Control
and Cybernetics
VOL. 12 (1983) No. 1–2

A phase I-phase II method for inequality constrained minimax problems

by

KRZYSZTOF CZESŁAW KIWIEL
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warszawa, Poland

We develop a class of methods for minimizing, subject to a finite number of inequality constraints, a nondifferentiable function which is the pointwise maximum of a finite collection of continuously differentiable functions. The methods proceed by solving iteratively quadratic programming problems to generate search directions. Several practical stepsize procedures are also introduced. The method does not require a feasible initial approximation to a solution. Global convergence of the algorithms is established. Under additional convexity assumptions the method is at least linearly convergent. The algorithm is conceptually simple and easy to implement. It generalizes several robust and efficient feasible point methods for standard nonlinear programming calculations.

1. Introduction

In this paper we present an implementable method for solving optimization problems of the following type: minimize $f_0(x)$ subject to $f(x) \leq 0$, where $x \in R^N$ and f_0 and f are real-valued functions that are the pointwise maxima of two finite collections of continuously differentiable functions. The problem is usually called a constrained minimax problem in the literature and abounds with applications [2, 3, 6, 9, 15].

Recently much research has been conducted in the area of minimax optimization. As expected, almost all the methods proposed until now tend to exploit the knowledge which is available for the differentiable case, i.e. when f_0 is continuously differentiable. Demyanov's minimax methods [2] are based on Zoutendijk's [16] and Polak's [12] feasible direction methods for solving standard nonlinear programming problems. The algorithm of Madsen and Schjaer-Jacobsen [6] for linearly constrained minimax optimization generalizes one of Zoutendijk's feasible direction methods [11, 16]. In [9] Panin presented an extension of Pshenichnyi's method of linearizations [14]. In this paper we introduce methods that generalize three well-known methods for standard nonlinear programming calculations: Pironneau and Polak's method of centers [10] and their feasible direction method [11], and the phase I — phase II feasible direction method of Polak, Trahan and Mayne [14]. These last three methods are considered to be the most robust and efficient among the feasible point methods for standard nonlinear programming calculations [13, 14]. Therefore our methods are potentially superior to the methods of Demyanov [2] and Madsen and Schjaer-Jacobsen [6].

Our methods combine, extend and modify ideas contained in [3, 10, 14, 15]. Their derivation is based on the application of quadratic approximation methods [3, 15] to the improvement function used in the modified method of centers [10, 12]. The algorithms are iterative in nature. They have search direction finding subproblems that are quadratic programming problems obtained by making a natural piecewise linearization of the problem functions. The matrices in the quadratic programming subproblems are preferably updated according to the rules used in variable metric algorithms [3, 15]. Specific variable metric updating schemes are not discussed here; they are a subject of on-going research.

When an initial approximation to a solution is feasible, the algorithms proceed as feasible point methods and the objective function f_0 need not be evaluated at infeasible points. This is important in certain applications [6]. When the initial approximation is infeasible, the methods decrease constraint violation at each iteration while not completely ignoring the objective function. Each method generates a sequence of points whose accumulation points satisfy the necessary conditions of optimality if some regularity assumption on the gradients of constraint functions outside the feasible set holds. However, we do not require that the optimization problem be normal, cf. [3], which is necessary for convergence of Panin's method [9] that uses an exact penalty function. Under additional convexity assumptions on the problem, we show that the methods construct a sequence of points converging to the solution with a linear rate. This seems to be the first such result for feasible point methods for nonlinearly constrained minimax optimization.

A further modification of the ideas presented in this paper has lead to a new implementable algorithm [5] for solving nondifferentiable problems of a more general nature, when f_0 and f are semismooth [7].

In section 2 we state the problem considered and its necessary conditions of optimality. The search direction finding subproblems are discussed in section 3. In section 4 we present the method and comment on its possible implementations. Section 5 contains results on global convergence of the methods. In section 6 we introduce additional regularity assumptions on the problem and establish linear rate of convergence of the algorithms. Section 7 provides some modifications. Finally, we have a conclusion section.

R^N denotes the N-dimensional Euclidean space with the usual inner product $\langle \cdot, \cdot \rangle$ and the associated norm $|\cdot|$. $\|\cdot\|$ denotes the associated norm of a matrix. We use x_i to denote the i-th component of the vector x. Superscripts are used to denote different vectors, e.g. x^1 and x^2. All vectors are column vectors and a row vector is denoted by the superscript T. However, for convenience a column vector in R^{N+n} is sometimes denoted by (x, y) even though x and y are column vectors in R^N and R^n respectively. If F is a real valued function on R^N, then F' denotes the gradient of F and F'' the Hessian.

2. Problem statement

Consider the following optimization problem

$$\text{minimize } f_0(x), \text{ subject to } f(x) \leqslant 0, \tag{2.1}$$

where

$$f_0(x) = \max_{i \in I_0} f_{0,i}(x), \quad f(x) \equiv \max_{i \in I} f_i(x) \tag{2.2}$$

and $f_{0,i}$: $R^N \to R^1$, $i \in I_0 = \{1, 2, \ldots, n\}$, and f_i: $R^N \to R^1$, $i \in I = \{1, 2, \ldots, m\}$, are continuously differentiable; $N, m, n < +\infty$.

REMARK 2.1. If the initial approximation x^1 to a solution of (2.1) is feasible, i.e. $x^1 \in S = \{x \in R^N : f(x) \leqq 0\}$, then we may only require that $f_{0,i}$, $i \in I_0$, be defined and continuously differentiable on S.

Let

$$W = \Big\{ w \in R^{m+n} : w \geqslant 0 \text{ and } \sum_{i=1}^{m+n} w_i = 1 \Big\}, \tag{2.3}$$

We say that a point $\bar{x} \in R^N$ satisfies the Fritz-John necessary conditions of optimality for the problem (2.1) if there exists a Lagrange multiplier $\bar{w}$ satisfying

$$\sum_{i \in I_0} \bar{w}_i f'_{0,i}(\bar{x}) + \sum_{i \in I} \bar{w}_{i+n} f'_i(\bar{x}) = 0, \tag{2.4a}$$

$$w \in W, \tag{2.4b}$$

$$\bar{w}_i [f_{0,i}(\bar{x}) - f_0(\bar{x}) - f(\bar{x})_+] = 0, \quad i \in I_0, \tag{2.4c}$$

$$\bar{w}_{i+n} [f_i(\bar{x}) - f(\bar{x})_+] = 0, \quad i \in I, \tag{2.4d}$$

$$f(\bar{x})_+ = 0, \tag{2.4e}$$

where $f(x)_+ \equiv \max\{f(x), 0\}$. Recall that (2.4) is a necessary condition for $\bar{x}$ to solve (2.1), cf. [1, 2]. We denote by $W(\bar{x})$ the set of all multipliers $\bar{w}$ satisfying (2.4).

We shall also consider an auxiliary problem

$$\text{minimize } f(x). \tag{2.5}$$

If $\bar{x}$ is a solution of the above problem, then there exists a multiplier $\bar{w}$ satisfying

$$\sum_{i\in I} \bar{w}_{i+n} f_i'(\bar{x})=0, \tag{2.6a}$$

$$\bar{w}\in W, \quad \bar{w}_i=0, \quad i\in I_0, \tag{2.6b}$$

$$\bar{w}_{i+n}[f_i(\bar{x})-f(\bar{x})]=0, \quad i\in I. \tag{2.6c}$$

Note that when $(\bar{x}, \bar{w})$ satisfies (2.6) and $f(\bar{x})=0$, then $(\bar{x}, \bar{w})$ is a Fritz-John point satisfying (2.4). To exclude this situation, one may employ the following Cottle constraint qualification, cf. [4]:

$$\text{if } f(\bar{x})\geqq 0 \text{ then there does not exist any } \bar{w} \text{ satisfying (2.6)}. \tag{2.7}$$

If the Cottle constraint qualification holds at $\bar{x}$, then the Fritz-John conditions (2.4) reduce to the Kuhn-Tucker conditions [1, 4]: there exists a Kuhn-Tucker multiplier $(\bar{u}, \bar{v})\in R^n\times R^m$ satisfying

$$\sum_{i\in I_3} \bar{u}_i f_{0,i}'(\bar{x})+\sum_{i\in I} \bar{v}_i f_i'(\bar{x})=0, \tag{2.8a}$$

$$(\bar{u}, \bar{v})\in UV, \tag{2.8b}$$

$$\bar{u}_i[f_{0,i}(\bar{x})-f_0(\bar{x})]=0, \quad i\in I_0, \tag{2.8c}$$

$$\bar{v}_i f_i(\bar{x})=0, \quad i\in I, \tag{2.8d}$$

$$f(\bar{x})\leqq 0, \tag{2.8e}$$

where $UV=\{(u, v)\in R^n\times R^m: (u, v)\leqq 0, \sum_{i=1}^{n} u_i=1\}$. To see this, note that (2.4) and (2.7) imply that

$$\bar{w}_0=\sum_{i\in I_0} \bar{w}_i>0 \text{ for any } \bar{w}\in W(\bar{x}), \tag{2.9}$$

hence we may put

$$\bar{u}_i=\bar{w}_i/\bar{w}_0, \quad i\in I_0, \quad \bar{v}_i=\bar{w}_{i+n}/\bar{w}_0, \quad i\in I, \tag{2.10}$$

to obtain (2.8) from (2.4). We denote by $UV(\bar{x})$ the set of all Kuhn-Tucker multipliers satisfying (2.8). Note that under the constraint qualification (2.7) this set is nonempty and bounded at any Kuhn-Tucker point $\bar{x}$ [8].

