Mathematical Programming 69 (1995) 111–147

New variants of bundle methods

Claude Lemaréchal [a,*], Arkadii Nemirovskii [b], Yurii Nesterov [b]

[a] *Institut National de Recherche en Informatique et en Automatique, B.P. 105, 78153 Le Chesnay, France*
[b] *Central Economical and Mathematical Institute, 32 Krasikova Street, Moscow 117481, Russia*

Received 15 September 1992; revised manuscript received 16 April 1994

Abstract

In this paper we describe a number of new variants of bundle methods for nonsmooth unconstrained and constrained convex optimization, convex–concave games and variational inequalities. We outline the ideas underlying these methods and present rate-of-convergence estimates.

Keywords: Nonsmooth convex optimization; Bundle methods

0. Introduction

0.1.

Consider the basic problem of minimizing a convex function f over a "simple" convex set $Q \subset \mathbb{R}^n$. Having generated the iterates $x_1, \ldots, x_i \in Q$ and using an *oracle* to compute function-values $f(x)$ and subgradient-values $f'(x)$, a fruitful object is the *cutting-plane model*

$$f_i(x) = \max\{f(x_j) + (f'(x_j))^{\mathrm{T}}(x - x_j) \mid 1 \leqslant j \leqslant i\} ,$$

underestimating f. To exploit it, the very first idea is the classical *cutting-plane algorithm* of [2,6] in which x_{i+1} minimizes f_i over Q; it is known as very slow, from both the theoretical and practical viewpoints; see [16] for example.

More recently, some refinements of this idea have been proposed, under the wording of *bundle methods*. In their simplest form [7,11,15], the next iterate is

$$x_{i+1} = \operatorname{argmin}\{f_i(x) + \tfrac{1}{2}u_i |x - x_i^+|^2 \mid x \in Q\} , \tag{0.1}$$

where the *current prox-center* x_i^+ is a certain point from the set $\{x_1, \ldots, x_i\}$ and u_i is the

* Corresponding author.

SSDI 0025-5610(94)00067-0

current penalty parameter. If $f(x_{i+1})$ turns out to have "sufficiently decreased" (*descent step*), the prox-center is updated to x_{i+1}; otherwise (*null step*), $x_{i+1}^+ = x_i^+$. This idea looks natural: the model accumulates all the information about f obtained so far, and the penalty term reduces the influence of the model's inaccuracy, thereby reducing instabilities. A bundle method is thus determined by two rules: (1) to define a "sufficient" decrease, and (2) to select the penalty parameter. Satisfactory rules have been developed for (1), based on a comparison between the actual value $f(x_{i+1})$ and the "ideal" value $f_i(x_{i+1})$ of the model. As for (2), the question is not so clear: the simplest choice $u_i \equiv 1$ is theoretically possible but experience demonstrates that efficiency requires "on line" adjustments, as in [8,21].

0.2.

Alternatives to (0.1) can be considered, which have the same stabilizing effect. Let us mention two of them: the "trust-region approach"

$$x_{i+1} = \operatorname{argmin}\{f_i(x) \mid x \in Q,\ |x - x_i^+| \leqslant \tau_i\},$$

which does not seem to have been studied, and the proposal of Lemaréchal et al. [13], in which the control parameter is a certain ϵ_i, whose choice implies a detour in the dual space. In what follows, we study a fourth variant: instead of u_i, τ_i or ϵ_i, we control the value of the model at the next iterate: we choose a *level* l_i and replace (0.1) by

$$x_{i+1} = \operatorname{argmin}\{\tfrac{1}{2}|x - x_i^+|^2 \mid x \in Q, f_i(x) \leqslant l_i\}. \tag{0.2}$$

It turns out that the level-sets of the model are rather "stable", so that extremely simple rules can be used for updating the level l_i. This property also allows us to forget about the concepts of prox-center and null-step: x_i^+ may be systematically set to the last iterate x_i in (0.2).

Our basic strategy works as follows: at the ith step, compute the minimal value $f_*(i)$ of the model over Q (assumed bounded); also, let

$$f^*(i) = \min\{f(x_j) \mid 1 \leqslant j \leqslant i\} = f(x_i^*)$$

be the best value of the objective obtained during the first i steps, and call

$$\Delta(i) = f^*(i) - f_*(i) \tag{0.3}$$

the ith *gap* (x_i^* certainly minimizes f within $\Delta(i)$ and our aim is to force $\Delta(i) \to 0$). Then, having $\lambda \in (0, 1)$, solve (0.2) with the value

$$l_i = \lambda f^*(i) + (1-\lambda) f_*(i) = f_*(i) + \lambda\Delta(i). \tag{0.4}$$

0.3.

The value $\lambda = 1$ in (0.4) would typically result in $l_i = f^*(i) = f_i(x_{i+1}) \leqslant f(x_{i+1})$: no progress would be obtained in the objective function. By contrast, $\lambda = 0$ would yield the convergent (even though slow) pure cutting–plane methods; this suggests that small values should be less dangerous than large values of λ, i.e., of the level.

An arbitrary but fixed $\lambda \in (0, 1)$ gives the following efficiency estimate: to obtain a gap smaller than ϵ, it suffices to perform

$$M(\epsilon) \leqslant c\left(\frac{LD}{\epsilon}\right)^2 \tag{0.5}$$

iterations (here, L and D are the Lipschitz constant of f and the diameter of Q respectively, c is a constant depending only on λ). Such an efficiency is optimal in a certain sense (see [16]): suppose Q is a ball of radius $\frac{1}{2}D$, and the dimension is $n \geqslant \frac{1}{4}(LD/\epsilon)^2$; take an arbitrary method but use at most $\frac{1}{4}(LD/\epsilon)^2$ evaluations of f and f' (and no other information from the problem); then, there exists a function for which this method does not obtain an accuracy better than ϵ. As a result, our method cannot be improved uniformly with respect to the dimension by more than an absolute constant factor.

To obtain the estimate (0.5), the key argument is as follows: consider, for a given i_0, the maximal sequence $I = \{i_0, i_0+1, \ldots, i_1\}$ of iterations (we call it a *group*), at the end of which the gap has not been reduced much, namely,

$$\Delta(i_1) \geqslant (1-\lambda)\Delta(i), \quad \text{for all } i \in I.$$

Then, all level-sets characterizing (0.2) with $i \in I$ have a point in common. This crucial property allows the following majoration of the number of iterations in the group:

$$|I| \leqslant c\left(\frac{LD}{\Delta(i_1)}\right)^2,$$

where c is a constant depending only on λ. Then, using the fact that the gap is reduced by $(1-\lambda)$ at the iteration i_1+1, repeated use of this argument results in the majoration (0.5).

In Section 2 we present a number of variants of the above algorithm, all enjoying the same efficiency property (0.5).

0.4.

The subsequent sections are devoted to problems for which the same idea can be considered. After all, the above "level" principle gives an implementable mechanism to solve a system of inequalities (via a method resembling Newton's method, see [19]): we want to find $x \in Q$ such that

$$f(x') + (f'(x'))^{\mathrm{T}}(x - x')\,[\leqslant f(x)] \leqslant f^* \quad \text{for all } x' \in Q.$$

Here, there are infinitely many indices, so they are accumulated one after the other: $x' = x_1$, $x_2, \ldots$; and f^* is unknown, so the level-strategy takes care of it.

The essential feature to make the method work is to define an appropriate nonnegative *gap* as in (0.3), which is 0 when the problem is solved. The whole approach is therefore to minimize this gap, an idea which can actually be extended to several problems.

(A) *Saddle-point problems* (Section 3). Given a convex–concave function $f(x, y)$ defined on the direct product of Q and H (convex and compact), find a *saddle point* $(x^*, y^*) \in Q \times H$, i.e., a point satisfying

$$\max\{f(x^*, y) \mid y \in H\} = f(x^*, y^*) = \min\{f(x, y^*) \mid x \in Q\} .$$

This just amounts to minimizing the convex function

$$F(x, y) = \max_H f(x, \cdot) - \min_Q f(\cdot, y)$$

over $Q \times H$. The difficulty is that we have no oracle computing the values and the subgradients of F; nevertheless, a set of iterates $\{(x_j, y_j) \mid 1 \leqslant j \leqslant i\}$ yields the *model*

$$F_i(x, y) = \bar{f}_i(x) - \underline{f}_i(y) , \tag{0.6}$$

where the standard first-order information is used:

$$\bar{f}_i(x) = \max\{f(x_j, y_j) + (f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j) \mid 1 \leqslant j \leqslant i\} ,$$

$$\underline{f}_i(y) = \min\{f(x_j, y_j) + (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j) \mid 1 \leqslant j \leqslant i\} ;$$

thus, F_i underestimates F. We know that the minimal value of F is zero; the minimal value of each F_i is therefore nonpositive and provides the gap $\Delta_i = -\min_{Q \times H} F_i$. This enables to define a method of the type (0.2) for saddle point problems with the efficiency estimate (0.5).

It is interesting to note the decomposed property of the model (0.6): to minimize it, it suffices to solve successively the two linearized optimization problems

$$\min_Q \bar{f}_i(x) \quad \text{and then} \quad \max_H \underline{f}_i(y) .$$

This suggests an interpretation of our approach in terms of games: there are two players x and y, in charge of minimizing f and $-f$, respectively; $\bar{f}_i$ and $(-\underline{f}_i)$ can be interpreted as under-approximations of their worst-case loss-functions.

We recall that the usual algorithms for saddle points are based on subgradient optimization [3]. In [1], approaches similar to ours were considered, but of course based on pure cutting-plane approximations.

(B) *Convex constrained problems* (Section 4). Given the function G, convex on the compact convex set Q, our approach to solve

$$\min\{f(x) \mid G(x) \leqslant 0, x \in Q\}$$

is via the equivalent problem

$$\min\{\max[f(x) - f^*, G(x)] \mid x \in Q\} . \tag{0.7}$$

(It is to alleviate notations that we assume just one inequality constraint.)

The optimal value f^* of the constrained problem is of course unknown, which introduces a new difficulty: no oracle can compute the function value in (0.7). We therefore under-estimate f^* by the optimal value $f_*(i)$ of

$$\min\{f_i(x) \mid G_i(x) \leqslant 0,\ x \in Q\}$$

(G_i being the cutting-plane approximation of G), and we propose two approaches.

First, duality theory tells us that (0.7) is equivalent to

$$\max\{h(\alpha) - \alpha f^* \mid 0 \leqslant \alpha \leqslant 1\}, \tag{0.8}$$

where

$$h(\alpha) = \min\{\alpha f(x) + (1-\alpha)G(x) \mid x \in Q\}$$

can be overestimated by the function

$$h_i(\alpha) = \min\{\alpha f(x_j) + (1-\alpha)G(x_j) \mid 1 \leqslant j \leqslant i\}.$$

Thus, a gap is obtained:

$$\Delta_i = \max\{h_i(\alpha) - \alpha f_*(i) \mid 0 \leqslant \alpha \leqslant 1\},$$

which must be reduced to the optimal value in (0.8), i.e., in (0.7), namely 0.

In our second approach, f^* is replaced by a parameter t, and the problem is to solve the equation

$$\kappa(t) = \min\{\max\{f(x) - t,\ G(x)\} \mid x \in Q\} = 0.$$

(This is close to the method of "loaded functional" [9] and Huard's method of centers [5].) Here again, κ cannot be computed exactly. A gap is therefore defined, by way of cutting-plane approximations in κ, and t is updated to the current $F_*(t)$ whenever this gap diminishes by a sufficient amount.

In both methods, the need to identify f^* while solving the saddle–point problem (0.7) is paid by an extra cost in the efficiency estimate, which becomes as follows: to reach a point x satisfying

$$f(x) \leqslant f^* + \epsilon \quad \text{and} \quad G(x) \leqslant \epsilon,$$

it suffices to perform

$$M(\epsilon) \leqslant c\,\frac{LD^2}{\epsilon}\ln\frac{LD}{\epsilon}$$

iterations. Note, however, that no Slater assumption is needed; as a result, the efficiency is not affected by large Lagrange multipliers, as is the case with methods involving exact penalty.

(C) *Variational inequalities with monotone operators* (Section 5). These also admit a solution procedure of the type (0.2) with efficiency estimate (0.5). Indeed, consider again Section 0.1: in the definition of the model f_i, replace the values $f(x_i)$ by the current best value $f^*(i)$. The result is a further underestimate of the model:

$$\phi_i(x) = f^*(i) + \max\{(f'(x_j))^{\mathrm{T}}(x - x_j) \mid 1 \leqslant j \leqslant i\} \leqslant f_i(x),$$

so a variant of the level algorithm is readily obtained if we replace the function f_i by ϕ_i (note the similarity with the *conjugate subgradient* approach of [10,22]). The interest of this variant is that function values are no longer involved, so it can be used to solve the problem

$$\text{find } x \in Q \text{ s.t. } (F(x'), x'-x) \geqslant 0, \quad \text{for all } x' \in Q. \tag{0.9}$$

(F is a (possibly multivalued) monotone mapping and Q is again closed and convex.) Here, the monotone mapping F plays the role of f' and ϕ_i allows the definition of a gap Δ_i associated with the function

$$f(x) = \sup\{(F(x'), x-x') \mid x' \in Q\}.$$

The resulting method is reminiscent of [14], but continuity of $F(\cdot)$ is not assumed (although we require both F and Q to be bounded).

