The principle of feasible directions for nonlinear approximants and infinitely many constraints

E. Blum *, W. Oettli **

[43]

* Universidad Nacional de Ingeniería, Departemento de Matemáticas, Lima, Perú

** Universität Mannheim, Fakultät für Mathematik und Informatik, 68 Mannheim, Germany.

Problem statement and characterization of solutions

The programming problem we are going to consider has the following form:

(P) $$\min \{F(x) \mid x \in C,\ f_t(x) \le 0\ \forall t \in T\}\ .$$

Let S denote the admissible domain of P, i.e.,

$$S = \{x \mid x \in C,\ f_t(x) \le 0\ \forall t \in T\}\ .$$

Let us introduce functions $\Phi(x,\xi)$, $\varphi_t(x,\xi)$ $(t \in T)$, defined on $S \times C$, which will be used as substitutes for $F(\xi)$, $f_t(\xi)$. We make once and for all the following assumption:

(A1)
- $C \subset \mathbb{R}^n$ is a closed, convex set; T is a compact metric space;
- $F(x)$ is continuous on C; $f_t(x)$ is continuous on $T \times C$;
- $\Phi(x,\xi)$ is continuous on $S \times C$; $\varphi_t(x,\xi)$ is continuous on $T \times S \times C$.

In what follows we shall be interested in properties of a certain point

$$\hat{x} \in S\ .$$

Concerning this point $\hat{x} \in S$ we make the following assumption:

(A2)
- The functions $\Phi(\hat{x},\xi)$ and $\varphi_t(\hat{x},\xi)$ are convex with regard to ξ;
- $\Phi(\hat{x},\hat{x}) = F(\hat{x})$, $|\Phi(\hat{x},\xi) - F(\xi)| \le o(\xi-\hat{x})$;
- $\varphi_t(\hat{x},\hat{x}) = f_t(\hat{x})$, $|\Phi_t(\hat{x},\xi) - f_t(\xi)| \le o(\xi-\hat{x})$,
- where the Landau-bound $o(\bullet)$ *is independent of* t.

Always for $\hat{x} \in S$ define

$$\hat{T} = \{t \in T \mid f_t(\hat{x}) = 0\}\ ,$$

and let $\tilde{T}$ satisfy

$$\hat{T} \subset \tilde{T} \subset T\ .$$

We consider the following system in ξ:

(1) $$\xi \in C,\ \Phi(\hat{x},\xi) - F(\hat{x}) < 0,\ \varphi_t(\hat{x},\xi) < 0\ \forall t \in \tilde{T}\ .$$

Lemma 1: Let ξ be a solution of (1). Then for any $r > 0$ there exists $x \in [\hat{x}, \xi]$ satisfying

(2) $$x \in C,\ |x-\hat{x}| \le r,\ F(x) - F(\hat{x}) < 0,\ f_t(x) < 0\ \forall t \in T .$$

Proof: The compactness of T and the continuity of $f_t(\hat{x})$ with regard to t imply the compactness of $\hat{T}$. Since $\hat{T}$ is compact, since $\varphi_t(\hat{x},\xi) < 0\ \forall t \in \hat{T}$, and since $\varphi_t(\hat{x},\xi)$ is continuous with regard to t, we have

$$\varphi_t(\hat{x},\xi) \le -\delta < 0\ \forall t \in \hat{T} .$$

Define an open neighbourhood of $\hat{T}$ by means of

$$U = \{t \in T \mid \varphi_t(\hat{x},\xi) < -\tfrac{\delta}{2}\} .$$

$T \setminus U$ is again compact; since $f_t(\hat{x}) < o\ \forall t \in T \setminus U$, and since $f_t(\hat{x})$ is continuous with regard to t, we have

$$f_t(\hat{x}) \le -\varepsilon < 0\ \forall t \in T \setminus U .$$

Likewise

$$\varphi_t(\hat{x},\xi) \le M < \infty\ \forall t \in T \setminus U .$$

For $0 < \lambda \le 1$ let $x_\lambda = \hat{x} + \lambda(\xi - \hat{x})$. Then $x_\lambda \in C$, and by (A2) we obtain

