U.S.S.R. Comput.Maths.Math.Phys.,Vol.25,No.2,pp.21-30,1985
Printed in Great Britain

0041-5553/85 $10.00+0.00
Pergamon Journals Ltd.

OPTIMAL METHODS OF SMOOTH CONVEX MINIMIZATION*

A.S. NEMIROVSKII and YU.E. NESTEROV

Convergence-rate-optimal methods are constructed for minimizing a smooth convex function with respect to type L_p Banach space, and the superposition of a convex scalar and convex vector-function with respect to a convex subset of Hilbert space.

Introduction.

We shall consider iterative methods of solving problems of the type

$$\min\{f(x)\,|\,x\in Q\}. \tag{1}$$

where $f(x)$ is a convex function in Banach space E, and Q is a closed convex subset of E. We assume that either f itself is smooth (or strongly convex), or that a method is known a priori for "assembling" it from smooth functions (e.g., $f=\max_{1\le i\le n} f_i$).

There is a vast literature on such problems, at least in the elementary case when E is Hilbert, $Q=E$, and f is strongly convex; and many numerical methods of solution have been proposed. Below we extend still further the stock of methods; this is justified, since the convergence rate of the new methods is essentially unimprovable, which is more than can be claimed for most of the old methods.

Let us explain the essential unimprovability of the new methods as applied to the most typical situation. Let E be Hilbert, $Q=E$, and all we know about f is that it is twice continuously differentiable and the spectrum of the Hessian of f at any point of E belongs to the interval $[1/\Lambda, 1]$ (i.e., f is strongly convex with index of conditionality Λ), while the normalization condition $f(0)-\min f\le 1$ holds. When solving problem (1), f and f' can be computed at (any) points of E; we wish to find the approximate solution of (1) with a given error (with respect to the function) $\varepsilon>0$. In accordance with the normalization condition, only the case $\varepsilon<1$ is non-trivial.

Assume that we have a method of solving problems of the family described. Denote by $M(\varepsilon)$ the least number of computations of f and f', with which the method guarantees that we find the approximate solution of the problem of minimizing f with an error (with respect to the function) not greater than ε, no matter what the function f of our family. We know /1/ that, with $\dim E=\infty$ (or, if convenient, uniformly with respect to E with $\dim E<\infty$), for any method $M(\varepsilon)\ge c\Lambda^{1/2}\ln\varepsilon^{-1}$, $\varepsilon<1$, $\Lambda\ge 2$; here and below, c is any positive absolute constant. For the conventional methods of smooth convex minimization, the respective characteristics are worse in the sense of the order of dependence on conditionality Λ. For instance, for the gradient method (both with constant step, and with total relaxation) $M(\varepsilon)\ge c\Lambda\ln\varepsilon^{-1}$ even with $E=R^2$. For some versions of the supposedly very efficient method of conjugate gradients, the lower bound of $M(\varepsilon)$ (uniformly with respect to $\dim E<\infty$) is no better than in the gradient method /1/; for other versions of the conjugate gradients method, the question of the dependence of $M(\varepsilon)$ on Λ remains open; but as far as we know, there is none for which it has been proved that $M(\varepsilon)$ is less than for the gradient method.

Methods for which $M(\varepsilon)\le c\Lambda^{1/2}\ln\Lambda\ln\varepsilon^{-1}$ were constructed in /1, 2/; the superfluous factor $\ln\Lambda$, compared with the lower bound, was the cost of solving at each step auxiliary problems of two-dimensional (in /1/) or one-dimensional (in /2/) minimization. Finally, a method with $M(\varepsilon)\le c\Lambda^{1/2}\ln\varepsilon^{-1}$ was constructed in /3/; this was shown to be unimprovable (uniformly with respect to $\dim E$) by more than an absolutely constant number of times.

In the present paper we extend the method described in /3/ with the aid of the devices of /4/ to a wider class of problems while preserving (under quite general assumptions about the situation) the property of essential unimprovability referred to above.

Our basic notation below is as follows: R^1 is the real axis; E and $\|\cdot\|$ are reflexive real Banach space and its norm; E^* and $\|\cdot\|_*$ are the adjoint space and dual norm; $\langle\varphi, x\rangle$ is the value of $\varphi\in E^*$ at $x\in E$. For $\varphi\in E^*$, we denote by $e(\varphi)$ the unit vector of E such that $\langle\varphi, e(\varphi)\rangle=\|\varphi\|_*$.

All the problems (1) considered will be assumed solvable; we denote the set of solutions of a problem by X^*, and the optimal value of its target functional by f^*. With $x\in E$, we denote $\inf\{\|x-y\|\,|\,y\in X^*\}$ by $\rho(x)$.

We denote by $C^{1,\nu}(E)$ the class of all continuously differentiable functions $f:E\to R^1$, for which there exists the constant $L_f<\infty$ such that $\|f'(x)-f'(y)\|_*\le L_f\|x-y\|^\nu$ for all x, y of E. We denote the minimum of these L_f by $L(f)$. Throughout, $0<\nu\le 1$, $\sigma=1+\nu$. Clearly, with $f\in C^{1,\nu}(E)$ for all $x, y\in E$

$$f(x+y)\le f(x)+\langle f'(x), y\rangle+\sigma^{-1}L(f)\|y\|^\sigma. \tag{2}$$

---

*Zh. vychisl. Mat. mat. Fiz., 25, 3, 356-369, 1985

We easily see from (2) that, with $\lambda \geq L(f)$ and $f \in C^{1,\nu}(E)$, for all $x \in E$,

$$f(x-\lambda^{-1/\nu}\|f'(x)\|_*^{1/\nu}e(f'(x))) \leq f(x)-(\nu/\sigma)\lambda^{-1/\nu}\|f'(x)\|_*^{\sigma/\nu}. \tag{3}$$

1. Unconditional minimization in Banach space.

1. We construct a method for solving problem (1) with $Q=E$ and with a function $f \in C^{1,\nu}(E)$, convex in E. We assume that the space E is r-smooth, i.e., the function $V(\varphi)=r^{-1}\|\varphi\|_*^r \in C^{1,r-1}(E^*)$, where $r \in (1,2]$ is a parameter.

Elementary examples of r-smooth spaces are the space $L_p(\Omega, \mu)$ of p-th degree summable real functions in measurable space Ω with σ-finite σ-additive measure μ, furnished with natural norms. For $1<p<\infty$, space $(L_p, \|\cdot\|_p)$ is r-smooth with

$$r=\min\{2, p/(p-1)\}, \qquad c_{L_p}=\begin{cases} 5, & p>2, \\ 4/(p-1), & 1<p\leq 2. \end{cases} \tag{1.1}$$

We can prove (1.1) by starting e.g., from /1, Sect.5, Chapter 3/.

Notice that, since the space E is reflexive, the mapping $V'(\cdot)$, i.e., the derivative of the function $V(\cdot)$, can be regarded as a mapping from E^* into E.

2. Let us describe the method $\mathfrak{M}(\bar{x}, \bar{R}, \bar{L}_0, N)$ of solving problem (1) with $Q=E$. It depends on the parameters: the initial approximation $\bar{x} \in E$, the real numbers $\bar{R}>0$ and $\bar{L}_0>0$, which are the a priori estimates respectively of $\rho(\bar{x})$ and $L(f)$, and on the number (integer) N of iterations, used in the method. The parameter ν, corresponding to the smoothness of f, is assumed given a priori.

The method $\mathfrak{M}(\cdot,\cdot,\cdot,\cdot)$ is constructed as follows.

Step 0. Initiation. We put $x_1=\bar{x}$,

$$\alpha=c_E^{-\sigma/r}\left[\frac{\sigma(r-1)}{r+\sigma-r\sigma}\bar{R}^{r/(r-1)}N^{-1}\right]^{(r+\sigma-r\sigma)/r}$$

(if $r+\sigma-r\sigma=0$, or what amounts to the same thing, $r=\sigma=2$, then $\alpha=1/c_E$).

