

ON ITERATED MINIMIZATION IN NONCONVEX OPTIMIZATION*

H. TH. JONGEN,[†] T. MÖBERT[‡] AND K. TAMMER[‡]

In dynamic programming and decomposition methods one often applies an iterated minimization procedure. The problem variables are partitioned into several blocks, say x and y . Treating y as a parameter, the first phase consists of minimization with respect to the variable x . In a second phase the minimization of the resulting optimal value function depending on y is considered. In this paper we treat this basic idea on a local level. It turns out that strong stability (in the sense of Kojima) in the first phase is a natural assumption. In order to show that the iterated local minima of the parametric problem lead to a local minimum for the whole problem, we use a generalized version of a positive definiteness criterion of Fujiwara-Han-Mangasarian.

1. Introduction. For the optimization problem (P):

$$\min\{f(x, y) | (x, y) \in M\}, \tag{P}$$

a possible solution method consists of two phases.

The first phase will be the parametric problem (P_y):

$$\phi(y) = \inf_x \{f(x, y) | x \in M(y)\} \quad \text{with} \tag{P_y}$$

$$M(y) = \{x | (x, y) \in M\}. \tag{1.1}$$

As a logical consequence we note that $\phi(y) = \infty$ in case that $M(y)$ is empty. The second phase will be an optimization problem in the variable y :

$$\min\{\phi(y) | y \in N\} \quad \text{where } N = \{y | \phi(y) < \infty\}. \tag{1.2}$$

This type of iterated minimization procedures is often applied within the area of dynamic programming and decomposition methods (cf. Bank et al. [1], Beer [2], Bellman [3], Benders [4], Lasdon [10]). The usual assumption in these applications is the *global* optimal solution of both problem (P_y) as well as problem (1.2). Then the underlying idea is the following simple theorem (cf. [1], [10]).

THEOREM 1.1. *A point (\hat{x}, \hat{y}) is a solution of problem (P) if and only if \hat{x} is optimal for ($P_{\hat{y}}$) and, in addition, \hat{y} is optimal for problem (1.2).*

We emphasize that Theorem 1.1 is of *global* nature and it is valid without additional assumptions on problem (P). However, if problem (P) is not a convex one, then the

*Received October 15, 1984; revised July 2, 1985.

AMS 1980 subject classification. Primary 90C30; Secondary 65K05.

IAOR 1973 subject classification. Main: Programming: Nonlinear.

OR/MS Index 1978 subject classification. Primary: 653 Programming/Nonlinear.

Key words. Nonconvex programming, dynamic programming, decomposition methods, iterated minimization, two-stage programming.

[†]Twente University of Technology.

[‡]Technische Hochschule Leipzig.

problems (P_y) and (1.2) may also be nonconvex and the usual methods in nonlinear optimization yield only stationary solutions, resp. local minimal solutions in general.

In this paper we treat the idea of decomposition for nonconvex problems at a local level. This leads to a study of the following two basic problems.

Problem A. Let $\hat{x}(y)$ be a stationary solution for problem (P_y) for all y in some neighborhood $U(\hat{y})$, and let \hat{y} be a stationary solution for $\hat{\phi}(y) := f(\hat{x}(y), y)$.

Under what condition will (\hat{x}, \hat{y}) , where $\hat{x} := \hat{x}(\hat{y})$, be a stationary solution for problem (P) .

Problem B. The same formulation as Problem A. However, every entry “stationary solution” is replaced by “local minimum”.

At first glance, it seems to be necessary from the nature of nonconvex problems that the mapping $\hat{x}(y)$ in Problems A, B need to be continuous. However, continuity is not sufficient as the following example w.r.t. Problem B shows.

EXAMPLE. Put $M = \mathbb{R}^2$ and $f(x, y) = x^2y^2(x - \frac{1}{2}y^2)(x - 2y^2) + x^6 + y^{12}$. For every fixed y the point $\hat{x}(y) := 0$ is a strict local minimum for f with respect to x . Moreover, the resulting function $\hat{\phi}(y) = f(\hat{x}(y), y) = y^{12}$ has a strict local minimum at $y = 0$. However, the point $(\hat{x}, \hat{y}) = (0, 0)$ is *not* a local minimum for f since the function f has a local maximum at $(0, 0)$ along the parabola $x = y^2$.

We note that a first one but trivial result to solve Problem B is given in Luderer [11], where it was assumed that $\hat{x}(y)$ is a global optimal solution of (P_y) but \hat{y} only a local one.

2. Problem A: Iterated stationary points. In the sequel we consider the following optimization problem:

$$\min \{ f(x, y) \mid h_i(x, y) = 0, i \in I, g_j(x, y) \leq 0, j \in J \}, \quad (x, y) \in \mathbb{R}^{n_1+n_2}. \quad (P)$$

The index sets I, J are assumed to be finite and $f, h_i, g_j \in C^1(\mathbb{R}^{n_1+n_2})$. By Df , resp. $D_x f, D_y f$, we denote the row vector of first partial derivatives of f , resp. with respect to x, y . The index set I (and also J) will be partitioned into two parts: $I = I_1 \cup I_2, I_1 \cap I_2 = \emptyset$, where $I_2 = \{i \in I \mid D_x h_i(x, y) \equiv 0\}$. In fact, the constraint functions $h_i, i \in I_2$, do not depend on x . The set of stationary points (\hat{x}, \hat{y}) corresponding to (P) will be the projection of the set of Kuhn-Tucker points ψ_{KT} on the (x, y) -space, where

$$\psi_{KT} = \left\{ (x, y, \lambda, u) \left| \begin{array}{l} D_x \left[f(x, y) + \sum_{i \in I} \lambda_i h_i(x, y) + \sum_{j \in J} u_j g_j(x, y) \right] = 0, \\ D_y \left[f(x, y) + \sum_{i \in I} \lambda_i h_i(x, y) + \sum_{j \in J} u_j g_j(x, y) \right] = 0, \\ h_i(x, y) = 0, \quad i \in I, \\ g_j(x, y) \leq 0, \quad j \in J, \\ u_j g_j(x, y) = 0, \quad j \in J, u_j \geq 0, j \in J. \end{array} \right. \right\} \quad (2.1)$$

On the analogy of 1 we consider two phases.