REMARK 2.2. Instead of the Cottle constraint qualification, one may assume a weaker condition at $\bar{x}$ satisfying the Fritz-John necessary condition of optimality:

$$w_0=\min\Big\{\sum_{i=I_0} \bar{w}_i: \bar{w}\in W(\bar{x})\Big\}>0. \tag{2.11}$$

As noted above, (2.7) implies (2.9), which in turn is equivalent to (2.11), since W and $W(\bar{x})$ are compact. It is straightforward to check that if (2.11) holds, then the necessary conditions of optimality (2.4) and (2.8) are equivalent in the sense that one may put $W(\bar{x})$ and $UV(\bar{x})$ into correspondence by using (2.10) and

$$\bar{w}_i=\bar{u}_i/\Big(\sum_{j\in I_0}\bar{u}_j+\sum_{j\in I}\bar{v}_j\Big),\quad i\in I_0,\quad \bar{w}_{i+n}=\bar{v}_i/\Big(\sum_{j\in I_0}\bar{u}_j+\sum_{j\in I}\bar{v}_j\Big). \tag{2.12}$$

REMARK 2.3. If the constraint functions f_i are convex, then the Cottle constraint qualification reduces to the Slater constraint qualification (cf. [2,4]):

$$\text{there exists a point } \tilde{x} \text{ satisfying } f(\tilde{x})<0. \tag{2.13}$$

If additionally f_0 is convex, e.g. $f_{0,i}$, $i\in I_0$, are convex, then any Fritz-John point $\bar{x}$ is a Kuhn-Tucker point and any Kuhn-Tucker point solves the problem (2.1), see [2].

We end this section by remarking that only certain subsequent results require any constraint qualification. This is always explicitly stated. On the other hand, various other assumptions are assumed to hold implicitly throughout the remainder of the paper.

3. The search direction finding subproblem

The method to be presented in the next section uses search directions generated as follows. Let $x\in R^N$ and $\delta>0$ be given. We introduce two activity sets

$$I_0(x,\delta)=\{i\in I_0: f_{0,i}(x)-f_0(x)\geqq f(x)_+-\delta\}, \tag{3.1}$$

$$I(x,\delta)=\{i\in I: f_i(x)\geqq f(x)_+-\delta\}.$$

Let B be a positive definite symmetric $N\times N$-matrix. Then the following search direction finding subproblem

$$\underset{(d_0,d)\in R^1\times R^N}{\text{minimize}}\ \frac{1}{2}\langle Bd,d\rangle+d_0 \tag{3.2}$$

$$\text{subject to } f_{0,i}(x)-f_0(x)-f(x)_++\langle f'_{0,i}(x),d\rangle\leqq d_0,\quad i\in I_0(x,\delta),$$

$$f_i(x)-f(x)_++\langle f'_i(x),d\rangle\leqq d_0,\quad i\in I(x,\delta),$$

is a quadratic programming problem which satisfies the Slater constraint qualification, see [14, p. 259] and [15]; its solution $(d_0(x), d(x))$ exists and is uniquely determined by the following set of conditions:

$$d(x)=-B^{-1}\Big\{\sum_{i\in I_0(x,\delta)}\hat{w}_i f'_{0,i}(x)+\sum_{i\in I(x,\delta)}\hat{w}_{i+n}f'_i(x)\Big\}, \tag{3.3}$$

$$d_0(x)=\max\Big\{\max_{i\in I_0(x,\delta)}[f_{0,i}(x)-f_0(x)-f(x)_++\langle f'_{0,i}(x),d(x)\rangle],$$

$$\max_{i\in I(x,\delta)}[f_i(x)-f(x)_++\langle f'_i(x),d(x)\rangle]\Big\},$$

where $\hat{w}_i$, $i \in I_0(x,\delta)$, $\hat{w}_{i+n}$, $i \in I(x,\delta)$, are the possibly nonunique Lagrange multipliers for (3.2), which satisfy

$$\hat{w}_i \geqq 0, \quad i \in I_0(x,\delta), \quad \hat{w}_{i+n} \geqq 0, \quad i \in I(x,\delta), \tag{3.4}$$

$$\sum_{i \in I_0(x,\delta)} \hat{w}_i + \sum_{i \in I(x,\delta)} \hat{w}_{i+n} = 1,$$

$$\hat{w}_i [f_{0,i}(x) - f_0(x) - f(x)_+ + \langle f'_{0,i}(x), d(x)\rangle - d_0(x)] = 0, \quad i \in I_0(x,\delta),$$

$$\hat{w}_{i+n} [f_i(x) - f(x)_+ + \langle f'_i(x), d(x)\rangle - d_0(x)] = 0, \quad i \in I(x,\delta).$$

If we define

$$\hat{w}_i = 0, \quad i \in I_0 \setminus I_0(x,\delta), \quad \hat{w}_{i+n} = 0, \quad i \in I \setminus I(x,\delta), \tag{3.5}$$

then we obtain from (3.3) and (3.4), cf. [15],

$$d(x) = -B^{-1} \sum_{i \in I_0} \hat{w}_i f'_{0,i}(x) + \sum_{i \in I} \hat{w}_{i+n} f'_i(x), \tag{3.6}$$

$$\hat{w} \in W,$$

$$\hat{w}_i [f_{0,i}(x) - f_0(x) - f(x)_+ + \langle f'_{0,i}(x), d(x)\rangle - d_0(x)] = 0, \quad i \in I_0,$$

$$\hat{w}_{i+n} [f_i(x) - f(x)_+ + \langle f'_i(x), d(x)\rangle - d_0(x)] = 0, \quad i \in I,$$

and

$$-d_0(x) = \langle Bd(x), d(x)\rangle + \sum_{i \in I_0} \hat{w}_i [f_0(x) - f_{0,i}(x) + \sum_{i \in I} \hat{w}_{i+n} [f(x)_+ - f_i(x)]. \tag{3.7}$$

The Lagrange multiplier $\hat{w}$ also solves the following dual search direction subproblem

$$\underset{w \in R^{m+n}}{\text{minimize}} \frac{1}{2} \Big| \sum_{i \in I_0(x,\delta)} w_i f'_{0,i}(x) + \sum_{i \in I(x,\delta)} w_{i+n} f'_i(x) \Big|_H^2 +$$

$$+ \sum_{i \in I_0(x,\delta)} w_i [f_0(x) - f_{0,i}(x) + f(x)_+] + \sum_{i \in I(x,\delta)} w_{i+n} [f(x)_+ - f_i(x)],$$

$$\text{subject to } w \in W, \quad w_i = 0, \quad i \in I_0 \setminus I_0(x,\delta), \quad w_{i+n} = 0, \quad i \in I \setminus I(x,\delta),$$

where $H = B^{-1}$ and $|x|_H^2 = \langle Hx, x\rangle$ for any $x \in R^N$. Thus it may be more efficient to solve (3.8) and to recover $d_0(x)$ and $d(x)$ from (3.6) and (3.7). Hence it may be easier to work with $H = B^{-1}$ rather than with B.

Now we compare the above search direction finding subproblems with existing methods. When $n = 1$. i.e. when (2.1) is a standard nonlinear programming problem, then the subproblems (3.2) and (3.8) are modifications of the subproblems employed in the phase I-phase II feasible direction methods of Polak, Trahan and Mayne [14], which have $B = H = I$ (an identity matrix) and $\delta = +\infty$. If additionally $f(x) \leqq 0$, then we obtain the subproblems of Pironneau and Polak's methods of centers [10] and feasible direction methods [11]. Next, suppose that

$n \geqq 2$ and $m=0$, i.e. (2.1) is an unconstrained minimax problem. Then we have $f(x)_+=0$ and $I=I(x,\delta)=\varnothing$. In this case, the direction $d(x)$ is computed as in Pshenichnyi's method of linearizations [14] if $B=I$; when B approximates the Hessian of an appropriate Lagrange function for (2.1), then $d(x)$ is equal to the direction obtained in Wierzbicki's quadratic approximation method [16]; if additionally $\delta=+\infty$, i.e. $I_0(x,\delta)=I_0$, then we have the subproblem of Han's variable metric minimax method [3].

In general, (3.2) may be viewed as a quadratic approximation subproblem for the function

$$\bar{p}(\tilde{x}; x)=\max\{f_0(\tilde{x})-f_0(x), f(\tilde{x})\}, \tag{3.9}$$

which is the improvement function of the method of centers [10, 12], and $d_0(x)$ may be interpreted as an approximate directional derivative of $\bar{p}$ at x in the direction $d(x)$, cf. [3, 14, 15].

