JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 71, No. 2, NOVEMBER 1991

New Variable-Metric Algorithms for Nondifferentiable Optimization Problems

S. P. URYAS'EV [1]

Communicated by O. L. Mangasarian

Abstract. This paper deals with new variable-metric algorithms for nonsmooth optimization problems, the so-called adaptive algorithms. The essence of these algorithms is that there are two simultaneously working gradient algorithms: the first is in the main space and the second is in the space of the matrices that modify the main variables. The convergence of these algorithms is proved for different cases. The results of numerical experiments are also given.

Key Words. Nonsmooth optimization, nondifferentiable programming, variable-metric algorithms.

1. Introduction

Many efficient algorithms have been developed for nonsmooth optimization problems (see, for example, Refs. 1–8). Most of these algorithms require the solution of some complicated linear or quadratic programming subproblem at each algorithm iteration. For high-dimensional problems it is, as a rule, impossible to solve high-dimensional subproblems at each iteration of the main method. The first subgradient algorithm proposed by Shor (Ref. 9) requires only subgradient calculations and projection operations at each iteration. It appears that this algorithm has a low convergence rate for ill-conditioned problems. Variable-metric algorithms lie somewhere in between: they require not too many calculations at each iteration, but nevertheless have a good practical rate of convergence for ill-conditioned functions. Variable-metric algorithms are widely used for smooth optimization problems (see, for example, Ref. 10). As a rule, these algorithms cannot be generalized to nonsmooth optimization problems. The difficulties are connected with the fact that, even if the first and second derivatives exist at

[1] Research Scholar, International Institute for Applied Systems Analysis, Laxenburg, Austria.

some point, they do not give the full local description of the function. Because the function is nonsmooth, a point of nonsmoothness can be arbitrarily close to the point where derivatives exist.

These difficulties have led to the appearance of new ideas in the construction of variable-metric algorithms. In the works of Shor and his coworkers (see, for example, Ref. 11), so-called space-dilation algorithms were developed. Such an approach offers the opportunity to construct practical and effective algorithms, but the most effective algorithm from this family, the r-algorithm, is not sufficiently understood from the theoretical point of view. References to this and related topics can be found in Ref. 12.

This author proposes an alternative adaptive approach that is applicable to optimization and game theoretic problems. The first variable-metric algorithms in the framework of this approach are discussed in Ref. 13 for stochastic quasigradient algorithms. Different optimization algorithms for stochastic optimization problems can be found in Ref. 14 and others. The results of this paper are based upon Ref. 15.

2. Essence of the Approach

Let us consider a convex optimization problem,

$$f(x) \to \min_{x \in R^n}, \tag{1}$$

where the function $f(x)$ is convex on the Euclidean space R^n. We use the following recurrent algorithm to solve this problem:

$$x^{s+1} = x^s - \rho_s H^s g^s, \qquad s = 0, 1, \ldots; \tag{2}$$

here, s is the iteration number; $\rho_s > 0$ is a stepsize (scalar value); H^s is an $n \times n$ matrix; and g^s is a subgradient from the subdifferential $\partial f(x)$ of the function $f(x)$ at the point x^s, i.e., $g^s \in \partial f(x^s)$. We recall that the subdifferential of the function $f(x)$ at the point $y \in R^n$ is given by the formula (see, for example, Ref. 16)

$$\partial f(y) = \{g \in R^n : f(x) - f(y) \geq \langle g, x - y \rangle, \forall x \in R^n\}.$$

At the s^{th} *iteration, the natural criterion defining the best choice of the matrix* H^s is via the function

$$\varphi_s(H) = f(x^s - \rho_s H g^s).$$

The best matrix is a solution of the problem

$$\varphi_s(H) \to \min_{H \in R^{n \times n}}. \tag{3}$$

It is easy to see that problem (3) is a reformulation of the source problem (1), since if H^* is a solution of (3), then the point $x^s - \rho_s H^* g^s$ is a solution of (1). Moreover, problem (3) is more complex than (1) because the dimension of problem (3) is n times higher than that of (1). However, at the sth iteration of algorithm (2), we do not need the optimal matrix; it is enough to correct (update) the matrix H^s. If we already have some matrix H_0^s, then the direction of adaptation can be defined by differentiating, in the general sense, the function $\varphi_s(H)$ at the point H_0^s. If the function $f(x)$ is a convex function, then the function $\varphi_s(H)$ is also convex. We can use the following formula (Ref. 17) for the differentiation of the complex function φ_s:

$$\partial \varphi_s(H_0^s) = -\rho_s\{g g^{s^T} : g \in \partial f(x^s - \rho_s H_0^s g^s)\};$$

here and below, the superscript T means transposition. If $g_0^s \in \partial f(x^s - \rho_s H_0^s g^s)$, then $-\rho_s g_0^s g^{s^T} \in \partial \varphi(H_0^s)$. With respect to the matrix H, in the direction $g_0^s g^{s^T}$ one can do a step of the generalized gradient method,

$$H_1^s = H_0^s + \lambda_0^s g_0^s g^{s^T}, \qquad \lambda_0^s > 0.$$

It is possible either to take $H^s = H_1^s$ or to continue the iterations of the generalized gradient algorithm with respect to H,

$$H_{i+1}^s = H_i^s + \lambda_i^s g_i^s g^{s^T}, \qquad \lambda_i^s > 0, \tag{4}$$

where

$$g_i^s \in \partial f(x^s - \rho_s H_i^s g^s) \quad \text{and} \quad -\rho_s g_i^s g^{s^T} \in \partial \varphi_s(H_i^s).$$

For some $i(s) \geq 1$, assume that $H^s = H_{i(s)}^s$. At the next iteration $H_0^{s+1} = H^s$. The number $i(s)$ can be taken independently of s, for example $i(s) = 1$ for all s. Generally speaking, algorithm (2) is not monotone with respect to the objective function $f(x)$. However, one can choose $i(s)$ such that

$$f(x^s - \rho_s H_{i(s)}^s g^s) < f(x^s - \rho_s H_0^s g^s)$$

and, with each iteration, the objective function decreases.

Note that matrix updating requires additional calculations of the objective function subgradients. This can be avoided by taking

$$g^{s+1} = g_0^s, \qquad i(s) = 1,$$

and using the matrix H_1^s at iteration $s+1$. Therefore, we propose the following formula for matrix updating:

$$H^{s+1} = H^s + \lambda_s g^{s+1} g^{s^T}, \qquad \lambda^s > 0. \tag{5}$$

In formula (5), additional subgradient calculations are not required. We supposed in formulas (4), (5) that the objective function is convex, but the same approach can be used also for nonconvex functions.

3. Convergence for Smooth Functions

First, let us investigate the convergence of algorithms (2) and (5) for the case of a differentiable function $f(x)$. Denote $g^s=\nabla f(x^s)$. For algorithm (2), as the direction of motion we choose the normalized subgradient,

$$\xi^s=\begin{cases} g^s/\|g^s\|, & \|g^s\|\neq 0,\\ 0, & \|g^s\|=0.\end{cases}$$

The algorithm can be rewritten as follows:

$$x^{s+1}=x^s-\rho_s H^s\xi^s, \tag{6a}$$

$$H^{s+1}=H^s+\rho_s g^{s+1}\xi^{sT}, \tag{6b}$$

$$H^0=I, \tag{6c}$$

where I is the unit matrix. Thus, for the parameter λ_s in formula (5), we choose the value $\rho_s/\|g^s\|$. Denote

$$\Theta^s=\sum_{l=0}^{s}\rho_l g^{l+1}\xi^{lT}, \qquad D(K)=\max_{y,x\in K}\|x-y\|,$$

$$f^*=\min_{x\in R^n} f(x), \qquad \bar f_s=\min_{0\le l\le s} f(x^l);$$

i.e., $D(K)$ is the diameter of the set K, and $\bar f_s$ is a record of the lowest value of the function f during the previous s iterations. We also denote by $\mathrm{Tr}(Q)$ the trace of the matrix Q.

Let us now formulate a theorem about the convergence of algorithm (6)

Theorem 3.1. Let $f: R^n \to R$ be a convex smooth Lipschitz function

$$f(x)-f(y)\le L_1\|x-y\|, \qquad x,y\in R^n, \tag{7}$$

with Lipschitz gradient

$$\|\nabla f(x)-\nabla f(y)\|\le L_2\|x-y\|, \qquad x,y\in R^n, \tag{8}$$

and let there exist a compact set K such that

$$\|\nabla f(x)\|\ge\delta_K>0, \qquad x\notin K. \tag{9}$$

If the stepsizes ρ_s, $s=0,1,\dots$, in algorithm (6) are nonnegative and if, for all s greater than some $\bar S$, we have

$$\delta_K\ge\left\{\sum_{l=0}^{s}\rho_l^2\left[L_1^2 2^{-1}+L_2\left(n+2(f(x^0)-f^*)+L_1^2\sum_{l=0}^{l-1}\rho_l^2\right)^{1/2}\right]+f(x^0)-f^*\right\}\left(\sum_{l=0}^{s}\rho_l\right)^{-1}, \tag{10}$$

then the following results hold:

(i) for $s>\bar S$,

$$\bar f_s-f^*\le\left\{\sum_{l=0}^{s}\rho_l^2\left[L_1^2 2^{-1}+L_2\left(n+2(f(x^0)-f^*)+L_1^2\sum_{l=0}^{l-1}\rho_l^2\right)^{1/2}\right]+f(x^0)-f^*-2^{-1}\mathrm{Tr}(\Theta^s\Theta^{sT})\right\}D(K)\left(\sum_{l=0}^{s}\rho_l\right)^{-1};$$

(ii) for $\rho_l=(s+1)^{-1/2}$, $l=0,\dots,s$ and $s>\bar S$, where $c=\text{const}$,

$$\begin{aligned}\bar f_s-f^*&\le[L_1^2 2^{-1}+L_2(n+2(f(x^0)-f^*)+L_1^2)^{1/2}\\&\quad+f(x^0)-f^*-2^{-1}\mathrm{Tr}(\Theta^s\Theta^{sT})]D(K)(s+1)^{-1/2}\\&\le c(s+1)^{-1/2};\end{aligned}$$