In the first phase we have to determine, for every y in a suitable open set Y , a stationary point of the problem:

$$\min_x \{ f(x, y) \mid h_i(x, y) = 0, i \in I_1, g_j(x, y) \leq 0, j \in J_1 \}. \quad (P_y)$$

Let $\hat{x}(y)$ be a vector valued function which assigns a stationary point to each parameter y . Then, for $\hat{x}(y)$ as well as for the associated Lagrange multipliers $\hat{\lambda}(y)$, $\hat{u}(y)$ the following holds:

$$(\hat{x}(y), \hat{\lambda}(y), \hat{u}(y)) \in \psi_{KT}(y), \text{ where}$$

$$\psi_{KT}(y) = \left\{ \begin{array}{l} (x, \lambda^{(1)}, u^{(1)}) \in \mathbb{R}^p \\ p = n_1 + |I_1| + |J_1| \end{array} \left| \begin{array}{l} D_x \left[f(x, y) + \sum_{i \in I_1} \lambda_i^{(1)} h_i(x, y) + \sum_{j \in J_1} u_j^{(1)} g_j(x, y) \right] = 0, \\ h_i(x, y) = 0, \quad i \in I_1, \\ g_j(x, y) \leq 0, \quad j \in J_1, \\ u_j^{(1)} g_j(x, y) = 0, \quad j \in J_1, \\ u_j^{(1)} \geq 0, \quad j \in J_1 \end{array} \right. \right\} \tag{2.2}$$

In the second phase we have to find a stationary point for the problem

$$\min \{ \hat{\phi}(y) | y \in Y \cap N_0 \}, \tag{2.3}$$

where $\hat{\phi}(y) = f(\hat{x}(y), y)$ and $N_0 = \{ y \in \mathbb{R}^{n_2} | h_i(y) = 0, i \in I_2, g_j(y) \leq 0, j \in J_2 \}$.

In general the function $\hat{\phi}(y)$ will not be differentiable. In that case a stationary point has to be declared by using a generalized subdifferential (cf. [5],[15]). On the other hand, the next lemma provides a sufficient condition in order that $\hat{\phi}(y)$ is differentiable at $y = \hat{y}$.

LEMMA 2.1. *Let $f, h_i, g_j \in C^1(\mathbb{R}^{n_1+n_2})$, $i \in I_1, j \in J_1$. Suppose that $\hat{x}(y), \hat{\lambda}(y), \hat{u}(y)$ are vector valued functions, $\hat{u}(y)$ being continuous at \hat{y} , and $\hat{x}(y)$ locally Lipschitz continuous at \hat{y} such that in addition $(\hat{x}(y), \hat{\lambda}(y), \hat{u}(y)) \in \psi_{KT}(y)$ for all y in the neighborhood $U_{\delta_0}(\hat{y})$. Then, the function $\hat{\phi}(y) := f(\hat{x}(y), y)$ is differentiable at \hat{y} with derivative:*

$$D\hat{\phi}(\hat{y}) = D_y \left[f(*) + \sum_{i \in I_1} \hat{\lambda}_i(\hat{y}) h_i(*) + \sum_{j \in J_1} \hat{u}_j(\hat{y}) g_j(*) \right]^1 \tag{2.4}$$

where $(*)$ stands for $(\hat{x}(\hat{y}), \hat{y})$.

PROOF. For arbitrary $y \in U_{\delta_0}(\hat{y})$ we put $x = \hat{x}(y)$ and $\hat{x} = \hat{x}(\hat{y})$, whereas the multipliers $\lambda, \hat{\lambda}$, resp. u, \hat{u} are defined in an analogous way. The vector valued functions h , resp. g stand for $h_i, i \in I_1$, resp. $g_j, j \in J_1$. In view of the differentiability assumption on f, h, g , and the local Lipschitz condition of $\hat{x}(y)$ at \hat{y} , i.e. $\|x - \hat{x}\| \leq c\|y - \hat{y}\|$ with $0 < c < \infty$, for every $\epsilon > 0$ there exist $\delta_i > 0, i = 1, 2, 3, 4$,

¹To abbreviate the formulas we use the above representation instead of

$$D_y \left[f(x, y) + \sum_{i \in I_1} \lambda_i h_i(x, y) + \sum_{j \in J_1} u_j g_j(x, y) \right] \Bigg|_{\substack{x = \hat{x}(\hat{y}), \quad y = \hat{y}, \\ \lambda = \hat{\lambda}(\hat{y}), \quad u = \hat{u}(\hat{y}).}}$$

Similar abbreviations will also appear in the following.

such that:

$$\begin{aligned} \left| f(x, y) - f(\hat{x}, \hat{y}) - Df(\hat{x}, \hat{y}) \begin{pmatrix} x - \hat{x} \\ y - \hat{y} \end{pmatrix} \right| &\leq \frac{\epsilon}{3(1+c)} (\|x - \hat{x}\| + \|y - \hat{y}\|) \\ &\leq \left[\frac{\epsilon c}{3(1+c)} + \frac{\epsilon}{3(1+c)} \right] \cdot \|y - \hat{y}\| \\ &= \frac{\epsilon}{3} \|y - \hat{y}\| \quad \text{for all } y \in U_{\delta_1}(\hat{y}). \end{aligned}$$

$$\begin{aligned} &\left| \left[g(x, y) - g(\hat{x}, \hat{y}) - Dg(\hat{x}, \hat{y}) \begin{pmatrix} x - \hat{x} \\ y - \hat{y} \end{pmatrix} \right]^\top \hat{u} \right| \\ &\leq \left\| g(x, y) - g(\hat{x}, \hat{y}) - Dg(\hat{x}, \hat{y}) \begin{pmatrix} x - \hat{x} \\ y - \hat{y} \end{pmatrix} \right\| \cdot \|\hat{u}\| \\ &\leq \frac{\epsilon}{3(1+c)(1+\|\hat{u}\|)} (\|x - \hat{x}\| + \|y - \hat{y}\|) \|\hat{u}\| \\ &\leq \frac{\epsilon}{3} \|y - \hat{y}\| \quad \text{for all } y \in U_{\delta_2}(\hat{y}). \end{aligned}$$

For the function $h(x, y)$ a similar estimate holds for all $y \in U_{\delta_3}(\hat{y})$, by using $\hat{\lambda}$ instead of \hat{u} .