Step 1. The k-th iteration, $k=1, 2, \ldots, N$. If $f'(x_k)=0$, stop: the exact solution is found. Otherwise:

a) we compute by successive inspection the least number $i=i_k \geq 0$, for which the inequality

$$f(x_k)-f\left(x_k-\frac{t_k^{1/\nu}e(f'(x_k))}{L^{1/\nu}}\right) \geq \frac{\nu}{\sigma}\frac{t_k^{\sigma/\nu}}{L^{1/\nu}} \tag{1.2}$$

holds for $L=2^iL_{k-1}$; here and henceforth $t_k=\|f'(x_k)\|_*$;

b) we put

$$L_k=2^{i_k}L_{k-1}, \qquad y_k=x_k-\frac{t_k^{1/\nu}e(f'(x_k))}{L_k^{1/\nu}}, \tag{1.3a}$$

$$p_k=y_k-x_1+V'\left(\sum_{j=1}^{k}\frac{\alpha}{L_j}\left(\frac{j}{\sigma}\right)^{\nu}f'(x_j)\right), \qquad x_{k+1}=y_k-\frac{\sigma}{k+1}p_k. \tag{1.3b}$$

The iteration is ended.

The result $\bar{y}$ of applying method $\mathfrak{M}(\bar{x}, \bar{R}, \bar{L}_0, N)$ to problem (1) is either the exact solution (if it is found at one of the first N iterations), or y_N.

Notes. 1. In view of (3), inequality (1.2) holds for all $L \geq L(f)$. It follows from this that, first, the method of finding i_k in method $\mathfrak{M}(\cdot,\cdot,\cdot,\cdot)$ is correct, and second, that

$$L_k \leq \bar{L}(f)=\max\{\bar{L}_0, 2L(f)\} \qquad \forall k=1, 2, \ldots, N. \tag{1.4}$$

2. With $\sigma=r=2$ (i.e., when minimizing a convex function with Lipschitz gradient in a space as smooth as Hilbert), the parameters $\bar{R}$ and N have no influence on the trajectory of the method (N only limits the number of iterations).

3. If we know a constant $L>0$ such that $L \geq L(f)$, then step 1 of method $\mathfrak{M}(\cdot,\cdot,\cdot,\cdot)$ can be replaced by the following.

Step 1'. The k-th iteration, $k=1, 2, \ldots, N$. If $f'(x_k)=0$, stop. Otherwise we put

$$y_k=x_k-\frac{t_k^{1/\nu}e(f'(x_k))}{L^{1/\nu}},$$

$$p_k=y_k-x_1+V'\left(\sum_{j=1}^{k}\frac{\alpha}{L}\left(\frac{j}{\sigma}\right)^{\nu}f'(x_j)\right), \quad x_{k+1}=y_k-\frac{\sigma}{k+1}p_k.$$

The iteration ends.

3. The possibilities of the above method are characterized by:

Theorem 1. Under the assumptions of Para.1, for the result $\bar{y}$ of applying method $\mathfrak{M}(\bar{x}, \bar{R}, \bar{L}_0, N)$ to problem (1), we have:

1) with $\sigma+r<4$:

$$f(\bar{y})-f_*\leqslant c_1\bar{L}(f)\bar{R}^\sigma[1+\Gamma(\bar{R})]N^{-\omega}, \tag{1.5}$$

where $\omega=(2r\sigma-r-\sigma)/r$, $\Gamma(\bar{R})=[\rho(\bar{x})/\bar{R}]^{r/(r-1)}$,

$$c_1=c_E^{\sigma/r}\frac{\sigma^\sigma(r-1)}{r}\left[\frac{r+\sigma-r\sigma}{\sigma(r-1)}\right]^{(r+\sigma-\sigma r)/r}$$

and $\bar{L}(f)$ is given by (1.4);

2) if $\sigma=r=2$, then

$$f(\bar{y})-f^*\leqslant 2c_E\bar{L}(f)\rho^2(\bar{x})N^{-2}; \tag{1.5'}$$

3) after N iterations of the method $\mathfrak{M}(\bar{x}, \bar{R}, \bar{L}_0, N)$, $f'(x)$ is evaluated at most N times, and $f(x)$ at most $N_f=2N+\lfloor\log_2(\bar{L}(f)/\bar{L}_0)\rfloor$ times.

Proof. Let x^* be the point of X^* nearest to $\bar{x}$, and $V^*(\varphi)=V(\varphi)+\langle\varphi, x_1-x^*\rangle$. We put

$$b_k=\alpha\left(\frac{k}{\sigma}\right)^\nu,\quad a_k=\frac{kb_k}{\sigma},\quad v_k=V^*\left(\sum_{j=1}^k\frac{b_j}{L_j}f'(x_j)\right),$$

$$\Delta_k=f(x_k)-f(y_k).$$

It can be assumed without loss of generality that $f^*=0$ and $\bar{y}=y_N$.

Applying (2) to the function $V^*(\cdot)$, we obtain

$$v_{k+1}\leqslant v_k+\frac{b_{k+1}}{L_{k+1}}\langle f'(x_{k+1}), x^*-x_1\rangle+\frac{c_E}{r}\left(\frac{b_{k+1}}{L_{k+1}}\right)^r t_{k+1}^r+$$

$$\frac{b_{k+1}}{L_{k+1}}\left\langle f'(x_{k+1}), V'\left(\sum_{j=1}^k\frac{b_j}{L_j}f'(x_j)\right)\right\rangle.$$

It follows from (1.3) that

$$V'\left(\sum_{j=1}^k\frac{b_j}{L_j}f'(x_j)\right)=x_1-x_{k+1}+\left(\frac{k+1}{\sigma}-1\right)(y_k-x_{k+1}).$$

Hence

$$v_{k+1}\leqslant v_k+\frac{b_{k+1}}{L_{k+1}}\Big[\langle f'(x_{k+1}), x^*-x_{k+1}\rangle+ \tag{1.6}$$

$$\left(\frac{k+1}{\sigma}-1\right)\langle f'(x_{k+1}), y_k-x_{k+1}\rangle\Big]+\frac{c_E}{r}\left(\frac{b_{k+1}}{L_{k+1}}\right)^r t_{k+1}^r.$$

Notice that $0\leqslant a_{k+1}-b_{k+1}\leqslant a_k$, $L_{k+1}\geqslant L_k$ for all $k\geqslant1$. Moreover, (1.2) gives an upper bound of t_{k+1} in terms of Δ_{k+1}. Using consecutively the convexity of f, the relation $f^*=0$, and (1.2), we obtain from (1.6):

$$v_{k+1}\leqslant v_k+\frac{b_{k+1}}{L_{k+1}}\left\{-f(x_{k+1})+\left(\frac{a_{k+1}}{b_{k+1}}-1\right)[f(y_k)-f(x_{k+1})]\right\}+$$

$$c_Er^{-1}\left(\frac{b_{k+1}}{L_{k+1}}\right)^r t_{k+1}^r\leqslant v_k-\frac{a_{k+1}}{L_{k+1}}f(x_{k+1})+$$

$$\frac{a_{k+1}-b_{k+1}}{L_{k+1}}f(y_k)+\frac{c_E}{r}\left(\frac{\sigma}{\nu}L_{k+1}^{1/\nu}\Delta_{k+1}\right)^{r\nu/\sigma}\left(\frac{b_{k+1}}{L_{k+1}}\right)^r.$$

