

ESCALAMIENTO MUTIDIMENSIONAL CON RESTRICCIONES USANDO SOBRECALENTAMIENTO SIMULADO

Jorge González Varela¹ y Javier Trejos Zelaya²

CIMPA - Escuela de Matemáticas, Universidad de Costa Rica, San José, Costa Rica

RESUMEN

El Escalamiento Multidimensional con restricciones en la configuración (Multidimensional Scaling with Restrictions on the configuration o MDSR) tiene como objetivo minimizar el Stress sujeto a restricciones sobre la matriz de configuración. De Leeuw y Heiser (1980) proponen un algoritmo para calcular el mínimo de Stress con restricciones en la configuración. Nuestro objetivo es mostrar un procedimiento alternativo utilizando Sobrecalentamiento Simulado, en el caso que las restricciones sean lineales. Utilizando los datos "Expresiones Faciales" presentadas por Borg & Groenen (1987) calculamos, con ambos algoritmos, el stress y la configuración correspondiente. Posteriormente repetimos los cálculos en un juego de 20 datos obtenidos de modificaciones de los anteriores. En todos los casos obtuvimos, con ambos algoritmos, el mismo Stress e idénticas configuraciones (salvo rotaciones). Además con los datos sobre "Esfuerzos Bacteriales" de Mathar (1989) obtuvimos una diferencia del 10 %, a favor del sobrecalentamiento simulado.

Palabras clave: Stress, escalamiento multidimensional con restricciones en la configuración, análisis de proximidades, sobrecalentamiento simulado, método Smacof.

ABSTRACT

Multidimensional Scaling (MDS) with restriction on the configuration (MDSR) aims to minimize Stress subject to some restrictions on the configuration matrix. De Leeuw & Heiser (1980) propose an algorithm to compute the minimum of Stress with restrictions on the configuration. Our objective is to show an alternate procedure based on Simulated Annealing, in the case that restrictions are linear. Using the "Facil Expressions" data, presented in Borg & Groenen (1987), we calculate Stress with both methods, and the corresponding configuration. We repeat the computations in a set of 20 data obtained from a modification of the preceding data. In all cases we obtained, with both algorithms, the same Stress and identical configuration (except but rotations). Moreover, with the "Bacterial Strain" data (Mathar (1989)) we obtained for Simulated Annealing better results than Smacof in 10 % of the cases.

Key words: Stress, multidimensional scaling with restrictions on the configuration, multidimensional scaling, simulated annealing, Smacof.

AMS: 62H99

1. INTRODUCCION

El problema básico de *MultiDimensional Scaling (MDS)* es que dada una matriz de disimilitud δ , de tamaño $n \times n$, se trata de encontrar una matriz X de dimensión $n \times p$ que minimice la función de Stress:

$$\sigma(X) = \sum_{i < j}^n w_{ij} (\delta_{ij} - d_{ij}(X))^2 \quad (1)$$

donde $d_{ij}(X)$ es la matriz de distancias euclídeas entre las filas de X y $W = (w_{ij})$ una matriz de pesos de tamaño $n \times n$.

Se han propuesto varios métodos para estimar la configuración X , entre los cuales se encuentra el método SMACOF ("Scaling by MAjorizing a COmplicated Function") que se basa en el algoritmo de Mayorización Iterativa (Iterative Majorization algorithm) (ver por ejemplo Groenen (1993)) y el método que se fundamenta en el algoritmo de Sobrecalentamiento Simulado (*Simulated Annealing*), el cual abreviamos como SS.

SMACOF usa un algoritmo muy sencillo de programar y de convergencia muy rápida, pero con el inconveniente de que no se garantiza un óptimo local.

¹Fax: +(506) 207 4397; E-mail: jgonzale@cariari.ucr.ac.cr

²Fax: +(506) 207 4397; E-mail: jtrejos@cariari.ucr.ac.cr

Por su parte el SS es un método de Optimización Estocástica, introducido por Kirkpatrick **et al.**, (1983), el cual utilizando un algoritmo conocido como la regla de Metrópolis garantiza una convergencia asintótica a un óptimo global (ver Aarts & Korst (1988)).

Un estudio comparativo de ambos métodos, en el marco del MDS, realizado por Villalobos (1988, pp. 57-65), muestra las ventajas del SS cuando la función de stress tiene varios mínimos locales. Nuestro propósito es extender el estudio al caso en que hay restricciones lineales sobre la configuración.