Recall that the standard formulation of a variational inequality is

$$\text{find } x \in Q \text{ s.t. } (F(x), x'-x) \geqslant 0, \quad \text{for all } x' \in Q, \tag{0.10}$$

which is *not* the same as (0.9). It can be proved, however, that (0.9) and (0.10) are equivalent in the maximal monotone case (see Section 1 for precise formulations).

An important computational advantage of (0.9) as compared to (0.10) is that we have to minimize the function f which is convex, but so would not be the case when dealing with the gap

$$f^{\#}(x) = \sup\{(F(x), x-x') \mid x' \in Q\},$$

associated with (0.10).

A final comment: solving the applications described above was made possible thanks to the introduction of levels into the bundle approach. In return, the same applications can be solved by the other variants of bundle methods, such as those alluded to in Sections 0.1, 0.2. This may be useful to remove any compactness assumptions; furthermore, the similarity between bundle methods and *sequential quadratic programming* (see [18]) opens the way to attractive alternatives to the exact penalty approach (cf. the end of (B) above).

In this paper, we consider the four above-mentioned classes of problems: constrained and unconstrained minimization, saddle points, variational inequalities. For each of them we present several algorithms close to each other, all based on the level idea (0.2), and we state their efficiency estimate. To prevent the paper from becoming unduly lengthy, we prove these estimates for one algorithm only (in each of the four classes). All proofs not given here can be found in [12].

1. Problems

Main notations. $|\cdot|$ denotes the standard Euclidean norm on $\mathbb{R}^n$. If Q is a nonempty closed convex subset in $\mathbb{R}^n$ and $x \in \mathbb{R}^n$, then $\pi(x, Q)$ denotes the (unique) point of Q closest to x.

We consider the following four problems.

(Min) Minimize $f(x)$ s.t. $x \in Q$.

Notation and assumptions on the data. f is *convex Lipschitz continuous* on the *bounded closed convex* set $Q \subset \mathbb{R}^n$. L denotes the Lipschitz constant of f, D denotes the diameter of Q with respect to the norm $|\cdot|$ and $V = LD$. f^* denotes the minimal value of f on Q.
Oracle. Given an input $x \in Q$, it computes $f(x)$ and a support functional $f'(x)$ of f at x, $|f'(x)| \leqslant L$.
Accuracy measure.

$$\epsilon(x) = \begin{cases} +\infty, & x \notin Q, \\ f(x) - \min_Q f, & x \in Q. \end{cases}$$

(Sad) Find a saddle point of $f(x, y)$ on $Q \times H$.
Notation and assumptions on the data. f is *convex in* $x \in Q$, *concave in* $y \in H$ *and Lipschitz continuous* on the direct product of *bounded closed convex* sets $Q \subset \mathbb{R}^n$, $H \subset \mathbb{R}^{n'}$. L_x (L_y) denotes the Lipschitz constant of f with respect to x (respectively y); D_x (D_y) denotes the diameter of Q (respectively H) with respect to the norm $|\cdot|$; V denotes the quantity $L_x D_x + L_y D_y$.
Oracle. Given an input $(x, y) \in Q \times H$, it computes $f(x, y)$ and two support functionals $f'_x(x, y)$ of $f(\cdot, y)$ at x and $f'_y(x, y)$ of $f(x, \cdot)$ at y, $|f'_x(x, y)| \leqslant L_x$, $|f'_y(x, y)| \leqslant L_y$.
Accuracy measure.

$$\epsilon(x, y) = \begin{cases} +\infty, & (x, y) \notin Q \times H, \\ \max_H f(x, \cdot) - \min_Q f(\cdot, y), & (x, y) \in Q \times H. \end{cases}$$

(CMin) Minimize $f(x)$ s.t. $x \in Q$, $g_i(x) \leqslant 0$, $i = 1, \ldots, m$.
Notation and assumptions on the data. f is *convex Lipschitz continuous* on the *bounded closed convex* set $Q \subset \mathbb{R}^n$; g_i, $i = 1, \ldots, m$, are *convex Lipschitz continuous* on Q. L denotes the maximum of the Lipschitz constants of f, $g_1, \ldots, g_m$; D denotes the diameter of Q with respect to the norm $|\cdot|$; $V = DL$, $G = \max\{g_1, \ldots, g_m\}$. The problem is assumed to be consistent, and f^* denotes the optimal value of the objective over the feasible set.
Oracle. Given an input $x \in Q$, it computes $f(x)$, $g_1(x), \ldots, g_m(x)$ and support functionals $f'(x)$, $g'_1(x), \ldots, g'_m(x)$ of f, $g_1, \ldots, g_m$ at x such that $|f'(x)| \leqslant L$, $|g'_i(x)| \leqslant L$, $i = 1, \ldots, m$.
Accuracy measure.

$$\epsilon(x) = \begin{cases} +\infty, & x \notin Q, \\ \max\{f(x) - f^*, G(x)\}, & x \in Q. \end{cases}$$

(Var) Find $x \in Q$ such that $F^{\mathrm{T}}(y)(y - x) \geqslant 0$, $y \in Q$.
Notation and assumptions on the data. F is a *monotone bounded-valued operator* on the *bounded closed convex* set $Q \subset \mathbb{R}^n$. L denotes the quantity $\sup_Q |F(\cdot)|$, D denotes the diameter of Q with respect to the norm $|\cdot|$, and V denotes the quantity LD.
Oracle. Given an input $x \in Q$, it computes $F(x)$.
Accuracy measure.

$$\epsilon(x)=\begin{cases}+\infty, & x\notin Q,\\ \max\{F^{\mathrm{T}}(y)(x-y)\mid y\in Q\}, & x\in Q.\end{cases}$$

Comment. The standard definition of a solution to the variational inequality $V(F, Q)$ generated by a closed convex set Q and a monotone multi-valued operator F ($\mathrm{Dom}\{F\}\supseteq \mathrm{ri}\ Q$) is as follows: $x\in\mathrm{Dom}\{F\}\cap Q$ is a solution, if there exists $\xi\in F(x)$ such that $\xi^{\mathrm{T}}(y-x)\geqslant 0$ for any $y\in Q$. A solution to (Var) (let us call it a *weak solution* to $V(F, Q)$) is defined in another way: it is a point of Q such that $F(y)^{\mathrm{T}}(y-x)\geqslant 0$ for all $y\in\mathrm{Dom}\{F\}\cap Q$. It is easily seen that every solution to $V(F, Q)$ is a weak solution as well, but the converse statement is false in general. It can be proved (see [12]) that under very mild restrictions (F is maximal monotone on $\mathrm{Dom}\{F\}\supseteq\mathrm{ri}\ Q$) this converse statement becomes true, and this justifies, to certain extent, our setting of the problem. From the computational viewpoint, the advantage of this setting is that the associated residual function $\epsilon(x)$ (this function vanishes on the set of weak solutions to $V(F, Q)$ and is positive outside this set) is convex, which is not the case for the natural residual responsible for the set of "strong" solutions.

2. Methods for (Min)

2.1. Notation

Assume we have called the oracle at the points $x_1, \ldots, x_i\in Q$. Then the following objects are defined.

Model. $f_i(x)=\max\{f(x_j)+(f'(x_j))^{\mathrm{T}}(x-x_j)\mid 1\leqslant j\leqslant i\}$.

Remark 2.1.1. Clearly,

$$f_1(x)\leqslant f_2(x)\leqslant\cdots\leqslant f_i(x)\leqslant f(x),\quad x\in Q, \tag{2.1}$$

all f_j are Lipschitz continuous with Lipschitz constant L and

$$f(x_j)=f_i(x_j),\quad 1\leqslant j\leqslant i. \tag{2.2}$$

ϵ-subdifferential of the model at $x\in Q$.

$$\begin{aligned}\partial_\epsilon f_i(x)&\equiv\{p\mid f_i(y)\geqslant f_i(x)-\epsilon+p^{\mathrm{T}}(y-x),\ \forall y\in\mathbb{R}^n\}\\ &=\Bigg\{p=\sum_{j=1}^{i}t_jf'(x_j)\ \Bigg|\ t_j\geqslant 0,\ \sum_{j=1}^{i}t_j=1,\\ &\qquad\qquad \sum_{j=1}^{i}t_j\{f(x_j)+(f'(x_j))^{\mathrm{T}}(x-x_j)\}\geqslant f_i(x)-\epsilon\Bigg\}.\end{aligned}$$

Remark 2.1.2. From (2.1), (2.2), it follows immediately that

$$\partial_\epsilon f_i(x_i) \subset \partial_\epsilon f(x_i) \,. \tag{2.3}$$

Model's best value. $f_*(i) = \min_Q f_i(\cdot)$.
Function's best value. $f^*(i) = \min\{f(x_1), \ldots, f(x_i)\}$.
Gap. $\Delta(i) = f^*(i) - f_*(i)$.
Best point. $x_i^* \in \operatorname{Argmin}\{f(x) \mid x \in \{x_1, \ldots, x_i\}\}$.

Remark 2.1.3. In view of (2.1), one has

$$f_*(1) \leqslant f_*(2) \leqslant \cdots \leqslant f_*(i) \leqslant f^*, \qquad f^*(1) \geqslant f^*(2) \geqslant \cdots \geqslant f^*(i) \geqslant f^* \,. \tag{2.4}$$

Remark 2.1.4. In view of (2.4), we have

$$f(x_i^*) - f^* \leqslant \Delta(i) \tag{2.5}$$

and

$$\Delta(1) \geqslant \Delta(2) \geqslant \cdots \geqslant \Delta(i) \geqslant 0 \,. \tag{2.6}$$

Truncated model. $\phi_i(x) = \max\{(f'(x_j))^{\mathrm{T}}(x - x_j) \mid 1 \leqslant j \leqslant i\}$.

Remark 2.1.5. Clearly,

$$\phi_1(x) \leqslant \phi_2(x) \leqslant \cdots, \tag{2.7}$$

and all $\phi_j(\cdot)$ are Lipschitz continuous with Lipschitz constant L.

Truncated model's best value. $\phi_*(i) = \min_Q \phi_i(\cdot)$.
Truncated gap. $\delta(i) = -\phi_*(i)$.

Remark 2.1.6. The following relations hold:

$$\delta(1) \geqslant \delta(2) \geqslant \cdots \geqslant \delta(i) \geqslant 0 \,, \tag{2.8}$$

$$\phi_i(x_i) \geqslant 0 \,, \quad f(x_i^*) - f^* \leqslant \delta(i) \,. \tag{2.9}$$

Monotonicity of $\delta(\cdot)$ immediately follows from (2.7). To prove nonnegativity of $\delta(i)$, let x^* be an optimal solution to (Min). Then $(f'(x_j))^{\mathrm{T}}(x^* - x_j) \leqslant 0$ for all j, so that $\phi_i(x^*) \leqslant 0$. (2.8) is proved. The first relation in (2.9) is evident. To prove the second relation, note that

$$f(x^*) \geqslant f(x_j) + (f'(x_j))^{\mathrm{T}}(x^* - x_j) \geqslant f(x_i^*) + (f'(x_j))^{\mathrm{T}}(x^* - x_j) \,, \quad j = 1, \ldots, i \,,$$

whence $f(x^*) \geqslant f(x_i^*) + \phi_i(x^*) \geqslant f(x_i^*) + \phi_*(i)$.

2.2. Methods

2.2.1. Level Method

The Level Method (LM) is defined as follows.

Parameters. $\lambda \in (0, 1)$.

Initialization. x_1 is an arbitrary point of Q.

ith step.

(1) Call the oracle, x_i being the input.

(2) Compute $f_*(i), f^*(i), x_i^*$.

(3) Set

$$l(i) = f_*(i) + \lambda\, \Delta(i) , \qquad x_{i+1} = \pi(x_i, \{x \mid x \in Q, f_i(x) \leqslant l(i)\}) .$$

Theorem 2.2.1. *For the above method,*

$$\epsilon(x_i^*) \leqslant \Delta(i) , \qquad i > c(\lambda)\left(\frac{V}{\epsilon}\right)^2 \quad \Rightarrow \quad \epsilon(x_i^*) \leqslant \epsilon ,$$

where

$$c(\lambda) = (1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1} .$$

(*Note that* $\min c(\cdot) = 4 = c(0.292\,89\ldots)$.)

Proof. In six steps.

(1) The efficiency estimate

$$\epsilon(x_i^*) \leqslant \Delta(i) \tag{2.10}$$

was established in (2.5).

(2) Set $S_i = [f_*(i), f^*(i)]$. Then (see (2.4)),

$$S_1 \supseteq S_2 \supseteq \cdots , \quad |S_i| = \Delta(i) , \tag{2.11}$$

where $|S|$ denotes the length of the segment S.

