JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 34, No. 1, MAY 1981

Algorithms for a Class of Nondifferentiable Problems[1]

G. PAPAVASSILOPOULOS[2]

Communicated by D. G. Luenberger

Abstract. A nonlinear programming problem with nondifferentiabilities is considered. The nondifferentiabilities are due to terms of the form $\min(f_1(x), \ldots, f_n(x))$, which may enter nonlinearly in the cost and the constraints. Necessary and sufficient conditions are developed. Two algorithms for solving this problem are described, and their convergence is studied. A duality framework for interpretation of the algorithms is also developed.

Key Words. Nonlinear programming, nondifferentiable optimization, algorithms, min–max problems, duality.

1. Introduction

The present paper deals with algorithms for finding the minimum of a problem with nondifferentiable cost functional and constraints. All the functions involved have domains and ranges in finite-dimensional Euclidean spaces.

Much research has been conducted in the area of nondifferentiable optimization, and more remains to be done. As expected, almost all the methods proposed until now tend to exploit the knowledge which is available for the differentiable case. Some of them use generalizations of notions which exist for the differentiable case. The subgradient and ϵ-subgradient methods (Refs. 7, 14) use the notions of subgradient and ϵ-subgradient of a function at a point, which generalize the familiar notion of the derivative,

[1] This work was supported in part by the National Science Foundation under Grant No. ENG-74-19332 and Grant No. ECS-79-19396, in part by the U.S. Air Force under Grant AFOSR-78-3633, and in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract N00014-79-C-0424.

[2] Graduate Student, Decision and Control Laboratory, Coordinated Science Laboratory and Department of Electrical Engineering, University of Illinois, Urbana, Illinois. Presently, Assistant Professor, Department of Electrical Engineering, University of Southern California, Los Angeles, California.

0022-3239/81/0500-0041$03.00/0 © 1981 Plenum Publishing Corporation

and they treat nondifferentiable problems in a direct way. Some other methods, like the one that we propose here, try to reduce the whole problem to a differentiable one, treating it in an indirect way. Optimality conditions have been derived by several authors (Refs. 9, 11, 12), and it has been shown that convexity theory is particularly helpful in establishing such conditions. In this context, Clarke's work (Ref. 9) can be pointed out as a representative one. Algorithmic procedures have also been proposed and analyzed (Refs. 4, 5, 7, 13, 17–19). Many of these methods are of limited applicability, due to restrictive assumptions, such as convexity, or are applicable only to problems of a specific structure. Also, many of them, although theoretically interesting and intellectually pleasing, are quite cumbersome and complicated in practice. For this reason, the increasing demand for general-purpose methods makes algorithms for nondifferentiable optimization a topic of current interest.

The algorithms that we propose and analyze deal with problems of the form

$$\text{(NP) minimize } g[x, \max\{f_{11}(x), \ldots, f_{1m}(x)\}, \ldots, \max\{f_{k1}(x), \ldots, f_{km}(x)\}],$$

$$\text{subject to } h_j[x, \max\{f_{11}(x), \ldots, f_{1m}(x)\}, \ldots, \max\{f_{k1}(x), \ldots, f_{km}(x)\}] = 0, \quad j = 1, \ldots, q,$$

where the functions

$$f_{ij}: R^n \to R, \qquad g: R^{n+k} \to R, \qquad h_j: R^{n+k} \to R$$

are continuously differentiable. The nondifferentiability of g and h_j with respect to x is due to the presence of the terms

$$\max\{f_{i1}(x), \ldots, f_{im}(x)\}, \qquad i = 1, \ldots, k.$$

It is clear that quite a large number of cases of practical interest is covered by Problem (NP). The algorithms are conceptually simple to understand and practically easy to implement. They are related closely to the methods of multipliers. For a fairly complete account of multiplier methods,[3] we suggest two recent survey papers (Refs. 3 and 21). The basic idea on which our algorithms operate was introduced in Refs. 4–5; and hence, we consider this work as a continuation of Refs. 4–5.

For the sake of avoiding the use of complicated formulas and keeping the exposition simple, we give the proofs only for a simplified version of problem (NP). These proofs can be generalized easily for (NP). The results and the algorithm for (NP) are given in Section 5.

[3] Also called augmented Lagrangian methods.

We start, in Section 2, by stating the problem and developing necessary and sufficient conditions for optimality. These conditions, besides being of interest in their own right, are used in the sequel to establish certain results. In Section 3, we develop a duality framework. In Section 4, we introduce the algorithms, interpreting them as gradient methods (approximate steepest ascent and Newton's method) for solving the dual problem, and we prove convergence and convergence-rate results. In Section 5, we state the results and the algorithms for Problem (NP) without proofs. At the end, we have a conclusion section.

Abbreviations

w.r.t.	with respect to
w.l.o.g.	without loss of generality
(NP)	nondifferentiable problem
(NPS)	simplified nondifferentiable problem
(DP)	decomposed problem corresponding to (NP)
(DPS)	decomposed problem corresponding to (NPS)
(AS)	algorithm with steepest descent update
(AN)	algorithm with Newton update

Notation

R^n	denotes the n-dimensional real Euclidean, space and its elements are considered to be column vectors;
	denotes transposition of vector or matrix;
$\|\cdot\|$	for $x \in R^n$, $\|x\| = \sqrt{(x_1^2 + \cdots + x_n^2)}$;
$\|\cdot\|$	for A, $n \times m$ real matrix, $\|A\| = \sup\{\|Ax\| \mid x \in R^m, \|x\| = 1\}$;
$S(x, \epsilon)$	for $x \in R^n$ and $\epsilon > 0$, $S(x, \epsilon)$ denotes the open ball in R^n centred at x with radius ϵ, i.e., $S(x, \epsilon) = \{y \in R^n \mid \|x - y\| < \epsilon\}$;
$\bar{S}(x, \epsilon)$	denotes the closure of $S(x, \epsilon)$ in R^n, i.e., $\bar{S}(x, \epsilon) = \{y \in R^n \mid \|x - y\| \leq \propto\}$;
C^1, C^2	C^1 denotes the set of all continuously differentiable functions from one Euclidean space to another, and C^2 denotes the set of twice continuously differentiable functions.

Our definitions of local, strict local, global, strict global minimum of a real-valued function are the standard ones, see Ref. 15.

For a function $f: R^n \to R$, $f \in C^1$, we denote the gradient of f at $x \in R^n$ by $\nabla f(x)$, and we consider it to be a column vector in R^n. If in addition $f \in C^2$, we denote the Hessian of f at x by $\nabla^2 f(x)$.

For a function

$$f: R^n \to R^m, \qquad f(x) = (f_1(x), \ldots, f_m(x))', \qquad f \in C^1,$$

we denote the $n \times m$ matrix $[\nabla f_1(x) \vdots \cdots \vdots \nabla f_m(x)]$ by $\nabla f(x)$. Using this notation, we have that, if

$$g(x) = f(h(x)),$$

where

$$f: R^k \to R^m, \qquad h: R^n \to R^k, \qquad g = R^n \to R^m, \qquad g, f, h \in C^1,$$

then

$$\nabla g(x) = \nabla h(x) \nabla f(x).$$

If a function $f: R^{n+k} \to R$ is in C^1, $x \in R^n$, $z \in R^k$, we write

$$\nabla_x f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_n)', \qquad \nabla_z f = (\partial f/\partial z_1, \ldots, \partial f/\partial z_k), \qquad \nabla_i f = \partial f/\partial z_i.$$

If in addition $f \in C^2$, we write

$$\nabla_{xz} f = \begin{bmatrix} \partial^2 f/\partial x_1 \partial z_1 & \cdots & \partial^2 f/\partial x_1 \partial z_k \\ \vdots & & \vdots \\ \partial^2 f/\partial x_n \partial z_1 & \cdots & \partial^2 f/\partial x_n \partial z_k \end{bmatrix}, \qquad n \times k \text{ matrix},$$

$$\nabla_{zx} f = (\nabla_{xz} f)'.$$

If A is an $n \times m$ matrix, then $\mathcal{N}(A)$ and $\mathcal{R}(A)$ denote the null space and the range of space of A, respectively, i.e.,

$$\mathcal{N}(A) = \{x \mid x \in R^m \text{ and } Ax = 0\},$$

$$\mathcal{R}(A) = \{x \mid x \in R^n \text{ and } x = Aw \text{ for some } w \in R^m\}.$$

We also write

$$A = (a_{ij}),$$

where a_{ij} are the entries of A.

By $\{x^s\}_{s=1}^{\infty}$, we denote the sequence $x^1, x^2, \ldots$ of elements of R^n.

If A and M are matrices, vectors, or scalars and A depends on M,

$$A = O(M)$$

means that A is of order M; i.e., for some $K > 0$, $\delta > 0$, it holds that

$$\|A\|/\|M\| \leq K, \qquad \text{for all } M \neq 0, \quad \|M\| \leq \delta.$$

The symbol 0 is used to denote the zero of R, the zero vector of R^n, or a zero matrix.

2. Problem Statement, Optimality Conditions

2.1. Problem Statement. The problem that we are concerned with is the following:

$$\text{(NP)} \qquad \text{minimize } g[x, \gamma[f_1(x)], \ldots, \gamma[f_k(x)]],$$
$$\text{subject to } h_j[x, \gamma[f_1(x)], \ldots, \gamma[f_k(x)]] = 0, \qquad j = 1, \ldots, q,$$

where the functions

$$f_i : R^n \to R^m, \qquad i = 1, \ldots, k, \qquad g : R^{n+k} \to R, \qquad h_j : R^{n+k} \to R, \qquad j = 1, \ldots, q,$$

are in C^1. For the f_i's, we have

$$f_i = (f_{i1}, \ldots, f_{im})', \qquad i = 1, \ldots, k, \tag{1}$$

where

$$f_{ij} : R^n \to R, \qquad f_{ij} \in C^1, \qquad i = 1, \ldots, k, \qquad j = 1, \ldots, q.$$

The function $\gamma : R^n \to R^m$ is defined by

$$\gamma[t] = \max\{t_1, \ldots, t_m\}, \tag{2}$$

where

$$t = (t_1, \ldots, t_m)' \in R^m.$$

Our standing assumptions for Problem (NP) are that the optimum value g^* is finite and that f_i, g, h_j are in C^1. The function γ is not everywhere differentiable, which is the cause of the nondifferentiable character of the problem. We will refer to γ defined in (2) as a *kink*.

Let us also introduce the simplified unconstrained version of (NP)

$$\text{(NPS)} \quad \text{minimize } g[x, \gamma[f_1(x)], \ldots, \gamma[f_k(x)]],$$
$$\text{subject to } x \in R^n,$$

where the functions f_i,

$$f_i : R^n \to R, \qquad i = 1, \ldots, k, \tag{3}$$

and $g : R^{n+k} \to R$ are in C^1, and where $\gamma : R \to R$ is defined by

$$\gamma[t] = \max\{0, t\}, \qquad \text{for all } t \in R. \tag{4}$$

Our standing assumptions for Problem (NPS) are that the optimum value g^* is finite and that f_i, g are in C^1. We will refer to γ defined in (4) as a *simple kink*. We write

$$\gamma_i(x) = \gamma[f_i(x)] = \max\{f_{i1}(x), \ldots, f_{im}(x)\}, \qquad i = 1, \ldots, k, \tag{5}$$

for the kink γ of problem (NP) and

$$\gamma_i(x) = \gamma[f_i(x)] = \max\{0, f_i(x)\}, \qquad i = 1, \ldots, k, \tag{6}$$

for the simple kink γ of Problem (NPS).

In the case of (NP), we say that the function f_{ij} is *active at* x if

$$\gamma_i(x) = f_{ij}(x);$$

it is *inactive at* x if

$$\gamma_i(x) > f_{ij}(x).$$

For $x \in R^n$, we denote

$$I_i(x) = \{j \mid f_{ij}(x) = \gamma_i(x), j = 1, \ldots, m\}, \qquad i = 1, \ldots, k. \tag{7}$$

In the case of (NPS), we say that the function f_i is *active at* x if

$$f_i(x) = 0;$$

it is *inactive at* x if

$$f_i(x) \neq 0.$$

For $x \in R^n$, we denote

$$I(x) = \{j \mid f_i(x) = 0, i = 1, \ldots, k\}. \tag{8}$$

Notice that, for given x, $I(x)$ may be empty. Although we could consider as definition of active function, for the (NPS) case, the same that we gave for the (NP) case, we prefer to give the different but essentially equivalent definition (7), because it leads to simpler formulas.

We comment now on the formulation of Problem (NP). There is no loss of generality in requiring that all the f_i's take values in R^m, since a kink of *length* less than m can be transformed to a kind of length m. For example,

$$\max\{t_1, t_2, t_3\} = \max\{t_1, t_2, t_3, t_2 - 10, t_1 - 2\}.$$

We also assume w.l.o.g. that the functions $g, h_1, \ldots, h_q$ contain the same kinks, since we may add zero multiples of any missing kinks to $g, h_1, \ldots, h_q$.

2.2. Equivalences with Nonlinear Programming Problems. In Sections 2, 3, 4 we deal exclusively with Problem (NPS). Consequently, by f_i and γ we mean those defined in (3) and (4), respectively, for (NPS).