$$f_t(x_\lambda) \le \varphi_t(\hat{x},x_\lambda) + o(x_\lambda - \hat{x}) \le \lambda\varphi_t(\hat{x},\xi) + (1-\lambda)\varphi_t(\hat{x},\hat{x}) + o(\lambda)$$
$$= \lambda\varphi_t(\hat{x},\xi) + (1-\lambda)f_t(\hat{x}) + o(\lambda) .$$

This gives

for all $t \in T \setminus U$: $f_t(x_\lambda) \le \lambda M + (1-\lambda)(-\varepsilon) + o(\lambda) < 0$ if $\lambda \in (0,\lambda_1)$;

for all $t \in U$: $f_t(x_\lambda) \le \lambda(-\frac{\delta}{2}) + o(\lambda) < 0$ if $\lambda \in (0,\lambda_2)$.

Also

$$F(x_\lambda) - F(\hat{x}) \le \lambda\Phi(\hat{x},\xi) + (1-\lambda)\Phi(\hat{x},\hat{x}) - F(\hat{x}) + o(\lambda)$$
$$= \lambda\big(\Phi(\hat{x},\xi) - F(\hat{x})\big) + o(\lambda) < 0 \quad \text{if } \lambda \in (0,\lambda_3) .$$

Therefore $x = x_\lambda$ satisfies (2) for all sufficiently small $\lambda > 0$. q.e.d.

The following theorem generalizes Kolmogorov's criterion for best Chebyshev-approximations [1].

Theorem 1: a) A necessary condition for $\hat{x} \in S$ to solve P is that (1) has no solution. b) This condition is also sufficient, if $F(\xi)$ and $f_t(\xi)$ are convex on C, and if $\exists \tilde{x} \in C$: $f_t(\tilde{x}) < 0 \ \forall t \in T$.

Proof: a) If (1) has a solution, then, by lemma 1, there exists $x \in S$ such that $F(x) < F(\hat{x})$, and $\hat{x}$ is not optimal for P. b) If $\hat{x}$ is not optimal for P there exists x satisfying

$$x \in C, \ f_t(x) \leq 0 \ \forall t \in T, \ F(x) < F(\hat{x}) \ .$$

Since $f_t(\tilde{x}) < 0 \ \forall t \in T$ and since $f_t(\cdot)$ is convex we have then $f_t(\xi) < 0$ for all $\xi \in (x,\tilde{x}]$ and for all $t \in T$. Also, since $F(x) < F(\hat{x})$ and $F(\cdot)$ is convex, we have $F(\xi) < F(\hat{x})$ for all $\xi \in (x,\tilde{x}]$ sufficiently close to x. Therefore the system

$$\xi \in C, \ F(\xi) - F(\hat{x}) < 0, \ f_t(\xi) < 0 \ \forall t \in T$$

has a solution. Now adapt lemma 1, with the rôles of $f(\cdot)$ and $\varphi(\hat{x},\cdot)$ interchanged, to conclude that (1) has a solution. q.e.d.

Concerning statement b) we mention an alternative assumption, under which the inconsistency of (1) is sufficient for $\hat{x} \in S$ to be a solution of P: $\Phi(\hat{x},\xi) \leq F(\xi)$, $\varphi_t(\hat{x},\xi) \leq f_t(\xi)$, $\exists \tilde{x} \in C$: $\varphi_t(\hat{x},\tilde{x}) < 0 \ \forall t \in T$. The proof in this case does not need lemma 1.