(3) First we need the following lemma.

Lemma. *Let* $i'' > i'$ *be such that*

$$\Delta(i'') \geqslant (1-\lambda)\Delta(i') . \tag{2.12}$$

Then,

$$f_*(i'') \leqslant l(i') . \tag{2.13}$$

Proof of the Lemma. Indeed, the length of the segment $\{s \in S_i \mid s \geqslant l(i')\}$ is $(1-\lambda)\Delta(i')$ and, since $S_{i'} \supseteq S_{i''}$ (2.11), the converse of (2.13) would imply $\Delta(i'') = |S_{i''}| < (1-\lambda) \cdot \Delta(i')$, which is impossible. □

(4) Let us fix $\epsilon > 0$ and assume that for certain N and all $i \leqslant N$ we have $\Delta(i) > \epsilon$. Let us split the integer segment $I = 1, \ldots, N$ in groups $I_1, \ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\Delta(i) \leqslant (1-\lambda)^{-1}\Delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \leqslant j_2$, for which $\Delta(i) \leqslant (1-\lambda)^{-1}\Delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \leqslant j_3$ satisfying $\Delta(i) \leqslant (1-\lambda)^{-1}\Delta(j_3)$, and so on. Let $u(i)$ be the minimizer of the function $f_{j_l}(\cdot)$ over Q. The preceding lemma, applied with an arbitrary $i' \in I_l$ and with $i'' = j_l$, demonstrates that $f_*(j_l) = f_{j_l}(u(l)) \leqslant l(i)$ for all $i \in I_l$. Formula (2.1) shows that $f_j(u(l)) \leqslant l(i)$ for all $i, j \in I_l$. Thus, we have established the following:

The (clearly convex) level sets $Q_i = \{x \in Q \mid f_i(x) \leqslant l(i)\}$

associated with $i \in I_l$, have a common point (namely, $u(l)$). (2.14)

(5) By virtue of standard properties of the projection mapping, (2.14) imply ($\operatorname{dist}(x, Q) \equiv \min_{y \in Q}|x-y|$):

$$\tau_{i+1} \equiv |x_{i+1} - u(l)|^2 \leqslant \tau_i - \operatorname{dist}^2\{x_i, Q_i\}, \quad i \in I_l. \tag{2.15}$$

We also have $f_i(x_i) - l(i) = f(x_i) - l(i) \geqslant f^*(i) - l(i) = (1-\lambda)\Delta(i)$ and $f_i(x_{i+1}) \leqslant l(i)$. From the Lipschitz property of f_i, it follows that

$$\operatorname{dist}\{x_i, Q_i\} = |x_i - x_{i+1}| \geqslant L^{-1}|f_i(x_i) - f_i(x_{i+1})| \geqslant L^{-1}(1-\lambda)\Delta(i).$$

Thus,

$$\tau_{i+1} \leqslant \tau_i - L^{-2}(1-\lambda)^2\Delta^2(i) \leqslant \tau_i - L^{-2}(1-\lambda)^2\Delta^2(j_l), \quad i \in I_l.$$

Because $0 \leqslant \tau_i \leqslant D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \leqslant D^2L^2(1-\lambda)^{-2}\Delta^{-2}(j_l). \tag{2.16}$$

(6) Form the definitions of N and of a group, we have

$$\Delta(j_1) = \Delta(N) > \epsilon, \qquad \Delta(j_{l+1}) > (1-\lambda)^{-1}\Delta(j_l).$$

These relations combined with (2.16) imply

$$N = \sum_{l \geqslant 1} N_l \leqslant d^2L^2(1-\lambda)^{-2}\sum_{l \geqslant 1}\epsilon^{-2}(1-\lambda)^{2(l-1)} = \left(\frac{V}{\epsilon}\right)^2(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}.$$

□

2.2.2. Proximal Level Method

The Proximal Level Method (PLM) is defined as follows.

Parameters. $\lambda \in (0,1)$; $\mu = (1-\lambda)$.

Initialization. x_1 is an arbitrary point of Q; $\Delta'(0) = \infty$.

ith step.
(1) Call the oracle, x_i being the input.
(2) Compute $f_*(i), f^*(i), x_i^*$ (keeping $x_i^* = x_{i-1}^*$ in the case of $f^*(i) = f^*(i-1)$).
(3) Set

$$l(i) = f_*(i) + \lambda \Delta(i) ,$$

$$l'(i) = \begin{cases} l(i) , & \text{if } \Delta(i) < \mu\, \Delta'(i-1) , \\ \min\{l(i), l'(i-1)\} , & \text{otherwise} , \end{cases}$$

$$\Delta'(i) = \begin{cases} \Delta(i) , & \text{if } \Delta(i) < \mu\, \Delta'(i-1) , \\ \Delta'(i-1) , & \text{otherwise} , \end{cases}$$

$$x_{i+1} = \pi(x_i^*, \{x \mid x \in Q, f_i(x) \leqslant l'(i)\}) .$$

Remark. The difference between PLM and LM is first that, in PLM, x_{i+1} is the projection of the ith best point x_i^* (and not the ith iterate x_i) onto the level set of the ith model f_i; second, the levels defining the above level sets are different: in LM this quantity, $l(i)$, divides in a fixed ratio the segment $[f_*(i), f^*(i)]$, and it can increase as well as decrease, as i varies, while in PLM the corresponding quantity is forbidden to increase until the gap $f^*(i) - f_*(i)$ decreases "substantially".

Theorem 2.2.2. *For the above method,*

$$\epsilon(x_i^*) \leqslant \Delta(i) , \qquad i > c(\lambda)\left(\frac{V}{\epsilon}\right)^2 \;\Rightarrow\; \epsilon(x_i^*) \leqslant \epsilon ,$$

$$c(\lambda) = (1-\lambda)^{-4}(2-\lambda)^{-1}\lambda^{-1}.$$

(*Note that* $\min c(\cdot) = 6.75 = c(0.183\,50\ldots)$.)

2.2.3. Dual Level Method

We also define a dual version of our level approach. It is directly inspired from [13], with a special value of e_i – here denoted $e^+(i)$ – and of the stepsize.

The Dual Level Method (DLM) is defined as follows.

Parameters. $\lambda, \mu \in (0, 1)$.

Initialization. x_1 is an arbitrary point of Q.

ith step.
(1) Call the oracle, x_i being the input.
(2) Compute $f^*(i), f_*(i), x_i^*$.
(3) Set

$$l(i) = f_*(i) + \lambda\, \Delta(i) \quad (= f^*(i) - (1-\lambda)\, \Delta(i)) ,$$

$$\epsilon^+(i) = f(x_i) - l(i) - \mu(1-\lambda)\, \Delta(i) .$$

(Note that $\epsilon^+(i) \geqslant 0$, since $f(x_i) - l(i) \geqslant f^*(i) - l(i) = (1-\lambda)\, \Delta(i)$.) Define p_i as the solution to the problem

$$\mathrm{P}(i): \text{minimize } |p|^2 \quad \text{subject to} \quad p \in \partial_{\epsilon^+(i)} f_i(x_i)$$

and set

$$x_{i+1} = \pi(x_i - \mu(1-\lambda)\Delta(i)|p_i|^{-2}p_i, Q) .$$

Theorem 2.2.3. *For the above method,*

$$\epsilon(x_i^*) \leqslant \Delta(1) , \qquad i > c(\lambda, \mu)\left(\frac{V}{\epsilon}\right)^2 \;\Rightarrow\; \epsilon(x_i^*) \leqslant \epsilon ,$$

$$c(\lambda, \mu) = \mu^{-2}(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1} .$$

(*Note that* $\min_\lambda c(\lambda, \mu) = 4\mu^{-2} = c(0.292\,89\ldots, \mu)$.)

2.2.4. Truncated Level Method

The Truncated Level Method (TLM) looks as follows.

Parameters. $\lambda \in (0, 1)$.

Initialization. x_1 is an arbitary point of Q.

ith step.

(1) Call the oracle, x_i being the input.

(2) Compute $\phi_*(i), f^*(i), x_i^*$.

(3) Set

$$l(i) = -(1-\lambda)\delta(i) , \qquad x_{i+1} = \pi(x_i, \{x \mid x \in Q, \phi_i(x) \leqslant l(i)\}) .$$

Remark. The difference between LM and TLM is that the latter method uses an artificial model which involves only subgradients, not the values of the objective. This feature of TLM is not valuable in the case of (Min), but it will be useful for (Var).

Theorem 2.2.4. *For the above method,*

$$\epsilon(x_i^*) \leqslant \delta(i) , \qquad i > c(\lambda)\left(\frac{V}{\epsilon}\right)^2 \;\Rightarrow\; \epsilon(x_i^*) \leqslant \epsilon ,$$

$$c(\lambda) = (1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1} .$$

(*Note that* $\min c(\cdot) = 4 = c(0.292\,89\ldots)$.)

3. Methods for (Sad)

3.0. Initial scaling

In what follows we assume that the diameters of Q and H coincide; D denotes their (common) value. This assumption can be provided by an appropriate isotropic scaling of, say, the y-variable. Note that the quantity $L_x D_x + L_y D_y$ remains invariant under this scaling. We denote $L = \max\{L_x, L_y\}$.

3.1. Notation

Denote

$$\bar{f}(x) = \max_H f(x, \cdot) : Q \to \mathbb{R}, \qquad \underline{f}(y) = \min_Q f(\cdot, y) : H \to \mathbb{R}.$$

(These are, respectively, the worst-case payment of the player choosing x and the worst-case income of the player choosing y in the game associated with f.)

Assume we have called the oracle at the points $(x_1, y_1), \dots, (x_i, y_i) \in Q$. Then the following objects are defined:

Models.

x-model.

$$\bar{f}_i(x) = \max\{f(x_j, y_j) + (f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j) \mid 1 \leqslant j \leqslant i\} : Q \to \mathbb{R}.$$

y-model.

$$\underline{f}_i(y) = \min\{f(x_j, y_j) + (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j) \mid 1 \leqslant j \leqslant i\} : H \to \mathbb{R}.$$

model.

$$f_i(x, y) = \bar{f}_i(x) - \underline{f}_i(y) : Q \times H \to \mathbb{R}.$$

Remark 3.1.1. Clearly, $\bar{f}_i$ is convex, $\underline{f}_i$ is concave,

$$\bar{f}_1(x) \leqslant \bar{f}_2(x) \leqslant \cdots \leqslant \bar{f}_i(x) \leqslant \bar{f}(x), \quad x \in Q, \tag{3.1}$$

$$\underline{f}_1(y) \geqslant \underline{f}_2(y) \geqslant \cdots \geqslant \underline{f}_i(y) \geqslant \underline{f}(y), \quad y \in H, \tag{3.2}$$

$\bar{f}_i, \underline{f}_i$ are Lipschitz continuous with Lipschitz constant L. Consequently,

$$f_1(x, y) \leqslant f_2(x, y) \leqslant \cdots \leqslant f_i(x, y) \leqslant \bar{f}(x) - \underline{f}(y), \quad (x, y) \in Q \times H, \tag{3.3}$$

and f_i is Lipschitz continuous with Lipschitz constant $2^{1/2}L$.

ϵ-subdifferential of the model at $x \in Q$.

$$\partial_\epsilon f_i(x, y) \equiv \{p \in \mathbb{R}^n \times \mathbb{R}^{n'} \mid f_i(u, v) \geqslant f_i(x, y) - \epsilon + p^{\mathrm{T}}((u, v) - (x, y))$$
$$\forall (u, v) \in \mathbb{R}^n \times \mathbb{R}^{n'}\}.$$

Model's best value. $f_*(i) = \min_{Q \times H} f_i(\cdot, \cdot)$.

Gap. $\Delta(i) = -f_*(i)$.

Remark 3.1.2. The following relations hold:

$$\Delta(1) \geqslant \Delta(2) \geqslant \cdots \geqslant \Delta(i) \geqslant 0; \quad f_i(x_i, y_i) \geqslant 0. \tag{3.4}$$

Indeed, the monotonicity of $\Delta(\cdot)$ follows from (3.3). Let us prove that $\Delta(\cdot)$ is nonnegative. Let $f^* = \min_Q \bar{f}(\cdot)$; by von Neumann's lemma, one also has $f^* = \max_H \underline{f}(\cdot)$. It follows that $\min_{Q\times H}(\bar{f}(x) - \underline{f}(y)) = f^* - f^* = 0$, and the first relation in (3.4) follows from (3.3). On the other hand, clearly $\bar{f}_i(x_i) \geqslant f(x_i, y_i)$, $\underline{f}_i(y_i) \leqslant f(x_i, y_i)$, which implies the second relation in (3.4).