4. The method

In this section we describe the method for solving the problem (2.1) and comment on its possible implementations.

The algorithm

Step 0 (Initial data). Select a starting point $x^1 \in R^N$, a final accuracy tolerance $\varepsilon_f \geqq 0$, an activity lower bound $\bar{\delta}>0$. a desired rate of convergence parameter $\gamma \in [0,1)$, line search parameters $m \in (0,1)$ and $0<m_1<m_2<1$. Choose initial values of an activity variable $\delta^1 \geqq \bar{\delta}$ and a convergence variable $\eta^1 \geqq 0$. Let $B_1=H_1=I$ (the identity matrix). Set $k=1$.

Step 1 (Direction finding). Compute $(d_0^k, d^k)=(d_0(x^k), d(x^k))$ by solving (3.2) with $x=x^k$, $B=B_k$ and $\delta=\delta^k$ or, equivalently, by solving (3.8) with $x=x^k$, $H=H_k$ and $\delta=\delta^k$ and then using (3.6) and (3.7).

Step 2 (Convergence test). If $|d_0^k| \leqq \varepsilon_f$, terminate.

Step 3 (Testing direct prediction). Let the improvement function at the k-th iteration be defined by

$$\rho_k(x)=\max\{f_0(x)-f_0(x^k), f(x)\} \quad \text{if} \quad f(x^k) \leqq 0, \tag{4.1}$$

$$\rho_k(x)=f(x) \quad \text{if} \quad f(x^k)>0.$$

If $\rho_k(x^k+d^k)<\rho_k(x^k)$ and $|d_0^k| \leq \gamma\eta^k$, set the stepsize coefficient $t^k=1$ and go to Step 6. Otherwise go to Step 4.

Step 4 (Line search). Compute a stepsize coefficient $\tilde{t}^k>0$ satisfying at least one of the following requirements:

(i) (Armijo's rule). $\tilde{t}^k=2^{-i_k}$ where i_k is the smallest number $i=0, 1, 2, \ldots,$ for which $\rho_k(x^k+2^{-i} d^k) \leq \rho_k(x^k)+m\, 2^{-i} d_0^k$;

(ii) (Goldstein's rule). If $f(x^k)>0$, then $\tilde{t}^k$ satisfies either

$$\rho_k(x^k)+m_2\,\tilde{t}^k d_0^k \le \rho_k(x^k+\tilde{t}^k d^k) \le \rho_k(x^k)+m_1\,\tilde{t}^k d_0^k, \tag{4.2}$$

or $f(x^k+\tilde{t}^k d^k)\le 0$. If $f(x^k)\le 0$, then $\tilde{t}^k$ satisfies (4.2).

Step 5 (Approximate or exact minimization at line search). Let $\tilde{x}^k=x^k+\tilde{t}^k d^k$. Compute the stepsize coefficient $t^k>0$ satisfying

$$f(x^k+t^k d^k)\le f(\tilde{x}^k) \quad \text{if} \quad f(\tilde{x}^k)>0, \tag{4.2a}$$

$$\max\{f_0(x^k+t^k d^k)-f_0(x^k), f(x^k+t^k d^k)\}\le \max\{f_0(\tilde{x}^k)-f_0(x^k), f(\tilde{x}^k)\}$$
$$\text{if} \quad f(\tilde{x}^k)\le 0. \tag{4.2b}$$

Step 6. Set $x^{k+1}=x^k+t^k d^k$ and $\eta^{k+1}=\max\{\gamma\eta^k, \min(\eta^k, -d_0^k)\}$. Choose $\delta^{k+1}\ge\bar{\delta}$ and a positive definite symmetric $N\times N$ matrix B_{k+1} or $H_{k+1}=B_{k+1}^{-1}$. Increase k by 1 and go to Step 1.

A few comments on the implementation of the algorithm are in order.

(a) The magnitude of δ^k controls the number of constraints in the direction finding subproblems (3.2) and (3.8). Note that the gradients $f'_{0,i}(x^k)$ and $f'_i(x^k)$ must be computed only for δ^k-active functions, i.e. $i\in I_0(x^k,\delta^k)$ and $i\in I(x^k,\delta^k)$ respectively. Thus small values of δ^k can reduce the computational effort per iteration. On the other hand, very small δ^k may result in too small a stepsize. The lower bound $\bar{\delta}>0$ prevents this jamming. It also establishes a threshold for determining the functions probably active at a solution. This suggests the following strategy for changing δ (cf. [16]):

$$\delta^{k+1}=\max\{\bar{\delta}, \zeta(\eta^{k+1})^{1/2}\},$$

with $\zeta>0$ being a scaling parameter.

(b) Now we comment on the line search rules. It is easy to check that, since $d_0^k\le 0$ by (3.7) (B_k is positive definite), if the algorithm does not terminate then $d_0^k<0$; hence the line search rules yield

$$f(x^{k+1})<f(x^k) \quad \text{if} \quad f(x^k)>0, \tag{4.3a}$$

$$f(x^{k+1})_+=0 \quad \text{if} \quad f(x^k)\le 0. \tag{4.3b}$$

Thus if x^k is infeasible, then the algorithm tries to find a feasible x^{k+1}; once a feasible approximation x^k is found, then the next approximations stay feasible owing to (4.3b), and the phase II of the method works. Note that even at the phase I, when $f(x^k)>0$, the direction d^k is influenced by the objective function if $f(x^k)\le\delta^k$ since then $I_0(x^k,\delta^k)$ is nonempty. On the other hand, the objective function values are ignored at line searches until a feasible approximation is found. In both phases, we try to reduce the number of problem function evaluations at line searches. The requirements of Step 4 are generalizations of the well-known convergent stepsize choices for standard algorithms, cf. [12, 14]. They can be met by various efficient stepsize procedures, e.g. Wierzbicki's procedure from [16]. The existence of a finite i_k at Step 4 (i) follows from the next section's results.

Under an additional assumption that the function $\rho_k(x^k+td^k)$ is bounded from below for $t>0$, finite termination of Wierzbicki's procedure from [15] can be easily established, thus providing a method for Step 4 (ii) (see also [12, 2.1.33 and 2.1.36]). At Step 3 we check if $t^k=1$ is acceptable. In general, small d_0^k indicates that x^k is a good approximation to a solution, hence we try to set $t^k=1$ at Step 3 if the algorithm makes good progress, as measured by $|d_0^k|\leq\gamma\,|d_0^{k-1}|$. Step 5 allows us to accept t^k better than $\tilde{t}^k$; obviously, we may always take $t^k=\tilde{t}^k$. In practice, one should monitor the requirements (4.2) while searching for $\tilde{t}^k$ at Step 4.

(c) In this paper we do not consider details of the choice of the variable metric matrices $\{B_k\}$. Our global convergence analysis requires that we have

$$\beta_1\,|d^k|^2\leq\langle B_k d^k, d^k\rangle \quad \text{and} \quad |B_k d^k|\leq\beta_2\,|d^k| \tag{4.4}$$

for some positive constants β_1 and β_2, for all k. Note that this property is much weaker than the usual assumption required for showing convergence of quasi-Newton methods, viz. that the matrices $\{B_k\}$ should be uniformly positive definite and bounded, cf. [3]. The simplest way to update B_k is to use the identity matrix. However, in order to obtain faster convergence, it is preferable to use B_k that estimates the Hessian $L''(x^k, w^k)$ with respect to x of the Lagrangian function for (2.1):

$$L(x, w)=\sum_{i\in I_0} w_i f_{0,i}(x)+\sum_{i\in I} w_{i+n} f_i(x),$$

where w^k denotes the Lagrange multiplier $\hat{w}$ of the k-th direction finding subproblem, cf. section 3. Some specific updating schemes are discussed in [15]. We may add that it is possible to ensure (4.4) by choosing initially some fixed numbers $\beta_1\in(0,1)$ and $\beta_2\geq 1$ and resetting B_k to the identity matrix if (4.4) is violated; in this case Step 1 should be repeated wuth $B_k=I$.

5. Convergence

In this section we analyze global convergence of the method. In the absence of convexity, we will content ourselves with finding feasible points that satisfy the Fritz-John necessary conditions of optimality (2.4). Under additional constraint qualifications, these points will also satisfy the Kuhn-Tucker necessary conditions (2.8). Naturally, asymptotic convergence results assume that the final accuracy tolerance $\varepsilon_f=0$.

