APPLICATION OF SIMULATED ANNEALING IN METRIC MULTIDIMENSIONAL SCALING

Mario Villalobos[1] y Javier Trejos[2], CIMPA-PIMAD, School of Mathematics, University of Costa Rica, San José, Costa Rica

ABSTRACT
In Multidimensional Scaling, we want an Euclidean representation of a set of points described by a dissimilarity table. Since no exact solution is known, there is a large number of methods that give an approximated solution minimizing some criterion. This criterion is usually a least squares one, called Stress, that compares the known dissimilarities to the Euclidean distances calculated in representation. The best known methods are gradient descent-type and lead to local optima of Stress. Some other methods, based in a majorizing function (SMACOF method) or the Tunneling method, also cannot guarantee a global optimum. Finally, there are also implementations of genetic algorithms that are quiet slow. We propose a simple implementation of simulated annealing that gives good results. We define a grid of the space of representatrion of the solution, and we go over this grid according to the Metropolis rule. The grid could be thiner as the control parameter, that plays the role of the temperature, tend to zero. We have compared the performances of our method, and its results are comparable and sometimes better than those obtained with other methods.

Key words: metric multidimensional scaling, analysis of vicenities, Stress, combinatorial optimization, Metropolis rule, discretization.

RESUMEN
En Escalamiento Multidimensional o Análisis de Proximidades, se quiere obtener una representación euclídea de un conjunto de puntos descritos por una tabla de disimilitudes. Como no existe una solución exacta, hay varios métodos propuestos para obtener una solución del problema minimizando un criterio. Este criterio es usualmente un criterio de mínimos cuadrados, llamado Stress, que compara las disimilitudes conocidas con las distancias euclídeas calculadas en la representación. Los métodos más conocidos son de tipo descenso de gradiente y conducen a óptimos locales del Stress. Algunos otros métodos, basados en una función de mayorización (método SMACOF) o el método de Tunneling, tampoco garantizan que se llegue a un óptimo global. Finalmente, existen también implementaciones de algoritmos genéticos que son un poco lentos. Se propone en este artículo una implementación simple del sobrecalentamiento simulado que da buenos resultados. Se define una malla del espacio de representación de la solución, y se recorre esta malla de acuerdo con la regla de Metrópolis. La malla se puede hacer más fina conforme avanzan las iteraciones y el parámetro de control, que juega el papel de la temperatura, tiende hacia cero. Hemos comparado el rendimiento de nuestro método, y los resultados son comparables o mejores que los de métodos conocidos.

Palabras clave: escalamiento multidimensional métrico, análisis de proximidades, Stress, optimización combinatoria, regla de Metropolis, discretización.

1. INTRODUCTION

Given a dissimilarity table $(\delta_{ij})_{n\times n}$ in a set of objects Ω = {1,…,n}, we want to find a set of vectors $\vec{x}_1,...,\vec{x}_n \in \mathbb{R}^p$, which form the columns of a matrix X, such that the following Stress is minimized

$$\begin{aligned}\sigma^2(X) &= \sum_{j=1}^{n}\sum_{i=1}^{j-1} w_{ij}[\delta_{ij} - d_{ij}(X)]^2 \\ &= \sum_{i<j} w_{ij}\delta_{ij}^2 + \sum_{i<j} w_{ij}\delta_{ij}^2(X) - 2\sum_{i<j} w_{ij}\delta_{ij}d_{ij}(X) \qquad (1)\\ &= \eta_{ij}^2 + \eta^2(X) - 2\rho(X)\end{aligned}$$

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

where $d_{ij}(X)$ is the Euclidean distance between $\vec{x}_i$ and $\vec{x}_j$, calculated in $\mathbb{R}^p$, and w_{ij} are weights associated to the dissimilarities; these weights can indicate, for example, some non available information with a null value. Eventually, the δ_{ij} could be distances. The dimension p should be small, say 2 or 3.

In case that the δ_{ij} are Euclidean, the solution of the problem is obtained with the diagonalization of the scalar products matrix $\vec{x}_i^t \vec{x}_j$ obtained from the δ_{ij} by the Torgerson form (see Bouroche **et al**. (1977) or Cox & Cox (1995)), p being the number of non-zero eigenvalues of the scalar products matrix.

In the following section we present the best known methods that are proposed to minimize $\sigma(X)$. All results presented in $\mathbb{R}^2$.

2. SOME KNOWN METHODS IN MULTIDIMENSIONAL SCALING

Kruskal (1964) proposed a method of the gradient descent-type for minimizing the normalized Stress, that is σ^2 divided by the sum of $d_{ij}^2(X)$. From a given initial configuration, it is improved by the translation of points $\vec{x}_i$ in the direction of the gradient, by choosing sdome number of parameters that measure how much we approach the new point.

The SMACOF method, De Leeuw (1977) and De Leeuw (1988), uses a majorization function of σ^2, decomposing in three terms. It can be proved that the sequence constructed by the majorization functions is not increasing and converges (to a local optimum).