Truncated model.

$$\phi_i(x, y)$$
$$= \max\{(f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j) - (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j) \mid 1 \leqslant j \leqslant i\} : Q \times H \to \mathbb{R}.$$

Remark 3.1.3. Clearly, ϕ_i is convex and Lipschitz continuous with Lipschitz constant $2^{1/2}L$, and

$$\phi_1(\cdot, \cdot) \leqslant \phi_2(\cdot, \cdot) \leqslant \cdots . \tag{3.5}$$

Truncated model's best value. $\phi_*(i) = \min_{Q\times H} \phi_i(\cdot, \cdot)$.
Truncated gap. $\delta(i) = -\phi_*(i)$.

Remark 3.1.4. We have

$$\delta(1) \geqslant \delta(2) \geqslant \cdots \geqslant 0 . \tag{3.6}$$

The monotonicity of $\delta(\cdot)$ follows from (3.5). Let us prove that $\delta(i) \leqslant 0$. Indeed, let (x^*, y^*) be a saddle point of f and let $(x, y) \in Q \times H$. We have

$$f(x^*, y) \geqslant (f'_x(x, y))^{\mathrm{T}}(x^* - x) + f(x, y),$$
$$f(x, y^*) \leqslant (f'_y(x, y))^{\mathrm{T}}(y^* - y) + f(x, y) ,$$

whence

$$f(x^*, y) - f(x, y^*) \geqslant (f'_x(x, y))^{\mathrm{T}}(x^* - x) - (f'_y(x, y))^{\mathrm{T}}(y^* - y) .$$

Since (x^*, y^*) is saddle point, $f(x^*, y) - f(x, y^*) \leqslant 0$, so that

$$(f'_x(x, y))^{\mathrm{T}}(x^* - x) - (f'_y(x, y))^{\mathrm{T}}(y^* - y) \leqslant 0 , \quad (x, y) \in Q \times H .$$

In other words, $\phi_i(x^*, y^*) \leqslant 0$.

3.2. Methods

3.2.1. Level Method

The Level Method (LM) is defined as follows.
Parameters. $\lambda \in (0, 1)$.
Initialization. (x_1, y_1) is an arbitrary point of $Q \times H$.
ith step.

(1) Call the oracle, (x_i, y_i) being the input.
(2) Compute $f_*(i)$, i.e., solve the pair of convex problems
$P_x(i)$: minimize

$$\bar{f}_i(x)=\max\{f(x_j, y_j)+(f'_x(x_j, y_j))^{\mathrm{T}}(x-x_j) \mid 1\leqslant j\leqslant i\}$$

subject to $x\in Q$,

and $P_y(i)$: maximize

$$\underline{f}_i(y)=\max\{f(x_j, y_j)+(f'_y(x_j, y_j))^{\mathrm{T}}(y-y_j) \mid 1\leqslant j\leqslant i\}$$

subject to $y\in H$.
(3) Set

$$l(i)=f_*(i)+\lambda\,\Delta(i)\,,$$

$$(x_{i+1}, y_{i+1})=\pi((x_i, y_i), \{(x, y) \mid (x, y)\in Q\times H, f_i(x, y)\leqslant l(i)\})\,.$$

ith approximate solution. It is defined as follows. When solving the problems $P_x(i)$ and $P_y(i)$, we find also optimal dual solutions, i.e., the quantities $\{t_i(j), s_i(j)\}_{1\leqslant j\leqslant i}$, satisfying

$$\sum_{j=1}^{i} t_i(j)=1\,,\quad t_i(j)\geqslant 0\,,$$

$$\min_{x\in Q}\sum_{j=1}^{i} t_i(j)\{f(x_j, y_j)+(f'_x(x_j, y_j))^{\mathrm{T}}(x-x_j)\}=\min_{Q}\bar{f}_i(\cdot)\,,$$

$$\sum_{j=1}^{i} s_i(j)=1\,,\quad s_i(j)\geqslant 0\,,$$

$$\max_{y\in H}\sum_{j=1}^{i} s_i(j)\{f(x_j, y_j)+(f'_y(x_j, y_j))^{\mathrm{T}}(y-y_j)\}=\max_{H}\underline{f}_i(\cdot)\,,$$

and the ith approximate solution is defined as

$$\left(x_i^*=\sum_{j=1}^{i} s_i(j)x_j,\ y_i^*=\sum_{j=1}^{i} t_i(j)y_j\right).$$

Theorem 3.2.1. *For the above method,*

$$\epsilon(x_i^*, y_i^*)\leqslant\Delta(i)\,,\qquad i>c(\lambda)\left(\frac{V}{\epsilon}\right)^2\ \Rightarrow\ \epsilon(x_i^*, y_i^*)\leqslant\epsilon\,,$$

where

$$c(\lambda)=4(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}\,.$$

(*Note that* $\min c(\cdot)=16=c(0.292\,89\ldots)$.)

Proof. In five steps.

(1) Let us fix $(x, y) \in Q \times H$. We have

$$f(x_j, y) \leqslant f(x_j, y_j) + (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j) ,$$
$$f(x, y_j) \geqslant f(x_j, y_j) + (f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j) .$$

It follows that

$$\sum_{j=1}^{i} t_i(j) f(x, y_j) - \sum_{j=1}^{i} s_i(j) f(x_j, y)$$
$$\geqslant \sum_{j=1}^{i} t_i(j) \{f(x_j, y_j) + (f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j)\}$$
$$- \sum_{j=1}^{i} s_i(j) \{f(x_j, y_j) + (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j)\} .$$

Since f is convex in x and concave in y, we have

$$\sum_{j=1}^{i} t_i(j) f(x, y_j) - \sum_{j=1}^{i} s_i(j) f(x_j, y) \leqslant f(x, y_i^*) - f(x_i^*, y) .$$

Thus,

$$f(x, y_i^*) - f(x_i^*, y) \geqslant \sum_{j=1}^{i} t_i(j) \{f(x_j, y_j) + (f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j)\}$$
$$- \sum_{j=1}^{i} s_i(j) \{f(x_j, y_j) + (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j)\} ,$$
$$(x, y) \in Q \times H .$$

Taking the minimum over $(x, y) \in Q \times H$ and using the definition of $t_i(j)$, $s_i(j)$, we obtain

$$\bar{f}(x_i^*) - \underline{f}(y_i^*) \leqslant \max_H \underline{f}_i(\cdot) - \min_Q \bar{f}_i(\cdot) = - \min_{Q \times H} f_i(\cdot, \cdot) .$$

In other words,

$$\epsilon(x_i^*, y_i^*) \leqslant \Delta(i) , \qquad (3.7)$$

as is required in the accuracy estimate.

(2) Set $S_i = [\phi_*(i), 0]$. Then (see (3.3), (3.4)), $S_i \neq \emptyset$ and

$$S_1 \supseteq S_2 \supseteq \ldots, \quad |S_i| = \Delta(i) , \qquad (3.8)$$

where $|S|$ denotes the length of the segment S.

(3) Let us fix $\epsilon > 0$ and assume that for certain N and all $i \leqslant N$ we have $\delta(i) > \epsilon$. Let us split the integer segment $I = 1, \ldots, N$ in groups $I_1, \ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which

$\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \leqslant j_2$, for which $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \leqslant j_3$ satisfying $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_3)$, and so on.

Let $(u(l), v(l))$ minimize the function $f_{j_l}(\cdot, \cdot)$ over $Q \times H$. For $i \in I_l$ from (3.8), the definition of $l(i)$ and the relation $\delta(j_l) \geqslant (1-\lambda)\delta(i)$, $i \in I_l$, it immediately follows that $f_*(j_l)) = f_{j_l}(u(l), v(l)) \leqslant l(i)$ for all $i \in I_l$. Eq. (3.3) shows that $f_j(u(l), v(l)) \leqslant l(i)$ for all $i, j \in I_l$. Thus, we have established the following:

$$\text{the (clearly convex) level sets } Q_i = \{z \in Q \times H \mid f_i(z) \leqslant l(i)\} \text{ associated with}$$
$$i \in I_l, \text{ have a common point (namely, } z(l) = (u(l), v(l))) \,. \tag{3.9}$$

(4) By virtue of the standard properties of the projection mapping, (3.9), under the notation $z_i = (x_i, y_i)$, implies

$$\tau_{i+1} \equiv |z_{i+1} - z(l)|^2 \leqslant \tau_i - \operatorname{dist}^2\{z_i, Q_i\}\,, \quad i \in I_l\,. \tag{3.10}$$

We also have $f_i(z_i) - l(i) \geqslant -l(i)$ (see (3.4)), whence $f_i(z_i) - l(i) \geqslant (1-\lambda)\delta(i)$, while $f_i(z_{i+1}) \leqslant l(i)$. Since f_i is Lipschitz continuous with constant $2^{1/2}L$, it follows that

$$\operatorname{dist}\{z_i, Q_i\} = |z_i - z_{i+1}| \geqslant 2^{-1/2}L^{-1}|f_i(z_i) - f_i(z_{i+1})| \geqslant 2^{-1/2}L^{-1}(1-\lambda)\delta(i)\,.$$

Thus,

$$\tau_{i+1} \leqslant \tau_i - 2^{-1}L^{-2}(1-\lambda)^2\delta^2(i) \leqslant \tau_i - 2^{-1}L^{-2}(1-\lambda)^2\delta^2(j_l)\,, \quad i \in I_l\,.$$

Because $0 \leqslant \tau_i \leqslant 2D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \leqslant 4D^2L^2(1-\lambda)^{-2}\delta^{-2}(j_l)\,. \tag{3.11}$$

(5) From the definitions of N and of a group, we have

$$\delta(j_1) = \delta(N) > \epsilon\,, \quad \delta(j_{l+1}) > (1-\lambda)^{-1}\delta(j_l)\,.$$

These relations combined with (3.11) imply

$$N = \sum_{l \geqslant 1} N_l \leqslant 4D^2L^2(1-\lambda)^{-2} \sum_{l \geqslant 1} \epsilon^{-2}(1-\lambda)^{2(l-1)}$$
$$= 4\left(\frac{V}{\epsilon}\right)^2 (1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}\,. \quad \square$$

3.2.2. Dual Level Method

Because no oracle is available to measure the decrease of the actual gap function $F(x, y) = \bar{f}(x) - \underline{f}(y)$, no proximal version of the method can be constructed. We therefore study directly the Dual Level Method (DLM), defined as follows.

Parameters. $\lambda, \mu \in (0, 1)$.

Initialization. (x_1, y_1) is an arbitrary point of $Q \times H$.
ith step.
(1) Call the oracle, (x_i, y_i) being the input.
(2) Compute $f_*(i)$, $\{t_i(j), s_i(j)\}_{1 \leqslant j \leqslant i}$ as in the Level Method 3.2.1.
(3) Set

$$l(i) = f_*(i) + \lambda\, \Delta(i) \quad (= -(1-\lambda)\Delta(i)),$$

$$\epsilon^+(i) = f_i(x_i, y_i) - l(i) - \mu(1-\lambda)\, \Delta(i).$$

(Note that $f_i(x_i, y_i) \geqslant 0$, see (3.4), so that $\epsilon^+(i) \geqslant 0$.)
Define $p_i \in \mathbb{R}^n \times \mathbb{R}^{n'}$ as the solution to the problem
P(i): minimize $|p|^2$ subject to $p \in \partial_{\epsilon^+(i)} f_i(x_i, y_i)$
and set

$$(x_{i+1}, y_{i+1}) = \pi((x_i, y_i) - \mu(1-\lambda)\Delta(i)\,|p_i|^{-2} p_i,\ Q \times H).$$

ith approximate solution. It is defined as

$$\left(x_i^* = \sum_{j=1}^{i} s_i(j) x_j,\ y_i^* = \sum_{j=1}^{i} t_i(j) y_j\right),$$

where $\{s_i(j)\}_j$ and $\{t_i(j)\}_j$ are the same as in Section 3.2.1, namely, the optimal dual solutions to $\mathrm{P}_y(i)$, $\mathrm{P}_x(i)$, respectively.

Theorem 3.2.2. *For the above method,*

$$\epsilon(x_i^*, y_i^*) \leqslant -f_*(i), \qquad i \geqslant c(\lambda, \mu)\left(\frac{V}{\epsilon}\right)^2 \ \Rightarrow\ \epsilon(x_i^*, y_i^*) \leqslant \epsilon,$$

where

$$c(\lambda, \mu) = 4\mu^{-2}(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}.$$

(*Note that* $\min_\lambda c(\lambda, \mu) = 16\mu^{-2} = c(0.292\,89\ldots, \mu)$.)