Throughout the paper we assume that the variable metric matrices satisfy

$$\beta_1\,|d^k|^2\leq\langle B_k d^k, d^k\rangle, \tag{5.1a}$$

$$|B_k d^k|\leq\beta_2\,|d^k|, \tag{5.1b}$$

for some fixed positive numbers β_1 and β_2, for all k. We shall also use the following convention. Let $w^k\in R^{m+n}$ denote the Lagrange multiplier w of the k-th direction finding subproblem (3.2) solved at Step 1 of the method, cf. section 3. Define

$$p^k = \sum_{i \in I_0} w_i^k f'_{0,i}(x^k) + \sum_{i \in I} w_{i+n}^k f'_i(x^k). \tag{5.2}$$

By (3.3), (3.6) and (3.7), we always have

$$d^k = -B_k^{-1} p^k = -H_k p^k \tag{5.3}$$

$$w^k \in W, \tag{5.4a}$$

$$w_i^k [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+ + \langle f'_{0,i}(x^k), d^k \rangle - d_0^k] = 0, \quad i \in I_0, \tag{5.4b}$$

$$w_{i+n}^k [f_i(x^k) - f(x^k)_+ + \langle f'_i(x^k), d^k \rangle - d_0^k] = 0, \quad i \in I, \tag{5.4c}$$

$$-d_0^k = \langle B_k d^k, d^k \rangle + \sum_{i \in I_0} w_i^k [f_0(x^k) - f_{0,i}(x^k) + f(x^k)_+] + \sum_{i \in I} w_{i+n}^k [f(x^k)_+ - f_i(x^k)]. \tag{5.5}$$

Now, (5.1a) and (5.3) give

$$\beta_1 |d^k|^2 \le \langle B_k d^k, d^k \rangle = \langle -p^k, d^k \rangle \le |p^k|\, |d^k|,$$

hence ($\beta_1 > 0$) we always have

$$|d^k| \le |p^k| / \beta_1. \tag{5.6}$$

On the other hand, from (5.1b) and (5.3) we immediately get

$$|p^k| \le \beta_2 |d^k|. \tag{5.7}$$

Since $w^k \ge 0$, (5.5) yields $-d_0^k \ge \langle B_k d^k, d^k \rangle$, which together with (5.1a) implies

$$-d_0^k \ge \beta_1 |d^k|^2. \tag{5.8}$$

From (3.3), we have

$$d_0^k = \max\{ \max_{i \in I_0(x^k, \delta^k)} [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+ + \langle f'_{0,i}(x^k), d^k \rangle], \max_{i \in I(x^k, \delta^k)} [f_i(x^k) - f(x^k)_+ + \langle f'_i(x^k), d^k \rangle] \}. \tag{5.9}$$

If $w \in W$ and $\bar{x}$ is a Fritz-John point, let

$$\bar{\rho}(w, W(\bar{x})) = \min \{ |w - \bar{w}| : \bar{w} \in W(\bar{x}) \}.$$

First we consider the case when the algorithm terminates.

PROPOSITION 5.1. If the algorithm terminates at the k-th iteration, then either $(\bar{x}, \bar{w})$ is a Fritz-John point satisfying (2.4), i.e. $\bar{\rho}(w^k, W(x^k)) = 0$, or x^k is infeasible and $(\bar{x}, \bar{w}) = (x^k, w^k)$ satisfies the necessary conditions of optimality (2.6) for the problem (2.5). If the Cottle constraint qualification (2.7) holds at $\bar{x} = x^k$, then x^k is feasible.

P r o o f. Since $|d_0^k| \le \varepsilon_f = 0$, (5.8) and (5.7) imply that $d^k = p^k = 0$. Then (5.2) and (5.4) yield that (2.4a—d) holds for $(\bar{x}, \bar{w}) = (x^k, w^k)$. If $f(x^k) \le 0$, then we also have (2.4e) and thus $(\bar{x}, \bar{w})$ is a Fritz-John point. On the other hand, if $f(x^k) = f(x^k)_+ > 0$, then $f_{0,i}(\bar{x}) - f_0(\bar{x}) - f(\bar{x})_+ \le -f(\bar{x}) < 0$; hence (2.4c) and $\bar{w} \ge 0$ yield

$$\bar{w}_i = 0, \quad i \in I_0. \tag{5.10}$$

But then (2.4a—d) reduces to (2.6). Now, if the Cottle constraint qualification holds at $\bar{x} = x^k$, then (2.6) cannot hold if $f(x^k) > 0$. This completes the proof. ■

From now on we assume that the algorithm does not stop and that it generates an infinite sequence $\{x^k\}$. Let $\bar{K} = \{1, 2, 3, \ldots\}$. If $\bar{x}$ is an accumulation point of x^k, i.e. $x^k \to \bar{x}$ as $k \to \infty$, $k \in K$, where K is an infinite subset of $\bar{K}$, we write $x^k \xrightarrow{K} \bar{x}$.

We start with the following auxiliary result.

PROPOSITION 5.2. Suppose $x^k \xrightarrow{K} \bar{x}$ and $d^k \xrightarrow{K} 0$. Then $d_0^k \xrightarrow{K} 0$.

P r o o f. Since $I_0(x^k, \delta^k) \in I_0$ and $f_0(x^k) = f_{0,i}(x^k)$ for some $i \in I_0$, we always have

$$\max_{i \in I_0(x^k, \delta^k)} [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+ + \langle f'_{0,i}(x^k), d^k \rangle] \le -f(x^k)_+ + |d^k| \max_{i \in I_0} |f'_{0,i}(x^k)|. \tag{5.11a}$$

On the other hand,

$$\max_{i \in I_0(x^k, \delta^k)} [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+ + \langle f'_{0,i}(x^k), d^k \rangle] \ge -f(x^k)_+ - |d^k| \max_{i \in I_0} |f'_{0,i}(x^k)|. \tag{5.11b}$$

Similarly, we deduce easily that

$$-[f(x^k)_+ - f(x^k)] - |d^k| \max_{i \in I} |f'_i(x^k)| \le \max_{i \in I(x^k, \delta^k)} [f_i(x^k) - f(x^k)_+ + \langle f'_i(x^k), d^k \rangle] \le -[f(x^k)_+ - f(x^k)] + |d^k| \max_{i \in I} |f'_i(x^k)|. \tag{5.12}$$

Then (5.9), (5.11) and (5.12) yield, since $\max\{-f(x^k)_+, -f(x^k)_+ + f(x^k)\} = -f(x^k)_+ + f(x^k)_+ = 0$, that

$$|d_0^k| \le |d^k| \max\{\max_{i \in I_0} |f'_{0,i}(x^k), \max_{i \in I} |f'_i(x^k)|\}. \tag{5.13}$$

Hence $d_0^k \xrightarrow{K} 0$ follows from (5.13) and the continuity of the problem function gradients. The proof is complete. ■

From the above proposition and (5.8), we deduce easily the following result.

PROPOSITION 5.3. Suppose $x^k \xrightarrow{K} \bar{x}$. Then $d^k \xrightarrow{K} 0$ if and only if $d_0^k \xrightarrow{K} 0$.

Next we have the following convergence result.

THEOREM 5.4. *Suppose that* $(x^k, w^k) \xrightarrow{K} (\bar{x}, \bar{w})$ *and* $d^k \xrightarrow{K} 0$. *Then either* $(\bar{x}, \bar{w})$ *is a Fritz-John point, or* $\bar{x}$ *is infeasible and* $(\bar{x}, \bar{w})$ *satisfies the necessary conditions of optimality* (2.6) *for the problem* (2.5). *If the Cottle constraint qualification* (2.7) *holds at* $\bar{x}$, *then* $\bar{x}$ *is feasible.*

Proof. By Proposition 5.3, we have $d_0^k \xrightarrow{K} 0$. Additionally, (5.7) yields $p^k \xrightarrow{K} 0$. Letting k approach infinity with $k \in K$ in (5.2) and (5.4), we obtain (2.4a—d.) Therefore one may use the arguments in the proof of Proposition 5.1 to complete the proof. ∎

Note that the above results do not depend on the line search properties. These properties are essential for showing that $d_0^k \xrightarrow{K} 0$ if $x^k \xrightarrow{K} \bar{x}$.

PROPOSITION 5.5. Suppose that $x^k \xrightarrow{K} \bar{x}$ and $d_0^k \leq \bar{d}_0 < 0$ for all $k \in K$. Then for any fixed number $\bar{m} \in (0, 1)$ there exists a number $\bar{t}(\bar{m}) > 0$ such that

$$\max\{f_0(x^k + td^k) - f_0(x^k), f(x^k + td^k)\} \leq f(x^k)_+ + \bar{m} t d_0^k \text{ for any } t \in [0, \bar{t}(\bar{m})] \quad (5.14)$$

and any $k \in K$.