<u>Iterative scheme and convergence</u>

Starting from an arbitrary point $x^o \in S$ a sequence $\{x^k\} \subset S$ is recursively defined as follows: Fix $\alpha > 0$, $\beta > 0$. Suppose $x^k \in S$ is given. Let

$$T_\alpha^k = \{t \in T \mid f_t(x^k) \geq -\alpha\} ,$$

$$H_\alpha^k(\xi) = \max \{\Phi(x^k,\xi) - F(x^k),\ \varphi_t(x^k,\xi)(t \in T_\alpha^k)\} .$$

Define ξ^k as being a solution of

(3) $$\min \{H_\alpha^k(\xi) \mid \xi \in C,\ |\xi - x^k| \leq \beta\} .$$

Define x^{k+1} as being a solution of

(4) $$\min \{F(x) \mid x \in [x^k, \xi^k] \cap S\} .$$

This completes one step of the iteration. Since T_α^k is compact and since $\varphi_t(x^k,\xi)$ is continuous on $T_\alpha^k \times C$, $H_\alpha^k(\xi)$ is continuous on C. Therefore (3) has a solution. Also (4) has a solution. Obviously x^{k+1} is again in S. We mention that T_α^k may be replaced by T. Also the requirement $|\xi - x^k| \leq \beta$ may be dropped from (3) if C is already compact. But α cannot be set equal to zero.

Theorem 2: Let $\hat{x}$ be a cluster point of the sequence $\{x^k\}$. Then $\hat{x} \in S$, and (1) has no solution (i.e., $\hat{x}$ satisfies the necessary optimality condition of theorem 1).

We have already defined $H_\alpha^k(\xi)$. For the proof let us define in addition

$$H_\infty(x,\xi) = \max \{\Phi(x,\xi) - F(x),\ \varphi_t(x,\xi)(t \in T)\} ,$$

$$\hat{H}(x,\xi) = \max \{\Phi(x,\xi) - F(x),\ \varphi_t(x,\xi)(t \in \hat{T})\} ,$$

where $\hat{T} = \{t \in T \mid f_t(\hat{x}) = 0\}$. Both functions are continuous on $S \times C$ by (A1). Note that assumption (A2) is supposed to hold at $\hat{x}$.

Proof of theorem 2:

—— We have $\hat{x} \in S$, as S is closed by (A1). According to (4) we have $F(x^k) \geq F(x^{k+1})$, and this monotonicity implies

(5) $$F(x^k) \geq F(\hat{x})\ \forall k .$$

Since $\hat{x}$ is a cluster of $\{x^k\}$ and since $|\xi^k - x^k| \leq \beta$ there exists $\hat{\xi} \in C$ and a subsequence $\{\kappa\} \subset \{k\}$ such that

$$x^\kappa \to \hat{x}, \; \xi^\kappa \to \hat{\xi} \; .$$

— There exists K_1 such that

$$\hat{T} \subset T_\alpha^\kappa \; \forall \kappa \geq K_1 \; . \tag{6}$$

Indeed, $\psi(x) = \min_{t \in \hat{T}} f_t(x)$ is continuous by the compactness of $\hat{T}$ and the continuity of $f_t(x)$ on $\hat{T} \times C$. Since $\psi(\hat{x}) = 0$ and $x^\kappa \to \hat{x}$, we have $\psi(x^\kappa) \geq -\alpha$ for all sufficiently large κ. This proves (6).

— We have

$$\hat{H}(\hat{x},\hat{\xi}) \leq \min_{\substack{\xi \in C \\ |\xi - \hat{x}| \leq \frac{\beta}{2}}} H_\infty(\hat{x},\xi) \; . \tag{7}$$

Indeed, suppose there exists $\xi \in C$ such that

$$\hat{H}(\hat{x},\hat{\xi}) > H_\infty(\hat{x},\xi), \; |\xi - \hat{x}| \leq \frac{\beta}{2} \; .$$

By continuity then for all $\kappa \geq K_2$

$$\hat{H}(x^\kappa,\xi^\kappa) > H_\infty(x^\kappa,\xi), \; |\xi - x^\kappa| \leq \beta \; .$$

