Publicacions Matemàtiques, Vol 32 (1988), 275–287.

ON THE BEST CHOICE OF A DAMPING SEQUENCE IN ITERATIVE OPTIMIZATION METHODS

L.N. VASERSTEIN

Abstract

Some iterative methods of mathematical programming use a damping sequence $\{\alpha_t\}$ such that $0 \leq \alpha_t \leq 1$ for all t, $\alpha_t \to 0$ as $t \to \infty$, and $\sum \alpha_t = \infty$. For example, $\alpha_t = 1/(t+1)$ in Brown's method for solving matrix games. In this paper, for a model class of iterative methods, the convergence rate for any damping sequence $\{\alpha_t\}$ depending only on time t is computed. This computation is used to find the best damping sequence.

1. Iterative procedure

Let L be a real affine space (so L with an origin fixed is the same as a real vector space). For any points x and y in L, let $[x, y]$ denote the closed interval with the ends x and y. For any real–valued function f on a subset X of L, let $f(X)$ denote its infimum on X.

On a non-empty subset X of L, we consider an iterative procedure of the form

$$x_t = (1-\alpha_t)x_{t-1} + \alpha_t y_t = x_{t-1} + \alpha_t(y_t - x_{t-1}), \tag{1}$$

where $0 \leq \alpha_t \leq 1$, $1 \leq t < T+1$, and $[y_t, x_{t-1}] \subset X$.

Here the total number T of iterations is either finite or infinite ($T = \infty$); in the second case t runs over all natural numbers.

The objective of the procedure, starting at a point x_0 of X, is to minimize a convex bounded from below function f on X. (We call f convex on X, if its restriction on every interval contained in X is convex).

To reach this objective, at each step t, one tends to choose y_t in X, so that f decreases when one starts to move from x_{t-1} to y_t. The choice of y_t depends, in general, on f, x_{t-1}, and t. We abstract ourselves from any concrete rule of choosing y_t, and just assume that the choice was good enough. Namely, we fix a number θ in the interval $0 \leq \theta < 1$ and consider the class of iterative methods such that

$$f([x_{t-1}, y_t]) - f(X) \leq \theta(f(x_{t-1}) - f(X)) \tag{2}$$

for all integers t in the interval $1 \leq t < T+1$.

Note that according to (1), after a good direction $y_t - x_{t-1}$ is chosen, we do note minimize f on the interval $[x_{t-1}, y_t]$, but make a step from x_{t-1} in direction to y_t, with a "stepsize" α_t depending only on t.

Iterative procedures of the form (1) can be used not only for minimization of convex functions (see, for example, [4]). Sometimes they can be used for minimization of a not necessary convex bounded from below function g on X, because, for an arbitrary g, its infimum $g(X)$ is equal to $f(X)$, where f is the largest convex function on X such that $f \leq g$ everywhere on X. This f exists for any g, because the supremum of any set of convex functions on an interval is convex. This approach is feasible, if directions satisfying the condition (2) can be easily chosen.

Also the procedures of the form (1) can be used to search for a convex subset X_∞ of X. For example, this X_∞ could be a point where a function on X reaches a critical value. The search for X_∞ can be reduced to minimization of a convex function f as follows. Pick a distance ρ on L invariant under all translations and such that $\rho(x, x+(y-x)\alpha) = \rho(x,y)\alpha$ for all x and y in L and all real numbers $\alpha \geq 0$. (So, when an origin 0 in L is fixed, $(L, \rho(0,\cdot))$ is a linear normed space in the sense of Day [3].) Then $f = \rho(X_\infty, \cdot)$ is a convex non-negative function on L and X_∞ consists of the points which minimize f.

The condition (2) for such f takes the form

$$(3) \qquad \rho(X_\infty, [x_{t-1}, y_t]) \leq \theta\rho(X_\infty, x_{t-1}) \text{ for all } t, 1 \leq t < T+1.$$

The distance $\rho(Y, Z)$ between two subsets of a metric space is defined to be the infimum of all $\rho(y, z)$, where $y \in Y$ and $z \in Z$.

Speaking of the convergence rate, minimization on the interval $[x_{t-1}, y_t]$ under the condition (2) would give the exponential convergence

$$f(x_t) - f(X) \leq \theta^t (f(x_0) - f(X)).$$

Avoiding computation of stepsize (there is no line search in (1)), we will obtain (for the best damping sequence) a slower convergence

$$f(x_t) - f(X) \leq C(1-\theta)^{-2} t^{-1} (\text{see Theorem 3 } (b,c) \text{ below}).$$

One cannot get a better convergence assuming that X is convex, and f is a convex function defined on the whole L (see the remarks in Sections 2 and 3 below).

Slow convergence of methods of the form (1) is sometimes compensated by their resistance to errors and data perturbations. The methods can be useful when data are uncertain and a precise solution is not feasible. See, for example, Belen'ky et al [1], where Robinson [5] result on the convergence of Brown's method [2] is generalized and applications to linear programming are given.

