, dado que la mortalidad infantil en nuestro país tiene valores relativamente altos y una importante proporción de causas de muerte postneonatales, los factores epidemiológicos y ambientales, no contemplados en nuestro modelo por falta de información, tienen un papel más determinante que en el caso de la esperanza de vida al nacimiento.

6. El análisis combinado de cluster y componentes principales permite identificar distintos tipos de zonas sanitarias:
- (a) las V, VI y VII que constituyen el conglomerado bonaerense, claramente destacadas por el peso del componente socioeconómico (altos niveles de urbanización con mayor peso de población con NBI y escasa infraestructura hospitalaria oficial);
  - (b) las zonas III, IX X y IV, de las cuales las 3 primeras están comprendidas en la Pampa Deprimida (la región socialmente más polarizada del interior de la PBA, aunque paradójicamente cuenta con mayor PBI y depósitos bancarios per cápita). En el dendograma, la zona IV se une con las restantes en un nivel posterior;
  - (c) las regiones I y VIII aunque en el gráfico presentan un comportamiento "promedio" en términos de urbanización y desarrollo. Por último, las zonas II y XI presentan un comportamiento atípico, influida por su buen nivel económico no correspondido con el nivel educativo en el primer caso, y con un nivel de instrucción muy destacado en el segundo. Debe señalarse asimismo que la división en zonas sanitarias no refleja adecuadamente las condiciones socio-económicas de la provincia ya que algunas de ellas se caracterizan por una alta dispersión en la distribución de sus indicadores. No cabe duda, por otra parte, que la heterogeneidad observada en el nivel provincial se repite en el interior de cada partido, tanto en términos sociales como espaciales.

## 5 Comentarios finales

Como bien ha señalado Daniel Noin en su estado del arte sobre las diferencias espaciales, "la explicación sigue siendo el punto débil de las investigaciones sobre la mortalidad", especialmente en los estudios de escala intermedia donde en función del conocido fenómeno de la falacia ecológica, las correlaciones realizadas suelen ser sencillamente "enigmáticas" (Noin, 1990: 368-369). Estas apreciaciones generales son igualmente válidas en el estudio que aquí presentamos.

Independientemente de sus limitaciones, los resultados son coincidentes en mostrar una mejor situación relativa de las zonas sanitarias más urbanizadas, a pesar de los valores negativos que caracterizan algunos de sus indicadores sociales. Nuestra hipótesis general sobre las aparentes paradojas mostradas en este estudio postula que, a diferencia de lo que ocurre en la mayor parte de las áreas metropolitanas de los países en vías de desarrollo o de lo ocurrido en la ciudad industrial decimonónica, las condiciones socio-sanitarias del área metropolitana de Buenos Aires son considerablemente mejores que las prevaletentes en

el interior rural de la Pampa Deprimida, hipótesis para cuya verificación serán necesarios nuevos estudios, dado el estado actual de los estudios de mortalidad en la Argentina.

Sin duda hemos arribado a nada más que un punto de partida para nuevos desarrollos e investigaciones. Estas deberían profundizar el análisis de las diferencias espaciales por sexos y grupos de edades específicos y, sobre todo, el análisis geográfico según la distribución de las causas de deceso. En este punto parece indispensable adoptar un enfoque etiológico atento a los agentes productores del proceso de morbi-mortalidad. Los enfoques espaciales basados en la distribución de causas "evitables" de mortalidad (Holland, 1988) pueden ser aquí de suma utilidad.

Si bien pueden y deben esperarse avances significativos en las áreas mencionadas, la principal piedra "dura" a remover en el futuro se halla en la calidad, cobertura, confiabilidad y disponibilidad de los datos relativos a factores exógenos a la mortalidad. Cuando éstos existen, responden mal a las necesidades de la investigación, especialmente en escalas geográficas intermedias. Por otra parte, ciertos datos vitales para la comprensión del proceso de morbi-mortalidad (como, por ejemplo, el consumo de alcohol y tabaco o el consumo alimentario en general, los factores climáticos, la contaminación ambiental o los riesgos laborales) son de muy difícil o imposible obtención. La búsqueda de indicadores más sensibles de tales factores constituirá sin duda uno de los pasos fundamentales de las investigaciones futuras.