Since $T \supset T_\alpha^\kappa$ we have by the definition of H_∞ and of H_α^κ

$$H_\infty(x^\kappa,\xi) \geq H_\alpha^\kappa(\xi) \; .$$

By (6) for all $\kappa \geq K_1$ we have $T_\alpha^\kappa \supset \hat{T}$, implying, by the definition of $\hat{H}$, that

$$H_\alpha^\kappa(\xi^\kappa) \geq \hat{H}(x^\kappa, \xi^\kappa) \; .$$

Thus for all sufficiently large κ we would have

$$H_\alpha^\kappa(\xi^\kappa) > H_\alpha^\kappa(\xi), \; |\xi - x^\kappa| \leq \beta \; .$$

This contradicts the definition of ξ^κ in (3). Thus (7) must hold.

— Suppose now that (1) has a solution. It follows by a simple application of lemma 1 that also the system

$$\xi \in C, \; |\xi - \hat{x}| \leq \frac{\beta}{2}, \; \Phi(\hat{x},\xi) - F(\hat{x}) < 0, \; \varphi_t(\hat{x},\xi) < 0 \; \forall t \in T$$

has a solution. This means that $\min_{\substack{\xi\in C \\ |\xi-\hat{x}|\le\frac{\beta}{2}}} H_\infty(\hat{x},\xi) < 0$; and by (7) this implies $\hat{H}(\hat{x},\hat{\xi}) < 0$, i.e.,

$$\Phi(\hat{x},\hat{\xi})-F(\hat{x}) < 0,\ \varphi_t(\hat{x},\hat{\xi}) < 0\ \forall t\in\hat{T}\ .$$

Then by lemma 1 there exists $\chi\in[\hat{x}\ \hat{\xi}]$ satisfying

$$F(\chi)-F(\hat{x}) < 0,\ f_t(\chi) < 0\ \forall t\in T\ .$$

Let $\chi = \hat{x}+\lambda(\hat{\xi}-\hat{x})$, $0 \le \lambda \le 1$, and set

$$\chi^\kappa = x^\kappa+\lambda(\xi^\kappa-x^\kappa) \in [x^\kappa,\xi^\kappa] \subset C\ .$$

Then $\chi^\kappa \to \chi$, and by the continuity and compactness assumptions of (A1)

$$F(\chi^\kappa)-F(\hat{x}) < 0,\ f_t(\chi^\kappa) < 0\ \forall t\in T$$

for all sufficiently large κ. This means that $\chi^\kappa\in[x^\kappa,\xi^\kappa] \cap S$, and by (4) implies $F(x^{\kappa+1}) \le F(\chi^\kappa)$. But since $F(\chi^\kappa) < F(\hat{x})$ we have $F(x^{\kappa+1}) < F(\hat{x})$, contradicting (5). The assumption that (1) has a solution was wrong. q.e.d.

It was already assumed in the proof of theorem 2 that $\Phi(\hat{x},\hat{x}) = F(\hat{x})$, $\varphi_t(\hat{x},\hat{x}) = f_t(\hat{x})$. Assume for the moment, that also $\Phi(x,x) \le F(x)$, $\varphi_t(x,x) \le f_t(x)\ \forall x\in S$. Then $H^k_\alpha(x^k) \le 0$, and by (3) $H^k_\alpha(\xi^k) \le H^k_\alpha(x^k)$. Therefore $H^k_\alpha(\xi^k) \le 0$. In the proof of statement (7) it was shown that $H^\kappa_\alpha(\xi^\kappa) \ge \hat{H}(x^\kappa,\xi^\kappa)\ \forall\kappa \ge K_1$. By continuity $\hat{H}(x^\kappa,\xi^\kappa) \to \hat{H}(\hat{x},\hat{\xi})$. Since (1) has no solution, $\hat{H}(\hat{x},\xi) \ge 0\ \forall\xi\in C$. In conclusion we obtain $\hat{H}(\hat{x},\hat{\xi}) = 0$, $H^\kappa_\alpha(\xi^\kappa) \to 0$.

