Comput Optim Appl (2010) 46: 333–346
DOI 10.1007/s10589-009-9267-0

A new line search inexact restoration approach for nonlinear programming

Andreas Fischer · Ana Friedlander

Received: 9 February 2009 / Published online: 24 June 2009

Abstract A new general scheme for Inexact Restoration methods for Nonlinear Programming is introduced. After computing an inexactly restored point, the new iterate is determined in an approximate tangent affine subspace by means of a simple line search on a penalty function. This differs from previous methods, in which the tangent phase needs both a line search based on the objective function (or its Lagrangian) and a confirmation based on a penalty function or a filter decision scheme. Besides its simplicity the new scheme enjoys some nice theoretical properties. In particular, a key condition for the inexact restoration step could be weakened. To some extent this also enables the application of the new scheme to mathematical programs with complementarity constraints.

Keywords Nonlinear programming · Inexact restoration · Line search · Penalty function · Complementarity constraints

1 Introduction

Let us consider the optimization problem

$$F(x) \to \min \quad \text{subject to } x \in G := \{x \in \Omega \mid H(x) = 0\} \tag{1}$$

This work was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.510/2006 - APQ1), FAPESP (Grants 2006/53768-0 and 2008/00875-9) and CNPq.

A. Fischer (✉)
Institute of Numerical Mathematics, Department of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany
e-mail: andreas.fischer@tu-dresden.de

A. Friedlander
Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas, SP, Brazil
e-mail: friedlan@ime.unicamp.br

with given functions $F : R^n \to \mathbb{R}$ and $H : \mathbb{R}^n \to \mathbb{R}^m$ and a given compact and convex set $\Omega \subset \mathbb{R}^n$. The functions F and H are assumed to be at least continuous on Ω. Additional smoothness assumptions will be made when needed.

Modern Inexact Restoration (IR) methods for Nonlinear Programming begin with the algorithm of Martínez and Pilotta [15]. The common features of this and other IR methods are the following:

(I) Given the current iterate $x^k \in \mathbb{R}^n$, an intermediate more feasible point y^k is computed using an arbitrary procedure which, in practice, is chosen according to the problem characteristics. This is the Restoration Phase of the method.
(II) A trial point z is computed on the "tangent set" that passes through y^k, in such a way that an optimality measure improves at z with respect to y^k.
(III) If the point z is acceptable for a criterion that combines feasibility and optimality, one defines the new iterate $x^{k+1} = z$. Otherwise, the trial point z is chosen in a smaller trust region around y^k.

The optimality improvement (II) involved in the choice of z can be done by a line search with respect to the objective function [9, 15] or the Lagrangian [14]. The acceptability of z in (III) depends, in [14, 15], on a function that combines feasibility and optimality. In [9] the acceptability of z was decided on the basis of a filter strategy. The filter strategy was developed and modified in further papers, see [11, 20, 21]. Assumptions for the local convergence of an IR method were analyzed by Birgin and Martínez [4]. Based on this and [14], Kaya and Martínez [12] suggested to apply an IR method to a class of discretized optimal control problems and provide results on local convergence properties of the method, see also Sect. 5 of the expanded version [13] of [12]. The use of an IR technique for solving bilevel problems was studied in [1]. An IR method in which inexact restoration appears as the natural generalization of the spectral projected gradient method [5–7] to minimization with nonlinear constraints was given in [8].

In the present paper we aim to simplify and extend the applicability of the IR approach in several ways. For this purpose we introduce an IR Model Algorithm where the new iterate (in an approximation of the tangent set) is obtained by means of a single line-search procedure that only involves a penalty function. Therefore, it is not necessary to begin with a line search involving optimality and completing the iteration with a criterion that combines optimality and feasibility.

Our main result proves that any sequence of search directions generated by the Model Algorithm tends to zero. With adequate choices of the search direction this implies a necessary optimality condition to hold at an accumulation point of the sequence $\{x^k\}$.

The Model Algorithm poses general conditions for the inexact restoration and for the choice of search directions. These conditions are weaker than existing ones and, in particular, allow to some extent the application of the Model Algorithm to problems with complementarity constraints.

Moreover, the sequences of penalty parameters and step lengths generated by the Model Algorithm are proved to be bounded away from zero. This simplifies the proofs and enables us to show that every accumulation point is stationary (and not just one of them as in [14, 15]).