Noting that $f(y_k)\geqslant0$, $(a_{k+1}-b_{k+1})/L_{k+1}\leqslant a_k/L_k$, $f(x_{k+1})=f(y_{k+1})+\Delta_{k+1}$, we obtain from the last relation

$$v_{k+1}\leqslant v_k+\frac{a_k}{L_k}f(y_k)-\frac{a_{k+1}}{L_{k+1}}f(y_{k+1})+$$

$$\left[\frac{c_E}{r}\left(\frac{b_{k+1}}{L_{k+1}}\right)^r\left(\frac{\sigma}{\nu}L_{k+1}^{1/\nu}\Delta_{k+1}\right)^{r\nu/\sigma}-\frac{a_{k+1}}{L_{k+1}}\Delta_{k+1}\right]=$$

$$v_k+\frac{a_k}{L_k}f(y_k)-\frac{a_{k+1}}{L_{k+1}}f(y_{k+1})-\left\{\frac{c_E}{r}b_{k+1}^r\left(\frac{\sigma}{\nu}\right)^{r\nu/\sigma}\left(\frac{\Delta_{k+1}}{L_{k+1}}\right)^{r\nu/\sigma}-a_{k+1}\frac{\Delta_{k+1}}{L_{k+1}}\right\}.$$

We shall strengthen this relation by replacing the expression in braces by its maximum with respect to all $\Delta_{k+1}\geqslant0$. As a result, we have

$$v_{k+1}\leqslant v_k+\frac{a_kf(y_k)}{L_k}-\frac{a_{k+1}f(y_{k+1})}{L_{k+1}}+\frac{r-1}{r}\bar{R}^{r/(r-1)}N^{-1}. \tag{1.7}$$

We can show in exactly the same way that

$$v_1\leqslant-\frac{a_1f(y_1)}{L_1}+\frac{r-1}{r}\bar{R}^{r/(r-1)}N^{-1}.$$

On adding inequalities (1.7) for $k=1, 2, \dots, N-1$, and adding the previous inequality, we obtain

$$\frac{a_N f(y_N)}{L_N} \leqslant \frac{r-1}{r} \bar{R}^{r/(r-1)} - v_N, \text{ but } V^*(\varphi) \geqslant -\frac{r-1}{r} \rho^{r/(r-1)}(\bar{x})$$

(Hölder's inequality). Hence $-v_N \leqslant [(r-1)/r] \rho^{r/(r-1)}(x)$, and we arrive at

$$\frac{a_N}{L_N} f(y_N) \leqslant \frac{r-1}{r} [\bar{R}^{r/(r-1)} + \rho^{r/(r-1)}(\bar{x})].$$

Hence, using the expression for a_N and α, we obtain

$$f(y_N) \leqslant c_1 L_N \bar{R}^\sigma [1+\Gamma(\bar{R})] N^{-\omega}. \tag{1.8}$$

Recalling (1.4), we immediately obtain from (1.8) inequality (1.5).

Further, with $\sigma=r=2$, the trajectory of the method is independent of $\bar{R}$, and (1.8) admits passage to the limit as $\bar{R} \to +0$; hence we obtain inequality (1.5').

Let us prove Para. 3) of Theorem 1. In view of (1.4), we have $L_k = 2^{i_1+\ldots+i_k} \bar{L}_0 \leqslant \bar{L}(f)$. i.e., $i_1+\ldots+i_k \leqslant [\log_2(\bar{L}(f)/\bar{L}_0)]$. But, in method $\mathfrak{M}(\cdot, \cdot, \cdot, \cdot)$, $f'(\cdot)$ is evaluated once at the k-th iteration, while $f(\cdot)$ is evaluated $2+i_k$ times. These remarks prove Para. 3), and hence Theorem 1.

Note 4. If we know a number $L \geqslant L(f)$, then, on applying to function $f(\cdot)$ method $\mathfrak{M}(\cdot, \cdot, \cdot, \cdot)$ with step 1', we obtain the point $\bar{y}$ for which (1.5) (or (1.5')) holds. Notice that, in this method, the function $f(\cdot)$ need not be evaluated $(N_f=0)$.

4. Method $\mathfrak{M}(\cdot, \cdot, \cdot, \cdot)$ allows passage to the limit as $\nu \to +0$. In fact, let $(E, \|\cdot\|)$ be r-smooth as above, and let the function $f(\cdot)$ be convex and Lipschitz with constant $L(f)$. Let us describe the method $\overline{\mathfrak{M}}(\bar{x}, \bar{R}, N)$ of minimizing the function f (in the description we mean by $f'(x)$ any support functional of the function f at the point x).

Step 0. Initiation. We put $x_1=\bar{x}$, $\alpha = c_E^{1/r}[(r-1)\bar{R}^{r/(r-1)} N^{-1}]^{1/r}$.

Step 1. The k-th iteration, $k=1, 2, \ldots, N$. We put $L_k = \max\{\|f'(x_j)\|_*, 1 \leqslant j \leqslant k\}$,

$$x_{k+1} = x_k - \frac{1}{k+1}\left(x_k - x_1 + V'\left(\sum_{j=1}^{k} \frac{\alpha}{L_j} f'(x_j)\right)\right).$$

The iteration ends.

Notice that the construction of point x_N requires N evaluations of $f'(\cdot)$. On arguing in the same way as in the proof of Theorem 1, we can show that

$$f(x_N) - f^* \leqslant c_E^{1/r} L(f) \bar{R} [1+\Gamma(\bar{R})] N^{-(r-1)/r}.$$

If an a priori upper bound of $\rho(\bar{x})$ is taken as $\bar{R}$, this estimate is the same, up to a factor (dependent only on c_E and r), as in the method of mirror descent associated with $(E, \|\cdot\|)$, see /1, Chapter 3/, applied to the problem $\min\{f(x) \mid \|x-\bar{x}\| \leqslant \bar{R}\}$ (since $\bar{R} \geqslant \rho(\bar{x})$, the solution of the latter problem is in fact the solution of (1)).

5. Now let the function $f(\cdot)$ in problem (1) satisfy, along with the conditions of Para. 1, the further strict minimum condition

$$f(x) - f^* \geqslant \sigma^{-1} l(f) \rho^\tau(x) \tag{1.9}$$

for all x such that $f(x) \leqslant f(\hat{x})$. Here, $\tau \geqslant \sigma$, $l(f) > 0$, and $\hat{x}$ is an a priori given point of E.

In this case method $\mathfrak{M}(\cdot, \cdot, \cdot, \cdot)$ can be used in the "scheme with renewal" and we can arrange to improve the order of the convergence rate. The following is what we mean by the "scheme with renewal."

We specify a positive number $\bar{L}_0$, a sequence of integers $\{M_i\}_{i=0}^\infty$, and a sequence of positive numbers $\{\bar{R}_i\}_{i=0}^\infty$. We define a sequence of points $\{\bar{x}^{(i)}\}_{i=0}^\infty$ as follows: $\bar{x}^{(0)} = \hat{x}$, $\bar{x}^{(i+1)}$ is the result of applying to the function f method $\mathfrak{M}_i = \mathfrak{M}(\bar{x}^{(i)}, \bar{R}_i L^{(i)}, M_i)$, if this result is better with respect to the value of f, than $\bar{x}^{(i)}$; otherwise $\bar{x}^{(i+1)} = \bar{x}^{(i)}$. Here, $L^{(0)} = \bar{L}_0$, and $L^{(i)}$ is the estimate of $L(f)$, constructed at the last iteration of method $\mathfrak{M}_{i-1}$ ($L_{M_{i-1}}$ in the notation for describing method $\mathfrak{M}$), $i \geqslant 1$.

Notice that the construction of the point $\bar{x}^{(i)}$ in this scheme requires not more than $N_i = M_0+\ldots+M_{i-1}$ evaluations of $f'(\cdot)$ and not more than $2N_i + [\log_2(\bar{L}(f)/\bar{L}_0)]$ evaluations of f, i.e., as many as a single application of method $\mathfrak{M}(\cdot, \bar{L}_0, N_i)$. However, it turns out that, in situation (1.9), given a sensible choice of sequences $\{M_i\}_{i=0}^\infty$, $\{\bar{R}_i\}_{i=0}^\infty$, we can arrange to accelerate the convergence, i.e., obtain a faster decrease of the error $f(\bar{x}^{(i)}) - f^*$ in the function of N_i than is prescribed by the right-hand side of (1.5) with $N=N_i$.

