Appl. Math. and Comp. Sci., 1997, Vol.7, No.3, 521–543

SUB-GRADIENT ALGORITHMS FOR SOLVING MULTI-DIMENSIONAL ANALYSIS PROBLEMS OF DISSIMILARITY DATA

ADNAN YASSINE *

The multi-dimensional scaling (MDS) problem, extensively addressed in data analysis, has been investigated in significant works (e.g. De Leeuw, 1977; 1988; De Leeuw and Pruzansky, 1976; Kruskal, 1964; Shepard, 1974). It consists in determination of a configuration x^* such that the matrix elements of distances between the points are required to be those of a given matrix called the proximity or dissimilarity matrix or, if this is impossible, it reduces to the nearest optimization problem in which a function (called the loss function) is to be minimized. In this paper, the stability and regularity of the Lagrangian duality in convex maximization (non-convex minimization) are considered. We present some convergence results of the DC (Difference of Convex functions) optimization algorithms which are based on DC duality and local optimality conditions for DC optimization. Various regularization techniques are studied in order to improve the quality (robustness, stability, convergence rate) of the DC algorithm (DCA). For solving MDS problems, sub-gradient algorithms (involving or not regularization techniques) in DC optimization are presented. Some numerical applications for large-scale problems are also provided.

1. Introduction

In recent years, active research has been conducted regarding the following class of non-convex and non-differentiable optimization problems:

$$(PNC): \ \inf\{g(x) - h(x) : x \in X\}$$

where g and h are convex, $X = \mathbb{R}^n$. The problem (PNC) is called the DC optimization problem and its particular structure makes significant developments in both qualitative and quantitative studies possible (e.g. Hiriart, 1989; Hiriart and Lemaréchal, 1990; Yassine, 1995).

As regards convex approaches to non-differentiable non-convex optimization (as opposed to combinatorial approaches to global optimization), we present here main results concerning DC optimization and algorithms for the DC optimization problems (DCA). The DC duality was firstly introduced by Toland (1979) in the context of

* Université Henri Poincaré (Nancy I), Institut Elie Cartan, UMR 9973, B.P. 239, 54506 Vandoeuvre-les-Nancy cedex, France, e-mail:yassine@loria.fr, yassine@iecn.u-nancy.fr.

variational calculus in mechanics, and generalized by Tao (1975; 1976) for convex maximization programming.

Owing to their relative simplicity of implementation, DCA's permit to solve large-scale real world DC optimization problems. Due to their local character, they cannot guarantee the globality of computed solutions to general DC optimization problems. In general, DCA converges to a local solution, but we observe quite often its convergence to a global one. A DC objective function has infinitely many decompositions which may exert strong influence on the quality (robustness, stability, rate of convergence and globality of sought solutions) of DCA.

The Multi-Dimensional Scaling problem (MDS) plays a leading role in statistics on account of various applications in different fields, e.g. social sciences (Bick *et al.*, 1977), biochemistry (Crippen, 1977; 1978), psychology (Levelt *et al.*, 1966), mathematical psychology (Beals *et al.*, 1977; Shepard, 1974), etc. A mathematical formulation of this important problem, due to Kruskal (1964a; 1964b), may be written down as follows:

Let ϕ be a norm on $\mathbb{R}^p$ and d be its corresponding distance. Let two symmetric matrices $\Delta = (\delta_{ij})$ and $W = (W_{ij})$ of order n be given such that

$$\delta_{ij} = \delta_{ji} > 0, \;\; W_{ij} = W_{ji} > 0 \quad \forall i \neq j, \quad \delta_{ii} = W_{ii} = 0 \quad \forall i = 1, \ldots, n$$

We call Δ the dissimilarity matrix and W the weight matrix. The MDS problem consists in finding n points $X_1, X_2, \ldots, X_n \in \mathbb{R}^p$ such that

$$d(X_i, X_j) \simeq \delta_{ij} \qquad \forall i, j = 1, \ldots, n$$

Denote by $M_{n,p}(\mathbb{R})$ the set of real $n \times p$ matrices. Let ϕ be a norm on $\mathbb{R}^p$. For all $X \in M_{n,p}(\mathbb{R})$ we consider the semi-norms $d_{ij}(X)$ defined by:

$$d_{ij}(X) = \phi\big[(X_{i.})^t - (X_{j.})^t\big]$$

where $X_{i.}$ denotes the i-th row of X. We set $\rho(X) = \sum_{i<j} W_{ij}\, \delta_{ij}\, d_{ij}(X)$, $\eta^2(X) = \sum_{i<j} W_{ij}\, d_{ij}^2(X)$ and $\eta_\delta^2 = \sum_{i<j} W_{ij}\, \delta_{ij}^2$.

Note that $d_{ij}(X)$ is convex, positively homogeneous and non-negative. Moreover, ρ and η are two semi-norms on $M_{n,p}(\mathbb{R})$. The MDS problem may be formulated as

$$(P_1): \;\; \min \left\{ \sigma(X) = \frac{1}{2} \sum_{i<j} W_{ij}\, (d_{ij}(X) - \delta_{ij})^2 \right\}$$

The problem (P_1) is equivalent to the DC optimization problem

$$(P_2): \;\; \min \left\{ \frac{1}{2}\eta^2(X) - \rho(X): \quad X \in M_{n,p}(\mathbb{R}) \right\}$$

Thus DCA can be directly used to solve (P_2). This is an important development in numerical methods for MDS.

The aim of this paper is to investigate the following optimization problems:

1. The stability of the Lagrangian duality in the problem of convex maximization

$$(P): \quad \max\{f(x) : \phi(x) \leq 1\}$$

where $f, \phi \in \Gamma_0(\mathbb{R}^n)$ are positively homogeneous and bounded everywhere. One of the conditions is assumed:

a) f is non-negative and ϕ has any norm on $\mathbb{R}^n$,

b) f and ϕ are two semi-norms on $\mathbb{R}^n$ such that $\phi^{-1}(0) \subset f^{-1}(0)$.

2. Duality and sub-gradient algorithms regularized in DC optimization, as well as their application to solving MDS problems.

The latter involves a study in the fields of DC optimization, sub-gradient methods, and regularization techniques (in order to accelerate the convergence of these methods). Applications generally concern large-scale numerical simulations related with multi-dimensional scale analysis.

In Section 2 the duality problem in DC optimization is presented in relation to sub-gradient algorithms. The stability of the Lagrangian duality in convex optimization is examined in Section 3. The multi-dimensional scaling problem for dissimilarity data (MDS) is discussed in Section 4. Section 5 presents comparative numerical experiments of solving the related MDS problem. Some final remarks and conclusions are given in Section 6.

2. DC Optimization

2.1. Introduction

Let $X = \mathbb{R}^n$. The dual space Y of X can then be identified with X itself ($Y \equiv \mathbb{R}^n$). Denote by $\Gamma_0(X)$ the cone of proper, lower semi-continuous, convex functions on X, and consider the following optimization problem:

$$(P): \quad \alpha = \inf\{g(x) - h(x) : x \in X\}, \quad g, h \in \Gamma_0(X)$$

Since g and h can become infinite simultaneously, we assume that $(+\infty) - (+\infty) = +\infty$ to avoid an indetermination problem.

The DC duality may be defined by using conjugate mappings g^* and h^* such that

$$(D): \quad \beta = \inf\{h^*(y) - g^*(y) : y \in Y\}$$

where $g^*(y) = \sup\{\langle x, y\rangle - g(x) : x \in X\}$ is the conjugate mapping of $g \in \Gamma_0(X)$ with values in $\Gamma_0(Y)$. Problem (D) is the dual of (P) and $\alpha = \beta$.

If α is finite, then $\operatorname{dom} g \subset \operatorname{dom} h$ and only the values of $g - h \in \operatorname{dom} g$ are involved in the search for global and local solutions to (P). This DC duality was first studied by Toland (1979) in a more general framework.