(iii) for $\sum_{l=0}^{\infty}\rho_s^2=\infty$,

$$\varliminf_{s\to\infty}(f(x^s)-f^*)\left(\sum_{l=0}^{s-1}\rho_l^2\right)^{-1/2}\rho_s^{-1}\le 2L_1L_2D(K);$$

and if $\rho_s=s^{-1/2}$, then

$$\varliminf_{s\to\infty}(\bar f_s-f^*)(\log(s))^{-1/2}s^{1/2}\le\text{const}.$$

Lemma 3.1. For the Euclidean norm of the matrix H^s, the following inequality holds:

$$\|H^{s+1}\|\le\left(n+2(f(x^0)-f^*)+L_1^2\sum_{l=0}^{s}\rho_l^2\right)^{1/2}. \tag{11}$$

Proof. The relation (6) implies that

$$\begin{aligned} H^{s+1}H^{s+1^T} &= (H^s+\rho_s g^{s+1}\xi^{s^T})(H^{s^T}+\rho_s\xi^s g^{s+1^T}) \\ &= H^sH^{s^T}+\rho_s(g^{s+1}\xi^{s^T}H^{s^T}+H^s\xi^s g^{s+1^T}) \\ &\quad +\rho_s^2\|\xi^s\|^2 g^{s+1}g^{s+1^T}. \end{aligned}$$

Since the function $f(x)$ satisfies the condition (7), then

$$\|g^s\|\leq L_1, \qquad s=0,1,\ldots.$$

Denote

$$\Delta x^{s+1}=x^{s+1}-x^s.$$

Taking into account the convexity of the function $f(x)$, we can evaluate the trace of $H^sH^{s^T}$,

$$\begin{aligned} \mathrm{Tr}(H^{s+1}H^{s+1^T}) &= \mathrm{Tr}(H^sH^{s^T})+2\rho_s\langle g^{s+1}, H^s\xi^s\rangle+\rho_s^2\|\xi^s\|^2\|g^{s+1}\|^2 \\ &\leq \mathrm{Tr}(H^sH^{s^T})-2\langle g^{s+1},\Delta x^{s+1}\rangle+\rho_s^2L_1^2 \\ &\leq \mathrm{Tr}(H^sH^{s^T})+2(f(x^s)-f(x^{s+1}))+\rho_s^2L_1^2 \\ &\leq \mathrm{Tr}(H^0H^{0^T})+2(f(x^0)-f(x^{s+1}))+L_1^2\sum_{l=0}^{s}\rho_l^2 \\ &\leq n+2(f(x^0)-f^*)+L_1^2\sum_{l=0}^{s}\rho_l^2. \end{aligned} \tag{12}$$

Since

$$\|H^{s+1}\|^2\leq \mathrm{Tr}(H^{s+1}H^{s+1^T}),$$

then

$$\|H^{s+1}\|\leq\left(n+2(f(x^0)-f^*)+L_1^2\sum_{l=0}^{s}\rho_l^2\right)^{1/2}.$$

The lemma is proved. □

Proof of Theorem 3.1. Inequality (12) implies that

$$\begin{aligned} \mathrm{Tr}(H^{s+1}H^{s+1^T}) &= \mathrm{Tr}\left[\left(\sum_{l=0}^{s}\rho_l g^{l+1}\xi^{l^T}+I\right)\left(\sum_{l=0}^{s}\rho_l\xi^l g^{l+1^T}+I\right)\right] \\ &= \mathrm{Tr}(\Theta^s\Theta^{s^T})+2\sum_{l=0}^{s}\rho_l\langle g^{l+1},\xi^l\rangle+n \\ &\leq n+2(f(x^0)-f^*)+L_1^2\sum_{l=0}^{s}\rho_l^2. \end{aligned}$$

Consequently,

$$\sum_{l=0}^{s}\rho_l\langle g^{l+1},\xi^l\rangle\leq 2^{-1}L_1^2\sum_{l=0}^{s}\rho_l^2-2^{-1}\mathrm{Tr}(\Theta^s\Theta^{s^T})+f(x^0)-f^*. \tag{13}$$

Since the gradient of the function $f(x)$ satisfies the Lipschitz condition, then applying Lemma 3.1, we see that

$$\begin{aligned} \|g^{l+1}-g^l\| &\leq L_2\|x^{l+1}-x^l\|=L_2\|\rho_lH^l\xi^l\|\leq\rho_lL_2\|H^l\| \\ &\leq\rho_lL_2\left(n+2(f(x^0)-f^*)+L_1^2\sum_{i=0}^{l-1}\rho_i^2\right)^{1/2}. \end{aligned} \tag{14}$$

Using (13) and (14), it is easy to establish that

$$\begin{aligned} \sum_{l=0}^{s}\rho_l\|g^l\| &= \sum_{l=0}^{s}\rho_l\langle g^l,\xi^l\rangle \\ &= \sum_{l=0}^{s}\rho_l\langle g^l-g^{l+1},\xi^l\rangle+\sum_{l=0}^{s}\rho_l\langle g^{l+1},\xi^l\rangle \\ &\leq \sum_{l=0}^{s}\rho_l\|g^l-g^{l+1}\|+\sum_{l=0}^{s}\rho_l\langle g^{l+1},\xi^l\rangle \\ &\leq L_2\sum_{l=0}^{s}\rho_l^2\left(n+2(f(x^0)-f^*)+L_1^2\sum_{i=0}^{l-1}\rho_i^2\right)^{1/2} \\ &\quad +2^{-1}L_1^2\sum_{l=0}^{s}\rho_l^2-2^{-1}\mathrm{Tr}(\Theta^s\Theta^{s^T})+f(x^0)-f^*. \end{aligned} \tag{15}$$

This function is ill-conditioned. Consequently,

$$\begin{aligned} \min_{0\leq l\leq s}\|g^l\| &\leq\left(\sum_{l=0}^{s}\rho_l\|g^l\|\right)\left(\sum_{l=0}^{s}\rho_l\right)^{-1} \\ &\leq\left\{\sum_{l=0}^{s}\rho_l^2\left[L_1^2 2^{-1}+L_2\left(n+2(f(x^0)-f^*)+L_1^2\sum_{i=0}^{l-1}\rho_i^2\right)^{1/2}\right]\right. \\ &\quad \left.-2^{-1}\mathrm{Tr}(\Theta^s\Theta^{s^T})+f(x^0)-f^*\right\}\left(\sum_{l=0}^{s}\rho_l\right)^{-1}. \end{aligned} \tag{16}$$

Combining Ineq. (10) with the last inequality, we get

$$\min_{1\leq l\leq s}\|g^l\|\leq\delta_K, \qquad s>\bar{S}.$$

Thus, for $l(s)$ such that

$$g^{l(s)} = \min_{1 \le l \le s} \|g^l\|,$$

the inclusion $x^{l(s)} \in K$ holds due to the definition of δ_K.

Let x^* be a minimum point, i.e., $f(x^*) = f^*$. Using the convexity of the function $f(x)$, we see that

$$\bar{f}_s - f^* \le f(x^{l(s)}) - f(x^*) \le \langle g^{l(s)}, x^{l(s)} - x^* \rangle$$
$$\le \|g^{l(s)}\| \, \|x^{l(s)} - x^*\| \le \|g^{l(s)}\| D(K) \le D(K) \min_{1 \le l \le s} \|g^l\|.$$

Applying this last relation and (16), we get Statement (i) of the theorem.

Statement (ii) of the theorem can be obtained by substituting the values $\rho_l = (s+1)^{-1/2}$, $l = 0, \ldots, s$, into Statement (i).

Let us now prove Statement (iii) of the theorem. It is sufficient to show that

$$\overline{\lim_{s \to \infty}} \|g^s\| \left(\sum_{l=0}^{s-1} \rho_l^2 \right)^{-1/2} \rho_s^{-1} \le 2 L_1 L_2, \tag{17}$$

because

$$\|g^s\| D(K) \ge f(x^s) - f^*.$$

If Ineq. (17) does not hold, then there exists a number $\hat{S}$ such that, for $s > \hat{S}$,

$$\|g^s\| \ge 2 L_1 L_2 \rho_s \left(\sum_{l=0}^{s-1} \rho_l^2 \right)^{1/2}.$$

Substituting the last inequality into the left side of Ineq. (15), we get a contradiction for $s \to \infty$, because the left side of the inequality tends to infinity faster than the right side. □

If the number of iterations of the algorithm is chosen before the start, then Statement (ii) of the theorem implies that the convergence rate of algorithm (6) is not worse than the generalized gradient algorithm with matrix $H^s = I$, $s = 0, 1, \ldots$. For $H^s = I$, $s = 0, 1, \ldots$, the following estimate (see, for example, Ref. 18) is known:

$$\bar{f}_s - f^* \le (s+1)^{-1/2} d(x^0, X^*),$$

where

$$d(x, X^*) = \min_{y \in X^*} \|x - y\|, \qquad X^* = \{x^* : f(x^*) = f^*\}. \tag{18}$$

Note that, in the estimate in Statement (ii) of the theorem, there exists an additional term $\mathrm{Tr}(\Theta^s \Theta^{s^T})$ that increases the convergence rate.

4. Convergence for Nonsmooth Functions

Let us consider algorithm (2). We assume that, at the sth iteration of the main algorithm, formula (4) is used $i(s)$ times for the updating matrix H_0^s. At iteration $s+1$, we take $H_0^{s+1} = H_{i(s)}^s$.

At iteration zero, $H_0^0 = I$, where I is the unit matrix. Let us fix some $\epsilon > 0$. We choose $i(s)$ to be the minimal number such that

$$f(x^s - \rho_s H_{i(s)}^s g^s) \le f(x^s) - \epsilon.$$

It is convenient to normalize the test vector g_i^s, therefore, denote

$$\xi_i^s = \begin{cases} 0, & \text{if } g_i^s = 0, \\ g_i^s \|g_i^s\|^{-1}, & \text{otherwise}. \end{cases} \tag{19}$$

For each $s = 0, 1, \ldots$, let the sequence $\{\lambda_{si}\}$, $i = 0, 1, \ldots$, of positive values be given. We express the algorithm in more detail.

Algorithm 4.1.

Step 1. Initialization. Set

$$s = 0,\ x^0 = x_{\text{init}},\ g^0 \in \partial f(x^0), \qquad i = -1,\ H_0^{-1} = I.$$

Step 2. (i) Set $H_0^s = H_{i+1}^{s-1}$, $i = 0$.
(ii) Set $x_i^s = x^s - \rho_s H_i^s g^s$.
(iii) Compute $g_i^s \in \partial f(x_i^s)$. If $g_i^s = 0$, then stop; otherwise, set $\xi_i^s = g_i^s \|g_i^s\|^{-1}$.
(iv) Set $H_{i+1}^s = H_i^s + \lambda_{si} \xi_i^s g^{s^T}$.
(v) If $f(x_i^s) \le f(x^s) - \epsilon$, then set $i(s) = i$, and go to Step 3.
(vi) Set $i = i+1$, return to (ii).