3.2.3. Truncated Level Method

The Truncated Level Method (TLM) is defined as follows
Parameter. $\lambda \in (0, 1)$.
Initialization. (x_1, y_1) is an arbitrary point of $Q \times H$.
ith step.
(1) Call the oracle, (x_i, y_i) being the input.
(2) Compute $\phi_*(i)$, i.e., solve the convex programming problem
$\mathrm{P}_{x,y}(i)$: minimize

$$\phi_i(x, y) = \max\{(f'_x(x_j, y_j))^{\mathrm{T}}(x - x_j) - (f'_y(x_j, y_j))^{\mathrm{T}}(y - y_j) \mid 1 \leqslant j \leqslant i\}$$

subject to $(x, y) \in Q \times H$.
(3) Set

$$l(i) = -(1-\lambda)\delta(i) ,$$

$$(x_{i+1}, y_{i+1}) = \pi((x_i, y_i) , \{(x, y) \mid (x, y) \in Q \times H, \phi_i(x, y) \leqslant l(i)\}) .$$

ith approximate solution. It is defined as follows:

$$\left(x_i^* = \sum_{j=1}^{i} r_i(j) x_j, y_i^* = \sum_{j=1}^{i} r_i(j) y_j\right),$$

where the quantities $\{r_i(j)\}_{1 \leqslant j \leqslant i}$ form an optimal dual solution to $P_{x,y}(i)$, i.e., these quantities satisfy the relations

$$\sum_{j=1}^{i} r_i(j) = 1 , \quad r_i(j) \geqslant 0 ,$$

$$\min_{(x,y) \in Q \times H} \sum_{j=1}^{i} r_i(j) \{ (f_x'(x_j, y_j))^{\mathrm{T}} (x - x_j) - (f_y'(x_j, y_j))^{\mathrm{T}} (y - y_j) \} = \min_{Q \times H} \phi_i(\cdot, \cdot) .$$

Theorem 3.2.3. *For the above method,*

$$\epsilon(x_i^*, y_i^*) \leqslant \delta(i) , \qquad i > c(\lambda) \left(\frac{V}{\epsilon}\right)^2 \Rightarrow \epsilon(x_i^*, y_i^*) \leqslant \epsilon ,$$

where

$$c(\lambda) = 4(1-\lambda)^{-2} \lambda^{-1} (2-\lambda)^{-1} .$$

(*Note that* $\min c(\cdot) = 16 = c(0.292\,89\ldots)$.)

4. Methods for (CMin)

4.0. Additional assumption

In what follows we assume that there exists $x \in 0$ with $G(x) > 0$, so that the problem is really a constrained one.

4.1. Notation

Assume we have called the oracle at the points $x_1, \ldots, x_i \in Q$. Then the following objects are defined.

Model of f.

$$f_i(x) = \max\{f(x_j) + (f'(x_j))^{\mathrm{T}} (x - x_j) \mid 1 \leqslant j \leqslant i\} .$$

Model of G.

$$G_i(x) = \max\{g_k(x_j) + (g_k'(x_j))^{\mathrm{T}} (x - x_j) \mid 1 \leqslant j \leqslant i, 1 \leqslant k \leqslant m\} .$$

Remark 4.1.1. Clearly,

$$f_1(x) \leqslant f_2(x) \leqslant \cdots \leqslant f_i(x) \leqslant f(x), \quad x \in Q, \tag{4.1}$$

$$G_1(x) \leqslant G_2(x) \leqslant \cdots \leqslant G_i(x) \leqslant G(x), \quad x \in Q, \tag{4.2}$$

$$f_i(x_j) = f(x_j), \quad G_i(x_j) = G(x_j), \qquad 1 \leqslant j \leqslant i, \tag{4.3}$$

and the functions f_i, G_i are Lipschitz continuous with Lipschitz constant L.

Model's best value. $f_*(i) = \min\{f_i(\cdot) \mid x \in Q, G_i(x) \leqslant 0\}$.

Remark 4.1.2. From Remark 4.1.1 it follows immediately that $f_*(i)$ is well-defined and

$$f_*(1) \leqslant f_*(2) \leqslant \cdots \leqslant f_*(i) \leqslant f^*. \tag{4.4}$$

Admissible set. $T(i) = \{(f(x_j), G(x_j)) \mid 1 \leqslant j \leqslant i\} \subset \mathbb{R}^2$.
Completed admissible set. $C(i) = (\operatorname{Conv} T(i)) + \mathbb{R}^2_+$.

In what follows we will use the natural order in $\mathbb{R}^2$:

$$(u_1, v_1) \leqslant (u_2, v_2) \quad \text{if } u_1 \leqslant u_2 \text{ and } v_1 \leqslant v_2.$$

4.2. Constrained Level Method (CLM)

4.2.1. Preliminary remarks

Assume we have called the oracle at the points $x_1, \ldots, x_i \in Q$. Then, besides the objects described in Section 4.1, we can define also the following.
Support function.

$$\begin{aligned} h_i(\alpha) &\equiv \min\{\alpha(f(x_j) - f_*(i)) + (1-\alpha)G(x_j) \mid 1 \leqslant j \leqslant i\} \\ &= \min\{\alpha(u - f_*(i)) + (1-\alpha)v \mid (u, v) \in C(i)\} : [0, 1] \to \mathbb{R}. \end{aligned}$$

Gap.

$$\Delta(i) = \max\{h_i(\alpha) \mid 0 \leqslant \alpha \leqslant 1\}.$$

Best point. Let $(u(i), v(i)) \in \operatorname{Argmin}\{\rho(u - f_*(i), v) \mid (u, v) \in C(i)\}$, where

$$\rho(p, q) = \max\{(p)_+, (q)_+\}.$$

Then there clearly exists a convex combination $\sum_{j=1}^{i} r_i(j)(f(x_j), G(x_j))$ of points belonging to $T(i)$, such that $\sum_{j=1}^{i} r_i(j)(f(x_j), G(x_j)) \leqslant (u(i), v(i))$. Set

$$x_i^* = \sum_{j=1}^{i} r_i(j)\, x_j;$$

this is *the best point* associated with $x_1, \ldots, x_i$.

Remark 4.2.1.1. (1) We have

$$x_i^* \in Q, \quad \epsilon(x_i^*) \leqslant \min\{\rho(u-f_*(i), v) \mid (u, v) \in C(i)\} = \Delta(i). \tag{4.5}$$

The inclusion in (4.5) is evident. The inequality follows from the relations

$$\begin{aligned}(f(x_i^*)-f_*(i), G(x_i^*)) &\leqslant \sum_{j=1}^{i} r_i(j)\ (f(x_j)-f_*(i), G(x_j)) \leqslant (u(i)-f_*(i), v(i)) \\ &\leqslant \min\{\rho(u-f_*(i), v) \mid (u, v) \in C(i)\}(1,1).\end{aligned}$$

(We have taken into account the convexity of f and G.) Since $f^* \geqslant f_*(i)$ (see (4.4)), the resulting inequality implies the inequality in (4.5).

Now let us prove that

$$\Delta(i) = \min\{\rho(u-f_*(i), v) \mid (u, v) \in C(i)\}. \tag{4.6}$$

Indeed, $\rho(p, q) = (\max\{\alpha p + (1-\alpha)q \mid 0 \leqslant \alpha \leqslant 1\})_+$, whence

$$\begin{aligned}&\min_{(u,v)\in C(i)} \left(\max_{\alpha\in[0,1]} \{\alpha(u-f_*(i)) + (1-\alpha)v\}\right)_+ \\ &= \left(\min_{(u,v)\in C(i)} \max_{\alpha\in[0,1]} \{\alpha(u-f_*(i)) + (1-\alpha)v\}\right)_+ \\ &= \left(\max_{\alpha\in[0,1]} \min_{(u,v)\in C(i)} \{\alpha(u-f_*(i)) + (1-\alpha)v\}\right)_+ \\ &= \left(\max_{\alpha\in[0,1]} h_i(\alpha)\right)_+ = (\Delta(i))_+.\end{aligned}$$

It remains to verify that $\Delta(i) \geqslant 0$. We have

$$\begin{aligned}\Delta(i) = \max_{0\leqslant\alpha\leqslant 1} h_i(\alpha) &= \max_{0\leqslant\alpha\leqslant 1} \min_{(u,v)\in C(i)} \{\alpha(u-f_*(i)) + (1-\alpha)v\} \\ &= \min_{(u,v)\in C(i)} \max_{0\leqslant\alpha\leqslant 1} \{\alpha(u-f_*(i)) + (1-\alpha)v\}\end{aligned}$$

(we have used von Neumann's lemma), and to prove nonnegativity of $\Delta(i)$ it suffices to verify that if $(u, v) \in C(i)$ and $v \leqslant 0$, then $u \geqslant f_*(i)$. Indeed, if $(u, v) \in C(i)$, then

$$(u, v) \geqslant \left(\sum_{j\leqslant i} t_j f(x_j), \sum_{j\leqslant i} t_j g(x_j)\right),$$

with certain $t_j \geqslant 0$, $\sum_{j\leqslant i} t_j = 1$. If $v \leqslant 0$, then $\sum_{j\leqslant i} t_j g(x_j) \leqslant 0$, whence

$$G\left(\sum_{j\leqslant i} t_j x_j\right) \leqslant 0,$$

and therefore

$$G_i\Big(\sum_{j\leqslant i} t_j x_j\Big) \leqslant 0 .$$

We see that $\sum_{j\leqslant i} t_j x_j$ is feasible in the problem defining $f_*(i)$; hence

$$f_*(i) \leqslant f_i\Big(\sum_{j\leqslant i} t_j x_j\Big) \leqslant f\Big(\sum_{j\leqslant i} t_j x_j\Big) \leqslant \sum_{j\leqslant i} t_j f(x_j) \leqslant u .$$

(2) One has

$$h_1(\alpha) \geqslant h_2(\alpha) \geqslant \cdots, \quad \alpha \in [0, 1] , \tag{4.7}$$

and $h_i(\cdot)$ is a concave Lipschitz continuous function with Lipschitz constant $2V$.

The monotonicity of $h_i(\cdot)$ in i immediately follows from (4.1), (4.2) and (4.4). Since f_i is Lipschitz continuous with constant L and $f(x_j) = f_i(x_j)$, $j \leqslant i$, we have $|f(x_j) - f_*(i)| \leqslant V$, and since G is Lipschitz continuous with the same constant and takes on Q positive (see Section 4.0) as well as nonpositive (since the problem is consistent) values, we have $|G(x_j)| \leqslant V$, so that $h_i(\cdot)$ is Lipschitz continuous with the constant $2V$. The concavity of h is evident.

4.2.2. Description of CLM

Parameters. $\lambda, \mu \in (0, 1)$.

Initialization. x_1 is an arbitrary point of Q.

ith step.

(1) Call the oracle, x_i being the input.

(2) Compute $f_*(i)$, $h_i(\cdot)$, $\Delta(i)$, x_i^*.

(3) Define $\alpha_{\min}(i)$ as the smallest, and $\alpha_{\max}(i)$ the largest of $\alpha \in [0, 1]$ such that $h_i(\alpha) \geqslant 0$.

For $i = 1$ set

$$\alpha(1) = \tfrac{1}{2}(\alpha_{\min}(1) + \alpha_{\max}(1)) ,$$

and for $i > 1$ set

$$\alpha(i) = \begin{cases} \tfrac{1}{2}(\alpha_{\min}(i) + \alpha_{\max}(i)) , & \text{if } \dfrac{\alpha(i-1) - \alpha_{\min}(i)}{\alpha_{\max}(i) - \alpha_{\min}(i)} \notin [\tfrac{1}{2}\mu, 1 - \tfrac{1}{2}\mu] , \\[2ex] \alpha(i-1) , & \text{otherwise} . \end{cases}$$

(4) Set

$$w(i) = \alpha(i) f_*(i) , \quad W(i) = \min_{1\leqslant j\leqslant i} (\alpha(i) f(x_j) + (1 - \alpha(i)) G(x_j)) ,$$

$$l(i) = w(i) + \lambda(W(i) - w(i)) ,$$

$$x_{i+1} = \pi(x_i, \{x \mid x \in Q, \alpha(i) f_i(x) + (1 - \alpha(i)) G_i(x) \leqslant l(i)\}) .$$

Theorem 4.2.2. *For the above method,*

$$\epsilon(x_i^*) \leqslant \Delta(i) ,$$

and if $0<\epsilon<V$, *then the following implication holds*:

$$i>c(\lambda, \mu)\left(\frac{2V}{\epsilon}\right)^2 \ln\left(\frac{2V}{\epsilon}\right) \Rightarrow \epsilon(x_i^*) \leqslant \epsilon ,$$

where

$$c(\lambda, \mu)=2(\ln 2)^{-1}\mu^{-2}\left\{\ln\left(\frac{2}{1+\mu}\right)\right\}^{-1}(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1} .$$

(*Note that* $\min c(\cdot, \cdot)=c(0.292\,89\ldots, 0.644\,53\ldots) \leqslant 142$.)

Proof. In six steps.

(1) The efficiency estimate

$$\epsilon(x_i^*) \leqslant \Delta(i)$$

was established in Remark 4.2.1.1.

(2) Let $\epsilon>0$ and let N be such that $\Delta(N)>\epsilon$. Let us split the integer segment $I=1, \ldots, N$ into sequential groups $J_1, \ldots, J_k$ in such a way that $\alpha(i) \equiv \alpha_l$ is constant for $i \in J_l$ and $\alpha_l \neq \alpha_{l+1}$. Let p_l be the first, and q_l be the last element of J_l.