Moreover, for all $y \in U_{\delta_4}(\hat{y})$ we have the zero-relation:

$$(g(x, y) - g(\hat{x}, \hat{y}))^\top \hat{u} = 0. \tag{2.5}$$

In fact, if all \hat{u}_j vanish, then (2.5) obviously holds. On the other hand, if $\hat{u}_j > 0$, then $u_j > 0$ in a neighborhood of \hat{y} (use the continuity of u_j at \hat{y}). Since the stationary point x satisfies the complementarity condition, we have $g_j(x, y) = 0$. This proves the zero-relation (2.5).

Since x satisfies the equality constraints we also have:

$$(h(x, y) - h(\hat{x}, \hat{y}))^\top \hat{\lambda} = 0 \quad \text{for all } y \in U_{\delta_0}(\hat{y}).$$

Next, we put $\delta = \min_{i=0, \dots, 4} \delta_i$ and obtain for all $y \in U_\delta(\hat{y})$:

$$\begin{aligned} &|\hat{\phi}(y) - \hat{\phi}(\hat{y}) - D_y [f(\hat{x}, \hat{y}) + \hat{\lambda}^\top h(\hat{x}, \hat{y}) + \hat{u}^\top g(\hat{x}, \hat{y})](y - \hat{y})| \\ &= |f(x, y) - f(\hat{x}, \hat{y}) + \hat{\lambda}^\top [h(x, y) - h(\hat{x}, \hat{y})] + \hat{u}^\top [g(x, y) - g(\hat{x}, \hat{y})] \\ &\quad - D_y [f(\hat{x}, \hat{y}) + \hat{\lambda}^\top h(\hat{x}, \hat{y}) + \hat{u}^\top g(\hat{x}, \hat{y})](y - \hat{y}) \\ &\quad - D_x [f(\hat{x}, \hat{y}) + \hat{\lambda}^\top h(\hat{x}, \hat{y}) + \hat{u}^\top g(\hat{x}, \hat{y})](x - \hat{x})| \\ &\leq 3 \cdot \frac{\epsilon}{3} \|y - \hat{y}\| = \epsilon \|y - \hat{y}\|. \end{aligned}$$

The above estimate shows that (2.4) is derivative of $\hat{\phi}$ at \hat{y} . ■

REMARK 2.1. The result of Lemma 2.1 is a special case of the results in [7] in the case that $\hat{x}(y)$ is a global optimal solution of (P_y) . In contradistinction to [7] we do not

use optimal properties of the vectorfunction $\hat{x}(y)$, but merely the fact that the Kuhn-Tucker conditions hold.

REMARK 2.2. According to [9] the local Lipschitz continuity of $\hat{x}(y)$ at \hat{y} is guaranteed if $\hat{x} := \hat{x}(\hat{y})$ is strongly stable in the sense of Kojima [8] and the linear independence constraint qualification holds.

An immediate consequence of Lemma 2.1 is the following theorem which gives a solution to Problem A.

Recall that the functions $h_i, i \in I_2, g_j, j \in J_2$, do not depend on x .

THEOREM 2.1. *If the assumptions of Lemma 2.1 hold and if, in addition, the point \hat{y} is a stationary point of $\hat{\phi}(y)$ with respect to N_0 , i.e. there exist $(\hat{\lambda}^{(2)}, \hat{u}^{(2)}) \in \mathbb{R}^{|I_2|+|J_2|}$ with*

$$\left\{ \begin{array}{l} D_y \left[\hat{\phi}(\hat{y}) + \sum_{i \in I_2} \hat{\lambda}_i^{(2)} h_i(\hat{y}) + \sum_{j \in J_2} \hat{u}_j^{(2)} g_j(\hat{y}) \right] = 0, \\ \qquad \qquad \qquad h_i(\hat{y}) = 0, \qquad i \in I_2, \\ \qquad \qquad \qquad g_j(\hat{y}) \leq 0, \qquad j \in J_2, \\ \qquad \qquad \hat{u}_j^{(2)} g_j(\hat{y}) = 0, \qquad j \in J_2, \\ \qquad \qquad \qquad \hat{u}_j^{(2)} \geq 0, \qquad j \in J_2, \end{array} \right.$$

then (\hat{x}, \hat{y}) is a stationary point of problem (P), where $\hat{x} = \hat{x}(\hat{y})$.

3. **Problem B: Iterated local minima.** In order to reduce the notations we will delete the equality constraints, i.e. $I = \emptyset$. We emphasize that all results can be extended straightforwardly to the case $I \neq \emptyset$. So, we consider the problem:

$$\min \{ f(x, y) | g_j(x, y) \leq 0, j \in J_1, g_j(y) \leq 0, j \in J_2 \}, \tag{P}$$

where the functions f, g_j are C^2 -functions.

In the sequel D^2f , resp. D_x^2f, D_y^2f , stands for the Hessian matrix of f , resp. the Hessian matrix with respect to x, y .

Let Y be an open set and let $\hat{x}(t)$ be a continuous vector function on Y such that $\hat{x}(y)$ is a local minimum for the problem (P_y):

$$\min_x \{ f(x, y) | g_j(x, y) \leq 0, j \in J_1 \}, \tag{P_y}$$

with active index set $J_1^0(y) := \{ j \in J_1 | g_j(\hat{x}(y), y) = 0 \}$.