Step 3. Set $x^{s+1} = x_i^s$, $g^{s+1} = g_i^s$.

Step 4. Set $s = s+1$, and return to Step 2.

We now formulate a theorem about the convergence of Algorithm 4.1.

Theorem 4.1. Let $f: R^n \to R$ be a convex function; let the set of minimum points X^* of the function $f(x)$ be nonempty and bounded; let $\{\rho_s\}$ be a sequence of positive numbers; and let $\{\lambda_{si}\}$, $s = 0, 1, \ldots$ and $i = 0, 1, \ldots$, be a given sequence of positive numbers satisfying

$$\sum_{i=0}^{\infty} \lambda_{si} = \infty, \quad \sum_{i=0}^{\infty} \lambda_{si}^2 < \infty, \qquad s = 0, 1, \ldots. \tag{20}$$

Then, there exists $\bar{s}$ such that $f(x^{\bar{s}}) \le \epsilon + f^*$ and $d(x_i^{\bar{s}}, X^*) \to 0$, $f(x_i^{\bar{s}}) \to f$; for $i \to \infty$, where f^* equals the minimum value of f on R^n.

Proof. Let us first prove that, if

$$f(x^s)-f^*-\epsilon=\delta>0,$$

then there exists $i(s)$ such that

$$f(x^s_{i(s)})\leq f(x^s)-\epsilon. \tag{21}$$

Let $x^*\in X^*$. Using the formulas of Step 2(ii) and (iv), we get

$$\begin{aligned}\|x^*-x^s_{i+1}\|^2&=\|x^*-x^s+\rho_s H^s_{i+1}g^s\|^2\\&=\|x^*-x^s+\rho_s(H^s_i+\lambda_{si}\xi^s_i g^{s^T})g^s\|^2\\&=\|x^*-x^s+\rho_s H^s_i g^s+\rho_s\lambda_{si}\|g^s\|^2\xi^s_i\|^2\\&=\|x^*-x^s_i+\rho_s\lambda_{si}\|g^s\|^2\xi^s_i\|^2\\&\leq\|x^*-x^s_i\|^2+2\rho_s\lambda_{si}\|g^s\|^2\langle\xi^s_i,x^*-x^s_i\rangle+(\rho_s\lambda_{si}\|g^s\|^2)^2\\&=\|x^*-x^s_0\|^2+2\rho_s\|g^s\|^2\sum_{l=0}^{i}\lambda_{si}\langle\xi^s_i,x^*-x^s_i\rangle\\&\quad+\rho_s^2\|g^s\|^4\sum_{l=0}^{i}\lambda^2_{sl}.\end{aligned}\tag{22}$$

We prove by contradiction that there exists $i(s)$ satisfying Ineq. (21). For all $i>0$, let

$$f(x^s_i)>f^*+\delta.$$

Since $\langle\xi^s_i,x^*-x^s_i\rangle\leq 0$, it follows from (22) and (20) that

$$\|x^*-x^s_{i+1}\|^2\leq\|x^*-x^s_0\|^2+\rho_s^2\|g^s\|^4\sum_{l=0}^{\infty}\lambda^2_{sl}\leq C=\text{const}.$$

The function $f:R^n\rightarrow R$ is convex; consequently, on the compact set $\{x\in R^n\colon\|x^*-x\|\leq C\}$, it satisfies the Lipschitz condition with a constant L. Therefore,

$$\|g^s_i\|\leq L,\qquad i\geq 0.$$

Using the convexity of the function $f(x)$, we obtain

$$\begin{aligned}2\rho_s\|g^s\|^2\langle\xi^s_i,x^*-x^s_i\rangle&\leq 2\rho_s\|g^s\|^2\|g^s_i\|^{-1}(f(x^*)-f(x^s_i))\\&\leq-2\rho_s\|g^s\|^2L^{-1}\delta=-\sigma<0.\end{aligned}$$

Substituting this inequality into relation (22), we see that

$$\begin{aligned}\|x^*-x^s_{i+1}\|^2&\leq\|x^*-x^s_0\|^2-\sigma\sum_{l=0}^{i}\lambda_{sl}+\rho_s^2\|g^s\|^4\sum_{l=0}^{i}\lambda^2_{sl}=\|x^*-x^s_0\|^2\\&\quad+\left(\sum_{l=0}^{i}\lambda_{sl}\right)\left(-\sigma+\rho_s^2\|g^s\|^4\left(\sum_{l=0}^{i}\lambda_{sl}\right)^{-1}\sum_{l=0}^{i}\lambda^2_{sl}\right).\end{aligned}\tag{23}$$

Applying (20), we have

$$\left(\sum_{l=0}^{i}\lambda_{sl}\right)^{-1}\sum_{s=0}^{i}\lambda^2_{sl}\rightarrow 0,\qquad i\rightarrow\infty.$$

Using (20), we obtain

$$\left(\sum_{l=0}^{i}\lambda_{sl}\right)\left(-\sigma+\rho_s^2\|g^s\|^4\left(\sum_{l=0}^{i}\lambda_{sl}\right)^{-1}\sum_{s=0}^{i}\lambda^2_{sl}\right)\rightarrow-\infty,\qquad i\rightarrow\infty,$$

and this contradicts (23).

It follows from (21) that

$$f(x^{s+1})\leq f(x^s)-\epsilon,\qquad \text{if } f(x^s)-f^*-\epsilon>0.$$

Consequently, there exists $\bar{s}$ such that

$$f(x^{\bar{s}})\leq\epsilon+f^*.$$

Since ϵ is arbitrary, then (21) implies that there exists a subsequence i_k for which

$$f(x^{\bar{s}}_{i_k})\rightarrow f^*,\qquad k\rightarrow\infty. \tag{24}$$

Let us prove that the convergence of this subsequence leads to the convergence of the sequence. Take some $\beta>0$; if

$$f(x^{\bar{s}}_i)-f^*\geq\beta,$$

then

$$2\rho_{\bar{s}}\|g^{\bar{s}}\|^2\langle\xi^{\bar{s}}_i,x^*-x^{\bar{s}}_i\rangle\leq-q<0.$$

It follows from (20) that there exists $\bar{I}$ such that, for all $i>\bar{I}$, the inequality

$$\lambda^{\bar{s}}_i\leq q\rho_{\bar{s}}^{-2}\|g^{\bar{s}}\|^{-4}$$

holds. Using Ineq. (22), we have

$$\begin{aligned}\|x^*-x^{\bar{s}}_{i+1}\|^2&=\|x^*-x^{\bar{s}}_i\|^2+\lambda_{\bar{s}i}(2\rho_{\bar{s}}\|g^{\bar{s}}\|^2\langle\xi^{\bar{s}}_i,x^*-x^{\bar{s}}_i\rangle+\lambda_{\bar{s}i}\rho^2_{\bar{s}}\|g^{\bar{s}}\|^4)\\&\leq\|x^*-x^{\bar{s}}_i\|^2+\lambda_{\bar{s}i}(-q+q)\\&\leq\|x^*-x^{\bar{s}}_i\|^2,\end{aligned}\tag{25}$$

for i such that $i>\hat{I}$ and $f(x_i^{\bar{s}})-f^*\geq\beta$. Denote

$$U(\mu)=\{x: d(x, X^*)\leq\mu\}, \qquad Q_\beta=\{x: f(x)<f^*+\beta\};$$

see (18). Let $\mu(\beta)$ be a minimal number such that $Q_\beta\subset U(\mu(\beta))$. Since the function $f: R^n\to R$ is convex and the set X^* is compact, then $\mu(\beta)\to 0$ for $\beta\to 0$. Applying (25), we see that, if $x_i^{\bar{s}}\notin U(\mu(\beta))$, then

$$\|x^*-x_{i+1}^{\bar{s}}\|<\|x^*-x_i^{\bar{s}}\|,$$

for $i>\hat{I}$. Using Step 2(ii) and (iv), we obtain

$$\|x_{i+1}^{\bar{s}}-x_i^{\bar{s}}\|=\|x^{\bar{s}}-\rho_{\bar{s}}H_{i+1}^{\bar{s}}g^{\bar{s}}-x_i^{\bar{s}}\|$$
$$=\|x_i^{\bar{s}}-\rho_{\bar{s}}\lambda_{\bar{s}i}\xi_i^{\bar{s}}\|g^{\bar{s}}\|^2-x_i^{\bar{s}}\|\leq\lambda_{\bar{s}i}\rho_{\bar{s}}\|g^{\bar{s}}\|^2.$$

Relation (24) implies that, beginning with some $k>\bar{K}$, the inclusion $x_{i_k}^{\bar{s}}\in U_{\mu(\beta)}$ holds. Taking into account the two previous inequalities, we get

$$x_i^{\bar{s}}\in U\left(\mu(\beta)+\rho_{\bar{s}}\|g^{\bar{s}}\|^2\max_{i_k\leq l\leq i_{k+1}}\lambda_{\bar{s}l}\right) \tag{26}$$

for $i>\max\{\hat{I}, i_{\bar{K}}\}$, $i_k\leq i<i_{k+1}$. It follows from assumption (20) of the theorem that

$$\max_{i_k\leq l\leq i_{k+1}}\lambda_{\bar{s}l}\to 0, \qquad k\to\infty;$$

therefore, for sufficiently large k, the inequality

$$\rho_{\bar{s}}\|g^{\bar{s}}\|^2\max_{i_k\leq l\leq i_{k+1}}\lambda_{\bar{s}i}\leq\mu(\beta)$$

holds. Substituting this estimate into (26), we have

$$x_i^{\bar{s}}\in U(2\mu(\beta)),$$

for sufficiently large i. Since β can be arbitrarily small and $\mu(\beta)\to 0$ for $\beta\to 0$, then $d(x_i^{\bar{s}}, X^*)\to 0$ for $i\to\infty$. The function $f: R^n\to R$ is convex; consequently, it is continuous on R^n. For this reason, the convergence $d(x_i^{\bar{s}}, X^*)\to 0$ implies $f(x_i^{\bar{s}})\to f^*$. The theorem is proved. □

Algorithm 4.1 has a substantial deficiency connected with the fact that the stepsize ρ_s does not change in the internal iterations $i=1,\ldots,i(s)$. Let us consider an algorithm with a steepest-descent control of ρ_s at each iteration $i=1,\ldots,i(s)$. Such a modification improves considerably the algorithm.