## Bibliografía

- [1] ARGENTINA. INSTITUTO NACIONAL DE ESTADÍSTICAS Y CENSOS (1982): *Censo Nacional de Población y Vivienda*, 1980.
- [2] BUENOS AIRES, MINISTERIO DE SALUD, SUB-DIRECCION DE INFORMACION SISTEMATIZADA, *Nacimientos de la Provincia de Buenos Aires, 1976-1983; Defunciones de la Provincia de Buenos Aires, 1980-1982*.
- [3] CASELLI, G.; EGIDI, V. (1988): "Les variations géographiques de la mortalité", en VALLIN, J.; MESLE, F.: *Les causes de décès en France de 1925 à 1978*, Paris, INED.
- [4] CELTON, D. (1993): "La declinación de la mortalidad en la Argentina", en *II Jornadas de la Asociación Argentina de Estudios de la Población*, Buenos Aires.
- [5] CUADRAS, C. M (1981): *Métodos de Análisis Multivariante*. Barcelona, Eunibar.
- [6] CHESNAIS, J. (1986): *La transition démographique*. París, INED.
- [7] HOLLAND, W. (ed.) (1988): *European Community Atlas of 'Avoidable Death'* Oxford University Press.
- [8] MARDIA, K; KENT, J; BIBBY, J (1979): *Multivariate Analysis*. New York, Academic Press.
- [9] MYCHASZULA, S; ACOSTA, L. (1990): *La mortalidad infantil en la Argentina, 1976-1981*. Buenos Aires, CENEP.



- [10] NOIN, D. (1990): "L'étude géographique de la mortalité: bilan et problèmes", en UNION GEOGRAPHIQUE INTERNATIONALE: *Les inégalités géographiques de la mortalité*, Colloque UGI, Lille, 24-28 avril.
- [11] ORTEGA, A. (1987): *Tablas de mortalidad*. San José, Costa Rica, CELADE.
- [12] OTERO, H.; VELAZQUEZ, G. (1995): *Tablas de mortalidad de la Provincia de Buenos Aires por partidos y zonas sanitarias, 1980-1982*. Tandil, Secretaría de Ciencia y Técnica, UNICEN.
- [13] PRESSAT, R.(dir.)(1985): *Manuel d'analyse de la mortalité*. Paris, OMS-INED.
- [14] REED, L; MERREL, M. (1939): "A Short Method for Constructing an Abridged Life Table", en *The American Journal of Hygiene* **30**(2).
- [15] REHER, D.S.; ORTEGA-OSONA, J.A. (s.f.): "Short run economic fluctuations and demographic behaviour: Some examples from twentieth century South America".
- [16] SEBER, G.A (1984): *Multivariate Observations*. New York, John Wiley & Sons.
- [17] SOMOZA, J; MÜLLER, M (1988): *Tablas de mortalidad. 1980-1981. Total y jurisdicciones*. Buenos Aires, INDEC.
- [18] TONNELIER, F. (1989): *Inégalité en matière de santé: Aspects géographiques*. CIRSH, Université de Lille.
- [19] VALLIN, J; D'SOUZA, S; PALLONI, A. (éds.) (1988): *Mesure et analyse de la mortalité. Nouvelles approches*. Paris, INED.
- [20] VELAZQUEZ, G. (1993): *Desigualdades geo-sociales de la mortalidad. El caso del partido de Tandil (Buenos Aires)*. Universidad de Buenos Aires, Tesis Doctoral.
- [21] VELAZQUEZ, G.; LIBERALI, A. (1989): "Geodiversidades ante la vida y la muerte. La 'postergación' extrapampeana y el 'privilegio' pampeano. Análisis de casos", en *Segundas Jornadas de Investigación, Universidad Nacional de Mar del Plata*.
- [22] WUNSCH, G. (1981): *Analyse multidimensionnelle de la mortalité. Etudes exploratoires des données*. Working Paper N° 97, Département de Démographie, Université Catholique de Louvain.

### Agradecimiento

Agradecemos la colaboración de la Dra. Nélica Winzer, Directora del Grupo de Estadística del NUCOMPA.