2. Convergence when a damping sequence is fixed

We fix the total number $T \geq 1$ (T is an integer or ∞) of iterations, and a real number F. We impose the following condition on the function f and the procedure (1):

$$(4) \qquad f(y_t) - f(X) \leq F \text{ for } 0 \leq t < T+1, \quad \text{where } y_0 = x_0.$$

When f is bounded from above (as well as from below) this condition holds automatically for a sufficiently large F.

Theorem 1. *Fix θ in the interval $0 \leq \theta < 1$, $T \leq \infty$, $F > 0$, and a sequence $\{\alpha_t\} \in [0,1]^T$. Set $d_0 := 1$ and*

$$d_t := \max(d_{t-1}(1 - \alpha_t + \alpha_t\theta), \theta d_{t-1}(1-\alpha_t) + \alpha_t), \quad 1 \leq t < T+1.$$

Then for any L and X, any f convex on X, and any iterative procedure (1) *satisfying the conditions* (2) *and* (4), *we have*

$$f(x_t) - f(X) \leq Fd_t, \quad 0 \leq t < T+1.$$

Moreover, there are L, X, f as above and a procedure (1) *satisfying* (2), (4) *such that for all t*

$$f(x_t) - f(X) = Fd_t.$$

Proof: Note that a "procedure (1)"is determined by a starting point x_0 and a sequence y_t in X such that $[y_t, x_{t-1}] \subset X$, since the sequence $\{\alpha_t\}$ is fixed. We prove the first conclusion by induction on t. When $t = 0$, $f(x_0) - f(X) \leq Fd_0$ by (4).

Let now $t \geq 1$ and we have proved that $f(x_{t-1}) - f(X) \leq Fd_{t-1}$. The function $g(\alpha) := f(x_{t-1} + \alpha(y_t - x_{t-1}))$ on the interval $[0,1]$ is convex and $g(0) = f(x_{t-1}) \leq f(X) + Fd_{t-1}$. Moreover $g(1) = f(y_t) \leq f(X) + F$ by (4) and $g([0,1]) \leq f(X) + \theta Fd_{t-1}$ by (2).

It follows that

$$f(x_{t-1} + \alpha(y_t - x_{t-1})) = g(\alpha) \leq$$
$$\leq \max((1-\alpha)g([0,1]) + \alpha g(1), (1-\alpha)g(0) + \alpha g([0,1])) \leq$$
$$\leq f(X) + F\max((1-\alpha)\theta d_{t-1} + \alpha, (1-\alpha)d_{t-1} + \alpha\theta d_{t-1}) \text{ for all } \alpha.$$

In particular, when $\alpha = \alpha_{t'}$ we obtain our conclusion: $f(x_t) - f(X) \leq Fd_t$. See Figure 1, where $f(X) = 0$.

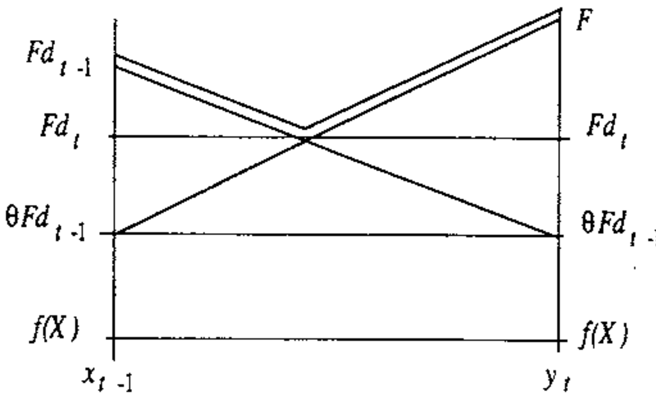

Figure 1: Upper bound for f on the interval $[x_{t-1}, y_t]$

Now we want to construct L, X, f, and a procedure such that $f(x_t) - f(X) = = Fd_t$ for all t. Let L be a $(T+1)$–dimensional real affine space, spanned by its $T+2$ points $y_t, -1 \leq t < T+1$, in general position (so y_t does not belong to the affine subspace spanned by all y_s with $s < t$, where $-1 < t < T+1$). Set $x_0 := y_0$ and $x_t := x_{t-1} + \alpha_t(y_t - x_{t-1})$ for $1 \leq t < T+1$. Set $X := \{y_{-1}\} \cup (\cup_{1 \leq t < T+1}[x_{t-1}, y_t])$. Set $f(y_{-1}) = 0$ and $f(y_0) = F$. For any α in the interval $0 < \alpha \leq 1$ and any t in the interval $0 \leq t < T+1$, we set

$$f((1-\alpha)x_t + \alpha y_{t+1}) = \begin{cases} (1-\alpha)\theta F d_t + F \text{ when } \theta d_t(1-\alpha_{t+1}) + \alpha_{t+1} \geq \\ \qquad \geq d_t(1 - \alpha_{t+1}(1-\alpha)), \\ ((1-\alpha)d_t + \alpha\theta d_t)F \text{ otherwise }. \end{cases}$$

