Partial Linearization Methods in Nonlinear Programming[1]

M. Patriksson[2]

Communicated by D.G. Luenberger

[1]The author wishes to thank Drs. K. Holmberg, T. Larsson, and A. Migdalas for their helpful comments.
[2]Lic. Phil., Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden.

Abstract. In this paper, we characterize a class of feasible direction methods in nonlinear programming through the concept of partial linearization of the objective function. Based on a feasible point, the objective is replaced by an arbitrary convex and continuously differentiable function, and the error is taken into account by a first order approximation of it. A new feasible point is defined through a line search with respect to the original objective, towards the solution of the approximate problem. Global convergence results are given for exact and approximate line searches, and possible interpretations are made. We present some instances of the general algorithm, and discuss extensions to nondifferentiable programming.

Key Words. Feasible direction methods, partial linearization, regularization, nondifferentiable programming.

1. Introduction

The purpose of this paper is to unify a number of feasible direction methods in nonlinear programming. Examples of these are the method of Frank and Wolfe and constrained Newton methods. The problem studied is

$$\text{(P)} \qquad \min_{x \in \mathbf{X}} T(x),$$

where $T : \Re^n \mapsto \Re$ is continuous, and $\mathbf{X}$ is assumed compact and convex with a nonempty interior. These conditions ensure the existence of a solution to (P).

Let the function $f(x, y)$ be convex and continuously differentiable with respect to x, and continuous with respect to y. The idea of the partial linearization method is to use the function $f\left(x, x^{(l)}\right)$ in iteration l instead of the original objective T, supposedly with an easier problem than (P) as a result. The error made, $T(x) - f\left(x, x^{(l)}\right)$, is taken into account by approximating it with a first order Taylor expansion around the feasible point $x^{(l)}$. The solution to this approximate problem defines a descent direction for the original objective T; a line search in this direction gives a new iteration point.

We are now ready to state the algorithm, and establish its global convergence.

2. The Partial Linearization Algorithm

In order to establish the properties of the partial linearization algorithm, it will first be stated for a problem over a compact and convex feasible set, where the original objective T is pseudoconvex and continuously differentiable.

The general algorithm works as follows.

In iteration l, the objective function is expressed as

$$T(x) = f\left(x, x^{(l)}\right) + \left(T(x) - f\left(x, x^{(l)}\right)\right). \tag{1}$$

The second term can be seen as expressing the error when replacing the original objective, T, by the function $f\left(\cdot, x^{(l)}\right)$. The idea is then to take this error into account by replacing it with a first order Taylor expansion around $x^{(l)}$.

The subproblem solved in iteration l then is

$$\text{(PL-SUB}^{(l)}\text{)} \qquad \min_{x \in \mathbf{X}} \; T^{(l)}(x),$$

where

$$T^{(l)}(x) = f\left(x, x^{(l)}\right) + T\left(x^{(l)}\right) - f\left(x^{(l)}, x^{(l)}\right) + \left[\nabla T\left(x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}} \left(x - x^{(l)}\right),$$

and where $\nabla_x f(x, y)$ denotes the gradient of $f(x, y)$ with respect to x.

An optimal solution to this convex problem, $\overline{x}^{(l)}$, defines a direction $d^{(l)} \stackrel{\text{def}}{=} \overline{x}^{(l)} - x^{(l)}$, which is a feasible descent direction with respect to T, if not $x^{(l)}$ solves (PL-SUB$^{(l)}$), in which case $x^{(l)}$ is an optimal solution to (P). [This is to be shown in Theorem 2.1.] A line search with respect to T is then made in the direction $d^{(l)}$, and a new iteration point $x^{(l+1)}$ is obtained. Under the assumptions made on (P), any accumulation point of the sequence $\left\{x^{(l)}\right\}_{l=1}^{\infty}$ is optimal to (P), for any initial feasible point $x^{(0)}$. [This is shown in Theorem 2.2.]

Theorem 2.1 Assume that $x^{(l)}$ is a feasible point in (P), and that $\overline{x}^{(l)}$ solves (PL-SUB$^{(l)}$). If $x^{(l)}$ solves (PL-SUB$^{(l)}$), then $x^{(l)}$ is optimal in (P). Otherwise, the direction $d^{(l)} = \overline{x}^{(l)} - x^{(l)}$ is a feasible descent direction with respect to T.

Proof If $x^{(l)}$ does not solve (PL-SUB$^{(l)}$), then by the convexity of $T^{(l)}$, and the fact that $\nabla T\left(\mathbf{x}^{(l)}\right) = \nabla T^{(l)}\left(x^{(l)}\right)$, we have that

$$\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right) < 0,$$

and by the convexity of $\mathbf{X}$ the direction $d^{(l)} = \overline{x}^{(l)} - x^{(l)}$ is a feasible descent direction with respect to T.

Since $\overline{x}^{(l)}$ solves (PL-SUB$^{(l)}$), the following variational inequality holds [Bazaraa and Shetty, Ref. 1, Theorem 3.4.3]:

$$\left[\nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) + \nabla T\left(x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}}\left(x - \overline{x}^{(l)}\right) \geq 0, \ \forall\ x \in \mathbf{X}. \tag{2}$$

If $x^{(l)}$ also solves the subproblem, then by replacing $\overline{x}^{(l)}$ with $x^{(l)}$ above, we obtain

$$\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(x - x^{(l)}\right) \geq 0, \ \forall\ x \in \mathbf{X},$$

which, by the pseudoconvexity of T, yields that $x^{(l)}$ solves (P). □

Theorem 2.2 The partial linearization algorithm either terminates in a finite number of main iterations or it generates an infinite sequence $\left\{x^{(l)}\right\}$ such that any accumulation point is an optimal solution to (P).

Proof The proof is a modification of the ones provided by Larsson and Migdalas (Ref. 2) and Larsson and Patriksson (Ref. 3), which in turn are based on the convergence proof of the Frank-Wolfe algorithm, provided by Martos (Ref. 4).

If the algorithm terminates in some iteration l, we have from Theorem 2.1 that

$$T^{(l)}\left(\overline{x}^{(l)}\right) = T^{(l)}\left(x^{(l)}\right)$$

and $x^{(l)}$ then is optimal in (P). Assume therefore that $T^{(l)}\left(\overline{x}^{(l)}\right) \neq T^{(l)}\left(x^{(l)}\right)$ for all iterations l.