Generally speaking, different strategies of renewal are possible. We take the simplest of them.

Assume that the values of σ and τ are known. We denote by $c_i(d_i)$ the functions σ, r, c_E, positive and continuous in the domains $\{1<\sigma, r \leqslant 2, c_E > 0\}$, $\{1<\sigma, r\leqslant 2, c_E>0, \tau>\sigma\}$ and accordingly, σ, r, c_E, τ. Two cases are possible.

Case a. $\tau > \sigma$. We put $(i \geqslant 0)$

$$M_i = 2^i, \quad \bar{R}_i = M_i^{-\omega/(\tau-\sigma)} \bar{R}, \tag{1.10}$$

where $\bar{R} > 0$ is given a priori. It turns out that, with this renewal strategy and a regular choice of $\bar{R}$, for all $i \geqslant 0$ we have the inequality

$$f(\bar{x}^{(i)})-f^*\leqslant aN_i^{-\omega\tau/(\tau-\sigma)}. \tag{1.11}$$

where $a=a(L(f), l(f), \sigma, \tau, r, c_E, \bar{R})$ is an explicitly written function. The construct on $\bar{R}$ is necessary only when $\tau\leqslant r/(r-1)$ and we have

$$\bar{R}\geqslant d_1[L(f)/l(f)]^{1/(\tau-\sigma)}, \tag{1.12}$$

where d_1 is an explicitly expressible function. Notice that, if we choose $\bar{R}=d_1(L/l)^{1/(\tau-\sigma)}$ with $L\geqslant L(f)$, $l\leqslant l(f)$, then, regardless of the ratio of τ to $r/(r-1)$, the factor a in (1.11) will have the form $d_2(L^\tau/l^\sigma)^{1/(\tau-\sigma)}$.

If $\sigma=r=2$, then regardless of the choice of $\bar{R}$, which now has no effect on the trajectory of the method, we have

$$f(\bar{x}^{(i)})-f^*\leqslant d_3[L^\tau(f)/l^2(f)]^{1/(\tau-2)}N_i^{-2\tau/(\tau-2)},$$

where the expression for d_3 can also be written explicitly.

Thus, the order of the convergence rate is improved by a factor $\tau/(\tau-\sigma)$ as a result of renewal.

Case b. $\tau=\sigma$. Here the renewal enables a geometric instead of a power convergence rate to be obtained. Assume that estimates $L\geqslant L(f)$ and $l\leqslant l(f), l>0$, are known a priori. We put

$$\bar{F}_0=\frac{1}{2}\left(\frac{\sigma}{l}\right)^{1/\nu}\|f'(\hat{x})\|_*^{\sigma/\nu}, \qquad \bar{R}_0=(2\sigma\bar{F}_0/l)^{1/\sigma}, \tag{1.13a}$$

$$M_i=M=]\,(4l)^{1/\omega}(4\sigma^\sigma C_E^{\sigma/r})^{1/\omega}[, \qquad \bar{R}_{i+1}=2^{-1/\sigma}\bar{R}_i, \qquad L_0=L. \tag{1.13b}$$

It turns out that, with renewal strategy (1.13), with $\sigma+r<4$, for all $i\geqslant0$ we have $f(\bar{x}^{(i)})-f^*\leqslant2^{-i}\bar{F}_0=2^{-N_i/M}\bar{F}_0$. With $\sigma=r=2$ we can take

$$L_0=L, \qquad M_i=M=]\,2(2c_EL/l)^{1/2}[\tag{1.14}$$

(the choice of $\bar{R}_i$ does not matter) and we can obtain, for all $i\geqslant1$,

$$f(\bar{x}^{(i)})-f^*\leqslant[f(\bar{x}^{(i-1)})-f^*]/2\leqslant2^{-N_i/M}[f(\hat{x})-f^*]. \tag{1.15}$$

Notice that the commonly considered problems of unconstrained minimization of a strongly convex function in Hilbert space reduce to the situation $\tau=\sigma=r=2$. As applied to this case, the method $\mathfrak{M}(\cdot,\cdot,\cdot,\cdot)$ with renewal (1.14) is similar to that described in /3/.

To conclude this section, we note the proofs of the results mentioned. They are all based on the same scheme. We shall confine ourselves to renewal (1.11) with $\sigma+r<4$.

Put $\varkappa=\sigma/\tau$, $\mu_i=f(\bar{x}^{(i)})-f^*$. Then, by (1.5) and (1.9), we have

$$\begin{aligned}\mu_{i+1}&\leqslant c_1L(f)\bar{R}_i^\sigma\left\{1+\left[\frac{\rho(\bar{x}^{(i)})}{\bar{R}_i}\right]^{r/(r-1)}\right\}M_i^{-\omega}=\\ &c_1L(f)\bar{R}^\sigma M_i^{-\omega/(1-\varkappa)}\left\{1+\left[\frac{\rho(\bar{x}^{(i)})}{\bar{R}_i}\right]^{r/(r-1)}\right\}\leqslant\\ &c_1L(f)\bar{R}^\sigma M_i^{-\omega/(1-\varkappa)}\left[1+\left(\frac{\sigma\mu_i}{l\bar{R}_i^\tau}\right)^{r/[\tau(r-1)]}\right]=\\ &c_1L(f)\bar{R}^\sigma\left(\frac{M_{i+1}}{2}\right)^{-\omega/(1-\varkappa)}\left[1+\left(\frac{\sigma\mu_i}{l\bar{R}^\tau}M_i^{\omega/(1-\varkappa)}\right)^{r/[\tau(r-1)]}\right].\end{aligned}$$

Put $F_i=\mu_iM_i^{\omega/(1-\varkappa)}$. Then,

$$F_{i+1}\leqslant b_1+b_2F_i^{r/[\tau(r-1)]}=G_{\bar{R}}(F_i), \tag{1.16}$$

where the constants b_1 and b_2 can be seen from the previous inequality. Further, by (1.9) and (2), $\sigma^{-1}l(f)\rho^\tau(\hat{x})\leqslant\sigma^{-1}L(f)\rho^\sigma(\hat{x})$, which gives the inequality

$$F_0=\mu_0\leqslant\frac{L(f)}{\sigma}\rho^\sigma(\hat{x})\leqslant\frac{L(f)}{\sigma}\left[\frac{L(f)}{l(f)}\right]^{\sigma/(\tau-\sigma)}=\bar{F}. \tag{1.17}$$

If now $\tau>r/(r-1)$, then, from recurrence inequality (1.16) and (1.17), we see that the sequence $\{F_i\}_{i=0}^\infty$ is upper-bounded by a constant. This implies (in the light of $M_i\geqslant N_i/2$) inequality (1.11).

Assume that $\tau\leqslant r/(r-1)$. In this case we easily see that $G_{\bar{R}}(F)\leqslant F$ for all F such that

$$\bar{R}\geqslant d_1\max\left\{\left[\frac{F}{L(f)}\right]^{1/\sigma}, \ \frac{L(f)}{F}\left[\frac{F}{l(f)}\right]^{r/[\tau(r-1)]}\right\}.$$

In particular, the condition $G_{\bar{R}}(\bar{F})\leqslant\bar{F}$ holds for all $\bar{R}$ satisfying (1.12) with a suitable constant d_1. With these $\bar{R}$, by (1.16), (1.17), and the relation $G_{\bar{R}}(\bar{F})\leqslant\bar{F}$, we have $F_i\leqslant\bar{F}$ for all i, which again leads to (1.11). It follows from the same arguments that, with $\bar{R}=d_1(4l)^{1/(\tau-\sigma)}$, $L\geqslant L(f)$, $l\leqslant l(f)$, we have $a=d_2(L^\tau/l^\sigma)^{1/(\tau-\sigma)}$.