In the following we assume a suitable constraint qualification at $\hat{x} = \hat{x}(\hat{y})$ with $\hat{y} \in Y$ which guarantees the existence of a vector function $\hat{u}(y)$ such that the following Kuhn-Tucker relations hold:

$$\left. \begin{array}{l} D_x \left[f(\hat{x}(y), y) + \sum_{j \in J_1} \hat{u}_j(y) g_j(\hat{x}(y), y) \right] = 0, \\ \left. \begin{array}{l} g_j(\hat{x}(y), y) \leq 0, \\ \hat{u}_j(y) \cdot g_j(\hat{x}(y), y) = 0, \\ u_j(y) \geq 0, \end{array} \right\} \quad j \in J_1 \end{array} \right\}$$

Then $\hat{x}(y)$ is a stationary point of (P_y) and we will assume, in addition, that \hat{x} is strongly stable for P_y in the sense of Kojima (cf. [8]).

REMARK 3.1. The strong stability of a stationary point \hat{x} for $(P_{\hat{y}})$ locally secures the existence, continuity and even the uniqueness of a vector function $\hat{x}(y)$ with $\hat{x}(\hat{y}) = \hat{x}$ and $\hat{x}(y)$ stationary point for (P_y) . If, moreover, \hat{x} is a local minimum for $(P_{\hat{y}})$ then $\hat{x}(y)$ is also a local minimum for (P_y) . Consequently, the basic assumption of this section is then automatically fulfilled.

For a first answer to Problem B we assume the Mangasarian-Fromovitz constraint qualification at \hat{x} :

$$\{v | D_x^T g_j(\hat{x}, \hat{y})v < 0, j \in J_1^0(\hat{y})\} \neq \emptyset. \tag{V1}$$

THEOREM 3.1. *Let \hat{x} be a strongly stable local minimum for $(P_{\hat{y}})$ at which condition (V1) is fulfilled. Let further Y be an open neighborhood of \hat{y} and let $\hat{x}(y)$ be a continuous vector function on Y such that $\hat{x}(y)$ is a stationary point for (P_y) and $\hat{x}(\hat{y}) = \hat{x}$.*

Then (\hat{x}, \hat{y}) is a local minimum for (P) if \hat{y} is a local minimum for problem (2.3).

PROOF. According to [8] the strong stability of \hat{x} for $(P_{\hat{y}})$ and (V1) imply the existence of bounded open neighborhoods U of \hat{y} and V of \hat{x} as well as the existence of a continuous vector function $\bar{x}: U \rightarrow V$ such that

- (a) $\bar{x}(\hat{y}) = \hat{x}$,
- (b) $\bar{x}(y)$ is a stationary point for (P_y) which is unique in the closure $\text{cl } V$ of the set V ,
- (c) $\bar{x}(y)$ is a strongly stable local minimum for (P_y) ,
- (d) $f(\hat{x}, \hat{y}) < f(x, \hat{y})$ if $x \in (\text{cl } V) \cap M(\hat{y})$ and $x \neq \hat{x}$,
- (e) Mangasarian-Fromovitz constraint qualification is satisfied at every point $x \in (\text{cl } V) \cap M(y)$, for every $y \in U$.

Because of (a)–(c) we have $\hat{x}(y) = \bar{x}(y)$ for $y \in U \cap Y$, and, moreover, $\text{cl}(V) \cap M(y) \neq \emptyset$ for $y \in U$. In particular, for every $y \in U$ the auxiliary problem

$$\min_x \{f(x, y) | x \in (\text{cl } V \cap M(y))\}$$

has at least one minimal point which we denote by $\bar{x}(y)$. In addition, we consider the function $\bar{\phi}$:

$$\bar{\phi}(y) = \inf_x \{f(x, y) | x \in (\text{bd } V) \cap M(y)\},$$

where $\text{bd } V$ denotes the boundary of V .

The function $\bar{\phi}$ is lower-semicontinuous at \hat{y} . This follows from the fact that for every sequence $(y^i) \subset U$ with $y^i \rightarrow \hat{y}$ and every sequence $(x^i) \subset (\text{bd } V) \cap M(y^i)$ with $f(x^i, y^i) = \bar{\phi}(y^i)$ we have: (x^i) has accumulation points, every accumulation point \bar{x} of (x^i) belongs to $(\text{bd } V) \cap M(\hat{y})$ and satisfies $f(\bar{x}, \hat{y}) \geq \bar{\phi}(\hat{y})$; hence, $\liminf_{i \rightarrow \infty} \bar{\phi}(y^i) \geq \bar{\phi}(\hat{y})$ is valid. (Note that the special case where $(\text{bd } V) \cap M(\hat{y}) = \emptyset$ is automatically included.)

Because of the continuity of f and $\hat{x}(y)$, the lower-semicontinuity of $\bar{\phi}$ and the inequality $f(\hat{x}, \hat{y}) < \bar{\phi}(\hat{y})$ (compare (d)), there is a neighborhood $U' \subset U$ of \hat{y} such that $f(\hat{x}(y), y) < \bar{\phi}(y)$ holds for $y \in U'$. Consequently, $\bar{x}(y) \notin \text{bd } V$ and hence, $\bar{x}(y)$ is a local minimum for (P_y) for $y \in U'$. This implies in view of (e) that $\bar{x}(y)$ is a stationary point for (P_y) , and because of (b) that even $\bar{x}(y) = \hat{x}(y)$ holds for $y \in U'$. Consequently, there is a neighborhood W of (\hat{x}, \hat{y}) such that the chain

$$f(\hat{x}, \hat{y}) = \hat{\phi}(\hat{y}) \leq \hat{\phi}(y) = f(\hat{x}(y), y) = f(\bar{x}(y), y) \leq f(x, y)$$

holds for $(x, y) \in W \cap M$. This completes the proof. ■

In the following we assume the linear independence constraint qualification at \hat{x} , which is stronger than (V1):

$$\text{The set } \{ D_x g_j(\hat{x}, \hat{y}), j \in J_1^0(\hat{y}) \} \text{ is linearly independent.} \tag{V2}$$

Under the assumption of (V2) we can sharpen Theorem 3.1 in the sense that the validity of the well-known strong second-order sufficient condition SSOSC (cf. [14]) in (P_y) and a *generalized* SSOSC in problem (2.3) even implies the SSOSC for the original problem (P).