2 The inexact restoration model algorithm

As tools for describing the Model Algorithm and its analysis we will make use of functions $h : \Omega \to [0, \infty)$ and $\Phi : \Omega \times [0, 1] \to \mathbb{R}$. The function h is assumed to be continuous on Ω. Moreover, h is required to be a majorant for $\|H\|$ on Ω, i.e.,

$$\|H(x)\| \le h(x) \quad \text{for all } x \in \Omega. \tag{2}$$

The advantage of using h instead of $\|H\|$ will become clear in Sect. 4 when we deal with complementarity constraints. The penalty function Φ is defined by

$$\Phi(x, p) := pF(x) + (1 - p)h(x) \quad \text{for all } (x, p) \in \Omega \times [0, 1].$$

We are now going to describe a simple frame for an inexact restoration algorithm. On the one hand, this frame allows the user to apply several concrete methods within both the restoration phase and the optimization phase. On the other hand, the frame is detailed enough for a concise global convergence analysis.

Model Algorithm

Let $r \in [0, 1)$ and $\beta, \gamma, \bar{\gamma}, \tau > 0$ be fixed.

Step 0: *Initialization.*
Choose $x^0 \in \Omega$ and $p_0 \in (0, 1)$. Set $k := 0$.

Step 1: *Inexact restoration.*
Compute $y^k \in \Omega$ so that

$$h(y^k) \le rh(x^k), \tag{3}$$

$$F(y^k) \le F(x^k) + \beta h(x^k). \tag{4}$$

Step 2: *Penalty parameter.*
Determine $p_{k+1} \in \{2^{-i} p_k \mid i \in \{0, 1, 2, \dots\}\}$ as large as possible so that

$$\Phi(y^k, p_{k+1}) - \Phi(x^k, p_{k+1}) \le \frac{1}{2}(1 - r)(h(y^k) - h(x^k)). \tag{5}$$

Step 3: *Search direction for optimization.*
Compute $d^k \in \mathbb{R}^n$ so that $y^k + d^k \in \Omega$ and

$$F(y^k + td^k) \le F(y^k) - \gamma t\|d^k\|^2, \tag{6}$$

$$h(y^k + td^k) \le h(y^k) + \bar{\gamma} t^2 \|d^k\|^2 \tag{7}$$

holds for all $t \in [0, \tau]$.

Step 4: *Line search.*
Determine $t_k \in \{2^{-i} \mid i \in \mathbb{N}\}$ as large as possible so that

$$\Phi(y^k + t_k d^k, p_{k+1}) - \Phi(x^k, p_{k+1}) \le \frac{1}{2}(1 - r)(h(y^k) - h(x^k)). \tag{8}$$

Step 5: *Update.*
Set $x^{k+1} := y^k + t_k d^k$ and $k := k + 1$. Go to Step 1.

The conditions on y^k in Step 1 are weaker than those used in existing IR methods, see Sect. 4 for a discussion. The remaining part of this section is devoted to the well-definedness of the Steps 3 and 1 of the Model Algorithm. For Steps 2 and 4 this question is answered by Lemma 3 in Sect. 3.

Remark 1 The conditions on d^k in Step 3 of the Model Algorithm can always be satisfied by setting $d^k := 0$. Obviously, this is not a useful choice in general. In Sect. 3 it will be shown that $\{d^k\}$ converges to zero if Step 1 is always well defined. Therefore, it has to be guaranteed that $\lim_{k\to\infty} d^k = 0$ has a reasonable meaning, i.e., that an accumulation point of the sequence $\{y^k\}$ converges to a stationary point of problem (1). Therefore, appropriate directions d^k which also satisfy (6) and (7) need to be used. Based on the following lemma, Remark 2 shows that this is possible at least if Ω is a polyhedron and h is given by $\|H\|$.

To proceed let us define the set (of tangent directions)

$$T(y) := \{d \in \mathbb{R}^n \mid y + d \in \Omega, \nabla H(y)^\top d = 0\} \quad \text{for all } y \in \Omega.$$

Obviously, $T(y)$ is always nonempty (since $0 \in T(y)$), convex, and closed. Moreover, for $y \in \Omega$, let

$$d(y) := \mathrm{Proj}_{T(y)}(-\nabla F(y))$$

denote the Euclidean projection of $-\nabla F(y)$ onto $T(y)$.

Lemma 1 *Suppose that F and H are differentiable on Ω and that ∇F and ∇H are Lipschitz continuous on Ω with modulus $L > 0$. Then, there are $\gamma, \bar{\gamma}, \tau > 0$ so that*

$$F(y + td(y)) \le F(y) - \gamma t\|d(y)\|^2 \tag{9}$$

and

$$\|H(y + td(y))\| \le \|H(y)\| + \bar{\gamma} t^2 \|d(y)\|^2 \tag{10}$$

hold for all $y \in \Omega$ and all $t \in [0, \tau]$.