We now introduce a class of nonlinear programming problems which is closely related to Problem (NPS). Let $x^* \in R^n$, $I(x^*)$ be as in (8), and let J be any subset of $i(x^*)$, including the empty set. let

$$g_J(x) = g[x, \delta_1(x;J), \ldots, \delta_k(x;J)],$$

where

$$\delta_i(x;J) = \begin{cases} f_i(x), & \text{if } f_i(x^*) > 0, \\ 0, & \text{if } f_i(x^*) < 0, \\ 0, & \text{if } i \in I(x^*) \text{ and } i \in J, \\ f_i(x), & \text{if } i \in J. \end{cases}$$

Consider the problem

$$\text{(DPS-J)} \qquad \text{minimize } g_J(x),$$

$$\text{subject to } f_i(x) \geq 0, \qquad \text{if } i \in J,$$

$$f_i(x) \leq 0, \qquad \text{if } i \in I(x^*),\ i \notin J.$$

The following lemma shows the relationship between problems (NPS) and (DPS-J). The proof is straightforward and is left to the reader.

Lemma 2.1. A vector $x^* \in R^n$ is a strict local minimum of (NPS) iff x^* is a strict local minimum of (DPS-J), for every $J \subset I(x^*)$. Also, $g^* = g_J(x^*)$, for every $J \subset I(x^*)$.

We now proceed to obtain conditions for optimality for Problem (NPS) by exploiting its relation with Problem (DPS-J).

2.3. First-Order and Second-Order Necessary Conditions for Optimality. We first give a theorem which resembles the one given in Ref. 4 as Proposition 3.2.

Theorem 2.1. Let x^* be a local minimum of Problem (NPS). Then, there exist scalars $y_1^*, \ldots, y_k^*$, such that

$$\begin{aligned} &\nabla_x g + \sum_{i=1}^{k} y_i^* \nabla_i g \nabla f_i = 0, \\ &0 \leq y_i^* \leq 1, \qquad i = 1, \ldots, k, \end{aligned} \tag{9}$$

$$y_i^* = \begin{cases} 0, & \text{if } f_i(x^*) < 0, \\ 1, & \text{if } f_i(x^*) > 0\,. \end{cases} \tag{10}$$

If in addition the vectors ∇f_i, for which $\nabla_i g \neq 0$ and $f_i(x^*) = 0$, are linearly independent, then the scalars y_i^* are unique. All the gradients are calculated at x^*.

Proof. The proof is a direct application of Proposition 2.1 of Ref. 12. First, we show that the generalized gradients of g at any point x where

$$f_1(x)=0,\dots,f_n(x)=0,$$
$$f_{n+1}(x)>0,\dots,f_{n+\lambda}(x)>0,$$
$$f_{n+\lambda+1}(x)<0,\dots,f_k(x)<0$$

is a subset of

$$\left\{\nabla_x g+\sum_{i=1}^{n} y_i\nabla f_i+\sum_{j=1}^{\lambda}\nabla f_{n+j},\, 0\le y_i\le 1, i=1,\dots,n\right\},$$

and then we use the necessary condition proved in Ref. 12 that, if x^* is a local minimum of (NPS), then the zero vector must belong to the generalized subgradient of g at x^*. □

Theorem 2.2. Let x^* be a local minimum of Problem (NPS), at which the gradients $\nabla f_i(x^*)$, $i\in I(x^*)$, are linearly independent. Then, there are real numbers $y_1^*,\dots,y_k^*$, such that

$$\nabla_x g+\sum_{i=1}^{k} y_i^*\nabla_i g\nabla f_i=0,\qquad 0\le y_i^*\le 1,\quad i=1,\dots,k, \tag{11}$$

$$y_i^*=0,\quad \text{for all } i\notin I(x^*),\quad f_i(x^*)<0,\qquad y_i^*=1,\quad \text{for all } i\notin I(x^*),\quad f_i(x^*)>0, \tag{12}$$

$$\nabla_i g\ge 0,\qquad \forall i\in I(x^*). \tag{13}$$

furthermore, $y_1^*,\dots,y_k^*$ are such that the scalars $y_i^*\nabla_i g$, $i=1,\dots,k$, are unique. All the gradients are calculated at x^*.

Proof. Assume w.l.o.g. that every f_i is active at x^*. then, by Lemma 2.1, we have that x^* is a local minimum of the following two problems:

$$\text{minimize } g[x,f_1(x),\dots,f_k(x)],\qquad \text{subject to } f_i(x)\ge 0,\quad i=1,\dots,k, \tag{14}$$

and

$$\text{minimize } g[x,0,\dots,0],\qquad \text{subject to } f_i(x)\le 0,\quad i=1,\dots,k. \tag{15}$$

Since the gradients of the contraints of these two problems are linearly independent by hypothesis, we can write the corresponding first-order necessary Kuhn–Tucker conditions

$$\nabla_x g + \sum_{i=1}^{k} \nabla_i g \nabla f_i - \sum_{i=1}^{k} \mu_i \nabla f_i = 0, \tag{16}$$

$$\nabla_x g + \sum_{i=1}^{k} \lambda_i \nabla f_i = 0, \tag{17}$$

where μ_i, λ_i, $i = 1, \ldots, k$, are nonnegative real numbers. Using the linear independence of the ∇f_i's, $i = 1, \ldots, k$, and using (16)–(17), we obtain

$$\nabla_i g = \mu_i + \lambda_i, \qquad i = 1, \ldots, k, \tag{18}$$

which implies (13). Since $\mu_i \geq 0$ and $\lambda_i \geq 0$, (18) implies also that there is a $y_i^* \in R$ satisfying

$$0 \leq y_i^* \leq 1, \qquad \lambda_i = y_i^* \nabla_i g, \qquad \mu_i = (1 - y_i^*) \nabla_i g, \qquad i = 1, \ldots, k. \tag{19}$$

It is also clear that, if $\nabla_i g \neq 0$, then y_i^* is unique. □

It is evident that the linear independence of the gradients of the active f_i's at x^* is equivalent to the regularity of x^* for Problems (DPS-J), see Ref. 15. With this in mind, we give the following definition.

Definition 2.1. A point $x^* \in R^n$ is said to be a regular point of the f_i's, $i = 1, \ldots, k$, if the gradients $\nabla f_i(x^*)$, $i \in I(x^*)$, are linearly independent.

By first-order necessary conditions, we mean the conditions of Theorem 2.2 (i.e., under regularity, unless stated otherwise).

The next theorem gives second-order necessary conditions for optimality of a point $x^* \in R^n$ for Problem (NPS).

Theorem 2.3. Assume that x^* satisfies the hypothesis of Theorem 2.2 and that $g, f_1, \ldots, f_k \in C^2$. Let $y_1^*, \ldots, y_k^*$ be as in Theorem 2.2, and denote

$$Y^* = \begin{bmatrix} y_1^* & & 0 \\ & \ddots & \\ 0 & & y_k^* \end{bmatrix}, \qquad k \times k \text{ matrix}, \tag{20}$$

$$y^* = (y_1^*, \ldots, y_k^*)' \in R^k, \tag{21}$$

$$f = (f_1, \ldots, f_k)', \tag{22}$$

$$\begin{aligned}\Xi(x^*, y^*) = \nabla_{xx} g + \sum_{i=1}^{k} y_i^* \nabla_i g \nabla^2 f_i + \nabla f Y^* \nabla_{zx} g + \nabla_{xz} g Y^* \nabla f' \\ + \nabla f Y^* \nabla_{zz} g Y^* \nabla f'.\end{aligned} \tag{23}$$

Then,

$$w'\Xi(x^*, y^*)w \geq 0, \qquad \text{for all } w \in R^n, \text{ such that } w'\nabla f_i = 0, \quad \text{for all} \quad i \in I(x^*). \tag{24}$$

All derivatives are calculated at x^*.

Proof. Assume w.l.o.g. that

$$I(x^*) = \{1, \ldots, p\},$$
$$f_{p+1}(x^*) > 0, \ldots, f_q(x^*) > 0,$$
$$f_{q+1}(x^*) < 0, \ldots, f_k(x^*) < 0.$$

Let

$$\tilde{g}[x, \gamma[f_1(x)], \ldots, \gamma[f_p(x)]] = g[x, \gamma[f_1(x)], \ldots, \gamma[f_p(x)], f_{p+1}(x), \ldots, f_q(x), 0, \ldots, 0]. \tag{25}$$

By Lemma 2.1, x^* is a local minimum of the problem

$$\begin{aligned} &\text{minimize } \tilde{g}[x, 0, \ldots, 0], \\ &\text{subject to } f_i(x) \leq 0, \qquad i \in I(x^*). \end{aligned} \tag{26}$$

By hypothesis, ∇f_i, $i \in I(x^*)$, are linearly independent; hence, there are nonnegative real numbers $\lambda_1, \ldots, \lambda_p$, such that

$$\nabla_x \tilde{g} + \sum_{i=1}^{p} \lambda_i \nabla f_i = 0, \tag{27}$$

$$w'\left[\nabla_{xx}\tilde{g} + \sum_{i=1}^{p} \lambda_i \nabla^2 f_i\right] w \geq 0, \qquad \text{for all } w \in R^n,\ w'\nabla f_i = 0, \quad i \in I(x^*). \tag{28}$$

By differentiation of $\tilde{g}$ in (26), we obtain

$$\nabla_x \tilde{g} = \nabla_x g + \sum_{i=p+1}^{q} \nabla_i g \nabla f_i, \tag{29}$$

$$\nabla_{xx}\tilde{g} = \nabla_{xx} g + \sum_{i=p+1}^{q} \{\nabla_{xi} g \nabla f_i' + \nabla f_i \nabla_{ix} g + \nabla_i g \nabla^2 f_i\} + \sum_{i,j=p+1}^{q} \nabla f_i \nabla_{ij} g \nabla f_j'. \tag{30}$$

From (27), (29), (11), we obtain

$$\lambda_i = y_i^* \nabla_i g, \qquad i \in I(x^*). \tag{31}$$

The desired result follows from (28), (30), (31). □

Notice that the contribution of $\nabla^2 f_i$, ∇f_i, $i = q+1, \ldots, k$, to the quantity (23) is zero, since they are multiplied by $y_i^* = 0$.

It should be pointed out, that although the Hessians of the Lagrangians of Problems (DPS-J), calculated at x^*, y^*, are different, they induce the same quadratic on the subspace

$$\mathcal{M} = \{w \mid w \in R^n,\ w'\nabla f_i = 0,\ i \in I(x^*)\}. \tag{32}$$

2.4. Second-Order Sufficiency Conditions for Optimality. The second-order sufficiency conditions can be proved similarly to the corresponding second-order necessary conditions.

Theorem 2.4. Assume that $g, f_1, \ldots, f_k \in C^2$ and that

$$x^* \in R^n, \qquad y^* = (y_1^*, \ldots, y_k^*)' \in R^k$$

satisfy

$$\nabla_x g + \sum_{i=1}^{k} y_i^* \nabla_i f_i = 0,$$

$$y_i^* = 1, \qquad \text{if } f_i(x^*) > 0,$$

$$y_i^* = 0, \qquad \text{if } f_i(x^*) < 0,$$

$$0 < y_i^* < 1, \qquad \text{if } i \in I(x^*), \tag{33}$$

$$\nabla_i g > 0, \qquad \text{if } i \in I(x^*), \tag{34}$$

$$w'\Xi(x^*, y^*)w > 0, \qquad \text{for all } w \in M,\ w \neq 0. \tag{35}$$

Then, x^* is a strict local minimum of (NPS). All derivatives are calculated at x^*.

It is clear that the second-order sufficiency conditions correspond to those given for the classical nonlinear programming problem under strict complementarity. Slightly different second-order sufficiency conditions would have resulted from assuming that

$$0 \leq y_i \leq 1, \qquad i = 1, \ldots, k, \qquad \nabla_i g \geq 0, \qquad i \in I(x^*),$$

instead of (33)–(34), and by considering

$$\mathcal{M} = \{w \mid w'\nabla f_i = 0,\ i \in I(x^*),\ y_i \nabla_i g > 0\},$$

instead of $\mathcal{M}$ as in (32).

Finally, note that, if x^*, y^* satisfy the second-order sufficiency conditions, then the problem

$$\text{minimize } g(x, z_1, \ldots, z_k),$$

$$\text{subject to } f_i(x) \leq z_i, \qquad i = 1, \ldots, k,$$

has a strict local minimum at (x^*, y^*), where

$$z_i^* = \gamma[f_i(x^*)],$$

with associated Kuhn–Tucker vector

$$\mu^* = (y_1^* \nabla_1 g, \ldots, y_k^* \nabla_k g).$$

3. Duality

In this section, we develop a duality framework which will serve to motivate and interpret the algorithms. Let us introduce the function $P_c(\cdot, y): R \to R$ of t,

$$P_c(t, y) = \inf\{\gamma[t-u] + yu + (c/2)u^2 \mid u \in R\}, \tag{36}$$

where y and c are fixed real numbers and $c > 0$. This function was originally introduced and studied in Refs. 4–5 in a more general framework. It is real valued, convex, and continuously differentiable in t. The infimum in (36) is achieved at a unique point u^* for every $t \in R$; thus, we can use minimum, instead of infimum, in (36). The function $p_c(\cdot, y)$ (see Fig. 1) and its derivative can be calculated explicitly (see Refs. 4–5):

$$p_c(t, y) = \begin{cases} t - (1-y)^2/2c, & \text{if } t \geq (1-y)/c, \\ -y^2/2c, & \text{if } t \leq -y/c, \\ yt + (c/2)t^2, & \text{if } -y/c \leq t \leq (1-y)/c, \end{cases} \tag{37}$$

$$\nabla_t p_c(t, y) = \begin{cases} 1, & \text{if } t \geq (1-y)/c, \\ 0, & \text{if } t \leq -y/c, \\ y + ct, & \text{if } -y \leq t \leq (1-y)/c. \end{cases} \tag{38}$$

By using (37), it is easily verified that, if $y \in \Lambda \subset R$, bounded, then

$$K/c \leq p_c(t, y) - \gamma(t) \leq 0, \qquad \text{for all } t \in R,\ y \in \Lambda, \tag{39}$$

where K is a fixed scalar which depends on Λ.