Let $v>0$ be a given number, and let $\{\lambda_j\}_0^\infty$ be a sequence of positive numbers.

Algorithm 4.2.

Step 1. Initialization. Set

$$s=0,\ H_0^{-1}=I,\ j=-1,\ i=-1,\ x^0=x_{\text{init}},\ g^0\in\partial f(x^0).$$

Step 2. (i) Set $H_0^s=H_{i+1}^{s-1}$, $i=0$.
(ii) Set $\rho_{si}=\arg\min_{\rho>0}f(x^s-\rho H_i^s g^s)$.
(iii) Set $j=j+1$, $j(s,i)=j$.
(iv) Set $x_i^s=x^s-\rho_{si}H_i^s g^s$.
(v) Compute $g_i^s\in\partial f(x_i^s)$ such that $\langle g_i^s, H_i^s g^s\rangle\leq 0$.
(vi) Set $H_{i+1}^s=H_i^s+\lambda_j\xi_i^s g^{s^T}$.
(vii) If $\|x_i^s-x^s\|\geq v$, then set $i(s)=i$, and go to Step 3.
(viii) $i=i+1$, return to (ii).

Step 3. Set $x^{s+1}=x_i^s$, $g^{s+1}=g_i^s$.

Step 4. Set $s=s+1$; return to Step 2.

Let us introduce some additional designations. Let

$$T(x)=\{y\in R^n: f(y)\leq f(x)\},$$

and let L be a Lipschitz constant of the function f on the set $T(x^0)$. Recall that the function f is considered strictly convex on the set $T(x^0)$ if

$$\alpha_1 f(x)+\alpha_2 f(y)>f(\alpha_1 x+\alpha_2 y),$$

for all α_1, α_2, x, y such that

$$\alpha_1+\alpha_2=1, \qquad \alpha_1>0, \qquad \alpha_2>0, \qquad x\in T(x^0), \qquad y\in T(x^0).$$

We next formulate a theorem about the convergence of Algorithm 4.2.

Theorem 4.2. Let a function $f: R^n\to R$ be strictly convex, possibly nonsmooth, on the compact set $T(x^0)$; let $v>0$ be given; and let the sequence $\{\lambda_j\}$, $j=0,1,\ldots,$ of positive numbers satisfy the conditions

$$\sum_{i=0}^\infty\lambda_j=\infty, \qquad \sum_{i=0}^\infty\lambda_j^2<\infty.$$

Then, there exists $\bar{s}$ such that

$$f(x^{\bar{s}})-f^*\leq 2vL.$$

Proof. Prior to proving Theorem 4.2, let us prove several lemmas. First, let us prove that the norm of the matrix H_i^s is uniformly bounded for all $s>0$, $i>0$.

Lemma 4.1. The inequalities

$$\|H_i^s\|^2 \le \mathrm{Tr}(H_i^s H_i^{s^T}) \le n + L^2 \sum_{l=0}^{\infty} \lambda_l^2 \tag{27}$$

hold.

Proof. The inequality

$$\|H_i^s\| \le \mathrm{Tr}(H_i^s H_i^{s^T})$$

follows from the definitions of norm and trace of a matrix. Step 2(vi) implies that

$$H_{i+1}^s H_{i+1}^{s^T} = H_i^s H_i^{s^T} + \lambda_{j(s,i)}(\xi_i^s g^{s^T} H_i^{s^T} + H_i^s g^s \xi_i^{s^T}) + \lambda_{j(s,i)}^2 \xi_i^s \xi_i^{s^T} \|g^s\|^2.$$

Using this equality and taking into account that, due to the construction of the algorithm,

$$\langle \xi_i^s, H_i^s g^s \rangle \le 0, \qquad \|\xi_i^s\| \le 1,$$

we have

$$\begin{aligned}
\mathrm{Tr}(H_{i+1}^s H_{i+1}^{s^T}) &= \mathrm{Tr}(H_i^s H_i^{s^T}) + 2\lambda_{j(s,i)} \langle \xi_i^s, H_i^s g^s \rangle + \lambda_{j(s,i)}^2 \|\xi_i^s\|^2 \|g^s\|^2 \\
&\le \mathrm{Tr}(H_i^s H_i^{s^T}) + \lambda_{j(s,i)}^2 \|g^s\|^2 \\
&\le \mathrm{Tr}(H_0^s H_0^{s^T}) + \|g^s\|^2 \sum_{l=j(s,0)}^{j(s,i)} \lambda_l^2 \\
&\le \mathrm{Tr}(H_{i(s-1)+1}^{s-1} H_{i(s-1)+1}^{s-1^T}) + L^2 \sum_{l=j(s,0)}^{j(s,i)} \lambda_l^2 \\
&\le \mathrm{Tr}(H_0^{-1} H_0^{-1^T}) + L^2 \sum_{l=0}^{j(s,i)} \lambda_l^2 \\
&= n + L^2 \sum_{l=0}^{j(s,i)} \lambda_l^2 \\
&\le n + L^2 \sum_{l=0}^{\infty} \lambda_l^2.
\end{aligned}$$

Lemma 4.1 is proved. □

Lemma 4.2. There exists $\bar{s}$ such that $\|x_i^{\bar{s}} - x^{\bar{s}}\| < \nu$ for all $i \ge 0$.

Proof. The statement of this lemma follows from the following lemma (Ref. 7).

Lemma 4.3. Suppose that the function $f(x)$ is strictly convex on R^n, the set $T(x^0)$ is bounded, and there exists a sequence $\{x^s\}$ such that $\{x^s\} \subset T(x^0)$ and

$$f(x^{s+1}) = \min_{\alpha \in [0,1]} f(x^s + \alpha(x^{s+1} - x^s)). \tag{28}$$

Then,

$$\lim_{s \to \infty} \|x^{s+1} - x^s\| = 0.$$

If Lemma 4.2 does not hold, then Step 2(ii) and (iv) imply that the sequence $\{x^s\}$ satisfies condition (28) and that, beginning with some s^*, the inequality

$$\|x^{s+1} - x^s\| < \nu \tag{29}$$

holds. This contradicts Step 2(vii), since

$$\|x^{s+1} - x^s\| = \|x_{i(s)}^s - x^s\| \ge \nu,$$

due to the construction of the algorithm. This contradiction proves Lemma 4.2. □

To complete the proof of the theorem, we need one more lemma. Recall that the set

$$\partial_\epsilon f(z) = \{g \in R^n : f(x) - f(z) \ge \langle g, x - z \rangle - \epsilon, \forall x \in R^n\}$$

is called the ϵ-subdifferential of the convex function $f(x)$ at the point $z \in R^n$ (Refs. 7, 16). A vector $g \in \partial_\epsilon f(z)$ is an ϵ-subgradient of the function f at a point z.

Lemma 4.4. Suppose that $g \in \partial f(z)$, $z \in T(x^0)$, $y \in T(x^0)$. If $\|z - y\| \le \nu$, then $g \in \partial_{2\nu L} f(y)$.

Proof. Using the definition of subdifferential, we have

$$\begin{aligned}
f(x) &\ge \langle g, x - z \rangle + f(z) \\
&= \langle g, x - y \rangle + f(y) + \langle g, y - z \rangle + f(z) - f(y) \\
&\ge \langle g, x - y \rangle + f(y) - \|g\| \, \|z - y\| - |f(z) - f(y)|.
\end{aligned} \tag{30}$$

Since the function $f(x)$ satisfies the Lipschitz condition with a constant L, then

$$\|g\| \le L, \qquad |f(z) - f(y)| \le L\|z - y\|.$$

Consequently,

$$\|g\|\,\|z-y\|+|f(z)-f(y)|\leq 2L\|z-y\|<2Lv.$$

It follows from (30) and the last inequality that

$$f(x)\geq\langle g, x-y\rangle+f(y)-2Lv.$$

□

Let us prove that $\bar{s}$ from Lemma 4.2 satisfies the statement of the theorem. Suppose that it does not hold, i.e.,

$$f(x^{\bar{s}})-f^*>2vL. \tag{31}$$

Since $\|x_i^{\bar{s}}-x^{\bar{s}}\|<v$, then according to Lemma 4.4,

$$g_i^{\bar{s}}\in\partial_{2vL}f(x^{\bar{s}}).$$

It follows from (31) that $0\notin\partial_{2vL}f(x^{\bar{s}})$; see, for example, Lemma 8.1 of Ref. 7. The set $\partial_{2vL}f(x^{\bar{s}})$ is convex, closed, and bounded (Ref. 7). Denote by q some vector satisfying

$$q\in\partial_{2vL}f(x^{\bar{s}}),\qquad \|q\|=\min_{g\in\partial_{2vL}f(x^{\bar{s}})}\|g\|>0.$$

By the definition of matrix norm,

$$\|H_{i+1}^{\bar{s}}\|=\max_{\|u\|\leq 1}\max_{\|v\|\leq 1}\langle u, H_{i+1}^{\bar{s}}v\rangle\geq\|q\|^{-2}\langle q, H_{i+1}^{\bar{s}}q\rangle. \tag{32}$$

By construction of the algorithm,

$$H_{i+1}^{\bar{s}}=H_i^{\bar{s}}+\lambda_{j(s,i)}\xi_i^{\bar{s}}g^{\bar{s}^T}=H_0^{\bar{s}}+\sum_{l=0}^{i}\lambda_{j(s,l)}\xi_l^{\bar{s}}g^{\bar{s}^T}.$$

Inequality (32) implies that

$$\|H_{i+1}^{\bar{s}}\|\geq\left\langle q,\left(H_0^{\bar{s}}+\sum_{l=0}^{i}\lambda_{j(s,l)}\xi_l^{\bar{s}}g^{\bar{s}^T}\right)q\right\rangle$$
$$=\langle q, H_0^{\bar{s}}q\rangle+\sum_{l=0}^{i}\lambda_{j(s,l)}\langle q,\xi_l^{\bar{s}}\rangle\langle g^{\bar{s}}, q\rangle. \tag{33}$$

Since $g_i^{\bar{s}}\in\partial_{2vL}f(x^{\bar{s}})$, $i\geq 0$, $g^{\bar{s}}\in\partial_{2vL}f(x^{\bar{s}})$, and the set $\partial_{2vL}f(x^{\bar{s}})$ is convex, then there exists a positive number α such that

$$\langle q,\xi_i^{\bar{s}}\rangle\geq\alpha,\qquad\langle q, g^{\bar{s}}\rangle\geq\alpha.$$