6. Finally, let us state the results on the essential unimprovability of our methods. We fix an r-smooth space $(E,\|\cdot\|)$, a point $\bar{x}\in E$, and numbers $\bar{R}>0$, $L>0$, $\sigma\in(1,2]$, $\tau\geqslant\sigma$. We introduce the families of problems $\mathcal{F}_{\bar{x},\sigma}^{L,\bar{R}}(E)$, $\mathcal{F}_{\bar{x},\sigma,\tau}^{L,l}(E)$. The first family is formed by all

problems (1) with $Q=E$ and convex minimizing functions f of class $C^{1,\sigma-1}(E)$ such that $L(f)\leqslant L$ and $\rho(\bar{x})\leqslant \bar{R}$. The second is formed by all the convex solvable problems (1) with $Q=E$ and functions f such that $f\in C^{1,\sigma-1}(E)$, satisfies (1.9) for $x=\bar{x}$, $l\leqslant l(f)$, and $L(f)\geqslant L$. With $\sigma=\tau$ we also impose on f the normalization condition $f(\bar{x})-f^*\leqslant 1$. We supply the families with a local determinate oracle, which certainly communicates f and f', and a normalizing mapping $r(f)=1$ (for the terminology used in this section, see /1, Chapter 1/).

To solve problems of class $\mathscr{F}^{L,\bar{R}}_{\bar{x},\sigma}(E)$, the method $\mathfrak{M}(\bar{x}, \bar{R}, L, N)$ with step 1′ can be used (see Note 3). By (1.5) and (1.5′), the labour used by the method to obtain an error $\varepsilon>0$ is

$$M=M(\varepsilon)=]\,c_2(L\bar{R}^\sigma/\varepsilon)^{1/\omega}[. \tag{1.18}$$

In other words, if we choose $N=M(\varepsilon)$, the method solves any problem of the class with an error with respect to the functional not exceeding ε, and then requires not more than $M(\varepsilon)$ evaluations of f and f'.

To solve problems of class $\mathscr{F}^{L,l}_{\bar{x},\sigma,\tau}(E)$, the method with renewal of Para.5 can be used, namely: with renewal (1.13) for $\tau=\sigma$, and with renewal (1.10) when $\tau>\sigma$; in the latter case, $\bar{R}=d_1(4l)^{1/(\tau-\sigma)}$ (cf. (1.12)), and in all cases $L_0=L$. From which was said in Para.5, the labour of solving problems of this case by this method with error $\varepsilon>0$ is not greater than

$$M_\tau(\varepsilon)=]\,d_5(L^\tau/l^\sigma)^{1/(\omega\tau)}\varepsilon^{-(\tau-\sigma)/(\omega\tau)}[\tag{1.19}$$

(notice that $d_5\to\infty$ as $\tau\to\sigma+0$) and

$$\bar{M}_\sigma(\varepsilon)=]\,c_3(L/l)^{1/\omega}\ln_+\varepsilon^{-1}[,\qquad \tau=\sigma. \tag{1.19′}$$

Now let $\sigma, r\in(1,2]$ and $\tau\geqslant\sigma$ be given; we put $p=r/(r-1)$, and let $(E,\|\cdot\|)=(L_p(\Omega,\mu),\|\cdot\|_p)$, while $n\leqslant\infty$ is the linear dimensionality of E. By (1.1), $(E,\|\cdot\|)$ is r-smooth with $c_E=5$. It turns out that the complexity $N(\varepsilon)$ of class $\mathscr{F}^{L,\bar{R}}_{\bar{x},\sigma}(E)$ satisfies the estimate

$$N(\varepsilon)\geqslant c_4\min\{n,\ (\bar{R}L^\sigma/\varepsilon)^{1/\omega}\}, \tag{1.20}$$

while the complexity $N_\tau(\varepsilon)$ of class $\mathscr{F}^{L,l}_{\bar{x},\sigma,\tau}(E)$, with $\tau>\sigma$, satisfies

$$N_\tau(\varepsilon)\geqslant d_6\min\{n,\ (L^\tau/l^\sigma)^{1/(\omega\tau)}\varepsilon^{-(\tau-\sigma)/(\omega\tau)}\}. \tag{1.21}$$

Comparing (1.18) with (1.20), and (1.19) with (1.21), we see that, in these classes of problems (in the second, when $\tau>\sigma$), our methods are essentially unimprovable with respect to convergence rate (at least, in certain spaces E of the given smoothness). The complexity of class $\mathscr{F}^{L,l}_{\bar{x},\sigma,\tau}$ with $\sigma=\tau$ is studied only in the most interesting case of Hilbert space E, $\dim E=n\leqslant\infty$, and $\tau=\sigma=2$. The complexity $\bar{N}_2(\varepsilon)$ in these conditions has the bound

$$\bar{N}_2(\varepsilon)\geqslant c_5\min\{n,\ (L/l)^{1/2}\ln_+\varepsilon^{-1}\}, \tag{1.21′}$$

which, in association with (1.19′), again implies the essential unimprovability of the method.

Notice that proofs of lower bounds (1.20), (1.21), (1.21′) for the most important cases ($\sigma=r=2$ and $\sigma=\tau=r=2$) respectively) are given in /1, Chapter 7/.

2. Minimization of a composition of smooth convex functions in Hilbert space.

1. In this section we construct an iterative method for solving problem

$$\min\{f(x)=F(\bar{f}(x))\quad |x\in Q\}, \tag{2.1}$$

where Q is a convex closed subset of Hilbert space $(E,\|\cdot\|)$; $F(u)$: $R^m\to R^1$ is a convex function; $\bar{f}(x)=(f_1(x),\dots,f_m(x))$ is a set of convex functions of class $C^{1,1}(E)$. We assume that F is non-decreasing with respect to any coordinate u, to which corresponds the non-linear component $\bar{f}$, and that F is Lipschitz with respect to any such coordinate (with a constant independent of the other coordinates). Under these assumptoins, the function $f(x)$ which has to be minimized in Q is convex.

Elementary examples of such functions are

$$F(u)=u:\ R^1\to R^1,\qquad F(u)=\max\{u^{(i)}\mid 1\leqslant i\leqslant m\},$$

$$F(u)=\sum_{i=1}^m (u^{(i)})_+,\quad \text{where}\quad a_+=\max\{0,a\}.$$

Notice also that, if the functions F_1, F_2 have this property, then $F(\cdot)=\alpha F_1(\cdot)+\beta F_2(\cdot)$ likewise has the property for any $\alpha,\beta\geqslant 0$.

Consider the situation when, during the numerical solution of (2.1), only local information can be obtained about $\bar{f}$ (at each iteration, $\bar{f}$, $\bar{f}'$ can be evaluated at the point of E of interest); whereas F is given a priori. Accordingly, we assume that the solution of auxiliary problems, "non-local with respect to F" (see Sect.3) is possible. In short, though f may prove to be essentially unsmooth, it is built up in a known way from smooth convex functions. If we know this way, we can construct a method of minimizing f which is much more efficient than is possible for arbitrary convex functions.

2. We start with some preliminary constructions. Suppose $J=\{j\,|\,f_j$ is non-linear$\}$. $K=\{e\in R^m|e_i\geqslant 0\ \forall i,\ e_i=0\ \forall i\notin J\}$. For $e\in K$ we put $\bar{F}(e)=\sup\{[F(u+te)-F(u)]/t|u\in R^m,\ t>0\}$. By the assumptions of Para.1 about F being Lipschitz with respect to the variables u_i, $i\in J$, we have

$F(e)<\infty$ for any $e\in K$.

Put $\bar{L}(f)=(L(f_1),\dots,L(f_m))$; then $\bar{L}(f)\in K$. Let $A(f)=\bar{F}(\bar{L}(f))$; then $0\leqslant A(f)<\infty$.