2. ESCALAMIENTO MULTIDIMENSIONAL CON RESTRICCIONES EN LA CONFIGURACION

El MDSR, en el caso de restricciones lineales tiene como objetivo minimizar la función:

$$\sigma(X) = \sum_{i < j}^n w_{ij} (\delta_{ij} - d_{ij}(X))^2 \text{ con } X = YC \quad (2)$$

donde Y es una matriz de $n \times q$ dada y C es una matriz de tamaño $q \times p$ a determinar.

De Leeuw y Heiser (1980), proponen de manera análoga a SMACOF un algoritmo para calcular el mínimo del Stress con restricciones en la configuración. Para ello se expresa el stress como:

$$\sigma(X) = \eta^2(\delta) + \eta^2(X) - 2\rho(X) \quad (3)$$

con

$$\eta^2(\delta) = \sum_{i < j} w_{ij} \delta_{ij}^2 \quad \eta^2(X) = \sum_{i < j} w_{ij} d_{ij}^2(X) \quad \rho(X) = \sum_{i < j} w_{ij} \delta_{ij} d_{ij}(X).$$

Si Z es una matriz de $n \times p$, se define V y B(Z) como:

$$V_{ii} = \sum_{j \neq i}^n w_{ij} \quad y \quad V_{ij} = -w_{ij}$$

$$b_{ij} = \begin{cases} -\frac{w_{ij} \delta_{ij}}{d_{ij}(Z)} & i \neq j, \quad d_{ij}(Z) \neq 0 \\ 0 & i \neq j, \quad d_{ij}(Z) = 0 \end{cases} \quad b_{ii} = -\sum_{j \neq i}^n b_{ij}$$

Usando V y B(X) el Stress se expresa como:

$$\sigma(X) = \eta(\delta^2) + \text{tr}(X^t V X) - 2\text{tr}(X^t B(X) X).$$

Si $r(X, Z) = \eta^2(\delta) + \text{tr}(X^t V X) - 2\text{tr}(X^t B(Z) Z)$ entonces

$$\sigma(X) = r(X, X) \leq r(X, Z) \text{ para todo } Z.$$

A partir de esta última desigualdad se deduce el siguiente teorema.

Teorema 2.1. Sea Y matriz de tamaño $n \times q$. El mínimo sobre X de $r(X, Z)$ con Z fijo de tamaño $n \times p$, sujeto a la restricción $X = YC$ con C de tamaño $q \times p$, se alcanza en $X^u = YC^u$ con $C^u = (Y^t V Y)^{-1} Y^t V \bar{Z}$, donde $\bar{Z} = V^+ B(Z) Z$ es la transformada de Guttman de Z y V^+ es la inversa de Moore-Penrose de V.

El teorema permite definir el algoritmo descrito a continuación.

2.1. Algoritmo basado en SMACOF para minimizar el Stress con la restricción $X = YC$

Sea X_0 una configuración inicial y $X_{m+1} = X_m C_m^u$. Del teorema se tiene que:

$$\sigma(X_m) = r(X_m, X_m) \leq r(X_m, X_{m-1}) \leq r(X_{m-1}, X_{m-1}) = \sigma(X_{m-1})$$

luego la sucesión $(\sigma(X_m))$ es decreciente y acotada, por lo tanto convergente. Un algoritmo de cálculo, que denominamos MDSR, basado en el descrito en De Leeuw y Heiser (1980), se define a continuación.

1. X_0 configuración al azar; $m = 0$.
2. $\bar{X}_m = V^*B(X_m)X_m$, $C_m^u = (Y^t V Y)^{-1} Y^t \bar{X}_m$
3. Haga $X_{m+1} = Y C_m^u$
4. Regrese a (2) hasta que $\sigma(X_m) - \sigma(X_{m+1}) < \epsilon$ ó $m = n_i$ con n_i número máximo de iteraciones.

3. MDS CON RESTRICCIONES Y SOBRECALENTAMIENTO SIMULADO

Los aspectos fundamentales del método que proponemos son los siguientes:

3.1. Sistema de Generación de Estados

3.1.1. Mallado

Como se trata de minimizar la función del Stress sobre la matriz C de tamaño $q \times p$, realizamos una discretización del espacio \mathbb{R}^q , construyendo un mallado formado por la unión de cuadrados de lado h .