**ANEXO:****Tabla de datos básicos por zonas sanitarias**

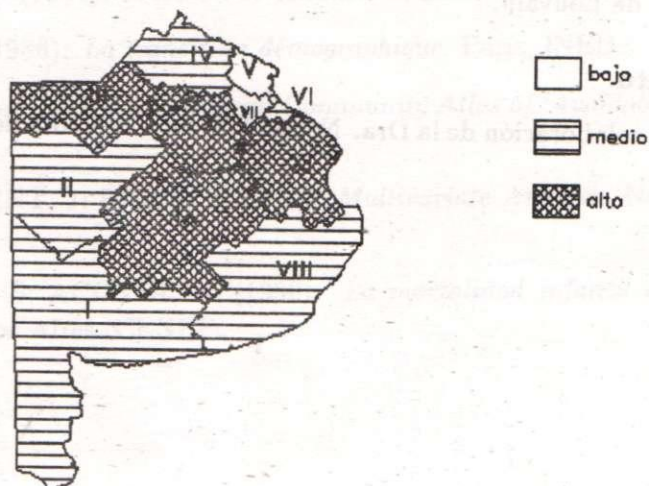
	NMI	Población	PBI	Camas	Pob NBI	Pob s/Ins	Pob c/SyT	Pob Ur	Depósitos	Eo
Zona 1	343	550404	154405	1587	90637	21571	139123	499823	6241,68	70,81
Zona 2	112	214320	99097	928	37659	11345	35733	163810	1880,77	70,20
Zona 3	145	247427	74017	841	44130	11348	44687	186402	2304,59	70,07
Zona 4	274	507809	147131	968	90679	19343	85297	378273	6037,70	70,55
Zona 5	1491	2291378	83779	2100	538147	95005	591692	2260274	15760,34	72,04
Zona 6	1716	2657032	32872	3959	743351	114298	619353	2629632	14865,50	71,21
Zona 7	1738	2514959	40121	5839	659912	104690	581619	2456218	10934,82	72,99
Zona 8	401	795445	119719	1799	153930	29727	198296	724939	8889,69	71,17
Zona 9	167	271770	111559	1138	50508	11937	52040	215198	2737,09	70,13
Zona 10	149	270831	78969	906	48692	13080	48156	219087	2477,16	69,75
Zona 11	522	780649	143751	5049	150278	26804	225990	753449	2412,57	69,68

**Matriz de rangos de Spearman**

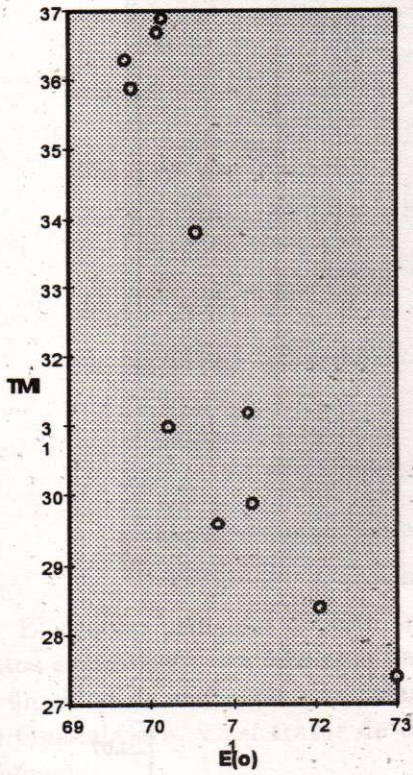
	NMI	Pob	PBI	Camas	Pob NBI	Pob s/Ins	Pob c/SyT	Pob Ur	Depósitos	Eo
Zona 1	6	6	11	6	5	6	6	6	6	7
Zona 2	1	1	6	3	1	1	1	1	1	5
Zona 3	2	2	3	1	2	2	2	2	2	3
Zona 4	5	5	10	4	6	5	5	5	5	6
Zona 5	9	9	5	8	9	9	10	9	11	10
Zona 6	10	11	1	9	11	11	11	11	10	9
Zona 7	11	10	2	11	10	10	9	10	9	11
Zona 8	7	8	8	7	8	8	7	7	7	8
Zona 9	4	4	7	5	4	3	4	3	4	4
Zona 10	3	3	4	2	3	4	3	4	3	2
Zona 11	8	7	9	10	7	7	8	8	8	1