Applying (33), we see that

$$\|H_{i+1}^{\bar{s}}\|>\langle q, H_0^{\bar{s}}q\rangle+\alpha^2\sum_{l=0}^{i}\lambda_{j(s,l)}.$$

By the conditions of the theorem,

$$\sum_{l=0}^{i}\lambda_{j(s,l)}\to\infty,\qquad i\to\infty;$$

consequently,

$$\|H_{i+1}^{\bar{s}}\|\to+\infty,\qquad i\to\infty.$$

This last statement contradicts Lemma 4.1, and the proof of Theorem 4.2 is complete. □

5. Algorithm with Symmetric Matrices

The algorithms discussed above have the following deficiency: one must store an $n\times n$ matrix, where n is the dimension of the source problem. We next propose an algorithm with symmetric matrices; to store such matrices requires only $(n^2+n)/2$ numbers. The function $\varphi_s(H)=f(x^2-\rho_sHg^s)$ characterizes the choice of a matrix H. Denote by G the set of symmetric $n\times n$ matrices. The set G is a linear space. For the adaptation of the matrix, we can consider the problem

$$\varphi_s(H)\to\min_{H\in G}.$$

Analogous to (4), one can use the gradient algorithm with projection onto the set G,

$$H_{i+1}^s=\Pi_G(H_i^s+\lambda_i^sg_i^sg^{s^T}),\qquad\lambda_i^s>0,$$

where $g_i^s\in\partial f(x^s-\rho_sH_i^sg^s)$ and Π_G is the projection operation onto the set G.

Lemma 5.1. If $H_i^s\in G$, then

$$\Pi_G(H_i^s+\lambda_i^sg_i^sg^{s^T})=H_i^s+2^{-1}\lambda_i^s(g_i^sg^{s^T}+g^sg_i^{s^T}).$$

Proof. It is evident that the matrix $H_i^s+2^{-1}\lambda_i^s(g_i^sg^{s^T}+g^sg_i^{s^T})$ is symmetric if $H_i^s\in G$. To prove the lemma, it is enough to show that the matrix

$$[H_i^s+2^{-1}\lambda_i^s(g_i^sg^{s^T}+g^sg_i^{s^T})]-[H_i^s+\lambda_i^sg_i^sg^{s^T}]=2^{-1}\lambda_i^s(g^sg_i^{s^T}-g_i^sg^{s^T})$$

is orthogonal to any symmetric matrix H. But

$$\langle H, g^sg_i^{s^T}-g_i^sg^{s^T}\rangle=\langle H, g^sg_i^{s^T}\rangle-\langle H, g_i^sg^{s^T}\rangle$$
$$=\langle H, g^sg_i^{s^T}\rangle-\langle H^T, g^sg_i^{s^T}\rangle=0.$$

□

Thus, to update the matrix H_0^s, one can use the algorithm

$$H_{i+1}^s = H_i^s + \lambda_i^s(g_i^s g^{s^T} + g^s g_i^{s^T}). \tag{34}$$

It is convenient to normalize the vector g_i^s; therefore, we rewrite formula (34) as

$$H_{i+1}^s = H_i^s + \lambda_i^s(\xi_i^s g^{s^T} + g^s \xi_i^{s^T}). \tag{35}$$

The symmetric formula for matrix modification can be combined with Algorithms 4.1 and 4.2. Theorem 4.2 can be proved for this algorithm without any difference. For this reason, we do not dwell on convergence proofs for algorithms with the matrix modification formula (35).

6. Algorithm with Positive Matrices

In the algorithms described above, the matrix H^s can be, generally speaking, nonpositive. If the function $f(x)$ is convex and $g^s \in \partial f(x^s)$, then the minimum point of problem (1) belongs to the subspace

$$A_s = \{x \in R^n : \langle x - x^s, g^s \rangle \leq 0\}.$$

It is possible that the point $x^{s+1} = x^s - \rho_s H g^s$ does not belong to the subspace A_s if H^s is nonpositive. To guarantee the positiveness of the matrix H^s, let us consider the case where the matrix H^s is represented as follows:

$$H^s = B^s B^{s^T},$$

where B^s is an $n \times n$ matrix. In this case, the iteration of the algorithm is given by the formula

$$x^{s+1} = x^s - \rho_s B^s B^{s^T} g^s,$$

where $g^s \in \partial f(x^s)$. The function $\varphi_s(B) = f(x^s - \rho_s B B^T g^s)$ defines the choice of the matrix B. If the function $f(x)$ is convex, then it can be proved that the function $\varphi_s(B)$ is weakly convex. Next, we use the family of weakly convex functions that was investigated in Ref. 19. Other analogous families of functions can also be used; see, for example, Ref. 20.

Let X be a convex subset of R^n, possibly $X = R^n$. A continuous function f on X is called weakly convex on the set X if, for all $x \in X$, the set $\partial f(x)$ consisting of vectors g such that

$$f(y) - f(x) \geq \langle g, y - x \rangle + \zeta(x, y), \qquad y \in X,$$

is nonempty, where $\zeta(x, y)$ is uniformly small with respect to $\|x - y\|$ on each compact subset $K \subset X$; i.e., for each $\epsilon > 0$, there exists $\delta > 0$ such that

$$\|\zeta(x, y)\| / \|x - y\| < \epsilon, \qquad x, y \in K, \qquad \|x - y\| < \delta.$$

Lemma 6.1. Let the function $f: R^n \to R$ be convex on R^n; and let the set $\mathscr{B} \in R^{n \times n}$ be convex. Then, the function $\varphi(B) = f(x - \rho B B^T \xi)$ is weakly convex on $\mathscr{B}$ and

$$\partial \varphi(B) = \{-\rho(\eta \xi^T + \xi \eta^T) B : \eta \in \partial f(x - \rho B B^T \xi)\}.$$

Proof. Let K be a compact subset of $\mathscr{B} \subset R^{n \times n}$; and let $B \in K$, $\Delta B \in \mathscr{B}$, $B + \Delta B \in K$, and $\eta \in \partial f(x - \rho B B^T \xi)$. Denote

$$X_{\mathscr{B}} = \{y \in R^n : y = x - \rho B B^T \xi,\ B \in K\}.$$

The function $f: R^n \to R$ is Lipschitz with some constant $L_{\mathscr{B}}$ on $X_{\mathscr{B}}$, because the function f is convex on R^n and the set $X_{\mathscr{B}}$ is compact. Using the Lipschitz and convexity properties of $f(x)$, we have

$$\begin{aligned}
\varphi(B + \Delta B) &= f(x - \rho(B + \Delta B)(B^T + \Delta B^T)\xi) \\
&= f(x - \rho(BB^T + B\Delta B^T + \Delta B B^T + \Delta B^T \Delta B^T)\xi) \\
&\geq f(x - \rho(BB^T + B\Delta B^T + \Delta B B^T)\xi) - L_{\mathscr{B}} \|\rho \Delta B \Delta B^T \xi\| \\
&\geq f(x - \rho B B^T \xi) + \langle -\rho(B\Delta B^T + \Delta B B^T)\xi, \eta \rangle \\
&\quad - L_{\mathscr{B}} \|\rho \Delta B \Delta B^T \xi\| \\
&= \varphi(B) - \rho \langle (\xi \eta^T + \eta \xi^T) B, \Delta B \rangle - L_{\mathscr{B}} \|\rho \Delta B \Delta B^T \xi\|.
\end{aligned}$$

Since the value $\|\Delta B \Delta B^T g\|$ is uniformly small with respect to $\|\Delta B\|$, the lemma is proved. □

Lemma 6.1 gives a formula for the subdifferential of the function $\varphi_s(B)$. For the adaptation of the matrix B^s, the following gradient method can be used:

$$B_{i+1}^s = B_i^s + \gamma_{si}(\xi_i^s g^{s^T} + g^s \xi_i^{s^T}) B_i^s, \qquad i = 0, 1, \ldots, \tag{36}$$

where ξ_i^s denotes the normalized vector $g_i^s \in \partial f(x_i^s)$; see (19). Analogous to Algorithm 4.2, we write an algorithm with the matrix modification formula (36).

Algorithm 6.1.

Step 1. Initialization. Set

$$s=0, \qquad B_0^{-1}=I, \qquad i=-1, \qquad j=-1,$$

$$x^0=x_{\text{init}}, \qquad g^0\in\partial f(x^0).$$

Step 2. (i) Set $B_0^s=\begin{cases} I, & \text{if } \|B_{i+1}^{s-1}g^s\|=0,\\ B_{i+1}^{s-1}, & \text{otherwise.}\end{cases}$

(ii) Set $i=0$.
(iii) Set $\rho_{si}=\arg\min_{\rho>0} f(x^s-\rho B_i^s B_i^{s^T} g^s)$.
(iv) Set $j=j+1, j(s,i)=j$.
(v) Set $x_i^s=x^s-\rho_{si}B_i^sB_i^{s^T}g^s$.
(vi) Compute $g_i^s\in\partial f(x_i^s)$ such that $\langle g_i^s, B_i^sB_i^{s^T}g^s\rangle\le 0$.
(vii) Set $B_{i+1}^s=B_i^s+\lambda_j(\xi_i^s g^{s^T}+g^s\xi_i^{s^T})B_i^s$.
(viii) If $\|x_i^s-x^s\|\ge v$, then set $i(s)=i$, and go to Step 3.
(ix) Set $i=i+1$, and return to (iii).

Step 3. Set $x^{s+1}=x_i^s$, $g^{s+1}=g_i^s$.

Step 4. Set $s=s+1$, and return to Step 2.

We formulate a theorem about the convergence of Algorithm 6.1 for smooth objective functions.

Theorem 6.1. Let the function $f:R^n\to R$ be strictly convex and smooth; let L_1 be the Lipschitz constant of the function f on the compact set $T(x^0)$; and let L_2 be the Lipschitz constant for the gradient $\nabla f(x)$ on the set

$$T_v(x^0)\overset{\text{def}}{=}\left\{x:\ \min_{y\in T(x^0)}\|x-y\|\le v\right\}.$$

Let there be given a value $v>0$ and a sequence of positive numbers $\{\lambda_j\}$ satisfying

$$\sum_{j=0}^{\infty}\lambda_j=\infty, \qquad \sum_{j=0}^{\infty}\lambda_j^2<\infty.$$

Then, for Algorithm 6.1, there exists $\bar{s}$ such that

$$\|g^{\bar{s}}\|\le 2vL_2.$$

Proof. Prior to proving Theorem 6.1, we need two lemmas. First, we evaluate the norm of the matrix B_i^s.