Proof. It follows from (5.2) and (5.6) that $\{d^k\}_{k \in K}$ are uniformly bounded. Combining this fact with the continuous differentiability of the problem function gradients, we obtain, by [2, Appendix III, Section 3, Note 2], that the following estimates are valid for $k \in K$:

$$f_{0,i}(x^k + td^k) - f_{0,i}(x^k) \leq t\langle f'_{0,i}(x^k), d^k\rangle + o_1(t), \quad i \in I_0, \quad (5.15a)$$

$$f_i(x^k + td^k) - f_i(x^k) \leq t\langle f'_i(x^k), d^k\rangle + o_1(t), \quad i \in I, \quad (5.15b)$$

where $o_1(t)/t \to 0$ as $t \to 0+$. By (5.9) and (5.15a), we have for $t \in [0, 1]$ and $k \in K$:

$$f_{0,i}(x^k + td^k) \leq f_{0,i}(x^k) + t[d_0^k - f_{0,i}(x^k) + f_0(x^k) + f(x^k)_+] + o_1(t) \leq$$

$$\leq f_0(x^k) + f(x^k)_+ + td_0^k + o_1(t), \quad i \in I_0(x^k, \delta^k); \quad (5.16a)$$

similarly, (5.9) and (5.15b) yield for $t \in [0, 1]$ and $k \in K$ that

$$f_i(x^k + td^k) \leq f(x^k)_+ + td_0^k + o_1(t), \quad i \in I(x^k, \delta^k). \quad (5.16b)$$

Let $C = \sup\{|d^k| \max(|f'_{0,i}(x^k)|, |f'_j(x^k)|): i \in I_0, j \in I, k \in K\}$. Note that $C < +\infty$, since $\{d^k\}_{k \in K}$ is uniformly bounded and $x^k \xrightarrow{K} \bar{x}$. By (5.15), the definition of the activity sets and the fact that $\delta^k \geq \bar{\delta}$, we have for $k \in K$

$$f_{0,i}(x^k + td^k) - f_0(x^k) \leq f(x^k)_+ - \bar{\delta} + tC + o_1(t), \quad i \in I_0 \setminus I_0(x^k, \delta^k), \quad (5.17a)$$

$$f_i(x^k + td^k) \leq f(x^k)_+ - \bar{\delta} + tC + o_1(t), \quad i \in I \setminus I(x^k, \delta^k). \quad (5.17b)$$

The argument leading to (5.15) may be applied to the functions $-f_{0,i}$ and $-f_i$; hence we obtain for $k \in K$

$$f_{0,i}(x^k + td^k) - f_{0,i}(x^k) \geq t\langle f'_{0,i}(x^k), d^k\rangle + o_2(t), \quad i \in I_0, \quad (5.18a)$$

$$f_i(x^k + td^k) \geq f_i(x^k) + t\langle f'_i(x^k), d^k\rangle + o_2(t), \quad i \in I, \quad (5.18b)$$

where $o_2(t)/t \to 0$ as $t \to 0+$. Since $f_0(x^k) = f_{0,i}(x^k)$ for some $i \in I_0(x^k, \delta^k)$, we obtain from (5.18)

$$\max_{i \in I_0(x^k, \delta^k)} f_{0,i}(x^k + td^k) \geq f_0(x^k) - tC + o_2(t), \tag{5.19a}$$

$$\max_{i \in I(x^k, \delta^k)} f_i(x^k + td^k) \geq f(x^k) - tC + o_2(t). \tag{5.19b}$$

We may choose a number $\hat{t} \in [0, 1]$ such that

$$-\bar{\delta} + tC + o_1(t) < -tC + o_2(t) \quad \text{for any} \quad t \in [0, \hat{t}], \tag{5.20}$$

since $\bar{\delta} > 0$. Then (5.16), (5.17), (5.19) and (5.20) imply

$$\max\{f_0(x^k + td^k) - f_0(x^k), f(x^k + td^k)\} \leq f(x^k)_+ + td_0^k + o_1(t) \tag{5.21}$$

for any $t \in [0, \hat{t}]$ and $k \in K$. Since $\bar{m} \in (0, 1)$ and $d_0^k \leq \bar{d}_0 < 0$ for all $k \in K$, it is possible to choose $\bar{t}(\bar{m}) \in (0, \hat{t})$ such that for $k \in K$

$$o_1(t)/t \leq (1 - \bar{m})(-\bar{d}_0) \leq (1 - \bar{m})(-d_0^k) \quad \text{for all} \quad t \in [0, \bar{t}(\bar{m})],$$

which implies

$$td_0^k + o_1(t) \leq \bar{m}td_0^k \quad \text{for any} \quad t \in [0, \bar{t}(\bar{m})], \quad k \in K \tag{5.22}$$

Combining (5.21) and (5.22), we obtain the desired relation (5.14). The proof is complete. ∎

Proposition 5.6. Suppose that $x^k \xrightarrow{K} \bar{x}$. Then $d^k \xrightarrow{K} 0$.

Proof. In view of Proposition 5.3, we only need to show that $d_0^k \xrightarrow{K} 0$. Since $d_0^k \leq 0$, this is equivalent to showing that for any fixed $\bar{d}_0 < 0$ we have $d_0^k \leq \bar{d}_0$ for only finitely many $k \in K$. Assume that this is not true. We will deduce from it a contradiction. Thus, with no loss of generality, suppose that $d_0^k \leq \bar{d}_0 < 0$ for all $k \in K$.

(i) Suppose that $t^k = 1$ is accepted at Step 3 for infinitely many $k \in K$. At such k, we have $|d_0^k| \leq \gamma\eta^k$ and $\eta^{k+1} = \gamma\eta^k$. Since $\gamma = (0, 1)$ and $\{\eta^k\}$ is nonincreasing by construction, we must have $|d_0^k| < |\bar{d}_0| > 0$ for some $k \in K$. This contradicts $d_0^k \leq \bar{d}_0 < 0$. Therefore we may suppose that the stepsize coefficient t^k is computed at Steps 4 and 5 for all $k \in K$.

(ii) Suppose that $f(x^{k_0}) \leq 0$ for some k_0. Then (4.3b) yields $f(x^k)_+ = 0$ for all $k \geq k_0$. Let $\bar{m} = [1 + \max\{m_1, m_2\}]/2$. Then Proposition 5.5 implies that there exists a number $\bar{t} > 0$ such that for almost all $k \in K$

$$\max\{f_0(x^k + td^k) - f_0(x^k), f(x^k + td^k)\} \leq \bar{m}td_0^k \quad \text{for any} \quad t \in [0, \bar{t}] \tag{5.23}$$

since $f(x^k)_+ = 0$ for $k \geq k_0$. If $\tilde{t}^k$ passes the Armijo test, then either $\tilde{t}^k = 1$ or $\rho_k(x^k + 2\tilde{t}^k d^k) = \max\{f_0(x^k + 2\tilde{t}^k d^k) - f_0(x^k), f(x^k + 2\tilde{t}^k d^k)\} > \rho_k(x^k) + m\, 2\tilde{t}^k d_0^k = m\, 2\tilde{t}^k d_0^k > \bar{m}\, 2\tilde{t}^k d_0^k$, hence (5.23) implies that $\tilde{t}^k \geq \bar{t}/2$. If $\tilde{t}^k$ satisfies the Goldstein test, then $\max\{f_0(x^k + \tilde{t}^k d^k) - f_0(x^k), f(x^k + \tilde{t}^k d^k)\} \geq m_2 \tilde{t}^k d_0^k > \bar{m}\tilde{t}^k d_0^k$, hence

(5.23) yields $\bar{t}^k \geq \bar{t}$. Therefore in both cases we have $\bar{t}^k \geq \bar{t}/2$ for almost all $k \in K$. Now, $f(x^k)_+ = 0$ and the line search rules yield

$$\max\{f_0(x^{k+1}) - f_0(x^k), f(x^{k+1})\} \leq \min\{m, m_1\}\, \bar{t}^k d_0^k, \tag{5.24}$$

which implies, since $\bar{t}^k \geq \bar{t}/2$ and $d_0^k \leq \bar{d}_0 < 0$, that we have

$$f_0(x^{k+1}) \leq f_0(x^k) + \min\{m, m_1\}\, \bar{t}\bar{d}_0/2 \tag{5.25}$$

for almost all $k \in K$. Then (5.24) yields $f_0(x^{k+1}) \leq f_0(x^k)$ for all $k \geq k_0$, therefore, since f_0 is continuous ([2]) and $x^k \xrightarrow{K} \bar{x}$, we have

$$f_0(x^k) \to f_0(\bar{x}) \quad \text{as} \quad k \to \infty. \tag{5.26}$$

Since $\min\{m, m_1\}\, \bar{t}\bar{d}_0/2 < 0$ is fixed, (5.25) contradicts (5.26).

(iii) Now suppose that $f(x^k) > 0$ for all k. Then the line search rules of Steps 4 and 5 imply $f(\tilde{x}^k) > 0$ for all k. One may argue as in part (ii) of the proof above to show that $t^k \geq \bar{t}/2$ and

$$f(x^{k+1}) \leq f(x^k) + \min\{m, m_1\}\, \bar{t}\bar{d}_0/2 \tag{5.27}$$

for almost all $k \in K$, and that $f(x^{k+1}) \leq f(x^k)$ for almost all k, which leads to

$$f(x^k) \to f(\bar{x}) \quad \text{as} \quad k \to \infty. \tag{5.28}$$

But then (5.27) contradicts (5.28). The proof is complete. ■

For the algorithm's starting point x^1, let

$$S(x^1) = \{x \in R^N : 0 < f(x) \leq f(x^1)\}. \tag{5.29}$$

Note that $S(x^1)$ is empty if x^1 is feasible.

Now we state our principal result.