Let us introduce the function F defined as

$$F(x, y, c) = g[x, p_c[f_1(x), y_1], \ldots, p_c[f_k(x), y_k]], \tag{40}$$

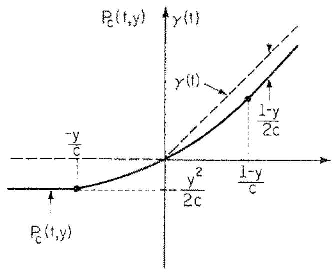

Fig. 1. The functions $p_c(t, y)$ and $\gamma(t)$.

where

$$y = (y_1, \ldots, y_k)' \in R^k.$$

The gradient of F w.r.t. x is

$$\nabla_x F(x, y, c) = \nabla_x g + \sum_{i=1}^{k} \tilde{y}_i \nabla_i g \nabla f_i, \tag{41}$$

where

$$\tilde{y}_i = y_i(x) = \begin{cases} 1, & \text{if } f_i(x) \geq (1 - y_i)/c, \\ 0, & \text{if } f_i(x) \leq -y_i/c, \\ y_i + cf_i(x), & \text{if } -y_i/c \leq f_i(x) \leq (1 - y_i)/c. \end{cases} \tag{42}$$

Theorem 3.1. Assume that $x^* \in R^n$ and $y^* \in R^k$, $y^* = (y_1^*, \ldots, y_k^*)'$ satisfy the first-order necessary conditions (Theorem 2.2) for x^* to be a local minimum of Problem (NPS). Then, for all $c > 0$,

$$\nabla_x F(x^*, y^*, c) = 0. \tag{43}$$

Proof. The hypothesis and (12), (37) yield

$$p_c[f_i(x^*), y_i^*] = \gamma[f_i(x^*)]. \tag{44}$$

The result now follows from (11) and (41). □

Although F is continuously differentiable, it is not twice continuously differentiable w.r.t. x, as (41) and (42) show. Nonetheless, we have the following theorem.

Theorem 3.2. Let x^*, y^* be as in Theorem 3.1, and assume in addition that they satisfy the second-order sufficiency conditions for x^* to be a strict local minimum of Problem (NPS). Then, there exists an $\epsilon_1 = \epsilon_1(c) > 0$, such that, for all $(x, y) \in S((x^*, y^*), \epsilon_1)$, the function $F(x, y, c)$ is twice continuously differentiable w.r.t. x. Also, there exist scalars $c^* \geq 0$ and $\epsilon > 0$, such that, for all $c \in [c^*, \bar{c}]$ and $(x, y) \in S((x^*, y^*), \epsilon)$, the Hessian $\nabla_{xx}F(x, y, c)$ is positive definite, where $\bar{c}$ is an arbitrarily large fixed constant.

Proof. By hypothesis, (33) holds. The continuity of the f_i's together with (33) and (42), guarantees that, for (x, y) sufficiently close to (x^*, y^*),

$$\tilde{y}_i = \tilde{y}_i(x) = \begin{cases} 1, & \text{if } f_i(x^*) > 0, \\ 0, & \text{if } f_i(x^*) < 0, \\ y_i + cf_i(x), & \text{if } f_i(x^*) = 0, \\ 0 < \tilde{y}_i < 1, & \text{if } f_i(x^*) = 0. \end{cases} \tag{45}$$

Consequently, $\nabla_x F(x, y, c)$ is differentiable w.r.t. x for $(x, y) \in S((x^*, y^*), \epsilon_1)$, for some $\epsilon_1 = \epsilon_1(c) > 0$. Similarly as in (45), we have

$$\nabla_x p_c(f_i(x), y_i) = \begin{cases} \nabla f_i(x), & \text{if } f_i(x^*) > 0, \\ 0, & \text{if } f_i(x^*) < 0, \\ (y_i + cf_i(x))\nabla f_i(x), & \text{if } f_i(x^*) = 0. \end{cases} \tag{46}$$

Direct calculation yields

$$\nabla_{xx}F(x, y, c) = \nabla_{xx}g + \sum_{i=1}^{k} \tilde{y}_i \nabla_i g \nabla^2 f_i + \nabla f \tilde{Y} \nabla_{zx} g + \nabla_{xz} g \tilde{Y} \nabla f'$$

$$+ \nabla f \tilde{Y} \nabla_{zz} g \tilde{Y} \nabla f' + c \sum_{i \in I(x^*)} \nabla_i g \nabla f_i \nabla f_i', \tag{47}$$

where

$$\tilde{Y} = \begin{bmatrix} \tilde{y}_1 & & 0 \\ & \ddots & \\ 0 & & \tilde{y}_k \end{bmatrix}, \qquad k \times k \text{ matrix}, \tag{48}$$

and all the derivatives are calculated at the current point (x, y). Using (47) and (23), we have

$$\nabla_{xx}F(x^*, y^*, c) = \Xi(x^*, y^*) + c \sum_{i \in I(x^*)} \nabla_i g \nabla f_i \nabla f_i' |_{(x^*, y^*)}. \tag{49}$$

By hypothesis, (34) holds. By Theorem 2.10 of Ref. 8, there exists $c^* \geq 0$, such that, for all $c \geq c^*$, the matrix $\nabla_{xx}F(x^*, y^*, c)$ is positive definite. By choosing ϵ' sufficiently small, we have that (48) and (49) hold for all $(x, y) \in S((x^*, y^*), \epsilon')$ and $c \in [c^*, \bar{c}]$. The continuity of $\nabla_{xx}F(x^*, y^*, c)$ for $(x, y) \in S((x^*, y^*, c), \epsilon')$ guarantees that there exists an $S((x^*, y^*), \epsilon)$ in which $\nabla_{xx}F(x, y, c)$ is positive definite for all $c \in [c^*, \bar{c}]$. □

The next theorem is the cornerstone of the duality framework that we wish to establish. It is very similar to Proposition 1 of Ref. 1, and its proof is the same if one uses the results and expressions of our problem, instead of the ones used in Ref. 1.

Theorem 3.3. Let x^*, y^*, c^* be as in Theorem 3.2. then, there exist positive scalars ϵ^* and δ^* such that, for all $y \in S(y^*, \delta^*)$ and all $c \in [c^*, \bar{c}]$, the problem

$$\begin{aligned} &\text{minimize } F(x, y, c), \\ &\text{subject to } x \in S(x^*, \epsilon^*), \end{aligned} \tag{50}$$

has a unique solution $x(y, c)$. Furthermore, for every ϵ with $0 < \epsilon \leq \epsilon^*$, there exists a δ with $0 < \delta \leq \delta^*$, such that

$$x(y, c) \in S(x^*, \epsilon), \qquad \text{for all } y \in S(y^*, \delta), c \in [c^*, \bar{c}].$$

The following corollary is an easy consequence of Theorem 3.3 (see Corollary 1.1 in Ref. 1).

Corollary 3.1. Let M be such that

$$\|f(x) - f(y)\| \leq M\|x - y\|, \qquad \text{for all } x, y \in \bar{S}(x^*, \epsilon^*), \tag{51}$$

and let ϵ^*, δ^* be as in Theorem 3.3. Then, for every ϵ with $0 < \epsilon \leq \epsilon^*$, there exists a δ with $0 < \delta \leq \delta^*$, such that

$$x(y, c) \in S(x^*, \epsilon), \qquad \tilde{y}(x(y, c)) \in S(y^*, \delta + \bar{c}M\epsilon), \tag{52}$$

for all

$$y \in S(y^*, \delta), \qquad c \in [c^*, \bar{c}],$$

where

$$\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_k)'.$$

Proof. The proof proceeds as the one of Corollary 1.1 of Ref. 1 if one notes that, for ϵ^*, δ^* sufficiently small, (45) holds for every $c \in [c^*, \bar{c}]$. □

We are now ready to define the dual functional $q_c(y)$. The function $F(x, y, c)$ has a locally convex structure under the assumptions of Theorem 3.2, and thus the dual character of $q_c(y)$ will be local. The definition is given under the assumption that the hypothesis of Theorems 3.2 and 3.3 holds. We define

$$\begin{aligned} q_c(y) = \min\ & F(x, y, c), \\ &\text{subject to } x \in S(x^*, \epsilon^*), \end{aligned} \tag{53}$$

for all $y \in S(y^*, \delta^*)$ and $c \in [c^*, \bar{c}]$, where the minimum over the open ball $S(x^*, \epsilon^*)$ is attained by Theorem 3.3. The results already obtained concerning $F(x, y, c)$ guarantee that ϵ^* and δ^* can be chosen so that $q_c(y)$ is twice continuously differentiable in $S(y^*, \delta^*)$ for all $c \in [c^*, \bar{c}]$; see Refs. 8, 15 for the corresponding result for the classical nonlinear programming problem. We assume that ϵ^* and δ^* have been so chosen.

We could have defined the dual functional in a different way. For fixed $c \geq c^*$, by using Theorem 3.2 and the implicit function theorem, we obtain that the system of equations

$$\nabla_x F(x, y, c) = 0$$

has an implicit function solution $x(y, c)$, which is a strict local minimum of $F(x, y, c)$. We could set

$$q_c(y) = F(x(y, c), y, c).$$

However, the domain of definition of q_c will then depend on c. On the other hand, in (53) the domain of definition of q_c is the same for all $c \in [c^*, \bar{c}]$. This is better suited to our purposes, since we intend to vary c in our algorithms. The restriction $c \leq \bar{c}$ does not lead to a great loss of generality especially for practical purposes.

We will now calculate $\nabla q_c(y)$ and $\nabla^2 q_c(y)$. From Theorem 3.3 and (53), we have

$$q_c(y) = F(x(y, c), y, c),$$

from which

$$\nabla q_c(y) = \nabla_y x(y, c) \nabla_x F(x(y, c), y, c) + \nabla_y F(x(y, c), y, c). \tag{54}$$

Theorem 3.3 yields

$$\nabla_x F(x(y, c), y, c) \equiv 0, \tag{55}$$

and thus

$$\nabla q_c(y) = \nabla_y F(x(y, c), y, c).$$

Calculating $\nabla_y F(x(y, c), y, c)$, we obtain

$$\nabla q_c(y) = (1/c) \begin{bmatrix} \nabla_1 g & & 0 \\ & \ddots & \\ 0 & & \nabla_k g \end{bmatrix} (\tilde{y}(x(y, c)) - y), \tag{56}$$

which can also be written as

$$\nabla q_c(y) = \begin{bmatrix} \nabla_1 g \nabla_{y_1} p_1 \\ \vdots \\ \nabla_k g \nabla_{y_k} p_k \end{bmatrix} = G \nabla_y p, \tag{57}$$

where

$$\nabla_{y_i} p_i = \begin{cases} (1-y_i)/c, & \text{if } f_i(x(y,c)) \geq (1-y_i)/c, \\ f_i(x(y,c)), & \text{if } -y_i/c \leq f_i(x(y,c)) \leq (1-y_i)/c, \\ -y_i/c, & \text{if } f_i(x(y,c)) \leq -y_i/c. \end{cases} \tag{58}$$

Let us assume that ϵ^* and δ^* have been chosen sufficiently small, so that, for all $i = 1, \ldots, k$ and all $y \in S(y^*, \delta^*)$ and $c \in [c^*, \bar{c}]$,

$$\begin{aligned} f_i(x(y,c)) &> (1-y_i)/c, && \text{if } f_i(x^*) > 0, \\ f_i(x(y,c)) &< -y_i/c, && \text{if } f_i(x^*) < 0, \\ -y_i/c < f_i(x(y,c)) &< (1-y_i)/c, && \text{if } i \in I(x^*), \end{aligned}$$

and thus (45) holds. Consequently, $\nabla^2 q_c(y)$ exists for all $y \in S(y^*, \delta^*)$, $c \in [c^*, \bar{c}]$, if ϵ^* and δ^* are sufficiently small. Let

$$\begin{aligned} &I(x^*) = \{1, \ldots, p\}, \\ &f_i(x^*) > 0, \qquad \text{for } i = p+1, \ldots, q, \\ &f_i(x^*) < 0, \qquad \text{for } i = q+1, \ldots, k. \end{aligned}$$