REMARK 3.2. (cf. [8],[9]). Under the condition (V2) the vector function $\hat{u}(y)$ is unique and both $\hat{x}(y)$ and $\hat{u}(y)$ are Lipschitzian on a neighborhood Y of \hat{y} in case of strong stability. Moreover, the *strong regularity* of a stationary point \hat{x} of $(P_{\hat{y}})$ in the sense of Robinson [14] (the Kuhn-Tucher relation to $(P_{\hat{y}})$ can be written as a generalized equation) is just equivalent to the *strong stability* of \hat{x} , again under the condition (V2). If, in addition, \hat{x} is even a local minimum, then under (V2) the strong regularity as well as the strong stability is equivalent to the SSOSC.

Let $J_1^+ := \{ j \in J_1^0(y) | \hat{u}_j(y) > 0 \}$ denote the index set of strictly binding constraints.

Following [8], the strong stability of $\hat{x} := \hat{x}(\hat{y})$ implies that every point $\hat{x}(y)$ for y in a neighborhood of \hat{y} is again a strongly stable local minimum for (P_y) . Also, the Jacobian matrix with respect to (x, u) of the system

$$\left. \begin{aligned} D_x \left[f(x, y) + \sum_{j \in J_1^+(\hat{y})} u_j g_j(x, y) \right] &= 0, \\ g_j(x, y) &= 0, \quad j \in J_1^+(\hat{y}), \end{aligned} \right\}$$

is regular at $(\hat{x}, \hat{y}, \hat{u})$. Hence, in view of the implicit function theorem, in an open neighborhood $U(\hat{y})$ there is a unique C^1 -vector valued function $(x^+(y), u^+(y))$ with $\hat{x}(\hat{y}) = x^+(\hat{y})$ and

$$\hat{u}_j(\hat{y}) = \begin{cases} u_j^+(\hat{y}), & j \in J_1^+(\hat{y}), \\ 0, & \text{for other } j. \end{cases}$$

Moreover, $x^+(y)$ is precisely a strongly stable local minimum for the reduced problem (P_y^+) :

$$\min_x \{ f(x, y) | g_j(x, y) = 0, j \in J_1^+(\hat{y}) \}. \tag{P_y^+}$$

Now we have two (locally marginal) functions, namely

$$\phi^+(y) := f(x^+(y), y) \quad \text{and} \quad \hat{\phi}^+(y) := f(\hat{x}(y), y). \tag{3.1}$$

The function ϕ^+ is a C^2 -function with the following derivatives:

$$D\phi^+(y) = D_y [f(x^+(y), y) + u^+(y)^T g^+(x^+(y), y)], \tag{3.2}$$

where $g^+(x, y) = (g_j(x, y), j \in J_1^+(\hat{y}))^T$, being a column vector

$$\begin{aligned} D^2\phi^+(y) &= D_x D_y^T [f(x^+(y), y) + u^+(y)^T g^+(x^+(y), y)] \cdot Dx^+(y) \\ &\quad + D_y^2 [f(x^+(y), y) + u^+(y)^T g^+(x^+(y), y)] \\ &\quad + [D_y^T g^+(x^+(y), y)] \cdot Du^+(y) \end{aligned} \tag{3.3}$$

The function $\hat{\phi}$ in (3.1) serves as a function in the “second phase”. In view of Remark 3.1 and Lemma 2.1 we see that $\hat{\phi}(y)$ is differentiable at \hat{y} . However, $\hat{\phi}$ need not be a C^2 -function. The next lemma provides a sufficient condition in order that \hat{y} is a local minimum for $\hat{\phi}(y)$ with respect to the set N_0 , $N_0 = \{y | g_j(y) \leq 0, j \in J_2\}$.

LEMMA 3.1. *Under the assumption of condition (V2), let $\hat{x}(y)$ with associated Lagrange multiplier $\hat{u}(y)$ be a strongly stable local minimum for (P_y) , for all y in an open neighborhood $U(\hat{y})$ of \hat{y} . Suppose that there exists a multiplier $\hat{u}^{(2)}$ such that (3.4), (3.5) hold (generalized SSOSC):*

$$D_y \left[\hat{\phi}(\hat{y}) + \sum_{j \in J_2} \hat{u}_j^{(2)} g_j(\hat{y}) \right] = 0, \tag{3.4.a}$$

$$\left. \begin{aligned} g_j(\hat{y}) &\leq 0, \\ \hat{u}_j^{(2)} \cdot g_j(\hat{y}) &= 0, \\ \hat{u}_j^{(2)} &\geq 0, \end{aligned} \right\} \quad j \in J_2, \tag{3.4.b}$$

$$\tag{3.4.c}$$

$$\tag{3.4.d}$$

$$z^T \left[D_2 \phi^+(\hat{y}) + \sum_{j \in J_2} \hat{u}_j^{(2)} D^2 g_j(\hat{y}) \right] z > 0 \quad \text{for all } z \in T_{\hat{y}}, z \neq 0 \quad \text{where} \tag{3.5}$$

$$T_{\hat{y}} = \{z | Dg_j(\hat{y})z = 0, j \in J_2^+\}, \quad J_2^+ = \{j \in J_2^0 | \hat{u}_j^{(2)} > 0\},$$

$$J_2^0 = \{j \in J_2 | g_j(\hat{y}) = 0\}.$$

Then, the point \hat{y} is a strict local minimum for $\hat{\phi}(y)$ with respect to the set N_0 .