By Theorem 2.1, the sequence $\left\{d^{(l)}\right\}$ consists of directions of descent. Therefore,

$$T\left(x^{(l+1)}\right) = \min_{\alpha_l \in [0,\overline{\alpha}_l]} T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) < T\left(x^{(l)}\right), \tag{3}$$

where $\overline{\alpha}_l$ is the maximal steplength in the direction $d^{(l)}$. Thus, $\left\{T\left(x^{(l)}\right)\right\}$ is a monotonically decreasing sequence which is bounded from below since $\mathbf{X}$ is compact and T continuous. Furthermore, according to Weierstrass' theorem, the sequence $\left\{x^{(l)}\right\}$ must have at least one accumulation point, say $x^* \in \mathbf{X}$, which is then a limit point of a convergent subsequence $\left\{x^{(l)}\right\}_{l\in L}$. We then have the following.

$$\lim_{l\in L} T\left(x^{(l)}\right) = T\left(x^*\right) \tag{4}$$

By (4) we have that

$$\lim_{l\in L}\left[T\left(x^{(l+1)}\right) - T\left(x^{(l)}\right)\right] = 0. \tag{5}$$

Again, according to Weierstrass' theorem, the sequence $\left\{\overline{x}^{(l)}\right\}$ has at least one accumulation point, say $\overline{x}$, being a limit point of a sequence $\hat{L} \subset L$. By (3) we have that

$$T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) \geq T\left(x^{(l+1)}\right), \ \forall\ \alpha_l \in [0, \overline{\alpha}_l].$$

Then,

$$T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right) \geq T\left(x^{(l+1)}\right) - T\left(x^{(l)}\right),$$

for all $\alpha_l \in [0, \overline{\alpha}_l]$. By (5),

$$\lim_{l\in\hat{L}}\left[T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right)\right] \geq 0,$$

and, by choosing α_l so that $\lim_{l\in\hat{L}} \alpha_l = 0$, we obtain

$$\lim_{l\in\hat{L}}\left[\frac{T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right)}{\alpha_l}\right] \geq 0,$$

and the limit exists due to the strict differentiability of T [Rockafellar, Ref. 5, Theorem 4F]. The limit is equal to the directional derivative of $T\left(x^*\right)$ in the direction $\overline{x} - x^*$ [Ref. 5, Equation (3.5)], and hence

$$\nabla T\left(x^*\right)^{\mathrm{T}}\left(\overline{x} - x^*\right) \geq 0. \tag{6}$$

Since $\overline{x}^{(l)}$ solves (PL-SUB$^{(l)}$), then from (2) we obtain [Ref. 1, Theorem 3.4.3]

$$\left[\nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) + \nabla T\left(x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}}\left(x - \overline{x}^{(l)}\right) \geq 0, \ \forall\ x \in \mathbf{X}.$$

Taking the limit over $\hat{L}$ yields

$$\left[\nabla_x f\left(\overline{x}, x^*\right) + \nabla T\left(x^*\right) - \nabla_x f\left(x^*, x^*\right)\right]^{\mathrm{T}}\left(x - \overline{x}\right) \geq 0, \ \forall\ x \in \mathbf{X}. \tag{7}$$

Choosing $x = x^*$, and noting the convexity of $f(x, y)$ with respect to x, we obtain that

$$\nabla T\left(x^*\right)^{\mathrm{T}}\left(x^* - \overline{x}\right) \geq 0.$$

This implies, by (6), that

$$\nabla T(x^*)^{\mathrm{T}}(x^* - \overline{x}) = 0. \tag{8}$$

If either component is zero, x^* solves (P), by the pseudoconvexity of T and by (7), respectively. Assume therefore that $x^* \neq \overline{x}$, and that $\nabla T(x^*) \neq 0$.

According to (7), $\overline{x}$ solves the subproblem defined by x^*. If we denote the subproblem objective by $T^*(x)$, then we have that $T^*(\overline{x}) \leq T^*(x^*)$, i.e., that

$$f(\overline{x}, x^*) - f(x^*, x^*) + \left[\nabla T(x^*) - \nabla_x f(x^*, x^*)\right]^{\mathrm{T}} (\overline{x} - x^*) \leq 0,$$

which by (8) and the convexity of f yields that

$$f(\overline{x}, x^*) = f(x^*, x^*) + \nabla_x f(x^*, x^*)^{\mathrm{T}} (\overline{x} - x^*).$$

But this in turn implies that

$$\begin{aligned} T^*(\overline{x}) &= \lim_{l \in \hat{L}} T^{(l)}\left(\overline{x}^{(l)}\right) \\ &= f(\overline{x}, x^*) + T(x^*) - f(x^*, x^*) + \left[\nabla T(x^*) - \nabla_x f(x^*, x^*)\right]^{\mathrm{T}} (\overline{x} - x^*) \\ &= T(x^*). \end{aligned}$$

But since $T^*(x^*) = T(x^*)$, we have that x^* *also* solves the subproblem defined by x^*. Replacing $\overline{x}$ by x^* in (7), we obtain that

$$\nabla T\left(x^*\right)^{\mathrm{T}} \left(x - x^*\right) \geq 0, \;\; \forall\, x \in \mathbf{X},$$

which is the desired result. □

Under the additional assumption that $T(x)$ has a *Lipschitzian gradient*, i.e., that there exists a positive constant K such that

$$\left\|\nabla T\left(x^1\right) - \nabla T\left(x^2\right)\right\| \leq K \left\|x^1 - x^2\right\|, \;\forall\, x^1, x^2 \in \mathbf{X}, \tag{9}$$

then it is possible to show that the partial linearization algorithm is globally convergent even when line searches are made only approximately. [A sufficient condition for Lipschitz continuity of ∇T is continuous differentiability of T, since the feasible set is assumed compact.] As an example, consider the case of replacing it with an Armijo (Ref. 6) steplength rule where the point $x^{(l+1)}$ is calculated as $x^{(l+1)} = x^{(l)} + \beta^{-\bar{\imath}}\left(\overline{x}^{(l)} - x^{(l)}\right)$, where $\bar{\imath}$ is the smallest integer i for which

$$T\left(x^{(l)} + \beta^{-i}\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right) \leq \varepsilon \beta^{-i} \nabla T\left(x^{(l)}\right)^{\mathrm{T}} \left(\overline{x}^{(l)} - x^{(l)}\right), \tag{10}$$

where $\varepsilon \in (0,1)$ and $\beta > 1$. Below, we establish the global convergence of the modified algorithm.

