SIAM J. CONTROL AND OPTIMIZATION
Vol. 26, No. 6, November 1988

003

SMOOTH OPTIMIZATION METHODS FOR MINIMAX PROBLEMS*

R. A. POLYAK†

Abstract. The classical discrete minimax problem is considered. It is transformed into an equivalent problem by a monotone transformation of the initial functions. It was found that the classical Lagrangian of the equivalent problem has a number of important properties both in primal and dual spaces in convex as well as in nonconvex cases.

In particular, the classical Lagrangian of the equivalent problem, being as smooth as the initial functions, has the main advantages of augmented Lagrangians. This makes it possible to construct a multiplier method for the minimax problem and a general method for the simultaneous solution of the primal and the dual problems.

These methods are based on the theory of methods of smooth optimization and preserve the main advantages of the latter for nonsmooth minimax problems.

Key words. minimax problem, monotone transformation, multiplier method, smooth optimization, dual problems

AMS(MOS) subject classification. 90C20

1. Introduction. A growing interest in problems of nonsmooth optimization and, in particular, in minimax problems is due to the special role these problems have in modern optimization theory (see, for example, [11], [24], [25]). Of the large number of papers that have appeared recently on this subject (see [10], [15], [16], [25], [26], [35]) the research summarized in [35] plays an important role. In it methods of generalized gradient and their variants are studied and developed. However, the rate of convergence of these methods is not high. Even in the case of ellipsoid methods ([18], [23], [35]) that have polynomial complexity applied to linear minimax problems (see [10] for a proof), the rate of convergence is estimated (see [35]) by the ratio $q_n = 1-(2n^2)^{-1}$ (n is the space dimension), i.e., $q_n \to 1$ as $n \to \infty$. In these methods the properties of convexity and smoothness of functions that appear in the minimax problem cannot essentially be used for acceleration of convergence since the gradient of $F(x) = \max\{f_i(x) | i = \overline{1, m}\}$ is not smooth, not to mention the absence of higher-order smoothness even when $f_i(x)$, $i = \overline{1, m}$ are sufficiently smooth.

Therefore our goal is to construct methods for solving minimax problems that will preserve the convergence rate of smooth optimization methods (assuming some properties of smoothness and convexity of $f_i(x)$, $i = \overline{1, m}$) without substantially increasing the number of computations needed at each step.

This is achieved by application of a monotone transformation to initial functions and subsequent use of the classical Lagrangian of the equivalent problem.

In this paper it is established that the classical Lagrangian of the equivalent problem shares all the advantages of augmented Lagrangians (see [7], [19], [29], [34]) in both convex and nonconvex cases. Moreover, it has the same order of smoothness as $f_i(x)$.

It allows us to use all means of smooth optimization techniques for the solution of minimax problems, including methods of Newtonian and quasi-Newtonian type.

* Received by the editors August 7, 1985; accepted for publication (in revised form) December 22, 1987.

† c/o Dr. Claude Lemarechal, Institut Nationale de Recherche en Informatique et Automatique, Domaine de Voluceau, Rocquencourt, Le Chesnay, France.

In addition, smoothness properties of the Lagrangian on the initial space give a deeper insight into convexity and smoothness properties of the dual function that can be used for construction of a method for simultaneous solution of the primal and the dual problem. The rate of convergence of the dual problem is defined by the product of the primal and dual spaces.

The main results of the paper were obtained in 1980 and 1981 and announced in [30] and [31]. In 1983 we learned from [3] about the possibility of developing the multiplier method for the minimax problem using the function

$$Q_c(g(x), \mu) = c^{-1} \log (\sum \mu_i \exp (cg_i(x))).$$

However, our main results are not presented in [3], or in the subsequent papers [4]–[6].

It seems impossible to obtain these results by a direct reformulation of the minimax problem as a constrained optimization problem, and still preserve the initial condition.

Besides, the corresponding class of augmented Lagrangians $\hat{P}_1$, as was mentioned in ([3, p. 309]), had been insufficiently investigated.

The class of functions $\hat{P}_1$ was first carefully studied in convex as well as nonconvex cases in [33]. In particular, using a monotone transformation and ordinary Lagrangian for the equivalent problem it was possible to essentially improve the barrier and the center methods.

Some other approaches to solving discrete minimax problems based on the replacement of $F(x)$ by sequences of smooth functions are considered in [2], [8], [16], [32].

2. Monotone transformation of the minimax problem. In this section we introduce a monotone transformation of $f_i(x)$, $i = \overline{1, m}$ and study properties of the classical Lagrangian of the equivalent problem.

We suppose that $f_i(x) \in C^1$, $i = \overline{1, m}: R^n \to R^1$ and consider the following problem:

$$x^* \in \operatorname{Argmin} \{F(x) | x \in R^n\}. \tag{1}$$

The existence of x^* is guaranteed by, for example, the following condition:

(2) There exists $c > 0$ such that $\Omega = \{x: F(x) \leqq c\}$ is compact.

Let $\Psi(t): R^1 \to R^1$ be a strictly convex and increasing function. Then functions $\tilde{f}_i(x, k) = k^{-1}\Psi(kf_i(x))$, where $k > 0$ define a minimax problem equivalent to the original one, i.e.,

$$x^* \in \operatorname{Argmin} \left\{ \max_{1 \leqq i \leqq m} \tilde{f}_i(x, k) | x \in R^n \right\}.$$

For $k > 0$, consider on $R \times S_m$, $S_m = \{u: \sum u_i = 1, u_i \geqq 0, i = \overline{1, m}\}$ the ordinary Lagrangian function

$$A(x, u, k) = k^{-1} \sum_{i=1}^{m} u_i \Psi(kf_i(x)).$$

It appears that $A(x, u, k)$ has a number of advantages over the Lagrangian $L(x, u) = \sum_{i=1}^{m} u_i g_i(x)$ of the initial problem. For definiteness we shall take $\psi(t) = \exp t$, so that

$$A(x, u, k) = k^{-1} \sum_{i=1}^{m} u_i \exp (kf_i(x)).$$

We observe that $M(x, k) = \sum_{i=1}^{m} \exp (kf_i(x))$ was already considered by Motzkin in 1952 [22].

The results that follow will show that the relation between $A(x, u, k)$ and $M(x, k)$ is the same as between augmented Lagrangians and penalty functions, despite the fact that $A(x, u, k)$ is an ordinary Lagrangian for an equivalent problem.

In what follows we suppose that there exists a Kuhn-Tucker point $z^* = (x^*, u^*)$:

$$L'_x(z^*) = \sum u_i^* f_i'(x^*) = 0, \quad u_i^*(F(x^*) - f_i(x^*)) = 0, \quad i = \overline{1, m}, \tag{3}$$

$$\sum_{i=1}^{m} u_i^* = 1, \qquad u_i^* \geqq 0.$$

Let $f(x) = (f_1(x), \cdots, f_m(x))$, $I^* = \{i: F^* = F(x^*) = f_i(x^*)\} = \{1, \cdots, r\}$, $\bar{f}(x) = (f_1(x), \cdots, f_r(x))$ be a vector of active functions, let $f'(x) = J(f(x))$ be its Jacobi matrix and let $\bar{f}'(x) = J(\bar{f}(x))$. Set $e = (1, \cdots, 1) \in R^n$, $\bar{e} = (1, \cdots, 1) \in R^r$. Sometimes we shall use the condition

$$\text{Rank}\,(\bar{f}'(x^*), -\bar{e}^T) = r, \quad u_i^* > 0, \quad i = \overline{1, r}, \tag{4}$$

which together with the condition

$$(L''_{xx}(z^*)y, y) \geqq \lambda \|y\|^2, \quad \lambda > 0 \quad \forall y: \bar{f}'(x^*)y = 0 \tag{5}$$

is a second-order sufficient condition in the minimax problems.

Set $S(x^*, \varepsilon) = \{x: \|x - x^*\| \leqq \varepsilon\}$, $S(u^*, \varepsilon) = \{u \in S_m: \|u - u^*\| \leqq \varepsilon\}$, and $S(z^*, \varepsilon) = S(x^*, \varepsilon) \times S(u^*, \varepsilon)$. Sometimes we shall also use the condition

$$\|f_i''(x) - f_i''(y)\| \leqq L\|x - y\| \quad \forall (x, y) \in S(x^*, \varepsilon) \times S(x^*, \varepsilon). \tag{6}$$

We shall use the following version of the well-known theorem of Debreu [1].