THEOREM 5.7. *Suppose that the Cottle constraint qualification (2.7) holds at any $\bar{x} \in S(x^1)$. Then every accumulation point $\bar{x}$ of the sequence $\{x^k\}$ generated by the algorithm satisfies the Fritz-John necessary optimality condition (2.4). Moreover, if $x^k \xrightarrow{K} \bar{x}$, then*

$$\bar{\rho}(w^k, \bar{W}(\bar{x})) \xrightarrow{K} 0. \tag{5.30}$$

Proof. Suppose that $x^k \xrightarrow{K} \bar{x}$. By Proposition 5.6, $d^k \xrightarrow{K} 0$. Owing to (5.4a), $\{w^k\} \subset W$. The compactness of W (see (2.3)) implies that any accumulation point of $\{w^k\}_{k \in K}$ lies in W and at least one such point exists. Let $\bar{w} \in W$ be any accumulation point of $\{w^k\}_{k \in K}$. In view of the algorithm's rules, we always have $f(x^{k+1}) \leq f(x^k)_+$. Hence $f(\bar{x}) \leq f(x^1)_+$ (f is continuous). Suppose that $f(\bar{x}) > 0$. Then (2.7) holds at $\bar{x}$, hence Theorem 5.4 yields $f(\bar{x}) \leq 0$ — a contradiction. Consequently, $\bar{x}$ is feasible. Therefore Theorem 5.4 implies that $(\bar{x}, \bar{w})$ is a Fritz-John point, i.e. $\bar{w} \in \bar{W}(\bar{x})$. Since $\bar{w}$ was arbitrary and W is compact, this proves (5.30) and completes the proof. ■

Let

$$w_0^k = \sum_{i \in I_0} w_i^k \tag{5.31}$$

and define estimates of the Kuhn-Tucker multipliers by

$$u_i^k = w_i^k / w_0^k, \quad i \in I_0, \quad \text{and} \quad v_i^k = w_{i+n}^k / w_0, \quad i \in I \tag{5.32}$$

whenever $w_0^k > 0$. Let $\bar{S}(x^1) = \{x \in R^N : f(x) \le f(x^1)_+\}$. Then we have the following result on convergence of the Kuhn-Tucker estimates (u^k, v^k).

THEOREM 5.8. *Suppose that the Cottle constraint qualification* (2.7) *holds at any* $\bar{x} \in \bar{S}(x^1)$. *Then every accumulation point* $\bar{x}$ *of the sequence* $\{x^k\}$ *constructed by the method satisfies the Kuhn-Tucker condition* (2.8). *Moreover, if* $x^k \xrightarrow{K} \bar{x}$, *then*

$$\bar{\rho}((u^k, v^k), UV(\bar{x})) \xrightarrow{K} 0. \tag{5.33}$$

Proof. Since $S(x^1) \in \bar{S}(x^1)$, Theorem 5.7 yields (2.4) for any accumulation point $\bar{w}$ of $\{w^k\}_{k \in K}$, whenever $x^k \xrightarrow{K} \bar{x}$. If we show that $\bar{w}_0 = \sum_{i \in I_0} \bar{w}_i > 0$, i.e. (2.9) holds, then the desired conclusion (5.33) will follow from (5.30—32) and the results of section 2. To obtain a contradiction, assume (with no loss of generality) that $w^k \xrightarrow{K} \bar{w}$ with $\bar{w}_0 = 0$. By Theorem 5.7, we have $\bar{w} \in W(\bar{x})$. Since $\bar{S}(x^1)$ is closed and $x^k \in \bar{S}(x^1)$, we obtain that the Cottle constraint qualification holds at $\bar{x} \in \bar{S}(x^1)$. By the results of section 2, $\bar{w}_0 \ge \hat{w}_0 > 0$, cf. (2.11), and we have a contradiction with $\bar{w}_0 = 0$. The proof is finished. ■

REMARK 5.9. Since $d_0^k < 0$ at Step 4, it follows from the proof of Proposition 5.5 that (5.14) holds for $\bar{m} = m$ and $\bar{t}_k(\bar{m}) > 0$, which proves that $i_k < \infty$ at Step 4(i). A similar approach may be used for proving finite termination of Wierzbicki's stepsize procedure from [15] for Step 4(ii) of the method.

6. Rate of convergence

In this section we show that under favourable conditions our algorithm converges at least R-linearly. Our analysis generalizes the results of Pironneau and Polak from [10], where the case of a smooth objective function f_0, i.e. $n=1$, is considered. To save space, we continually refer to [10] and use its notation, providing here essential modifications only.

For ease of reference, we list certain assumptions under the following hypothesis. Hypothesis 6.1. The functions $f_{0,i}$, $i \in I_0$, and f_i, $i \in I$, are convex and twice continuously differentiable. The starting point x^1 is feasible and such that the set

$$B = \{x \in R^N : f_0(x) \le f_0(x^1) \text{ and } f(x) \le 0\}$$

is compact. Moreover, f_0 is strictly convex in B, e.g. $f_{0,i}$, $i \in I_0$, are strictly convex, and the Slater constraint qualification (2.13) holds, i.e. the set $C = \{x \in R^N : f(x) < 0\}$ is nonempty.

In this section we shall always assume that Hypothesis 6.1 is fulfilled. It follows that there exists a unique $\bar{x}$ solving (2.1). By Remark 2.3 and the results of section 5, we have $\{x^k\} \subset B$ and $x^k \to \bar{x}$. Moreover, if we denote for $\varepsilon > 0$

$$N(W(\bar{x}), \varepsilon) = \{w \in W : \bar{\rho}(w, W(\bar{x})) \le \varepsilon\},$$

then we have, by (5.30), that for any $\varepsilon > 0$

$$w^k \in N(W(\bar{x}), \varepsilon) \quad \text{for almost all} \quad k. \tag{6.1}$$

It follows from Propositions 5.3 and 5.6 that

$$d_0^k \to 0 \quad \text{as} \quad k \to \infty. \tag{6.2}$$

We shall also assume that there exist constants $\varepsilon > 0$ and $m_0 \in (0, 1)$ such that

$$m_0 |y-x|^2 \le \langle y-x, L''(x, w)(y-x)\rangle \quad \text{for all} \quad x, y \in B(\bar{x}, \varepsilon)$$
$$\text{and} \quad w \in N(W(\bar{x}), \varepsilon), \tag{6.3}$$

where $B(\bar{x}, \varepsilon) = \{x \in S : |x - \bar{x}| \le \varepsilon\}$ and the Lagrangian L for (2.1) is defined by

$$L(x, w) = \sum_{i \in I_0} w_i f_{0,i}(x) + \sum_{i \in I} w_{i+n} f_i(x). \tag{6.4}$$

We assume that the algorithm constructs the infinite sequence $\{x^k\}$ with $\gamma = 0$, i.e. that no direct prediction steps are taken. Since x^1 is feasible, only the phase II of the method works and we always have

$$\rho_k(x) = \max\{f_0(x) - f_0(x^k), f(x)\}. \tag{6.5}$$

We shall start by estimating

$$c^k = \rho_k(x^{k+1}) - \rho_k(x^k). \tag{6.6}$$

Proposition 6.2. There exists a constant $\bar{t} > 0$ such that

$$c^k \le \bar{t} d_0^k \quad \text{for all} \quad k. \tag{6.7}$$

P r o o f. Let $M_1 = \max\{\|f_{0,i}''(x)\| + \|f_j''(x)\| : x \in B, i \in I_0, j \in I\}$.

Since

$$f_{0,i}(x^k + td^k) = f_{0,i}(x^k) + t\langle f_{0,i}'(x^k), d^k\rangle + t^2 \langle d^k, f_{0,i}''(\tilde{x}) d^k\rangle \le f_{0,i}(x^k) +$$
$$+ t\langle f_{0,i}'(x^k), d^k\rangle + t^2 M_1 |d^k|^2, \tag{6.8}$$

where $\tilde{x}$ lies on the segment joining x^k and $x^k + td^k$, we obtain from (5.1a), (5.9) and (6.8) that for $t \in [0, 1]$ and $i \in I_0(x^k, \delta^k)$

$$f_{0,i}(x^k + td^k) \le f_{0,i}(x^k) + t[d_0^k - f_{0,i}(x^k) + f_0(x^k) + f(x^k)_+ +$$
$$+ t^2 M_1 |d^k|^2 \le f_0(x^k) + f(x^k)_+ + td_0^k + t^2 M_1 |d^k|_{B_k}^2 / \beta_1.$$

In the same manner we obtain for $t \in [0, 1]$ and $i \in I(x^k, \delta^k)$

$$f_i(x^k + td^k) \le f(x^k)_+ + td_0^k + t^2 M_1 |d^k|_{B_k}^2 / \beta_1.$$

Now let $M_2 = \max\{|f'_{0,i}(x)| + |f'_j(x)| : x \in B, i \in I_0, j \in I\}$. By (5.8) and $\delta^k \geq \bar{\delta}$, if $i \in I_0 \setminus I_0(x^k, \delta^k)$ then

$$f_{0,i}(x^k + td^k) = f_{0,i}(x^k) + t\langle f'_{0,i}(\tilde{x}), d^k\rangle \leq f_{0,i}(x^k) + tM_2\,|d^k| \leq f_0(x^k) + \\ + f(x^k)_+ - \bar{\delta} + tM_2(-d_0^k)^{\frac{1}{2}}/(\beta_1)^{\frac{1}{2}}.$$