Theorem 2.3 Under the additional condition (9), the partial linearization algorithm is globally convergent with exact line searches replaced by the Armijo steplength rule (10).

Proof We will show that any convergent subsequence $\hat{L}$ of the sequences $\left\{x^{(l)}\right\}_{l=1}^{\infty}$ and $\left\{\overline{x}^{(l)}\right\}_{l=1}^{\infty}$ satisfies

$$\lim_{l \in \hat{L}} \nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right) = \nabla T(x^*)^{\mathrm{T}}(\overline{x} - x^*) = 0,$$

i.e., equivalent to (8). The proof then follows using exactly the same arguments as in the proof of Theorem 2.2.

By Theorem 2.1, the sequence $\left\{d^{(l)}\right\}$ consists of feasible directions of descent. Hence,

$$\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right) < 0,$$

and by (10),

$$T\left(x^{(l+1)}\right) \leq T\left(x^{(l)}\right), \ \forall\ l.$$

Indeed, by the construction of the steplength rule,

$$T\left(x^{(l+1)}\right) \leq T\left(x^{(l)}\right) + \varepsilon\alpha_l \nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right). \tag{11}$$

This inequality is valid, since (10) can always be satisfied within a finite number of trial steps. The proof of this fact, given below, is a modification of a part of the convergence proof for the Frank-Wolfe algorithm of Pshenichny and Danilin (Ref. 7, p. 172).

Using Taylor's formula and the condition (9),

$$\begin{aligned} T\left(x^{(l)} + \alpha\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right) &= \int_0^{\alpha} \nabla T\left(x^{(l)} + s\left(\overline{x}^{(l)} - x^{(l)}\right)\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right) ds \\ &\leq \alpha \nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right) + \int_0^{\alpha} sK \left\|\overline{x}^{(l)} - x^{(l)}\right\|^2 ds \\ &= \alpha \nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right) + \frac{1}{2}\alpha^2 K \left\|\overline{x}^{(l)} - x^{(l)}\right\|^2. \end{aligned} \tag{12}$$

From (12) we obtain, by choosing

$$\alpha \leq 2(\varepsilon - 1)\frac{\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right)}{K\left\|\overline{x}^{(l)} - x^{(l)}\right\|^2},$$

that

$$T\left(x^{(l)} + \alpha\left(\overline{x}^{(l)} - x^{(l)}\right)\right) \leq T\left(x^{(l)}\right) + \varepsilon\alpha \nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right).$$

Replacing α with β^{-i}, we see that $\bar{\imath}$ is the smallest integer to satisfy

$$\beta^{-i} \leq 2(\varepsilon - 1)\frac{\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)} - x^{(l)}\right)}{K\left\|\overline{x}^{(l)} - x^{(l)}\right\|^2}.$$

Hence,

$$\beta\alpha_l = \beta^{-(\bar{\imath}-1)} > 2(\varepsilon-1)\frac{\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)}{K\left\|\overline{x}^{(l)}-x^{(l)}\right\|^2},$$

i.e.,

$$\alpha_l > \frac{2(\varepsilon-1)}{\beta}\frac{\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)}{K\left\|\overline{x}^{(l)}-x^{(l)}\right\|^2}. \tag{13}$$

Hence, the Inequality (10) will be satisfied after a finite number of steps with a steplength satisfying (13).

Noting that $\left\|\overline{x}^{(l)}-x^{(l)}\right\|$ is bounded from above by a positive constant, say C, because of the compactness of $\mathbf{X}$, by (11) and (13) we obtain

$$T\left(x^{(l+1)}\right)-T\left(x^{(l)}\right) \leq \frac{2\varepsilon(\varepsilon-1)}{\beta}\frac{\left[\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)\right]^2}{KC^2}. \tag{14}$$

Summing (14) for all $l \leq m-1$, we obtain

$$T\left(x^{(m)}\right)-T\left(x^{(0)}\right) \leq -\frac{2\varepsilon(\varepsilon-1)}{\beta KC^2}\sum_{l=1}^{m-1}\left[\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)\right]^2.$$

Using the fact that $T(x)$ is bounded from below on $\mathbf{X}$ by, say $\underline{T}$,

$$\begin{aligned}\sum_{l=1}^{m-1}\left[\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)\right]^2 &\leq \frac{\beta KC^2}{2\varepsilon(\varepsilon-1)}\left(T\left(x^{(0)}\right)-T\left(x^{(m)}\right)\right)\\ &\leq \frac{\beta KC^2}{2\varepsilon(\varepsilon-1)}\left(T\left(x^{(0)}\right)-\underline{T}\right).\end{aligned}$$

It follows that the series

$$\sum_{l=1}^{\infty}\left[\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)\right]^2$$

is convergent, and hence,

$$\lim_{l\to\infty}\nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(\overline{x}^{(l)}-x^{(l)}\right)=0. \tag{15}$$

By using the same arguments as in Theorem 2.2, we may choose subsequences of $\left\{x^{(l)}\right\}$ and $\left\{\overline{x}^{(l)}\right\}$ so that

$$\lim_{l\in L} x^{(l)} = x^* \quad \text{and} \quad \lim_{l\in\hat{L}}\overline{x}^{(l)}=\overline{x}.$$

By (15),

$$\nabla T(x^*)^{\mathrm{T}}(\overline{x}-x^*)=0,$$

which completes the proof. □

3. Interpretations and Some Instances

The convergence proof of Theorem 2.2 implies several properties of the partial linearization algorithm.

A nice property of the method of Frank and Wolfe (Ref. 8), which may be recognized by the choice of the function $f(x,y) \equiv 0$, is that, for convex problems, a termination criterion is available through the lower bound on the optimal objective value, obtained from the linear subproblem. In the partial linearization algorithm, this termination criterion is, in general, not valid since the error function, $T(x) - f(x,y)$, is not convex. However, by Theorem 2.2,

$$T\left(x^{(l)}\right) - T^{(l)}\left(\overline{x}^{(l)}\right) > 0$$

whenever $x^{(l)}$ is not a solution to (P), and this difference tends towards zero when the iterations proceed. Hence, the difference between the objective value and the corresponding optimal subproblem value may still be utilized in a termination criterion.