Then f is a convex function on X, $f(X) = 0$, and $f(x_t) = d_t$ for all t. ∎

Remark. We could give a similar example with X in plane L, see Figure 2. But the $(T+1)$–dimensional example above can be easily modified to an example with a convex X. Namely our function f on X can be extended to a convex function f' on the convex hull

$$X' = \{x \in L : x' = \sum \gamma(x)x\}$$

(here and below γ ranges over all non–negative functions on X taking only finitely many non–zero values and such that $\sum \gamma(x) = 1$, so $\sum \gamma(x)x$ is a convex linear combination of points in X) of X as follows:

$$f'(x') := \inf(\sum \gamma(x)f(x) : \sum \gamma(x)x = x').$$

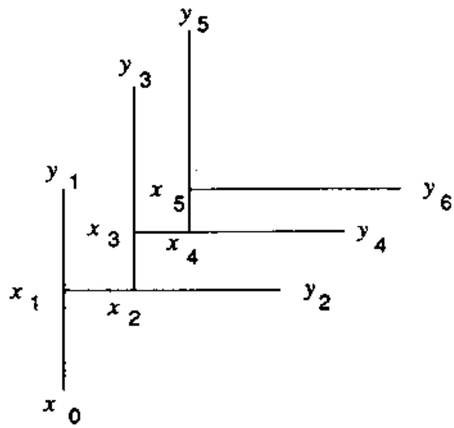

Figure 2 (see the remark)

At a small cost, an example with a convex f defined on the whole L can be constructed. Namely, for any $\varepsilon > 0$ we can construct a convex function $f = f_\varepsilon$ on L and a procedure (1), (2), (4) such that $f(x_t) - f(X) \geq F(d_t - \varepsilon)$ for all t. Indeed, let L and $\{y_t\}_{-1 \leq t < T+1}$, $\{x_t\}_{0 \leq t < T+1}$ be as above. We define f on the line $R_{-1} = \{(1-\alpha)y_{-1} + \alpha y_0 : \alpha \text{ real } \}$ as follows: $f((1-\alpha)y_{-1} + \alpha y_0) = \max(0, \alpha F)$. For any t in the interval $0 \leq t < T+1$, we define a convex function f on the line $R_t = \{(1-\alpha)x_t + \alpha y_{t+1} : \alpha \text{ real } \}$ as follows.

When $\theta d_t(1-\alpha_{t+1}) + \alpha_{t+1} \geq d_t(1 - \alpha_{t+1}(1-\theta))$, we set $f((1-\alpha)x_t + \alpha y_{t+1}) = \max(f(x_t) - \alpha f(x_t)(1-\theta)(F - \theta f(x_t) + \varepsilon F)/(\varepsilon F),\ F\alpha + (1-\alpha)(\theta f(x_t) - F\varepsilon))$ (see Figure 3).

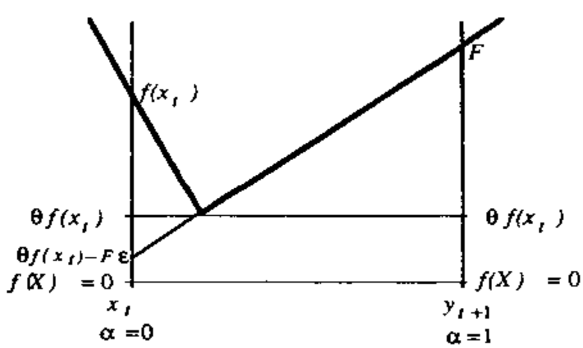

Figure 3

Otherwise, we set $f((1-\alpha)x_t + \alpha y_{t+1}) = \max(((1-\alpha)+\alpha\theta)f(x_t), f(x_t)\theta)$ (see Figure 4).

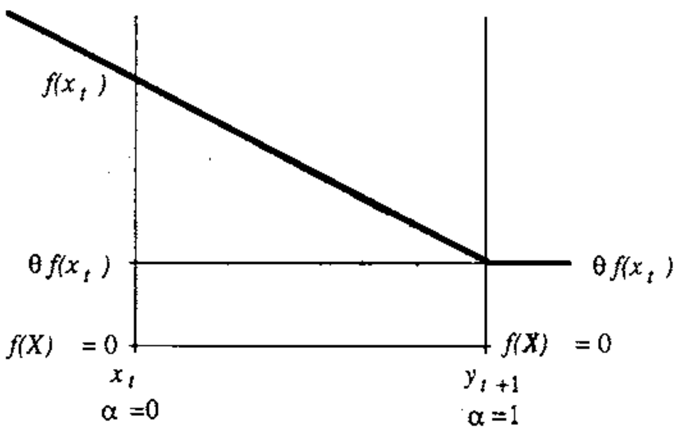

Figure 4

Then $|f(x_t) - Fd_t| \le F\varepsilon$ for all t. Now we can extend f from the union $X = \cup_{0\le t<T+1} X_t$ of all lines X_t to a convex function f on the affine space L spanned by the union. Replacing X by its convex hull, we can arrange X to be convex.