(3) Let us prove that the amount k of groups satisfies the relation

$$k \leqslant \left\{\ln\left(\frac{2}{1+\mu}\right)\right\}^{-1} \ln\left(\frac{2V}{\epsilon}+1\right)+1 . \tag{4.8}$$

Indeed, let $T_0=[0, 1]$, $T_i=[\alpha_{\min}(i), \alpha_{\max}(i)]$, $i \geqslant 1$. Then $T_i \supseteq T_{i+1}$ (see (4.7)) and $h_i(\cdot)$ is negative outside T_i. Note that α_l is the center of T_{p_l} and for $l<k$, either α_l does not belong to T_{q_l+1}, or this segment is divided by α_l into parts such that at least one of them is less than $\frac{1}{2}\mu|T_{q_l+1}|$. Since $T_{q_l+1} \subset T_{p_l}$, it follows that $|T_{q_l+1}| \leqslant \frac{1}{2}(1+\mu)|T_{p_l}|=\frac{1}{2}(1+\mu)|T_{q_{l-1}+1}|$, where $q_0=0$. Thus, if $k>1$, then $|T_N| \leqslant |T_{q_{k-1}+1}| \leqslant (\frac{1}{2}(1+\mu))^{k-1}$. Since $h_N(\cdot)$ is negative outside T_N and is Lipschitz continuous with the constant $2V$ (Remark 4.2.1.1 (2)), it follows that in the case of $k>1$ we have $\Delta(N)=\max_{0 \leqslant \alpha \leqslant 1} h_N(\alpha) \leqslant 2V(\frac{1}{2}(1+\mu))^{k-1}$. Since $\Delta(N)>\epsilon$, we obtain in the case of $k>1$: $k \leqslant \{\ln(2/(1+\mu))\}^{-1}\ln(2V/\epsilon)+1$, which implies (4.8).

(4) Now let us prove that the amount of elements, M_l, in the group J_l satisfies the relation

$$M_l \leqslant (2/\mu)^2(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1}\left(\frac{V}{\epsilon}\right)^2 . \tag{4.9}$$

Let $\delta(i)=h_i(\alpha_l)$ $(=W(i)-w(i))$. Since $\alpha(i)=\alpha_l$, $i \in J_l$, from definition of $W(i)$, $w(i)$ and (4.4) it follows immediately that

$$W(p_l) \geqslant W(p_l+1) \geqslant \cdots \geqslant W(q_l) , \tag{4.10}$$

$$w(p_l) \leqslant \cdots \leqslant w(q_l) , \tag{4.11}$$

so that

$$\delta(p_l) \geqslant \delta(p_l+1) \geqslant \cdots \geqslant \delta(q_l) . \tag{4.12}$$

Let us prove that

$$\delta(q_l) \geqslant \tfrac{1}{2}\mu\epsilon . \tag{4.13}$$

Indeed, α_l splits the segment T_{ql} in two parts, each not shorter than $\frac{1}{2}\mu|T_{ql}|$; $h_{ql}(\cdot)$ is nonnegative on T_{ql} and concave, so that

$$\max_{T_{ql}} h_{ql}(\cdot) \leqslant (2/\mu) h_{ql}(\alpha_l) = (2/\mu)\,\delta(q_l) .$$

Outside T_{ql} the function $h_{ql}(\cdot)$ is negative; thus,

$$\Delta(q_l) = \max_{T_{ql}} h_{ql}(\cdot) \leqslant (2/\mu)\,\delta(q_l) .$$

Since $\Delta(q_l) > \epsilon$, we obtain (4.13).

(5) Let us split the integer segment J_l into groups $I_1, \ldots, I_s$ as follows. The last element of I_1 is $j_1 \equiv q_l$, and I_1 consists precisely of those $i \in J_l$ for which $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_1)$. The largest $i \in J_l$ which does not belong to I_1, if such an i exists, is the last element, j_2, of the second group I_2, and I_2 consists precisely of those $i \in J_l$, $i \leqslant j_2$, for which $\delta(i) \leqslant (1-\lambda)^{-1} \cdot \delta(j_2)$. The last element of J_l which does not belong to $I_1 \cup I_2$, if such an element exists, is the last element, j_3, of I_3, and I_3 consists of those $i \in J_l$, $i \leqslant j_3$, for which $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_3)$, and so on. Let us prove that the number of elements, N_r, in the group I_r, satisfies the relation

$$N_r \leqslant D^2L^2(1-\lambda)^{-2}\delta^{-2}(j_r) . \tag{4.14}$$

Indeed, let $i \in I_r$ and let $S_i = [w(i), W(i)]$. Then (see (4.10)–(4.12)) S_i are nonempty segments, $|S_i| = \delta(i)$; besides this, $S_{i+1} \subseteq S_i$, $i+1 \in I_l$. Let $\phi^l(x) = \alpha_l f(x) + (1-\alpha_l)G(x)$, and let $\phi_i(x) = \alpha_l f_i(x) + (1-\alpha_l)G_i(x)$. Then clearly

$$\phi_{i_r}(\cdot) \leqslant \phi_{i_r+1}(\cdot) \leqslant \cdots \leqslant \phi_{j_r}(\cdot) \leqslant \phi^l(\cdot) , \tag{4.15}$$

where i_r is the first element of I_r, and

$$\phi_i(x_i) \geqslant W(i) , \quad i \in I_r , \quad \min_Q \phi_i(\cdot) \leqslant w(i) . \tag{4.16}$$

Let $u(r)$ minimize $\phi_{j_r}(\cdot)$ over Q. Then (see (4.16)) $\phi_{j_r}(u(r)) \leqslant w(j_r)$, so that (see (4.15)) $\phi_i(u(r)) \leqslant w(j_r)$, $i \in I_r$. On the other hand, for $i \in I_r$ we have $l(i) = w(i) + \lambda(W(i) - w(i)) = W(i) - (1-\lambda)\delta(i) \geqslant W(j_r) - \delta(j_r) = w(j_r)$ (we have taken into account that $W(i) \geqslant W(j_r)$ and $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_r)$, $i \in I_r$). Thus, $\phi_i(u(r)) \leqslant l(i)$, $i \in I_r$. We have proved that

$$\text{the (clearly convex) level sets } Q_i = \{x \in Q \mid \phi_i(x) \leqslant l(i)\},\ i \in I_r, \tag{4.17}$$

have a common point (namely, $u(r)$) .

Now, $x_{i+1}=\pi(x_i, Q_i)$, $i\in I_r$. In view of the standard properties of the projection mapping, we have

$$\tau_{i+1}\equiv |x_{i+1}-u(r)|^2\leqslant \tau_i-\mathrm{dist}^2\{x_i, Q_i\}\,. \tag{4.18}$$

Furthermore, $\phi_i(x_i)\geqslant W(i)$ (see (4.16)) and $\phi_i(x_{i+1})\leqslant l(i)$, so that $\phi_i(x_i)-\phi_i(x_{i+1})\geqslant(1-\lambda)\delta(i)$. Clearly, $\phi_i(\cdot)$ is Lipschitz continuous with the constant L, and we obtain that $|x_i-x_{i+1}|=\mathrm{dist}\{x_i, Q_i\}\geqslant L^{-1}(1-\lambda)\delta(i)$.

Thus, (4.18) implies

$$\tau_{i+1}\leqslant\tau_i-L^{-2}(1-\lambda)^2\delta^2(i)\leqslant\tau_i-L^{-2}(1-\lambda)^2\delta^2(j_r)\,,\quad i\in I_r\,.$$

Since clearly $\tau_i\leqslant D^2$, (4.14) follows.

It remains to note that $\delta(j_{r+1})>(1-\lambda)^{-1}\delta(j_r)$, so that

$$M_l=|J_l|=\sum_r N_r\leqslant D^2L^2\delta^{-2}(j_1)(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1}\,,$$

which combined with (4.13) proves (4.9).

(6) (4.9) combined with (4.8) imply the required efficiency estimate. □

4.3. *Constrained Newton Method (CNM)*

4.3.1. *Preliminary remarks*

Denote

$$F_t(x)=\rho(f(x)-t,\, G(x))\,,$$

where, as above, $\rho(u,v)=\max\{(u)_+,(v)_+\}$, and let

$$\kappa(t)=\min_Q F_t(\cdot)\,.$$

The idea underlying this section is to solve the equation $\kappa(t)=0$ by a Newton method. Remarkably enough, this method is linearly convergent, even in the degenerate case (which occurs when (CMin) does not satisfy a Slater assumption); this is due to the convexity of the function κ. Naturally, $\kappa(t)$ cannot be computed exactly, but we construct an implementation which preserves the efficiency of the approach.

Assume we have called the oracle at the points $x_1, \ldots, x_i\in Q$. Then, besides the objects described in Section 4.1, we can define the following.

Upper distance function.

$$\kappa^*(i;t)\equiv\min\{\rho(u-t,v)\ \mid\ (u,v)\in C(i)\}\,.$$

Lower distance function.

$$\kappa_*(i;t)\equiv\min\{\rho(f_i(x)-t,\, G_i(x))\ \mid\ x\in Q\}\,.$$

Remark 4.3.1.1. With respect to $t\in\mathbb{R}$, the functions $\kappa(t)$, $\kappa^*(i;t)$, $\kappa_*(i;t)$ are nonincreasing convex Lipschitz continuous with Lipschitz constant 1, and

$$\kappa_*(1;t) \leqslant \kappa_*(2;t) \leqslant \cdots \leqslant \kappa_*(i;t) \leqslant \kappa(t)\ , \tag{4.19}$$

$$\kappa^*(1;t) \geqslant \kappa^*(2;t) \geqslant \cdots \geqslant \kappa^*(i;t) \geqslant \kappa(t)\ . \tag{4.20}$$

$\rho(\cdot,\cdot)$ is monotone and convex on $\mathbb{R}^2$; therefore for convex $p(\cdot), q(\cdot): Q \to \mathbb{R}$ the function $\rho(p(x)-t, q(x))$ is convex on $Q \times \mathbb{R}$, so that $\min_Q \rho(p(\cdot)-t, q(\cdot))$ is convex on $\mathbb{R}$ (and clearly Lipschitz continuous with constant 1). These remarks prove the convexity and the Lipschitz continuity of κ, κ^*, κ_*. The monotonicity of κ^* and κ_* in i immediately follow the monotonicity of ρ combined with (4.1), (4.2) and the (evident) inclusions $C(1) \subset C(2) \subset \cdots \subset C(i)$. (4.1), (4.2) and the monotonicity of ρ imply also the inequality $\kappa_*(i;t) \leqslant \kappa(t)$. Convexity of f and G implies immediately that for every $(u,v) \in C(t)$ there exists a convex combination x of the points $x_1, \ldots, x_i$ such that $(f(x), G(x)) \leqslant (u,v)$, and this observation combined with the monotonicity of ρ, leads to the inequality $\kappa^*(i;t) \geqslant \kappa(t)$.

Best point. Let $(u_i(t), v_i(t)) \in \operatorname{Argmin}\{\rho(u-t, v) \mid (u,v) \in C(i)\}$. Then there clearly exists a convex combination $\sum_{j=1}^{i} r_i(j;t)(f(x_j), G(x_j))$ of points belonging to $T(i)$, such that $\sum_{j=1}^{i} r_i(j;t)(f(x_j), G(x_j)) \leqslant (u_i(t), v_i(t))$. Set

$$x_i^*(t) = \sum_{j=1}^{i} r_i(j;t) x_j\ .$$

Remark 4.3.1.2. Let $t \leqslant f^*$. Then,

$$\epsilon(x_i^*(t)) \leqslant \kappa^*(i;t)\ . \tag{4.21}$$

Indeed, we have

$$(f(x_i^*(t)) - t, G(x_i^*(t))) \leqslant \sum_{j=1}^{i} r_i(j;t)(f(x_j) - t, G(x_j)) \leqslant (u_i(t) - t, v_i(t))\ ,$$

so that $\rho(f(x_i^*(t)) - t, G(x_i^*(t))) \leqslant \rho(u_i(t) - t, v_i(t)) = \kappa^*(i;t)$. It remains to note that $t \leqslant f^*$, so that

$$\begin{aligned}\epsilon(x_i^*(t)) &= \rho(f(x_i^*(t)) - f^*, G(x_i^*(t))) \\ &\leqslant \rho(f(x_i^*(t)) - t, G(x_i^*(t))) \leqslant \kappa^*(i;t)\ .\end{aligned}$$

4.3.2. Description of CNM

Parameters. $\lambda \in (0,1)$, $\mu \in (\frac{1}{2}, 1)$.

Initialization. x_1 is an arbitrary point of Q.

ith step.

(1) Call the oracle, x_i being the input.

(2) Compute $f_*(i)$, $\kappa^*(i;\cdot)$, $\kappa_*(i;\cdot)$.