If the function $f(x,y)$ is chosen *strictly* convex with respect to x, then the subproblem (PL-SUB$^{(l)}$) is strictly convex, and the solution, $\overline{x}^{(l)}$, is unique. This implies the possibility of utilizing dual techniques for the solution of (PL-SUB$^{(l)}$); furthermore, the termination criterion $T^{(l)}\left(\overline{x}^{(l)}\right) - T^{(l)}\left(x^{(l)}\right) = 0$ is replaced by $\overline{x}^{(l)} = x^{(l)}$, and it is easily shown that any accumulation point of the sequence of subproblem solutions is a solution to (P). Hence, the choice of strictly convex functions $f(x,y)$ corresponds to making better and better approximations of the original objective, in the sense that in the limit the original problem is solved by the subproblem solution.

If the function $f(x,y)$ is chosen so that

$$\nabla_x f(x,y) = 0 \quad \text{if} \quad x = y, \tag{31}$$

then the subproblem of the partial linearization algorithm is equivalent to

$$\min_{x \in \mathbf{X}} \ \nabla T\left(x^{(l)}\right)^{\mathrm{T}}\left(x - x^{(l)}\right) + f\left(x, x^{(l)}\right).$$

Algorithms of this type are studied by Migdalas (Ref. 9), under the name of *regularized Frank-Wolfe algorithms*. The above expression provides a nice interpretation of the algorithmic class as *partial* linearization methods as opposed to *complete* linearization methods, such as the Frank-Wolfe algorithm. In the Frank-Wolfe algorithm, the first order Taylor expansion, which is valid only locally around $x^{(l)}$, is used globally in the subproblem phase. The subproblem solved in a partial linearization method introduces a regularization term in the objective function of the Frank-Wolfe subproblem, restricting the difference between the current point $x^{(l)}$ and the subproblem solution $\overline{x}^{(l)}$ to some measured distance. By retaining the nonlinearity of the original objective function, partial linearization methods will therefore avoid the *tailing-off* phenomena inherent in the Frank-Wolfe method, caused by the generation of extreme point solutions. This was the original idea behind the first formulation of a partial linearization method (Ref. 3), where the objective is described as the sum of a strictly convex and differentiable function and a differentiable but not necessarily convex function. The subproblem then is defined by a linearization of the nonconvex part. Also, it is pointed out the possibility of choosing the strictly convex function separable to utilize separability in the constraints. This is demonstrated in the case of traffic equilibrium problems in Refs. 10 and 11.

The partial linearization method may also be used to regularize *not* strictly convex problems by iteratively adding a strictly convex function to the original objective, as shown below.

Assume that the objective function $T(x)$ is convex but *not* strictly convex. Assume also that the function $f(x, y)$ is strictly convex with respect to x. Let

$$\overline{f}(x, y) = T(x) + f(x, y),$$

which then is strictly convex with respect to x. Rewrite the objective as in (1), i.e., let

$$\begin{aligned} T(x) &= \overline{f}(x, y) + \left(T(x) - \overline{f}(x, y)\right) \\ &= T(x) + f(x, y) + (-f(x, y)). \end{aligned}$$

The subproblem solved in the partial linearization method then is to minimize

$$T^{(l)}(x) = T(x) + f\left(x, x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)^{\mathrm{T}}\left(x - x^{(l)}\right),$$

and if f is chosen to satisfy (31), the subproblem objective equals

$$T^{(l)}(x) = T(x) + f\left(x, x^{(l)}\right),$$

i.e., the subproblem introduces a *regularization* of the original objective by adding to it a strictly convex function. [For a general discussion on regularization methods in nonlinear programming, see, e.g., Ref. 12.]

As an example, consider the function

$$f(x, y) = \frac{c}{2}(x - y)^{\mathrm{T}}\mathbf{I}(x - y) = \frac{c}{2}\left\|x - y\right\|_2^2,$$

where $c > 0$. The subproblem, given a feasible point $x^{(l)}$, then is equal to the minimization of the function

$$T(x) + \frac{c}{2}\left\|x - x^{(l)}\right\|_2^2,$$

which is equivalent to the subproblem of the proximal point algorithm (Ref. 13).

This also suggests the possibility of using partial linearization methods for nondifferentiable problems; this is further discussed in Section 4.

3.1. Some Instances of the Partial Linearization Algorithm

By choosing the function $f(x, y)$ in various ways, well known feasible direction methods are recognized. Below we list a few such methods. For further discussions, we refer to Ref. 11.

$f(x,y)$	Algorithm
0	Frank-Wolfe
$\frac{1}{2}x^{\mathrm{T}}\nabla^2 T(y)x$	constrained Newton
$\frac{1}{2}x^{\mathrm{T}}\mathbf{B}(y)x$	constrained quasi-Newton
$T(x)+\dfrac{c}{2}\left\|x-y\right\|^2$	proximal point
$\displaystyle\sum_{i=1}^{n}\int_0^{\sum_{j\neq i} y_j + x_i} \frac{\partial T(s)}{\partial x_i} ds$	nonlinear Jacobi

Note that when $T(x)$ is only pseudoconvex, the Hessian $\nabla^2 T(x)$ is not necessarily positive semidefinite; however constrained Newton methods can always be modified by adding a positive definite matrix to the Hessian, e.g., let

$$f(x,y) = \frac{1}{2}x^{\mathrm{T}}\left(\nabla^2 T(y) + \varepsilon\mathbf{I}\right)x,$$

where $\varepsilon > 0$ is sufficiently large so that the matrix $\nabla^2 T(y) + \varepsilon\mathbf{I}$ is (at least) positive semidefinite.

4. Nondifferentiable Programming

In this section we extend the partial linearization algorithm to nondifferentiable programming. Let the objective $T(x)$ be given as

$$T(x) = g(x) + h(x), \tag{48}$$

where $g(x) : \Re^n \mapsto (-\infty, +\infty]$ is a closed and proper convex function and $h(x) : \Re^n \mapsto (-\infty, +\infty]$ is continuously differentiable on $\mathbf{X}$. The problem of minimizing T over the compact and convex set $\mathbf{X}$ will be referred to as (NP).