Proof Let $y \in \Omega$ be arbitrarily chosen. Then, because $0 \in T(y)$, we get

$$\|\nabla F(y) + d(y)\|^2 = \| - \nabla F(y) - \mathrm{Proj}_{T(y)}(-\nabla F(y))\|^2 \le \|\nabla F(y)\|^2$$

and, thus,

$$\nabla F(y)^\top d(y) \le -\frac{1}{2}\|d(y)\|^2.$$

Using the Lipschitz continuity of ∇F on Ω, y, $y + d(y) \in \Omega$, and the convexity of Ω we obtain by Taylor's formula

$$F(y + td(y)) = F(y) + t\nabla F(y)^\top d(y) + t\int_0^1 (\nabla F(y + \sigma td(y)) - \nabla F(y))^\top d(y)\, \mathrm{d}\sigma$$

$$\leq F(y) - \frac{1}{2}t\|d(y)\|^2 + \frac{1}{2}t^2 L\|d(y)\|^2$$

for all $t \in [0, 1]$. Therefore, (9) is valid for all $t \in [0, \tau]$ with $\tau := \min\{1, \frac{1}{2L}\}$ and $\gamma := \frac{1}{4}$. Similarly, using $\nabla H(y)^\top d(y) = 0$ due to $d(y) \in T(y)$, we have

$$H(y + td(y)) = H(y) + t\int_0^1 (\nabla H(y + \sigma t d(y)) - \nabla H(y))^\top d(y)\, \mathrm{d}\sigma$$

and

$$\|H(y + td(y))\| \leq \|H(y)\| + \frac{1}{2}Lt^2\|d(y)\|^2$$

for all $t \in [0, 1]$. Thus, (10) holds for all $t \in [0, 1]$ with $\bar{\gamma} := \frac{1}{2}L$.

Hence, (9) and (10) are valid for all $y \in \Omega$ and all $t \in [0, \tau]$ with $\gamma, \bar{\gamma}, \tau$ as defined in this proof. □

Remark 2 Let Ω be defined by linear inequalities and equations. Assume that $h := \|H\|$ and d^k in Step 3 of the Model Algorithm is given by $d^k := d(y^k)$. Moreover, let us assume that

$$\lim_{k\to\infty} d^k = 0$$

is satisfied. Since $\{y^k\}$ lies in the compact set Ω a subsequence of $\{y^k\}$ converges to some $y^* \in \Omega$. According to Lemma 1, it follows that $y^* \in G$ (by (3)) and

$$\lim_{k\to\infty} \mathrm{Proj}_{T(y^k)}(-\nabla F(y^k)) = 0. \tag{11}$$

This implies that y^* is feasible and satisfies the AGP optimality condition associated to problem (1). AGP (Approximate Gradient Projection) is a strong *sequential* necessary optimality condition introduced by Martínez and Svaiter [16]. If a feasible point satisfies the AGP condition and the Mangasarian-Fromovitz constraint qualification (MFCQ), or even the weaker constant positive linear dependence (CPLD) constraint qualification from [18], then this point is a Karush-Kuhn-Tucker point, see Andreani, Martínez, and Schuverdt [3].

Hence, with Lemma 1, we have shown that Step 3 of the Model Algorithm can generate directions d^k for which $\lim_{k\to\infty} d^k = 0$ has a reasonable meaning in respect to an optimality condition.

Based on the next lemma, sufficient conditions under which Step 1 (inexact restoration) of the Model Algorithm is well defined will be provided in Remark 3.

Lemma 2 *Suppose that Ω is a bounded box, i.e., $\Omega = \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i$ for $i = 1, \dots, n\}$ with $-\infty < l_i < u_i < \infty$ for $i = 1, \dots, n$. Moreover, suppose that the feasible set G is nonempty, that H is continuously differentiable on Ω, and that MFCQ holds for all $x \in G$. Then, there is $\zeta > 0$ so that for each $x \in \Omega$ there is $y(x)$, so that*

$$y(x) \in G \tag{12}$$

and

$$\|x - y(x)\| \leq \zeta \|H(x)\|. \tag{13}$$

Proof Taking into account that G is compact and that the MFCQ holds for all points in G we can apply Corollary 1 in Robinson's paper [19]. This yields that there is $\zeta_0 > 0$ so that for each $z \in G$ there is some $\delta(z) > 0$ so that

$$\operatorname{dist}[x, G] \leq \zeta_0 \|H(x)\| \quad \text{for all } x \in \Omega \cap \mathcal{B}(z, \delta(z)),$$

where $\mathcal{B}(z, \delta(z))$ denotes the ball around z with radius $\delta(z)$. By the compactness of G we easily get that there is $\delta > 0$ so that

$$\operatorname{dist}[x, G] \leq \zeta_0 \|H(x)\| \quad \text{for all } x \in \Omega \cap (G + \delta\mathcal{B}) \tag{14}$$

with $\mathcal{B}$ as the unit ball in $\mathbb{R}^n$. Let us assume that there is $x \in \bar{\Omega} := \Omega \setminus (G + \delta\mathcal{B})$. Since G and $\bar{\Omega}$ are nonempty and compact and $\|H\|$ is continuous we have

$$c_1 := \sup_{x \in \bar{\Omega}} \operatorname{dist}[x, G] < \infty \quad \text{and} \quad c_2 := \inf_{x \in \bar{\Omega}} \|H(x)\| > 0.$$