Lemma 6.2. The inequalities

$$\|B_i^s\|^2\le \operatorname{Tr}(B_i^sB_i^{s^T})\le n\prod_{l=0}^{\infty}(1+4L_1^2\lambda_l^2)<\infty \tag{37}$$

hold for all integers $s>0$, $i\ge 0$.

Proof. It follows from Step 2(vii) that

$$\begin{aligned} B_{i+1}^sB_{i+1}^{s^T}&=B_i^sB_i^{s^T}+\lambda_{j(s,i)}[(\xi_i^s g^{s^T}+g^s\xi_i^{s^T})B_i^sB_i^{s^T}+B_i^sB_i^{s^T}(\xi_i^sg^{s^T}+g^s\xi_i^{s^T})]\\ &\quad+\lambda_{j(s,i)}^2(\xi_i^sg^{s^T}+g^s\xi_i^{s^T})B_i^sB_i^{s^T}(\xi_i^sg^{s^T}+g^s\xi_i^{s^T}). \end{aligned}\tag{38}$$

We denote by $k(s)$ the maximal iteration number such that $k(s)\le s$ and $B_0^{k(s)}=I$. Using (38) and taking account of the fact that, due to the construction of the algorithm,

$$\langle \xi_i^s, B_i^sB_i^{s^T}g^s\rangle=0 \quad\text{and}\quad \|\xi_i^s\|\le 1,$$

we obtain

$$\begin{aligned} \operatorname{Tr}(B_{i+1}^sB_{i+1}^{s^T})&=\operatorname{Tr}(B_i^sB_i^{s^T})+\lambda_{j(s,i)}4\langle\xi_i^s, B_i^sB_i^{s^T}g^s\rangle\\ &\quad+\lambda_{j(s,i)}^2[\langle\xi_i^s,g^s\rangle\langle\xi_i^s,B_i^sB_i^{s^T}g^s\rangle+\|\xi_i^s\|^2\langle g^s,B_i^sB_i^{s^T}g^s\rangle\\ &\quad+\|g^s\|^2\langle\xi_i^s,B_i^sB_i^{s^T}\xi_i^s\rangle+\langle g^s,\xi_i^s\rangle\langle\xi_i^s,B_i^sB_i^{s^T}g^s\rangle]\\ &\le\operatorname{Tr}(B_i^sB_i^{s^T})+\lambda_{j(s,i)}^2[\|\xi_i^s\|^2\|g^s\|^2\|B_i^sB_i^{s^T}\|\\ &\quad+\|\xi_i^s\|^2\|g^s\|^2\|B_i^sB_i^{s^T}\|+\|g^s\|^2\|\xi_i^s\|^2\|B_i^sB_i^{s^T}\|\\ &\quad+\|g^s\|^2\|\xi_i^s\|^2\|B_i^sB_i^{s^T}\|]\\ &\le\operatorname{Tr}(B_i^sB_i^{s^T})+4\lambda_{j(s,i)}^2\|g^s\|^2\|B_i^sB_i^{s^T}\|\\ &\le\operatorname{Tr}(B_i^sB_i^{s^T})+4\lambda_{j(s,i)}^2L_1^2\operatorname{Tr}(B_i^sB_i^{s^T})\\ &=\operatorname{Tr}(B_i^sB_i^{s^T})(1+4L_1^2\lambda_{j(s,i)}^2)\\ &=\operatorname{Tr}(B_0^sB_0^{s^T})\prod_{l=0}^{i}(1+4L_1^2\lambda_{j(s,l)}^2)\\ &=\operatorname{Tr}(B_{i(s-1)+1}^{s-1}B_{i(s-1)+1}^{s-1^T})\prod_{l=0}^{i}(1+4L_1^2\lambda_{j(s,l)}^2)\\ &\le\operatorname{Tr}(B_0^{k(s)}B_0^{k(s)^T})\prod_{l=j(k(s),0)}^{j(s,i)}(1+4L_1^2\lambda_l^2)\\ &\le n\prod_{l=0}^{\infty}(1+4L_1^2\lambda_l^2). \end{aligned}$$

The inequality

$$\prod_{i=0}^{\infty} (1+4L_1^2\lambda_i^2) < \text{const}$$

follows from the convergence of the series $\sum_0^\infty \lambda_i^2$ in the conditions of the theorem. □

Lemma 6.3. There exists $\bar{s}$ such that $\|x_i^{\bar{s}} - x^{\bar{s}}\| < \nu$ for all $i \geq 0$.

Proof. We prove the lemma by contradiction. Suppose that the statement of the lemma does not hold. By the construction of the algorithm, the sequence $\{x^s\}$ satisfies the assumptions of Lemma 4.3. Consequently,

$$\|x^{s+1} - x^s\| \to 0, \qquad s \to \infty.$$

Applying Step 2(viii), we see that

$$\|x^{s+1} - x^s\| \geq \nu,$$

and obtain a contradiction. □

Now let us prove, by contradiction, the statement of the theorem. We show that, for $\bar{s}$ from Lemma 6.3, the statement of the theorem holds. Suppose that this is not the case, i.e.,

$$\|g^{\bar{s}}\| \geq 2\nu L_2. \tag{39}$$

Since $\|x_i^{\bar{s}} - x^{\bar{s}}\| < \nu$, then $\|g_i^{\bar{s}} - g^{\bar{s}}\| \leq L_2\nu$, because the gradient of the function $f(x)$ satisfies a Lipschitz condition. Write the following inequalities

$$\begin{aligned}\langle \xi_i^{\bar{s}}, g^{\bar{s}}\rangle &= \langle g_i^{\bar{s}}/\|g_i^{\bar{s}}\|, g^{\bar{s}}\rangle = \langle (g_i^{\bar{s}} - g^{\bar{s}} + g^{\bar{s}})/\|g_i^{\bar{s}}\|, g^{\bar{s}}\rangle \\ &\geq \|g^{\bar{s}}\|^2/\|g_i^{\bar{s}}\| - \|g_i^{\bar{s}} - g^{\bar{s}}\| \cdot \|g^{\bar{s}}\|/\|g_i^{\bar{s}}\| \\ &= (\|g^{\bar{s}}\| - \|g_i^{\bar{s}} - g^{\bar{s}}\|)\|g^{\bar{s}}\|/\|g_i^{\bar{s}}\| \\ &\geq (\|g^{\bar{s}}\| - L_2\nu)\|g^{\bar{s}}\| \, \|g_i^{\bar{s}}\|^{-1} \\ &\geq L_2\nu\|g^{\bar{s}}\|L_1^{-1} \geq 2L_2^2\nu^2L_1^{-1}.\end{aligned} \tag{40}$$

We evaluate from below the value $\|B_{i+1}^{\bar{s}}B_{i+1}^{\bar{s}T}\|$. Using the relations $\langle \xi_i^{\bar{s}}, B_i^{\bar{s}}B_i^{\bar{s}T}g^{\bar{s}}\rangle = 0$, (38), and (40), we get

$$\begin{aligned}\|g^{\bar{s}}\|^2\|B_{i+1}^{\bar{s}}B_{i+1}^{\bar{s}T}\| &\geq \langle g^{\bar{s}}, B_{i+1}^{\bar{s}}B_{i+1}^{\bar{s}T}g^{\bar{s}}\rangle \\ &= \|B_{i+1}^{\bar{s}T}g^{\bar{s}}\|^2 = \langle g^{\bar{s}}, B_i^{\bar{s}}B_i^{\bar{s}T}g^{\bar{s}}\rangle \\ &+ \lambda_{j(\bar{s},i)}\langle g^{\bar{s}}, [(\xi_i^{\bar{s}}g^{\bar{s}T} + g^{\bar{s}}\xi_i^{\bar{s}T})B_i^{\bar{s}}B_i^{\bar{s}T} \\ &+ B_i^{\bar{s}}B_i^{\bar{s}T}(\xi_i^{\bar{s}}g^{\bar{s}T} + g^{\bar{s}}\xi_i^{\bar{s}T})]g^{\bar{s}}\rangle \\ &+ \lambda_{j(\bar{s},i)}^2\langle g^{\bar{s}}, (\xi_i^{\bar{s}}g^{\bar{s}T} + g^{\bar{s}}\xi_i^{\bar{s}T})B_i^{\bar{s}}B_i^{\bar{s}T}(\xi_i^{\bar{s}}g^{\bar{s}T} + g^{\bar{s}}\xi_i^{\bar{s}T})g^{\bar{s}}\rangle \\ &\geq \|B_i^{\bar{s}T}g^{\bar{s}}\|^2 + 2\lambda_{j(\bar{s},i)}\langle g^{\bar{s}}, \xi_i^{\bar{s}}\rangle\langle g^{\bar{s}}, B_i^{\bar{s}}B_i^{\bar{s}T}g^{\bar{s}}\rangle \\ &\geq (1 + 2\lambda_{j(\bar{s},i)}\langle g^{\bar{s}}, \xi_i^{\bar{s}}\rangle)\|B_i^{\bar{s}T}g^{\bar{s}}\|^2 \\ &\geq (1 + 4\lambda_{j(\bar{s},i)}L_2^2\nu^2L_1^{-1})\|B_i^{\bar{s}T}g^{\bar{s}}\|^2 \\ &\geq \|B_0^{\bar{s}T}g^{\bar{s}}\|^2 \prod_{l=0}^{i}(1 + 4\lambda_{j(\bar{s},l)}L_2^2\nu^2L_1^{-1}).\end{aligned}$$

Since

$$\sum_0^\infty \lambda_j = \infty, \qquad \lambda_i > 0, \qquad i = 0, 1, \ldots,$$

then

$$\prod_{l=0}^{i}(1 + 4\lambda_{j(\bar{s},l)}L_2^2\nu^2L_1^{-1}) \to +\infty, \qquad i \to \infty.$$

Consequently, $\|B_{i+1}^{\bar{s}}B_{i+1}^{\bar{s}T}\| \to +\infty$, and this contradicts Lemma 6.2. □

7. Results of Numerical Experiments

Here, we describe two practical algorithms based on Algorithm 6.1. These algorithms have some heuristic features that are not yet rigorously proven. However, the results of numerical experiments seem to verify their effectiveness. In practical situations, the steepest descent is impossible in all cases, and one is forced to use approximations. These algorithms use rather rough approximations. Some heuristic formulas are also proposed for stepsize control with respect to matrices.