2.2. Duality in DC Optimization

Theorem 1. (Tao, 1986) *Let* $(\wp)$ *and* (Δ) *be the solution sets of problems* (P) *and* (D)*, respectively. Then:*

1. $\partial h(x) \subset \partial g(x) \qquad \forall x \in (\wp)$
2. $\partial g^*(y) \subset \partial h^*(y) \quad \forall y \in (\Delta)$
3. $\cup\{\partial g^*(y) : y \in (\Delta)\} \subset (\wp)$ *(an equality if* h *is sub-differentiable in* $(\wp)$*)*
4. $\cup\{\partial h(x) : x \in (\wp)\} \subset (\Delta)$ *(an equality if* g^* *is sub-differentiable in* (Δ)*)*

Definition 1. A point x^* of X is a *local minimum* of $(g-h)$ if $g(x^*)$ and $h(x^*)$ are bounded and if $g(x) - h(x) \geq g(x^*) - h(x^*)$ for each x in a neighbourhood U of x^*. Consequently, $\operatorname{dom} g \cap U \subset \operatorname{dom} h$.

Definition 2. A point x^* of X is a *critical point* of $(g-h)$ if $\partial h(x^*) \cap \partial g(x^*) \neq \emptyset$.

Theorem 2. *If a point* x^* *admits a neighbourhood* U *such that*

$$\partial h(x) \cap \partial g(x^*) \neq \emptyset \quad \forall x \in U$$

then $g(x) - h(x) \geq g(x^*) - h(x^*) \forall x \in U$ *(i.e.* x^* *is a local minimizer of* $g-h$*).*

Proof. We have $h(x^*) \geq h(x) + \langle x^* - x, y\rangle, \forall x \in X, \forall y \in \partial h(x)$. In particular $h(x) - h(x^*) \leq \langle x - x^*, y\rangle, \forall x \in U, \forall y \in \partial h(x) \cap \partial g(x^*)$. But $g(x) - g(x^*) \geq \langle x - x^*, y\rangle, \forall x \in U, \forall y \in \partial h(x) \cap \partial g(x^*)$. Hence

$$g(x) - g(x^*) \geq h(x) - h(x^*) \qquad \forall x \in U \qquad \blacksquare$$

If x^* is a local minimum of $(g-h)$, then $\partial h(x^*) \subset \partial g(x^*)$ (Hiriart, 1989). This necessary condition is also sufficient for several non-differentiable DC problems (Tao, 1981; 1984), in particular for a polyedral h (Hiriart, 1989).

The sub-gradient algorithms presented in the following enable us to obtain a point x^* such that $\partial h(x^*) \subset \partial g(x^*)$. The local minimum property of $g-h$ for x^* is likely.

Let $\wp_1$ (resp. Δ_1) be the set of points verifying the necessary conditions of local optimality for (P) (resp. for (D)), i.e.

$$\wp_1 = \{x \in X : \partial h(x) \subset \partial g(x)\}, \quad \Delta_1 = \{y \in Y : \partial g^*(y) \subset \partial h^*(y)\}$$

For every point x^* in X (resp. y^* in Y), the problems

$$S(x^*) = \inf \{h^*(y) - g^*(y) : y \in \partial h(x^*)\}$$

and

$$T(y^*) = \inf \{g(x) - h(x) : x \in \partial g^*(y^*)\}$$

are defined.

We denote by $s(x^*)$ (resp. $\tau(y^*)$) the set of solutions to $S(x^*)$ (resp. to $T(y^*)$).

Theorem 3. (Toland, 1979; Yassine, 1995)

$$x^* \in \wp_1 \text{ iff } y^* \in \Delta_1 \text{ s.t. } x^* \in \partial g^*(y^*)$$

$$y^* \in \Delta_1 \text{ iff } x^* \in \wp_1 \text{ s.t. } y^* \in \partial h(x^*)$$

Corollary 1. *If* $x^* \in \wp_1$ *(resp.* $y^* \in \Delta_1$*), then:*

i) $s(x^*) = \partial h(x^*)$ *(resp.* $\tau(y^*) = \partial g^*(y^*)$*)*

ii) $h^*(y) - g^*(y) = g(x^*) - h(x^*) \quad \forall y \in \partial h(x^*)$ *(resp.* $g(x) - h(x) = h^*(y^*) - g^*(y^*) \quad \forall x \in \partial g^*(y^*)$*)*

These results constitute the basis of DCA to be studied in Section 2.3. In general, DCA converges to a local solution. However, it would be interesting to formulate sufficient conditions for local optimality.

2.3. Sub-Gradient Algorithms (DCA Algorithms)

2.3.1. Complete Form

In the sub-gradient algorithm, two sequences $\{x^k\}$ and $\{y^k\}$ verifying Theorems 1 and 3 are constructed schematically as follows: Starting from any element x^0 of X, the algorithm creates two sequences $\{x^k\}$ and $\{y^k\}$ defined by

$$y^k \in s(x^k), \quad x^{k+1} \in \tau(y^k)$$

Theorem 4. (Toland, 1979) *Let us assume that the sequences* $\{x^k\}$ *and* $\{y^k\}$ *are well-defined. Then we have:*

1. $g(x^{k+1}) - h(x^{k+1}) \leq h^*(y^k) - g^*(y^k) \leq g(x^k) - h(x^k)$

   The equality $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ *is fulfilled iff* $x^k \in \partial g^*(y^k)$ *and* $y^k \in \partial h(x^k)$*. Then* $x^k \in \wp_1$ *and* $y^k \in \Delta_1$.

2. *If* α *is bounded, then*

   $\lim_{k \to +\infty} \{g(x^k) - h(x^k)\} = \lim_{k \to +\infty} \{h^*(y^k) - g^*(y^k)\} = \alpha^* \geq \alpha$

3. *If* α *is bounded and if the sequences* $\{x^k\}$ *and* $\{y^k\}$ *are bounded, then* $\forall x^* \in \Omega(x^k)$ *(resp.* $\forall y^* \in \Omega(y^k)$*) there exists* $y^* \in \Omega(y^k)$ *(resp.* $x^* \in \Omega(x^k)$*) such that:*

   i) $x^* \in \wp_1$ *and* $g(x^*) - h(x^*) = \alpha^* \geq \alpha$

   ii) $y^* \in \Delta_1$ *and* $h^*(y^*) - g^*(y^*) = \alpha^* \geq \alpha$

   iii) $\lim_{k \to +\infty} \{g(x^k) + g^*(y^k)\} = g(x^*) + g^*(y^*) = \langle x^*, y^* \rangle$

   iv) $\lim_{k \to +\infty} \{h(x^k) + h^*(y^k)\} = h(x^*) + h^*(y^*) = \langle x^*, y^* \rangle$

   where $\Omega(z^k)$ *is the set of the accumulation points of* $\{z^k\}$.

From a pratical point of view, although this algorithm uses a DC decomposition mentioned above, Problems $(S(x^k))$ and $(T(x^k))$ remain DC optimization programmes. Calculation of y^k and x^{k+1} is therefore still a difficult task. In practice, the sub-gradient algorithms are generally used on the simplified form presented in what follows.

2.3.2. Simple Form

Starting from an arbitrary point $x^0 \in X$, we define the two sequences $\{x^k\}$ and $\{y^k\}$ by taking

$$y^k \in \partial h(x^k), \quad x^{k+1} \in \partial g^*(y^k)$$

In this case, all the assumptions of Theorem 4 are still satisfied. Moreover, one could expect to obtain the properties $\partial h(x^*) \subset \partial g(x^*)$ and $\partial g^*(y^*) \subset \partial h^*(y^*)$, but we only have $\partial h(x^*) \cap \partial g(x^*) \neq \emptyset$ and $\partial g^*(y^*) \cap \partial h^*(y^*) \neq \emptyset$.