Rate of convergence

This section deals with the rate of convergence of $F(x^k)$ in some special cases. In addition to (A1) we require:

$$\text{(A3)}\quad \begin{cases} F(\xi), f_t(\xi), \Phi(x,\xi), \varphi_t(x,\xi) \text{ are convex with respect to } \xi\,; \\ \Phi(x,x) = F(x), \varphi_t(x,x) = f_t(x)\ \forall x\in S\,; \text{ the set} \\ \qquad S_o = \{x\in C \mid f_t(x) \le 0\ \forall t\in T,\ F(x) \le F(x^o)\} \\ \text{is bounded; } \exists \tilde{x}\in C\colon f_t(\tilde{x}) < 0\ \forall t\in T;\ \text{(A2) holds} \\ \text{for all } \hat{x}, \text{ which are cluster points of } \{x^k\}\,. \end{cases}$$

The compactness of S_o implies that P has a solution. Also the sequence $\{x^k\} \subset S_o$ has cluster points. Under (A3) each cluster point is a solution of P, by theorem 2 and theorem 1b). Furthermore, since each subsequence of $\{x^k\}$ contains a subsequence which converges to a solution of P, the remark following the proof of theorem 2 implies that $H_\alpha^k(\xi^k) \to 0$ for the entire sequence.

We shall use in the following the abbreviations

$$\tau^k = H_\alpha^k(\xi^k),\quad \delta^k = F(x^k) - \hat{F}\,,$$

where $\hat{F}$ is the optimal value of P. Then

$$\tau^k \le 0,\ \tau^k \to 0;\ \delta^k \ge 0,\ \delta^k \to 0\,.$$

With $\sigma\in\mathbb{R}$ define

$$K(\xi,\sigma) = \max\{F(\xi)-\sigma,\ f_t(\xi)(t\in T)\}\,.$$

Lemma 2: There exists $\rho_1 > 0$ such that $\min\limits_{\substack{\xi\in C \\ |\xi - x^k| \le \beta}} K(\xi, F(x^k)) \le \rho_1(-\delta^k)$.

Proof:

— Set $F^k = F(x^k)$. Since $S_o \subset C$,

$$\text{(8)}\qquad \min_{\substack{\xi\in C \\ |\xi-x^k|\le\beta}} K(\xi,F^k) \le \min_{\substack{\xi\in S_o \\ |\xi-x^k|\le\beta}} K(\xi,F^k)\,.$$

— Let $0 < \theta \le \min\{1, \frac{\beta}{R}\}$, where R is the diameter of S_o. Then for any $x\in S_o$

there exists $\xi \in S_o$ such that

$$|\xi - x^k| \leq \beta, \; K(\xi, F^k) \leq \theta K(x, F^k) \; .$$

(Choose $\xi = x^k + \theta(x - x^k)$, and note that $K(\cdot, F^k)$ is convex, $K(x^k, F^k) = 0$.) This implies

$$\text{(9)} \qquad \min_{\substack{\xi \in S_o \\ |\xi - x^k| \leq \beta}} K(\xi, F^k) \leq \theta \cdot \min_{x \in S_o} K(x, F^k) \; .$$

— Since $F^k \leq F^o$, it is easy to verify that

$$\text{(10)} \qquad \min_{x \in S_o} K(x, F^k) = \min_{x \in C} K(x, F^k) \equiv v(F^k) \; .$$

— The function $v(\sigma) = \min_{x \in C} K(x, \sigma)$ is convex over $\mathbb{R}$ since $K(x, \sigma)$ is convex over $C \times \mathbb{R}$. Also $v(\sigma)$ is monotonically nonincreasing. $v(\hat{F}) = 0$, by the very definition of $\hat{F}$. Furthermore for $\sigma > \hat{F}$ the value of v is negative, since there exists $\tilde{x} \in C$ with $f_t(\tilde{x}) < 0 \; \forall t \in T$. In particular, $v(F^o) = \gamma < 0$. By convexity, for all $\sigma \in [\hat{F}, F^o]$, $v(\sigma) \leq \frac{\sigma - \hat{F}}{F^o - \hat{F}} \gamma$. Therefore

$$\text{(11)} \qquad v(F^k) \leq \delta^k \frac{\gamma}{F^o - \hat{F}} \; .$$

Inequalities (8) - (11) in succession prove the lemma. q.e.d.