Differentiating (56) w.r.t. y, we get

$$\begin{aligned} \nabla^2 q_c(y) = {} & \nabla_y x(y,c)\{\nabla_{xz} g(1/c)(\tilde{Y} - Y) + \nabla f \tilde{Y} \nabla_{zz} g(1/c)(\tilde{Y} - Y) + \nabla f \tilde{G}\} \\ & + (1/c)(\tilde{Y} - Y)\nabla_{zz} g(1/c)(\tilde{Y} - Y) + (1/c)(\tilde{G} - G), \end{aligned} \tag{59}$$

where $\tilde{Y} = \tilde{Y}(x(y,c))$,

$$G = \begin{bmatrix} \nabla_1 g & & 0 \\ & \ddots & \\ 0 & & \nabla_k g \end{bmatrix}, \qquad k \times k \text{ matrix}, \tag{60}$$

$$\tilde{G} = \begin{bmatrix} \nabla_1 g & & & & & \\ & \ddots & & & 0 & \\ & & \nabla_p g & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{bmatrix}, \qquad k \times k \text{ matrix}, \tag{61}$$

and all the quantities are calculated at $(x(y,c), y)$. Differentiating (55) w.r.t. y (total derivative), we obtain

$$\nabla_y x(y,c) = -\nabla_{yx} F[\nabla_{xx} F]^{-1}, \tag{62}$$

since by Theorem 3.2 $[\nabla_{xx}F]^{-1}$ exists. By direct calculation, we also have

$$\nabla_{yx}F = (1/c)(\tilde{Y} - Y)\nabla_{zx}g + \tilde{G}\nabla f' + (1/c)(\tilde{Y} - Y)\nabla_{zz}g\tilde{Y}\nabla f'. \tag{63}$$

Let

$$L = [\nabla_{xx}F]^{-1}.$$

Substituting (63) in (62) and (62) in (59), we have finally

$$\begin{aligned}\nabla^2 q_c(y) = &-\tilde{G}\nabla f' L \nabla f \tilde{G} - (1/c)(G - \tilde{G})\\ &+ (1/c)(\tilde{Y} - Y)\{\nabla_{zz}g - \nabla_{zx}gL\nabla_{xz}g - \nabla_{zz}g\tilde{Y}\nabla f' L \nabla f \tilde{Y}\nabla_{zz}g\\ &-\nabla_{zx}gL\nabla f\tilde{Y}\nabla_{zz}g + \nabla_{zz}\tilde{Y}\nabla f' L\nabla_{xz}g\}(1/c)(\tilde{Y} - Y)\\ &-(1/c)(\tilde{Y} - Y)\{\nabla_{zx}gL\nabla f + \nabla_{zz}g\tilde{Y}\nabla f' L \nabla f\}\tilde{G}\\ &-\tilde{G}\{\nabla f' L\nabla_{xz}g + \nabla f' L \nabla f \tilde{Y}\nabla_{zz}g\}(1/c)(\tilde{Y} - Y).\end{aligned} \tag{64}$$

Using (45), we can write

$$(1/c)(\tilde{Y} - Y) = \begin{bmatrix} f_1(x) & & & & & & & \\ & \ddots & & & & & 0 & \\ & & f_p(x) & & & & & \\ & & & (1-y_{p+1})/c & & & & \\ & & & & \ddots & & & \\ & & & & & (1-y_q)/c & & \\ & 0 & & & & & -y_{q+1}/c & \\ & & & & & & & \ddots \\ & & & & & & & -y_k/c \end{bmatrix}_{x = x(y,c)}. \tag{65}$$

So, for x and y sufficiently close to x^* and y^*, respectively, (64) takes the form

$$\nabla^2 q_c(y) = -\tilde{G}\nabla f' L \nabla f \tilde{G} - (1/c)(G - \tilde{G}) + O(\|y - y^*\|). \tag{66}$$

An immediate consequence of (66) is that $q_c(y)$ has a locally concave structure around y^*, on the subspace

$$V^\perp = \mathcal{R}(\tilde{G}\nabla f'|_{x^*}).$$

The characterization of $q_c(y)$ as a dual functional of $F(x, y, c)$ under the assumptions stated is now justified. Since, for x near x^*, only the active f_i's are of importance, we can consider that all the f_i's are active, or equivalently restrict our attention on $M^\perp$, where $q_c(y)$ is strictly concave. The point y^* is a strict local maximum of $q_c(y)$ on $M^\perp$, and

$$q_c(y^*) = F(x(y^*, c), y^*, c),$$

since

$$x(y^*, c) = x^*.$$

Therefore, we define the problem

$$\begin{aligned} &\text{maximize } q_c(y), \\ &\text{subject to } y \in V^\perp, \end{aligned} \tag{67}$$

to be the *locally-dual problem* of the problem

$$\begin{aligned} &\text{minimize } F(x, y, c), \\ &\text{subject to } x \in R^n. \end{aligned} \tag{68}$$

In the next section, we introduce the algorithms for solving Problem (NPS), interpreting them as gradient methods for solving the dual problem.

4. Algorithms, Convergence Results

4.1. Algorithms. In this section, we describe two algorithms, denoted as Algorithms (AS) and (AN). Algorithm (AS) was originally introduced in Ref. 4 for Problem (NPS) and extended further in Ref. 5. In the next section, we will extend Algorithm (AS) for Problem (NP). Algorithm (AN) is introduced here for the first time. The description of Algorithm (AN) for (NPS) is given in this section, and for (NP) in the next section. They both take the following general (and imprecise) form.

Step 0. Choose a vector $y = y^\circ = (y_1^\circ, \ldots, y_k^\circ)' \in R^k$ and a scalar $c = c^\circ > 0$.

Step 1. Find a perhaps local minimum $x^s = x(y^s, c^s)$ of the problem

$$\begin{aligned} &\text{minimize } F(x, y^s, c^s), \\ &\text{subject to } x \in R^n. \end{aligned} \tag{69}$$

Step 2. Update y^s and c^s in a certain way and get y^{s+1}, c^{s+1}, with $c^{s+1} \geq c^s$. Set $s = s + 1$, and go to Step 1.

The cost function of Problem (69) is differentiable w.r.t. x; thus, any of the known suitable techniques can be used. The difference between Algorithms (AS) and (AN) lies in Step 2. Before completing the description of Algorithms (AS) and (AN), we make this remark. If any update for y^s is used in Step 3, with $y^s \in \Lambda$ for every $s = 0, 1, \ldots$, where Λ is a bounded subset of R^k, we have that Problem (69) is an approximate version of Problem (NPS) and corresponds to an iteration of the penalty method for the classical nonlinear programming problem. So, one can expect some kind of convergence of $\{x^s\}_{s=1}^\infty$ to an optimum of (NPS), under certain assumptions.

By updating y in an intelligent way, our algorithm will enjoy the advantages of the multiplier methods over the penalty method.

The update that Algorithm (AS) uses is the following.

Step 2. Algorithm (AS):

$$y_i^{s+1} = \begin{cases} 1, & \text{if } (1-y_i^s)/c^s \le f_i(x^s), \\ y_i^s + c^s f_i(x^s), & \text{if } -y_i^s/c^s \le f_i(x^s) \le (1-y_i^s)/c^s, \\ 0, & \text{if } f_i(x^s) \le -y_i^s/c^s, \end{cases} \tag{70}$$

which is a steepest ascent-type iteration for solving the dual problem.

The update that Algorithm (AN) uses is the following.

Step 2. Algorithm (AN):

$$y^{s+1} = y^s - [H^s]^{-1} G^* \nabla_y p, \tag{71}$$

where $\nabla_y p$ is given in (58),

$$H^s = -\tilde{G}^* \nabla f' L \nabla f \tilde{G}^* - (1/c)(G^* - \tilde{G}^*)|_{x^s, y^s, c^s}, \tag{72}$$

$$G^* = (a_{ij}^*), \qquad \tilde{G}^* = (b_{ij}^*), \qquad k \times k \text{ diagonal matrices,}$$

$$a_{ii}^* = \begin{cases} \nabla_i g, & \text{if } \nabla_i g \ne 0, \\ 1, & \text{if } \nabla_i g = 0, \end{cases}$$

$$b_{ii}^* = \begin{cases} \nabla_i g, & \text{if } -y_i/c < f_i(x) < (1-y_i)/c \text{ and } \nabla_i g \ne 0, \\ 1, & \text{if } -y_i/c < f_i(x) < (1-y_i)/c \text{ and } \nabla_i g = 0, \\ 0, & \text{otherwise,} \end{cases}$$

and everything is calculated at x^s, y^s, c^s. If H^s as given in (72) is not invertable, set $H^s = I$. This iteration is a quasi-Newton iteration for solving the dual problem. We call it quasi-Newton iteration, because H^s corresponds to an appropriate invertible approximation of the Hessian. If the assumptions of Theorem 3.3 hold and

$$I(x^*) = \{1, 2, \ldots, p\},$$

then for x, y sufficiently close to x^*, y^*, respectively, the iteration (71) becomes

$$\begin{aligned} & y_2^{s+1} = 1, \qquad \text{if } (1-y_i^s)/c^s \le f_i(x^s),\ i \notin I(x^*), \\ & y_2^{s+1} = 0, \qquad \text{if } f_i(x^s) \le -y_i^s/c^s,\ i \notin I(x^*), \\ & \begin{bmatrix} y_1^{s+1} \\ \vdots \\ y_p^{s+1} \end{bmatrix} = \begin{bmatrix} y_1^s \\ \vdots \\ y_1^s \end{bmatrix} - \begin{bmatrix} \nabla_1 g & & 0 \\ & \ddots & \\ 0 & & \nabla_p g \end{bmatrix}^{-1} A^{-1} \begin{bmatrix} f_1(x^s) \\ \vdots \\ f_p(x^s) \end{bmatrix}, \end{aligned} \tag{73}$$

where A is the upper left $p \times p$ minor of $\nabla f' L \nabla f$.

4.2. Convergence Results. Here, we deal with convergence results for the algorithms. Theorem 4.1 is similar to Proposition 2 of Ref. 1. The functions f and g are assumed to be in C^2 throughout this section.

Theorem 4.1. Assume that the hypothesis of Theorem 3.3 holds. Let

$$I(x^*)=\{1,\dots,p\}.$$

Assume that the $p\times p$ matrix $D(x, y)$ defined as

$$D_0=\left[\begin{array}{c|c} D(x,y) & 0 \\ \hline 0 & 0\end{array}\right]=-\tilde{G}\nabla f'[\Xi(x,y)]^{-1}\nabla f\tilde{I}, \tag{74}$$

where

$$\tilde{I}=\left[\begin{array}{cc} I_{p\times p} & 0 \\ \hline 0 & 0\end{array}\right], \qquad k\times k \text{ matrix},$$

is defined and invertible in a set $S(x^*, \epsilon)\times S(y^*, \delta+\bar{c}M\epsilon)$, where ϵ and δ are positive scalars, such that

$$x(y, c)\in S(x^*, \epsilon), \qquad \tilde{y}(x(y, c))\in S(y^*, \delta+\bar{c}M\epsilon),$$

for all $y\in S(y^*, \delta)$ and all $c\in[c^*, \bar{c}]$, in accordance with Theorem 3.3 and Corollary 2.1, and ϵ^*, δ^* are assumed to be sufficiently small, so that (45) holds. Assume also that Algorithm (AS) yields a sequence $\{(x^s, y^s)\}_{s=1}^{\infty}$ converging to (x^*, y^*) and that, after some index $\bar{s}$, the (x^s, y^s) are contained in $S(x^*, \epsilon)\times S(y^*, \delta)$. Then, we have

$$\|y^{s+1}-y^*\|\le r_s\|y^s-y^*\|, \qquad \text{for all } s\ge\bar{s}, \tag{75}$$

where

$$r_s=r_s^*+2K(\delta+\bar{c}M\epsilon),$$

$$r_s^*=\max_{(x,y)\in\bar{S}(x^*,\epsilon)\times\bar{S}(y^*,\delta+\bar{c}M\epsilon)}|1/\{1-c^s e_i[D(x,y)]\}|, \qquad i=1,\dots,k. \tag{76}$$

K is some constant with $K\ge 0$, and $e_i[D(x, y)]$ denotes the ith eigenvalue of $D(x, y)$.