First, we describe an algorithm with stepsize control via gradient calculations.

Algorithm 7.1.

Step 0. Parameters: $\alpha_1 = 1.25$, $\alpha_2 = 0.7$, $\alpha_3 = 0.65$.
Data: x, ρ, Δ_*, G_*.

Step 1. Initialization. Set
$s = 1, \quad s_f = 1, \quad l = 0, \quad l_b = 0, \quad \rho_p = \rho, \quad B = I.$

Step 2. Function and subgradient calculation:
$f = f(x), \quad g \in \partial f(x).$

Step 3. If $\|g\| \leq G_*$, then stop.

Step 4. Subgradient normalization: $g = g/\|g\|$.

Step 5. Set $g_d = g$.
Step 6. Set $g_b = B^T g$, $g_m = BB^T g$.
Step 7. Set $x_p = x$, $f_p = f$, $g_p = g$.
Step 8. Set $x = x - (\rho / \|g_b\|) g_m$.
Step 9. If $\|x - x_p\| \leq \Delta_*$, then stop.
Step 10. Function and subgradient computation:
$f = f(x)$, $g \in \partial f(x)$.
Step 11. Set $s_f = s_f + 1$.
Step 12. If $\langle g_m, g \rangle > 0$, then set $l = l + 1$; if $s = 1$, then set $\rho = 1.5\rho$; if $l > 1$ and $s > 1$, then set $\rho = \alpha_1 \rho$; go to Step 7.
Step 13. If $l = 0$, then set $\rho = \alpha_2 \rho$.
Step 14. Set $l_b = l_b + 1$.
Step 15. If $l = 0$ and $l_b \leq 3$, then set $\rho = \max(\rho_p, \rho)$.
Step 16. If $l > 0$, then set $\rho_p = \rho$; set $l_b = 0$.
Step 17. If $\|g\| \leq G_*$, then stop.
Step 18. Subgradient normalization: $g = g / \|g\|$.
Step 19. Set $B = B + \alpha_3 (g_d g^T B + g g_d^T B)$.
Step 20. Return to the point with the smallest function value: If $f_p < f$, then set $x = x_p$; set $f = f_p$; if $l > 0$, then {if $\|g_p\| \leq G_*$ then stop; set $g_p = g_p / \|g_p\|$}; set $g = g_p$.
Step 21. Set $l = 0$, $s = s + 1$.
Step 22. Go to Step 5.

In the next algorithm, stepsize control takes place via the objective function calculation.

Algorithm 7.2.

Step 0. Parameters: $\alpha_1 = 1.25$, $\alpha_2 = 0.8$, $\alpha_3 = 0.55$.
Data: x, ρ, Δ_*, G_*.
Step 1. Initialization. Set
$s = 1$, $s_f = 0$, $l = 0$, $B = I$.
Step 2. Function and subgradient calculation:
$f = f(x)$, $g \in \partial f(x)$.
Step 3. If $\|g\| \leq G_*$, then stop.
Step 4. Subgradient normalization: $g = g / \|g\|$.
Step 5. Set $g_d = g$.
Step 6. Set $g_b = B^T g$, $g_m = B g_b$.
Step 7. Set $x_p = x$, $f_p = f$.
Step 8. Set $x = x - (\rho / \|g_b\|) g_m$.
Step 9. If $\|x - x_p\| \leq \Delta_*$, then stop.
Step 10. Function computation: $f = f(x)$.

Step 11. Set $s_f = s_f + 1$.
Step 12. If $f < f_p$, then set $l = l + 1$; if $s = 1$, then set $\rho = 1.5\rho$; if $l > 1$ and $s > 1$, then set $\rho = \alpha_1 \rho$; go to Step 7.
Step 13. If $l = 0$, then set $\rho = \alpha_2 \rho$.
Step 14. Set $g \in \partial f(x)$.
Step 15. If $\|g\| \leq G_*$, then stop.
Step 16. Subgradient normalization: $g = g / \|g\|$.
Step 17. Set $B = B + \alpha_3 (g_d g^T B + g g_d^T B)$.
Step 18. Set $x = x_p$, $f = f_p$.
Step 19. If $l = 0$, then $g = g_d$.
Step 20. Set $l = 0$, $s = s + 1$.
Step 21. Go to Step 7.

In Algorithm 7.1, s is the number of main algorithm iterations and s_f is the number of subgradient and objective function value calculations. The subgradients and objective function values are calculated at the same points.

In Algorithm 7.2, s gives the amount of subgradient calculation and s_f is the number of objective function value calculations.

We consider the computation results for the following three examples.

Example 7.1. The objective function is

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2.$$

The optimal point is $x^* = (1, 1)$, $f(x^*) = 0$. The initial point is $x_{\text{init}} = (-1, 1)$, $f(x_{\text{init}}) = 2$. The initial stepsize is $\rho = 0.1$.

Algorithm 7.1 gives

$$x^{66} = (1.00000, 1.00000), \qquad f(x^{66}) = 0.63 \cdot 10^{-13},$$

$$s = 66, \qquad s_f = 100.$$

Algorithm 7.2 gives

$$x^{155} = (1.00000, 1.00000), \qquad f(x^{155}) = 0.22 \cdot 10^{-13},$$

$$s = 155, \qquad s_f = 271.$$

Example 7.2. See Ref. 11. The objective function is

$$f(x) = \max_{i = 1,10} \varphi_i(x), \qquad x \in R^5,$$

$$\varphi_i(x) = d_i \sum_{j=1}^{5} (x_j - c_{ij})^2,$$

where

$$x=\{x_1,\dots,x_5\},$$

$$\{d_1,\dots,d_{10}\}=\{1, 5, 10, 2, 4, 3, 1.7, 2.5, 6, 3.5\},$$

and the transposed matrix $\{c_{ij}\}$ is

$$\{c_{ij}\}^T=\begin{bmatrix} 0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

The optimal point is

$$x^*=(1.12434, 0.97945, 1.47770, 0.92023, 1.12429),$$

$$f(x^*)=22.60016.$$

The initial point is

$$x_{\text{init}}=(0, 0, 0, 0, 1), \qquad f(x_{\text{init}})=80.$$

The initial stepsize is $\rho=0.1$.

Algorithm 7.1 yields the optimal point x^* with $s=57$, $s_f=87$.

Algorithm 7.2 yields the optimal point x^* with $s=76$, $s_f=123$.

Example 7.3. This problem investigates the optimal control of a spring, mass, and damper system. It is adapted from Ref. 21. While it is acknowledged that there may be simple ways of solving the problem by taking specific advantage of the nature of the constraints, it serves the present purpose of providing a large nonsmooth optimization problem. The original problem is

$$\min \quad \varphi(x, y, u)=(1/2)\sum_{t=0}^{T} x_t^2, \tag{41}$$

$$\text{s.t.} \quad x_{t+1}=x_t+0.2y_t, \tag{42a}$$

$$y_{t+1}=y_t-0.01y_t^3-0.004x_t+0.2u_t, \tag{42b}$$

$$x_0=10, \quad y_0=0, \quad y_T=0, \tag{43a}$$

$$-0.2\le u_t\le 0.2, \tag{43b}$$

$$y_t\ge -1.0, \tag{43c}$$

with $t=0,\dots,T-1$. We use a nonsmooth exact penalty function to reduce this problem to the following optimization problem:

$$\begin{aligned}\min \quad f(x, y, u)=(1/2)\sum_{t=0}^{T} x_t^2+k_1 &\\ \times \sum_{t=0}^{T-1}\max\{-0.2-u_t;\, u_t-0.2;\, 0\}+k_2|y_T|+k_3&\\ \times\sum_{t=1}^{T}\max\{-1-y_t, 0\},&\end{aligned} \tag{44}$$

$$\text{s.t.} \quad (42) \text{ and } x_0=10,\ y_0=0.$$

The constraints (42) imply that

$$x=x(u), \qquad y=y(u),$$

and we can reformulate (44) as follows:

$$\min \quad F(u)=f(x(u), y(u), u). \tag{45}$$

This problem has dimension 100 with respect to the variables $u_0,\dots,u_{T-1}$. The function $F(u)$ is locally Lipschitz over u. The subgradient of the locally Lipschitz function $F(u)$ can be computed by the following formulas (Refs. 23, 23), where $t=0,\dots,T-1$:

$$g(u)=(g_0(u),\dots,g_{T-1}(u))\in\partial F(u),$$

$$g_t(u)=-0.2\lambda_{2,t+1}+k_1\begin{cases}-1, & \text{if } u_t<-0.2,\\ +1, & \text{if } u_t>0.2,\\ 0, & \text{otherwise},\end{cases}$$

$$\begin{aligned}\begin{bmatrix}\lambda_{1t}\\ \lambda_{2t}\end{bmatrix}&=\begin{bmatrix}\lambda_{1,t+1}-0.004\lambda_{2,t+1}\\ 0.2\lambda_{1,t+1}+(1-0.03y_t^2)\lambda_{2,t+1}\end{bmatrix}-\begin{bmatrix}x_t\\ 0\end{bmatrix}\\ &\quad -k_3\begin{bmatrix}0\\ \begin{cases}-1, & \text{if } y_t<-1\\ 0, & \text{otherwise}\end{cases}\end{bmatrix}\end{aligned} \tag{46}$$

$$\lambda_{1T}=-x_T,$$

$$\lambda_{2T}=-k_2\begin{cases}1, & \text{if } y_T>0\\ -1, & \text{otherwise}\end{cases}-k_3\begin{cases}-1, & \text{if } y_T<-1,\\ 0, & \text{otherwise}.\end{cases} \tag{47}$$

To calculate the subgradient $g(u)$, one should first solve the difference equation (42) with initial conditions $x_0=10$, $y_0=0$, then the difference equation

(46) with initial conditions (47). The constants k_1, k_2, k_3 should be large enough, but they should not tend to infinity due to the exactness of the nonsmooth penalties. The numerical experiments show that we can take $k_1 = 500$, $k_2 = 100$, $k_3 = 500$.