Theorem 2. *Under the conditions of Theorem 1 assume that* $T = \infty$. *Set* $\alpha_\infty := \limsup \alpha_t$ *and* $d_\infty := \limsup d_t$. *Then:*

(a) if $\sum_{t=1}^{\infty} \alpha_t = +\infty$, *then* $d_\infty \le \frac{\alpha_\infty}{1-\theta+\alpha_\infty\theta}$;
in particular, when $\alpha_\infty = 0$, *i.e.* $\alpha_t \to 0$, *then* $d_\infty = 0$. *i.e.* $d_t \to 0$, *i.e.* $f(x_t) \to f(X)$ *uniformly for all* L, X, f *as above and all procedures* (1), (2), (4) *with fixed* θ *and* F;

(b) if $\sum_{t=1}^{\infty} \alpha_t < \infty$, *then* $0 < d_\infty = \sup_{0\le t<+\infty} d_t \Pi_{s=t+1}^{\infty}(1-\alpha_s+\alpha_s\theta)$;

(c) if $\alpha_\infty > 0$, *then* $d_\infty \ge \alpha_\infty > 0$; *when, moreover,* $\alpha_t \to \alpha_\infty > 0$, *then* $d_t \to d_\infty = \alpha_\infty/(1-\theta+\alpha_\infty\theta)$;

(d) if X *contains an interval* $[x,y]$ *and a point* z *strictly inside* $[x,y]$ *such that* $f(x/2, y/2) = f(X) \ne f(z)$, *then in the cases (b) and (c) there is a procedure* (1), (2), (4) *such that* $f(x_t)$ *does not converge to* $f(X)$.

Proof: (a) We take any α such that $\alpha > \alpha_t$ for all $t > t_0$. We have to prove that $d_\infty \le \alpha/(1-\theta+\alpha\theta)$. Let $t > t_0$.

If $d_t \ge \alpha/(1-\theta+\alpha\theta)$, then using that $\alpha > \alpha_{t+1}$ we conclude that

$$d_t \ge \alpha_{t+1}/(1-\theta+\alpha_{t+1}\theta), \quad \text{i.e. } d_t\theta(1-\alpha_{t+1}) + \alpha_{t+1} \le d_t.$$

On the other hand, by the definition,

$$d_{t+1} := \max(d_t(1-\alpha_{t+1}+\alpha_{t+1}\theta), \theta d_t(1-\alpha_{t+1})+\alpha_{t+1}) \le \theta d_t(1-\alpha_{t+1})+\alpha_{t+1}.$$

So $d_{t+1} \le d_t$.

If $d_t \le \alpha/(1-\theta+\alpha\theta)$, then

$$\theta(1-\alpha_{t+1})d_t + \alpha_{t+1} \le \theta(1-\alpha_{t+1})\alpha/(1-\theta+\alpha\theta) + \alpha_{t+1} \le$$
$$\le \theta(1-\alpha)\alpha/(1-\theta+\alpha\theta) + \alpha = \alpha/(1-\theta+\alpha\theta) \text{ (using again that } \alpha > \alpha_{t+1}).$$

So $d_{t+1} \le \theta d_t(1-\alpha_{t+1}) + \alpha_{t+1} \le \alpha/(1-\theta+\alpha\theta)$.

Thus, either $d_t \le \alpha/(1-\theta+\alpha\theta)$ for all sufficiently large t (which implies the inequality $d_\infty \le \alpha/(1-\theta+\alpha\theta)$ which we are proving), or the sequence $\{d_t\}$ is monotone for all $t > t_0$ and consequently, $d_t \to d_\infty$.

In the last case, for any limit point α_0 of the sequence $\{\alpha_t\}$ (of course, $\alpha_0 \le \alpha_\infty \le \alpha$) we have:

$$d_\infty = \max(d_\infty(1-\alpha_0+\alpha_0\theta), \theta d_\infty(1-\alpha_0)+\alpha_0).$$

When $\alpha_0 \ne 0$, it follows that $d_\infty = \theta d_\infty(1-\alpha_0)+\alpha_0$, hence $d_\infty = \alpha_0/(1-\theta+\alpha_0\theta)$ and $\alpha_0 = \alpha_\infty$.

Therefore only the following case remains: $d_t \to d_\infty$ and $\alpha_t \to 0$. We have to prove now that $d_\infty = 0$. Assuming that $d_\infty \ne 0$, we obtain that $d_\infty = 0$ (a contradiction) as follows. For all sufficiently large t,

$$d_{t+1} := \max(d_t(1-\alpha_{t+1}+\alpha_{t+1}\theta), \theta d_t(1-\alpha_{t+1})+\alpha_{t+1}) =$$
$$= d_t(1-\alpha_{t+1}+\alpha_{t+1}\theta) \ge \theta d_t(1-\alpha_{t+1})+\alpha_{t+1},$$

hence

$$d_\infty = d_t \Pi_{s=t+1}^{\infty}(1-\alpha_s(1-\theta)) = 0,$$

because the series $\sum \alpha_t$ diverges.