Proof. Since, for ϵ and δ sufficiently small, (45) holds, the y_i's for which $i\notin I(x^*)$ will converge in a finite number of iterations. So w.l.o.g. we assume that all the f_i's are active at x^*. Then,

$$D(x, y)=-G\nabla f'[\Xi(x, y)]^{-1}\nabla f,$$

and

$$\tilde{y}(x(y^s, c^s))=\tilde{y}(x^s)=y^{s+1}=y^s+c^s f(x^s).$$

We have

$$\|y^{s+1}-y^*\|=\|y^s-y^*+c^s f(x(y^s,c^s))\|.$$

Since

$$x(y^*,c^s)=x^*, \qquad f(x^*)=0,$$

we have

$$\begin{aligned}\|y^{s+1}-y^*\|&=\|y^s-y^*+c^s[f(x(y^s,c^s))-f(x(y^*,c^s))]\|\\&=\|y^s-y^*+c^s\int_0^1 \nabla_y f(x(y,c^s))'(y^s-y^*)\,dt\|\\&=\|\int_0^1 \{I+c^s\nabla_y f(x(y,c^s))\}'(y^s-y^*)\,dt\|,\end{aligned} \tag{77}$$

where

$$y=y^*+t(y^s-y^*), \qquad 0\le t\le 1.$$

Using (62) and (63), we obtain

$$I+c^s\nabla_y f(x(y,c^s))=I-c^sG\nabla f'L\nabla f-(\tilde{Y}-Y)[\nabla_{zx}g+\nabla_{zz}g\tilde{Y}\nabla f']L\nabla f, \tag{78}$$

where

$$L=[\Xi+c^s\nabla fG\nabla f']^{-1},$$

and all the quantities are calculated at $x(y,c^s)$, y. A well-known matrix identity implies that

$$I-c^sG\nabla f'[\Xi+c^s\nabla fG\nabla f']^{-1}\nabla f=[I+c^sG\nabla f'\Xi^{-1}\nabla f]^{-1}. \tag{79}$$

Notice that, since all f_i's are active at x^* and the second-order sufficiency conditions hold, the matrix G will have positive entries for ϵ and δ sufficiently small. Equations (77), (78), (79) yield

$$I+c^s\nabla_y f(x(y,c^s))=[I-c^sD]^{-1}-(\tilde{Y}-Y)F, \tag{80}$$

where

$$F=F(x(y,c^s),y)=[\nabla_{zx}g+\nabla_{zz}g\tilde{Y}\nabla f']L\nabla f.$$

Since $F\to\bar{F}$, $\bar{F}$ a constant matrix and $\tilde{Y}-Y\to 0$, as $s\to+\infty$, the result follows. □

We now obtain the following local convergence result (see also Corollary 2 in Ref. 1).

Corollary 4.1. Assume that the hypothesis of Theorem 3.3 holds and that the matrix $D(x, y)$ and ϵ, δ are as in Theorem 4.1. Assume also that ϵ and δ are sufficiently small and c^s is sufficiently large, so that, for some constant μ,

$$c^s \geq \mu > \max\{0, 2/e_i[D(x, y)]\}, \qquad \text{for all } s > 0,$$

for all eigenvalues $e_i[D(x, y)]$ of $D(x, \lambda)$ over $\bar{S}(x^*, \epsilon) \times \bar{S}(y^*, \delta + \bar{c}M\epsilon)$ and that $y_0 \in S(y, \bar{\delta})$, where $\bar{\delta}$ is sufficiently small, so as to have $\bar{\delta} + \bar{c}M\epsilon < \delta$, for ϵ sufficiently small. Then, the sequence $\{y^s\}_{s=1}^{\infty}$ generated by the iteration (70) remains in $S(y^*, \delta)$ and converges to y^* for ϵ and δ sufficiently small.

Proof. The proof goes as the proof of Corollary 2.1 in Ref. 1. Notice that, if

$$I(x^*) = \{1, \ldots, p\}$$

and $y_{\bar{i}}^{\circ} \neq 0$ or 1, for some $\bar{i} \notin I(x^*)$, then $y_{\bar{i}}^{1} = 0$ or 1 if

$$f_{\bar{i}}(x^*) < 0 \qquad \text{or} \qquad f_{\bar{i}}(x^*) > 1,$$

respectively, and $y^1 \in S(y^*, \delta)$. So, we can assume w.l.o.g. that $y^{\circ} \in S(y^*, \delta)$ and $y_{\bar{i}}^{\circ} = 0$ or 1 if

$$f_{\bar{i}}(x^*) < 0 \qquad \text{or} \qquad f_{\bar{i}}(x^*) > 0,$$

respectively, and forget about $\bar{\delta}$.

We first prove that, for some constant ρ^* with $0 \leq \rho^* < 1$, it holds that

$$r_s^* \leq \rho^* < 1, \qquad \text{for } s \geq 0.$$

By taking ϵ and δ sufficiently small, we make the quantity $2K(\delta + \bar{c}M\epsilon)$ arbitrarily small, so that, for some constant ρ and for all $s \geq 0$,

$$0 \leq r_s \leq \rho < 1. \qquad \square$$

The next theorem is similar to Proposition 1 of Ref. 2.

Theorem 4.2. Assume that x^*, y^* satisfy the first-order necessary and the second-order sufficiency conditions for x^* to be a strict local minimum of (NPS) and that, in a neighbourhood $S(x^*, \epsilon)$ of x^*, $f_1, \ldots, f_k$ are twice continuously differentiable and $\nabla^2 f_1, \ldots, \nabla^2 f_k$ are Lipschitz continuous. Assume also that, in a neighbourhood $S((x^*, z^*), \epsilon)$ of (x^*, z^*), where

$$z^* = (\gamma_1(x^*), \ldots, \gamma_k(x^*))' \in R^k,$$

g is twice continuously differentiable and $\nabla^2 g$ is Lipschitz continuous. Let B be any bounded subset of R^k. Then, there exists a scalar $c^* \geq 0$ depending on B, such that, for every $c > c^*$ and $y \in B$, Problem (68) has a unique minimizing point $x(y, c)$ with respect to x within some open ball centered at x^*. Furthermore, for some scalar $N \geq 0$, we have

$$\|x(y, c) - x^*\| \leq N\|y - y^*\|/c, \qquad \text{for all } c > c^* \text{ and } y \in B, \tag{81}$$

$$\|\tilde{y} - y^*\| \leq N\|y - y^*\|/c, \qquad \text{for all } c > c^* \text{ and } y \in B, \tag{82}$$

where the vector

$$\tilde{y} = \tilde{y}(x(y, c) \in R^k$$

has components $\tilde{y}_i(x(y, c))$ as in (45).

Proof. The proof is similar to the proof of Proposition 1 in Ref. 2. For $x \in S(x^*, \epsilon)$ and any fixed $y \in B$, $c > 0$, we consider the auxiliary variables

$$p = x - x^*, \qquad q = \tilde{y}(x) - y^*, \qquad s = \begin{bmatrix} p \\ q \end{bmatrix},$$

where $\tilde{y}(x)$ is as in (45). Using (39) and the fact that $f_i \in C^1$ for $\|p\| < \epsilon$, we obtain

$$\begin{aligned} |p_c(f_i(x), y_i) - z_i^{*8}| &\leq p_c(f_i(x), y_i) - \gamma(f_i(x))| + |\gamma(f_i(x)) - \gamma(f_i(x^*))| \\ &\leq K_0/c + |f_i(x) - f_i(x^*) \leq K_0/c + L_i\|p\|, \end{aligned} \tag{83}$$

where K_0, $L_i \geq 0$ are constants which depend on B and ϵ, respectively. The relation (83) holds for $\|p\| < \epsilon$. From (83), we obtain that, for $\|p\| < \epsilon_1$ and $c > \tilde{c}$, for some appropriately chosen ϵ_1, $\tilde{c}$ with $0 < \epsilon_1 \leq \epsilon$, $\tilde{c} \geq 0$, we will be working in the domain where the differentiability and Lipschitz assumptions hold. In the rest of the proof, we assume that

$$\|p\| < \epsilon_1 \qquad \text{and} \qquad c \geq \tilde{c}.$$

We consider now

$$q = \tilde{y}(x) - y^*.$$

If $i \notin I(x^*)$, then, for $\|p\|$ sufficiently small and c sufficiently large, it holds that $q_i = 0$ for every $y \in B$. Consequently, we consider ϵ_1 and $\tilde{c}$ sufficiently small and large, respectively, and thus we assume w.l.o.g. that

$$I(x^*) = \{1, \ldots, k\}.$$

If x is a local minimum of $F(x, y, c)$, it will hold that

$$\nabla_x F(x, y, c) = \left(\nabla_x g + \sum_{i=1}^{k} \tilde{y}_i(x) \nabla_i g \nabla f_i\right)\Big|_{x,y,c} = 0.$$

It also holds that

$$\left(\nabla_x g + \sum_{i=1}^{k} y_i^* \nabla_i g_i f_i\right)\Big|_{x^*,y^*,c} = 0.$$

Since, for $\|p\| \to 0$, we have that

$$\nabla_x g|_{x,y,c} \to \nabla_x g|_{x^*,y^*,c}, \qquad \nabla_i g|_{x,y,c} \to \nabla_i g|_{x^*,y^*,c}, \qquad \nabla f_i(x) \to \nabla f_i(x^*),$$

we can prove, by using Proposition 1.2 and the lemma used there of Ref. 23, that

$$\tilde{y}(x)\,|_{x,y,c} \to y^*, \qquad \text{as } \|p\| \to 0.$$

It is clear that $\tilde{y}(x) \to y^*$, as $\|p\| \to 0$, uniformly in c, i.e., for all $c \geq c_1$. If we vary c, say $c \to \infty$, then we facilitate the convergence of $\tilde{y}(x)$ to y^*. We conclude that, for ϵ_1 sufficiently small, q will be given by

$$q = y + cf(x) - y^*,$$

from which

$$\nabla f(x^*)'p - q/c = (y^* - y)/c - r_5(p), \tag{84}$$

where

$$f(x) = f(x^*) + \nabla f(x^*)'p + r_5(p),$$

$$r_5(0) = 0, \tag{85}$$

$$\|\nabla r_5(p)\| \leq K_5 \|p\|, \tag{86}$$

and $K_5 \geq 0$ is a constant depending on ϵ.

For every $y \in B$, there holds that

$$\nabla_x g|_{x,y,c} = \nabla_x g|_{x^*,y^*,c} + (\nabla_{xx} g + \nabla_{xz} g Y^* \nabla f')|_{x^*,y^*,c} p + r_1(s). \tag{87}$$

Here, $Z\,|_{x,y,c}$ means that A is calculated at x, y, c. For example,

$$\nabla_x g\,|_{x,y,c} = \nabla_x g\,|_{x,y,c} = \nabla_x g[x, p_c(f_1(x), y_1), \ldots, p_c(f_k(x), y_k)].$$

We will also denote in this proof $A\,|_{x^*,y^*,c}$ by A^*; and we will denote $A\,|_{x,y,c}$ by A.

$$r_1(0) = 0. \tag{88}$$

Using the Lipschitz assumption, we can show that

$$\|\nabla r_1(s)\| \leq K_1 \|s\|, \tag{89}$$

where K_1 is constant which depends on ϵ and B.

Similarly, we have, for $i = 1, \ldots, k$,

$$\nabla_i g\,|_{x,y,c} = \nabla_i g\,|_{x^*,y^*,c} + (\nabla_{ix} g + \nabla_{iz} g Y^* \nabla f')\,|_{x^*,y^*,c} p + r_2^i(s), \tag{90}$$

$$r_2^i(0) = 0, \tag{91}$$

$$\|\nabla r_2^i(s)\| \le K_{2i} \|s\|, \tag{92}$$

$$\nabla f_i(x) = \nabla f_i(x^*) + \nabla^2 f_i(x^*) p + r_3^i(p), \tag{93}$$

$$r_3^i(0) = 0, \tag{94}$$

$$\|\nabla r_3^i(p)\| \le K_{3i} \|p\|, \tag{95}$$

where $K_{2i}, K_{3i} \ge 0$ are constants which depend on B and ϵ.

We consider now $F(x, y, c)$. Using (44) and (87)–(95), we obtain

$$\begin{aligned}\nabla_x F(x, y, c) &= \nabla_x g\Big|_{x,y,c} + \sum_{i=1}^{k} \tilde{y}_i(x) \nabla_i g\Big|_{x,y,c} \nabla f_i(x)\\ &= \nabla_x g + \nabla_{xx} g p + \nabla_{xz} g Y^* \nabla f' p + r_1 + \sum_{i=1}^{k} (q_i + y_i^*)[\nabla_i g + \nabla_{ix} g p\\ &\quad + \nabla_{iz} g Y^* \nabla f' p + r_2^i][\nabla f_i + \nabla^2 f_i p + r_3^i];\end{aligned}$$

all derivatives in the last expression are calculated at x^*, y^*, c; equivalently,

$$\nabla_x F(x, y, c) = \Xi(x^*, y^*) p + \nabla f G q + r_4(s), \tag{96}$$

where

$$r_4(0) = 0, \tag{97}$$

$$\|\nabla r_4(s)\| \le K_4 \|s\|; \tag{98}$$

here, $K_4 \ge 0$ is a constant depending on B and ϵ.

Combining now (96), (97), (98), (84)–(86), we have that, in order for a point $x \in S(x^*, \epsilon_1)$ to satisfy

$$\nabla_x F(x, y, c) = 0,$$

it is necessary and sufficient that the corresponding point s satisfies the equation

$$Bs = a + r(s), \tag{99}$$

where

$$B = \begin{bmatrix} \Xi & \nabla f G \\ \nabla f' & -I/c \end{bmatrix}, \qquad a = \begin{bmatrix} 0 \\ (y^* - y)/c \end{bmatrix}, \qquad r(s) = \begin{bmatrix} -r_4(s) \\ -r_5(s) \end{bmatrix}, \tag{100}$$

and I is the $k \times k$ identity matrix. Concerning $r(s)$, we have

$$r(0)=0, \qquad \|\nabla r(s)\| \leq K\|s\|, \tag{101}$$

where $K \geq 0$ is a constant depending on B and ϵ.