PROOF. Note that $\hat{u}(y)$ is continuous. Consequently, we obtain the following obvious chain (locally): $J_1^+(\hat{y}) \subset J_1^0(y) \subset J_1^0(\hat{y})$. Hence, for the functions $\phi^+(y)$, $\hat{\phi}(y)$ we obtain, using the fact that both $x^+(\hat{y})$ is strongly stable for $(P_{\hat{y}}^+)$ and $\hat{x}(\hat{y})$ is strongly stable for $(P_{\hat{y}})$

$$\phi^+(y) \leq \hat{\phi}(y) \quad (\text{locally}) \quad \text{and} \tag{3.6.a}$$

$$\phi^+(\hat{y}) = \hat{\phi}(\hat{y}). \tag{3.6.b}$$

Putting $\Delta(y) = \hat{\phi}(y) - \phi^+(y)$, we see that $\Delta(y)$ has a local minimum at \hat{y} . Consequently, $D\Delta(\hat{y}) = 0$, which means:

$$D\phi^+(\hat{y}) = D\hat{\phi}(\hat{y}). \tag{3.7}$$

If we substitute (3.7) into (3.4.a), then (3.4), (3.5) imply that the strong second-order sufficient condition holds for $\phi^+(y)$ at \hat{y} with respect to N_0 . Hence, \hat{y} is a strict local minimum for $\phi^+(y)$ with respect to N_0 . Taking (3.6) into account we see that \hat{y} is also a strict local minimum for $\hat{\phi}(y)$ with respect to N_0 . ■

For the second derivative of $\phi^+(y)$ at \hat{y} we need an explicit formula. In view of formula (3.3) we have to calculate $Dx^+(\hat{y})$ and $Du^+(\hat{y})$.

As an abbreviation we put:

$$A = D_x^2 [f(x^+(\hat{y}), \hat{y}) + u^+(\hat{y})^\top g^+(x^+(\hat{y}), \hat{y})], \tag{3.8.a}$$

$$B = D_x^\top g^+(x^+(\hat{y}), \hat{y}), \tag{3.8.b}$$

$$C = D_y D_x^\top [f(x^+(\hat{y}), \hat{y}) + u^+(\hat{y})^\top g^+(x^+(\hat{y}), \hat{y})], \tag{3.8.c}$$

$$D_0 = D_y^\top g^+(x^+(\hat{y}), \hat{y}), \tag{3.8.d}$$

$$E_0 = D_y^2 [f(x^+(\hat{y}), \hat{y}) + u^+(\hat{y})^\top g^+(x^+(\hat{y}), \hat{y})], \tag{3.8.e}$$

$$F = [D_y^\top g_j(\hat{y}), j \in J_2^+]. \tag{3.8.f}$$

From the definition of $(x^+(y), u^+(y))$ by means of the implicit function theorem we obtain:

$$\begin{pmatrix} Dx^+(\hat{y}) \\ Du^+(\hat{y}) \end{pmatrix} = - \begin{pmatrix} A & B \\ B^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} C \\ D_0^\top \end{pmatrix}. \tag{3.9}$$

For an explicit formula of the inverse matrix in (3.9) we use the following (more general) lemma. The symbol “ \dagger ” will denote the Moore-Penrose inverse (cf. [13]).

LEMMA 3.2. *Let A be a symmetric (n, n) matrix and B an (n, m) matrix with $m \leq n$. Let \bar{B} denote an $(n, n - l)$ matrix whose columns form a basis for the linear space $T_x = \{x | B^\top x = 0\}$, where $l = \text{rank}(B)$.*

a. *If $\text{rank}(B) = m$, then we have:*

$$R := \begin{pmatrix} A & B \\ B^\top & 0 \end{pmatrix} \text{ is regular iff } \bar{B}^\top A \bar{B} \text{ is regular.}$$

b. *If $\bar{B}^\top A \bar{B}$ is regular, then we have:*

$$\begin{pmatrix} A & B \\ B^\top & 0 \end{pmatrix}^\dagger = \begin{pmatrix} \bar{B}(\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top & [I - \bar{B}(\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top A](B^\dagger)^\top \\ B^\dagger [I - A \bar{B}(\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top] & -B^\dagger A [I - \bar{B}(\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top A](B^\dagger)^\top \end{pmatrix}, \tag{3.10}$$

where I stands for the identity matrix.

PROOF. If $\text{rank}(B) = m$, then the matrices

$$Z_1 = \begin{pmatrix} B & \bar{B} & 0 \\ 0 & 0 & I \end{pmatrix}, \quad Z_2 = \begin{pmatrix} 0 & \bar{B} & B \\ I & 0 & 0 \end{pmatrix}, \quad B^\top B$$

are regular and statement a is following directly from the identity:

$$Z_1^\top R Z_2 = \begin{pmatrix} B^\top B & B^\top A \bar{B} & B^\top A B \\ 0 & \bar{B}^\top A \bar{B} & \bar{B}^\top A B \\ 0 & 0 & B^\top B \end{pmatrix}.$$

Moreover, after inversion of this identity and taking into account the following formula

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} \\ 0 & X_{22} & X_{23} \\ 0 & 0 & X_{33} \end{pmatrix}^{-1} = \begin{pmatrix} X_{11}^{-1} & -X_{11}^{-1}X_{12}X_{22}^{-1} & -X_{11}^{-1}[X_{13} - X_{12}X_{22}^{-1}X_{23}]X_{33}^{-1} \\ 0 & X_{22}^{-1} & -X_{22}^{-1}X_{23}X_{33}^{-1} \\ 0 & 0 & X_{33}^{-1} \end{pmatrix}$$

we get the inverse R^{-1} as a special case of (3.10) with $B^\dagger = (B^\top B)^{-1}B^\top$ (cf. McCormick [12, p. 243]).