Thus, it follows that

$$\operatorname{dist}[x, G] \leq c_1 = \frac{c_1}{c_2} c_2 \leq \frac{c_1}{c_2} \|H(x)\| \quad \text{for all } x \in \bar{\Omega}.$$

With $\zeta := \zeta_0 + \frac{c_1}{c_2}$ we obtain

$$\operatorname{dist}[x, G] \leq \zeta \|H(x)\| \quad \text{for all } x \in \Omega.$$

By (14), this is also valid if $\bar{\Omega} = \emptyset$.

Therefore and with the compactness of G we obtain for each $x \in \Omega$ that there is $y(x) \in G$ with $\|x - y(x)\| = \operatorname{dist}[x, G] \leq \zeta \|H(x)\|$, i.e., (12) and (13) are satisfied for all $x \in \Omega$. □

Remark 3 We are now going to explain that Lemma 2 provides sufficient conditions under which Step 1 of the Model Algorithm (inexact restoration) is well defined. To this end let h be given by $\|H\|$ and note that, given $x^k \in \Omega$, (12) implies

$$h(y(x^k)) = \|H(y(x^k))\| = 0 \leq r\|H(x^k)\| = rh(x^k)$$

for any $r \geq 0$. Furthermore, by the local Lipschitz continuity of F and (13) we have

$$F(y(x^k)) - F(x^k) \leq L_F \|y(x^k) - x^k\| \leq L_F \zeta \|H(x^k)\| = L_F \zeta h(x^k),$$

where $L_F > 0$ is some Lipschitz constant for F on (the compact set) Ω. Setting $y^k := y(x^k)$ and $\beta := L_F \zeta$ we finally see that y^k satisfies the conditions (3) and (4) within Step 1 of the Model Algorithm.

Note that the MFCQ is slightly weaker than the regularity condition (assumption A4) used by Martínez [14]. In contrast to Lemma 2 he shows that in a certain neighborhood of G a particular Newtonian choice of y^k satisfies conditions (3) and (4).

We finally refer the reader to [12, Lemma 4.1] where the Euler discretization of a certain optimal control problem has the property that (3) (with $h := H$) implies condition (16) (see Sect. 4) which, again, implies (4).

3 Convergence analysis

We will first prove that the Model Algorithm is well defined if its Step 1 is well defined. In particular, the penalty parameters p_k and the step lengths t_k are proved to remain bounded away from zero. Based on this, we will show in Theorem 1 that the sequences $\{h(x^k)\}$ and $\{H(x^k)\}$ tend to zero. Finally, Theorem 2 presents our main result that $\{d^k\}$ converges to zero.

Lemma 3 *Suppose that Step 1 of the Model Algorithm is well defined. Then, the Model Algorithm is well defined. Moreover, there are $k_0 \in \mathbb{N}$ and $\bar{t} \in \mathbb{R}$ so that*

$$
\begin{aligned}
&p_k = p_{k_0} > 0 \quad \textit{for all } k \geq k_0,\\
&t_k \geq \bar{t} > 0 \quad \textit{for all } k \in \mathbb{N}.
\end{aligned}
$$

Proof Let x^k and p_k be generated by the Model Algorithm. Then, using (3) and (4) we obtain, for any $p \in (0, 1)$,

$$
\begin{aligned}
\Phi(y^k, p) - \Phi(x^k, p) &= p(F(y^k) - F(x^k)) + (1-p)(h(y^k) - h(x^k))\\
&\leq p\beta h(x^k) - (1-p)(1-r)h(x^k)\\
&= h(x^k)(p(\beta + 1 - r) - (1-r)).
\end{aligned}
$$

Therefore, if

$$
0 \leq p \leq \tilde{p} := \frac{1-r}{2(\beta + 1 - r)}
$$

it follows that $p(\beta + 1 - r) - (1-r) \leq -\frac{1}{2}(1-r)$ and

$$
\Phi(y^k, p) - \Phi(x^k, p) \leq -\frac{1}{2}(1-r)h(x^k) \leq \frac{1}{2}(1-r)(h(y^k) - h(x^k)).
$$

Hence, Step 2 of the Model Algorithm is well defined and generates some p_{k+1}. More in detail, due to the rule of finding p_{k+1} as the largest value in $\{2^{-i} p_k \mid i \in \mathbb{N}\}$ which satisfies (5), we have

$$
p_{k+1} \geq \bar{p} := \min\left\{p_0, \frac{1}{2}\tilde{p}\right\} > 0 \quad \text{for all } k \in \mathbb{N}. \tag{15}
$$

Thus, since either $p_{k+1} = p_k$ or $p_{k+1} \leq \frac{1}{2} p_k$ for all $k \in \mathbb{N}$, there is $k_0 \in \mathbb{N}$ so that $p_k = p_{k_0} \geq \bar{p}$ for all $k \geq k_0$.