PROPOSITION 1. *Let* $A = A^T: R^n \to R^n$; $B: R^n \to R^r$,

$$U = \text{diag}\, u_i: R^r \to R^r, \qquad u = (u_1, \cdots, u_r) > 0.$$

$$Bx = 0 \Rightarrow (Ax, x) \geqq \lambda \|x\|^2, \qquad \lambda > 0. \tag{7}$$

Then for any $\gamma < \lambda$ *there exists a* $k_0 \geqq 0$ *such that for* $k \geqq k_0$

$$((A + kB^T U B)x, x) \geqq \gamma \|x\|^2 \quad \forall x \in R^n. \tag{8}$$

The proof is similar to that of Debreu's Theorem.

The following theorem describes the properties of $A(x, u, k)$.

THEOREM 1. *Assume that conditions* (3)-(5) *are satisfied and* $f_i \in C^2$. *Then there exist* $\varepsilon > 0$, $k_0 > 0$, *and a convex neighborhood* $V(k_0, \varepsilon)$ *of* x^* (*dependent on* k_0 *and* ε) *such that for* $k \geqq k_0$ *and for all* $u \in S(u^*, \varepsilon)$ *the following hold*:

(1) *There exists* $\hat{x} \equiv x_k(u) \in \text{int}\, V(k_0, \varepsilon)$ *such that*

$$A'_x(\hat{x}, u, k) = 0.$$

(2) $A(\cdot, u, k)$ *is strongly convex in* $V(k_0, \varepsilon)$ *so that*

$$\hat{x} = \underset{x = V(k_0, \varepsilon)}{\text{argmin}}\, A(x, u, k).$$

(3) *For* $\hat{x}$ *and* $\hat{u} \equiv \hat{u}_k(u) = (\hat{u}_1, \cdots, \hat{u}_m)$: $\hat{u}_i = u_i \exp(kf_i(\hat{x})) \cdot (\sum u_j \exp(kf_j(\hat{x})))^{-1}$, $i = \overline{1, m}$ *the following estimates hold*:

$$\|\hat{x} - x^*\| \leqq ck^{-1}\|u - u^*\|, \qquad \|\hat{u} - u^*\| \leqq ck^{-1}\|u - u^*\|, \tag{9}$$

where c *is independent of* k.

(4) *If* $f_i(x)$, $i = \overline{1, m}$ *are convex there exists* $\hat{x} = \text{argmin}\,\{A(x, u, k) | x \in R^n\}$ *and* (2)-(3) *are true for* $\hat{x}$ *and* $\hat{u}$.

Proof. For $(x, u) \in S(z^*, \varepsilon)$ and $\varkappa \in R^1$ we consider the functions

$$\hat{U}_i(x, u, \varkappa) = \begin{cases} u_i \exp(\cdot |\varkappa|^{-1} f_i(x))(\sum u_j \exp(|\varkappa|^{-1} f_j(x)))^{-1}, & \varkappa \neq 0, \\ 0, & \varkappa = 0. \end{cases}$$

If $\varepsilon > 0$ is small enough, then the definition is correct, because $\sum u_j \exp(|\varkappa|^{-1} f_j(x)) > 0$. The functions $\hat{U}_i(x, u, \varkappa) \equiv \hat{U}_i(\cdot)$, $i > r$ are smooth on $S(z^*, \varepsilon) \times R^1$ and $\hat{U}'_{ixu\varkappa}(x, u, 0) = 0$, $i > r$. In fact, there is a $\sigma > 0$: $F^* - f_i(x^*) > \sigma$, $i > r$, $u_i^* \geqq \sigma$, $i \leqq r$. Choose an $\varepsilon > 0$ such that $|f_i(x) - f_i(x^*)| \leqq \delta$, $|u_i - u_i^*| \leqq \delta$, for all $(x, u) \in S(z^*, \varepsilon)$, $\delta < \sigma(2m)^{-1}$. Then for $\varkappa \neq 0$ we have

$$\begin{aligned} &\sum u_j \exp(|\varkappa|^{-1} f_j(x)) \\ &\quad = \sum_{j=1}^{r} u_j \exp(|\varkappa|^{-1} f_j(x)) + \sum_{j=r+1}^{m} u_j \exp(|\varkappa|^{-1} f_j(x)) \\ &\quad \geqq (\sigma - \delta) r \exp(|\varkappa|^{-1}(F^* - \delta)) - \delta(m - r) \exp(|\varkappa|^{-1}(F^* - \sigma + \delta)) \\ &\quad \geqq 2^{-1}(\sigma - \delta) r \exp(|\varkappa|^{-1}(F^* - \delta)). \end{aligned}$$

For $i > r$ we have

$$\begin{aligned} 0 &< \exp(|\varkappa|^{-1} f_i(x))(\sum u_j \exp(|\varkappa|^{-1} f_j(x)))^{-1} \\ &\leqq \exp(|\varkappa|^{-1}(F^* - \sigma + \delta))[2^{-1}(\sigma - \delta) r \exp(|\varkappa|^{-1}(F^* - \delta))]^{-1} \\ &= 2(\sigma - \delta)^{-1} r^{-1} \exp(-|\varkappa|^{-1}(\sigma - 2\delta)). \end{aligned}$$

Therefore there exists $c > 0$ independent of $\varkappa$ and such that $\hat{U}_i(x, u, \varkappa) \leqq c \exp(-c|\varkappa|^{-1})$, $\varkappa \neq 0$, so $\hat{U}_i(\cdot)$ is continuous in $S(z^*, \varepsilon) \times (-\varepsilon, \varepsilon)$, $i > r$. Further,

$$\begin{aligned} &\hat{U}'_{ix}(\cdot) = |\varkappa|^{-1} \hat{U}_i(\cdot)(f_i'(x) - \sum \hat{U}_j(\cdot) f_j'(x)), \\ &\hat{U}_{iu_i}(\cdot) = (1 - \hat{U}_i(\cdot)) \exp(|\varkappa|^{-1} f_i(x))(\sum u_j \exp(|\varkappa|^{-1} f_j(x)))^{-1}, \\ &\hat{U}'_{iu_j}(\cdot) = -\hat{U}_i(\cdot) \exp(|\varkappa|^{-1} f_j(x))(\sum u_j \exp(|\varkappa|^{-1} f_j(x)))^{-1}, \qquad j \neq i, \\ &\hat{U}'_{i\varkappa}(\cdot) = -(\varkappa|\varkappa|)^{-1} \hat{U}_i(\cdot)(f_i(x) - \sum \hat{U}_j(\cdot) f_j(x)). \end{aligned}$$

Therefore, taking into account the estimate for $\hat{U}_i(\cdot)$, we obtain for the full derivative $\hat{U}'_{ixu\varkappa} = (\hat{U}'_{ix}(\cdot); \hat{U}'_{iu}(\cdot); \hat{U}'_{i\varkappa}(\cdot))$ that there is a $c > 0$ independent of $\varkappa$ such that $\|\hat{U}'_i(\cdot)\| \leqq c|\varkappa|^{-2} \exp(-c|\varkappa|)$. Therefore $\hat{U}'_i(x, u, 0)$ exists and $\hat{U}'_i(x, u, 0) = 0$, so $U'_i(x, u, \varkappa)$ is continuous. Set $\rho(x, u, \varkappa) = \sum_{j=r+1}^{m} \hat{U}_j(\cdot) f_j'(x)$; then $\rho(x, u, \varkappa)$ is smooth on $S(z^*, \varepsilon) \times R$ and $\rho(x, u, 0) = 0$, $\rho'(x, u, 0) = 0$. Let $\bar{u}^* = (u_1^*, \cdots, u_r^*)$, and $S(\bar{u}^*, \varepsilon) = \{u = (u_1, \cdots, u_r) : \|u - \bar{u}^*\| \leqq \varepsilon\}$. On $S(x^*, \varepsilon) \times S(\bar{u}^*, \varepsilon) \times (-\varepsilon, \varepsilon) \times S(u^*, \varepsilon) \times (-\varepsilon, \varepsilon)$ we consider the map $\Phi(x, \hat{u}, t, u, \varkappa) : R^{n+r+m+2} \to R^{n+r+1}$ defined by $\Phi(x, \hat{u}, t, u, \varkappa) = (\sum_{i=1}^{r} \hat{u}_i f_i'(x) + \rho(x, u, \varkappa);$ $f_i(x) - F^* + t - \varkappa \ln \hat{u}_i u_i^{-1},$ $i = \overline{1, r};$ $\sum_{i=1}^{r} \hat{u}_1 + \sum_{i=r+1}^{m} \hat{U}_j(x, u, \varkappa) - 1)$. Since $f_i(x^*) = F^*$, $i = \overline{1, r}$, $\sum u_i^* = 1$, $\sum_{i=1}^{r} u_i^* f_i'(x^*) = 0$, $\rho(x^*, u^*, 0) = 0$, $\sum_{j=r+1}^{m} \hat{U}_j(x^*, u^*, 0) = 0$ we have $\Phi(x^*, \bar{u}^*, 0, u^*, 0) = 0$. Then setting $\bar{f}'(x^*) = \bar{f}'$, $L''_{xx}(z^*) = L''_{xx}$ we get