(3) Set

$$t_i = \begin{cases} f_*(i), & i=1 \text{ or if } \kappa_*(i; t_{i-1}) > \mu\kappa^*(i; t_{i-1}), \\ t_{i-1}, & \text{otherwise}, \end{cases}$$

$$w(i) = \kappa_*(i; t_i), \quad W(i) = \kappa^*(i; t_i), \qquad l(i) = w(i) + \lambda(W(i) - w(i)),$$

$$x_{i+1} = \pi(x_i, \{x \mid x \in Q, \rho(f_i(x) - t_i, G_i(x)) \leqslant l(i)\}).$$

Theorem 4.3.3. *For the above method,*

$$\epsilon(x_i^*(t_i)) \leqslant \kappa^*(i; t_i),$$

and if $0 < \epsilon < V$, *then the following implication holds*:

$$i > c(\lambda, \mu)\left(\frac{V}{\epsilon}\right)^2 \ln\left(\frac{18V}{\epsilon}\right) \Rightarrow \epsilon(x_i^*(t_i)) \leqslant \epsilon,$$

where

$$c(\lambda, \mu) = 2\{\ln(2\mu)\}^{-1}(1-\mu)^{-2}(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1}.$$

(*Note that* $\min c(\cdot, \cdot) = c(0.292\,89\ldots, 0.652\,52\ldots) \leqslant 249$.)

5. A Method for (Var)

5.1. Notation

Assume we have called the oracle at the points $x_1, \ldots, x_i \in Q$. Then the following objects are defined.

Model.

$$\phi_i(x) = \max\{(F(x_j))^{\mathrm{T}}(x - x_j) \mid 1 \leqslant j \leqslant i\}.$$

Model's best value.

$$\phi_*(i) = \min_Q \phi_i(x).$$

Gap.

$$\delta(i) = -\phi_*(i).$$

Optimal multipliers. They are the quantities $r_i(j)$, $1 \leqslant j \leqslant i$, such that $r_i(j) \geqslant 0$, $\sum_{j=1}^{i} r_i(j) = 1$, and

$$\min\left\{\sum_{j=1}^{i} r_i(j)(F(x_j))^{\mathrm{T}}(x - x_j) \mid x \in Q\right\} = \min_Q \phi_i(\cdot) = \phi_*(i). \tag{5.1}$$

Best point.

$$x_i^* = \sum_{j=1}^{i} r_i(j) x_j .$$

Remark 5.1.1. (1) We evidently have

$$\phi_1(x) \leqslant \phi_2(x) \leqslant \cdots , \tag{5.2}$$

and $\phi_i(\cdot)$ are convex and Lipschitz continuous with constant L.

(2) We have

$$\delta(1) \geqslant \delta(2) \geqslant \cdots \geqslant 0 . \tag{5.3}$$

Indeed, let x^* be a solution to (Var), so that $(F(x))^{\mathrm{T}}(x-x^*) \geqslant 0$, $x \in Q$, whence $\phi_i(x^*) \leqslant 0$ and therefore $\phi_*(i) = -\delta(i) \leqslant 0$. Thus, $\delta(\cdot)$ is nonnegative. The monotonicity of $\delta(i)$ in i follows from (5.2).

(3) We have

$$\epsilon(x_i^*) \leqslant \delta(i) . \tag{5.4}$$

Indeed, let $x \in Q$. Then $(F(x))^{\mathrm{T}}(x-x_i^*) = (F(x))^{\mathrm{T}} \sum_{j=1}^{i} r_i(j)(x-x_j) \geqslant \sum_{j=1}^{i} r_i(j)(F(x_j))^{\mathrm{T}}(x-x_j) \geqslant \min \{\sum_{j=1}^{i} r_i(j)\ (F(x_j))^{\mathrm{T}}(y-x_j) \mid y \in Q\} = \phi_*(i)$ (we have taken into account the monotonicity of $F(\cdot)$ and (5.1)). Thus, $\epsilon(x_i^*) = \max\{(F(x))^{\mathrm{T}}(x_i^* - x) \mid x \in Q\} \leqslant -\phi_*(i) = \delta(i)$.

5.2. Truncated Level Method

The Truncated Level Method (TLM) looks as follows.

Parameters. $\lambda \in (0, 1)$.

Initialization. x_1 is an arbitrary point of Q.

ith step.

(1) Call the oracle, x_i being the input.

(2) Compute $\phi_*(i)$ and x_i^*.

(3) Set

$$l(i) = -(1-\lambda)\delta(i) ,$$

$$x_{i+1} = \pi(x_i, \{x \mid x \in Q, \phi_i(x) \leqslant l(i)\}) .$$

Theorem 5.2. *For the above method,*

$$\epsilon(x_i^*) \leqslant \delta(i) , \qquad i > c(\lambda)\left(\frac{V}{\epsilon}\right)^2 \Leftrightarrow \epsilon(x_i^*) \leqslant \epsilon ,$$

where

$$c(\lambda) = (1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1} .$$

(*Note that* $\min c(\cdot) = 4 = c(0.292\,89)$.)

Proof. In six steps.

(1) The efficiency estimate

$$\epsilon(x_i^*) \leqslant \delta(i) \tag{5.5}$$

was established in (5.4).

(2) Set $S_i = [\phi_*(i), 0]$. Then (see (5.2), (5.3)), $S_i \neq \emptyset$ and

$$S_1 \supseteq S_2 \supseteq \cdots, \quad |S_i| = \delta(i), \tag{5.6}$$

where $|S|$ denotes the length of a segment S.

(3) Let us fix $\epsilon > 0$ and assume that for certain N and all $i \leqslant N$ we have $\delta(i) > \epsilon$. Let us split the integer segment $I = 1, \ldots, N$ in groups $I_1, \ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \leqslant j_2$, for which $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \leqslant j_3$ satisfying $\delta(i) \leqslant (1-\lambda)^{-1}\delta(j_3)$, and so on.

(4) From (5.6) it immediately follows that $\phi_*(j_l) \leqslant l(i)$, $i \in I_l$. Let $u(l)$ minimize the function $\phi_{j_l}(\cdot)$ over Q; then for $i \in I_l$ one has $\phi_i(u(l)) \leqslant \phi_{j_l}(u(l)) \leqslant l(i)$. Thus, we have established that

$$\text{the (clearly convex) level sets } Q_i = \{x \in Q \mid \phi_i(x) \leqslant l(i)\} \tag{5.7}$$
associated with $i \in I_l$, have a common point (namely, $u(l)$).

(5) By virtue of the standard properties of the projection mapping, (5.7) implies

$$\tau_{i+1} \equiv |x_{i+1} - u(l)|^2 \leqslant \tau_i - \operatorname{dist}^2\{x_i, Q_i\}, \quad i \in I_l. \tag{5.8}$$

We also have $\phi_i(x_i) - l(i) \geqslant -l(i)$ (see (2.9)), so that $\phi_i(x_i) - l(i) \geqslant (1-\lambda)\delta(i)$, and $\phi_i(x_{i+1}) \leqslant l(i)$. From the Lipschitz property of ϕ_i, it follows that

$$\operatorname{dist}\{x_i, Q_i\} = |x_i - x_{i+1}| \geqslant L^{-1}|\phi_i(x_i) - \phi_i(x_{i+1})| \geqslant L^{-1}(1-\lambda)\delta(i).$$

Thus,

$$\tau_{i+1} \leqslant \tau_i - L^{-2}(1-\lambda)^2\delta^2(i) \leqslant \tau_i - L^{-2}(1-\lambda)^2\delta^2(j_l), \quad i \in I_l.$$

Because $0 \leqslant \tau_i \leqslant D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \leqslant D^2L^2(1-\lambda)^{-2}\delta^{-2}(j_l). \tag{5.9}$$

(6) From the definitions of N and of a group, we have

$$\delta(j_1) = \delta(N) > \epsilon, \quad \delta(j_{l+1}) > (1-\lambda)^{-1}\delta(j_l).$$

These relations combined with (5.9) imply

$$N=\sum_{l\geqslant 1} N_l \leqslant D^2L^2(1-\lambda)^{-2}\sum_{l\geqslant 1}\epsilon^{-2}(1-\lambda)^{2(l-1)}=\left(\frac{V}{\epsilon}\right)^2(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}.$$

□

6. Computational results

All the test-problems described below are available from the authors.

6.1. Unconstrained minimization

We have tested the simplest method of those described in Section 2, namely the Level Method LM.

Our implementation used two features.

· An input parameter $f_*(0)$ was given to the algorithm, serving as a lower bound on the optimal value f_*. The algorithm could then be run without compactness assumption on Q.

· The two auxiliary problems to compute $f_*(i)$ and x_{i+1} were solved with the help of the code QL0001 of Schittkowsky, itself based on the algorithm of Powell [17]. In some of the experiments we used simplex codes of Borisova and Sokolov in order to compute $f_*(i)$.

In all our experiments reported below, the parameter λ was set to 0.5 and the algorithm was run until the gap became smaller than 10^{-6} (in relative accuracy). We used double-precision Fortran on a Sun Workstation. The test problems were the following.

· BADGUY. This is a hand-made function, illustrating worst-case behaviours; see [16]. It is organized so that the gap after in calls to the oracle (n is the dimension of the problem) cannot be reduced by more than the factor 2^{3i+1}. We used $n=30$ variables.

· MAXQUAD and TR48 are described in [11].

· MAXANAL is a regularization of MAXQUAD, where the objective $\max\{f_k(x)\}$ is replaced by

$$\max\left\{\sum \lambda_k f_k(x)+\epsilon\sum \ln(\lambda_k)\;\middle|\;\sum \lambda_k=1\right\}.$$

Here, $\epsilon=10^{-3}$.

· NET22h is the dual of a network problem, described by Goffin. It has 22 variables and is badly scaled.

· URY100 is a convex variant of a problem described by Uryas'ev. It is actually the sum of a piecewise linear function and of a quadratic, with $n=100$ variables bound by the box $-0.2\leqslant x_i\leqslant 0.2$.

· TSP is the dual of a traveling salesman problem, following the Lagrangian relaxation of Held and Karp [4]. The function to be minimized is therefore the maximum of a very large number of affine functions; we used datasets with $n=6$, 14, 29, 100, 120 and 442 variables respectively, coming from VLSI design.

Table 1

BADGUY30 $f_*(0)=-5.e3$		MAXQUAD $f_*(0)=-1.e1$		MAXANAL $f_*(0)=-1.e1$		TR48 $f_*(0)=-7.e5$	
Itr	Function	Itr	Function	Itr	Function	Itr	Function
1	−1792.000	1	5337.066 00	1	5337.035 0000	1	−464 816.0
63	−2034.666	5	324.279 00	5	324.278 9000	34	−622 699.5
96	−2045.494	8	51.929 33	8	51.927 9300	67	−637 023.0
126	−2047.840	15	2.3870 00	15	2.33 7940	102	−638 392.0
187	−2047.986	50	−0.81511 09	34	−0.376 7172	126	−638 548.3
189	−2047.997	57	−0.83655 71	52	−0.788 7634	129	−638 562.9
220	−2047.999	64	−0.84096 04	58	−0.828 9160	131	−638 564.4
		78	−0.84136 39	64	−0.830 4996		
		89	−0.84140 29	87	−0.830 7792		
		98	−0.84140 77	97	−0.830 8066		
				104	−0.830 8082		
				111	−0.830 8085		

The results are reported in Tables 1–4. Observe the quality of the performances, as compared to the simplicity of implementation. From a practical viewpoint, however, a severe drawback is the need to store the complete bundle, which may result in too large LP and QP problems. In order to overcome this, reasonable strategies are wanted to truncate or aggregate the bundle – as is done in the classical variants of the method, see, for example, [22]. In our experiments with the largest of our test problems, TSP442, we used the simplest such strategy, described as follows. The first iteration is called critical. Then the ith iteration is called critical if the gap $\Delta(i)$ is $\leqslant \frac{1}{2}\Delta(i')$, where i' is the latest critical iteration preceding the ith one. At each critical iteration, we eliminate from the bundle all cutting planes which are not active at the minimizer of the current model. It is easily seen that the theoretical efficiency estimates are preserved; as far as practical behaviour is concerned, this strategy

Table 2

NET22H [a] $f_*(0)=-2.e2$		URYconv [b] $f_*(0)=0.$		URYconv [c] $f_*(0)=0.$	
Itr	Function	Itr	Function	Itr	Function
1	1121.345 00	1	10 814.641	1	10 814.641
9	4.517 85	4	1886.694	4	1886.694
35	−94.052 40	11	1255.478	63	1277.539
63	−102.502 31	22	1218.168	204	1218.104
122	−103.301 34	58	1210.598	500	1210.921
157	−103.406 71	207	1209.984		
248	−103.410 94	323	1209.903		
306	−103.411 90	349	1209.896		
321	−103.411 98				

[a] $10^{-6} \leqslant x$.
[b] $-0.2 \leqslant x \leqslant 0.2$; $\mathrm{Itr}_{\max}=350$.
[c] Box penalized; $\mathrm{Itr}_{\max}=500$.