Obviously, Step 3 of the Model Algorithm is well defined, see Remarks 1 and 2. To show that Step 4 is also well defined we first note that, by (5), (6), (7), and (15),

$$\begin{aligned}
&\Phi(y^k + td^k, p_{k+1}) - \Phi(x^k, p_{k+1}) \\
&\quad = \Phi(y^k + td^k, p_{k+1}) - \Phi(y^k, p_{k+1}) + \Phi(y^k, p_{k+1}) - \Phi(x^k, p_{k+1}) \\
&\quad \le p_{k+1}(F(y^k + td^k) - F(y^k)) + (1 - p_{k+1})(h(y^k + td^k) - h(y^k)) \\
&\qquad + \frac{1}{2}(1 - r)(h(y^k) - h(x^k)) \\
&\quad \le -\bar{p}\gamma t\|d^k\|^2 + \bar{\gamma} t^2\|d^k\|^2 + \frac{1}{2}(1 - r)(h(y^k) - h(x^k)) \\
&\quad = t\|d^k\|^2(-\bar{p}\gamma + t\bar{\gamma}) + \frac{1}{2}(1 - r)(h(y^k) - h(x^k))
\end{aligned}$$

holds for all $t \in [0, \tau]$. Thus, if

$$0 \le t \le \tilde{t} := \min\{\tau, \bar{p}\gamma\bar{\gamma}^{-1}\}$$

we have $-\bar{p}\gamma + t\bar{\gamma} \le 0$ and

$$\Phi(y^k + td^k, p_{k+1}) - \Phi(x^k, p_{k+1}) \le \frac{1}{2}(1 - r)(h(y^k) - h(x^k)).$$

Hence, Step 4 of the Model Algorithm is well defined and provides t_k with

$$t_k \ge \bar{t} := \frac{1}{2}\tilde{t}$$

for all $k \in \mathbb{N}$. □

Remark 4 For previous IR methods having the same penalty function the boundedness of the penalty parameters away from zero is not known. This indicates a possible advantage of the Model Algorithm in respect to the speed of convergence.

Theorem 1 *Suppose that Step* 1 *of the Model Algorithm is well defined. Then, for any sequence* $\{x^k\}$ *generated by the Model Algorithm, there is* $\hat{\sigma} \ge 0$ *so that*

$$\sum_{k=0}^{\infty} h(x^k) = \hat{\sigma}.$$

Moreover,

$$\lim_{k\to\infty} h(x^k) = 0 \quad \text{and} \quad \lim_{k\to\infty} H(x^k) = 0$$

are valid.

Proof According to Lemma 3 the Model Algorithm generates infinite sequences $\{x^k\}$, $\{y^k\}$, $\{d^k\}$, $\{p_k\}$, and $\{t_k\}$. In particular, $p_k = p_{k_0} > 0$ holds for all $k \ge k_0$.

Therefore, exploiting (8) and (3), we obtain

$$\Phi(x^{k+1}, p_{k_0}) - \Phi(x^k, p_{k_0}) \le \frac{1}{2}(1-r)(h(y^k) - h(x^k)) \le -\frac{1}{2}(1-r)^2 h(x^k)$$

for all $k \ge k_0$. It follows immediately that, for any $\ell \in \mathbb{N}$ with $\ell > k_0$,

$$\begin{aligned}\Phi(x^\ell, p_{k_0}) - \Phi(x^{k_0}, p_{k_0}) &= \sum_{k=k_0}^{\ell-1} (\Phi(x^{k+1}, p_{k_0}) - \Phi(x^k, p_{k_0})) \\ &\le -\frac{1}{2}(1-r)^2 \sum_{k=k_0}^{\ell-1} h(x^k).\end{aligned}$$

Taking into account that the function $\Phi(\cdot, p_{k_0})$ is continuous on the compact set Ω, that $\{x^k\} \subset \Omega$, and that $\{h(x^k)\} \subset [0, \infty)$, we conclude that the sequence $\{\sigma_\ell\}$ given by

$$\sigma_\ell := \sum_{k=0}^{\ell} h(x^k) \quad \text{for } \ell \in \mathbb{N}$$

is monotonically increasing and bounded and, in turn, converges to some $\hat{\sigma} \ge 0$. Obviously, this implies $\lim_{k\to\infty} h(x^k) = 0$ and, due to (2), also $\lim_{k\to\infty} H(x^k) = 0$. □