Let us first consider the experimental results for Algorithm 7.2. In Table 1, s is the number of the algorithm iteration (amount of gradient calculation), s_f is the total amount of objective function calculation, ρ is the stepsize, $F(u^s)$ is the value of the objective function at point u^s, and the value $c_{\max}$ characterizes the constraint violations,

$$c_{\max} = \max\left\{500 \sum_{t=0}^{T-1} \max\{-0.2 - u_t;\, u_t - 0.2;\, 0\};\; 100|y_T|;\right.$$
$$\left. 500 \sum_{t=1}^{T} \max\{-1 - y_t;\, 0\}\right\}.$$

The optimal objective function value is $f(u^*) = 0.121284 \cdot 10^4$. We see from Table 1 that 300 iterations of the algorithm were sufficient to obtain a good approximation of the extremum point: the performance index $F(u)$ was within 0.1 percent of its final value, and the maximal constraint violation

Table 1. Calculation results for Algorithm 7.2.

s	s_f	ρ	$F(u^s)$	$c_{\max}$
1	1	0.10×10^{0}	0.317054×10^4	0.13×10^4
50	86	0.63×10^{-2}	0.124404×10^4	0.44×10^{0}
100	179	0.13×10^{-2}	0.123318×10^4	0.57×10^{-1}
300	568	0.14×10^{-3}	0.121960×10^4	0.73×10^{-2}
600	1171	0.34×10^{-3}	0.121796×10^4	0.60×10^{-2}
1000	1949	0.25×10^{-5}	0.121543×10^4	0.47×10^{-5}
2000	3967	0.14×10^{-3}	0.121472×10^4	0.15×10^{-2}
3000	5962	0.29×10^{-4}	0.121439×10^4	0.15×10^{-3}
4000	7968	0.17×10^{-3}	0.121362×10^4	0.37×10^{-3}
5000	9924	0.97×10^{-8}	0.121340×10^4	0.95×10^{-7}
6000	11956	0.12×10^{-4}	0.121338×10^4	0.21×10^{-3}
7000	13952	0.50×10^{-5}	0.121336×10^4	0.78×10^{-4}
8000	15949	0.25×10^{-5}	0.121332×10^4	0.13×10^{-4}
9000	17950	0.32×10^{-5}	0.121325×10^4	0.19×10^{-4}
10000	19945	0.67×10^{-6}	0.121322×10^4	0.74×10^{-5}
11000	21967	0.91×10^{-4}	0.121303×10^4	0.50×10^{-4}
12000	23944	0.53×10^{-6}	0.121284×10^4	0.81×10^{-6}
13000	25928	0.15×10^{-7}	0.121284×10^4	0.68×10^{-8}
14000	27923	0.49×10^{-8}	0.121284×10^4	0.44×10^{-7}

Table 2. Calculation results for Algorithm 7.1.

s	s_f	ρ	$F(u^s)$	$c_{\max}$
1	1	0.10×10^{0}	0.317054×10^4	0.13×10^4
50	83	0.12×10^{-2}	0.124521×10^4	0.79×10^{0}
100	171	0.39×10^{-3}	0.123207×10^4	0.49×10^{-1}
300	558	0.33×10^{-4}	0.121890×10^4	0.14×10^{-2}
600	1101	0.34×10^{-6}	0.121818×10^4	0.26×10^{-4}
1000	1844	0.14×10^{-5}	0.121666×10^4	0.10×10^{-3}
2000	3780	0.26×10^{-4}	0.121457×10^4	0.49×10^{-3}
3000	5714	0.61×10^{-5}	0.121411×10^4	0.28×10^{-3}
4000	7618	0.61×10^{-4}	0.121394×10^4	0.23×10^{-3}
5000	9487	0.13×10^{-7}	0.121385×10^4	0.10×10^{-5}
6000	11406	0.45×10^{-5}	0.121382×10^4	0.72×10^{-4}
7000	13308	0.36×10^{-5}	0.121378×10^4	0.47×10^{-4}
8000	15226	0.55×10^{-5}	0.121371×10^4	0.72×10^{-5}
9000	17147	0.29×10^{-6}	0.121351×10^4	0.39×10^{-5}
10000	19043	0.33×10^{-5}	0.121348×10^4	0.14×10^{-4}
11000	20976	0.10×10^{-6}	0.121346×10^4	0.48×10^{-5}
12000	22904	0.17×10^{-6}	0.121344×10^4	0.14×10^{-5}
13000	24840	0.17×10^{-6}	0.121343×10^4	0.17×10^{-6}
14000	26770	0.33×10^{-6}	0.121326×10^4	0.67×10^{-5}

was less than 10^{-4}. For Algorithm 7.1, the analogous table is Table 2. Algorithm 7.1 demonstrates practically the same rate of convergence as Algorithm 7.2 in the beginning of the optimization process. In the vicinity of the extremum, Algorithm 7.1 converges slower than Algorithm 7.2.

The results of the numerical experiments show that, for the objective functions considered, Algorithm 7.1 is better for low-dimensional problems and Algorithm 7.2 is better for high-dimensional ones.

References

1. Eaves, B. C., and Zangwill, W., *Generalized Cutting Plane Algorithms*, SIAM Journal on Control, Vol. 9, pp. 529–542, 1971.
2. Bertsekas, D. P., and Mitter, S. K., *A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals*, SIAM Journal on Control, Vol. 11, pp. 637–652, 1973.
3. Wolfe, P., *A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions*, Mathematical Programming Study, Vol. 3, pp. 145–173, 1975.
4. Lemarechal, C., Strodiat, J. J., and Bihain, A., *On a Bundle Algorithm for Nonsmooth Optimization*, Nonlinear Programming, Vol. 4, Edited by O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York, New York, pp. 245–282, 1981.

5. Mifflin, R., *A Modification and an Extension of Lemarechal's Algorithm for Nonsmooth Minimization*, Mathematical Programming Study, Vol. 17, pp. 77–90, 1982.
6. Kiwiel, K. C., *Methods of Descent for Nondifferentiable Optimization*, Springer-Verlag, Berlin, Germany, 1985.
7. Dem'janov, V. F., and Vasil'ev, L. V., *Nondifferentiable Optimization*, Springer, New York, New York, 1985.
8. Robinson, S. M., *Newton's Method for a Class of Nonsmooth Functions*, Preprint, Department of Industrial Engineering, University of Wisconsin–Madison, Madison, Wisconsin, 1988.
9. Shor, N. Z., *On the Structure of Algorithms for the Numerical Solution of Optimal Planning and Designing*, PhD Thesis, V. Glushkov Institute of Cybernetics, Kiev, USSR, 1964 (in Russian).
10. Dennis, J. M., and Moré, J. J., *Quasi-Newton Methods: Motivation and Theory*, SIAM Review, Vol. 19, pp. 46–89, 1977.
11. Shor, N. Z., *Minimization Methods for Nondifferentiable Functions*, Springer-Verlag, Berlin, Germany, 1985.
12. Poljak, B. T., *Subgradient Methods: A Survey of Soviet Research*, Proceedings of the IIASA Workshop on Nonsmooth Optimization, Edited by C. Lemarechal and R. Mifflin, Sopron, Hungary, pp. 5–29, 1977.
13. Uryas'ev, S. P., *Stochastic Quasigradient Algorithms with Adaptively Controlled Parameters*, Working Paper WP-86-32, IIASA, Laxenburg, Austria, 1986.
14. Ermoliev, Yu., and Wets, R. J. B., Editors, *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, Berlin, Germany, 1988.
15. Uryas'ev, S. P., *Adaptive Variable-Metric Algorithms for Nonsmooth Optimization Problems*, Working Paper WP-88-60, IIASA, Laxenburg, Austria, 1988.
16. Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
17. Pshenychnyi, B. N., *Necessary Conditions for an Extremum*, Marcel Dekker, New York, New York, 1971.
18. Nesterov, Yu, E., *Minimization Methods for Nonsmooth Convex and Quasiconvex Functions*, Ekonomika i Matematicheskie Metody, USSR, Vol. 20, pp. 519–531, 1984 (in Russian).
19. Nurminski, E., *Numerical Methods for Solving Deterministic and Stochastic Minimax Problems*, Naukova Dumka, Kiev, USSR, 1979 (in Russian).
20. Hoffman, A., *Weak Convex Functions, Multifunctions, and Optimization*, 27th Internationales Wissenschaftliches Kolloquium der Technische Hochschule Ilmenau, Vol. 5, pp. 33–36, 1982.
21. Ritch, P. S., *Discrete Optimal Control with Multiple Constraints, I: Constraint Separation and Transformation Technique*, Automatica, Vol. 9, p. 415–429, 1973.
22. Outrata, J. V., and Schindler, Z., *On Some Nondifferentiable Problems in Optimal Control*, Proceedings of the IIASA Workshop on Nondifferentiable Optimization, Edited by V. F. Demyanov and D. Pallaschke, Sopron, Hungary, pp. 118–128, 1984.
23. Ermoliev, Yu. M., Lyashko, I. I., Mikhalevich, V. S., and Tyuptya, V. I., *Mathematical Methods of Operation Research*, Visha Shkola, Kiev, USSR, 1979 (in Russian).

JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 71, No. 2, NOVEMBER 1991

TECHNICAL NOTE

Optimal Investment in the Reclamation of Eutrophic Water Bodies[1]

M. Gatto,[2] L. Ghezzi,[3] and S. Rinaldi[4]

Communicated by G. Leitmann

Abstract. Eutrophication, i.e., the abnormal growth of phytoplankton, is considered in this note, which focuses on the optimal treatment of eutrophic water bodies. The issue is addressed by the use of a nonlinear model where phytoplankton and the number of wastewater treatment plants in operation are the state variables. The decision maker is a governmental agency which has to define the time pattern of investment in new plants so as to minimize the present value of environmental and treatment costs. The optimal solution is shown to have the following features. First, the optimal size for the wastewater treatment system is attained in minimum time. Subsequently, investment replaces wornout treatment plants, and phytoplankton adjusts asymptotically to its optimal equilibrium value.

Key Words. Eutrophication, wastewater treatment, optimal reclamation, Pontryagin's principle, singular arcs.

1. Introduction

A water body (sea, lake, or artificial reservoir) is called eutrophic when it is rich in nutrients (nitrogen and phosphorus) and phytoplankton (microscopic algae). Although eutrophication is to some extent a slow and unavoidable natural process, it is very frequently induced or accelerated by man.

[1]This work was supported by Centro Teoria dei Sistemi—CNR, Milano and by Fondazione ENI E. Mattei.
[2]Professor, Dipartimento di Elettronica, Politecnico di Milano, Milano, Italy.
[3]Research Fellow, Dipartimento di Elettronica, Politecnico di Milano, Milano, Italy.
[4]Professor, Dipartimento di Elettronica, Politecnico di Milano, Milano, Italy.