**Mapa**

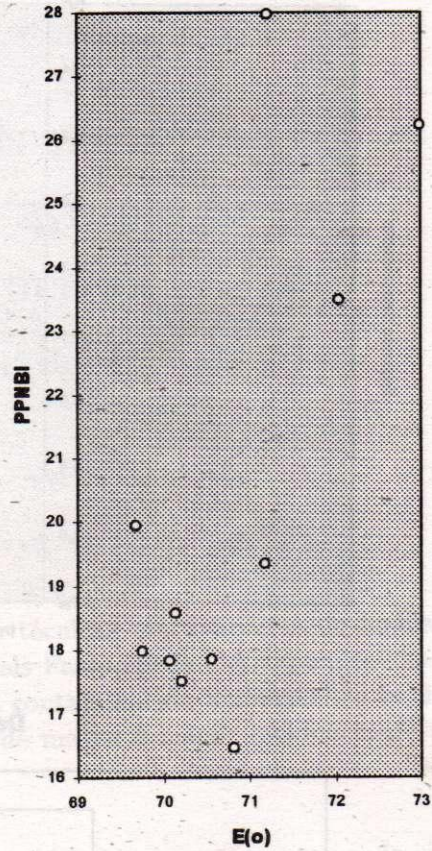
NIVELES DE MORTALIDAD DE LA PROVINCIA DE BUENOS AIRES.  
POR ZONAS SANITARIAS, 1980-1982



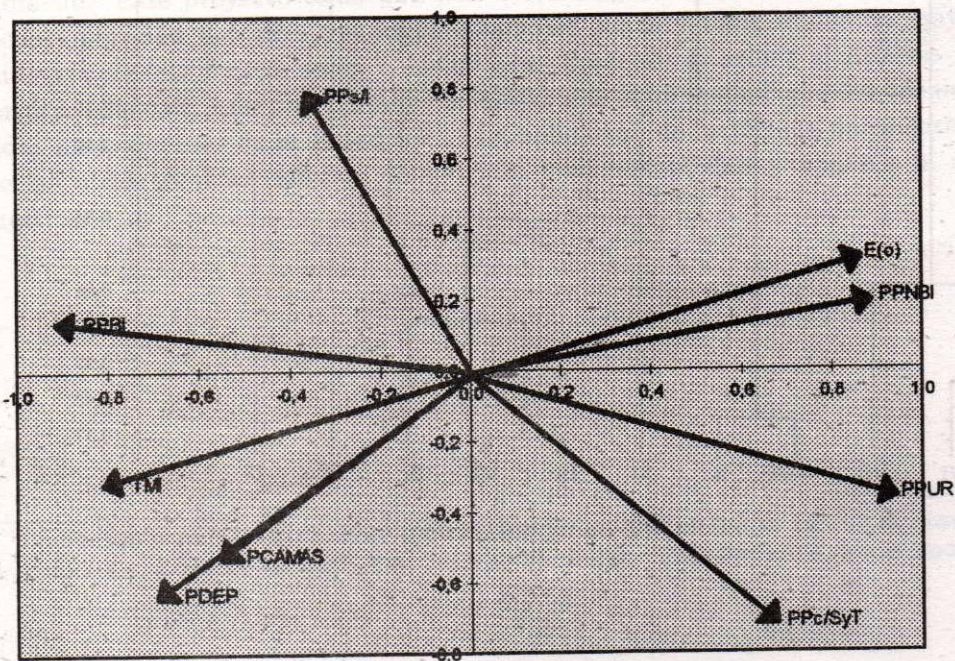
**Gráfico 1**  
Scatter-plot "E(o) vs. TMI"