Suppose for the moment we would have determined x^{k+1} as a solution of $\min \{K(x, F^k) \,|\, x \in C, |x - x^k| \leq \beta\}$. This would be the iteration rule for the method of centers [3]. Its main difference to the method of feasible directions [2] is that no line-minimization like (4) takes place. It is clear, that $x^{k+1} \in S$ under this rule. Furthermore, from lemma 2 we would have $\delta^{k+1} - \delta^k = F(x^{k+1}) - F^k \leq$ $\leq K(x^{k+1}, F^k) \leq \rho_1(-\delta^k)$, i.e., $\delta^{k+1} \leq (1 - \rho_1)\delta^k$. We return now to the method of feasible directions.

Lemma 3: If there exist constants $\mu \geq 0$, $0 < m \leq 1$, such that (i) $\Phi(x, \xi) - \mu|\xi - x|^2$ is convex with respect to ξ, (ii) $\varphi_t(x, \xi) - \mu|\xi - x|^2 \leq$ $\leq f_t(\xi)$, (iii) $\Phi(x, \xi) - (1 - m)\mu|\xi - x|^2 \leq F(\xi)$, then there exists $\rho_2 > 0$ such that $\tau^k \leq \rho_2(-\delta^k)$.

Proof:

— Let $C_k = \{\xi\in C \mid |\xi-x^k| \le \beta\}$. Problem (3), which leads to ξ^k and thereby to τ^k, may be written as

$$\min \{ \tau \mid \tau\in\mathbb{R},\ \xi\in C_k,\ -\tau + \Phi(x^k,\xi) - F(x^k) \le 0,\ -\tau + \max_{t\in T_\alpha^k} \varphi_t(x^k,\xi) \le 0\} .$$

This is a convex programming problem with two numerical constraints. Its optimal value is τ^k, and due to its special structure (Slater's regularity assumption is satisfied) the Kuhn-Tucker-conditions [4] hold: There exist $u^k \ge 0$, $v^k \ge 0$ such that

$$\tau^k \le \tau + u^k\left(-\tau + \Phi(x^k,\xi) - F(x^k)\right) + v^k\left(-\tau + \max_{t\in T_\alpha^k} \varphi_t(x^k,\xi)\right) \ \forall\tau\in\mathbb{R},\ \forall\xi\in C_k .$$

Since this holds for all $\tau\in\mathbb{R}$ we conclude readily that

$$u^k + v^k = 1,$$

(12) $$\tau^k \le u^k\left(\Phi(x^k,\xi) - F(x^k)\right) + v^k \left(\max_{t\in T} \varphi_t(x^k,\xi)\right) \forall\xi\in C_k .$$

— We show first that $\liminf_{k\to\infty} u^k > 0$. Otherwise there would exist a subsequence

$$u^\kappa \to 0,\ v^\kappa \to 1,\ x^\kappa \to \hat{x},\ \tau^\kappa \to 0 ,$$

and (12) would give in the limit for all $\xi\in C$ satisfying $|\xi-\hat{x}| \le \beta$ that

(13) $$0 \le \max_{t\in T} \varphi_t(\hat{x},\xi) .$$

However, since there exists $\tilde{x}\in C$ with $f_t(\tilde{x}) < 0\ \forall t\in T$, and since (A2) holds for $\hat{x}$, an adaption of lemma 1 shows that there exists a $\xi\in C$ which contradicts (13). Consequently we have K and $\bar{u} > 0$ such that $u^k \ge \bar{u}\ \forall k \ge K$. Let $\upsilon^k = mu^k$. Then

$$0 < \bar{\upsilon} \le \upsilon^k \le 1 .$$

— (12) implies that for all $\xi\in C_k$

$$\tau^k \le \mu|\xi-x^k|^2 + u^k\left[\Phi(x^k,\xi) - F(x^k) - \mu|\xi-x^k|^2\right] + v^k\left[\max_{t\in T} f_t(\xi)\right] .$$