Definition 3. A function f is *strongly convex* on X if there exists a real $\rho > 0$ (called the coercivity coefficient) such that

$$f\left[\lambda x + (1-\lambda)y\right] \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\lambda(1-\lambda)}{2}\rho\|x-y\|^2 \ \ \forall\lambda \in [0,1]; \ \ \forall x, y \in X$$

Theorem 5. (Auslender, 1976) *If f is strongly convex on X, then there exists a real $\rho > 0$ such that*

$$f(x') \geq f(x) + \langle y, x' - x\rangle + \rho\|x' - x\|^2 \quad \forall x, x' \in X; \quad \forall y \in \partial f(x)$$

The converse is true if f is sub-differentiable.

Theorem 6. (Tao, 1986; Yassine, 1995) *Let us assume that g and h are strongly convex and the sequences $\{x^k\}$ and $\{y^k\}$ are well-defined. Then we get the following properties:*

1. $$\begin{aligned} g(x^{k+1}) - h(x^{k+1}) &\leq h^*(y^k) - g^*(y^k) - \rho_h\|x^{k+1} - x^k\|^2 \\ &\leq g(x^k) - h(x^k) - (\rho_h + \rho_g)\|x^{k+1} - x^k\|^2 \end{aligned}$$

   where ρ_h and ρ_g are the respective coefficients of coercivity related to h and g.

2. $$\begin{aligned} h^*(y^{k+1}) - g^*(y^{k+1}) &\leq g(x^{k+1}) - h(x^{k+1}) - \rho_{g^*}\|y^{k+1} - y^k\|^2 \\ &\leq h^*(y^k) - g^*(y^k) - (\rho_{h^*} + \rho_{g^*})\|y^{k+1} - y^k\|^2 \end{aligned}$$

   where ρ_{h^*} and ρ_{g^*} are the respective coefficients related to h^* and g^*.

Corollary 2. (convergence of the simple form)

1. $g(x^{k+1}) - h(x^{k+1}) = h^*(y^k) - g^*(y^k) \iff y^k \in \partial h(x^{k+1})$ *and* $x^{k+1} = x^k$
   In this case, we get $y^k \in \partial h(x^k) \cap \partial g(x^k)$.
2. $h^*(y^k) - g^*(y^k) = g(x^k) - h(x^k) \iff x^k \in \partial g^*(y^k)$ *and* $y^{k-1} = y^k$
   Here, we get $y^k \in \partial h(x^k) \cap \partial g(x^k)$.
3. *If* α *is bounded and the sequences* $\{x^k\}$ *and* $\{y^k\}$ *are bounded, then* $\forall x^* \in \Omega(x^k)$ *(resp.* $\forall y^* \in \Omega(y^k)$*) there exists* $y^* \in \Omega(y^k)$ *(resp.* $x^* \in \Omega(x^k)$*) such that:*

   i) $g(x^k) - h(x^k) = h^*(y^k) - g^*(y^k) \longrightarrow [h^*(y^*) - g^*(y^*)] = \alpha^* \geq \alpha$ *as* $k \to +\infty$

   ii) $y^* \in \partial h(x^*) \cap \partial g(x^*)$ *and* $x^* \in \partial h^*(y^*) \cap \partial g^*(y^*)$

   iii) $\lim_{k\to+\infty} \|x^{k+1} - x^k\| = 0$ *and* $\lim_{k\to+\infty} \|y^{k+1} - y^k\| = 0$

   where $\Omega(z^k)$ *is the set of the accumulation points of* $\{z^k\}$.

Proof. This result follows immediately from Theorems 5 and 6.

3. Stability of the Lagrangian Duality in Non-Convex Optimization

3.1. Problem Statement

Let $X = \mathbb{R}^n$ and Y be its dual space (i.e. $Y \equiv \mathbb{R}^n$). In this section, we consider the stability of the Lagrangian duality in convex optimization problems of the form:

$$(P): \ \max\{f(x) : \phi(x) \leq 1\}$$

where $f \in \Gamma_0(\mathbb{R}^n)$, and it is positively homogeneous, non identically zero, and ϕ is any of the norms on X.

The problem (P) (called the *primary problem*) may be formulated as follows:

$$(P): \ \min\left\{ -f(x) : \frac{1}{2}\phi^2(x) \leq \frac{1}{2}\right\}$$

The Lagrangian function related to (P) is defined by

$$L(x,\lambda) = \begin{cases} -f(x) + \frac{\lambda}{2}\{\phi^2(x) - 1\} & \text{if } \lambda \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

We define

$$(P_\lambda): \ g(\lambda) = \inf\left\{L(x,\lambda) : x \in \mathbb{R}^n\right\} = \inf\left\{ -f(x) + \frac{\lambda}{2}(\phi^2(x) - 1) : x \in \mathbb{R}^n\right\}$$

The dual problem (D) related to (P) may be written down as follows:

$$(D): \ \beta = \sup\{g(\lambda) : \lambda \geq 0\}$$

Such an aproach is motivated by the following rationales:

1. It provides significant additional information to characterize primary and dual solutions. Consequently, we are able to obtain the primary solution from the dual solution and vice-versa.

2. The study permits to use problem (P_λ) whose solution leads to that of (P).

3.2. Study of Problem (P_λ)

We denote by $(\wp_\lambda)$ (resp. P and D) the set of solutions of (P_λ) (resp. (P) and (D)).

Proposition 1.

1. $\operatorname{dom} g =]0, +\infty[$

2. $(\wp_\lambda) \subset \left\{x \in \mathbb{R}^n : g(\lambda) = -\frac{f(x)}{2} - \frac{\lambda}{2} = -\frac{\lambda}{2}(\phi^2(x) + 1)\right\}$

Proof.

1. If f is positively homogeneous and non identically zero, then

$$g(0) = \inf\{ - f(x) : x \in X\} = -\infty \tag{1}$$

Hence $\operatorname{dom} g \subset]0, +\infty[$.

If f is finite, then $\forall x \in X, \exists b > 0$ such that $f(x) \leq b \cdot \phi(x)$. We have

$$L(x, \lambda) = -f(x) + \frac{\lambda}{2}\left\{\phi(x)^2 - 1\right\} \geq -b \cdot \phi(x) + \frac{\lambda}{2}\left\{\phi(x)^2 - 1\right\}$$

and

$$\lim_{\phi(x) \to +\infty} L(x, \lambda) = \lim_{\phi(x) \to +\infty} \frac{\lambda}{2}\, \phi(x)^2 = +\infty$$

Consequently, $\operatorname{dom} g =]0, +\infty[$.

2. $x \in (P_\lambda) \Rightarrow 0 \in \partial_x L(x, \lambda) \Rightarrow 0 \in -\partial f(x) + \lambda\phi(x)\partial\phi(x) \Rightarrow \partial f(x) \subset \lambda\phi(x)\partial\phi(x)$. Then

$$\forall y \in \partial f(x), \forall z \in \partial\phi(x), \quad x^t y = \lambda\phi(x) x^t z \tag{2}$$

If f is positively homogeneous and ϕ is a norm on X, then $\forall y \in \partial f(x), \forall z \in \partial\phi(x)$

$$x^t y = f(x), \quad x^t z = \phi(x) \tag{3}$$

Combining (2) and (3), we get

$$f(x) = \lambda\phi(x)^2 \tag{4}$$

Hence

$$g(\lambda) = -\frac{\lambda}{2} - f(x) + \frac{\lambda}{2}\phi(x)^2 = -\frac{\lambda}{2} - \frac{f(x)}{2} = -\frac{\lambda}{2}\{1 + \phi(x)^2\}$$

■

Corollary 3. *Let x be a solution to $(\wp_\lambda)$. Then we have*

i) $\phi(x) > 1 \Rightarrow -\lambda > g(\lambda) > -f(x)$

ii) $\phi(x) < 1 \Rightarrow -\lambda < g(\lambda) < -f(x)$

iii) $\phi(x) = 1 \Rightarrow -\lambda = g(\lambda) = -f(x)$ *and* $x \in \boldsymbol{P}$, $\lambda \in \boldsymbol{D}$.