A Frank-Wolfe type algorithm for nondifferentiable problems is defined through a linearization of the continuously differentiable part of the objective, i.e., in iteration l,

$$T^{(l)}(x) = g(x) + h\left(x^{(l)}\right) + \nabla h\left(x^{(l)}\right)^{\mathrm{T}}\left(x - x^{(l)}\right).$$

Mine and Fukushima (Ref. 14) propose such a descent method, coupled with exact line searches on T. The algorithm is applied to unconstrained problems; it is shown that, under a compactness assumption on the level set of $T\left(x^{(0)}\right)$ and a strict convexity assumption on

$g(x)$, any accumulation point of the sequence $\left\{x^{(l)}\right\}$ is a critical point, i.e., that 0 belongs to the Clarke (Ref. 15) generalized subdifferential of T at an accumulation point x^*, or

$$-\nabla h\left(x^*\right) \in \partial g\left(x^*\right).$$

The algorithm is extended by Fukushima and Mine (Ref. 16), where a proximal point type method is applied. The subproblem in iteration l is to minimize

$$g(x) + h\left(x^{(l)}\right) + \nabla h\left(x^{(l)}\right)^{\mathrm{T}}\left(x - x^{(l)}\right) + \frac{c_l}{2}\left\|x - x^{(l)}\right\|^2,$$

where $c_l > 0$. Furthermore, the exact line search is replaced by a modified Armijo steplength rule. They establish the convergence towards a critical point x^*, and also a linear rate of convergence, provided that $\nabla^2 h\left(x^*\right)$ is strictly positive definite and that $c_l \equiv c$ is chosen large enough.

Extending the partial linearization algorithm to nondifferentiable programming, let in iteration l the objective be described as

$$T(x) = \left(g(x) + f\left(x, x^{(l)}\right)\right) + \left(h(x) - f\left(x, x^{(l)}\right)\right),$$

and the subproblem is defined through a linearization of the second term.

We may identify the algorithm in Ref. 14 by the choice of $f \equiv 0$, and the proximal point algorithm in Ref. 16 is obtained by letting $f(x, y) = \dfrac{c}{2}\left\|x - y\right\|^2$. [The possibility of making subproblems strictly convex by adding and subtracting a strictly convex function is discussed in Ref. 14.]

Fukushima (Ref. 17) further develops the method in Ref. 16, for constrained problems. The subproblem used may be defined by letting $f\left(x, x^{(l)}\right) = \frac{1}{2}x^{\mathrm{T}}\mathbf{B}_l x$, and by linearizing nonlinear constraint functions. The matrix $\mathbf{B}_l$ is assumed positive definite. The method of Kiwiel (Ref. 18) resembles the method of Ref. 16, but includes polyhedral approximations of g in each step of the algorithm to simplify the computation of search directions.

We will show that the partial linearization algorithm described above extends the results obtained in Refs. 14 and 16 in the sense that the partial linearization algorithm is convergent towards critical points, and that approximate line searches may be used.

The results are given in three theorems. First we show that the subproblem solutions yield descent directions. Then we prove the convergence of two versions of the algorithm, where the line search is made exactly and by applying a modified Armijo steplength rule, respectively.

The first theorem is a generalization of Theorem 2.1.

Theorem 4.1 Assume that $x^{(l)}$ is feasible in (NP), and that $\overline{x}^{(l)}$ solves (PL-SUB$^{(l)}$). If $x^{(l)}$ solves (PL-SUB$^{(l)}$), then $x^{(l)}$ is a critical point in (NP). Otherwise, the direction $d^{(l)} = \overline{x}^{(l)} - x^{(l)}$ is a feasible descent direction with respect to T.

Proof The proof extends Lemma 3.2 in Ref. 14, where it is also assumed that the function g is strictly convex on its domain, and is based on Theorem 2.1.

We will show that for any $x^{(l)}$ the directional derivative of T in the direction $d^{(l)}$ is nonpositive, and that it is strictly negative if $x^{(l)}$ is not a critical point, i.e., that

$$T^{(l)}\left(\overline{x}^{(l)}\right) < T^{(l)}\left(x^{(l)}\right) \text{ implies } T'\left(x^{(l)}; d^{(l)}\right) < 0.$$

$$\begin{aligned} T'\left(x^{(l)}; d^{(l)}\right) &\stackrel{\text{def}}{=} \lim_{\lambda \to 0_+} \frac{1}{\lambda}\left[T\left(x^{(l)} + \lambda d^{(l)}\right) - T\left(x^{(l)}\right)\right] \\ &= \sup_{u \in \partial g\left(x^{(l)}\right)} \left[\nabla h\left(x^{(l)}\right) + u\right]^{\mathrm{T}} d^{(l)} \\ &= \nabla h\left(x^{(l)}\right)^{\mathrm{T}} d^{(l)} + \sup_{u \in \partial g\left(x^{(l)}\right)} u^{\mathrm{T}} d^{(l)}, \end{aligned}$$

where the second equality follows from Ref. 19, Theorem 23.4.

If $x^{(l)}$ does not solve the subproblem, then by the convexity of $f(x, y)$ with respect to x it is easy to show that

$$\nabla h\left(x^{(l)}\right)^{\mathrm{T}} d^{(l)} < g\left(x^{(l)}\right) - g\left(\overline{x}^{(l)}\right).$$

Using this fact it follows, by the convexity of g, that

$$T'\left(x^{(l)}; d^{(l)}\right) < g\left(x^{(l)}\right) - g\left(\overline{x}^{(l)}\right) + \sup_{u \in \partial g\left(x^{(l)}\right)} u^{\mathrm{T}} d^{(l)} \leq 0.$$

Hence, the direction $d^{(l)}$ is a feasible descent direction with respect to T. However, if $x^{(l)}$ solves the subproblem, we then have that the variational inequality

$$\left[u^{(l)} + \nabla h\left(x^{(l)}\right)\right]^{\mathrm{T}}\left(x - x^{(l)}\right) \geq 0, \ \forall\ x \in \mathbf{X}$$

holds for some $u^{(l)} \in \partial g\left(x^{(l)}\right)$. Hence, $x^{(l)}$ is a critical point for the problem (NP), and, in the case of convex functions h, this is a sufficient condition for global optimality of the point $x^{(l)}$ in the problem (NP) [Ref. 1, Theorem 3.4.3]. This completes the proof. □

Theorem 4.2 The partial linearization algorithm applied to (NP) either terminates in a finite number of main iterations or it generates an infinite sequence $\left\{x^{(l)}\right\}$ such that any accumulation point is a critical point in (NP). Especially, if the function $h(x)$ is convex, then the algorithm is globally convergent.

Proof The proof is a modification of the one given in Theorem 2.2.

According to Theorem 4.1, the algorithm terminates in a finite number of main iterations if the subproblem in iteration l is solved by $x^{(l)}$. Assume therefore that

$$T^{(l)}\left(\overline{x}^{(l)}\right) < T^{(l)}\left(x^{(l)}\right)$$

for all iterations l.