Now, if for arbitrary $x \in C_k$ we set $\xi = x^k + \upsilon^k(x-x^k)$, then $\xi \in C_k$, and since the expressions in brackets $[\cdot]$ are convex in ξ, and are nonpositive for $\xi = x^k$, we obtain

$$\tau^k \le \mu(\upsilon^k)^2|x-x^k|^2 + u^k\upsilon^k\left(\Phi(x^k,x) - F(x^k) - \mu|x-x^k|^2\right) + v^k\upsilon^k\left(\max_{t\in T} f_t(x)\right) \ \forall x \in C_k .$$

Noting that $\tau^k \le 0$ we have for all $x \in C_k$

$$\frac{\tau^k}{\bar{\upsilon}} \le \frac{\tau^k}{\upsilon^k} \le u^k\left(\Phi(x^k,x) - F(x^k) - (1-m)\mu|x-x^k|^2\right) + v^k\left(\max_{t\in T} f_t(x)\right)$$

$$\le u^k\left(F(x) - F(x^k)\right) + v^k\left(\max_{t\in T} f_t(x)\right)$$

$$\le \max\{F(x) - F(x^k),\ f_t(x)(t\in T)\}$$

$$= K(x,F(x^k)).$$

Thus $\tau^k \le \bar{\upsilon}\cdot\min_{x\in C_k} K(x,F(x^k))$. Lemma 2 then completes the proof: $\tau^k \le \bar{\upsilon}\rho_1(-\delta^k)$.

q.e.d.

Example: For F, f_t convex and differentiable with

$$F(\xi) - F(x) \ge (\xi-x)^T\nabla F(x) + m\mu|\xi-x|^2$$

choose

$$\Phi(x,\xi) = F(x) + (\xi-x)^T\nabla F(x) + \mu|\xi-x|^2 ,$$

$$\varphi_t(x,\xi) = f_t(x) + (\xi-x)^T\nabla f_t(x) + \mu|\xi-x|^2 .$$

Then (i) - (iii) are satisfied.

Lemma 4: a) If $F(\xi) \le \Phi(x,\xi) + M|\xi-x|^2$, $f_t(\xi) \le \varphi_t(x,\xi) + M|\xi-x|^2$, then $\delta^{k+1} - \delta^k \le -\gamma(\tau^k)^2$ for some $\gamma > 0$. b) If, in addition, there exists $\mu > 0$ such that $\Phi(x,\xi) - \mu|\xi-x|^2$ and $\varphi_t(x,\xi) - \mu|\xi-x|^2$ are convex with regard to ξ, then $\delta^{k+1} - \delta^k \le \gamma\tau^k$ for some $\gamma > 0$.

Proof: For $0 \le \lambda \le 1$ let $x_\lambda = x^k + \lambda(\xi^k - x^k)$.

a) We have for $t \notin T_\alpha^k$

$$f_t(x_\lambda) \le \varphi_t(x^k, x_\lambda) + M\lambda^2|\xi^k - x^k|^2$$

$$\le \varphi_t(x^k, x^k) + \lambda\left(\varphi(x^k,\xi^k) - \varphi(x^k,x^k)\right) + \lambda M|\xi^k - x^k|^2$$

$$\le -\alpha + \lambda b \le 0 \quad \text{for} \quad \lambda \in [0,\lambda_1]$$