The above properties are direct consequences of Proposition 1.

Corollary 4. *For every $\lambda > 0$, we get the following three properties which are mutually excluded:*

i) $(\wp_\lambda) \subset \{x \in \mathbb{R}^n : \phi(x) > 1\}$

ii) $(\wp_\lambda) \subset \{x \in \mathbb{R}^n : \phi(x) < 1\}$

iii) $(\wp_\lambda) \subset \{x \in \mathbb{R}^n : \phi(x) = 1\}$

This corollary follows immediately from Proposition 1.

Remark 1. If $(\wp_\lambda)$ is a singleton, then the above corollary is trivial.

Theorem 7.

1. $g(\lambda) = -\frac{\lambda}{2} + \frac{k}{\lambda}$, *where k is a negative constant depending on f and ϕ.*

2. $\boldsymbol{D} = \{\lambda^*\} = \{\sqrt{-2k}\}$ *is a singleton and we get*

i) $(\wp_\lambda) \subset \{x \in \mathbb{R}^n : \phi(x) = 1\}$

ii) $\alpha = \beta = g(\lambda^*) = -f(x^*) = -\lambda^*$

3. $\boldsymbol{P} = (\wp_\lambda)$

Proof.

1. $\partial g(\lambda) \subset \mathrm{co}\left\{\frac{\phi^2(x) - 1}{2}\right\}$, hence $\nabla g(\lambda) = g'(\lambda) = \frac{\phi^2(x) - 1}{2} = -1 - \frac{g(\lambda)}{\lambda}$, and then $g(\lambda) = -\frac{\lambda}{2} + \frac{k}{\lambda}$, where k is a real to be determined.

2. and 3. Let λ^* and x^* be the dual and primary solutions, respectively. Therefore: $g'(\lambda^*) = -\frac{1}{2} - \frac{k}{\lambda^{*2}} = 0$, whence $k = -\frac{\lambda^{*2}}{2}$ and finally $\lambda^* = \sqrt{-2k}$. Futhermore, $g'(\lambda^*) = (\phi^2(x^*)-1)/2 = 0 \Rightarrow \phi^2(x^*) = 1 \Rightarrow f(x^*) = \lambda^*\phi^2(x^*) = \lambda^* \Rightarrow g(\lambda^*) = -\lambda^*$, and then $\beta = g(\lambda^*) = -f(x^*) = \alpha = -\lambda^*$, $\phi(x^*) = 1$ and $\lambda^* = \sqrt{-2k}$. ∎

3.3. Stability of the Lagrangian Duality when f and ϕ are Two Semi-Norms

Let us consider the primal problem (P) when f and ϕ are two semi-norms defined on $X = \mathbb{R}^n$ such that $\boldsymbol{N}(\phi) \subset \boldsymbol{N}(f)$. We want to prove that the previous results are still consistent for this class of problems. If $\boldsymbol{A} = \boldsymbol{N}(\phi)$ and $\boldsymbol{B} = \boldsymbol{A}^{\perp}$, then X can be expressed as $X = \mathbf{A} \oplus \boldsymbol{B}$ and the following properties are established:

$$\phi(x+a) = \phi(x), \quad f(x+a) = f(x) \quad \forall a \in \boldsymbol{A}$$

Consequently, (P) can be expressed as

$$(Q): \ \max \{f(x) : x \in \boldsymbol{B}, \phi(x) \le 1\}$$

The semi-norms f and ϕ on X are norms on $\boldsymbol{B}$, hence problem (Q) always possesses a solution. If we denote by $\boldsymbol{Q}$ the set of solutions to (Q), then $\boldsymbol{P} = \boldsymbol{Q} + \boldsymbol{A}$ (i.e. if x is a solution to (Q), then $x^* = x + a$ is a solution to (P) $\forall a \in \boldsymbol{A}$). In addition, we get

$$(P_\lambda): \ g(\lambda) = \inf \Big\{ -f(x) + \frac{\lambda}{2}(\phi^2(x) - 1) : x \in \mathbb{R}^n \Big\}$$

Based on the previous remarks, problem (P_λ) may be formulated as follows:

$$(P_\lambda): \ g(\lambda) = \inf \Big\{ -f(x) + \frac{\lambda}{2}\phi^2(x) : x \in \boldsymbol{B} \Big\} - \frac{\lambda}{2}$$

Since $g(\lambda) \le -\lambda/2$, we can consider the set $\boldsymbol{E} = \{x \in \boldsymbol{B} : -f(x) + (\lambda/2)\phi^2(x) \le 0\}$, which means that the last problem considered is equivalent to

$$(P_\lambda): \ g(\lambda) = \inf \Big\{ -f(x) + \frac{\lambda}{2}\phi^2(x) : x \in \boldsymbol{E} \Big\} - \frac{\lambda}{2}$$

Since f is bounded everywhere, $\exists b > 0$ such that $f(x) \le b\,\phi(x)$. Hence (P_λ) can be written down in the form

$$(P_\lambda): \ g(\lambda) = \inf \Big\{ -f(x) + \frac{\lambda}{2}\phi^2(x) : x \in \boldsymbol{E}_1 \Big\} - \frac{\lambda}{2}$$

where $\boldsymbol{E}_1 = \{x \in \boldsymbol{B} : \phi(x) \le 2b/\lambda,\ \lambda > 0\}$. Since $\phi(x)$ is a norm on $\boldsymbol{B}$, we deduce that $\boldsymbol{E}_1$ is bounded and the previous stability results can be applied.

3.4. Solution of Problem (P)

The idea consists in solving the intermediate problem (P_λ) for a given λ^0, which leads to a value of the constant $k = \lambda^0\{g(\lambda^0) + (\lambda^0/2)\}$, since $\lambda^* = \sqrt{-2k}$ is a solution to (D). Problem (P_{λ^*}) is solved again to obtain a solution to (P). We then get the following algorithmic scheme:

1. Choose any $\lambda^0 > 0$.
2. Solve

$$(P_{\lambda^0}): \ g(\lambda^0) = \inf\{L(x,\lambda^0) : x \in \mathbb{R}^n\}$$
$$= \inf\Big\{-f(x) + \frac{\lambda^0}{2}\big(\phi^2(x) - 1\big) : x \in \mathbb{R}^n\Big\}$$

3. Compute the constants $k = \lambda^0\{g(\lambda^0) + (\lambda^0/2)\}, \quad \lambda^* = \sqrt{-2k}.$
4. Solve

$$(P_{\lambda^*}): \ \inf\{L(x,\lambda^*) : x \in \mathbb{R}^n\} = \inf\Big\{-f(x) + \frac{\lambda^*}{2}(\phi^2(x) - 1) : x \in \mathbb{R}^n\Big\}$$

The solution x^* to (P_{λ^*}) is also a solution to (P).