(b) By the definition of d_t, we have

$$d_{t+1} \ge d_t(1-\alpha_{t+1}+\alpha_{t+1}\theta),$$

hence

$$d_t \ge d_s \Pi_{i=s+1}^{\infty}(1-\alpha_i(1-\theta)) =: c_s$$

whenever $t \ge s$, and $c_{s+1} > c_s$ for all s. Therefore

$$d_\infty := \limsup d_t \ge \limsup c_t = \sup_{0\le t<\infty} c_t =: c_\infty.$$

Since $\sum \alpha_s$ converges, $c_s > 0$ for sufficiently large s and $d_s/c_s \to 1$. So $d_\infty = c_\infty$.

(c) By the definition of d_t,

$$d_{t+1} \ge \theta d_t(1-\alpha_{t+1}) + \alpha_{t+1} \ge \alpha_{t+1}, \text{ hence } d_\infty \ge \alpha_\infty.$$

Suppose now that $\alpha_t \to \alpha_\infty > 0$. In the view of (a), it remains to prove only that

$$\liminf d_t \geq \alpha_\infty/(1-\theta+\theta\alpha_\infty).$$

We pick any $\alpha < \alpha_\infty$ and want to show that

$$\liminf d_t \geq \alpha/(1-\theta+\theta\alpha).$$

Pick t_0 such that $\alpha_t \geq \alpha$ for all $t \geq t_0$. For $t \geq t_0$, if $d_t < \alpha/(1-\theta+\theta\alpha)$, then $d_t\theta(1-\alpha_{t+1}) + \alpha_{t+1} \geq d_t$, hence $d_{t+1} \geq d_t$.

On the other hand, if $d_t \geq \alpha/(1-\theta+\theta\alpha)$ for some t, then

$$d_t\theta(1-\alpha_t) + \alpha_t \geq \alpha/(1-\theta+\theta\alpha), \text{ hence } d_{t+1} \geq \alpha/(1-\theta+\theta\alpha).$$

Thus, either $d_t \geq \alpha/(1-\theta+\theta\alpha)$ for sufficiently large t, which implies the wanted inequality $\liminf d_t \geq \alpha/(1-\theta+\theta\alpha)$, or the sequence $\{d_t\}$ is monotone (non-decreasing).

In the last case, $d_t \to d_\infty \leq \alpha/(1-\theta+\theta\alpha) < \alpha_\infty(1-\theta+\theta\alpha_\infty)$, which contradicts to equality $d_\infty = \max(d_\infty(1-\alpha_0+\alpha_0\theta), \theta d_\infty(1-\alpha_0)+\alpha_0)$ with $\alpha_0 = \alpha_\infty$ observed in (a) above.

(d) Among those w in $[x,y]$ where the function f reaches its minimal value, i.e. $f(w) = f(X)$, we take the point x' closest to z.

Replacing, if necessary, the interval $[x,y]$ by a subinterval centered around x', we assume that $x' = x/2 + y/2$, $f(u) - f(X) \leq F$ for all u in $[x,y]$ and $f((1-\alpha)x + \alpha y) > f(X)$ whenever $1/2 < \alpha \leq 1$.

When the condition (b) holds, i.e. $\sum \alpha_t < \infty$, our procedure will take place in the interval $[x',y]$. Note that $f(x') = f([x',y]) = f(X)$ and $f(u) > f(X)$ for $u \neq x'$ in the interval $[x',y]$. Let s be the smallest t_0 such that $\alpha_t < 1$ for all $t \geq t_0$. If $s = 1$, we take $x_0 = y$. Otherwise, $\alpha_{s-1} = 1$, and we set $x_0 = y_t = x'$ for $t < s$ and $y_s = y$. In both cases, $x_{s-1} = y$. We set $y_t = x'$ for $t > s$. Then the condition (2) holds for all t with $\theta = 0$, and for all $t \geq s$ we have

$$x_t - x' = \Pi_{i=s}^{t}(1-\alpha_i)(z-x'), \text{ so } x_t \to x_\infty = x' + \Pi_{i=s}^{\infty}(1-\alpha_i)(z-x') \neq x'.$$

Thus, $f(x_t) \to f(x_\infty) > f(X)$.

If the condition (c) holds, i.e. $\limsup \alpha_t = \alpha_\infty > 0$, then we pick a sequence $\{s(i)\}$ of natural numbers such that: $s(i) < s(i+1)$,

$$\frac{\alpha_\infty}{2} \leq \alpha_{s(i)} \leq \frac{3\alpha_\infty}{2}, \text{ and } \Pi_{t=s(i)+1}^{s(i+1)-1}(1-\alpha_t) \leq \frac{1}{3} \text{ for all } i.$$