By the same arguments as in the proof of Theorem 2.2, the algorithm yields descent directions, and we may choose a subsequence $\hat{L}$ such that $x^{(l)}$ and $\overline{x}^{(l)}$ has limit points x^* and $\overline{x}$, respectively. It follows that

$$T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right) \geq T\left(x^{(l+1)}\right) - T\left(x^{(l)}\right),$$

for all $\alpha_l \in [0, \overline{\alpha}_l]$. Since $T\left(x^{(l+1)}\right) - T\left(x^{(l)}\right)$ tends to zero in $\hat{L}$, we have that

$$\lim_{l \in \hat{L}} \left[T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right)\right] \geq 0. \tag{49}$$

Choose the sequence of α_l so that α_l tends to zero in $\hat{L}$. The *upper subderivative* of T at x^* with respect to the direction $d^* = \overline{x} - x^*$ is defined as

$$T^{\uparrow}(x^*; d^*) = \limsup_{\substack{x^{(l)} \xrightarrow{T} x^* \\ l \in \hat{L}}} \inf_{d^{(l)} \to d^*} \frac{T\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - T\left(x^{(l)}\right)}{\alpha_l}.$$

The convex function g is finite everywhere, since it is continuous on the compact set $\mathbf{X}$. Hence by Theorem 3G of Ref. 5 the subderivative of g at x^* with respect to d^* satisfies the relation

$$g^{\uparrow}(x^*; \overline{x} - x^*) = g'(x^*; \overline{x} - x^*).$$

Now, the function h is continuously differentiable, and hence the limit of

$$\frac{h\left(x^{(l)} + \alpha_l\left(\overline{x}^{(l)} - x^{(l)}\right)\right) - h\left(x^{(l)}\right)}{\alpha_l}$$

exists and is equal to $\nabla h(x^*)^{\mathrm{T}}(\overline{x} - x^*)$ [Ref. 5, Theorem 4F]. By Theorem 4.4.a of Ref. 20,

$$T^{\uparrow}(x^*; d^*) = g'(x^*; \overline{x} - x^*) + \nabla h(x^*)^{\mathrm{T}}(\overline{x} - x^*), \tag{50}$$

which is nonnegative by (49).

Theorem 23.4 of Ref. 19 yields that

$$g'(x^*; \overline{x} - x^*) = \sup_{u \in \partial g(x^*)} u^{\mathrm{T}}(\overline{x} - x^*),$$

and the supremum is attained in, say $u^* \in \partial g(x^*)$, since $\partial g(x^*)$ is compact [see Ref. 19, Theorem 24.7 and Ref. 21, Theorem 1.7]. Hence, from (50) there exists a subgradient u^* of g in the point x^* such that

$$\left[u^* + \nabla h(x^*)\right]^{\mathrm{T}}(\overline{x} - x^*) \geq 0. \tag{51}$$

Hence, in the limit the subproblem does not yield a descent direction.

In iteration l, the subproblem is equivalent to the following variational inequality problem [Ref. 1, Theorem 3.4.3]:

there exists a subgradient $\overline{u}^{(l)}$ of g at $\overline{x}^{(l)}$ such that for all $x \in \mathbf{X}$,

$$\left[\overline{u}^{(l)} + \nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) + \nabla h\left(x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}}\left(x - \overline{x}^{(l)}\right) \geq 0. \tag{52}$$

Noting that if $\overline{u}^{(l)} \in \partial g\left(\overline{x}^{(l)}\right)$, then by the compactness of $\partial g\left(\overline{x}^{(l)}\right)$ for all l, a subsequence of $\hat{L}$ of the subgradients $\overline{u}^{(l)}$ converges towards a point $\overline{u}$, which by Theorem 24.4 of Ref. 19 satisfies $\overline{u} \in \partial g\left(\overline{x}\right)$.

Hence, taking the limit over this subsequence, (52) yields

$$\left[\overline{u} + \nabla_x f\left(\overline{x}, x^*\right) + \nabla h\left(x^*\right) - \nabla_x f\left(x^*, x^*\right)\right]^{\mathrm{T}}\left(x - x^*\right) \geq 0, \ \forall \ x \in \mathbf{X}, \tag{53}$$

which implies that $\overline{x}$ solves the subproblem of the partial linearization algorithm defined by the point x^*.

By the monotonicity of $\nabla_x f\left(x, y\right)$ with respect to x, (53) implies, by the choice of $x = \overline{x}$, that

$$\left[\overline{u} + \nabla h\left(x^*\right)\right]^{\mathrm{T}}\left(x^* - \overline{x}\right) \geq 0. \tag{54}$$

Summing (51) and (54), we obtain that

$$\left[\overline{u} - u^*\right]^{\mathrm{T}}\left(x^* - \overline{x}\right) \geq 0.$$

But by the definition of convexity of g in the points $\overline{x}$ and x^*, the inequality is reversed, so that, in fact,

$$\left[\overline{u} - u^*\right]^{\mathrm{T}}\left(x^* - \overline{x}\right) = 0.$$

Substituting this in (54),

$$\left[u^* + \nabla h\left(x^*\right)\right]^{\mathrm{T}}\left(x^* - \overline{x}\right) \geq 0,$$

which by (51) implies

$$\left[u^* + \nabla h\left(x^*\right)\right]^{\mathrm{T}}\left(x^* - \overline{x}\right) = 0.$$

Using the same arguments as in the proof of Theorem 2.2, we may conclude that

$$T^*\left(\overline{x}\right) = T^*\left(x^*\right),$$

so by (53) there exists a subgradient u^* of g at x^* so that

$$\left[u^* + \nabla h\left(x^*\right)\right]^{\mathrm{T}}\left(x - x^*\right) \geq 0, \ \forall \ x \in \mathbf{X}.$$

Note that if h is convex, then the variational inequality above is, by the convexity of T, both necessary and sufficient for global optimality of x^* in (NP) [Ref. 1, Theorem 3.4.3]. This completes the proof. □

If g is the indicator function of a closed and convex set $\mathbf{S} \subset \Re^n$, it follows from the above proof that the algorithm is globally convergent also when h is only pseudoconvex.