$$\Phi'_{x\hat{u}t} = \Phi'_{x\hat{u}t}(x^*, \bar{u}^*, 0, u^*, 0) = \begin{pmatrix} L''_{xx} & \bar{f}'^T & 0 \\ \bar{f}' & 0 & \bar{e}^T \\ 0 & \bar{e} & 0 \end{pmatrix}.$$

The matrix $\Phi'_{x\hat{u}t}$ is nonsingular. Indeed, set $w = (y, v, \tau)$, $y \in R^n$, $v \in R^r$, $\tau \in R$. Then $\Phi'_{x\hat{u}t} w = 0$ implies $L''_{xx} y + \bar{f}'^T v = 0$, $\bar{f}' y + \tau \bar{e} = 0$, $(\bar{e}, v) = 0$. Taking the inner product of the second equality with $\bar{u}^*$ and taking into account the Kuhn–Tucker relations, we obtain $(\sum_{i=1}^{r} u_i^* f_i'(x^*), y) + \tau = 0$, so $\tau = 0$. It implies $\bar{f}' y = 0$. Taking the inner product

at the first equality and y we obtain $(L''_{xx}y, y)+(\bar{f}'y, v)=0$, so $y=0$ by (5) and $\bar{f}'^{T}v=0$, $(v, \bar{e})=0$. The last equalities together with (4) give $v=0$. Since $\Phi'_{x\hat{u}t}w=0$ implies $w=0$ the matrix $\Phi'_{x\hat{u}t}$ is nonsingular in a neighborhood of $(x^*, \bar{u}^*, 0, u^*, 0)$. Since $f_i(x)\in C^2$, the implicit function theorem (see [20]) suggests that in this neighborhood there exists a unique smooth vector function $(x(u, \varkappa), \hat{U}(u, \varkappa), t(u, \varkappa))$:

$$\Phi(x(u, \varkappa), \hat{U}(u, \varkappa), t(u, \varkappa), u, \varkappa)\equiv\Phi(\cdot)\equiv 0$$

such that

$$\sum_{i=1}^{m}\hat{u}_i(x(u, \varkappa), u, \varkappa)f'_j(x(u, \varkappa))=0,$$

$$\hat{u}_i(x(u, \varkappa), u, \varkappa)=u_i\exp(|\varkappa|^{-1}f_i(x, u, \varkappa))(\sum u_j\exp(|\varkappa|^{-1}f_j(x(u, \varkappa))))^{-1}, \qquad i=\overline{1, m}.$$

For $\varkappa\neq 0$ equality $\Phi(\cdot)=0$ is equivalent to $A'_x(x, u, k)=0$, $k=\varkappa^{-1}$. Differentiating the map $\Phi(\cdot)$ with respect to u, we obtain $\Phi'_{x\hat{u}t}(\cdot)\times w(\cdot)+\Phi'_u(\cdot)=0$, where $w(\cdot)=(x'_u(u, \varkappa), u'_u(u, \varkappa), t'_u(u, \varkappa))$, i.e., $w(\cdot)=-\Phi'^{-1}_{x\hat{u}t}(\cdot)\Phi'_u(\cdot)$. Since $\Phi'_{x\hat{u}t}$ is a nonsingular matrix and $f_i(x)\in C^2$, there are $\varepsilon>0$ and $c_1>0$, $c_2>0$ that are independent of $\varkappa$ and such that the matrix $\Phi'_{x\hat{u}t}(\cdot)$ is nonsingular in $S(u^*, \varepsilon)\times(-\varepsilon, \varepsilon)$ and $\|\Phi'^{-1}_{x\hat{u}t}(\cdot)\|\leqq c_1$, $\|\Phi'_u(\cdot)\|\leqq c_2\varkappa$. Therefore there exists $c>0$ not dependent on $\varkappa$ and such that $\|w(\cdot)\|\leqq c\varkappa$. Since $\Phi(x^*, \bar{u}^*, 0, u^*, \varkappa)=0$ for all $\varkappa(|\varkappa|<\varepsilon)$ we have $x(u^*, \varkappa)=x^*$, $\hat{u}(u^*, \varkappa)=u^*$, so

$$\|x(u, \varkappa)-x^*\|\leqq c\varkappa\|u-u^*\|, \qquad \|\hat{u}(u, \varkappa)-u^*\|\leqq c\varkappa\|u-u^*\|.$$

Setting $k=\varkappa^{-1}>0$; $\hat{x}\equiv x_k(u)\equiv x(u, \varkappa)$; and $\hat{u}\equiv\hat{u}_k(u)=(\hat{u}(u, \varkappa), \hat{u}_i(x, u, \varkappa), i=\overline{r+1, m})$, we obtain the estimates (9).

Finally, the strong convexity of $A(x, u, k)$ in $\varkappa$ in the neighborhood of $\hat{x}=x(u, \varkappa)$ follows from $f_i(x)\in C^2$ and the relation

$$\begin{aligned}A''_{xx}(\hat{x}, u, k)&=\sum u_i\exp(kf_i(\hat{x}))f''_i(\hat{x})+k\sum u_i\exp(kf_i(\hat{x}))f'^{T}_i(\hat{x})f'_i(\hat{x})\\&=(\sum u_i\exp(kf_i(\hat{x}))(L''_{xx}(\hat{z})+k\sum\hat{u}_if'^{T}_i(\hat{x})f'_i(\hat{x})))\end{aligned}$$

if we take into account Proposition 1 and estimate (9) for $k>0$ sufficiently large. Due to strong convexity $A(x, u, k)$ in x, the necessary condition $A'_x(\hat{x}, u, k)=0$ is sufficient for $\hat{x}$ to be a minimizer. Observe that we have not assumed that $f_i(x)$ are convex. If so, the condition $A'_x(\hat{x}, u, k)=0$, $u\geqq 0$, $k>0$, together with the positive definiteness of matrix $A''_{xx}(\hat{x}, u, k)$, gives (4), and the proof of Theorem 1 is complete. □

The local results (1)-(3) of Theorem 1 are valid under weaker conditions than (4)-(5) despite the fact that these are the standard second-order sufficient condition for the minimax problem. Consider an example.

Put $I^*=\{1, 2, 3, 4\}$; $f_1(x_1, x_2)=f_1(\cdot)=(x_1-1)^2+x_2^2$; $f_2(\cdot)=(x_1+1)^2+x_2^2$; $f_3(\cdot)=x_1^2+(x_2-1)^2$; $f_4(\cdot)=x_1^2+(x_2+1)^2$. In this case condition (4) is not satisfied and the set $Y: f'(x^*)y=0$ consists of a single point $y=0$, which makes (5) meaningless. However, (1)-(3) of Theorem 1 remain true if we take $I_1=\{1, 2\}$ or $I_2=\{3, 4\}$ as I^*.