The Tunneling method used by Groenen (1993a) and Groenen (1993b), constructs “tunnels” in the SMACOF’s majorization function, so as to “go to the other side of the hill” in a horizontal search of a state with equal value of Stress, and follow the search of optima of $\sigma^2(X)$ by SMACOF descent.

Mathar (1995) studied an implementation of genetic algorithms, using a crossing function between two configurations X_1 and X_2 that computes the convex combination $\lambda X_1 + (1 - \lambda)X_2$. On the other hand, N’Gouenet (1995) uses also a genetic algorithm programmed in parallel.

Mathar & Žilinskas (1993) studied the convexity of the Stress function. Some studies of the optimality have been made in the simplest cases (see Pliner (1996) and Simantiraki (1996)).

3. THE PROPOSED IMPLEMENTATION

We propose the use of simulated annealing, that has shown good performance in many combinatorial optimization problems Kirkpatrick **et al**. (1983), Aarts & Korst (1988) and mainly in Data Analysis (see De Amorim **et al**. (1992), Trejos (1992), Trejos **et al**. (1998)). It is known that simulated annealing converges asymptotically to the global optimum of the function to be minimized, and that a good finite-time implementation is essential. Particularly, it has to be chosen a procedure for estimating the initial value of the control parameter c_0, a slow-decreasing law of the control parameter c_k and a criterion for stopping the algorithm; finally, it must be defined the maximum value L_k for the number of iterations associated with each value of c_k. The choice of these four parameters is called a *cooling program*.

In the implementation that we propose, a state I is a set of n vectors in a discretized space $\mathbb{R}^p$:

$$I = [\vec{x}_1^I, \ldots, \vec{x}_n^I]$$

where $\vec{x}_i^I \in (h\mathbb{Z})^p$, $h\mathbb{Z} = \{x \in \mathbb{R} / x = hr,\ r \in \mathbb{Z}\}$ and $h \in \mathbb{R}^+$. We will say that a state J is a neighbour of I if there exist $l \in \{1,\ldots,n\}$ and $j \in \{a,\ldots,p\}$ such that

$$\vec{x}_i^J = \begin{cases} \vec{x}_i^I & \text{if } i \neq l,\ i \in \{1,\ldots,n\} \\ \vec{x}_i^I + h\vec{e}_j & \text{if } i = l. \end{cases} \qquad (2)$$

$\vec{e}_j$ being the j-th canonical vector in $\mathbb{R}^p$. The cardinal of the set of neighbours if I is then 2np. The procedure for generating a new state will be:

1. Select at random $I \in \{1,\ldots,n\}$.
2. Select at random $j \in \{1,\ldots,p\}$.
3. Define J according to (2).

The algorithm SAMSCAL is an implementation of usual simulated annealing as described in Aarts & Korst (1988), using the procedure described for generating new configurations of points in $\mathbb{R}^p$.

4. RESULTS

We have applied the SAMSCAL algorithm over many examples, for comparing its performances with respect to the best known methods. We will call *attraction rate r* of a local minimum of Stress the percentage of times that this minimum is reached in a large number of executions of the program. For comparing our results to those obtained by other authors, we use the normalized Stress σ^2/η_δ^2, where η is the sum of square dissimilarities; this ratio is the normalized Stress and usualy appears in the results of many authors. We note σ^2_{opt} the best value of σ^2 found with the program SAMSCAL and # the number of times that the algorithm has been applied.

The fixed distances data

We consider a set of n points such that

$$\delta_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j. \end{cases}$$

For many values of n we obtained the following results:

	$\sigma^2_{opt}/\eta^2_\delta$	#	r	mean	$\sigma^2_{opt}/\eta^2_\delta$ for Smacof
4 points	0.028595479	80	100 %	0.02859548	0.02859547921
6 points	0.071454137	224	88.8 %	0.07197208	-
7 points	0.085922088	140	37.14 %	0.08599070	

In the case of 7 points, we also found the local minimum 0.086029671, 62.86 % of the runs.

Colas data

We consider the colas data reported in Green **et al**. (1989), that cross 10 beverages and shows the similarities between each couple, according to an experience with 38 students. Mathar (1995) found 152 local optima, whole Mathar & Žilinskas (1993) found 457. Using SAMSCAL, we found 30 local optima.

$\sigma^2_{opt}/\eta^2_\delta$	#	r	**SMACOF** Groenen **et al**. (1995)
0.0367837933	310	17.74 %	0.03678052

It should be noted that the mean error that we found is 0.0394416, while those found by Groenen **et al**. (1995) is 0.04145104, 0.4070994 and 0.04113443, according to different methods: Smacof, relaxed Smacof and Kruskal's, respectively.

In the following table, we present the 4 best minima and their frequencies of appearance in 310 executions of the program. It can be seen that these four minima are very similar and that they are found in nearlly 50 % of the times. Also, the corresponding configurations are quiet similar.

Minima	Frequency	Cumulated frequency	Percent (%)	Cumulated percent (%)
0.036786	55	55	17.74	17.74
0.036860	38	93	12.26	30.00
0.036975	24	117	7.74	37.74
0.037363	36	153	11.61	49.35
other	157		51.65	100.0

The 9 points example

In a plane, we consider 9 points that form a squared grid with 3 points equally spaced in each side, and we compute the usual Euclidean distance between the 9 points. In the table that follows we present some minima, as well as the results found with SAMSCAL, and those shown in Groenen (1993b).