Theorem 2 *Suppose that Step* 1 *of the Model Algorithm is well defined. Then,*

$$\lim_{k\to\infty} d^k = 0.$$

Proof Using (6) and (4), we obtain that, for all $k \in \mathbb{N}$,

$$\begin{aligned}F(x^{k+1}) - F(x^k) &= F(y^k + t_k d^k) - F(y^k) + F(y^k) - F(x^k) \\ &\le -\gamma t_k \|d^k\|^2 + \beta h(x^k).\end{aligned}$$

Hence, by Lemma 3, it follows that

$$F(x^{\ell+1}) - F(x^0) = \sum_{k=0}^{\ell} (F(x^{k+1}) - F(x^k)) \le -\gamma \bar{t} \sum_{k=0}^{\ell} \|d^k\|^2 + \beta \sum_{k=0}^{\ell} h(x^k)$$

is valid for all $\ell \in \mathbb{N}$. Theorem 1 and $\{h(x^k)\} \subset [0, \infty)$ now provide

$$F(x^{\ell+1}) - F(x^0) \le -\gamma \bar{t} \sum_{k=0}^{\ell} \|d^k\|^2 + \beta \hat{\sigma} \quad \text{for all } \ell \in \mathbb{N}.$$

Because F is continuous on the compact set Ω and $\{x^k\} \subset \Omega$ we have $\lim_{k\to\infty} d^k = 0$. □

4 Dealing with complementarity constraints

If complementarity constraints are present then the condition

$$\|y^k - x^k\| \le \tilde{\beta}\|H(x^k)\| \quad \text{for all } k \in \mathbb{N} \tag{16}$$

(with some $\tilde{\beta} > 0$) as used in previous IR algorithms [1, 14, 15] for the inexact restoration step becomes critical. Instead of requiring (16) the Model Algorithm assumes (4) to hold. If F is Lipschitz-continuous on the compact set Ω with modulus $L_F > 0$ condition (4) is implied by (16). More in detail, we have

$$F(y^k) - F(x^k) \le L_F\|y^k - x^k\| \le L_F\tilde{\beta}\|H(x^k)\| \le \beta h(x^k),$$

where (2) has been taken into account. Thus, condition (4) is weaker than (16). In what follows we will describe two cases to exploit this. The first case is treated in Sect. 4.1 and deals with a Mathematical Program with Complementarity Constraints (MPCC) whose complementarity variables do not appear in the objective function. In Sect. 4.2 we suggest a special choice of the majorant function h so that condition (4) can be satisfied in the presence of complementarity constraints. Although the ideas also work for more general settings the presentation in both cases is for simplicity based on the following MPCC

$$F(u, v) \to \min \quad \text{subject to } (u, v, w) \in G,$$

where G contains exactly those points $(u, v, w) \in \Omega$ which satisfy

$$H(u, v, w) := \begin{pmatrix} g(u, v) - w \\ v_1 w_1 \\ \vdots \\ v_m w_m \end{pmatrix} = 0. \tag{17}$$

The functions $F : \mathbb{R}^{p+m} \to \mathbb{R}$ and $g : \mathbb{R}^{p+m} \to \mathbb{R}^m$ are assumed to be sufficiently smooth. The set $\Omega \subset \mathbb{R}^{p+m+m}$ is a compact box, where the variables v_i and w_i have the lower bound 0 for $i = 1, \dots, m$. Thus, x in problem (1) can be identified with (u, v, w), where $n = p + 2m$.

4.1 MPCCs with a particular objective

Let us assume here that F depends on $u \in \mathbb{R}^p$ only. For example, such situations can be encountered for certain discretized optimal control problems where the objective only depends on the control. Optimal control problems of this kind can be found in [10, 17].

Let us now consider Step 1 of our Model Algorithm with $x^k = (u^k, v^k, w^k)$ given. If the inexact restoration is done in such a way that u^k remains fixed and only (v^k, w^k) can be modified so that $y^k = (u^k, \hat{v}^k, \hat{w}^k)$ and

$$h(y^k) \le r h(x^k)$$

it follows that condition (4) is obviously satisfied since

$$F(y^k) = F(u^k) = F(x^k) \leq F(x^k) + \beta h(x^k).$$

In contrast to this, due to $y^k - x^k = (0, \hat{v}^k - v^k, \hat{w}^k - w^k)$, condition (16) is not satisfied in general. It might be interesting to identify other situations where condition (4) holds but (16) does not.

4.2 The use of a special majorant function h

Before defining the continuous function $h : \Omega \to [0, \infty)$ with the property (2) we will provide an auxiliary result.