3.1.2. Vecindario

Sea $B = \{\pm e_i | e_i \text{ vector canónico de } \mathbb{R}^q\}$.

Denotemos como $C(l,v)$ la matriz obtenida de C al sustituir la columna l - ésima de C , que denotamos C^l por $C^l = C^l + hv$ con $v \in B$.

Definimos el vecindario de C como:

$$V(C) = \{C(l,v) | l = 1, \dots, p; v \in B\}.$$

Notemos que para cualquier C los vecindarios tienen la misma cardinalidad $|V(C)| = 2qp$.

3.1.3. Algoritmo

1. Se escoge una matriz C al azar, cada entrada con probabilidad uniforme en $[0,1]$.

Se calcula la temperatura inicial c_0 , se elige el largo de la cadena L_c y se inicia el contador en $t = 0$. Escoger $\gamma \in [0.8, 0.95]$.

2. Sea $t = t + 1$ y $c_t = \gamma c_{t-1}$. Repetir L_c veces los pasos (a) y (b).

(a) Elegir al azar una columna l de C y una dirección v de B y generar un vecino $C(l,v)$

(b) Calcular el incremento del Stress: $\Delta S = \text{Stress}(C) - \text{Stress}(C(l,v))$. Aceptar $C(l,v)$ de acuerdo con la regla de Metropolis, esto es con probabilidad igual a $\min\left\{1, e^{-\frac{\Delta S}{c_t}}\right\}$. Si es aceptado redefinir $C = C(l,v)$.

3. Si la temperatura c_t es menor que el umbral definido parar, en caso contrario regresar a 2.

3.1.4. Convergencia

Notemos que de acuerdo con el proceso de generación definido tenemos que la probabilidad de generar el estado j a partir del estado i es constante e igual a $G_{i,j} = \frac{1}{2qp}$, por lo que la reversibilidad $G_{i,j} = G_{j,i}$ y la conexidad de dos estados cualesquiera se satisfacen trivialmente, es decir, las condiciones de convergencia asintótica se satisfacen (Aarts & Korst (1988)). En adelante nos referiremos a este algoritmo como **ssMDSR**.

4. COMPARACION ENTRE MDSR Y ssMDSR

El propósito de esta sección es comparar los algoritmos descritos anteriormente MDSR y ssMDSR. Para ello utilizaremos los datos descritos por Borg y Groenen (1997) bajo el nombre de "Expresiones Faciales" y los reportados por Mathar (1989) de "Esfuerzos Bacteriales".

La comparación la realizamos calculando con ambos procedimientos el coeficiente de variación r , el cual se encuentra entre 0 y 1 y es una medida de la calidad de la aproximación de la matriz de disimilitud por la matriz de distancia obtenida.

$$r = 1 - \frac{\sigma(X)}{n^2(\delta)} = 1 - \frac{\sum_{i < j}^n w_{ij} (\delta_{ij} - d_{ij}(X))^2}{\sum_{i < j} w_{ij} \delta_{ij}^2} \quad (4)$$

4.1. Datos de expresiones faciales

Con el propósito de analizar si un sujeto puede identificar correctamente el mensaje emocional que una persona transmite a través de su expresión facial, Engen *et al.*, (1958) idearon un experimento que se fundamenta en una teoría desarrollada por ellos, la cual afirma que una buena clasificación de las diferentes expresiones faciales debe considerar las variables AD, ID, TD en los siguientes rangos:

AD: Agradable-Desagradable.
ID: Interesado-Desinteresado.
TD: Tenso-Dormido.

Tabla 1. Matriz de restricciones entre trece expresiones faciales.

	SITUACION	AD	ID	TD
1	Muerte madre	3.8	4.2	4.1
2	Comiendo galleta	5.9	5.4	4.8
3	Sorpresa agradable	8.8	7.8	7.1
4	Amor maternal	7.0	5.9	4.0
5	Cansancio físico	3.3	2.5	3.1
6	Problema en avión	3.5	6.1	6.8
7	Perro bravo	2.1	8.0	8.2
8	Empujando	6.7	4.2	6.6
9	Encuentro inesperado	7.4	6.8	5.9
10	Cambio repentino	2.9	3.0	5.1
11	Dolor extremo	2.2	2.2	6.4
12	El avión cae	1.1	8.6	8.9
13	Sueño ligero	4.1	1.3	1.0

Engel, Levy y Schlosberg usaron cuarenta y ocho fotografías de la expresión facial de una mujer ante trece situaciones diferentes para luego asignarles de manera empírica, usando una escala de nueve puntos, valores en las tres variables definidas anteriormente. Así se obtuvo la Tabla 1 de restricciones en la configuración.