If Ξ is not positive definite, we consider the problem

$$\text{minimize } g[x, \gamma[f_1(x)], \ldots, \gamma[f_k(x)]] + (c/2)\|f(x)\|^2. \tag{102}$$

It is easy to prove that x^*, y^* satisfy the first-order necessary and second-order sufficiency conditions for Problem (NP) iff they satisfy the first-order necessary and second-order sufficiency conditions for the Problem (102), assuming in both cases that

$$I(x^*) = \{1, \ldots, k\}.$$

To (102) corresponds a Ξ_c, which equals $\Xi(x^*, y^*) + c\nabla f(x^*)\nabla f(x^*)'$, and is positive definite for c sufficiently large (see Theorem 2.10 in Ref. 8). The proof now follows exactly as the one of Proposition 1 in Ref. 2, by using the following two lemmas. □

Lemma 4.1 is a modified version of Lemma 1 of Ref. 20 and can be proved by trivial modification of this lemma. Lemma 4.2 extends Lemma 2 of Ref. 20 and can be proved again by a trivial modification of the proof in Ref. 20.

Lemma 4.1. Consider the $(n+k)\times(n+k)$ matrix

$$B = \begin{bmatrix} \Xi & SG \\ S' & -I/c \end{bmatrix},$$

where Ξ is an $n \times n$ positive-definite matrix, S is an $n \times k$ matrix, with rank $M = k$, G is a $k \times k$ positive-definite, symmetric matrix, and I is the $k \times k$ identity matrix. Then, B is invertible for every $c > 0$ and B^{-1} is uniformly bounded for all $c > 0$; i.e., for some $c_1 > 0$ and all $c > 0$,

$$\|B^{-1}\| \leq c_1.$$

Lemma 4.2. Let E be a Hilbert space. Let B be a linear operator from E into E possessing an inverse and

$$\|B^{-1}\| \leq c_1.$$

Let $r(s)$ be an operator from E into E, such that

$$r(0) = 0 \qquad \text{and} \qquad \|\nabla r(s)\| \leq K\|s\|,$$

where $K \geq 0$ is a constant. Then, there exists a $c^* \geq 0$, such that, for all

$$c > c^* \qquad \text{and} \qquad \|a\| \leq 1/8c_1^2K,$$

the equation

$$Bs = a + r(s)$$

has in the sphere

$$\|s\| < 4c_1\|a\|$$

a unique solution s^*, where

$$\|s^*\| \leq (c_1/2)\|a\|.$$

We now consider Algorithm (AN), where the update (71) is employed.

Theorem 4.3. Assume that the hypothesis of Theorem 3.3 holds and that, for some scalars $\epsilon > 0$, $R > 0$, we have

$$\|\nabla^2 q_c(y) - \nabla^2 q_c(z)\| \leq R\|y - z\|, \qquad \text{for all } y, z \in S(y^*, \epsilon),$$
$$\text{for all } c \in [c^*, c].$$

Then, there exists a scalar $\epsilon' > 0$, such that, if $\|y^\circ - y^*\| < \epsilon'$, then $y^s \to y^*$ and $\{\|y^s - y^*\|\}_{s=1}^\infty$ converges to zero quadratically.

Proof. For ϵ' sufficiently small, y will be updated by (73). So, we assume w.l.o.g. that

$$I(x^*) = \{1, \ldots, k\}.$$

Then, (73) can be written as

$$y^{s+1} = y^s - D^{-1}\nabla q_{c^s}(y^s),$$

where

$$D = -G\nabla f' L \nabla f G|_{x^s, y^s, c^s}$$

[recall (57)]. Since

$$\nabla^2 q - D = O(\|y^s - y^*\|), \qquad \nabla q_{c^s}(y^*) = 0,$$

by using Taylor's expansion we have

$$\begin{aligned}\|y^{s+1}-y^*\| &= \|y^s - D^{-1}\nabla q_{c^s}(y^s) - y^*\| = \|D^{-1}[D(y^s-y^*)-\nabla q_{c^s}(y^s)]\| \\ &= \|D^{-1}[D(y^s-y^*) - \nabla^2 q_{c^s}(y^s)(y^s-y^*) - \int \{\nabla^2 q_{c^s}(y^s + t(y^*-y^s)) \\ &\quad -\nabla^2 q_{c^s}(y^s)\}(y^s-y^*)\, dt\| \leq \|D^{-1}\|\{\|D-\nabla^2 q_{c^s}(y^s)\|\|y^s-y^*\| \\ &\quad +(R/2)\|y^s-y^*\|^2\} = \|\{\nabla^2 q_{c^s}(y^s) + O(\|y^s-y^*\|)\}^{-1}\| \\ &\quad \cdot\{O(\|y^s-y^*\|^2)+(R/2)\|y^s-y^*\|^2\} \leq \bar{R}\|y^s-y^*\|^2,\end{aligned}$$

where $\bar{R}$ is a constant depending on R and ϵ. □

The following theorem is a rate of convergence result.

Theorem 4.4. Assume that the hypothesis of theorem 3.3 holds and that Algorithm (AN) generates the sequences $\{y^s\}_{s=0}^{\infty}$ and $\{x^s\}_{s=0}^{\infty}$ which converge to y^* and x^*, respectively. Then, we have

$$\lim[\|y^{s+1}-y^*\|/\|y^s-y^*\|]=0,$$

and hence the sequence $\{\|y^s-y^*\|\}_{s=0}^{\infty}$ converges to zero superlinearly.

Proof. The proof is a direct application of Proposition 1.14 of Ref. 23. Notice that the hypothesis of Theorems 4.3 and 4.4 are different. □

5. Results and Algorithms for Problem (NP)

The results that we have proved until now concern Problem (NPS). In this section, we state without proofs several results for Problem (NP), which correspond to results already proved for Problem (NPS). We also give the algorithms for this problem. We use the same sequence of theorems, lemmas, and corollaries concerning (NP) as that used for (NPS). The symbols $f_{ij}, f_i, \gamma, g, h_j, I_i(x)$ now denote the functions and the sets of active f_{ij}'s at x for (NP).

Let $x^* \in R^n$,

$$J(x^*) = I_1(x^*)\times\cdots\times I_k(x^*),$$

$$t=(\mu_1,\ldots,\mu_k)\in J(x^*),$$

$$g_t(x) = g[x, f_{1\mu_1}(x),\ldots,f_{k\mu_k}(x)],$$

$$h_{jt}(x) = h_j[x, f_{1\mu_1}(x),\ldots,f_{k\mu_k}(x)], \qquad j=1,\ldots,q.$$

Consider the problem

(DP-t) minimize $g_t(x)$,

$$\text{subject to } h_{1t}(x) = \cdots = h_{qt}(x) = 0,$$
$$f_{ij}(x) \le f_{i\mu_i}(x), \qquad j = 1, \ldots, m, j \neq \mu_i, i = 1, \ldots, k.$$

Lemma 5.1. A vector $x^* \in R^n$ is a strict local minimum of Problem (NP) iff x^* is a strict local minimum of Problem (DP-t), for every $t \in J(x^*)$. Also, $g^* = g_t(x^*)$ for every $t \in J(x^*)$.

The following notation refers to Theorem 2.2 given below. All gradients are evaluated at the point x^*.

$$\begin{aligned} &\nabla f_i = [\nabla f_{i1} \vdots \cdots \vdots \nabla f_{im}], \qquad i = 1, \ldots, k, \qquad n \times m \text{ matrices}, \\ &\tilde{\nabla} f_1 = [\nabla f_{i_1} \vdots \cdots \vdots \nabla f_{ip_i}], \qquad 1 \le p_i \le m, \qquad i = 1, \ldots, k, \qquad n \times m \text{ matrices}, \end{aligned} \tag{103}$$

$$\begin{aligned} \nabla f &= [\nabla f_1 \vdots \cdots \vdots \nabla f_k], && n \times (k \cdot m) \text{ matrix}, \\ \tilde{\nabla} f &= [\tilde{\nabla} f_1 \vdots \cdots \vdots \tilde{\nabla} f_k], && n \times (p_1 + \cdots + p_k) \text{ matrix}, \\ \nabla_x h &= [\nabla_x h_1 \vdots \cdots \vdots \nabla_x h_q], && n \times q \text{ matrix}, \\ \partial_z h &= [\nabla_z h_1 \vdots \cdots \vdots \nabla_z h_q], && k \times q \text{ matrix}, \end{aligned} \tag{104}$$

$$Y_0 = \begin{bmatrix} \left.\begin{matrix} -1 \\ \vdots \\ -1 \end{matrix}\right\} p_1 & \\ \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & 0 \\ & \ddots \\ 0 & \left.\begin{matrix} -1 \\ \vdots \\ -1 \end{matrix}\right\} p_k \\ & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \end{bmatrix} \left.\vphantom{\begin{matrix} -1 \\ \vdots \\ -1 \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} m, \qquad (k \cdot m) \times k \text{ matrix}, \tag{105}$$

$$\tilde{Y}_0 = \begin{bmatrix} \left.\begin{matrix} -1 \\ \vdots \\ -1 \end{matrix}\right\} p_1 & 0 \\ 0 & \left.\begin{matrix} -1 \\ \vdots \\ -1 \end{matrix}\right\} p_k \end{bmatrix}, \qquad (p_1 + \cdots + p_k) \times k \text{ matrix}, \tag{106}$$

$$A=\left[\begin{array}{c|c}\nabla_x h & \nabla f \\ \hline \nabla_z h & Y_0'\end{array}\right], \qquad (n+k)\times(q+km) \text{ matrix}, \tag{107}$$

$$\tilde{A}=\left[\begin{array}{c|c}\nabla_x h & \tilde{\nabla} f \\ \hline \nabla_z h & \tilde{Y}_0'\end{array}\right], \qquad (n+k)\times(q+p_1+\cdots+p_k) \text{ matrix}. \tag{108}$$

Theorem 5.1. Let x^* be a local minimum of Problem (NP). Assume for convenience that

$$I_i(x^*)=\{1,\ldots,p_i\}, \qquad i=1,\ldots,k,$$

where p_i, $i=1,\ldots,k$, are integers satisfying $1\le p_i\le k$; and assume that $\tilde{A}$ has full rank $q+p_2+\cdots+p_k$. Then, there are real numbers

$$y_{ij}^*, \qquad i=1,\ldots,k, j=1,\ldots,m,$$

$\lambda_1^*,\ldots,\lambda_q^*$, such that

$$\nabla_x g+\sum_{j=1}^{q}\lambda_j^*\nabla_x h_j+\left(\nabla_1 g+\sum_{j=1}^{q}\lambda_j^*\nabla_1 h_j\right)\left(\sum_{j=1}^{m}y_{1j}^*\nabla f_{1j}\right)$$

$$+\cdots+\left(\nabla_k g+\sum_{j=1}^{q}\lambda_j^*\nabla_k h_j\right)\left(\sum_{j=1}^{m}y_{kj}^*\nabla f_{kj}\right)=0, \tag{109}$$

$$0\le y_{ij}^*\le 1, \qquad \text{for all } i=1,\ldots,k, \text{ for all } j=1,\ldots,m,$$

$$y_{ij}^*=0, \qquad \text{if } j\notin I_i(x^*), \text{ for all } i=1,\ldots,k, \tag{110}$$

$$\sum_{j=1}^{m}y_{ij}^*=1, \qquad \text{for all } i=1,\ldots,k.$$

Also, for every $i=1,\ldots,k$, for which $p_i\ge 2$,

$$\nabla_i g+\sum_{j=1}^{q}\lambda_j^*\nabla_i h_j\ge 0. \tag{111}$$

Finally, the scalars

$$\lambda_1^*,\ldots,\lambda_q^*, \qquad \left(\nabla_i g+\sum_{j=1}^{q}\lambda_j^*\nabla_i h_j\right)y_{ij}^*, \qquad i=1,\ldots,k, \qquad j=1,\ldots,m,$$

are unique. All the gradients are evaluated at x^*.