We set $x_0 = x' = y_t$ for all t outside $\{s(i)\}, y_{s(2i+1)} = y$, and $y_{s(2i)} = x$ for all i. Then $x_t = \beta_t y + (1-\beta_t)x$, where: $\beta_t = 1/2$ when $t < s(1)$; $\beta_{s(1)} = (1+\alpha_{s(1)})/2$; $1/2 < \beta_t$ when $s(1) \leq t < s(2)$;

$$\beta_{s(2)-1} = \frac{1}{2} + \frac{\alpha_{s(1)}}{2}\Pi_{t=s(1)+1}^{s(2)-1}(1-\alpha_t) \leq \frac{1}{2} + \frac{\alpha_\infty}{4};$$
$$\beta_{s(2)} = \beta_{s(2)-1}(1-\alpha_{s(2)}) \leq (1-\alpha_\infty 2/4)/2.$$

Induction on i shows:

$1/2 \le \beta_t \le (1 + 3\alpha_\infty/2)/2$ for $s(2i-1) \le t < s(2i)$ and $\beta_{s(2i-1)} \le 1/2 + \alpha_\infty/4, (1 + 3\alpha_\infty/2)/2 \le \beta_t \le 1/2$ for $s(2i) \le t < s(2i+1)$ and $\beta_{s(2i+1)-1} \ge\ge 1/2 - \alpha_\infty/4$.

So $\alpha_\infty^2/8 \le |\beta_{s(i)} - 1/2| \le 3\alpha_\infty/4$ for all i.

Therefore $\limsup \beta_t \ge 1/2 + \alpha_\infty 2/8 > 1/2$, hence $\limsup f(x_t) > f(x') = f(X)$.

Thus, $f(x_t)$ does not converge to $f(X)$. ∎

Remark. If X does not contain any interval $[x, y]$ with $x \ne y$, then, evidently, $x_t = x_0$ for all t and any procedure (1); moreover, the condition (2) with $t = 1$ implies that $f(x_0) = f(X)$.

Example. Let $\theta = 0$ in Theorem 1. When $\alpha_t = 1/(t+1)$ for all t, then $d_t = 1/(t+1)$ for all t. In the next section we will see that this $\{\alpha_t\}$ is the best sequence when $\theta = 0$.

3. The best damping sequence

Now we want to find the best damping sequence $\{\alpha_t\}$, that is, the one which gives the minimal value for d_T (when T is finite) in Theorem 1. The following theorem claims, among other things, the existence and uniqueness of such a sequence and its independence on F and T (when T increases, new members are added to the sequence, but old members stay the same).

Theorem 3. *For a fixed θ in the interval $[0,1]$ let us define inductively a sequence $\{A_t\} = \{A_t(\theta)\}$ by $A_1 = 1/2$ and*

$$A_{t+1} = A_t(1 - A_t + \theta A_t)/(1 - (1-2\theta)A_t^2) \text{ for } t \ge 1.$$

Then:

(a) the sequence $\{A_t(\theta)\}$ strictly decreases to 0; when $\theta < 1$ we have

$$A_t(\theta) = (1-\theta)/((t+1)(1-\theta)^2 + \theta^2 \ln(t) + \theta c_t(\theta)/(1-\theta)^2)$$

with $-7 \le c_t(\theta) \le 3$ for all t and θ; when $\theta = 1$ we have

$$A_t(1) = (3/A_t(1/3) - 2)^{-1/2} = (2t + \ln(t)/2 + 27c_t(1/3)/8)^{-1/2};$$

(b) the sequence $D_t(\theta) := A_{t+1}(\theta)/(1 - \theta + A_{t+1}(\theta)(2\theta - 1))$ decreases strictly to 0 and has the form

$$D_t(\theta) = ((1-\theta)^2 t + \theta^2 \ln(t+1) - 2\theta + 2\theta^2 + 1 + \theta c_t(\theta)/(1-\theta)^2)^{-1}$$

when $\theta < 1$; otherwise, $D_t(1) = 1$ for all t;

(c) for any L, X, f, T, θ, F as above and any procedure (1),(2),(4), *with* $\{\alpha_t\} = A_t(\theta)$ *we have*

$$f(x_t) - f(X) \leq FD_t(\theta),\ 0 \leq t < T+1;$$

(d) for arbitrary $T, F > 0$, $\theta < 1$, $\{\alpha_t\} \in [0,1]^T$ there exists L, X, f, and a procedure (1),(2),(4) *such that $f(x_t) - f(X) \geq FD_t(\theta)$ for all t in the interval $0 \leq t < T+1$, and the inequality is strict for each t such that $\alpha_s \neq A_s(\theta)$ for some $s \leq t$.*

Proof: (a) Assume that for some t we have shown that $0 < A_t \leq 1/2$ (for $t = 1$, this is the case), and let us show that then $0 < A_{t+1} < A_t$. Indeed the inequalities $0 < A_t \leq 1/2$ and $0 \leq \theta \leq 1$ imply that $1 > (1-\theta)A_t > (1-2\theta)A_t^2$, hence

$$0 < A_{t+1}/A_t = (1 - A_t(1-\theta))/(1 - A_t^2(1-2\theta)) < 1, \text{ so } 0 < A_{t+1} < A_t.$$

Let A_∞ be the limit of the monotone sequence $\{A_t\}$. Then $A_\infty \geq 0$ and

$$A_\infty = A_\infty(1 - A_\infty(1-\theta))/(1 - A_\infty^2(1-2\theta)),$$

hence $A_\infty = 0$, because $1 - \alpha(1-\theta) < 1 - \alpha^2(1-2\theta)$ for any α in the interval $0 < \alpha \leq 1/2$.