Por otra parte Abelson y Sermat (1962) construyeron, usando las mismas fotografías, una matriz de disimilitud. Para ello recurrieron a un grupo de treinta estudiantes que opinaron sobre la diferencia entre el grado de alegría que reflejaban las diferentes expresiones faciales. Teniendo en cuenta estas evaluaciones construyeron la matriz de disimilitud mostrada en la Tabla 2.

Tabla 2. Matriz de similitud entre trece expresiones faciales.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0												
2	4.05	0											
3	8.25	2.54	0										
4	5.57	2.69	2.11	0									
5	1.15	2.67	8.98	3.78	0								
6	2.97	3.88	9.27	6.05	2.34	0							
7	4.34	8.53	11.87	9.78	7.12	1.36	0						
8	4.90	1.31	2.56	4.21	5.90	5.18	8.47	0					
9	6.25	1.88	0.74	0.45	4.77	5.45	10.20	2.63	0				
10	1.55	4.84	9.25	4.92	2.22	4.17	5.44	5.45	7.10	0			
11	1.68	5.81	7.92	5.42	4.34	4.72	4.31	3.79	6.58	1.98	0		
12	6.57	7.43	8.30	8.93	8.16	4.66	1.57	6.49	9.77	4.93	4.83	0	
13	3.93	4.51	8.47	3.48	1.60	4.89	9.18	6.05	6.55	4.12	3.51	12.65	0

4.2. Obtención de una configuración plana vía MDSR y SS

Para realizar un estudio de las Expresiones Faciales disponemos de una matriz de disimilitudes δ y de una tabla Y de expresiones faciales por variables, obtenidas de manera independiente una de otra. Sin embargo es natural plantearse que exista una relación entre ellas, cuestión que podemos verificar al asumir que es posible construir una configuración X de 13 puntos en el plano, de modo que las dos variables que la determinan sean combinaciones lineales de las tres variables definidas anteriormente y que además la matriz de distancias euclídeas $d(X)$ aproxime la matriz de disimilitud.

Estamos entonces ante un problema de Escalamiento Multidimensional con restricciones lineales en la configuración que resolveremos con los dos algoritmos que hemos definido.

En este caso, δ es la matriz de disimilitud de tamaño 13×13 de la tabla 2, Y es la matriz de 13×13 de la tabla 1 que tiene los valores que las expresiones faciales asumen en las tres variables empíricas definidas y necesitamos encontrar una matriz C de tamaño 3×2 que minimice la función de stress (2):

Estimamos la matriz C con los algoritmos MDSR y ssMDSR, ambos programados en *Mathematica* 3.0.

La matriz C, el valor del Stress y el gráfico de la configuración obtenida con ambos algoritmos se reportan en seguida. Además se incluyen los resultados obtenidos por Borg y Groenen (1997) para los mismos datos.

Tabla 3. Matrices C obtenidas por los métodos MDSR y ssMDSR y la reportada por Borg & Groenen (1997).

	MDSR		ssMDSR		Borg & Groenen	
Peso de AD	0.815	1.111	-0.481	1.329	0.219	0.031
Peso de ID	0.289	-0.287	0.474	0.086	-0.35	0.137
Peso de TD	0.453	-0.549	0.541	-0.017	-0.024	0.053

En el caso de ssMDSR se utilizaron los siguientes parámetros: largo de la cadena $L_c = 800$, $\gamma = 0.85$, temperatura inicial $c_0 = 285$ (estimada según se propone en Aarts & Kors (1988)), paso del mallado $h = 0.05$.

El coeficiente de variación r con ambos algoritmos fue de 0.955. En la Tabla 3 se indican las matrices C y en la Figura 1 las configuraciones correspondientes a MDSR, ssMDSR y la reportada por Borg & Groenen (1997).