Lemma 4 *For some $i \in \{1, \ldots, m\}$, let $q : \mathbb{R}^{p+m+m} \to \mathbb{R}$ be defined by*

$$q(u, v, w) := v_i w_i \quad \text{for all } (u, v, w) \in \mathbb{R}^{p+m+m}.$$

If, for some $(u, v, w) \in \mathbb{R}^{p+m+m}$ and some $d = (d_u, d_v, d_w) \in \mathbb{R}^{p+m+m}$,

$$v_i w_i \geq 0, \qquad (v_i, w_i) \neq (0, 0), \qquad \nabla q(u, v, w)^\top d = 0$$

hold, then

$$q(u + td_u, v + td_v, w + td_w) = (v + td_v)_i (w + td_w)_i \leq v_i w_i = q(u, v, w)$$

is valid for all $t \in \mathbb{R}$.

Proof Let us consider the case that $v_i \neq 0$. Then, since $\nabla q(u, v, w)^\top d = 0$ is the same as

$$w_i (d_v)_i + v_i (d_w)_i = 0,$$

we have

$$(d_w)_i = -(v_i)^{-1} w_i (d_v)_i.$$

With both equations and $v_i w_i \geq 0$ it follows that

$$\begin{aligned}
(v + td_v)_i (w + td_w)_i &= v_i w_i + t(v_i (d_w)_i + w_i (d_v)_i) + t^2 (d_v)_i (d_w)_i \\
&= v_i w_i - t^2 (v_i)^{-1} w_i (d_v)_i (d_v)_i \\
&\leq v_i w_i.
\end{aligned}$$

In analogy to this, the same inequality is obtained for the case that $w_i \neq 0$. Thus, by definition of q, the assertion of the lemma follows. □

Now, let us define the function $h : \Omega \to [0, \infty)$ by

$$h(u, v, w) := \|H(u, v, w)\| + \sqrt{v^\top w} \quad \text{for all } (u, v, w) \in \Omega. \tag{18}$$

Note that $(u, v, w) \in \Omega$ implies $u, v \in \mathbb{R}^n_+$. Obviously, (2) is valid.

Let $x^k = (u^k, v^k, w^k)$ and $y^k = (\hat{u}^k, \hat{v}^k, \hat{w}^k)$ denote the input and the output of Step 1 of the Model Algorithm. The use of h as defined by (18) has the advantage that condition (4) is now significantly weaker compared to the case when $h = \|H\|$ is employed. Condition (4) with h given by (18) reads as

$$F(y^k) - F(x^k) \le \beta \left(\|H(x^k)\| + \sqrt{(v^k)^\top w^k} \right), \tag{19}$$

where H is given by (17). In the presence of complementarity constraints it seems appropriate to require that

$$\|y^k - x^k\| \le \bar{\beta} \left(\|H(x^k)\| + \sqrt{(v^k)^\top w^k} \right) \tag{20}$$

holds for all k with some $\bar{\beta} > 0$. Note that

$$0 \le \sum_{i=1}^m \min\{v_i, w_i\} \le \sqrt{v^\top w} \quad \text{for all } v, w \ge 0.$$

If F is Lipschitz continuous on the compact set Ω with modulus $L_F > 0$ we have

$$\|F(y^k) - F(x^k)\| \le L_F \|y^k - x^k\|$$

so that (20) implies (19). Moreover, it would be possible to replace $\sqrt{v^\top w}$ within (18) by $(v^\top w)^\alpha$ with some $\alpha > 0$ to further weaken (19) and (20).

It seems that replacing $\|H\|$ by some h so that (4) becomes significantly weaker is quite simple. However, the difficulty with replacing H by h are linked to condition (7) in Step 3 of the Model Algorithm. Fortunately, for the definition of h given by (18) it is possible to show that (7) can be satisfied, at least by choosing $d^k := d(y^k)$ (see Sect. 2 for $d(y)$). According to Lemma 1 and Lemma 4 this choice leads to

$$\begin{aligned} h(y^k + td^k) &= \|H(y^k + td^k)\| + \sqrt{(v^k + td_v^k)^\top (w^k + td_w^k)} \\ &\le \|H(y^k)\| + \bar{\gamma} t^2 \|d^k\|^2 + \sqrt{(v^k)^\top w^k} \\ &\le h(y^k) + \bar{\gamma} t^2 \|d^k\|^2 \end{aligned}$$

for all $t \in [0, \tau]$. Hence, condition (7) is satisfied and the convergence theory for the Model Algorithm is applicable.

5 Final remarks

The philosophy of Inexact Restoration relies on recognizing a few obvious and practical facts of Nonlinear Programming. The first is that feasibility is generally more important than optimality and, so, deserves to be treated independently, using its improvement as an intermediate objective to be verified at each iteration. Nevertheless, being rigorous with feasibility from the beginning may lead to slow convergence when high nonlinearity of the constraints is present or when restoring feasibility is computationally costly. Finally, many problem structures suggest their own way for keeping or restoring feasibility and this possible advantage should be exploited in algorithms.