An alternative formulation of the conclusions of Theorem 5.2 is that there is a unique $(q+km)$-vector

$$\begin{bmatrix}\lambda\\ y\end{bmatrix}=(\lambda_1,\ldots,\lambda_q,y_{11},\ldots,y_{1m},\ldots,y_{k1},\ldots,y_{km})',$$

such that

$$\nabla g + A\begin{bmatrix}\lambda\\ y\end{bmatrix} = 0,$$

$$y_{ij} \geq 0, \qquad \text{for all } i=1,\ldots,k, j=1,\ldots,m, \text{ if } p_i \geq 2,$$

$$y_{ij} = 0, \qquad \text{if } j \notin I_i(x^*), i=1,\ldots,k.$$

Notice that

$$\lambda_j = \lambda_j^*, \qquad j=1,\ldots,q,$$

and

$$y_{ij} = \left(\nabla_i g + \sum_{j=1}^{q} \lambda_j \nabla_i h_j\right) y_{ij}^*, \qquad i=1,\ldots,k, j=1,\ldots,m.$$

Definition 5.1. A point $x^* \in R^n$ is said to be a regular point of the f_{ij}'s, $i=1,\ldots,k$, $j=1,\ldots,m$, if the matrix $\tilde{A}$ has full rank $q+p_1+\cdots+p_k$. We assume w.l.o.g. that

$$I_i(x^*) = \{1,\ldots,p_i\}, \qquad i=1,\ldots,k.$$

Let

$$\Pi(x,z) = g(x,z) + \sum_{j=1}^{q} \lambda_j^* h_j(x,z), \qquad x \in R^n, \qquad z \in R^k, \tag{112}$$

$$Y^* = \begin{bmatrix} \begin{matrix} y_{11}^* \\ \vdots \\ y_{1m}^* \end{matrix} & & 0 \\ & \ddots & \\ 0 & & \begin{matrix} y_{k1}^* \\ \vdots \\ y_{km}^* \end{matrix} \end{bmatrix}, \qquad (k \cdot m) \times k \text{ matrix}. \tag{113}$$

The conditions (109) and (111) can be written as

$$\nabla_x \Pi + \nabla f Y^* \nabla_z \Pi|_{(x^*,z^*)} = 0, \tag{114}$$

$$\nabla_i \Pi|_{(x^*,z^*)} \geq 0, \qquad i=1,\ldots,k, \text{ if } p_i \geq 2, \tag{115}$$

where

$$z^* = (\gamma_1(x^*),\ldots,\gamma_k(x^*))'.$$

Theorem 5.2. Assume that x^* satisfies the hypothesis of Theorem 5.1 and that $g, f_1, \ldots, f_k, h_1, \ldots, h_q \in C^2$. Let y_{ij}^*, λ_j^* be as in Theorem 5.1. Then, the matrix

$$\Gamma = \left[\begin{array}{c|c} \nabla_{xx}\Pi + \sum_{i=1}^{k} \nabla_i \Pi \sum_{j=1}^{m} y_{ij}^* \nabla^2 f_{ij} & \nabla_{xz}\Pi \\ \hline \nabla_{zx}\Pi & \nabla_{zz}\Pi \end{array}\right] \tag{116}$$

is positive semidefinite on $\mathcal{N}(\tilde{A}')$. All gradients are evaluated at (x^*, z^*).

Theorem 5.3. Assume that $g, f_1, \ldots, f_k, h_1, \ldots, h_q \in C^2$ and that $x^* \in R^n$, y_{ij}^*, $i = 1, \ldots, k$, $j = 1, \ldots, m$, $\lambda_1^*, \ldots, \lambda_q^*$ satisfy

$$\nabla_x g + \sum_{j=1}^{q} \lambda_j^* \nabla_x h_j + \left(\nabla_1 g + \sum_{j=1}^{q} \lambda_j^* \nabla_1 h_j\right)\left(\sum_{j=1}^{m} y_{1j}^* \nabla f_{1j}\right)$$

$$+ \cdots + \left(\nabla_k g + \sum_{j=1}^{q} \lambda_j^* \nabla_k h_j\right)\left(\sum_{j=1}^{m} y_{kj}^* \nabla f_{kj}\right) = 0,$$

$$y_{ij}^* = 0, \qquad \text{if } j \notin I_i(x^*),\ i = 1, \ldots, k,$$

$$y_{ij}^* > 0, \qquad \text{if } j \in I_i(x^*),\ i = 1, \ldots, k,$$

$$\sum_{j=1}^{m} y_{ij}^* = 1, \qquad \text{for all } i = 1, \ldots, k,$$

$$\nabla_i \pi = \nabla_i g + \sum_{j=1}^{q} \lambda_j^* \nabla_i h_j > 0, \qquad \text{if } p_i \geq 2,\ i = 1, \ldots, k.$$

Γ is positive definite on $\mathcal{N}(\tilde{A}')$. Then, x^* is a strict local minimum of Problem (NP). All gradients are evaluated at (x^*, z^*).

We introduce the function $p_c(\cdot, \lambda)$: $R^m \to R$ (see Ref. 5),

$$p_c(t, \lambda) = \text{infimum}\{\gamma[t-u] + \lambda' u + (c/2)\|u\|^2 \mid u \in R^m\}, \tag{117}$$

where $\lambda \in R^m$, $c > 0$ are fixed. We have (see Ref. 5)

$$p_c(t, y) = (1/2c) \sum_{i=1}^{m} \{[\max[0, y_i + c(t_i - \mu(t, y, c))]]^2 - y_i^2\} + \mu(t, y, c), \tag{118}$$

where

$$\mu = \mu(t, y, c)$$

is a scalar determined uniquely, for given t, y, c, by

$$\sum_{i=1}^{m} \max[0, y_i + c(t_i - \mu(t, y, c))] = 1, \tag{119}$$

$$\nabla_t p_c(t, y) = \begin{bmatrix} \max[0, y_1 + c(t_1 - \mu)] \\ \vdots \\ \max[0, y_m + c(t_m - \mu)] \end{bmatrix}. \tag{120}$$

The function $p_c(\cdot, y)$ is real valued and convex, and a relation similar to (39) holds for $\lambda \in \Lambda \subset R^k$, Λ bounded. Let

$$\begin{aligned} F(x, y, \lambda, c) = g[x, p_c[f_1(x), y_1], \ldots, p_c[f_k(x), y_k]] \\ + \sum_{j=1}^{q} \lambda_j h_j[x, p_c[f_1(x), y_1], \ldots, p_c[f_k(x), y_k]] \\ +(c/2) \sum_{j=1}^{q} (h_j[x, p_c[f_1(x), y_1], \ldots, p_c[f_k(x), y_k]])^2, \end{aligned} \tag{121}$$

where

$$y = (y_1', \ldots, y_k')' \in R^{km}, \qquad y_i \in R^m, \qquad i = 1, \ldots, k,$$
$$\lambda = (\lambda_1, \ldots, \lambda_q)' \in R^q.$$

We have

$$\begin{aligned} \nabla_x F(x, y, \lambda, c) = \nabla_x g + \sum_{j=1}^{q} \tilde{\lambda}_j \nabla_x h_j \\ + \sum_{i=1}^{k} \left[\left(\nabla_i g + \sum_{j=1}^{q} \tilde{\lambda}_j \nabla_i h_j\right) \sum_{j=1}^{m} \tilde{y}_{ij} \nabla f_{ij}\right], \end{aligned} \tag{122}$$

$$\tilde{y}_{ij}(x) = \max[0, y_{ij} + c(f_{ij}(x) - \mu_i(f_i(x), y_i, c))], \qquad i = 1, \ldots, k, \quad j = 1, \ldots, m, \tag{123}$$

$$\tilde{\lambda}_j(x) = \lambda_j + ch_j, \qquad j = 1, \ldots, q,$$

and μ_i satisfies, for $i = 1, \ldots, k$,

$$\sum_{j=1}^{m} \max[0, y_{ij} + c(f_{ij}(x) - \mu_i(f_i(x), y_i, c))] = 1. \tag{124}$$

Theorem 5.4. Assume that $x^* \in R^n$ and y_{ij}^*, $i = 1, \ldots, k$, $j = 1, \ldots, m$, $\lambda_1^*, \ldots, \lambda_q^*$ satisfy the hypothesis of Theorem 1.2. Then, for all $c > 0$,

$$\nabla_x F(x^*, y^*, \lambda^*, c) = 0. \tag{125}$$

Theorem 5.5. Let x^*, y_{ij}^*, $\lambda_1^*, \ldots, \lambda_q^*$ be as in Theorem 5.1, and assume in addition that they satisfy the hypothesis of Theorem 5.3. Then, there exists an $\epsilon_1 = \epsilon_1(c) > 0$, such that, for all $(x, y, \lambda) \in S((x^*, y^*, \lambda^*), \epsilon_1)$, the function $F(x, y, \lambda, c)$ is twice continuously differentiable w.r.t. x. Also, there exists scalars $c^* \geq 0$ and $\epsilon > 0$, such that, for all $c \in [c^*, \bar{c}]$ and $(x, y, \lambda) \in S((x^*, y^*, \lambda^*), \epsilon)$, the Hession $\nabla_{xx}F(x, y, \lambda, c)$ is positive definite, where $\bar{c}$ is an arbitrarily large fixed constant.

We have

$$\begin{aligned}\nabla_{xx}F(x, y, \lambda, c) = {} & \nabla_{xx}\tilde{\Pi} + \sum_{i=1}^{k} \nabla_i\tilde{\Pi} \sum_{j=1}^{m} \tilde{y}_{ij}\nabla^2 f_{ij} + \nabla f\tilde{Y}\nabla_{zx}\tilde{\Pi} + \nabla_{xz}\tilde{\Pi}\tilde{Y}'\nabla f' \\ & + \nabla f\tilde{Y}\nabla_{zz}\tilde{\Pi}\tilde{Y}'\nabla f' + c\{\bar{\tilde{\nabla}}f\tilde{\bar{G}}\bar{\tilde{\nabla}}f' \\ & + (\nabla_x h + \nabla f\tilde{Y}\nabla_z h)(\nabla_z h'\tilde{Y}'\nabla f' + \nabla_x h')\},\end{aligned} \tag{126}$$

where

$$\tilde{\Pi}(x, z) = g(x, z) + \sum_{j=1}^{q} \tilde{\lambda}_j h_j(x, z), \tag{127}$$

$$\tilde{Y} = \begin{bmatrix} \begin{matrix} \tilde{y}_{11} \\ \vdots \\ \tilde{y}_{1m} \end{matrix} & & 0 \\ & \ddots & \\ 0 & & \begin{matrix} \tilde{y}_{k1} \\ \vdots \\ \tilde{y}_{km} \end{matrix} \end{bmatrix}, \qquad km \times k \text{ matrix}, \tag{128}$$

$$\tilde{G} = \begin{bmatrix} \nabla_1\tilde{\Pi} & & & & & & 0 \\ & \ddots & & & & & \\ & & \nabla_1\tilde{\pi} & & & & \\ & & & \ddots & & & \\ & & & & \overbrace{\nabla_k\tilde{\Pi}}^{p_k} & & \\ & & & & & \ddots & \\ 0 & & & & & & \nabla_k\tilde{\Pi} \end{bmatrix}. \tag{129}$$

$\tilde{G}$ contains only those $\nabla_i\tilde{\Pi}$ for which $p_i \geq 2$; and $\bar{\tilde{\nabla}}f$ is like $\tilde{\nabla}f$ in (104), but contains only those ∇f_i's for which $p_i \geq 2$. All the quantities are assumed at x, y, λ, c and $z = (z_1, \ldots, z_k)'$, with

$$z_i = p_c[f_i(x), y_i].$$

Theorem 5.6. Let x^*, y_{ij}^*, $\lambda_1^*, \ldots, \lambda_q^*$, c^* be as in Theorem 5.5. Then, there exist positive scalars ϵ^* and δ^* such that, for all $(y, \lambda) \in S((y^*, \lambda^*), \delta)$ and all $c \in [c^*, \bar{c}]$, the problem

$$\text{minimize } F(x, y, \lambda, c),$$
$$\text{subject to } x \in S(x^*, \epsilon^*),$$

has a unique solution $x(y, \lambda, c)$. Furthermore, for every ϵ with $0 < \epsilon \leq \epsilon^*$, there exists a δ with $0 < \delta \leq \delta^*$, such that

$$x(y, \lambda, c) \in S(x^*, \epsilon), \qquad \text{for all } (y, \lambda) \in S((y^*, \lambda^*), \delta),\ c \in [c^*, \bar{c}].$$

Corollary 5.1. Let M be such that

$$\|h(y) - h(y')\| \leq M\|y - y'\|, \qquad \text{for all } y, y' \in \bar{S}(y^*, \epsilon^*),$$
$$\|f_i(x) - f_i(x')\| \leq M\|x - x'\|, \qquad i = 1, \ldots, k, \qquad \text{for all } x, x' \in \bar{S}(x^*, \epsilon^*),$$

where $y \in R^{n+k}$ and $h = (h_1, \ldots, h_q)'$; and let ϵ^*, δ^* be as in Theorem 5.6. Then, for every ϵ with $0 < \epsilon \leq \epsilon^*$, there exists a δ with $0 < \delta \leq \delta^*$, such that

$$x(y, \lambda, c) \in S(x^*, \epsilon^*),$$
$$(\tilde{y}(x(y, \lambda, c)), \tilde{\lambda}(x(y, \lambda, c))) \in S(((y^*, \lambda^*), \delta + \bar{c}M\epsilon),$$

for all

$$(y, \lambda) \in S((y^*, \lambda^*), \delta), \qquad c \in [c^*, \bar{c}].$$

Under the assumption that the hypotheses of Theorems 5.5 and 5.6 hold, we define the dual functional $q_c(y, \lambda)$ by

$$q_c(y, \lambda) = \min F(x, y, \lambda, c), \quad \text{subject to } x \in S(x^*, \epsilon^*), \tag{130}$$

for all

$$(y, \lambda) \in S((y^*, \lambda^*), \delta^*), \qquad c \in [c^*, \bar{c}].$$

We have

$$\nabla q_c(y, \lambda) = (1/c)G\begin{bmatrix} \tilde{y} \\ \tilde{\lambda} \end{bmatrix} - \begin{bmatrix} y \\ \lambda \end{bmatrix} = G\nabla \cdot p, \tag{131}$$

where

$$G=\left[\begin{array}{ccccccc:c}
\nabla_1\Pi & & & & & & & \\
& \ddots & & & 0 & & & 0\\
& & \nabla_1\Pi & & & & & \\
& & & \ddots & & & & \\
& & & & \nabla_k\Pi & & & \\
& 0 & & & & \ddots & & \\
& & & & & & \nabla_k\Pi & \\
\hdashline
& & 0 & & & & & I_{q\times q}
\end{array}\right], \tag{132}$$