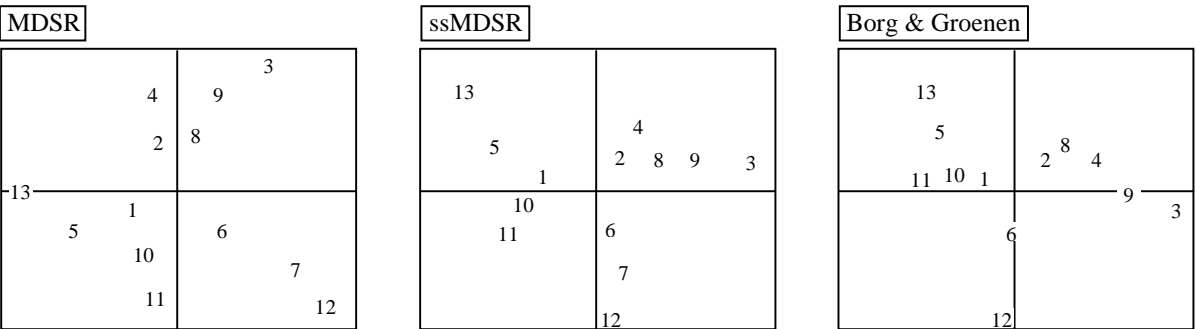


Figura 1. Configuraciones de las trece expresiones faciales obtenidas con los métodos MDSR, ssMDSR y la reportada por Borg & Groenen (1997).

Podemos observar que las configuraciones en los tres casos son iguales, salvo rotación. Por otra parte las variables sugeridas por Engen, Levy y Scholosberg nos dan una configuración análoga a las anteriores como podemos observar en el plano principal del Análisis en Componentes Principales realizado sobre la matriz Y (ver Figura 2), lo que indica, en alguna medida, que las tres variables definidas explican adecuadamente las expresiones faciales.

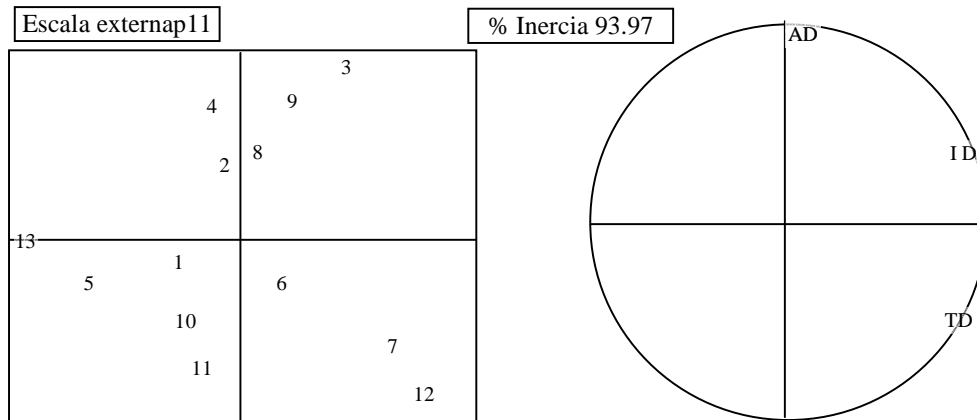


Figura 2. Plano principal y círculo de correlaciones del Análisis en Componentes Principales de la matriz Y.

4.3. Esfuerzos bacteriales

Estos datos fueron reportados por Mathar (1989) y corresponde a experimentos realizados en 1986 entre 17 individuos. La matriz de disimilitud tiene la particularidad que al realizar un MDS sin restricciones se han encontrado alrededor de 1000 mínimos locales.

En este caso usamos una matriz Y de tamaño 17 x 3 (ver Tabla 4), elegida de modo que produjera varios mínimos locales. Realizamos 60 corridas de cada algoritmo, en el caso del MDSR hicimos, por cada corrida, 10 tirajes al azar de la configuración inicial y elegimos el mejor r. Para ssMDSR se hizo un tiraje al azar por cada corrida, pues en este caso hay una menor dependencia de la configuración inicial y un mayor tiempo de ejecución. La tabla de disimilitudes se presenta en la Tabla 5.