In the first case, conditions similar to (4), (5) are satisfied for $z^*=(x^*; u^*)=(0; 0; \frac{1}{2}; \frac{1}{2}; 0; 0)$, and in the second case they are satisfied for $z^*=(x^*; u^*)=(0; 0; 0; 0; \frac{1}{2}; \frac{1}{2})$.

Moreover, the results of Theorem 1 remain true if we replace the convex functions $f_1(\cdot)$ and $f_2(\cdot)$ by the nonconvex functions $\tilde{f}_1(\cdot)=-(x_1-1)^2+x_2^2$; $\tilde{f}_2(\cdot)=-(x_1+1)^2+x_2^2$. This shows that (1)-(3) of Theorem 1 hold not only without convexity of $f_i(x)$, $i=\overline{1, r}$, but with conditions similar to (4)-(5) satisfied for any minimal set $I\subset I^*$.

The set $I\subset I^*$ is called minimal if

$$\min\left\{\left\|\sum_{i\in I}u_if'_i(x^*)\right\|\;\middle|\;\sum_{i\in I}u_i=1, u_i\geqq 0, i\in I\right\}=\left\|\sum_{i\in I}u_i^*f'_i(x^*)\right\|=0$$

and

$$\min\left\{\left\|\sum_{i\in I\setminus j} u_i f_i'(x^*)\right\| \,\middle|\, \sum_{i\in I\setminus j} u_i = 1,\ u_i \geqq 0,\ i\in I\setminus j\right\} > 0$$

for all $j\in I$.

It is easy to show that there is a one-to-one correspondence between minimal sets and vertices of the Kuhn-Tucker polyhedron $Q(x^*)=\{u\in S_r\colon \sum_{i=1}^r u_i f_i'(x^*)=0\}$. Moreover, for any minimal set I, the vectors $(f_i'(x^*), -1)$ are linearly independent and $u_i^*>0$, $i\in I$. Thus, condition (4) is always satisfied for the minimal set. Therefore (1)-(3) of Theorem 1 remain true if, instead of (4), (5), we assume that

$$(5') \qquad (L''_{xx}(z^*)y, y)\geqq \lambda\|y\|^2, \quad \lambda>0 \quad \forall y: f_i'(x^*)y=0, \quad i\in I,$$

where u^* are vertices of the Kuhn-Tucker polyhdron, which correspond to I.

3. Multiplier method. Theorem 1 allows us to realize the method of multipliers:

$$(10) \qquad \begin{aligned} &x^{S+1}=\operatorname*{argmin}_x \exp(-kF(x^S))A(x, u^S, k),\\ &u^{S+1}=(u_i^{S+1}=u_i^S\exp(kf_i(x^{S+1}))(\textstyle\sum u_j^S\exp(kf_j(x^{S+1})))^{-1},\ i=\overline{1, m}). \end{aligned}$$

Under the conditions of Theorem 1 we obtain the estimate

$$(11) \qquad \|x^S-x^*\|\leqq ck^{-S}, \qquad |u^S-u^*\|\leqq ck^{-S}$$

with $c>0$ independent of k.

In order to realize the method (10), if $f_i(x)$ are convex it is enough to know $u^0\in S(u^*, \varepsilon)$; if $f_i(x)$ are nonconvex we must know $z^0\in S(z^*, \varepsilon)$.

The next lemmas allow us to obtain $z^0\in S(z^*, \varepsilon)$. The second difficulty, which we must overcome in order to realize (10), is to change the infinite procedure of smooth optimization to find x^S to a finite one and preserve the estimates above.

Let $\mu\geqq 0$, $U(x, u, k)=(u_i(x, u, k)=u_i\exp(kf_i(x))\cdot(\sum u_j\exp(kf_j(x)))^{-1}$, $i=\overline{1, m}$).

PROPOSITION 2. *If the conditions of Theorem* 1 *are satisfied, there exist* $k_0>0$ *and* $c>0$ *independent of* k *such that for all* $k\geqq k_0$ *and* $\tilde z=(\tilde x, \tilde u)$

$$\|\exp(-kF(x))A_x'(\tilde x, u, k)\|\leqq \mu k^{-1}\|U(\tilde x, u, k)-u\|, \qquad \tilde u=U(\tilde x, u, k)$$

the following estimate holds:

$$\begin{aligned} &\|\tilde x-x^*\|\leqq c(1+\mu)k^{-1}\|u-u^*\|,\\ &\|\tilde u-u^*\|\leqq c(1+\mu)k^{-1}\|u-u^*\| \quad \forall u\in S(u^*, \varepsilon). \end{aligned}$$

The proof is as in Theorem 5 of [29].

Proposition 2 allows us, in principle, to overcome the second difficulty. Indeed, if $k>0$ is large enough and $f_i(x)\in C^2$ then it is sufficient, beginning from some step S_0, to make only one step of the Newton or quasi-Newton method of smooth optimization (see [27], [28]) to obtain $\tilde z^s=(\tilde x^S\tilde u^S)$ which satisfies (11). The next lemmas allow us to obtain $z\in S(z^*, \varepsilon)$.

4. Lemmas. Let $f_i(x)\in C$, $X^*=\{x\colon F(x)=F^*\}$, $d(x, x^*)=\min\{\|x-y\|\,|\,y\in X^*\}$.

If (2) is satisfied then $F^*>-\infty$. Set $N(x, k)=(M(x, k))^{1/k}$; $F_k^*=\ln\inf_{x\in R^n} N(x, k)$; $X_k^*=\{x|\ln N(x, k)=F_k^*\}$.

LEMMA 1. *If* (2) *is satisfied then*

$$F^* \leqq F_k^* \leqq k^{-1} \ln m + F^*, \tag{12}$$

$$X_k^* \neq \varnothing, \quad \text{if} \quad k > (C - F^*)^{-1} \ln m, \tag{13}$$

$$\lim_{k\to\infty} \lim \{d(x, X^*) | x \in X_k^*\} = 0. \tag{14}$$

Proof. We have

$$\exp(F(x)) \leqq N(x, k) \leqq m^{1/k} \exp(F(x)) \quad \forall x. \tag{15}$$

Let $x^* \in X^*$. Then

$$\exp(F_k^*) = \inf_{x\in R^n} N(x, k) \leqq N(x^*, k) \leqq m^{1/k} \exp(F(x^*)) = \exp(k^{-1} \ln m + F^*).$$

The latter implies $F_k^* \leqq k^{-1} \ln m + F^*$. For $\varepsilon > 0$ and $x \in \{x: N(x, k) \leqq \inf_x N(x, k) + \varepsilon\}$ we have $\exp(F^*) \leqq \exp(F(x)) \leqq N(x, k) \leqq \inf_x N(x, k) + \varepsilon \leqq \exp(F_k^*) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, $F^* \leqq F_k^*$. Therefore (12) is proved.

Set

$$\Omega_k = \{x: F(x) \leqq k^{-1} \ln m + F^*\},$$

$$\bar{\Omega}_k = \{x: N(x, k) \leqq m^{1/k} \exp(F^*)\}.$$

The following inclusions hold: $X^* \subset \bar{\Omega}_k \subset \Omega_k \subset \Omega$. Indeed, consider $x^* \in X^*$. Then $N(x^*, k) \leqq m^{1/k} \exp(F^*)$, i.e., $x^* \in \Omega_k$. If $x \in \bar{\Omega}_k$ then $\exp(F(x)) \leqq N(x, k) \leqq m^{1/k} \exp(F^*)$. So $F(x) \leqq k^{-1} \ln m + F^*$, i.e., $x \in \Omega_k$. Finally, $k > (C - F^*)^{-1} \ln m$ implies that $F(x) \leqq k^{-1} \ln m + F^* \leqq C$ for $x \in \Omega_k$. So $\Omega_k \subset \Omega$ in view of (2).

Let $d(k) = \max\{d(x, X^*) | x \in \Omega_k\}$. It is obvious that $d(k) \to 0$ as $k \to \infty$. Therefore $\Omega_k \to X^*$ in the Hausdorff metric. The inclusion $X^* \subset \bar{\Omega}_k \subset \Omega_k \subset \Omega$ implies that $\bar{\Omega}_k \neq \varnothing$ and is bounded. Therefore the continuity of $N(x, k)$ implies (13). Also $X_k^* \subset \bar{\Omega}_k$ implies $X_k^* \subset \Omega_k$. Therefore, taking into account that $d(k) \to 0$, we obtain (14).