Similarly, for $i \in I \setminus I(x^k, \delta^k)$ we obtain

$$f_i(x^k + td^k) \leq f(x^k)_+ - \bar{\delta} + tM_2(-d_0^k)^{\frac{1}{2}}/(\beta_1)^{\frac{1}{2}}.$$

Then (6.5) and the above four estimates yield

$$p_k(x^k + td^k) \leq p_k(x^k) + \bar{m}td_0^k \quad \text{for} \quad t \in [0, \bar{t}^k], \tag{6.9}$$

where $\bar{m} = \max\{m, m_1\}$ and

$$t^{-k} = \min\{1, \bar{\delta}/[(-d_0^k)^{\frac{1}{2}}(M_2/(\beta_1)^{\frac{1}{2}} + \bar{m}(-d_0^k), (1-\bar{m})\beta_1/M_1)\}. \tag{6.10}$$

Since no direct prediction steps are taken and $f(x^k)_+ = p_k(x^k) = 0$, we may argue as in part (ii) of the proof of Proposition 5.6 to deduce from (6.9) that $\hat{t}^k \geq \bar{t}^k/2$ and that

$$c^k = p_k(x^{k+1}) - p_k(x^k) \leq \min\{m, m_1\}\,\bar{t}^k d_0^k/2. \tag{6.11}$$

Next, it follows from (6.10) and $d_0^k \to 0$ that there exists a positive constant $\bar{t} < \min \times \times \{m, m_1\}\,\bar{t}^k/2$ for any k; hence (6.11) yields (6.7). The proof is complete. ■

Proceeding as in [10], let

$$\sigma(x^k) = \min\{\sigma : f_{0,i}(x) - f_0(x^k) - \sigma \leq 0,\ i \in I_0,\ f_i(x) - \sigma \leq 0,\ i \in I,\ x \in B\}. \tag{6.12}$$

The following proposition is a straightforward extension of Lemma 2.7 from [10], therefore we omit its proof.

Proposition 6.3. Let $\tilde{w}^k \in R^{m+n}$ be any solution of the dual of (6.12), i.e. of

$$\max_{w \geq 0}\Big[\min_{(\sigma, x) \in R^1 \times B}\Big\{\Big(1 - \sum_{i=1}^{m+n} w_i\Big)\sigma + \sum_{i \in I_0} w_i[f_{0,i}(x) - f_0(x^k)] + \\ + \sum_{i \in I} w_{i+n} f_i(x)\Big\}\Big] \tag{6.13}$$

Then $\tilde{w}^k \in W$ and $\tilde{\rho}(\tilde{w}^k, W(\bar{x})) \to 0$ as $k \to \infty$.

By convex duality theory, see Theorem 2.11 in [10], we obtain

$$\sigma(x^k) = \min_{x \in B}\Big\{\sum_{i \in I_0} \tilde{w}_i^k[f_{0,i}(x) - f_0(x^k)] + \sum_{i \in I} w_{i+n}^k f_i(x)\Big\}. \tag{6.14}$$

Upon replacing x by $\bar{x}$ in (6.14) and noting that $f(\bar{x}) \leq 0$, we obtain

$$\sigma(x^k) \leq \sum_{i \in I_0} \tilde{w}_i^k[f_0(\bar{x}) - f_0(x^k)]. \tag{6.15}$$

Next, from Proposition 6.3 and the results of section 2, cf. (2.11), we deduce that

$$\liminf_{k \to \infty} \sum_{i \in I_0} \tilde{w}_i^k \geq \hat{w}_0 > 0,$$

which implies that, given any $\tau \in (0, 1)$, there exists a $k_0(\tau)$ such that

$$\sum_{i \in I_0} \tilde{w}_i^k \geq \hat{w}_0 (1-\tau) \quad \text{for all} \quad k \geq k_0(\tau). \tag{6.16}$$

Combining (6.16) with (6.15) we now obtain

$$\sigma(x^k) \leq \hat{w}_0 (1-\tau) [f_0(x) - f_0(x^k)] \quad \text{for all} \quad k \geq k_0(\tau). \tag{6.17}$$

Generalizing Theorem 3.16 from [10], we get

PROPOSITION 6.4. Assume (with no loss of generality) that $m_0 \leq \beta_2^2/\beta_1$. Then

$$\sigma(x^k) \geq \beta_2^2 [d^k + \tfrac{1}{2}\langle B_k d^k, d^k\rangle/(\beta_1 m_0) \quad \text{for almost all} \quad k. \tag{6.18}$$

Proof. From the relation (3.23) in [10] we obtain that for almost all k

$$\sigma(x^k) = \max_{w \in W} \inf_{y \in B(\bar{x}, \varepsilon)} \Big\{ \sum_{i \in I_0} w_i f_{0,i}(y) - f_0(x^k) + \sum_{i \in I} w_{i+n} f_i(y) \Big\}.$$

Therefore for almost all k

$$\sigma(x^k) \geq \inf_{y \in B(\bar{x}, \varepsilon)} \Big\{ \sum_{i \in I_0} w_i^k [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+] + \sum_{i \in I} w_{i+n}^k [f_i(x^k) - f(x^k)_+] + \sum_{i \in I_0} w_i^k [f_{0,i}(y) - f_{0,i}(x^k)] + \sum_{i \in I} w_{i+n}^k [f_i(y) - f_i(x^k)] \Big\}.$$

Expanding $f_{0,i}(y) - f_{0,i}(x^k)$ and $f_i(y) - f_i(x^k)$ to second order terms and making use of (6.1) and (6.3), we obtain for almost all k

$$\sigma(x^k) \geq \sum_{i \in I_0} w_i^k [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+] + \sum_{i \in I} w_{i+n}^k [f_i(x^k) - f(x^k)_+] + \inf_{y \in B(\bar{x}, \varepsilon)} \Big\{ \Big\langle \sum_{i \in I_0} w_i^k f'_{0,i}(x^k) + \sum_{i \in I} w_{i+n}^k f'_i(x^k), y - x^k \Big\rangle + m_0 |y - x^k|^2/2 \Big\}. \tag{6.19}$$

By deleting the constraint $y \in B(\bar{x}, \varepsilon)$ in (6.19) and using (5.2), we get

$$\sigma(x^k) \geq \sum_{i \in I_0} w_i^k [f_{0,i}(x^k) - f_0(x^k) - f(x^k)_+] + \sum_{i \in I} w_{i+n}^k [f_i(x^k) - f_i(x^k)_+] + - |p^k|^2/(2m_0). \tag{6.20}$$

It follows from (5.1) and (5.3) that $|p^k|^2 \leq \beta_2^2 |d^k|^2 \leq \langle B_k d^k, d^k\rangle \beta_2^2/\beta_1$. Hence (6.20), the fact that $\beta_2^2/(\beta_1 m_0) \geq 1$ and that the first two terms in (6.20) are nonpositive, together with (5.5) and (5.1a), yield (6.18). This completes the proof. ∎

We are now ready to state the main rate of convergence result.

THEOREM 6.5. *Given any $\tau \in (0, 1)$, there exists a $k_0(\tau)$ such that for all $k \geq k_0(\tau)$*

$$f_0(x^{k+1}) - f_0(\bar{x}) \leq [1 - \hat{t} m_0 \hat{w}_0 (1-\tau) \beta_1/\beta_2^2] [f_0(x^k) - f_0(\bar{x})]. \tag{6.21}$$

Proof. Since $\rho_k(x^k) = f(x^k)_+ = 0$, Proposition 6.2 implies

$$f_0(x^{k+1}) - f_0(x^k) \leq \hat{t} d_0^k. \tag{6.22}$$

From (6.17) and (6.18) we obtain for $k \geq k_0(\tau)$

$$d_0^k \leq \beta_1 m_0 \sigma(x^k)/\beta_2^2 \leq [\beta_1 m_0 \hat{w}_0 (1-\tau)/\beta_2^2] [f_0(\bar{x}) - f_0(x^k)]. \tag{6.23}$$

Finally, from (6.22) and (6.23)

$$f_0(x^{k+1}) - f_0(x^k) \leq [\bar{t} m_0 \hat{w}_0 (1-\tau) \beta_1/\beta_2^2] [f_0(\bar{x}) - f_0(x^k)], \tag{6.24}$$

for $k \geq k_0(\tau)$. Rearranging (6.24), we obtain (6.21). The proof is complete.