Next, suppose that $\text{rank}(B) = l < m$. Then we write B as a product:

$$B = PQ^\top$$

where P , resp. Q is an (n, l) , resp. (m, l) matrix with $\text{rank}(P) = \text{rank}(Q) = l$. Moreover, Q is chosen such that $Q^\top Q = \text{Identity}$ (use a singular value decomposition of B). The matrix R can be written as a (congruence) product:

$$R = \begin{pmatrix} I & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} A & P \\ P^\top & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & Q^\top \end{pmatrix} = R_1 R_2 R_3.$$

Obviously, $\text{rank}(R_1) = \text{rank}(R_3) = n + l$ and we have:

$$R_1^\dagger = \begin{pmatrix} I & 0 \\ 0 & Q^\top \end{pmatrix}, \quad R_3^\dagger = \begin{pmatrix} I & 0 \\ 0 & Q \end{pmatrix}. \tag{3.11}$$

From the definition of P and Q we see that $T_x = \{x | P^\top x = 0\}$. Then application of statement a yields that the matrix R_2 is regular. Consequently, for the computation of the inverse R_2^{-1} we can apply Formula (3.10), replacing B by P . Since $R_1^\top R_1 = \text{Identity}$ and $R_3 = R_1^\top$, we have

$$R^\dagger = R_3^\dagger R_2^{-1} R_1^\dagger. \tag{3.12}$$

For B^\dagger we obtain the formula $B^\dagger = QP^\dagger$. Now, formula (3.10) is obtained by a combination of all partial results. ■

For the second derivative of $\phi^+(y)$ at \hat{y} we now obtain the following formula, by means of a combination of (3.3), (3.8), (3.9) and (3.10). The appearing matrix \bar{B} has the same meaning as in Lemma 3.2:

$$\begin{aligned} D^2\phi^+(\hat{y}) &= E_0 + C^\top \cdot Dx^+(\hat{y}) + D_0 \cdot Du^+(\hat{y}) \\ &= \left\{ \begin{array}{l} E_0 - C^\top \bar{B} (\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top C - C^\top M (D_0 B^\dagger)^\top \\ -D_0 B^\dagger M^\top C + D_0 B^\dagger A M (D_0 B^\dagger)^\top \end{array} \right\}, \end{aligned} \tag{3.13}$$

where $M = I - \bar{B} (\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top A$.

For the final theorem we need one more technical lemma which generalizes a result of Fujiwara et al. [6]. Note the similarity between (3.13) and the formula (3.14) below.

LEMMA 3.3. *Let $A(n_1, n_1), B(n_1, m), C(n_1, n_2), D(n_2, m), E(n_2, n_2), F(n_2, r)$ be matrices with dimensions as given within the parentheses.*

Let A, E be symmetric and $\text{rank}(B) = m$. Define the following linear subspaces:

$$T = \{(x, y) \in \mathbb{R}^{n_1+n_2} \mid B^\top x + D^\top y = 0, F^\top y = 0\},$$

$$T_x = \{x \in \mathbb{R}^{n_1} \mid B^\top x = 0\}, \quad T_y = \{y \in \mathbb{R}^{n_2} \mid F^\top y = 0\}.$$

Let \bar{B} be an $(n_1, n_1 - m)$ matrix whose columns form a basis for T_x , and let \bar{F} be an $(n_2, n_2 - l)$ matrix whose columns form a basis for T_y , where $l = \text{rank}(F)$.

Then the matrix

$$\begin{pmatrix} A & C \\ C^\top & E \end{pmatrix}$$

is positive definite on T iff both $\bar{B}^\top A \bar{B}$ and G are positive definite, where

$$G = \bar{F}^\top \tilde{G} \bar{F}, \quad \tilde{G} = \begin{pmatrix} E - C^\top \bar{B} (\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top C - C^\top M (DB^\dagger)^\top \\ -DB^\dagger M^\top C + DB^\dagger A M (DB^\dagger)^\top \end{pmatrix}, \quad (3.14)$$

where $M = I - \bar{B} (\bar{B}^\top A \bar{B})^{-1} \bar{B}^\top A$.

PROOF. Firstly, we decompose T into a direct sum, i.e. $T = T_1 + T_2, T_1 \cap T_2 = \{0\}$.

$$T = \{(x, y) \mid B^\top x + D^\top y = 0, F^\top y = 0, x = \bar{B}v + Bw, v \in \mathbb{R}^{n_1-m}, w \in \mathbb{R}^m\}$$

$$= \left\{ (x, y) \mid \begin{matrix} x = \bar{B}v \\ y = 0 \end{matrix} \right\} \oplus \left\{ (x, y) \mid \begin{matrix} x = Bw \\ B^\top x + D^\top y = 0, F^\top y = 0 \end{matrix} \right\} := T_1 \oplus T_2.$$

Next, we simplify the representation of T_2 :

$$T_2 = \{(x, y) \mid x = Bw, y = \bar{F}z, z \in \mathbb{R}^{n_2-l}, B^\top Bw + D^\top y = 0\}$$

$$= \left\{ (x, y) \mid \begin{matrix} x = Bw, w = -(B^\top B)^{-1} D^\top \bar{F}z \\ y = \bar{F}z \end{matrix} \right\}$$

$$= \left\{ (x, y) \mid \begin{matrix} x = -B(B^\top B)^{-1} D^\top \bar{F}z \\ y = \bar{F}z \end{matrix} \right\}.$$

Note that $B^\dagger = (B^\top B)^{-1} B^\top$. Hence, the columns of the following matrix form a basis for T :

$$\begin{pmatrix} \bar{B} & -(DB^\dagger)^\top \bar{F} \\ 0 & \bar{F} \end{pmatrix}.$$

Consequently, the blockmatrix

$$\begin{pmatrix} A & C \\ C^\top & E \end{pmatrix}$$

is positive definite on T iff the following matrix is positive definite:

$$\begin{pmatrix} \bar{B}^\top & 0 \\ -\bar{F}^\top DB^\dagger & \bar{F}^\top \end{pmatrix} \cdot \begin{pmatrix} A & C \\ C^\top & E \end{pmatrix} \cdot \begin{pmatrix} \bar{B} & -(DB^\dagger)^\top \bar{F} \\ 0 & \bar{F} \end{pmatrix}. \quad (3.15)$$

The product matrix in (3.15) is again a matrix with blockstructure, say

$$Z = \begin{pmatrix} U & V \\ V^\top & W \end{pmatrix}.$$

From Fujiwara et al. [6] we see that such a blockmatrix is positive definite iff both U and the Schur complement of U in $Z, W - V^\top U^{-1}V$, are positive definite. Now, the assertion of the lemma follows from a simple computation, noting that $\bar{B}^\top A \bar{B} = U$ and $G = W - V^\top U^{-1}V$. ■

Finally we can state and prove a theorem which gives an additional answer to Problem B.