COROLLARY. *Let*

$$x(k) = \operatorname{argmin}\{M(x, k) | x \in R^n\},$$

$$u(k) = (u_i(k) = \exp(kf_i(x(k)))(\textstyle\sum \exp(kf_i(x(k))))^{-1}, \ i = \overline{1, m}), \quad z(k) = (x(k), u(k)).$$

If $X^* = x^*$ *and condition* (4) *holds, i.e.,* $z^* = (x^*, u^*)$ *is unique, then* $\lim_{k\to\infty} z(k) = z^*$. *Since* $M_x'(x(k), k) = \sum \exp(kf_i(x(k)))f_i'(x(k)) = 0$, *we have* $\sum u_i(k) f_i'(x(k)) = 0$, $\{x(k)\} \subset \Omega$, $\{u(k)\} \subset S_m$. *Therefore, for all* $\{z(k_i)\}$: $\lim_{k_i\to\infty} z(k_i) = \tilde{z}$, *we have* $\sum \tilde{u}_i f_i'(\tilde{x}) = 0$, $\tilde{u}_i(F(\tilde{x}) - f_i(\tilde{x})) = 0$; *i.e.,* $\tilde{z} = z^*$. *Taking into account that* z^* *is unique we get* $\lim_{k\to\infty} z(k) = z^*$.

Remark. Lemma 1 and some facts stated in [32] allow us to attach global character to the local results of Theorem 1 if we can obtain $x(k)$ for sufficiently large $k > 0$.

The estimate of distance between $x(k) \in X_k^*$ and x^* is related to uniqueness conditions for $z^* = (x^*, u^*)$.

LEMMA 2. *Suppose* $f_i(x) \in C^2$, $i = \overline{1, m}$, *and conditions* (2)-(5) *are satisfied. Then there exist* $k_0 > 0$ *and* $c > 0$ *which do not depend on* $k \geqq k_0$ *and are such that for all* $k \geqq k_0$ *we have the following*:

(1) *The estimates*

$$\|x(k) - x^*\| \leqq ck^{-1}, \qquad \|u(k) - u^*\| \leqq ck^{-1} \tag{16}$$

are true.

(2) *The function $M(x, k)$ is strongly convex in x in a neighborhood of $x(k)$.*

Proof. It follows from Lemma 1 that $x(k)$ exists if $k_0 > (c - F^*)^{-1} \ln m$. Consider the following functions in a neighborhood of x^* and for $\varkappa \in R^1$:

$$V_i(x, \varkappa) = \begin{cases} \exp(|\varkappa|^{-1} f_i(x)) \left(\sum_{j=1}^{m} \exp(|\varkappa|^{-1} f_j(x)) \right)^{-1}, & \varkappa \neq 0, \\ 0, & \varkappa = 0. \end{cases}$$

If $i > r$, then there is a $c > 0$ independent of $\varkappa$ such that $V_i(x, \varkappa) \leqq c \exp(-c|\varkappa|^{-1})$ for all $(x, \varkappa) \in S(x^*, \varepsilon) \times (-\varepsilon, \varepsilon)$. Therefore $V_i(x, \varkappa)$ is continuous in this neighborhood $(x^*, 0)$. We have $\| V_i'(x, \varkappa) \| = \| (V_{ix}'(\cdot), V_{i\varkappa}'(\cdot)) \| = \| |\varkappa|^{-1} V_i(\cdot)(f_i'(x) - \sum V_j(\cdot) f_j'(x)), -(\varkappa|\varkappa|)^{-1} V_i(\cdot)(f_i(x) - \sum V_j(\cdot) f_j(x)) \| \leqq c|\varkappa^{-2}| \exp(-c|\varkappa|^{-1})$. Setting $V_i'(x, 0) = 0$, $i > r$; $\rho(x, \varkappa) = \sum_{i=r+1}^{m} V_i(\cdot) f_i'(x)$, we see that $V_i(x, \varkappa)$ is smooth as well as $\rho(x, \varkappa)$, and $\rho(x, 0) = 0$, $\rho'(x, 0) = 0$, for all $(x, \varkappa) \in S(x^*, \varepsilon) \times (-\varepsilon, \varepsilon)$. Now consider the map $\bar{\Phi}(x, v, t, \varkappa) = (\sum_{i=1}^{r} v_i f_i'(x) + \rho(x, \varkappa)$; $f_i(x) - F^* + t - \varkappa \ln v_i$, $i = \overline{1, r}$; $\sum_{i=1}^{r} v_i + \sum_{i=r+1}^{m} V_i(x, \varkappa) - 1) : R^{n+r+2} \to R^{n+r+1}$. Taking into account the Kuhn-Tucker relations and the equalities $f_i(x^*) = F^*$, $i = \overline{1, r}$, $\rho(x^*, 0) = 0$, $\rho'(x^*, 0) = 0$, we conclude that $\bar{\Phi}(x^*, \bar{u}^*, 0, 0) = 0$ and $\bar{\Phi}_{xvt}' = \Phi_{x\hat{u}t}'$, so $\bar{\Phi}_{xvt}'$ is a nonsingular matrix. Therefore the implicit function theorem gives us the following: if $\varepsilon > 0$ is small enough, then there is a unique smooth vector-function $y(\varkappa) = (x(\varkappa), v(\varkappa), t(\varkappa)) \equiv y(\cdot)$: $\bar{\Phi}(x(\cdot), v(\cdot), t(\cdot), \cdot) \equiv \bar{\Phi}(\cdot) \equiv 0$; that is, $\sum_{i=1}^{r} V_i(\cdot) f_i'(x(\cdot)) + \rho(x(\cdot), \cdot) = 0$, $f_i(x(\cdot)) - F^* + t(\cdot) - \varkappa \ln V_i(\cdot) = 0$, $i = \overline{1, r}$ and $\sum_{i=1}^{r} V_i(\cdot) + \sum_{j=r+1}^{m} V_j(x(\cdot), \cdot) = 1$, $v(0) = \bar{u}^*$, $t(0) = 0$. Furthermore, from $\bar{\Phi}(\cdot) = 0$ we get $\bar{\Phi}_{xvt}'(\cdot) y'(\cdot) + \Phi_{\varkappa}^{-1}(\cdot) = 0$. Taking into account that $\bar{\Phi}_{xvt} = \Phi_{x\hat{u}t}'$, $f_i(x) \in C^2$, we obtain $\| \bar{\Phi}_{xvt}'^{-1}(\cdot) \| \leqq c_1$; then $\| \bar{\Phi}_{\varkappa}'(\cdot) \| \leqq c_2$, and moreover c_1 and c_2 are independent of $\varkappa$. Therefore $\| y'(\cdot) \| \leqq c$ for $c = c_1 \cdot c_2$; that is, $\| x'(\varkappa) \| \leqq c$, $\| v'(\varkappa) \| \leqq c$. Using the inequalities and setting $k = \varkappa^{-1}$, we obtain (16).

Now we prove the strong convexity $M(x, k)$ in a neighborhood of $x(k)$.

Let $U^* = \operatorname{diag} u_i^* : R^r \to R^r$, $U(k) = \operatorname{diag} u_i(k) : R^m \to R^m$. Then $M_{xx}''(x, k) = k(\sum \exp(kf_i(x)) f_i''(x) + \sum \exp(kf_i(x) f_i'^T(x) f_i''(x)))$ so that

$$\begin{aligned} M_{xx}''(x(k), k) &= M_{xx}''(\cdot, k) \\ &= (k \sum \exp(kf_i(\cdot))) \cdot (\sum u_i(k) f_i''(\cdot) + k \sum u_i(k) f_i'^T(\cdot) f_i'(\cdot)) \\ &= k \sum \exp(kf_i(\cdot)) (L_{xx}''(\cdot) + kf'^T(\cdot) U(k) f'(\cdot)). \end{aligned}$$

Using estimate (16), Proposition 1, and the fact that $f_i \in C^2$, we have, for $k > 0$ large enough, $(M_{xx}''(\cdot, k)\xi, \xi) \approx kM(x^*, k)((L_{xx}''(z^*) + k\bar{f}'^T(x^*) U^* \bar{f}'(x^*))\xi, \xi) \geqq \gamma \|\xi\|^2$, $\gamma > 0$, for all $\xi \in R^n$; therefore $M(x, k)$ is strongly convex in a neighborhood of $x(k)$.