For $A>0$ and $x, y\in E$, let

$$\bar{f}(x,y)=(f_1(x,y),\dots,f_m(x,y)),\qquad f_i(x,y)=f_i(x)+\langle f_i'(x),\ y-x\rangle,$$

$$\Phi(x,A,y)=F(\bar{f}(x,y))+0.5A\|x-y\|^2.$$

$$\Phi^*(x,A)=\min\{\Phi(x,A,y)\,|\,y\in Q\},$$

$$T(x,A)=\operatorname{argmin}\{\Phi(x,A,y)\,|\,y\in Q\},\qquad g(x,A)=x-T(x,A).$$

Notice that $g(\cdot,\cdot)$ is a generalization of the "gradient mapping" constructed in /1, Chapter 8/, as applied to problem (2.1) with $F(\bar{f}(x))=\max\{f_1(x),\dots,f_m(x)\}$. We shall require:

Lemma. For all $x\in E, y\in Q, A>0$, we have

$$\Phi^*(x,A)+A\langle g(x,A),\ y-x\rangle+0.5A\|g(x,A)\|^2\leqslant f(y). \tag{2.2}$$

In addition, for $A\geqslant A(f)$,

$$\Phi^*(x,A)\geqslant f(T(x,A)). \tag{2.3}$$

Proof. We fix $x\in E, A>0$ and put for brevity $T=T(x,A)$, $g=g(x,A)=x-T$, $\varphi(y)=F(\bar{f}(x,y))$. Then, T is the minimum point of a convex (defined throughout E) function $\varphi(y)+0.5A\|y-x\|^2$ with respect to $y\in Q$. Thus, for some $p\in\partial\varphi(T)$, we have $\langle\bar{p},\ y-T\rangle\geqslant 0$ for all $y\in Q$, where $\bar{p}=p+A(T-x)=p-Ag$. Thus,

$$\langle p,\ y-T\rangle\geqslant A\langle g,\ y-T\rangle\qquad \forall y \text{ of } Q. \tag{2.4}$$

Now, since $p\in\partial\varphi(T)$,

$$\langle p,\ y-T\rangle\leqslant\varphi(y)-\varphi(T)=\varphi(y)-(\Phi^*(x,A)-0.5A\|g\|^2)=$$

$$\varphi(y)-\Phi^*(x,A)+0.5A\|g\|^2.$$

Further, since $\bar{f}(y)-\bar{f}(x,y)\in K$, we have $\varphi(y)=F(\bar{f}(x,y))\leqslant F(\bar{f}(y))=f(y)$, whence $\langle p,\ y-T\rangle\leqslant f(y)-\Phi^*(x,A)+0.5A\|g\|^2$. From this inequality, with $y\in Q$, it follows by (2.4) that $A\langle g,\ y-T\rangle\leqslant f(y)-\Phi^*(x,A)+0.5A\|g\|^2$. The left-hand side of the last relation is $A\langle g,\ y-x\rangle+A\langle g,\ x-T\rangle=A\langle g,\ y-x\rangle+A\|g\|^2$, and after suitable substitutions, it gives (2.2).

Using the standard treatment of vector inequalities, we have $\bar{f}(y)\leqslant\bar{f}(x,y)+0.5\bar{L}(f)\|x-y\|^2$, and $\bar{f}(y)-\bar{f}(x,y)$ lies in K along with $\bar{L}(f)$. Hence

$$f(y)\leqslant F(\bar{f}(y))\leqslant F(\bar{f}(x,y)+\frac{1}{2}\bar{L}(f)\|x-y\|^2)\leqslant F(\bar{f}(x,y))+$$
$$F(\bar{L}(f))\frac{1}{2}\|x-y\|^2.$$

Thus, with $A\geqslant A(f)$ we have $f(y)\leqslant F(\bar{f}(x,y))+0.5A\|x-y\|^2$, whence (2.3) follows at once. The lemma is proved.

3. Let us describe method $\mathfrak{M}(\bar{x},A_0)$ of solving problem (2.1); here, $\bar{x}\in E$ is the initial approximation, and $A_0>0$ is an a priori estimate of $A(f)$.

Step 0. Initiation. We put $x_1=\bar{x}, a_1=1, y_0=x_1$.

Step 1. k-th iteration, $k\geqslant 1$:

a) the successive inspection we evaluate the least number $i=i_k\geqslant 0$, for which the inequality $\Phi^*(x_k,A)\geqslant f(T(x_k,A))$ holds with $A=2^{i_k}A_{k-1}$ (notice that, by (2.3), this holds for all $A\geqslant A(f)$);

b) we put $A_k=2^{i_k}A_{k-1}$,

$$y_k=T(x_k,A_k), \tag{2.5}$$

$$a_{k+1}=[1+(1+4a_k^2)^{1/2}]/2, \tag{2.6}$$

and hence

$$a_{k+1}(a_{k+1}-1)=a_k^2,\qquad a_{k+1}\geqslant 1; \tag{2.6$'$}$$

we also put

$$x_{k+1}=y_k+\frac{a_k-1}{a_{k+1}}(y_k-y_{k-1}). \tag{2.7}$$

The iteration ends.

Theorem 2. Under the assumptions of Para.1, for the sequence of results $\{y_k\}_{k=1}^{\infty}$ constructed by method $\mathfrak{M}(\bar{x},A_0)$, we have:

1) $y_k\in Q\ \forall k\geqslant 1$ and

$$f(y_k)-f^*\leqslant 2A_k\rho^2(\bar{x})(k+1)^{-2}, \tag{2.8}$$

where

$$A_k\leqslant\bar{A}(f)=\max\{A_0,2A(f)\}; \tag{2.9}$$

2) the construction of $y_1,\dots,y_k$ requires: a) evaluation of $\bar{f}'(\cdot)$ at not more than k points of E; b) evaluation of $\bar{f}(\cdot)$ at not more than $2k+[\log_2(\bar{A}(f)/A_0)]$ points of E; c) solution of not more than $k+[\log_2(\bar{A}(f)/A_0)]$ auxiliary problems of the type $\min\{\Phi(\cdot,\cdot,y)\,|\,y\in Q\}$.

Proof. Relation (2.9) follows from (2.3) and the method of forming A_k. Further, at the k-th iteration, f' is evaluated once, and f is evaluated $2+i_k$ times, whereas the auxiliary problem is solved $1+i_k$ times. Moreover, $A_k=2^{i_1+\dots+i_k}A_0$: this, along with (2.9), is proved in Para.2).

Let us prove (2.8). Let $p_k=(a_k-1)(y_{k-1}-y_k)$, $g_k=g(x_k, A_k)=x_k-y_k$. Then, $p_{k+1}-y_{k+1}=(a_{k+1}-1)(y_k-y_{k+1})-y_{k+1}=a_{k+1}(y_k-y_{k+1})-y_k=a_{k+1}(y_k-x_{k+1}+g_{k+1})-y_k$. On expressing y_k-x_{k+1} in terms of p_k by means of (2.7), we obtain $a_{k+1}(y_k-x_{k+1}+g_{k+1})-y_k=p_k-y_k+a_{k+1}g_{k+1}$. Thus,

$$p_{k+1}-y_{k+1}=p_k-y_k+a_{k+1}g_{k+1}. \tag{2.10}$$

Let x^* be the point of X^* nearest to $\bar{x}$, and let $r_k^2=\|p_k-y_k+x^*\|^2$. By (2.10), we have

$$r_{k+1}^2-r_k^2=2a_{k+1}\langle g_{k+1}, p_k-y_k+x^*\rangle+a_{k+1}^2\|g_{k+1}\|^2. \tag{2.11}$$

On substituting into (2.11) the relation $y_k=x_{k+1}+a_{k+1}^{-1}p_k$, that follows from (2.7), we can rewrite (2.10) as

$$r_{k+1}^2-r_k^2=2(a_{k+1}-1)\langle g_{k+1}, p_k\rangle+2a_{k+1}\langle g_{k+1}, x^*-x_{k+1}\rangle+a_{k+1}^2\|g_{k+1}\|^2. \tag{2.12}$$