4. Multi-Dimensional Analysis of Dissimilarity Data

4.1. Introduction

We consider a set of statistical units $\{o_1, o_2, \ldots, o_n\}$ (called the objects) and ask some subjects $\{g_1, g_2, \ldots, g_k\}$ (called the judges) to settle, under certain experimental conditions (called occasions) and starting from a pre-definite scale, a list of numerical values, representing the proximities between the different objects. This list is denoted by $\{\delta_{ijm}, (i,j,m) \in L{\times}K\}$, where L is a part of $\{1,\ldots,n\} \times \{1,\ldots,n\}$, $K = \{1,\ldots,k\}$ and δ_{ijm} denotes the proximity between the objects o_i and o_j defined by the judge g_m. The numerical problem to be solved by MDS methods consists in looking for a set of points $\{X_1^{(m)}, X_2^{(m)}, \ldots, X_n^{(m)}\}$, in k-Euclidean spaces (Ω_m, d) of finite (and minimum) dimensions, for $m = 1,\ldots,k$, in such a way that the proximity $d(X_i^{(m)}, X_j^{(m)})$ can be as close as possible to the real approximate proximity using the terms $\{\delta_{ijm}, (i,j,m) \in L \times K\}$ defined by the judges. The choice of the representation space (Ω, d) of the set $\{\delta_{ijm}\}$, gives a general idea of proximity of different objects, but it is also necessary to take into account the importance given by every judge to each dimension of the representation space ω. When considering several representation spaces $\{(\Omega_m, d), m = 1,\ldots,k\}$, one can get an idea about every judge involved by the definition of the list of data $\{\delta_{i,j,m}, (i,j,m) \in L{\times}K\}$.

Practical examples of applications of multi-dimensional scaling methods have come from social sciences (Bick *et al.*, 1977), biochemistry (Crippen, 1977; 1978), psychophysics (Levelt *et al.*, 1966), mathematical psychology (Shepard, 1974), and many others.

Example 1. To easily understand the MDS problem, here is a concrete and simple example: We try to find a metric configuration in $\mathbb{R}^p$ ($p = 2$ or $p = 3$) which represents the "relationship of friendship" existing between the students from the same group. Let G be a group of six students. To establish the level of friendship between these students, each of them (S_i for $i = 1, \ldots, 6$ i.e. $n = 6$) is asked to give a mark (out of 20), denoted by N_{ij}, to each student S_j. The results are shown in Table 1.

Table 1. Marks $N = (N_{ij})$, $i, j = 1, \ldots, 6$.

	S_1	S_2	S_3	S_4	S_5	S_6
S_1		18	16	12	10	13
S_2	16.4		15	11	13	14
S_3	13.2	13.5		18	14	10
S_4	15	15	12.5		11	12
S_5	11.5	11	11.5	15		11
S_6	11	10.5	14.5	13	14.5	

Let $W = (W_{ij})$, $i, j = 1, \ldots, 6$ be the weight matrix which corresponds to the normal case:

$$W_{ii} = 0 \quad \text{and} \quad W_{ij} = 1 \quad \forall i \neq j$$

and $\Delta = (\delta_{ij})$, $i, j = 1, \ldots, 6$ be defined by

$$\delta_{ij} = |N_{ij} - N_{ji}| \quad \forall i, j = 1, \ldots, 6$$

Table 2. $\Delta = (\delta_{ij})$, $i, j = 1, \ldots, 6$.

0	1.6	2.8	3	1.5	2
1.6	0	1.5	4	2	3.5
2.8	1.5	0	5.5	2.5	4.5
3	4	5.5	0	4	1
1.5	2	2.5	4	0	3.5
2	3.5	4.5	1	3.5	0

We need to find six points $X_1, X_2, \ldots, X_6 \in \mathbb{R}^p$, such that the differences $|d(X_i, X_j) - \delta_{ij}|$ for every i and j are minimum (d is the distance in $\mathbb{R}^p$).

4.2. Mathematical Modelling of MDS Problems

Let us consider the Euclidean space $\mathbb{R}^p$ endowed with the norm ϕ and the corresponding distance d. Two symmetric n-matrices $\Delta = (\delta_{ij})$ and $W = (W_{ij})$ are given such that

$$\delta_{ij} = \delta_{ji} > 0, \quad W_{ij} = W_{ji} > 0 \quad \forall i \neq j;\ \delta_{ii} = W_{ii} = 0 \quad \forall i = 1, \dots, n$$

(Δ is called the dissimilarity matrix and W is the weight matrix).

The MDS problem consists in finding n points $X_1, X_2, \dots, X_n \in \mathbb{R}^p$ such that

$$d(X_i, X_j) \simeq \delta_{ij} \quad \forall i, j = 1, \dots, n$$

We denote by $M_{n,p}(\mathbb{R})$ the set of real $n{\times}p$ matrices. For all $X \in M_{n,p}(\mathbb{R})$ we consider the semi-norms $d_{ij}(X)$ defined by

$$d_{ij}(X) = \phi\big[(X_{i.})^t - (X_{j.})^t\big]$$

We set $\rho(X) = \sum_{i<j} W_{ij}\, \delta_{ij}\, d_{ij}(X)$, $\eta^2(X) = \sum_{i<j} W_{ij}\, d_{ij}^2(X)$ and $\eta_\delta^2 = \sum_{i<j} W_{ij}\, \delta_{ij}^2$.

Note that $d_{ij}(X)$ is convex, positively homogeneous and non-negative. Moreover, ρ and η are two semi-norms on $M_{n,p}(\mathbb{R})$. The relevant MDS problem may be written down in the form

$$(P_1):\ \min\Big\{\sigma(X) = \frac{1}{2}\sum_{i<j} W_{ij}\,(d_{ij}(X) - \delta_{ij})^2\Big\}$$

We conclude easily that (P_1) can be expressed in the following forms:

$$(P_2):\ \min\Big\{\frac{1}{2}\eta^2(X) - \rho(X) : X \in M_{n,p}(\mathbb{R})\Big\}$$

$$(P_3):\ \max\Big\{\frac{\rho(X)}{\eta(X)} : \eta(X) \neq 0\Big\}$$

$$(P_4):\ \max\Big\{\rho(X) : \eta(X) \leq 1\Big\}$$

$$(P_5):\ \min\Big\{\chi_C(X) - \rho(X) : X \in M_{n,p}(\mathbb{R})\Big\}$$

where $C = \{X \in M_{n,p}(\mathbb{R}) : \eta^2(X) \leq 1\}$.

According to the previous considerations, the Lagrangian stability applied to (P_4) can be easily observed. In the following, we shall present some results where an explicit form of the dual function related to (P_4) is given. This permits to solve the MDS problem of the form (P_4) with a complete study of the regularization in terms of the function $\mu\|\cdot\|^2/2$. We also analyze the importance of these regularizations when applying the sub-gradient algorithms to numerical solutions of large-scale MDS problems in the form (P_2) and (P_5).

4.3. Solving Problem (P_2)

Proposition 5. (DE Leeuw, 1977; Tao, 1981) *Let*

$$B(X) = (B_{ij}) : B_{ij} = \begin{cases} \dfrac{-\delta ij}{d_{ij}(X)} & \text{if } i \neq j \text{ and } d_{ij}(X) \neq 0 \\ \displaystyle\sum_{k=1}^{n} \frac{\delta_{ik}}{d_{ik}(X)} & \text{if } i = j \text{ and } d_{ik}(X) \neq 0 \quad \forall k = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

and

$$V = (V_{ij}) : V_{ij} = \begin{cases} -W_{ij} = -1 & \text{if } i \neq j \\ \displaystyle\sum_{k=1}^{n} W_{ik} = n-1 & \text{otherwise} \end{cases}$$

Then

1. *$B(X)$ and V are two symmetric, positive semi-definite, centred, $n \times n$ matrices of rank $\leq (n-1)$*
2. $\rho(X) = \text{tr}[X^t \cdot B(X) \cdot X]$ *and* $\frac{1}{2}\eta^2(X) = \text{tr}[X^t \cdot V \cdot X]$
3. $B(X) \cdot X \in \partial\big(\rho(X)\big)$ *and* $V \cdot X \in \partial\Big(\frac{1}{2}\eta^2(X)\Big)$.