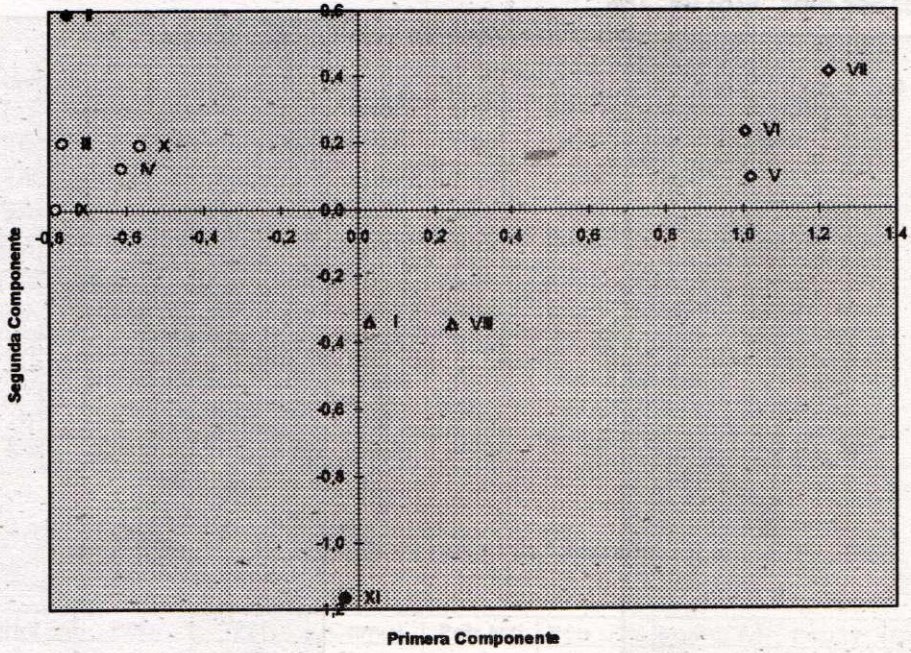
**Gráfico 2**  
Scatter-plot "E(o) vs. PPNBI"



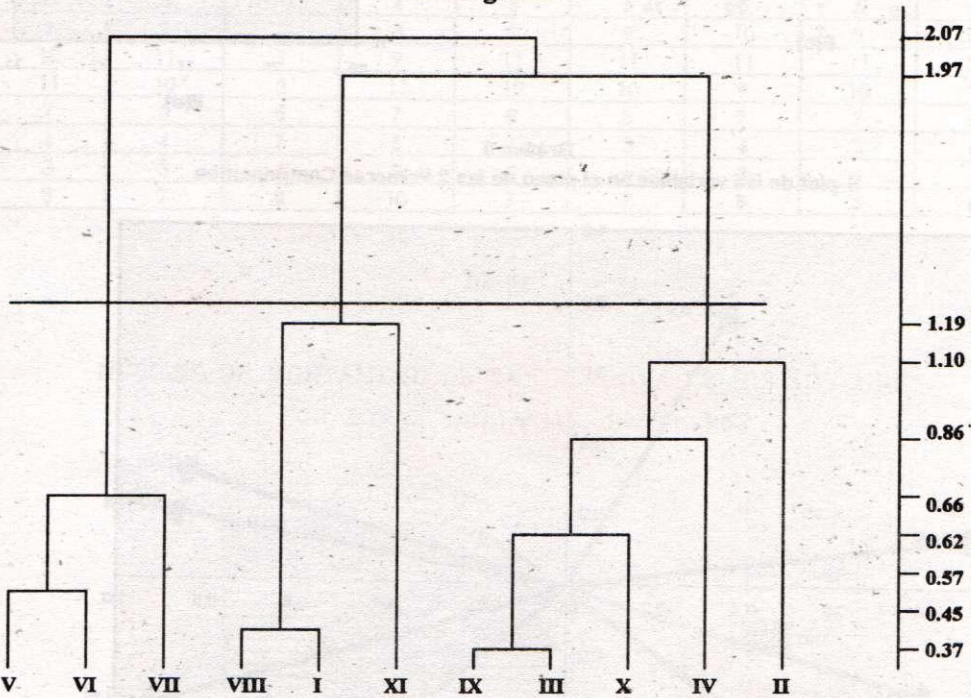
**Gráfico 3**  
H-plot de las variables en el plano de las 2 Primeras Componentes



**Gráfico 4**  
**Zonas en el plano de las 2 primeras Componentes Principales**



**Dendrograma**



APLICACIÓN DEL ANÁLISIS FACTORIAL DE  
CORRESPONDENCIAS AL ESTUDIO DE LA  
CONTAMINACIÓN DEL AIRE  
EN GUADALAJARA, JALISCO (MÉXICO)

LUZ MARÍA ZÚÑIGA MEDINA<sup>1</sup> - JESÚS E. SÁNCHEZ GARCÍA<sup>1</sup>  
IRINA E. TERESCHENKO<sup>1</sup>

---

El ojetivo principal de este trabajo es aplicar técnicas multivariadas de análisis de datos estadísticos –en este caso particular el Análisis Factorial de Correspondencias– con la finalidad de encontrar una caracterización de la contaminación ambiental en la ciudad de Guadalajara, y así tratar de establecer zonas de mayor o menor incidencia de dicho fenómeno.