We now substitute $A=A_{k+1}$, $x=x_{k+1}$ into (2.2) and note that $f(y_{k+1})\leqslant\Phi^*(x_{k+1}, A_{k+1})$. We find that, for any y of Q,

$$f(y_{k+1})+A_{k+1}\langle g_{k+1}, y-x_{k+1}\rangle+\frac{A_{k+1}}{2}\|g_{k+1}\|^2\leqslant f(y). \tag{2.13}$$

Let $\varphi(y)=f(y)-f^*$. We substitute $y=x^*$ into (2.13) and obtain

$$\langle g_{k+1}, x^*-x_{k+1}\rangle\leqslant-\frac{1}{A_{k+1}}\varphi(y_{k+1})-\frac{1}{2}\|g_{k+1}\|^2. \tag{2.14}$$

On substituting $y=y_k$ into (2.13), we obtain $\|g_{k+1}\|^2\leqslant2\langle g_{k+1}, x_{k+1}-y_k\rangle+2A_{k+1}^{-1}[\varphi(y_k)-\varphi(y_{k+1})]$, or, since $x_{k+1}-y_k=-a_{k+1}^{-1}p_k$,

$$\|g_{k+1}\|^2\leqslant-2a_{k+1}^{-1}\langle g_{k+1}, p_k\rangle+2A_{k+1}^{-1}[\varphi(y_k)-\varphi(y_{k+1})]. \tag{2.15}$$

Substituting (2.14) into (2.12) (which is possible since $a_{k+1}\geqslant1$) we obtain

$$\begin{aligned}r_{k+1}^2-r_k^2\leqslant{}&2(a_{k+1}-1)\langle g_{k+1}, p_k\rangle-2\frac{a_{k+1}}{A_{k+1}}\varphi(y_{k+1})-\\&a_{k+1}\|g_{k+1}\|^2+a_{k+1}^2\|g_{k+1}\|^2=2(a_{k+1}-1)\langle g_{k+1}, p_k\rangle-\\&2\frac{a_{k+1}}{A_{k+1}}\varphi(y_{k+1})+a_k^2\|g_{k+1}\|^2\end{aligned}$$

(the last, by (2.6')). Using (2.15) in the resulting inequality, we obtain

$$r_{k+1}^2-r_k^2\leqslant2\left[(a_{k+1}-1)-\frac{a_k^2}{a_{k+1}}\right]\langle g_{k+1}, p_k\rangle+2\frac{a_k^2}{A_{k+1}}[\varphi(y_k)-\varphi(y_{k+1})]-2\frac{a_{k+1}}{A_{k+1}}\varphi(y_{k+1}),$$

or, on again using (2.6'),

$$r_{k+1}^2-r_k^2\leqslant2\frac{a_k^2}{A_{k+1}}\varphi(y_k)-2\frac{a_k^2+a_{k+1}}{A_{k+1}}\varphi(y_{k+1})=2\frac{a_k^2}{A_{k+1}}\varphi(y_k)-2\frac{a_{k+1}^2}{A_{k+1}}\varphi(y_{k+1}).$$

Hence, since $A_{k+1}\geqslant A_k$, $\varphi(y_k)\geqslant0$, we finally obtain

$$r_{k+1}^2-r_k^2\leqslant2\frac{a_k^2}{A_k}\varphi(y_k)-2\frac{a_{k+1}^2}{A_{k+1}}\varphi(y_{k+1}). \tag{2.16}$$

From (2.16) it easily follows that

$$2\frac{a_k^2}{A_k}\varphi(y_k)\leqslant r_1^2+2\frac{a_1^2}{A_1}\varphi(y_1)=\|p_1-y_1+x^*\|^2+2\frac{a_1^2}{A_1}\varphi(y_1). \tag{2.17}$$

Since $a_1=1$, then $p_1=0$ and $\|y_1-x^*\|^2=\|x_1-x^*\|^2+2\langle g_1, x^*-x_1\rangle+\|g_1\|^2$ and (2.17) gives

$$2\frac{a_k^2}{A_k}\varphi(y_k)\leqslant\rho^2(\bar{x})+2\left[A_1\langle g_1, x^*-x_1\rangle+\frac{A_1}{2}\|g_1\|^2\right]A_1^{-1}. \tag{2.18}$$

By (2.2), the expression in brackets in (2.18) is not greater than $\varphi(x^*)=0$. Hence (2.18) gives

$$2\frac{a_k^2}{A_k}\varphi(y_k)\leqslant\rho^2(\bar{x}). \tag{2.19}$$

From (2.6) it easily follows that $a_{k+1}\geqslant(k+1)/2$, and (2.19) along with (2.9) gives (2.8). The theorem is proved.

4. Notice two important special cases of problem (2.1), in which the auxiliary problems can prove to be realtively simple.

Case a: minimization of a convex function of class $C^{1,1}(E)$ in a "simple" set. Here, $m=1$, $F(u)=u$. Let $\pi_Q(x)$ be the point of Q nearest to $x\in E$. Then, $T(x,A)=\pi_Q(x-A^{-1}f'(x))$ and $\Phi^*(x,A)=f(x)+\langle f'(x),T(x,A)-x\rangle+0.5A\|T(x,A)-x\|^2$, so that solution of the auxiliary problem requires a single evaluation of $\pi_Q(\cdot)$.

Case b: minimization in a polyhedron. Let the set in problem (2.1) be specified by: $Q=\{x\in E|\langle g_j,x\rangle\leqslant b_j, 1\leqslant j\leqslant l\}\neq\varnothing$. Let $F(\cdot)$ be a positive-homogeneous convex function (of degree 1) (in this case we have the identity $F(u)=\max\{\langle\lambda,u\rangle|\lambda\in\partial F(0)\}$, where $\partial F(0)$ is the subdifferential of the function F at zero). Under these assumptions, the auxiliary problem $\min\{\Phi(x,A,y)|y\in Q\}$ is equivalent to the following dual problem:

$$\max\Big\{-\frac{1}{2A}\Big\|\sum_{i=1}^m\lambda_i f_i'(x)+\sum_{j=1}^l\mu_j q_j\Big\|^2+ \tag{2.20}$$
$$\sum_{i=1}^m\lambda_i f_i(x)+\sum_{j=1}^l\mu_j(-b_j+\langle q_j,x\rangle)\,|\,(\lambda_1,\dots,\lambda_m)=$$
$$\lambda\in\partial F(0),\quad \mu_j\geqslant 0,\quad 1\leqslant j\leqslant l\Big\}.$$

If $\lambda(x),\mu(x)$ is the solution of problem (2.20), then

$$T(x,A)=x-A^{-1}\Big[\sum_{i=1}^m\lambda_i(x)f_i'(x)+\sum_{j=1}^l\mu_j(x)q_j\Big],$$
$$\Phi^*(x,A)=F(\bar f(x,T(x,A)))+0.5A\|T(x,A)-x\|^2.$$

For small values of $m+l$ and a simply constructed $\partial F(0)$, problem (2.20) is not very difficult. We recall that

$$\partial F_1(0)=\Big\{\lambda\in R^m|\lambda_i\geqslant 0,\ \sum_{i=1}^m\lambda_i=1\Big\}\ \text{ for }$$
$$F_1(u)=\max\{u^{(i)}|i=1,2,\dots,m\},$$
$$\partial F_2(0)=\{\lambda\in R^m|0\leqslant\lambda_i\leqslant 1\}\ \text{ for }\ F_2(u)=\sum_{i=1}^m(u^{(i)})_+,$$
$$\partial(\alpha F+\beta G)(0)=\{\lambda\in R^m|\lambda=\alpha\lambda'+\beta\lambda'',$$
$$\lambda'\in\partial F(0),\lambda''\in\partial G(0)\}\ \text{ with }\ \alpha,\beta\geqslant 0.$$

5. We now take the case when $f(x)=F(\bar f(x))$ has a strict minimum. For simplicity, we confine ourselves to the case when, for some $a>0$,

$$f(x)-f^*\geqslant 0.5a\rho^2(x),\qquad x\in Q,\qquad f(x)\leqslant f(\bar x). \tag{2.21}$$

Here, $\bar x\in Q$ is the given initial approximation.