Let $\boldsymbol{A}$ be the vector sub-space of constant configurations and $\boldsymbol{B} = \boldsymbol{A}^{\perp}$ the vector sub-space of centred configurations. Then $M_{n,p}(\mathbb{R}) = \boldsymbol{A} \oplus \boldsymbol{B} = \boldsymbol{A} \oplus \boldsymbol{A}^{\perp}$ and

$$\forall X \in \boldsymbol{A},\ \rho(X) = \eta(X) = 0 \Rightarrow \boldsymbol{A} = \rho^{-1}(0) = \eta^{-1}(0)$$

Since ρ and η are two semi-norms, we have $\rho(X + A) = \rho(X)$ and $\eta(X + A) = \eta(X)\, \forall A \in \boldsymbol{A}$, and if x^* is a solution to (P_2), then $x^* + \boldsymbol{A}$ is the set of solutions to (P_2). Futhermore, if $x^* \in M_{n,p}(\mathbb{R})$ is a solution to (P_2), then $x^*_{\boldsymbol{B}}$ is a solution to (P_2) and the space $\boldsymbol{B} = \boldsymbol{A}^{\perp}$ is to be considered.

To solve (P_2), we apply sub-gradient methods with and without regularization. We denote by DCA1 (resp. RDCA1) the sub-gradient algorithm without (resp. with) regularization techniques, introduced in order to solve (P_2). This means that

$$DCA1: \ \min\Big\{\frac{1}{2}\eta^2(X) - \rho(X) : X \in M_{n,p}(\mathbb{R})\Big\}$$

$$RDCA1: \ \min\Big\{\Big(\frac{\lambda}{2}\eta^2(x) + \frac{\mu}{2}\|X\|^2\Big) - \Big(\lambda\rho(X) + \frac{\mu}{2}\|X\|^2\Big) : X \in M_{n,p}(\mathbb{R})\Big\}$$

where λ is a strictly positive real number used to accelerate the convergence and μ is a strictly positive real number called the regularization parameter.

The algorithm DCA1 consists in defining two sequences $\{X^k\}$ and $\{Y^k\}$ such that

$$X^k \longrightarrow Y^k \in \partial\rho(X^k) \longrightarrow X^{k+1} \in \partial\Big(\frac{1}{2}\eta^2\Big)^*(Y^k),\ \ Y^k \in \partial\Big(\frac{1}{2}\eta^2(X^{k+1})\Big)$$

We get

$$X^k \longrightarrow Y^k = B(X^k)\cdot X^k \longrightarrow X^{k+1} = \Big(V + \frac{1}{n}e\cdot e^t\Big)^{-1}\cdot B(X^k)\cdot X^k$$

Then the calculation of the new iterate X^{k+1} requires to solve the linear system

$$\Big(V + \frac{1}{n}e\cdot e^t\Big)\cdot X^{k+1} = B(X^k)\cdot X^k$$

Since $V = nI - e\cdot e^t$, where $e = (1,1,\dots,1)^t$ and $ee^t\cdot X^{k+1} = \sum_{i=1}^{n} X_i^{k+1} = \sum_{i=1}^{n} Y_i^k = 0$ because the matrix Y^k is centred, we have $V\cdot X^{k+1} = n\cdot X^{k+1}$, which leads to

$$X^{k+1} = \frac{Y^k}{n} = \frac{B(X^k)\cdot X^k}{n}$$

The algorithm RDCA1 is the regularization of DCA1 applied to the problem. It consists in defining two sequences $\{X^k\}$ and $\{Y^k\}$ such that

$$X^k \longrightarrow Y^k \in \partial\Big(\lambda\rho + \frac{\mu}{2}\|\cdot\|^2\Big)(X^k) \longrightarrow X^{k+1} \in \partial\Big(\frac{\lambda}{2}\eta^2 + \frac{\mu}{2}\|\cdot\|^2\Big)^*(Y^k)$$

The calculations lead to the following formula:

$$X^k \longrightarrow Y^k = \lambda\cdot B(X^k)\cdot X^k + \mu\cdot X^k \longrightarrow Y^k = \lambda\cdot V\cdot X^{k+1} + \mu\cdot X^{k+1}$$

Hence

$$X^{k+1} = \frac{Y^k}{\mu + \lambda n}$$

4.4. Solving Problem (P_4)

The Lagrangian function related to (P_4) is defined by

$$L(X,\lambda) = \begin{cases} \frac{\lambda}{2}\{\eta^2(X) - 1\} - \rho(X) & \text{if } \lambda \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

We denote by **Dual1** the algorithm for solving (P_4), in which we use the stability of the Lagrangian duality. Its steps may be summarized as follows (cf. Section 4.4):

1. Select any $\lambda^0 \geq 0,\ k = 0$.
2. Solve (P_{λ^0}) : $\min\big\{L(X,\lambda^0) : X \in M_{n,p}(\mathbb{R})\big\}$.

3. Let X_1 be the solution to (P_{λ^0}). Then $k = \lambda^0\{(\lambda^0/2)\eta^2(X_1) - \rho(X_1)\}$ and $\lambda^* = \sqrt{-2k}$. In particular, if $\lambda^0 = 1$, then $k = (1/2)\eta^2(X_1) - \rho(X_1)$ and $\lambda^* = \sqrt{-2k}$.

4. Solve $(P_{\lambda^*}):\ \min\{L(X,\lambda^*) : X \in M_{n,p}(\mathbb{R})\}$.

The solution X^* to (P_{λ^*}) is a solution to (P_4) which verifies the conditions

$$\rho(X^*) = \lambda^* = -g(\lambda^*), \quad \eta(X^*) = 1$$

4.5. Procedure for Solving Problem (P_5)

We denote by DCA2 and RDCA2 the respective sub-gradient algorithms, without and with the regularization technique, for solving (P_5). Of course, (P_5) is of the same type as (P_2). We can write

$$DCA2:\ \min\{\chi_C(X) - \rho(X) : X \in M_{n,p}(\mathbb{R})\}$$

$$RDCA2:\ \min\Big\{[\lambda\cdot\chi_C(X) + \frac{1}{2}\|X\|^2] - [\lambda\cdot\rho(X) + \frac{1}{2}\|X\|^2)] : X \in M_{n,p}(\mathbb{R})\Big\}$$

The algorithm DCA2 consists in defining two sequences $\{X^k\}$ and $\{Y^k\}$ such that

$$X^k \longrightarrow Y^k \in \partial\rho(X^k) \longrightarrow X^{k+1} \in \partial(\chi_C)^*(Y^k),\ \ Y^k \in \partial(\chi_C)(X^{k+1})$$

The calculations yield

$$X^k \longrightarrow Y^k = B(X^k)\cdot X^k \longrightarrow Y^k = \lambda\cdot V\cdot X^{k+1},\ \ \lambda \geq 0,\ \ \eta(X^{k+1}) = 1$$

Hence

$$X^{k+1} = \frac{V^+\cdot Y^k}{\eta(V^+\cdot Y^k)} = \frac{B(X^k)\cdot X^k}{\eta[B(X^k)\cdot X^k]}$$

The algorithm RDCA2 involves definition of two sequences $\{X^k\}$ and $\{Y^k\}$ such that

$$X^k \longrightarrow Y^k \in \partial\Big(\lambda\rho + \frac{\mu}{2}\|\cdot\|^2\Big)(X^k) \longrightarrow X^{k+1} \in \partial(\mu I + \lambda\cdot\partial\chi_C)^*(Y^k)$$

Then we get

$$X^k \longrightarrow Y^k = \lambda\cdot B(X^k)\cdot X^k + \mu X^k \longrightarrow X^{k+1} = \mathrm{proj}_C(Y^k)$$

Hence

$$X^{k+1} = \begin{cases} Y^k & \text{if } \eta(Y^k) \leq 1 \\ Y^k/\eta(Y^k) & \text{otherwise} \end{cases}$$

5. Numerical Results

In this section we present some comparative numerical examples which concern all the algorithms applied when solving MDS problems. Two different cases can be considered, according to the choice of the dissimilarity matrix:

Case 1. The dissimilarities are considered as the distances between points of the vector space (in that case, it is known that the optimal value is zero).