Lemma 1 and 2 show that (10) can be realized with the estimate $\| z^s - z^* \| \leqq ck^{-s}$, starting, for example, from $u^0 = (m^{-1}, \cdots, m^{-1}) \in S_m$ if obtaining $x(k)$, $u(k)$ is possible for sufficiently large $k > 0$. Convexity $f_i(x)$, $i = \overline{1, m}$ is sufficient for this.

The rate of convergence can be improved by increasing k. But when k increases, the function $A(x, u, k)$ becomes ill-conditioned in x, which makes it more difficult to search for the minimum of $A(x, u, k)$. Therefore we cannot succeed in obtaining rapidly convergent processes using only $A(x, u, k)$. It appears that such processes can be obtained by applying important properties of the problem dual to (1). We shall now study these properties.

5. Dual problem. When proving Theorem 1 we found that there exists

$$x_k(u) = \operatorname*{argmin}_{x \in V^*} A(x, u, k)$$

where

$$V^* = \begin{cases} R^n & \text{when all the functions } f_i \text{ are convex} \\ V(k_0, \varepsilon) & \text{if not} \end{cases}$$

if k is sufficiently large and (3)-(5) are satisfied. Therefore we have a function $\varphi_k(u) = A(x_k(u), u, k)$ defined on $S(u^*, \varepsilon)$. For $u \in S_m \setminus S(u^*, \varepsilon)$ we set $\varphi_k(u) = \inf_x A(x, u, k)$. In the neighborhood $S(u^*, \varepsilon)$, smoothness properties of $\varphi_k(u)$ are determined by the corresponding properties of $f_i(x)$, $i = \overline{1, m}$. In particular,

$$\begin{aligned} \varphi'_{ku}(u) &= A'_x(x_k(u), u, k)x'_k(u) + A'_u(x_k(u), u, k) \\ &= A'_u(x_k(u), u, k) \\ &= k^{-1}(\exp(kf_1(x_k(u))), \cdots, \exp(kf_m(x_k(u)))), \end{aligned}$$

since $A'_x(x_k(u), u, k) = 0$. Furthermore, $\varphi''_{kuu}(u) = A''_{ux}(x_k(u), u, k)x'_k(u) = A''_{ux}(\cdot)x'_k(\cdot)$.

In order to determine $x'_k(u)$ we shall differentiate $A'_x(x_k(u), u, k) \equiv 0$ with respect to u. We obtain $A''_{xx}(\cdot)x'_k(u) + A''_{xu}(\cdot) = 0$. Therefore $x'_k(u) = (A''_{xx}(\cdot))^{-1}A''_{xu}(\cdot)$ and $\varphi''_{kuu}(u) = -A''_{ux}(\cdot)(A''_{xx}(\cdot))^{-1}A''_{xu}(\cdot)$. These formulas give the relation between smoothness properties of $\varphi_k(u)$ and corresponding properties of $f_i(x)$, $i = \overline{1, m}$. In particular, the continuity of $f''_i(x)$ implies the continuity of $\varphi''_{kuu}(u)$ and (6) implies that $\varphi''_{kuu}(u)$ satisfies a Lipschitz condition. Consider the problem dual to (1):

$$\tilde{u} = \operatorname{argmax}\{\varphi_k(u) | u \in S_m\}. \tag{17}$$

The following theorem holds.

THEOREM 2. *Let the conditions* (3)-(5) *be satisfied and* $f_i(x) \in C^2$, $i = \overline{1, m}$:

(1) *Then there exists* $k_0 > 0$ *such that for* $k \geqq k_0$ *the solution of the dual problem* (17) *exists and the strict form of sufficient optimality condition is satisfied for* (17).

(2) *There exists* $\varepsilon > 0$ *such that for* $k \geqq k_0$ *the function* $A(x, u, k)$ *is strongly convex in* $x \in S(x^*, \varepsilon)$, *concave in* u, *and has a unique saddle point on* $S(x^*, \varepsilon) \times S_m$, *that is,*

$$A(x, u^*, k) \geqq A(x^*, u^*, k) = \varphi_k(u^*) \geqq A(x^*, u, k). \tag{18}$$

(3) *If* $f_i(x)$, $i = \overline{1, m}$ *are convex, then* (18) *holds on* $R^n \times S_m$, *and if instead of* (5) *we assume strong convexity one of* $f_i(x)$, $i \in I^*$ *then conditions* (1)-(3) *hold for all* $k > 0$.

Proof. (1) Since $A(x, u, k)$ is strongly convex in x, there exists $x_k(u) = \operatorname{argmin}_{x \in V^*} A(x, u, k)$ and $\varphi'_k(u) = k^{-1}(\exp(kf_1(x_k(u))), \cdots, \exp(kf_m(x_k(u))))$ for all $u \in S(u^*, \varepsilon)$. Moreover, $x_k(u^*) = x^*$, $\varphi'_k(u^*) = k^{-1}(\exp kF^*, \cdots, \exp kF^*; \exp kf_{r+1}(x^*), \cdots, \exp kf_m(x^*))$. Consider the Lagrangian of (17):

$$L(u, \lambda) = \varphi_k(u) + \sum \lambda_i u_i + \lambda_0(\sum u_i - 1).$$

Setting $\lambda_0^* = -k^{-1} \exp kF^*$ and $\lambda_i^* = k^{-1}(\exp kF^* - \exp kf_i(x^*))$, we obtain that the Kuhn-Tucker relations for the problem (17) are satisfied at (x^*, λ^*):

$$k^{-1} \exp(kf_i(x^*)) + \lambda_i^* + \lambda_0^* = 0, \qquad i = \overline{1, m},$$

i.e.,

$$\mathcal{L}'_u(u^*, \lambda^*) = 0, \quad \lambda_i^* \geqq 0, \quad u_i^* \geqq 0, \quad \lambda_i^* u_i^* = 0, \quad \sum u_i^* = 1.$$

The function $\varphi_k(u)$ is concave no matter whether $f_i(x)$ are convex or not, so u^* is a solution of (17).

Let $e_i=(0,0,\cdots,0,\underline{1},0,\cdots,0)$, $e=(1,\cdots,1)\in R^m$. Since $F^*>f_i(x^*)$, $i=\overline{r+1,m}$, then $\lambda_i^*>0$, $i=\overline{r+1,m}$. The gradients of active constraints in (17), $e_{r+1},\cdots,e_m, e$ are linearly independent, i.e., for (17) the conditions of form (4) are satisfied. We shall now show that the conditions of form (5) are also satisfied for (17). They have the following form:

(19) $(e_i,v)=0$, $i=\overline{r+1,m}$, $(e,v)=0$ implies $-(\varphi''_{kuu}(u^*)v,v)\geqq\mu\|v\|^2$, $\mu>0$.

Since $\varphi''_{kuu}(u^*)=-A''_{ux}(A''_{xx})A''_{xu}$ at $x=x^*$ and $u=u^*$ we have $(-\varphi''_{kuu}(u^*)v,v)=(A''^{-1}_{xx}A''_{xu}v,A''_{xu}v)$. There (19) is equivalent to the condition

$$(e_i,v)=0,\quad i=\overline{r+1,m},\quad (e,v)=0 \text{ implies } \|A''_{xu}v\|\geqq\mu|v\|,\quad \mu>0,$$

i.e.,

$$v_i=0,\quad i=\overline{r+1,m},\quad \sum_{i=1}^{r}v_i=0 \text{ implies } \left\|\sum_{i=1}^{r}v_i\exp(kf_i(x^*))f_i'(x^*)\right\|\geqq\mu\|v\|.$$

But the last condition is equivalent to the condition

$$\sum_{i=1}^{r}v_i=0 \text{ implies } \left\|\sum_{i=1}^{r}v_if_i'(x^*)\right\|\geqq\mu\|v\|,\qquad \mu>0,$$

which is equivalent to (4), so the condition of type (5) for the dual problem (17) has been proved.