In Theorem 4.1 we in fact proved a more general result than stated, i.e., from the proof it follows that any decrease in the subproblem objective function $T^{(l)}$ from the point $x^{(l)}$ is sufficient to yield a descent direction with respect to the original objective T. This implies, in turn, the possibility of truncating the process of solving (PL-SUB$^{(l)}$) before reaching its optimal solution, and introduces a trade-off between solving the subproblem accurately and obtaining sufficiently steep descent directions. In Ref. 11 the global convergence of such a scheme is given in the differentiable case, under the condition that at least one step of a descent method is performed on (PL-SUB$^{(l)}$) in each iteration l, and that this descent method has a closed algorithmic map. This result immediately carries over to the nondifferentiable case. [This way of obtaining search directions finitely by truncating the subproblem algorithm may also be seen as an alternative to the approximation scheme of Kiwiel (Ref. 18).]

The partial linearization method may be extended to inexact line searches also in the case of nondifferentiability by introducing a Lipschitz condition on the differentiable part of the objective, and by restricting $f(x,y)$ to be strictly convex with respect to x. The Armijo steplength rule is modified as follows. The point $x^{(l+1)}$ is obtained by finding the smallest integer $\bar{\imath}$ such that

$$T\left(x^{(l)}+\beta^{-i}d^{(l)}\right)-T\left(x^{(l)}\right)\leq-\varepsilon\beta^{-i}\left[\nabla_x f\left(\overline{x}^{(l)},x^{(l)}\right)-\nabla_x f\left(x^{(l)},x^{(l)}\right)\right]^{\mathrm{T}}d^{(l)}, \tag{55}$$

where $d^{(l)}=\overline{x}^{(l)}-x^{(l)}$.

Theorem 4.3 Under the additional assumptions that the function $h(x)$ has a Lipschitzian gradient, and that the function $f(x,y)$ is strictly convex with respect to x, the partial linearization algorithm, applied to (NP) using the modified Armijo steplength formula (55), is convergent in the same sense as in Theorem 4.2.

Proof The Lipschitz continuity assumption on ∇h implies, by (12), that for all $\alpha\geq 0$,

$$h\left(x^{(l)}+\alpha d^{(l)}\right)-h\left(x^{(l)}\right)\leq\alpha\nabla h\left(x^{(l)}\right)^{\mathrm{T}}d^{(l)}+\frac{1}{2}\alpha^2K\left\|d^{(l)}\right\|^2, \tag{56}$$

where K denotes the Lipschitz constant, and where $d^{(l)}=\overline{x}^{(l)}-x^{(l)}$. Also, for all $\alpha\in[0,1]$,

$$\begin{aligned} g\left(x^{(l)}+\alpha d^{(l)}\right)-g\left(x^{(l)}\right) &= \int_0^\alpha g'\left(x^{(l)}+sd^{(l)};d^{(l)}\right)ds \\ &\leq -\alpha g'\left(x^{(l)}+d^{(l)};-d^{(l)}\right) \\ &= \alpha \inf_{u\in\partial g\left(\overline{x}^{(l)}\right)} u^{\mathrm{T}}d^{(l)} \\ &\leq \alpha\overline{u}^{(l)\mathrm{T}}d^{(l)}, \end{aligned} \tag{57}$$

where the first equality follows from Corollary 24.2.1 of Ref. 19 and the first inequality from Theorem 24.1 of the same reference. Note that the subgradient $\overline{u}^{(l)}$ is arbitrary in $\partial g\left(\overline{x}^{(l)}\right)$, but will be chosen equal to a particular one, derived below.

From

$$\overline{x}^{(l)}\in\arg\min_{x\in\mathbf{X}}T^{(l)}(x),$$

we have by (52) that there exists a subgradient $\overline{u}^{(l)}$ of g at the point $\overline{x}^{(l)}$ so that

$$\left[\overline{u}^{(l)} + \nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) + \nabla h\left(x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}} d^{(l)} \leq 0, \tag{58}$$

which implies, by using (56) and (57), that

$$T\left(x^{(l)} + \alpha d^{(l)}\right) - T\left(x^{(l)}\right) \leq -\alpha \left[\nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}} d^{(l)} + \frac{1}{2}\alpha^2 K \left\|d^{(l)}\right\|^2. \tag{59}$$

By replacing $\nabla T\left(x^{(l)}\right)^{\mathrm{T}} d^{(l)}$ by $-\left[\nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}} d^{(l)}$ in the proof of Theorem 2.3, we may conclude that the modified Armijo line search (55) is finite, and that the series

$$\sum_{l=1}^{\infty} \left(\left[\nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}} d^{(l)}\right)^2$$

is convergent.

Hence,

$$\lim_{l\to\infty} \left[\nabla_x f\left(\overline{x}^{(l)}, x^{(l)}\right) - \nabla_x f\left(x^{(l)}, x^{(l)}\right)\right]^{\mathrm{T}} d^{(l)} = 0,$$

and we may choose subsequences of $\left\{x^{(l)}\right\}$ and $\left\{\overline{x}^{(l)}\right\}$ so that

$$\lim_{l\in L} x^{(l)} = x^* \text{ and } \lim_{l\in \hat{L}} \overline{x}^{(l)} = \overline{x},$$

and we thereby obtain that

$$\left[\nabla_x f\left(\overline{x}, x^*\right) - \nabla_x f\left(x^*, x^*\right)\right]^{\mathrm{T}} \left(\overline{x} - x^*\right) = 0.$$

By the strict convexity of $f(x,y)$ with respect to x, this implies that

$$x^* = \overline{x}.$$

Using this in the limit of (58) [cf. Equation (53)], there exists a subgradient u^* of $\partial g\left(x^*\right)$ such that

$$\left[u^* + \nabla h\left(x^*\right)\right]^{\mathrm{T}} \left(x - x^*\right) \geq 0, \ \forall\, x \in \mathbf{X},$$

which completes the proof. □

Note that convergence cannot be ensured with the modified Armijo line search (55) if $h(x) \equiv 0$, i.e., for totally nondifferentiable convex programs, while the convergence proof for the algorithm under exact line searches still applies. If on the other hand, $g(x) \equiv 0$, then (NP) reduces to (P), and the algorithms in this section reduces to those described in Section 2. It is interesting to note that the modified Armijo line search (55), which is dependent on the choice of the approximating function $f(x,y)$, generalizes the one used in Ref. 16. This method may be identified by choosing $f(x,y) = \frac{c}{2}\|x-y\|^2$, and the corresponding line search (55) would then be to choose the smallest integer $\bar{\imath}$ such that

$$T\left(x^{(l)} + \beta^{-i} d^{(l)}\right) - T\left(x^{(l)}\right) \leq -c\varepsilon\beta^{-i} \left\|d^{(l)}\right\|^2,$$

and, indeed, this is the line search employed in Ref. 16.