Case 2. The dissimilarity data are positive real numbers (*a priori*, the optimal value is unknown).

The following examples related to large-scale MDS problems correspond to simulations tested on a SUN workstation. The examples enable us to compare the different methods. In Tables 3–6, $\sigma(X^*)$ represents the optimal value of the function $\sigma(X) = (1/2)\sum_{i<j} W_{ij}\,[\delta_{ij} - d_{ij}(X)]^2$ at a minimum point X^* and $g(X^*)$ denotes the norm of the gradient at X^*.

Table 3. Number of iterations (Case 1).

dim	DCA1	RDCA1	Dual1	RDCA2	DCA2	$\sigma(X^*)$	$g(X^*)$
50	95	99	97	95	95	28E-8	83E-7
100	100	102	102	100	100	11E-6	63E-7
200	105	106	107	105	105	44E-6	62E-7
300	108	109	110	108	108	10E-5	62E-7
400	110	111	112	110	110	17E-5	64E-7
500	112	112	114	112	112	28E-5	60E-7
600	113	114	115	113	113	39E-5	65E-7
800	115	116	118	116	115	71E-5	82E-7
1000	117	117	119	117	117	10E-4	93E-7
1200	119	119	121	119	119	12E-4	82E-7
1500	120	120	122	120	120	25E-4	94E-7

Table 4. Number of iterations (Case 2).

dim	DCA1	RDCA1	Duall	RDCA2	DCA2	$\sigma(X^*)$	$g(X^*)$
50	44	44	46	44	44	6.4798	79E-7
100	48	48	50	48	48	21.919	95E-7
200	52	52	54	52	52	87.676	86E-7
300	55	55	57	55	55	197.272	48E-7
400	56	57	58	56	56	350.707	80E-7
500	58	58	60	58	58	547.980	85E-7
600	59	59	61	59	59	711.404	87E-7
800	61	61	63	61	61	1302.633	78E-7
1000	62	62	64	62	62	2069.191	49E-7
1200	63	63	65	63	63	2845.619	87E-7
1500	64	64	66	64	64	4553.558	92E-7

Table 5. Computing time (Case 1).

dim	DCA1	RDCA1	Duall	RDCA2	DCA2
50	1.0	1.5	1.7	1.9	1.9
100	4.5	6.5	7.0	7.8	8.0
200	20	27	30	31	32
300	46	64	72	73	75
400	84	116	129	130	133
500	134	179	207	220	225
600	227	270	310	325	332
800	360	455	580	599	610
1000	535	730	765	812	835
1200	855	1129	1312	1391	1424
1500	1353	1751	2142	2234	2297

The number of iterations given in Tables 3 and 4 does not provide complete information taking into account the complexity of calculations for each iterative process. Consequently, it is interesting to present the CPU time related to each method applied to the problem MDS (see Tables 5 and 6).

Table 6. Computing time (Case 2).

dim	DCA1	RDCA1	Dual1	RDCA2	DCA2
50	0.5	0.9	0.9	0.9	0.9
100	2.5	4.0	4.0	4.0	4.0
200	11.5	14	16	17	17.5
300	27	36	38	40	41
400	50	68	72	75	76
500	75	110	116	121	123
600	115	153	170	180	182
800	216	298	314	328	331
1000	343	470	500	525	529
1200	516	718	755	785	790
1500	885	1183	1252	1296	1302

The graph of the function

$$g(\lambda) = -\frac{\lambda}{2} + \frac{k}{\lambda}$$

is given in Figs. 1–4, for several different cases.

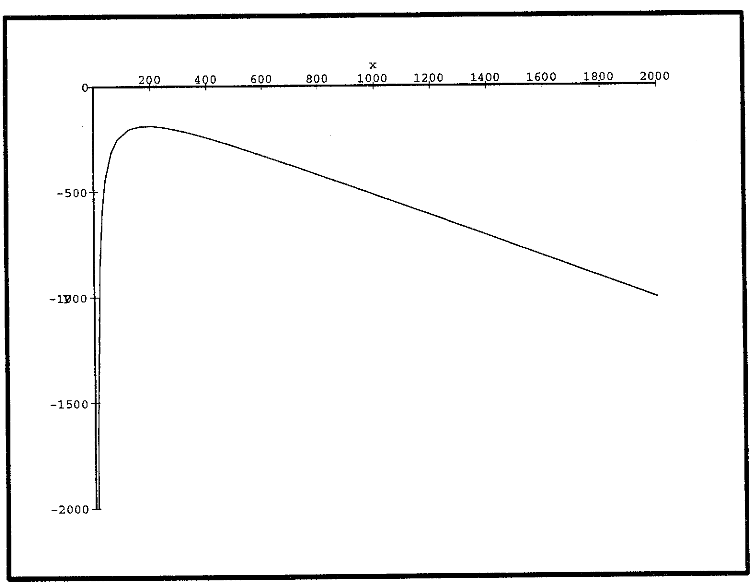

Fig 1. Graph of g (Case 1) for the dimension 1000, $\lambda^* = 187.8604$.

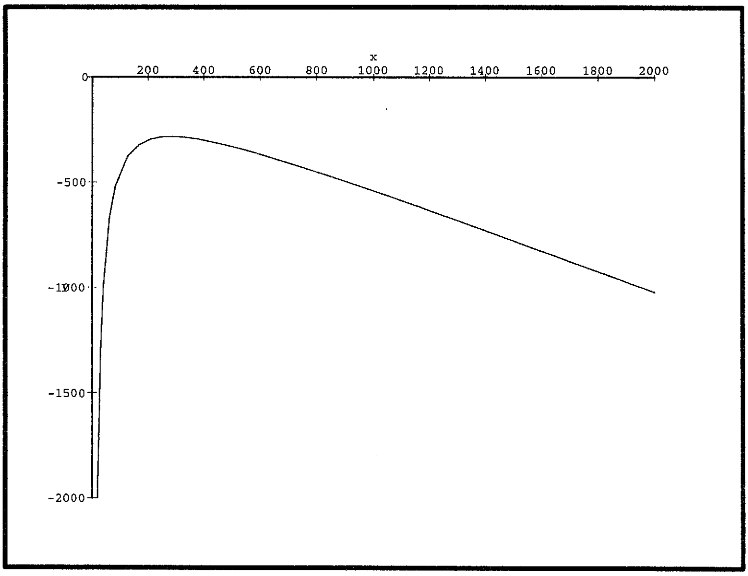

Fig 2. Graph of g (Case 1) for the dimension 1500, $\lambda^* = 281.7905$.

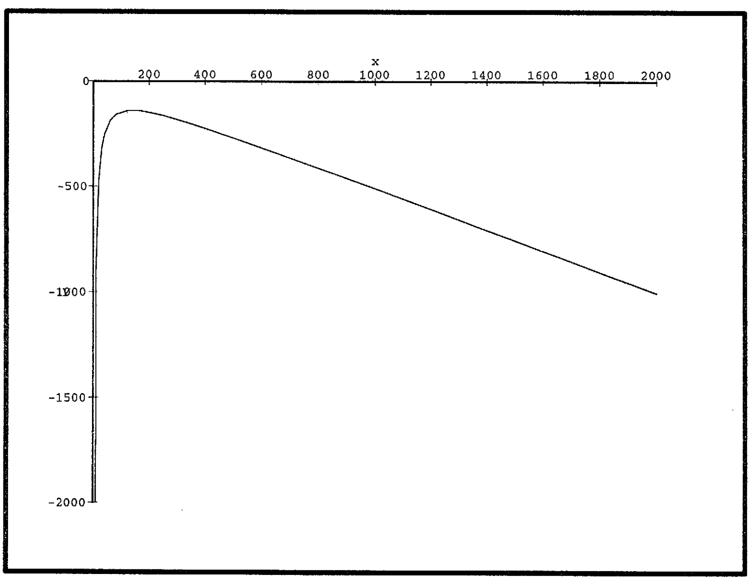

Fig 3. Graph of g (Case 2), for the dimension 1000, $\lambda^* = 138.4307$.