Table 3

TSP6 $f_*(0) = -2.e2$		TSP14 $f_*(0) = 0.$		TSP29 $f_*(0) = 0.$	
Itr	Function	Itr	Function	Itr	Function
1	−403.0000	1	−2633.000	1	−1666.000
		13	−3259.119	35	−1996.989
5	−611.5000	22	−3317.877	52	−2010.877
12	−617.0000	24	−3321.485	61	−2013.199
		25	−3322.000	65	−2013.478
				68	−2013.497
				69	−2013.498

Table 4

TSP100 $f_*(0) = -3.e4$		TSP120 $f_*(0) = -8.e3$		TSP442 $f_*(0) = -1.e5$	
Itr	Function	Itr	Function	Itr	Function
1	−18 993.070	1	−5840.000	1	−46 862.30
51	−20 749.554	82	−6812.904	144	−50 142.60
108	−20 922.238	186	−6901.119	436	−50 472.39
141	−20 936.037	246	−6910.327	683	−50 502.67
149	−20 937.738	276	−6911.150	817	−50 505.27
152	−20 937.909	286	−6911.238	857	−50 505.57
153	−20 937.927	288	−6911.246	861	−50 505.60
		289	−6911.247		

Table 5

SAD08			SAD16			SAD32		
Itr	Gap	Objective	Itr	Gap	Objective	Itr	Gap	Objective
2	208 364.123	26 892.39	2	117 494.921	34 661.43	2	5448 581.43	−963 479.9
5	18 244.583	55 997.28	4	9828.878	31 997.96	4	458 570.34	−1334 416.5
8	1523.993	58 236.56	8	1066.392	31 259.49	10	30 586.87	−1204 660.1
12	158.842	58 579.83	15	109.481	31 159.29	17	5408.69	−1203 882.3
16	19.760	58 642.40	23	10.621	31 142.88	29	441.19	−1200 394.7
20	1.754	58 644.54	34	0.929	31 143.37	40	39.31	−1200 375.9
24	0.289	58 646.00	40	0.109	31 142.99	51	4.38	−1200 375.6
36	0.083	58 644.60	43	0.052	31 142.99	60	1.23	−1200 372.1

Table 6

RAND20 [a] (CLM)			RAND20 [a] (CNM)		
Itr	Objective	Unfeasible	Itr	Objective	Unfeasible
2	−3104.446 42	1245.234 537	2	44.593 12	326.967 23
12	562.191 05	0.753 830	12	501.954 17	1.372 5472
22	519.525 05	−0.085 973	22	516.095 80	0.009 2386
32	516.019 06	−0.001 225	32	515.971 97	0.000 3822
42	515.955 84	−0.000 012	42	515.954 90	0.000 0085
47	515.955 36	−0.000 007	53	515.955 07	0.000 0003

[a] $f_*(0)=0.0$.

Table 7

RAND40 [a] (CLM)			RAND40 [a] (CNM)		
Itr	Objective	Unfeasible	Itr	Objective	Unfeasible
2	−6183.8717	782.028 72	2	−6134.4774	780.303 9700
12	−5067.7366	0.022 10	12	−5037.3119	1.136 8766
23	−5090.0951	−0.022 30	23	−5093.9608	0.012 0858
34	−5094.5703	−0.000 78	34	−5094.6278	0.000 0513
43	−5094.6251	−0.000 06	47	−5094.6314	0.000 0075
			53	−5094.6309	0.000 0004

[a] $f_*(0)=-1.e4$.

Table 8

CHAIN20 [a] (CLM)			CHAIN20 [a] (CNM)		
Itr	Objective	Unfeasible	Itr	Objective	Unfeasible
2	−14.565 95	0.668 2579	2	−14.987 64	0.690 4065
14	−9.132 62	0.023 5769	14	−18.504 36	0.092 1989
24	−9.103 42	0.000 0675	24	−9.103 28	0.000 1196
31	−9.103 98	0.000 0006	30	−9.103 98	0.000 0007
CHAIN40 [b] (CLM)			**CHAIN40 [b] (CNM)**		
Itr	Objective	Unfeasible	Itr	Objective	Unfeasible
2	−34.998 39	0.798 7866	2	−35.257 71	0.805 425
12	−67.293 07	0.388 6304	12	−97.892 55	0.565 003
23	−36.420 46	0.035 5375	23	−84.292 31	0.012 037
32	−36.348 09	0.005 2049	32	−36.542 75	0.000 399
41	−36.441 83	0.000 0100	41	−36.440 25	0.000 000
48	−36.440 17	0.000 0007			

[a] $f_*(0)=-1.e3$.
[b] $f_*(0)=-1.e4$.

in our tests increased the number of iterations by at most 10–15%. The cardinality of the bundle never exceeded $2n$, and in TSP442 ($n=442$) this maximal cardinality was 446.

6.2. *Saddle points*

We tested the Level Method on a number of randomly generated saddle–point problems of the following type:

find a saddle point of the quadratic function

$$f(x, y) = \tfrac{1}{2}(\boldsymbol{P}x, x) - \tfrac{1}{2}(\boldsymbol{Q}y, y) + (\boldsymbol{R}x, y)$$

under the constraints

$$\boldsymbol{A}x \leqslant a, \qquad \|x\|_\infty \leqslant r, \qquad \boldsymbol{B}y \leqslant b, \qquad \|y\|_\infty \leqslant r,$$

where x and y are both n-dimensional, $\boldsymbol{P}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$ are matrices of corresponding sizes, and $\boldsymbol{P}$ and $\boldsymbol{Q}$ are positive semidefinite. The numbers of rows in the constraint matrices $\boldsymbol{A}$, $\boldsymbol{B}$ are equal to m.

We used a simple generator of test problems. The input to the generator includes the sizes n, m, as well as the parameter dc used to control the condition numbers of $\boldsymbol{P}$ and $\boldsymbol{Q}$ and the range of Lagrange multipliers at the saddle point (i.e., the coefficients in the representation of f'_x, f'_y at the solution as linear combinations of the gradients of the linear constraints active at the solution). Table 5 corresponds to problems

SAD08 ($r=10, n=8, m=12$, dc = 100), saddle value: 58 644.621 053 471.

SAD16 ($r=10, n=16, m=24$, dc = 100), saddle value: 31 142.996 423 246.

SAD32 ($r=10, n=32, m=48$, dc = 100), saddle value: − 1200 372.085 7410.

The control parameter λ of the method was set to 0.5; the process was terminated when the current gap $\Delta(i)$ was reduced to 10^{-6} (in relative accuracy).

Note that theoretically $f(x_i, y_i)$ should not converge to the value of the game (recall that all we claim is that $\epsilon(x_i^*, y_i^*)$ tends to 0 at the rate prescribed by the theoretical efficiency estimate). Nevertheless, our tests demonstrate that the values $f(x_i, y_i)$ also behave themselves well.

6.3. *Constrained minimization*

We ran both methods of Section 4, i.e., CLM and CNM, on two sets of test problems. Problems of the first set were randomly generated problems of the form

$$\text{minimize} \quad f(x) = (c, x)$$

$$\text{subject to} \quad f_i(x) = \|\boldsymbol{Q}_i x - q_i\|_2 - \rho_i \leqslant 0, \quad 1 \leqslant i \leqslant m,$$

$$\boldsymbol{A}_1 x = b_1, \qquad \boldsymbol{A}_2 x \leqslant b_2, \qquad \|x\|_\infty \leqslant r,$$

where x is n-dimensional, $\boldsymbol{Q}_i$ are $k \times n$ matrices, and $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ are $m_e \times n$ and $m_i \times n$ matrices, respectively.

The random problems of the above type were created by a simple generator; the input to the generator includes the sizes (n, m, k, m_e, m_i), as well as r (size of the box) and the additional control parameters m_{ai}, m_{an} (the numbers of linear inequality constraints and nonlinear constraints active at the solution) and c, dc, ag (responsible for the condition numbers of Q_i, for the range of Lagrange multipliers at the solution and for the range of values of the constraints nonactive at the solution, respectively).

Tables 6 and 7 represent the behaviour of CLM and CNM on two instances:

RAND20 ($n=20, m=8, m_e=2, m_i=4, m_{ai}=2, m_{an}=4, k=10, r=100, c=10$, dc $=10$, ag $=0.1$), optimal value: 515.955 062 799 04.

RAND40 ($n=40, m=16, m_e=4, m_i=8, m_{ai}=4, m_{an}=8, k=20, r=100, c=10$, dc $=10$, ag $=0.01$), optimal value: -5094.631 101 0407.

The test problems of the second type were as follows. Consider a chain made of n weightless segments in the vertical plane, and assume that the first segment starts at $(0, 0)$ and the last ends at $(L, 0)$ (the x-axis is horizontal, the y-axis is vertical). The length of each segment is $l=cL/n$. At the end of the ith segment (or, which is the same, at the beginning of the the $(i+1)$th segment) there is a unit mass, and we minimize the potential energy of the resulting system. In other words, we should minimize the function

$$\sum_{i=1}^{n-1} y_i$$

under the constraints

$$(x_i-x_{i+1})^2+(y_i-y_{i+1})^2 \leqslant l^2, \quad 0 \leqslant i \leqslant n-1,$$

where $x_0=y_0=y_n=0$, $x_n=L$.

The above problem is defined by the data n, L, c. The results in Table 8 correspond to the problems CHAIN20 ($n=20, c=2, L=1$) and CHAIN40 ($n=40, c=2, L=2$).

References

[1] A. Auslender, *Problèmes de Minimax via l'Analyse Convexe et les Inegalités Variationnelles: Théorie et Algorithmes*, Lecture Notes in Economics and Mathematical Systems (Springer, Berlin, 1972).

[2] E.W. Cheney and A.A. Goldstein, "Newton's methods for convex programming and Tchebytcheff approximation," *Numerische Mathematik* 1 (1959) 253–268.

[3] Yu.M. Ermoliev, "Methods for solving nonlinear extremal problems," *Cybernetics* 4 (1966) 1–17 (in Russian).

[4] M. Held and R.M. Karp, "The traveling-salesman problem and minimum spanning trees: Part II," *Mathematical Programming* 1 (1971) 6–25.

[5] P. Huard, "Resolution of mathematical programming problems with nonlinear constraints by the method of centers," in: J. Abadie, ed., *Nonlinear Programming* (North-Holland, Amsterdam, 1967) 206–219.

[6] J.E. Kelley, "The cutting plane method for solving convex programs," *Journal of the SIAM* 8 (1960) 703–712.

[7] K.C. Kiwiel, "An aggregate subgradient method for nonsmooth convex minimization," *Mathematical Programming* 27 (1983) 320–341.

[8] K.C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable minimization," *Mathematical Programming* 46 (1990) 105–122.

[9] V.Yu. Lebedev, ''On the convergence of the method of loaded functional as applied to a convex programming problem,'' *Journal of Numerical Mathematics and Mathematical Physics* 12 (1977) 765–768 (in Russian).
[10] C. Lemaréchal, ''An extension of Davidon methods to non-differentiable problems,'' *Mathematical Programming Study* 3 (1975) 95–109.
[11] C. Lemaréchal, ''Nonsmooth optimization and descent methods,'' Research Report 78-4, IIASA, Laxenburg, Austria, 1978.
[12] C. Lemaréchal, A. Nemirovskii and Yu. Nesterov, ''New variants of bundle methods,'' Research Report # 1508, Institut National de Recherche en Informatique et en Automatique, Le Chesnay, 1991.
[13] C. Lemaréchal, J.J. Strodiot and A. Bihain, ''On a bundle algorithm for nonsmooth optimization,'' in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., *Nonlinear Programming 4* (Academic Press, New York, 1981) 245–282.
[14] P. Marcotte and J.P. Dussault, ''A sequential linear programming algorithm for solving monotone variational inequalities,'' *SIAM Journal on Control and Optimization* 27 (1989) 1260–1278.
[15] R. Mifflin, ''A modification and an extension of Lemaréchal's algorithm for nonsmooth minimization,'' *Mathematical Programming Study* 17 (1982) 77–90.
[16] A.S. Nemirovskij and D.B. Yudin, *Problem Complexity and Method Efficiency in Optimization* (Wiley-Interscience, New York, 1983).
[17] M.J.D. Powell, ''ZQPCVX, a Fortran subroutine for convex programming,'' Report NA17, DAMTP, Cambridge, 1983.
[18] B.N. Pshenichny and Yu.M. Danilin, *Numerical Methods for Extremal Problems* (Mir, Moscow, 1978).
[19] S.M. Robinson, ''Extension of Newton's method to nonlinear functions with values in a cone,'' *Numerische Mathematik* 9 (1972) 341–347.
[20] R.T. Rockafellar, ''On the maximality of sums of nonlinear monotone operators,'' *Transactions of the American Mathematical Society* 149 (1970) 75–88.
[21] H. Schramm and J. Zowe, ''A version of the bundle idea for minimizing a non-smooth function: conceptual idea, convergence analysis, numerical results,'' *SIAM Journal on Optimization* 2 (1992) 121–152.
[22] P. Wolfe, ''A method of conjugate subgradients for minimizing nondifferentiable functions,'' *Mathematical Programming Study* 3 (1975) 145–173.