$(km+q)\times(km+q)$ matrix,

and

$$\nabla p=(1/c)\left\{\begin{bmatrix}\tilde{y}\\ \tilde{\lambda}\end{bmatrix}-\begin{bmatrix}y\\ \lambda\end{bmatrix}\right\}=\begin{bmatrix}(\tilde{y}-y)/c\\ h(x)\end{bmatrix}. \tag{133}$$

Also,

$$\nabla^2 q_c(y,\lambda)=-G\Delta' L\Delta G-(1/c)PG+O\left\{\left\|\begin{bmatrix}y\\ \lambda\end{bmatrix}-\begin{bmatrix}y^*\\ \lambda^*\end{bmatrix}\right\|\right\}, \tag{134}$$

where

$$\Delta=[\nabla f^* \,\vdots\, \nabla_x h+\nabla f\tilde{Y}\nabla_z h],\qquad n\times(km+q)\text{ matrix},$$

$$\nabla f^*=\nabla f-\nabla f P_0,$$

$$p_0=\left[\begin{array}{cc}
\begin{array}{c:c}
\begin{array}{ccc}1/p_1 & \cdots & 1/p_1\\ \vdots & & \vdots\\ 1/p_1 & \cdots & 1/p_1\end{array} & 0\\
\hdashline
0 & \begin{array}{ccc}1 & & 0\\ & \ddots & \\ 0 & & 1\end{array}
\end{array} & 0\\
0 & \ddots \\
0 & \begin{array}{c:c}
\begin{array}{ccc}1/p_k & \cdots & 1/p_k\\ \vdots & & \vdots\\ 1/p_k & \cdots & 1/p_k\end{array} & 0\\
\hdashline
0 & \begin{array}{ccc}1 & & 0\\ & \ddots & \\ 0 & & 1\end{array}
\end{array}
\end{array}\right],$$

(each diagonal block is $m\times m$; the $1/p_i$ block is $p_i\times p_i$)

$(k\cdot m)\times(k\cdot m)$ matrix,

$$L=[\nabla_{xx}F(x(y,\lambda,c),y,\lambda,c)]^{-1},$$

$$P = \left[\begin{array}{c|c} P_0 & 0 \\ \hline 0 & 0_{q\times q} \end{array}\right], \qquad (km+q)\times(km+q) \text{ matrix.}$$

We define the problem

$$\begin{aligned} &\text{maximize } q_c(y, \lambda), \\ &\text{subject to } y \in V^{\perp}, \end{aligned} \tag{135}$$

as the locally *dual problem* of the problem

$$\begin{aligned} &\text{minimize } F(x, y, \lambda, c), \\ &\text{subject to } x \in R^n, \end{aligned} \tag{136}$$

where $V^{\perp} = \mathcal{R}(CA''|_{x^*})$, where

$$C = \left[\begin{array}{c|c} \tilde{G} & 0 \\ \hline 0 & I_{q\times q} \end{array}\right].$$

Algorithms (AS) and (AN) for Problem (NP) operate as follows.

Step 0. Choose vectors $y = y^{\circ} \in R^{km}$, $\lambda = \lambda^0 \in R^q$, and a scalar $c = c^0 > 0$.

Step 1. Find a perhaps local minimum $x^s = x(y^s, \lambda^s, c^s)$ of the problem

$$\begin{aligned} &\text{minimize } F(x, y,^s, \lambda^s, c^s), \\ &\text{subject to } x \in R^n. \end{aligned} \tag{137}$$

Step 2. Algorithm (AS). Update y^s, λ^s, c^s by

$$y_{ij}^{s+1} = \max[0, y_{ij}^s + c^s(f_{ij}(x^s) - \mu_i^s)], \qquad i = 1, \dots, k,\ j = 1, \dots, m, \tag{138}$$

where μ_i^s satisfies, for $i = 1, \dots, k$,

$$\sum_{j=1}^{m} \max[0, y_{ij}^s + c^s(f_{ij}(x^s) - \mu_i^s)] = 1,$$

$$\lambda_j^{s+1} = \lambda_j^s + c^s h_j[x^s, p_{c^s}[f_1(x^s), y_1^s], \dots, p_{c^s}[f_k(x^s), y_k^s]], \qquad j = 1, \dots, q, \tag{139}$$

$$c^{s+1} \geq c^s. \tag{140}$$

Step 2. Algorithm (AN). Update y^s, λ^s, c^s by

$$\begin{bmatrix} y^{s+1} \\ \lambda^{s+1} \end{bmatrix} = \begin{bmatrix} y^s \\ \lambda^s \end{bmatrix} - [H^s]^{-1} G^* \nabla p, \tag{141}$$

where H^s is an invertible matrix[4] which approximates the Hessian $\nabla^2 q_{c^s}(y^s, \lambda^s)$, ∇p is as in (133), G^* is defined similarly to G^* used by Algorithm (AN) for Problem (NPS),

$$c^{s+1} \geq c^s. \tag{142}$$

Set $s = s+1$, and go to Step 1.

Remarks similar to those made for the interpretation of Algorithms (AS) and (AN) in the (NPS) case hold here.

The convergence results (Theorems 4.1–4.4) hold, with the appropriate modifications. For example, we should use in Theorem 4.1

$$\begin{bmatrix} y \\ \lambda \end{bmatrix}, \qquad \text{instead of } y,$$

$$D_0 = -G^* \tilde{I} \tilde{\Delta}' \Xi(x, y, \lambda)^{-1} \tilde{\Delta} \tilde{I}, \qquad \text{instead of } D_0,$$

where

$$\tilde{I} = \begin{bmatrix} I_1 & & 0 \\ & \ddots & \\ 0 & & I_k \end{bmatrix},$$

$$I_i = m \times m \text{ matrix},$$

$$I_i = 0_{m \times m}, \qquad \text{if } p_i = 1,$$

$$I_i = \begin{bmatrix} \overbrace{\begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix}}^{p_i} & 0 \\ 0 & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{bmatrix}, \qquad \text{if } p_i \geq 2,$$

$$\tilde{\Delta} = [\nabla f \tilde{I} \,\vdots\, \nabla_x h + \nabla f \tilde{Y} \nabla_z h],$$

$$G^* = \begin{bmatrix} \overbrace{\begin{matrix} \nabla_1 G & & \\ & \ddots & \\ & & \nabla_1 G \end{matrix}}^{m} & & 0 \\ & \ddots & \\ 0 & & \overbrace{\begin{matrix} \nabla_k G & & \\ & \ddots & \\ & & \nabla_k G \end{matrix}}^{m} \end{bmatrix} \tilde{I}.$$

[4] The matrix H^s can be taken equal to the Hessian, if the Hessian is invertible, or equal to a matrix which is close to the Hessian, if the Hessian is not invertible or almost singular, like in the quasi-Newton methods for differentiable problems.

We assumed w.l.o.g. that

$$I_i(x^*) = \{1, \ldots, p_i\},$$

and $\Xi(x, y, \lambda)$ is the part of $\nabla_{xx}F(x, y, \lambda, c)$ which is not multiplied explicitly by c.

6. Conclusions

Our work in the previous sections has established certain results concerning the problem considered and the algorithms proposed. It should be clear that the results of this paper are very similar to results proved in Refs. 1–2 for multiplier methods. It is also clear that other results and remarks, given in the above-mentioned two papers, carry over to our case. For example, we can employ inexact minimization for the problems (69) or (137); see Proposition 2 in Ref. 2. We can also treat the case of inequality constraints $h_j \leq 0$ by introducing slack variables, although slack variables often introduce unnecessary numerical difficulties, see Ref. 24. It is felt that our main aim has been achieved, i.e., the establishment of a duality framework similar to the one holding for the multiplier methods and the demonstration that basic results concerning these methods hold for the problem and the algorithms considered here. The reader who is familiar with Ref. 5 can see that similar results would hold, should another type of γ function be considered. We did not present any implementation of the algorithms considered here, but the reader can find in Ref. 4 some implementation results of Algorithm (AS).

Recently, much attention has been focused on a class of methods called dual variable-metric algorithms (see Refs. 13, 20, 22, 6) and projected Langrangian algorithms (see Refs. 24–29). Although these methods were introduced for differentiable problems, it seems reasonable to expect that their basic philosophy is applicable to approximation methods for nondifferentiable problems, like Problem (NP).

References

1. BERTSEKAS, D. P., *Combined Primal-Dual and Penalty Methods for Constrained Minimization*, SIAM Journal on Control and Optimization, Vol. 13, No. 3, 1975.
2. BERTSEKAS, D. P., *On Penalty and Multiplier Methods for Constrained Minimization*, SIAM Journal on Control and Optimization, Vol. 14, No. 2, 1976.
3. BERTSEKAS, D. P., *Multiplier Methods: A Survey*, Automatica, Vol. 12, No. 1, 1976.

4. BERTSEKAS, D. P., *Nondifferentiable Optimization via Approximation*, Mathematical Programming Study 3, Edited by M. L. Balinski and P. Wolfe, North Holland Publishing Company, Amsterdam, Holland, pp. 1–25, 1975.
5. BERTSEKAS, D. P., *Approximation Procedures Based on the Method of Multipliers*, Coordinated Science Laboratory, University of Illinois, Urbana, Illinois, Working Paper, 1976.
6. BERTSEKAS, D. P., *On the Convergence Properties of Second-Order Multiplier Methods*, Coordinated Science Laboratory, University of Illinois, Urbana, Illinois, Working Paper, 1976.
7. BERTSEKAS, D. P., and MITTER, S. K., *A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functional*, SIAM Journal on Control and Optimization, Vol. 11, No. 4, 1973.
8. BUYS, J. D., *Dual Algorithms for Constrained Optimization*, Rijksuniversiteit de Leiden, PhD Thesis, 1972.
9. CLARKE, F. H., *Generalized Gradients and Applications*, Transactions of the American Mathematical Society, Vol. 205, No. 2, 1975.
10. CULLUM, J., *et al*, *An Algorithm for Minimizing Certain Nondifferentiable Convex Functions*, IBM Research, Report No. RC-4611, 1973.
11 DEM'YANOV, V. F., *Differentiability of a Maximum Function, I and II*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, Vol. 8, No. 1, 1968, and Vol. 9, No. 2, 1969.
12. GOLDSTEIN, A. A., *Optimization of Lipschitz Continuous Functions*, Mathematical Programming (to appear).
13. HAN, S. P., *Dual Variable-Metric Algorithms for Constrained Optimization*, SIAM Journal on Control and Optimization, Vol. 15, No. 4, 1977.
14. LEMARECHAL, C., *An Algorithm for Minimizing Convex Functions*, Proceedings of the 1974 IFIP Congress, North Holland Publishing Company, Amsterdam, Holland, 1974.
15. LUENBERGER, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1973.
16. BALINSKI, M. L., and WOLFE, P., Editors, *Nondifferentiable Optimization*, Mathematical Programming Study 3, North Holland Publishing Company, Amsterdam, Holland, 1975.
17. MIFFLIN, R., *An Algorithm for Nonsmooth Optimization*, School of Organization and Management, Yale University, Report No. 28, 1975.
18. POLYAK, B. T., *Minimization of Unsmooth Functionals*, Zhurnal Vychislitel'noi matematiki i Matematicheskoi Fiziki, Vol. 9, No. 3, 1969.
19. POLYAK, B. T., *The Convergence Rate of the Penalty Function Method*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, Vol. 11, No. 1, 1971.
20. POWELL, M. J. D., *Algorithms for Nonlinear Constraints That Use Langrangian Functions*, Ninth International Symposium on Mathematical Programming, Budapest, Hungary, 1976.
21. ROCKAFELLAR, R. T., *Solving a Nonlinear Programming Problem by Way of a Dual Problem*, Simposia Matematica (to appear).

22. TAPIA, R. A., *Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization*, Rice University, Houston, Texas, Department of Mathematical Sciences, 1976.
23. BERTSEKAS, D. P., *Class Notes on Optimization*, University of Illinois, Urbana, Illinois, Coordinated Science Laboratory, 1976.
24. SARGENT, R. W. H., *Reduced Gradient and Projection Methods for Nonlinear Programming*, Numerical Methods for Constrained Optimization, Edited by J. A. Gill and W. R. Murray, Academic Press, New York, New York, 1974.
25. WILSON, R. B., *A Simplicial Algorithm for Concave Programming*, Harvard University, Graduate School of Business Administration, PhD Thesis, 1963.
26. MURRAY, W., *Constrained Optimization*, National Physical Laboratory, Teddington, England, Report No. MA-79, 1969.
27. MURRAY, W., *An Algorithm for Constrained Minimization*, Optimization, Edited by R. Fletcher, Academic Press, New York, New York, 1969.
28. WRIGHT, M. H., *Numerical Methods for Nonlinearly Constrained Optimization*, Stanford Linear Accelerator Center, Report No. 193, 1976.
29. HAN, S. P., *Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems*, Mathematical Programming, Vol. 11, No. 1, 1976.