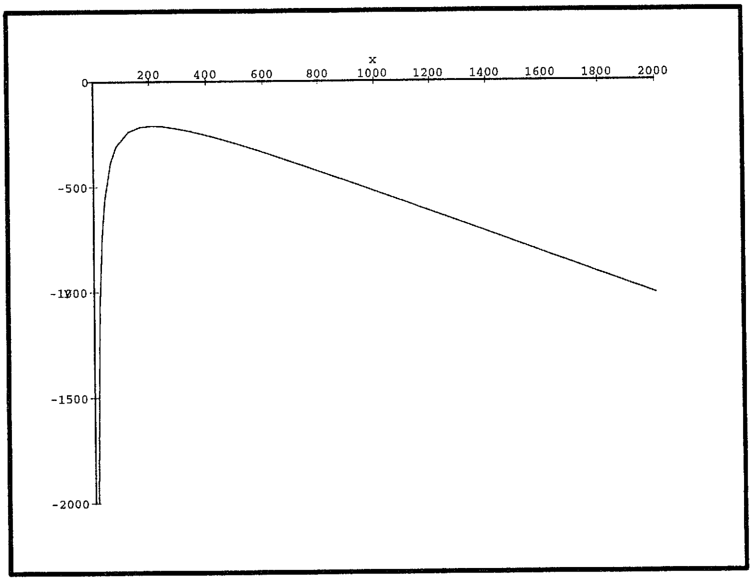

Fig 4. Graph of g (Case 2), for the dimension 1500, $\lambda^* = 210.1834$.

6. Concluding Remarks

The following comments regarding the present study can be made:

- The numerical results do not provide a complete comparison if we consider only the number of iterations, because the complexity of each iteration is not the same. The presentation of results also requires considering the computing time related to each method. This additional information enables us to evaluate more accurately the quality of each algorithm.
- The regularization applied to problem DCA2 slightly improves the speed of the computing algorithm. When applied to problem DCA1, it does not provide any advantage.
- The number of iterations for sub-gradient methods increases slightly with the size of the problem. The computing time turns out to be increasing exponentially.
- Several numerical tests revealed that the best values of λ and μ for regularization of DCA1 and DCA2 are $\lambda^* = \mu^* = 1$, except for RDCA1 in Case 2, where the best values of λ and μ are found to be $\lambda^* = 1$ and μ^* such that $50 < \mu < 150$. A very slow convergence is obtained if $0 < \mu < 1$ or $\mu > 150$.
- Numerous computer experiments show that DCA1 is the best algorithm for solving problem MDS, followed by its regularization algorithm RDCA1 and the algorithm Dual1. As regards the two other sub-gradient algorithms (DCA2 and

RDCA2), there are no significant differences in relation to the computing time or the number of iterations.

- The solutions obtained for Case 1 are global, since the optimal value is zero.

References

Auslender A. (1976): *Optimisation, méthodes numériques.* — Paris: Masson.

Beals R., Krantz D.H. and Tversky A. (1968): *Foundation of multidimensional scaling.* — Psychol. Rev., Vol.75, pp.127–142.

Bick W., Bauer H., Mueller P.J. and Gieseke O. (1977): *Multidimensional Scaling and Clustering Techniques (Theory and Applications in the Social Science).* — Institut fur Angewandte Sozialforchung, Universitat zu Koln.

Chine A. (1991): *Algorithmes robustes en optimisation non convexe. Codes et simulations numériques en grande dimension.* — Thèse de Doctorat, Université Joseph Fourier, Grenoble.

Crippen G.M. (1977): *A novel approach to calculation of conformation: distance geometry.* — J. Comp. Phys. Vol.24, pp.96–107.

Crippen G.M. (1978): *Rapid calculation of coordinates from distance matrices.* — J. Computational Phys. Vol.26, pp.449–452.

DE Leeuw J. and Pruzansky S. (1976): *A new computational method to fit the weighted Euclidean distance model.* — Report of University of Leiden, Germany.

DE Leeuw J. (1977): *Applications of convex analysis to multidimensional scaling*, In: Recent Developments in Statistics (J.R. Barra *et al.*, Ed.). — North-Holland: Elsevier, pp.133–145.

DE Leeuw J. (1988): *Convergence of the majorization method for multidimensional scaling.* — J. Classification, Vol.5, pp.163–180.

Guttman L. (1968): *A general nonmetric technique for finding the smallest coordinate space for a configuration of points.* — Psychometrika, Vol.33, pp.469–506.

Hiriart Urruty J.B. (1989): *How to regularize a difference of two functions?.* — Seminar of numerical analysis, Université Paul Sabatier, Toulouse.

Hiriart Urruty J.B and Lemaréchal C. (1990): *Testing necessary and sufficient conditions for global optimality in the problem of maximizing a convex quadratic over a convex polyhedron.* — Seminar of numerical analysis, Université Paul Sabatier, Toulouse.

Kruskal J.B. (1964a): *Multidimensional scaling by optimizing goodness-of-fit to a non metric hypothesis.* — Psychometrika, Vol.29, pp.1–28.

Kruskal J.B. (1964b): *Nonmetric multidimensional scaling: a numerical method.* — Psychometrika, Vol.29, pp.115–129.

Levelt W.J.M., Van de Geer J.P. and Plomp R. (1966): *Tridiac comparisons of musical intervals.* — British J. Math. Statist. Psych., Vol.19, pp.163–179.

Lingoes J.C. and Roskam E.E. (1973): *A mathematical and empirical analysis of two multidimensional scaling algorithms.* — Psychometrika, Vol.38, monograph supplement.

Shepard R.N. (1974): *Representation of structure in similarity data, problem and prospects.* — Psychometrika, Vol.39, pp.373–421.

Tao P.D. (1975): *Eléments homoduaux d'une matrice A relatifs à un couple des normes* (ϕ, ψ). *Applications au calcul de* $S_{\phi\psi}(A)$. — Séminaire d'analyse numérique, Grenoble, No.236.

Tao P.D. (1976): *Calcul du maximum d'une forme quadratique définie positive sur la boule unité de* ψ_∞. — Séminaire d'analyse numérique, Grenoble, No.247.

Tao P.D. (1981): *Contribution à la théorie de normes et ses applications à l'analyse numérique.* — Thèse de Doctorat d'Etat Es Sciences, USMG, Grenoble.

Tao P.D. (1984): *Convergence of subgradient method for computing the bound norm of matrices.* — Linear Alg. Its Appl., Vol.62, pp.163–182.

Tao P.D. (1986): *Algorithms for solving a class of non convex optimization problems. Methods of subgradients*, In: Fermat Days 85. Mathematics for Optimization (J.B. Hiriart Urruty, Ed.). — North-Holland: Elsevier.

Toland J.F. (1979): *A duality principle for non-convex optimization and calculus of variations.* — Arch. Rational. Mech. Analysis. Vol.71.

Yassine A. (1989): *Etudes adaptatives et comparatives de certains algorithmes en optimisation. Implémentations effectives et applications.* — Thèse de Doctorat, Université Joseph Fourier, Grenoble.

Yassine A. (1995): *Algorithmes de sous-gradients pour calculer les valeurs propres extrêmes d'une matrice symétrique réelle.* — Rapport de Recherche $n^\circ 5$, les Prépublications de l'Institut Élie Cartan, Nancy, France.

Received: 20 February 1997
Revised: 18 June 1997