Let us now find the asymptotic of the sequence$\{A_t\}$ when $\theta < 1$. Set $x_t = 1/A_t$. Then $x_1 = 2 \leq x_t$ for all t, and the defining equality for A_t takes the form

$$x_{t+1} = x_t + (1-\theta) + \theta^2/(x_t - 1 + \theta).$$

This can be rewritten as follows:

$$y_0 = \theta, y_{t+1} = y_t + \theta^2/((1-\theta)t + y_t), \tag{5}$$

where

$$y_t := x_t - (t+1)(1-\theta),\ y_1 = 2\theta,\ y_2 = \theta(2+3\theta)/(1+\theta).$$

It is clear from (5) that $y_t \geq 0$ for all t and that $y_{t+1} - y_t \leq \theta^2/(1-\theta)t$. So for $t \geq 2$ we obtain:

$$y_t = \sum_{s=3}^{t}(y_s - y_{s-1}) + y_2 \leq \frac{\theta^2}{1-\theta}\sum_{s=2}^{t-1} 1/s + \frac{\theta(2+3\theta)}{1+\theta} \leq$$
$$\leq \frac{\theta^2}{1-\theta}\int_1^{t-1}\frac{dt}{t} + 3\theta = \frac{\theta^2}{1-\theta}\ln(t-1) + 3\theta \leq \frac{\theta}{1-\theta}(\ln(t) + 3).$$

The obtained inequality $y_t \leq \theta(\ln(t) + 3)/(1-\theta)$ holds also for $t = 1$.

On the other hand, substituting this upper bound for y_t into (5), we obtain that

$$y_{t+1} - y_t = \theta^2/((1-\theta)t + y_t) = \theta^2/(1-\theta)t - \theta^2 y_t/(1-\theta)t((1-\theta)t + y_t) \geq$$
$$\geq \theta^2/(1-\theta)t - \theta^3(\ln\ (t) + 3)/t^2(1-\theta)^3.$$

So, for any $t \geq 1$, we have:

$$y_t \geq 2\theta + \sum_{s=1}^{t} \frac{\theta^2}{s(1-\theta)} - \theta^3 \sum_{s=1}^{t} \frac{\ln\ (s) + 3}{s^2(1-\theta)^3} \geq$$
$$\geq 2\theta + \theta^2 \int_1^t \frac{ds}{s(1-\theta)} - \frac{3\theta^3}{(1-\theta)^3} - \theta^3 \int_1^t \frac{\ln\ (s) + 3}{s^2(1-\theta)^3} =$$
$$= 2\theta + \theta^2 \ln\ (t)/(1-\theta) - 3\theta^3/(1-\theta)^3 + \theta^3((\ln\ (t) + 4)/t - 4)/(1-\theta)^3 \geq$$
$$\geq \theta^2 \ln\ (t)/(1-\theta) - 7\theta^3/(1-\theta)^3.$$

Therefore, for

$$c_t(\theta) := (1-\theta)^3(y_t - \theta^2 \ln\ (t)/(1-\theta))/\theta,$$

we obtain that

$$c_t(\theta) \leq (1-\theta)^3(\theta^2 \ln\ (t-1)/(1-\theta) + 3\theta - \theta^2 \ln\ (t)/(1-\theta)/\theta \leq$$
$$\leq (1-\theta)^3(3\theta)/\theta = 3(1-\theta)^3 \leq 3.$$

and

$$c_t(\theta) \geq (1-\theta)^3(-7\theta^3/(1-\theta)^3)/\theta = -7\theta^4 \geq -7.$$

When $t = 1$, we have

$$c_1(\theta) = (1-\theta)^3(2\theta)/\theta = 2(1-\theta)^3.$$

Thus, $-7 \leq c_t(\theta) \leq 3$ for all t and θ, hence $|c_t(\theta)|$ is bounded uniformly over all t and θ.

Let us find now $A_t(1)$ in the terms of $A_t(1/3)$. By the definition, $A_1 = 1/2$ and $A_{t+1}(1) = A_t(1)/(1 + A_t(1)^2)$ or

$$\frac{1}{\sqrt{y_{t+1}}} = \frac{\sqrt{y_t}}{1 + y_t},$$

where $y_t := A_t(1)^{-2}$. Therefore

$$y_1 = 4 \text{ and } y_{t+1} = y_t + 2 + 1/y_t \text{ for } t \geq 1.$$

On the other hand, for $x_t := 1/A_t(1/3)$ we have

$$x_1 = 2 \text{ and } x_{t+1} = x_t + 2/3 + 1/(9x_t - 6) \text{ for } t \geq 1.$$

So for $z_t := 3x_t - 2$ we obtain that

$$z_1 = 4 \text{ and } z_{t+1} = z_t + 2 + 1/z_t \text{ for } t \geq 1.$$

Thus, $z_t = y_t$ for all $t \geq 1$, i.e.

$$1/A_t(1) = (3/A_t(1/3) - 2)^{1/2} = (2t + \ln(t)/2 + 27c_t(1/3)/8)^{1/2}.$$

(b) This follows from the definition of D_t and the part (a).