**Resumen**

Debido al problema tan grande de contaminación del aire que se presenta en la ciudad de Guadalajara, se han hecho diversas investigaciones que den posibles soluciones a este fenómeno. Este proyecto tiene una nueva visión, pues se analiza el comportamiento de la contaminación del aire y se logra establecer zonas y temporadas de mayor o menor incidencia de este fenómeno en Guadalajara. Para ello se utiliza el Analisis Factorial de Correspondencias, que nos permite encontra más información que las técnicas multivariadas mas comunes de la Estadística. La aplicacion de la técnica da mejores resultados que las otras técnicas clásicas, debido a la interpretabilidad y ajuste a la realidad que muestran.

---

<sup>1</sup>UNIVERSIDAD DE GUADALAJARA, DEPARTAMENTO DE MATEMÁTICAS (CUCEI), REVOLUCIÓN 500, GUADALAJARA, JALISCO, MÉXICO; TEL.: 811-49-55; E-MAIL: LUCY@PHYSICS.FISMAT.UDG.MX

## PETRI NETS METHODOLOGY APPLIED FOR PROCESS AND PRODUCT MODELING

MILENA DIDIC<sup>1</sup> - ULRICH MAMPEL<sup>2</sup> - KARL-HEINZ STERNEMANN<sup>3</sup>

---

### Abstract

The application of a mathematical methodology called Petri Nets (PN) has been introduced recently in the scope of the cooperation between the University of Costa Rica and the German institutions listed above.

Petri Nets are formal and the models described by the Petri Nets based description technique are executable. Various verification methods have been developed around the world for their support.

We use a Timed Petri Net based approach to model environment relevant material and energy flow in production systems. It has to fulfill the needs of the ecological life cycle assessment of products as well as for the ecological analysis of the processes in the production sites.

As in all types of Petri Nets we use the three basic elements: place, transition and a directed arc connecting them. The technical processes are modeled in transitions. The relations between process inputs and outputs are described by mathematical functions and calculated for a given period of time. Thereby we apply discrete as well as continuous models of factory applications.

**Keywords:** Petri Net models, ecological life cycle assessment, material and energy flow

---

<sup>1</sup>ESCUELA DE INGENIERIA QUIMICA, UNIVERSIDAD DE COSTA RICA, 2060 SAN JOSÉ, COSTA RICA;  
MILENA@SCRATCHY.EMATE.UCR.AC.CR

<sup>2</sup>RESEARCH CENTER KARLSRUHE - TECHNOLOGY AND ENVIRONMENT, ALEMANIA

<sup>3</sup>FACHHOCHSCHULE OFFENBURG, ALEMANIA

## MODELADO DE MÁQUINAS ASÍNCRONAS E IDENTIFICACIÓN ÓPTIMA

SALVADOR DE LARA<sup>1</sup> - JEAN FAUCHER<sup>2</sup>

---

### Resumen

El estudio del comportamiento de máquinas asíncronas está ligado a una herramienta matemática o modelo, que permite calcular las características de las salidas a partir de los valores de entrada y de los parámetros del modelo matemático. Los autores presentan un sistema de identificación de parámetros para máquinas asíncronas. Los parámetros eléctricos son estimados utilizando datos de entrada-salida a partir de condiciones dinámicas de transitorios eléctricos. El sistema está basado en la metodología del modelo de referencia aplicada fuera de línea, la cual minimiza un criterio función del error entre las salidas de la máquina y del modelo, por un algoritmo de optimización del tipo Quasi-Newton.

Un artículo completo correspondiente a este resumen aparecerá próximamente publicado en la *Revista de Matemática: Teoría y Aplicaciones*.

---

<sup>1</sup>Instituto de investigaciones Eléctricas, Unidad de Control de Procesos, Div. Sistemas de Control, Av. Reforma No. 113, Colonia Palmira, Temixco, Morelos, 62490 México; e-mail: [sdelara@axp8.iie.org.mx](mailto:sdelara@axp8.iie.org.mx)

<sup>2</sup>Laboratoire d'Electronique et d'Electronique Industrielle, LEEI-ENSEEIH, U.R.A.-CNRS No. 847, 2 rue Camichel, 31071 Toulouse Cedex, Francia; e-mail: [faucher@leei.enseeiht.fr](mailto:faucher@leei.enseeiht.fr)