(the uniform bound b exists by (A1)). For $t \in T_\alpha^k$ we have

$$f_t(x_\lambda) \le \varphi_t(x^k, x_\lambda) + M\lambda^2|\xi^k - x^k|^2$$

$$\le (1-\lambda)\varphi_t(x^k,x^k) + \lambda\varphi_t(x^k,\xi^k) + \lambda^2 M\beta^2$$

$$\le \lambda\left(\varphi_t(x^k,\xi^k) + \lambda M\beta^2\right)$$

$$\le \lambda(\tau^k + \lambda c) \le 0 \quad \text{for } \lambda \in [0, -\frac{\tau^k}{c}] .$$

In the same way

$$F(x_\lambda) - F(x_k) \le \lambda\left(\Phi(x^k, \xi^k) - F(x^k) + \lambda M\beta^2\right)$$

$$\le \lambda(\tau^k + \lambda c) \le \lambda\frac{\tau^k}{2} \quad \text{for} \quad \lambda \in [0, -\frac{\tau^k}{2c}] .$$

Since $\tau^k \to 0$ we may assume that $-\frac{\tau^k}{2c} \le \lambda_1$. Then, if we choose $\lambda = -\frac{\tau^k}{2c}$, $x_\lambda \in S$ and $F(x_\lambda) - F(x^k) \le -\frac{(\tau^k)^2}{4c}$. By (4) this implies

$$F(x^{k+1}) - F(x^k) \le F(x_\lambda) - F(x^k) \le -\gamma(\tau^k)^2 .$$

b) For $t \notin T_\alpha^k$ we have, as under a), $f_t(x_\lambda) \le 0$ if $\lambda \in [0,\lambda_1]$. For $t \in T_\alpha^k$ we obtain

$$f_t(x_\lambda) \le \varphi_t(x^k, x_\lambda) + M\lambda^2|\xi^k - x^k|^2$$

$$= \varphi_t(x^k, x_\lambda) - \mu|x_\lambda - x^k|^2 + (M+\mu)\lambda^2|\xi^k - x^k|^2$$

$$\le \lambda\left(\varphi_t(x^k,\xi^k) - \mu|\xi^k - x^k|^2\right) + (M+\mu)\lambda^2|\xi^k - x^k|^2$$

$$\le \lambda\varphi_t(x^k,\xi^k) \le \lambda\tau^k \quad \text{for} \quad \lambda \in [0,\lambda_2] .$$

In the same way

$$F(x_\lambda) - F(x^k) \le \lambda\tau^k \quad \text{for} \quad \lambda \in [0,\lambda_2] .$$

With $\lambda = \gamma \le \min\{\lambda_1,\lambda_2\}$ we have $x_\lambda \in S$ and $F(x^{k+1}) - F(x^k) \le F(x_\lambda) - F(x^k) \le \gamma\tau^k$.

q.e.d.

Combining the results of lemma 3 and lemma 4 we obtain immediately

Theorem 3: If the assumptions of lemma 3 and lemma 4a) hold, then $\delta^{k+1} \le (1 - \rho\delta^k)\,\delta^k$ for some $\rho > 0$. If the assumptions of lemma 3 and lemma 4b) hold, then $\delta^{k+1} \le (1 - \rho)\,\delta^k$ for some $\rho > 0$.

This extends some results of [5], [6].

Bibliography

[1] G. Meinardus, Approximation von Funktionen und ihre numerische Behandlung. Berlin - Heidelberg - New York, 1964.

[2] G. Zoutendijk, Methods of Feasible Directions. Amsterdam, 1960.

[3] P. Huard, Programmation mathématique convexe. Rev. Française Informat. Recherche Opérationnelle 2 (1968), no. 7, 43-59.

[4] S. Karlin, Mathematical Methods and Theory in Games, Programming, and Economics, Vol. 1. Cambridge, Mass., 1959.

[5] M. Frank, P. Wolfe, An algorithm for quadratic programming. Naval Res. Logist. Quart. 3 (1956), 95-110.

[6] O. Pironneau, E. Polak, On the rate of convergence of certain methods of centers. Math. Programming 2 (1972), 230-257.