(c) By the definition,

$$D_t := D_t(\theta) = A_{t+1}/(1 - \theta + 2\theta A_{t+1} - A_{t+1}).$$

Expressing here $A_{t+1} = A_{t+1}(\theta)$ in terms of $A_t = A_t(\theta)$ and using the definition of D_{t-1}, we obtain:

$$\begin{aligned} D_t &= 1/((1-\theta)/A_{t+1} + 2\theta - 1) = A_t(1 - A_t + \theta A_t)/(1 - \theta + (2\theta - 1)A_t) = \\ &= D_{t-1}(1 - A_t + \theta A_t) = \theta D_{t-1}(1 - A_t) + A_t = \\ &= \max(D_{t-1}(1 - A_t + \theta A_t), \theta D_{t-1}(1 - A_t) + A_t). \end{aligned}$$

Moreover, $D_0(\theta) = 1$, since $A_1(\theta) = 1/2$.

Therefore the sequence $\{d_t\} := \{D_t(\theta)\}$ coincides with the sequence $\{d_t\}$ of Theorem 1 when α_t is taken to be $A_t(\theta)$ for all t. So the conclusion of Theorem 3(c) with $\theta < 1$ follows from Theorem 1. When $\theta = 1$ the conclusion coincides with the condition (4).

vglue 0.3 true cm (d) Let us define L, X, f, y_t, x_t as in the proof of Theorem 1. Then $f(x_t) - f(X) = Fd_t$, where $d_0 = 1$ and

$$d_{t+1} := \ max(d_t(1 - \alpha_{t+1} + \alpha_{t+1}\theta), \theta d_t(1 - \alpha_{t+1}) + \alpha_{t+1}) \text{ for } t \geq 0.$$

We want to prove now that $d_t \geq D_t(\theta)$ for all t, and that this inequality is strict for each t such that $\alpha_s \neq A_s(\theta)$ for some $s \leq t$.

Proceeding by induction on t, we assume, for some t, that $d_t \geq D_t(\theta)$ and in the case of equality every α_s with $s \leq t$ coincides with $A_s(\theta)$.

If $\alpha_{t+1} \leq A_{t+1}(\theta)$, then

$$d_{t+1} \geq d_t(1 - (1-\theta)\alpha_{t+1}) \geq D_t(\theta)(1 - (1-\theta)\alpha_{t+1}) = D_{t+1}(\theta).$$

Otherwise, i.e. when $\alpha_{t+1} > A_{t+1}(\theta)$, we have

$$d_{t+1} \geq \theta d_t(1 - \alpha_{t+1}) + \alpha_{t+1} \geq \theta D_t(\theta)(1 - A_{t+1}(\theta)) + A_{t+1}(\theta) = D_{t+1}(\theta).$$

Thus, $d_t \geq D_t(\theta)$ in both cases. Furthermore, the equality $d_{t+1} = D_{t+1}(\theta)$ implies, evidently, that $d_t = D_t(\theta)$ and $\alpha_{t+1} = A_{t+1}(\theta)$).

Remark. For any $\varepsilon > 0$, an example can be constructed (see the remark in the previous section) with a convex f defined on the whole L and a convex X such that $f(x_t) - f(X) \geq F(D_t(\theta) - \varepsilon)$ for all t in the interval $0 \leq t < T+1$, and $f(x_t) - f(X) > FD_t(\theta)$ for each t such that $\alpha_s \neq A_s(\theta)$ for some $s \leq t$. So the sequence $\{A_t(\theta)\}$ stays the best in this restricted class of functions f.

Acknowlegment. A referee made many corrections in this paper.

References

1. V.Z.BELEN'KY, V.A.VOLKONSKY,S.A.IVANKOV,A.B.POMANSKY, A. D. SHAPIRO, "*Iterative methods in game theory and programming* ," Nauca, Moscow, 1974. (in Russian)
2. G. W. BROWN, Some notes on computation of game solutions,, *RAND Report P-78,* (April 1949), The Rand Corporation, Santa Monica, California.
3. M. M. DAY, "*Normed linear spaces*," Springer-Verlag, 1958.
4. B. T. POLJAK, A general method for solving extremum problems, *Soviet Math. Dokl.* 8 (1967), 593-597.
5. J. ROBINSON, An iterative method of solving a game, *Ann. Math.* 54 (1951), 246-301.

Department of Mathematics
Penn State University
University Park, PA 16802, USA.
and
Institut d'Estudis Catalans
Centre de Recerca Matemàtica
Apartat 50-08193 Bellaterra, SPAIN.

Rebut el 17 de Maig de 1988