THEOREM 6.6. *The sequence $\{x^k\}$ converges to $\bar{x}$ at least linearly.*

Proof. Let $\bar{w} \in W(\bar{x})$. According to the Taylor expansion formula, for any x^k there exists a point y^k on the segment joining x^k and $\bar{x}$ such that

$$\sum_{i \in I_0} \bar{w}_i [f_{0,i}(x^k) - f_{0,i}(\bar{x})] + \sum_{i \in I} \bar{w}_{i+n} [f_i(x^k) - f_i(\bar{x})] =$$

$$= \left\langle x^k - \bar{x}, \sum_{i \in I_0} \bar{w}_i f'_{0,i}(\bar{x}) + \sum_{i \in I} \bar{w}_{i+n} f'_i(\bar{x}) \right\rangle + $$

$$+ \tfrac{1}{2} \langle x^k - \bar{x}, L''(y^k, \bar{w})(x^k - \bar{x}) \rangle. \tag{6.25}$$

Since $\bar{w}$ satisfies (2.4) and $x^k \to \bar{x}$, (6.3) and (6.25) give

$$\sum_{i \in I_0} \bar{w}_i [f_0(x^k) - f_0(\bar{x})] + \sum_{i \in I} \bar{w}_{i+n} [f(x^k)_+ - f(\bar{x})_+] \geq m_0/2 \, |x^k - \bar{x}|^2.$$

Therefore, for almost all k

$$|x^k - \bar{x}|^2 \leq (2\bar{w}_0/m_0) [f_0(x^k) - f_0(\bar{x})] \tag{6.26}$$

and our assertion follows from (6.21) and (6.26), thus ending the proof. ■

7. Modifications and extensions

In this section we discuss some modifications of the method.

We start by remarking that one may use $\delta^k = +\infty$ in the algorithm, i.e. $I_0(x^k, \delta^k) = I_0$ snd $I(x^k, \delta^k) = I$ for all k. Clearly, ths strategy is covered by our preceding analysis.

If some x^k happens to be feasible, e.g. x^1 is feasible, we may modify the method to obtain a feasible direction algorithm that generalizes the Pironneau--Polak feasible direction methods [11]. It suffices to re-define the improvement function ρ_k in the algorithm's description by putting

$$\rho_k(x) = f_0(x) - f_0(x_k)$$

and then to include additional stepsize requirements that $f(x^k + d^k) \leq 0$ at Step 3, $f(x^k + \bar{t}^k d^k) \leq 0$ at Step 4, and to substitute (4.26) by the following:

$$f_0(x^k + t^k d^k) \leq f_0(\bar{x}^k) \quad \text{and} \quad f(x^k + t^k d^k) \leq 0. \tag{4.26'}$$

These modified line search rules maintain feasibility of consecutive points. One may check easily that all the preceding convergence results hold for this modification; in particular — linear convergence is retained.

Another modification concerns the line search and the stopping criterion. It consists in replacing in the algorithm's description the variable d_0^k by the variable $\tilde{d}_0^k = -\langle B_n d^n, d^n \rangle = -\langle H_n p^n, p^n \rangle$. From (5.5), we always have $d_0^k \le \tilde{d}_0^k \le 0$. For unconstrained minimax problems, this choice of $\tilde{d}_0^k$ at Armijo-type line searches is strongly advocated by Han [3], who argues that it may provide larger stepsizes, hence faster convergence. It is straightforward to check that the results of section 5 still hold. However, we have not been able to establish linear rate of convergence for this modification.

8. Conclusions

We have presented a phase I-phase II method for inequality constrained minimax problems that does not require a feasible starting point. The method generalizes some of the most robust and efficient feasible point algorithms for standard nonlinear programming calculations [10, 11, 13, 14]. Global convergence of the method has been established. The algorithm converges at least linearly when the optimization problem is convex and certain regularity assumptions are fulfilled. To the best of our knowledge, this seems to be the first implementable method for nonlinearly constrained minimax problems which is both globally and linearly convergent.

References

[1] Clarke F. H. A new approach to Lagrange multipliers. *Mathematics of Operations Research* **1** (1976), 165—174.

[2] Demyanov V. F., Malozemov V. N. Introduction to minimax. New York, John Wiley & Sons, 1974.

[3] Man S. P. Variable metric methods for a class of nondifferentiable functions. *Mathematical Programming* **20** (1981), 1—13.

[4] Hiriart Urruty J. B. Refinements of necessary optimality conditions for nondifferentiable programming. I. *Applied Mathematics and Optimization* **5** (1979), 63—82.

[5] Kiwiel K. C. A variable metric method of centers for nonsmooth minimization. CP-81-23. International Institute for Applied Systems Analysis, Laxenburg, Austria, 1981.

[6] Madsen K., Schjaer-Jacobsen H. Linearly constrained minimax optimization, *Mathematical Programming* **14** (1978), 208—223.

[7] Mifflin R. An algorithm for constrained minimization with semismooth functions, *Mathematics of Operations Research* **2** (1977), 191—207.

[8] Nguyen V. H., Strodiot J. J., Mifflin R. On conditions to have bounded multipliers in locally Lipschitz programming. *Mathematical Programming* **18** (1980), 100—106.

[9] Panin V. M. A method of linearizations for the discrete minimax problem. *Kibernetika* **3** (1980), 86—90 (in Russian).

[10] Pironneau O., Polak E. On the rate of convergence of certain methods of centers. *Mathematical Programming* **2** (1972), 230—257.

[11] Pironneau O., Polak E. Rate of convergence of a class of methods of feasible directions. *SIAM Journal on Numerical Analysis* **10** (1973), 161—174.
[12] Polak E. Computational methods in optimization. A unified approach. New York. Academic Press. 1971.
[13] Polak E., Mayne D. Q. A robust secant method for optimization problems with inequality constraints. *Journal of Optimization Theory and Applications* **33** (1981), 463—477.
[14] Polak E., Trahan R., Mayne D. Q. Combined phase I-phase II methods of feasible directions. *Mathematical Programming* **17** (1979), 32—61.
[15] Pshenichnyi B. N., Danilin Yu. M. Numerical methods in extremal problems. Moscow. Nauka. 1975 (in Russian).
[16] Wierzbicki A. P. Lagrangian functions and nondifferentiable optimization, WP-78-63. International Institute for Applied Systems Analysis, Laxenburg, Austria, 1978.
[17] Zoutendijk G. Methods of feasible directions. Amsterdam. Elsevier. 1960.

Received, September 1982.

Dwufazowa metoda dla zadań minimaksowych z ograniczeniami nierównościowymi

Przedstawiono nową klasę algorytmów do minimalizacji, przy skończonej liczbie ograniczeń nierównościowych, nieróżniczkowalnej funkcji będącej punktowym maksimum ze skończonej rodziny funkcji różniczkowalnych w sposób ciągły. W metodach tych rozwiązuje się kolejne zadania programowania kwadratowego w celu generacji kierunków poszukiwań. Zaproponowano kilka praktycznych sposobów przeszukiwania kierunku. Metody nie wymagają dopuszczalnego punktu startowego. Wykazano globalną zbieżność proponowanych algorytmów. Przy dodatkowych założeniach wypukłości, szybkość zbieżności metod jest co najmniej liniowa. Algorytmy dają się łatwo zaprogramować. Uogólniają one szereg niezawodnych i efektywnych metod typu punktu dopuszczalnego dla standardowych zadań programowania nieliniowego.

Двухфазный метод для задачи дискретного минимакса с ограничениями в форме неравенсть

Представлен класс методов минимизации функции дискретного максимума при наличии конечного числа гладких ограничений в форме неравенсть. В методах итеративно решаются подзадачи квадратического программирования для нахождения направлений спуска. Введено несколько практических способов регулировки шага. Методы не требуют допустимого начального приближения. Установлено глобальную сходимость методов. При дополнительных предложениях выпуклости, доказано линейную скорость сходимости. Методы легко программируются на эвм. Они обобщают несколько робастных и эффективных методов типа допустимой точки для решения стандартных задач нелинейного программирования.

Control
and Cybernetics
VOL. 12 (1983) No. 1-2

News and Announcements

Seventh European Meeting on Cybernetics and Systems Research is organized by the Austrian Society for Cybernetic Studies, to be held on April 24—27, 1984, at the University of Vienna, Austria

The meeting offers plenary lectures and symposia on:
- — General Systems Methodology
- — System and Decision Theory
- — Cybernetics in Biology and Medicine
- — Cybernetics of Cognition and Learning
- — Cybernetics in Organization and Management
- — Economic and Social Systems
- — Ecological Systems
- — Health Care Systems
- — Fuzzy Sets
- — Communication and Computers
- — Artificial intelligence

Details of Submission:
Please send three copies of the full paper (up to 7 pages, single-spaced, maximum 2600 words) in English. Submissions should include the selected symposium. Papers must be received not later than September 1st, 1983.
Submissions and/or enquiries should be sent to the Chairman:

Professor Robert Trappl
Department of Medical Cybernetics
University of Vienna
Freyung 2/6
A-1010 Vienna
AUSTRIA

Mathematical Methods in Operations Research

Sofia, Bulgaria, October 24—29, 1983
organized by Institute of Mathematics with Computer Centre of the Bulgarian Academy of Sciences.

Topics:

— Theory of Optimization Problems;
— Numerical Methods for Solving Optimization Problems;
— Applications and Program Packages.

Abstracts (1 page in camera ready form) should be sent before June 1, 1983 to:

Dr. A. L. Dontchev (Institute of Mathematics 1090 Sofia P.O. Box 373) Bulgaria

Languages: Russian and English.