Therefore, for (17), the second-order sufficient conditions in a strict form (see, for example, [28, p. 47]) are satisfied if these conditions are satisfied for (1).

(2) The strong convexity of $A(x,u,k)$ in $x\in S(x^*,\varepsilon)$ for all $u\in S(u^*,\varepsilon)$ follows from Theorem 1 and the fact that $f_i(x)\in C^2$ if k is large enough.

In particular, $A''_{xx}(x^*,u^*,k)$ is a positive definite matrix, so together with $A'_x(x^*,u^*,k)=0$ it gives the left inequality (18) and uniqueness of x^*. Since $A(x,u,k)$ is a linear function of u, the function $\varphi_k(u)$ is concave in u no matter whether or not $f_i(x)$, $i=\overline{1,m}$ are convex and the uniqueness of u^* is a consequence of (4). The right-hand side of (18) follows from $A(x^*,u,k)\leqq k^{-1}\exp(kF^*)$, for all $u\in S_m$.

(3) The left inequality in (18) follows from the convexity $A(x,u^*,k)$ on x and (3); the right one follows from $A(x^*,u,k)\leqq k^{-1}\exp(kF^*)$, for all $u\in S_m$. The uniqueness of x^* follows from the strong convexity of $A(x,u^*,k)$ on x, and the uniqueness of u^* follows from (4).

COROLLARY. *The restriction of the cost function $\varphi_k(u)$ to the manifold of active constraints of the dual problem* (17) *is a strongly concave function, i.e., the restrictions of $\varphi_k(u)$ to the manifold $\{u=(u_1,\cdots,u_m)\geqq 0\colon u_i=0,\ i=\overline{r+1,m},\ \sum_{i=1}^{m}u_i=1\}$ are strongly concave.*

We shall consider this problem in more detail: Set

$$\bar\varphi_k(u)=\min_{x\in V^*}k^{-1}\sum_{i=1}^{r}u_i\exp(kf_i(x))=\min_{x\in V^*}A_r(x,u,k)\quad\forall u\in\bar S(u^*,\varepsilon),$$

$$\bar S(u^*,\varepsilon)=\{u=(u_1,\cdots,u_r)\geqq 0\colon \|u-u^*\|\leqq\varepsilon,\ u^*=(u_1^*,\cdots,u_r^*)\},$$

$$\mathfrak{L}=\{u=(u_i,\cdots,u_r)\colon \textstyle\sum u_i=1\}.$$

Let the matrix $P\colon R^r\to R^r$ be the orthogonal projector into $\mathfrak{L}$ of the vectors $u\in\bar S(u^*,\varepsilon)$ and let $\bar\varphi_{k\mathfrak{L}}(u)=\bar\varphi_k(Pu)$ be the restriction of the function $\bar\varphi_k(u)$ to $\mathfrak{L}$. Then $\bar\varphi'_{k\mathfrak{L}u}(u)=PA'_{ru}(\cdot)$ is the gradient of the restriction $\bar\varphi_k(u)$ to $\mathfrak{L}$. The Hessian of the restriction $\bar\varphi_k(u)$ to $\mathfrak{L}$ is defined by

$$\bar H_{k\mathfrak{L}}(u)=\bar H_{k\mathfrak{L}}(\cdot)=P\bar\varphi''_{kUU}(\cdot)P=-PA''_{rux}(\cdot)(A''_{rxx}(\cdot))^{-1}A''_{rxu}(\cdot)P.$$

In order for $-\bar{H}_{k\psi}(\cdot)$ to be positive definite it is sufficient (see [21]) that the matrix $A''_{rxx}(\cdot)$ be positive definite and $A''_{rxu}(\cdot)$ be nonsingular on Ω. The first property was proved in Theorem 1. Since $\bar{\varphi}_k(u)$ is considered in a small neighbourhood of u^*, we have $u_i > 0$, $i = \overline{1, r}$. Furthermore,

$$A''_{rxu}(\cdot) = \left(\sum_{i=1}^{r} u_i \exp(kf_i(x_k(u))) \right) T(u, k),$$

$$T(u, k) = (\hat{u}_1 u_1^{-1} f_1'(x_k(u)), \cdots, \hat{u}_r u_r^{-1} f_r'(x_k(u)))$$

so that for sufficiently large k we obtain $T(u^*, k) \approx (f_1'(x^*), \cdots, f_r'(x^*)) = \bar{f}'^T(x^*)$. Therefore the nonsingularity of the operator $A''_{rxu}(\cdot)$ on Ω is a consequence of condition (4).

Remark 1. Let I be any minimal set, and let u^* be a vertex of the Kuhn-Tucker polyhedron corresponding to I:

$$S_I(u^*, \varepsilon) = \left\{ u : \sum_{i \in I} u_i = 1, u_i \geqq 0, i \in I, u_i = 0, i \bar{\in} I, \|u - u^*\| \leqq \varepsilon \right\}.$$

Then the strong concavity property for $\varphi_{Ik}(u) = \min_{x \in V^*} k^{-1} \sum_{i \in I} u_i \exp(kf_i(x))$ on $S_I(u^*, \varepsilon)$ still holds if we replace (5) by (5').

Remark 2. The results of Theorems 1 and 2 are not true for the classical Lagrangian function $L(x, u)$ corresponding to the original problem (1), since (4)-(5) do not ensure in general that $L(x, u)$ is convex in x and $\operatorname{argmin}_{x \in V^*} L(x)$ may not exist or may not coincide with x^*.

Finally, note that $\varphi_k(u)$ is concave, and that if conditions (4)-(6) are satisfied, then $\varphi_k(u)$ is strongly concave on Ω in a neighbourhood of u^* with its Hessian satisfying a Lipschitz condition. These properties are used to find an approximation for the Lagrange multipliers. Consider this problem in more detail.

As a result of one step of (10) for sufficiently large k we can obtain $u \in S(u^*, \varepsilon)$ and isolate the set of active constraints I^*. After that, the search for u^* is reduced to the determination of $u^* = (u_1^*, \cdots, u_y^*)$

$$u^* = \operatorname{argmax} \{\bar{\varphi}_k(u) | u \in \bar{S}(u^*, \epsilon) \cap \Omega\}. \tag{20}$$

Every method of smooth optimization of $\bar{\varphi}_k(u)$ on Ω determines some relaxation operator $R : \bar{S}(u^*, \varepsilon) \cap \Omega \to \bar{S}(u^*, \varepsilon) \cap \Omega$, i.s., $\bar{\varphi}_k(Ru) > \bar{\varphi}_k(u)$ and $R^S u \to u^*$, for all $u \in \bar{S}(u^*, \varepsilon) \cap \Omega$. In particular, the gradient method has its relaxation operator defined by

$$Ru = u + tPA'_{ru}(\cdot), \qquad t > 0, \tag{21}$$

and the relaxation operator corresponding to the Newton method is

$$Ru = u + \xi \tag{22}$$

where ξ is the normal solution of the system

$$\bar{H}_{k\psi}(u)\xi = -PA'_{ru}(\cdot).$$

The properties of these operators are determined by the properties of the corresponding methods of smooth optimization. The convergence rates of the approximation of u_i^*, $i \in I^*$ and, consequently, x^* are determined not only by the rate of convergence of the corresponding method of smooth optimization but also by the estimate from (9) because the optimization of $\bar{\varphi}_k(u)$ on Ω is always accompanied by a step of (10).