Our main objective here has been to provide a simple framework for defining Inexact Restoration algorithms and proving their convergence. The Model Algorithm in Sect. 2 has at least as good theoretical properties (regarding global convergence) as those that are known for the IR Algorithms defined in [14, 15]. In addition, the Model Algorithm provides that *any* accumulation point of the sequence $\{y^k\}$ is an AGP point. To have the same for the sequence $\{x^k\}$ one could add an additional but very weak condition in Step 1, e.g. $\|x^k - y^k\| \le \tilde{\beta} h(x^k)^{1/4}$. Moreover, the boundedness of the penalty parameters and the step lengths away from zero could be shown. A number of research topics remains. They include the efficient choice of directions in Step 3 and local superlinear convergence.

The application of our algorithm to Mathematical Programs with Complementarity Constraints (or bilevel programming) is appealing. We gave some hints on this application in Sect. 4. In this respect, there are several open questions. In particular, what is the meaning of the AGP optimality condition in the presence of complementarity constraints. Note that Andreani and Martínez [2] analyzed this question successfully under a strict complementarity condition. Moreover, it seems interesting to identify in detail those classes of bilevel programs or MPCCs that can be treated by the ways suggested in Sects. 4.1 and 4.2. This is going to be a subject of future research.

References

1. Andreani, R., Castro, S.L.C., Chela, J., Friedlander, A., Santos, S.A.: An inexact-restoration method for nonlinear bilevel programming problems. Comput. Optim. Appl. (2007). doi:10.1007/s10589-007-9147-4
2. Andreani, R., Martínez, J.M.: On the solution of mathematical programming problems with equilibrium constraints. Math. Methods Oper. Res. **54**, 345–358 (2001)
3. Andreani, A., Martínez, J.M., Schuverdt, M.L.: On the relation between constant positive linear dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**, 473–485 (2005)
4. Birgin, E.G., Martínez, J.M.: Local convergence of an inexact-restoration method and numerical experiments. J. Optim. Theory Appl. **127**, 229–247 (2005)
5. Birgin, E.G., Martínez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM J. Optim. **10**, 1196–1211 (2000)
6. Birgin, E.G., Martínez, J.M., Raydan, M.: Algorithm 813: SPG—software for convex-constrained optimization. ACM Trans. Math. Softw. **27**, 340–349 (2001)
7. Birgin, E.G., Martinez, J.M., Raydan, M.: Inexact spectral projected gradient methods on convex sets. IMA J. Numer. Anal. **23**, 539–559 (2003)
8. Gomes-Ruggiero, M.A., Martínez, J.M., Santos, S.A.: Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. SIAM J. Sci. Comput. **31**, 1628–1652 (2009)
9. Gonzaga, C.C., Karas, E.W., Vanti, M.: A globally convergent filter method for nonlinear programming. SIAM J. Optim. **14**, 646–669 (2003)
10. Haslinger, J., Neittaanmäki, P.: Finite Element Approximation for Optimal Shape, Material and Topology Design, 2nd edn. Wiley, Chichester (1996)
11. Karas, E.W., Oening, A.P., Ribeiro, A.A.: Global convergence of slanting filter methods for nonlinear programming. Appl. Math. Comput. **200**, 486–500 (2008)

12. Kaya, C.Y., Martínez, J.M.: Euler discretization and Inexact restoration for optimal control. J. Optim. Theory Appl. **134**, 191–206 (2007)
13. Kaya, C.Y., Martínez, J.M.: Euler discretization and inexact restoration for optimal control. Technical Report (2007). http://www.maths.unisa.edu.au/~yalcin/techreports/iroc.pdf. Accessed 31 May 2009
14. Martínez, J.M.: Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. J. Optim. Theory Appl. **111**, 39–58 (2001)
15. Martínez, J.M., Pilotta, E.A.: Inexact restoration algorithms for constrained optimization. J. Optim. Theory Appl. **104**, 135–163 (2000)
16. Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. J. Optim. Theory Appl. **118**, 117–133 (2003)
17. Outrata, J., Kocvara, M., Zowe, J.: Nonsmooth Approach to Optimization Problems with Equilibrium Constraints. Kluwer, Dordrecht (1998)
18. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
19. Robinson, S.M.: Stability theory for systems of inequalities, part II: differentiable nonlinear systems. SIAM J. Numer. Anal. **13**, 497–513 (1976)
20. Silva, C.E.P., Monteiro, M.T.T.: A filter inexact-restoration method for nonlinear programming. TOP **16**, 126–146 (2008)
21. Silva, C.E.P., Monteiro, M.T.T.: A filter algorithm: comparison with NLP solvers. Int. J. Comput. Math. **85**, 667–689 (2008)