	SCAL_SS		Groenen (1993b), Multistart
minimum	**#**	**r**	**r**
0.0000	274	95.47	72.0
0.05671	1	0.35	-
0.07391	12	4.14	19.5
other	0	0.00	8.5

Example of 10 random points in [-1, 1]^5

This example is presented in Mathar & Žilinskas (1993). It consists on the Euclidean distances between 10 point generated randomly in [-1, 1]^5. These authors cite that they found 133 local minima.

	SCAL_SS	Mathar & Žilinskas (1993)
$\sigma^2_{opt}/\eta^2_\delta$	0.036026404	0.036162536
r	37.75 %	15.6 %
Mean	0.0401566	0.0417546

5. CONCLUSIONS AND FURTHER RESEARCH

It can be seen over the presented examples that the results obtained with our method are as good those obtained with other methods, not only in the rate of attraction, but also in the mean values.

Some deeper studies have to be made, mainly in the effect of the grid over the solutions. A Monte Carlo simulation will soon be made. On the other hand, we have also made some research on the use of Tabu Search in Multidimensional Scaling; some results can be consulted in Trejos & Villalobos (in press). Finally, some ideas of a genetic algorithm implementation are also undertaken; these ideas will be compared to the genetic algorithms of Mathar (1995) and N'Gouenet (1995), as well to the other methods already considered.

REFERENCES

AARTS, E.M. and J. KORST (1988): **Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Networks**, John Wiley & Sons, Chichester.

BORG, I. and P. GROENEN (1997): **Modern Multidimensional Scaling: Theory and Applications**, Springer, Berlín.

BOUROCHE, J.M. **et al**. (1977): **Analyse des Données en Marketing**, Masson, París.

CHARON, J.; A, GERMAN and O. HUDRY (1996): **Méthodes d'Optimisation Combinatoire**, Masson, París.

COX, T. and M. COX (1995): **Multidimensional Scalin**, Chapman & Hall, London.

De AMORIM, S.; J.P. BARTHELEMY and C. RIBEIRO (1992): "Clustering and clique partitioning: simulated annealing and tabu search approaches", **Journal of Classification** 9: 17-41.

De LEEUW, J. (1977): "Applications of convex analysis to multidimensional scaling", **Recent Developments in Statistics**, J.R. Barra **et al**. (Eds.), North Holland, Amsterdam.

_____ (1988): "Convergence of the majorization method for multidimensional scaling", **Journal of Classification** 5: 163-180.

DIDAY, E. **et al**. (1982): **Eléments d'Analyse des Données**, Dunod, Paris.

GREEN, P.E. **et al**. (1989): **MDS. Concepts and Applications**, Allyn Bacon, Boston.

GROENEN, P. (1993a): **The Majorization Approach to Multidimensional Scaling**, DSWO Press, Leiden.

_____ (1993b): "A comparison of two methods for global optimization in multidimensional scaling", **Information and Classification**, O. Opitz, B., Lausen & R. Klar (Eds.), Springer, Berlin.

HUBERT, L.; P. ARABIE and M. HENSON – Mc INNIS (1992): "Multidimensional scaling in the city-block metric: a combinatorial approach", **Journal of Classification** 9, 211-236.

KIRKPATRICK, S.; C.D. GELLAT and M.P. VECCHI (1983): "Optimization by simulated annealing", **Science** 220, 671-680.

KRUSKAL, J. (1964): "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", **Psychometrika** 29, 1-27.

MATHAR, R. and A. ŽILINSKAS (1993): "On global optimization in multidimensional scaling", **Acta Aplicandae Mathematicae** 33, 109-118.

MATHAR, R. (1995): "A genetic algorithm for multidimensional scaling", Internal Report, RWTH, Aachen.

N'GOUENET, R.F. (1995): Analyse Géométrique des Données de Dissimilarité par la Multidimensional Scaling: une Approche Paralléle Basée sur les Algorithmes Génétiques. Application aux Séquences Biologiques, Thèse de Doctorat, Université de Rennes I.

PLINER, V. (1996): "Metric unidimensional scaling and global optimization", **Journal of Classification** 13, 3-18.

SIMANTIRAKI, E: (1996): "Unidimensional scaling: a linear programming approach minimizing absolute deviations", **Journal of Classification** 13, 19-25.

TREJOS, J. (1993): "A simulated annealing implementation for oblique varimax rotations", **Applied Stochastic Models and Data Analysis**, II, J. Janssen & C.H. Skiadas (Eds.), World Scientific, Singapur: 981-989.

TREJOS, J.; A. MURILLO and E. PIZA (1998): "Global stochastic optimization for partitioning", **Advances in Data Science and Classification**, A. Rizzi **et al** (Eds.), Springer-Verlag, Berlín: 185-190.

TREJOS, J. and M. VILLALOBOS (in press): "Applications of tabu search in metric multidimensional scaling", to appear in **Investigación Operacional**.