Under the conditions of Theorem 1 the matrix $\bar{H}_{k\nu}(u)$ is continuous and has a negative spectrum. Therefore for the operator (22) we obtain a sequence $\{u^{S+1} = Ru^S\}_{S=0}^{\infty}$ such that $\|u^{S+1} - u^*\| \leqq q_S \|u^S - u^*\|$, $q_S \to 0$. If condition (6) is satisfied then there exists $\lambda > 0$: $\|u^{S+1} - u^*\| \leqq \lambda \|u^S - u^*\|^2$. The use of Newtonian and quasi-Newtonian methods allows us to obtain relaxation operators with corresponding properties of convergence. This makes it possible to formulate the following general method, which we shall consider for the convex $f_i(x)$, $i = \overline{1, m}$:

Let $u^0 = (m^{-1}, \cdots, m^{-1}) \in R^m$ and let $k > 0$ be large enough. Assume that $z^S = (x^S, u^S)$ has already been found. Then define $z^{S+1} = (x^{S+1}; u^{S+1})$:

$$x^{S+1} = \operatorname{argmin} \{A_r(x, Ru^S, k) | x \in R^n\}, \tag{23}$$

$$u^{S+1} \times \left(u_i^{S+1} = \bar{u}_i^S \exp\left(kf_i(x^{S+1})\right) \left(\sum_{i=1}^{r} \bar{u}_i^S \exp\left(kf_i(x^{S+1})\right) \right)^{-1}, \; i = \overline{1, r} \right) \tag{24}$$

where $\bar{u}^S = (\bar{u}_1^S, \cdots, \bar{u}_r^S) = Ru^S$.

Let operator R have one of the following properties:

$$(1°) \qquad \|Ru - u^*\| \leqq q|u - u^*\|, \qquad q < 1,$$

$$(2°) \qquad \|Ru - u^*\| \leqq q(u)\|u - u^*\|, \; q(u) \to 0 \quad \text{as } u \to u^*,$$

$$(3°) \qquad \|Ru - u^*\| \leqq \lambda \|u - u^*\|^2.$$

If $f_i(x)$, $i = \overline{1, m}$ is convex, (3)-(6) are satisfied, and k is large enough, then for the sequences $\{z^S\}_{S=0}^{\infty}$ generated by method (23), (24), using relaxation operators with properties (1°)-(3°) we obtain the following estimation:

$$(1^{\infty}) \qquad \|z^S - z^*\| \leqq (cqk^{-1})^S;$$

$$(2^{\infty}) \qquad \|z^S - z^*\| \leqq (ck^{-1})^S \prod_{i=1}^{S} q_i q_S \to 0 \quad \text{as } S \to \infty;$$

$$(3^{\infty}) \qquad \|z^S - z^*\| \leqq (c\lambda k^{-1})^{2^S - 1} q_0^{2^S}, \qquad q_0 < 1.$$

The last estimates follow directly from the properties of operators R and the inequalities

$$\|x^{S+1} - x^*\| \leqq \frac{c}{k} \|Ru^S - u^*\|, \qquad \|u^{S+1} - u^*\| \leqq \frac{c}{k} \|Ru^S - u^*\|$$

which follow from Theorem 1.

Acknowledgments. It is my pleasure to express my gratitude to Drs. G. D. Maistrovsky and M. E. Primak, and to the reviewers whose remarks contributes to the improvement of this work. I am grateful to Prof. D. Bertsekas, who kindly sent me many of his results.

REFERENCES

[1] K. Arrow, L. Hurwicz, and H. Uzava, *Studies in Linear and Nonlinear Programming*, Stanford University Press, Stanford, CA, 1958.

[2] D. Bertsekas, *Nondifferential optimization via approximation*, Math. Programming Stud., 3 (1975), pp. 1-25.

[3] ———, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1981.

[4] ———, *Approximation procedures based on the method of multipliers*, J. Optim. Theory Appl., 23 (1977), pp. 487-510.

[5] ———, *A new algorithm for solution of resistive networks involving diodes*, IEEE Trans. Circuits and Systems, CS-23 (1976), pp. 599-608.

[6] D. BERTSEKAS, G. LAUER, N. SANDELL, AND T. POSBERCH, *Optimal short-term scheduling of large scale power system*, IEEE Trans. Automat. Control, AC-28 (1983), pp. 432-443.

[7] D. BERTSEKAS, *A convergence analysis of the method of multipliers for nonconvex constrained optimization*, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1978.

[8] C. CHARALAMBOUS, *Nonlinear least Pth optimization and nonlinear programming*, Math. Programming, 12 (1977), pp. 195-225.

[9] ———, *Acceleration of the least Pth algorithm for nonlinear programming*, Math. Programming, 17 (1979), pp. 270-297.

[10] V. A. DAUGAVET AND V. N. MALOZYOMOV, *Quadratic convergence of some method of linearization for discrete minimax problems*, U.S.S.R. Comput. Math. and Math. Phys., 21 (1981), pp. 435-443.

[11] V. DEMYANOV AND V. MALOZYOMOV, *Introduction to Minimax*, John Wiley, New York, 1974.

[12] A. V. FIACCO AND G. MCCORMICK, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley, New York, 1968.

[13] E. G. GOLSHTEIN AND N. V. TRETYAKOV, *On augmented Lagrangians for convex programming*, Econom. Math. Methods, 16 (1980), pp. 772-777.

[14] K. GROSSMAN AND A. KAPLAN, *Nonlinear Programming Based on Unconstrained Optimization*, Nauka, Moscow, 1981. (In Russian.)

[15] S. P. HAN, *Variable metric methods for minimizing a class of nondifferentiable functions*, Math. Programming, 20 (1981), pp. 1-13.

[16] J. HALD AND K. MADSEN, *Combined LP and quasi Newton methods for minimax optimization*, Math. Programming, 20 (1981), pp. 49-62.

[17] L. V. KANTOROVICH AND G. P. AKILOV, *Functional Analysis*, Nauka, Moscow, 1973.

[18] L. G. KHACHIAN, *Polynomial algorithm in linear programming*, Soviet Math. Dokl., 244 (1979), pp. 1093-1096.

[19] B. KORT AND D. BERTESEKAS, *A New Penalty Function for Constrained Minimization*, Proc. IEEE Conference on Decision and Control, New Orleans, LA, 1972.

[20] ———, *Combined primal-dual penalty method for convex programming*, SIAM J. Control Optim., 14 (1976), pp. 268-294.

[21] T. D. MAISTROVSKY AND U. V. OLKHOVSKY, *On the convergence of the method of the steepest descent in constrained optimization problems*, U.S.S.R. Comput. Math. and Math. Phys., 15 (1975), pp. 844-859.

[22] T. S. MOTZKIN, *New technique for linear inequalities and optimization*, in Project SCOOP Symposium on Linear Inequalities and Programming, Planning Research Division, U.S. Air Force, Washington, D.C., 1952.

[23] A. S. NEMIROVSKY AND D. B. YUDIN, *Complexity of Problems and Effectiveness of Optimization Methods*, Nauka, Moscow, 1979. (In Russian.)

[24] ———, *Nonsmooth optimization*, in Proc. IIASA Workshop, C. Lemarechal and P. Mifflin, eds., Pergamon Press, Oxford, 1978.

[25] ———, *Nondifferentiable optimization*, Math. Programming Stud., 3 (1975).

[26] E. A. NURMINSKY, *Numerical Methods for Deterministic and Stochastic Minimax Problems*, Naukova Dumka, Kiev, 1979. (In Russian.)

[27] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative solution of Nonlinear Equations of Several Variables*, Academic Press, New York, 1970.

[28] V. N. PSHENICHNY AND U. M. DANILIN, *Numerical methods in extremal problems*, MIR, Moscow, 1978.

[29] B. T. POLYAK AND N. N. TRETYAKOV, *Penalty method for conditional extremum problems*, U.S.S.R. Comput. Math. and Math. Phys., 13 (1973), pp. 34-46.

[30] R. A. POLYAK, *Smooth optimization methods for solving nonlinear extremal and equilibrium problems with constrains*, Abstracts of the papers, Eleventh International Symposium on Mathematical Programming, Bonn, 1982.

[31] ———, *Smooth optimization method in discrete minimax problems*, Ann. New York Acad. Sci. (1983), pp. 133-135.

[32] ———, *On the best convex Chebyshev approximation*, Soviet Math. Dokl., 12 (1971), pp. 538-540.

[33] ———, *Controlled processes in the extremal and equilibrium problems*, Dept. Manuscript, Viniti, Moscow, 1986. (In Russian.)

[34] R. ROCKAFELLAR, *The multipliers method of Hestens and Powell applied to convex programming*, J. Optim. Theory Appl., 12 (1973), pp. 555-562.

[35] N. Z. SHOR, *Minimization Methods for Nondifferentiable Functions*, Lecture Notes in Computat. Math., Springer-Verlag, Berlin, New York, 1985.