THEOREM 3.2. *Let \hat{x} be a strongly stable local minimum for $(P_{\hat{y}})$ at which condition (V2) is fulfilled. Let further Y be an open neighborhood of \hat{y} and let $\hat{x}(y)$ be a continuous vector function on Y such that $\hat{x}(y)$ is a stationary point for (P_y) and $\hat{x}(\hat{y}) = \hat{x}$. If at \hat{y} the generalized SSOSC relations (3.4), (3.5) hold, then the strong second-order sufficient condition holds at (\hat{x}, \hat{y}) with respect to the original problem (P).*

PROOF. Firstly, we note that Theorem 2.1 is applicable in view of (3.4) and Remark 3.2. Hence, (\hat{x}, \hat{y}) is a stationary point for Problem (P).

Now we have to verify that the matrix

$$\begin{pmatrix} A & C \\ C^\top & E \end{pmatrix}$$

is positive definite on T , where

$$T = \left\{ (x, y) \mid \begin{matrix} B^\top x + D_0^\top y = 0 \\ F^\top y = 0 \end{matrix} \right\},$$

and A, B, C, D_0, F as in (3.8), whereas $E = E_0 + \sum_{j \in J_2} \hat{u}_j^{(2)} D^2 g_j(\hat{y})$.

In order to verify this restricted positive definiteness, we note that A is positive definite on $T_x = \{x \mid B^\top x = 0\}$, since $\hat{x}(\hat{y})$ is a strongly stable local minimum for $(P_{\hat{y}})$. Hence, $\bar{B}^\top A \bar{B}$ is positive definite, where \bar{B} is a matrix whose columns form a basis for T_x . Furthermore, we see that formula (3.5) is equivalent with the positive definiteness of the matrix H ,

$$H = \bar{F}^\top \left[D^2 \phi^+(\hat{y}) + \sum_{j \in J_2} \hat{u}_j^{(2)} D^2 g_j(\hat{y}) \right] \bar{F}, \tag{3.16}$$

where \bar{F} is a matrix whose columns form a basis for the space $T_y = \{y \mid F^\top y = 0\}$. By comparing (3.13), (3.14), (3.16), we see that the matrix H in (3.16) has the form of the matrix G in (3.14). So, G is positive definite, and together with the positive definiteness of $\bar{B}^\top A \bar{B}$, the above verification follows immediately from Lemma 3.3. ■

REMARK 3.3. If we add to the assumptions of Theorem 3.2 the linear independence condition (V2) applied to the local optimal solution \hat{y} for $\hat{\phi}$ with respect to the set N_0 then the point (\hat{x}, \hat{y}) is even a strongly stable local minimum for problem (P).

Acknowledgement. The authors are indebted to W. Schäfer for the initiative which has led to the origin of this paper. Moreover, the authors would like to thank P. Jonker, D. Klatte and the anonymous referees for several suggestions which consistently improved the original manuscript.

References

- [1] Bank, B., Mandel, R. and Tammer K. (1979). Parametrische Optimierung und Aufteilungsverfahren. In K. Lommatzsch (Ed.), *Anwendungen der linearen parametrischen Optimierung*. Akademie-Verlag, Berlin, 107–123.
- [2] Beer, K. (1977). *Lösung grosser linearer Optimierungsaufgaben*. Akademie-Verlag, Berlin.
- [3] Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton, N.J.
- [4] Benders, J. F. (1962). Partitioning Procedures for Solving Mixed-Variables Programming Problems. *Numer. Math.* **4** 238–252.
- [5] Clarke, F. H. (1975). Generalized Gradients and Applications. *Trans. Amer. Math. Soc.* **205** 247–262.
- [6] Fujiwara, O., Han, S.-P. and Mangasarian, O. L. (1984). Local Duality in Nonlinear Programs. *SIAM J. Control Optim.* **22** 162–169.
- [7] Gauvin, J. (1982). The Generalized Gradient of the Perturbation Function in Mathematical Programming. In A. V. Fiacco (Ed.), *Mathematical Programming with Data Perturbations. I*. Marcel Dekker Inc., New York, Basel.
- [8] Kojima, M. (1980). Strongly Stable Stationary Solutions in Nonlinear Programs. In S. M. Robinson (Ed.), *Analysis and Computation of Fixed Points*. Academic Press, New York.
- [9] _____ and Hirabayashi, R. (1980). Some Results on the Strong Stability in Nonlinear Programs. Technical Report No. 4, Tokyo Institute of Technology, Dept. Management Science and Engineering.
- [10] Lasdon, L. S. (1970). *Optimization Theory for Large Systems*. The Macmillan Co., New York.
- [11] Luderer, B. (1984). Über die Äquivalenz nichtlinearer Optimierungsaufgaben. *Wiss. Z. Tech. Hochsch. Karl-Marx-Stadt* **26**, 2 257–258.
- [12] McCormick, G. P. (1983). *Nonlinear Programming—Theory, Algorithms, and Applications*. John Wiley and Sons, New York.
- [13] Noble, B. (1969). *Applied Linear Algebra*. Prentice-Hall, Inc., Englewood Cliff, N.J.
- [14] Robinson, S. M. (1980). Strongly Regular Generalized Equations. *Math. Oper. Res.* **5** 43–62.
- [15] Rockafellar, R. T. (1981). *The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions*. Berlin, Heldermann-Verlag.

JONGEN: DEPARTMENT OF APPLIED MATHEMATICS, PO BOX 217, TWENTE UNIVERSITY OF TECHNOLOGY, ENSCHEDE, THE NETHERLANDS

MÖBERT & TAMMER: SEKTION MATHEMATIK UND RECHENTECHNIK, POB 66, TECHNISCHE HOCHSCHULE LEIPZIG, 7030 LEIPZIG, GERMAN DEMOCRATIC REPUBLIC