5. Discussion

In this paper we have introduced a classification of iterative algorithms for convex constrained programming based on partial linearizations of equivalent reformulations of the original objective function. This class of algorithms, which is referred to as partial linearization methods, includes the well known Frank and Wolfe method and constrained Newton-type methods, but also induces new algorithms. Under mild convexity assumptions the algorithm is shown to be globally convergent with or without exact line searches, and for both differentiable and nondifferentiable programming. The function f is arbitrary, and may therefore be chosen to utilize problem structure. An important example is for problems (P) with a nonseparable objective to be minimized over a Cartesian product of sets. Separability is then induced by choosing f separable with respect to this set, see Refs. 2, 10 and 11.

The Frank-Wolfe algorithm is known to have an arithmetic rate of convergence (Ref. 22), and constrained Newton methods have convergence rates ranging from linear to quadratic [Ref. 7, Section III.3]. Hence, instances of the partial linearization algorithm have rates of convergence ranging from arithmetic to quadratic. It would be of interest to further study the theoretical convergence rate of iterative methods in the framework of partial linearization algorithms, as well as studying practical choices of the function $f(x, y)$ for different types of problems.

One of the nice features of the algorithmic class is that efficient second order methods may be defined for nondifferentiable problems, by applying second order approximations to the differentiable part. Convex constrained problems with partially differentiable objectives arise when solving nonlinear continuous problems with Benders decomposition techniques (Refs. 23 and 24). Efficient methods for solving the Benders master problem can be studied in the context of partial linearization methods. Other applications are in dual approaches for large-scale linearly constrained convex programming, where the dual problem may be formulated as a convex and partially differentiable program, and for iterative use in exact penalty methods for differentiable convex programming.

The partial linearization algorithm has recently been extended to finite-dimensional variational inequalities, and shown to include well known methods such as projection type algorithms, linearization methods and diagonalization algorithms, disregarding the line search, see Refs. 3 and 11. Especially interesting is that partial linearization methods induce valid merit functions for asymmetric variational inequalities, and that, as a consequence, line searches with respect to these merit functions may be introduced in classical successive approximation methods to accelerate convergence. The recently presented descent method of Fukushima (Ref. 25) is an instance of this class of methods.

References

[1] Bazaraa, M.S., and Shetty, C.M., *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, New York, New York, 1979.

[2] Larsson, T., and Migdalas, A., *An Algorithm for Nonlinear Programs over Cartesian Product Sets*, Optimization, Vol. 21, pp. 535-542, 1990.

[3] Larsson, T., and Patriksson, M., *A Generalized Gap Function for Variational Inequalities*, Report LiTH-MAT-R-90-18, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden, 1990.

[4] Martos, B., *Nonlinear Programming Theory and Methods*, North-Holland, Amsterdam, Holland, 1975.

[5] Rockafellar, R.T., *The Theory of Subgradients and its Applications to Problems of Optimization: Convex and Nonconvex Functions*, Heldermann Verlag, Berlin, Germany, 1981.

[6] Armijo, L., *Minimization of Functions Having Lipschitz Continuous Partial Derivatives*, Pacific Journal of Mathematics, Vol. 16, pp. 1-3, 1964.

[7] Pshenichny, B.N., and Danilin, Yu. M., *Numerical Methods in Extremal Problems*, MIR Publishers, Moscow, Russia, 1978.

[8] Frank, M., and Wolfe, P., *An algorithm for Quadratic Programming*, Naval Research Logistics Quarterly, Vol. 3, pp. 95-110, 1956.

[9] Migdalas, A., *A Regularization of the Frank-Wolfe Algorithm*, Report LiTH-MAT-R-90-10, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden, 1990.

[10] Larsson, T., Migdalas, A., and Patriksson, M., *A Partial Linearization Method for the Traffic Assignment Problem*, Report LiTH-MAT-R-89-06, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden, 1989.

[11] Patriksson, M., *A Unified Description of Iterative Algorithms for Traffic Equilibria*, European Journal of Operational Research, forthcoming.

[12] Polyak, B.T., *Introduction to Optimization*, Optimization Software, New York, New York, 1987.

[13] Rockafellar, R.T., *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*, Mathematics of Operations Research, Vol. 1, pp. 97-116, 1976.

[14] Mine, H., and Fukushima, M., *A Minimization Method for the Sum of a Convex Function and a Continuously Differentiable Function*, Journal of Optimization Theory and Applications, Vol. 33, pp. 9-23, 1981.

[15] Clarke, F.H., *Generalized Gradients and Applications*, Transactions of the American Mathematical Society, Vol. 205, pp. 247-262, 1975.

[16] Fukushima, M., and Mine, H., *A Generalized Proximal Point Algorithm for Certain Non-Convex Minimization Problems*, International Journal of Systems Sciences, Vol. 12, pp. 989-1000, 1981.

[17] Fukushima, M., *A Successive Quadratic Programming Method for a Class of Constrained Nonsmooth Optimization Problems*, Mathematical Programming, Vol. 49, pp. 231-251, 1991.

[18] Kiwiel, K.C., *A Method for Minimizing the Sum of a Convex Function and a Continuously Differentiable Function*, Journal of Optimization Theory and Applications, Vol. 48, pp. 437-449, 1986.

[19] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[20] Rudin, W., *Principles of Mathematical Analysis, 3rd Edition*, McGraw-Hill, Singapore, Republic of Singapore, 1985.

[21] Shor, N.Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, Germany, 1985.

[22] Canon, M.D., and Cullum, C.D., *A Tight Upper Bound on the Rate of Convergence of the Frank-Wolfe Algorithm*, SIAM Journal on Control, Vol. 6, pp. 509-516, 1968.

[23] Florian, M., and Nguyen, S., *A Method for Computing Network Equilibrium with Elastic Demands*, Transportation Science, Vol. 8, pp. 321-332, 1974.

[24] Holmberg, K., and Jörnsten, K.O., *Cross Decomposition Applied to the Stochastic Transportation Problem*, European Journal of Operational Research, Vol. 17, pp. 361-368, 1984.

[25] Fukushima, M., *Equivalent Differentiable Optimization Problems and Descent Methods for Asymmetric Variational Inequality Problems*, Mathematical Programming, Vol. 53, pp. 99-110, 1992.