Tabla 4. Matriz de restricciones transpuestas (Y^t) para los datos de esfuerzos bacteriales.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	2.8	7.1	6.4	6.4	4.0	8.5	4.6	6.4	6.4	8.4	8.1	4.5	7.2	3.4	6.0	8.7	5.8
2	4.9	4.9	8.4	2.5	7.5	7.2	2.5	2.0	2.4	2.5	8.4	6.6	5.0	8.3	1.5	4.9	7.4
3	5.5	2.2	1.9	7.5	7.2	7.9	6.9	6.4	5.2	1.1	1.4	2.1	8.5	6.0	2.7	3.7	8.6

El mejor r obtenido en ambos algoritmos fue $r = 0.777712$, pero con MDSR en un 80 % de las veces contra un 90 % del ssMDSR. Los otros valores de r obtenidos en el caso del MDSR, con un 5 % de las veces fueron: 0.777702, 0.777691, 0.777559, 0.776885 y para el ssMDSR el otro 10 % fue para $r = 0.777703$.

Tabla 5. Matriz de disimilitudes para los datos de esfuerzos bacteriales.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	35	28	29	31	42	49	70	60	100	75	48	77	60	36	88	93
2	35	0	17	42	34	49	55	77	54	89	98	98	78	43	23	86	100
3	28	17	0	32	38	59	32	96	62	95	93	85	91	38	44	80	79
4	29	42	32	0	32	36	100	66	55	93	81	73	98	54	47	78	56
5	31	34	38	32	0	50	57	86	86	100	100	74	86	34	45	94	80
6	42	49	59	36	50	0	33	100	69	100	79	90	88	51	55	92	97
7	49	55	32	100	57	33	0	85	89	57	71	85	100	63	57	93	100
8	70	77	96	66	86	100	85	0	81	89	77	87	58	80	73	88	90
9	60	54	62	55	86	69	89	81	0	92	79	73	82	73	86	88	100
10	100	89	95	93	100	100	57	89	92	0	100	85	89	52	42	96	57

(...Continuación **Tabla 5**)

11	75	98	93	81	100	79	71	77	79	100	0	100	100	100	85	62	80
12	48	98	85	73	74	90	85	87	73	85	100	0	87	63	81	67	71
13	77	78	91	98	86	88	100	58	82	89	100	87	0	79	92	84	66
14	60	43	38	54	34	51	63	80	73	52	100	63	79	0	4	69	76
15	36	23	44	47	45	55	57	73	86	42	85	81	92	4	0	81	57
16	88	86	80	78	94	92	93	88	88	96	62	67	84	69	81	0	79
17	93	100	79	56	80	97	100	90	100	57	80	71	66	76	57	79	0

5. CONCLUSIONES

En los ejemplos mostrados se observa una vez más la tendencia reportada por otros autores cuando comparan los algoritmos clásicos con los obtenidos con sobrecalentamiento simulado, esto es que este último obtiene mejores resultados en los casos en que se presentan varios mínimos locales. Tiene la desventaja de que su tiempo de ejecución es un poco más del doble que el MDSR.

REFERENCIAS

- AARST, E. and J. KORST (1988): **Simulated Annealing and Boltzman Machines: a Stochastic Approach to Combinatorial Optimization and Neural Computin**, John Wiley & Sons, Chichester.
- ABELSON, R.P. and V. SERMAT (1962): "Multidimensional scaling of facial expressions", **Journal of Experimental Psychology**, 63:546-554.
- BORG, I. and P.J.F. GROENEN (1997): **Modern Multidimensional Scaling**, Springer, New York.
- DE LEEUW, J. and W. HEISER (1980): "Multidimensional scaling with restrictions on the configuration", En: P.R. Krishnaiah (Ed.) **Multivariate Analysis**, North-Holland Publishing Company, Amsterdam; 501-522.
- ENGEL, T.; N. LEVY and H. SCHLOSBERG (1958): "The dimensional analysis of a new series of facial expressions", **Journal of Experimental Psychology**, 55:454-458.
- MATHAR, R. (1989): "Algorithms in multidimensional scaling", In: O. Optiz (Ed.), **Conceptual and Numerical Analysis of Data**, Springer, Heidelberg: 159-177.
- KIRKPATRICK, S.; C.D. GELLAT and M.P. VECCHI (1983): "Optimization by simulated annealing", **Science**, 220:671-680.
- TREJOS, J. y M. VILLALOBOS (1998): "Análisis de proximidades usando sobrecalentamiento simulado", En: Castillo, W.; Trejos, J. (Eds.), **Estudios de Análisis de Datos y Estadística**, Universidad de Costa Rica, Instituto Tecnológico de Costa Rica: 41-52.
- VILLALOBOS, M. (1998): **Optimización Estocástica para el Análisis de Proximidades**, Tesis de Maestría, Universidad de Costa Rica.