Notice that (2.21) holds in the important special case of minimizing the maximum of a strongly convex function. In the latter situation, $A(f)$ is an upper bound (with respect to $x\in E$ and with respect to functions f_i of the set) of the maximum points of the spectra of Hessians $f_i''(x)$, while a in (2.21) is the lower bound of the minimum points of the spectra of these Hessians.

Returning to the general situation (2.1), (2.21), notice that here, in the same way as in Para.5 of Sect.1, it is sensible to introduce renewal into method $\mathfrak{M}(\bar x,A_0)$. Let the a of (2.21) and $A\geqslant A(f)$ be given a priori. We put

$$N=]2^{3/2}(A/a)^{1/2}[\tag{2.22}$$

and define the points $\bar x^{(i)}$ as: $\bar x^{(0)}=\bar x$ and $\bar x^{(i+1)}$ is the N-th point of $\{y_s\}_{s=1}^\infty$, formed by method $\mathfrak{M}(\bar x^{(i)},A)$ in problem (2.1); by (2.8) we have $f(\bar x^{(1)})-f^*\leqslant 2A\rho^2(\bar x)(N+1)^{-2}$. By (2.21), we have $\rho^2(x)\leqslant 2a^{-1}[f(\bar x)-f^*]$, which gives

$$f(\bar x^{(1)})-f^*\leqslant\frac{2A}{a(N+1)^2}[f(\bar x)-f^*].$$

In accordance with (2.22), we obtain from this for $i=0$

$$f(\bar x^{(i+1)})-f^*\leqslant[f(\bar x^{(i)})-f^*]/2. \tag{2.23}$$

But then (2.21) holds when $\bar x$ is replaced by $\bar x^{(1)}$, so that (2.23) in fact holds for $i=1$; by induction, (2.23) holds for all i. Notice that, under the assumptions made in this paragraph, the construction of $\bar x^{(1)},\dots,\bar x^{(i)}$ requires $N_i=iN$ evaluations of $\bar f$ and $\bar f'$ (and solution of N_i auxiliary problems).

Notice in conclusion that the error estimates for the methods of this Section, as functions of the number of iterations, are the same as those corresponding to the case $\sigma=r=2$ in Sect.1 (viz: (2.8) the same as (1.5'), and (2.23) the same as (1.15)).

Hence the methods of Sect.2, like those of Sect.1, are essentially unimprovable with respect to their rate of convergence.

REFERENCES

1. NEMIROVSKII A.S. and YUDIN D.B., Complexity of problems and the efficiency of optimization methods (Slozhnost' zadach i effektivnost' metodov optimizatsii), Nauka, Moscow, 1979.
2. NEMIROVSKII A.S. and YUDIN D.B., Informational complexity of mathematical programming, Izv. Akad. Nauk SSSR, Tekhn. kibernetika, No.1, 88-117, 1983.
3. NESTEROV YU.E., A method of solving convex programming problems with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSRm 269, No.3, 543-547, 1983.
4. NEMIROVSKII A.S., A unit vector method of smooth convex minimization, Izv. Akad. Nauk SSSR, Tekhn. kibernetika, No.2, 18-29, 1982.

Translated by D.E.B.

U.S.S.R. Comput.Maths.Math.Phys.,Vol.25,No.2,pp.30-37,1985 0041-5553/85 $10.00+0.00
Printed in Great Britain **Pergamon Journals Ltd.**

SOME INVERSE PROBLEMS OF MAGNETO-TELLURIC SOUNDING FOR OBLIQUELY INCIDENT PLANE WAVES. I.*

E.Y. BIDAIBEKOV and V.G. ROMANOV

Inverse problems of magneto-telluric sounding are formulated, and are reduced to converse problems for a non-stationary system of Maxwell equations with Cauchy data and a concentrated source on the boundary. The structure of solutions of the latter problem is analyzed, and the expressions needed to study the inverse problems are given.

In the present paper we study the unique solvability of some particular inverse problems of magneto-telluric sounding (m.t.s.), when the required parameters of the medium depend only on the space variable z. The general statement of these inverse problems is as follows: the parameters of the medium are assumed known in the half-space $z\leqslant 0$ and unknown in the half-space $z\geqslant 0$, where they have to be determined from observations of the reflection coefficients of obliquely incident sinusoidal plane waves.

Notice that inverse problems of m.t.s. were studied in various formulations in /1—4/. The results below are a direct development and continuation of the study method that we used in /4/.

1. Statement of inverse problems.

We take the model of an unbounded medium, filling the entire space of variables x, y, z. The properties of the medium filling the half-space $z\leqslant 0$ are assumed known, and are characterized by the constants $\varepsilon=\varepsilon_0^->0$, $\mu=\mu_0^->0$, $\sigma=\sigma_0^->0$. In the domain $z\geqslant 0$ the medium is simply anisotropic; the parameters ε, μ are unknown, depend only on the one coordinate z, and have the form

$$\varepsilon=\begin{Vmatrix}\varepsilon_1 & 0 & 0\\ 0 & \varepsilon_1 & 0\\ 0 & 0 & \varepsilon_2\end{Vmatrix},\qquad \mu=\begin{Vmatrix}\mu_1 & 0 & 0\\ 0 & \mu_1 & 0\\ 0 & 0 & \mu_2\end{Vmatrix},\qquad \sigma=\begin{Vmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_1 & 0\\ 0 & 0 & \sigma_2\end{Vmatrix}.$$

The parameters ε, μ, σ can have a finite discontinuity in the plane $z=0$.

The plane electromagnetic wave is written as the superposition of a plane-polarized and a normally polarized wave, when the electric and magnetic field-strengths $\mathbf{E}$ and $\mathbf{H}$ are independent of the y coordinate; in the case of parallel polarization they have the form $\mathbf{E}=(E_1,0,E_3)$, $\mathbf{H}=(0,H_2,0)$, and in the case of normal polarization: $\mathbf{E}=(0,E_2,0)$, $\mathbf{H}=(H_1,0,H_3)$. In /4/ we studied the inverse problem of m.t.s. for an iostorpic medium, or more precisely, the problem of finding the parameters of the medium $\varepsilon(z)>0$, $\sigma(z)\geqslant 0$ $(\mu=1)$ for $z\geqslant 0$ in the case of parallel polarization. In the present paper, to find the parameters, we use either one polarization separately, or both polarizations together.

The system of Maxwell equations is written in either case as

$$\frac{\partial \mathbf{\Pi}^\nu}{\partial t}+A_1^\nu\frac{\partial \mathbf{\Pi}^\nu}{\partial x}+A_2^\nu\frac{\partial \mathbf{\Pi}^\nu}{\partial z}+A_3^\nu\mathbf{\Pi}^\nu=0,\qquad \nu=1,2. \tag{1.1}$$

$$\mathbf{\Pi}^1=\begin{Vmatrix}E_1\\ E_3\\ H_2\end{Vmatrix},\qquad A_1^1=\begin{Vmatrix}0 & 0 & 0\\ 0 & 0 & -\varepsilon_2^{-1}\\ 0 & -\mu_1^{-1} & 0\end{Vmatrix},\qquad A_2^1=\begin{Vmatrix}0 & 0 & \varepsilon_1^{-1}\\ 0 & 0 & 0\\ \mu_1^{-1} & 0 & 0\end{Vmatrix},$$

$$A_3^1=\begin{Vmatrix}\sigma_1/\varepsilon_1 & 0 & 0\\ 0 & \sigma_2/\varepsilon_2 & 0\\ 0 & 0 & 0\end{Vmatrix},\qquad \mathbf{\Pi}^2=\begin{Vmatrix}E_2\\ H_1\\ H_3\end{Vmatrix},$$

*Zh.vychisl.Mat.mat.Fiz.,25,3,370-380,1985