Computational Optimization and Applications, , 1–13 ()

A Linearly Convergent Derivative-Free Descent Method for Strongly Monotone Complementarity Problems *

O.L. MANGASARIAN olvi@cs.wisc.edu
Computer Sciences Department,
University of Wisconsin,
1210 West Dayton Street,
Madison, WI 53706, U.S.A.

M.V. SOLODOV solodov@impa.br
Instituto de Matemática Pura e Aplicada,
Estrada Dona Castorina 110, Jardim Botânico,
Rio de Janeiro, RJ 22460-320, Brazil

Editor:

Abstract. We establish the first rate of convergence result for the class of derivative-free descent methods for solving complementarity problems. The algorithm considered here is based on the implicit Lagrangian reformulation [26, 35] of the nonlinear complementarity problem, and makes use of the descent direction proposed in [42], but employs a different Armijo-type linesearch rule. We show that in the strongly monotone case, the iterates generated by the method converge globally at a linear rate to the solution of the problem.

Keywords: Complementarity problems, implicit Lagrangian, descent algorithms, derivative-free methods, linear convergence

1. Introduction

The classical nonlinear complementarity problem [31, 6], NCP(F), is to find a point $x \in \Re^n$ such that

$$x \geq 0, \quad F(x) \geq 0, \quad \langle x, F(x) \rangle = 0, \tag{1}$$

where $F : \Re^n \to \Re^n$, and $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\Re^n$. Throughout this paper, we shall assume that $F(\cdot)$ is continuously differentiable. In the case when $F(\cdot)$ is affine, NCP(F) reduces to the linear complementarity problem [1].

Among popular approaches to solving NCP(F) which recently attracted attention, are derivative-free methods based on minimizing appropriate unconstrained or constrained minimization reformulations of the original problem. The literature on derivative-free algorithms is vast; see [11, 42, 15, 24, 14, 20, 34, 43]. To our knowledge, no rate of convergence results have been previously established for any of the

* Research of the first author is supported in part by NSF Grant CCR-9322479 and AFOSR Grant F49620-97-1-0326. Research of the second author is supported by CNPq Grant 300734/95-6 and by PRONEX–Optimization.

methods in this class. In this paper, we describe a derivative-free descent algorithm for minimizing the implicit Lagrangian [26, 35] merit function, and establish its linear convergence in the case when $F(\cdot)$ is strongly monotone.

It is known that NCP(F) can be cast as a problem of minimizing the following *implicit Lagrangian* function

$$M_\alpha(x) := \langle x, F(x)\rangle + \frac{1}{2\alpha}\left(\|(x-\alpha F(x))_+\|^2 - \|x\|^2 + \|(F(x)-\alpha x)_+\|^2 - \|F(x)\|^2\right), \tag{2}$$

where $\alpha > 1$ is a parameter and $(\cdot)_+$ denotes the orthogonal projection map onto the nonnegative orthant $\Re^n_+$. The implicit Lagrangian was introduced in [26], and further studied and extended in [23, 42, 18, 15, 33, 41, 43, 4, 36]; see [35] for a survey.

Let $r(\cdot)$ be the *natural residual* for NCP(F) [29], that is

$$r(x) := x - (x - F(x))_+ = \min\{x, F(x)\},$$

where the minimum is taken component-wise. It is well known that $r(x) = 0$ if and only if x is a solution of NCP(F). Below we summarize some of the properties of the implicit Lagrangian that will be used in the sequel.

THEOREM 1 *(See [26, 23]) Let $\alpha > 1$. Then the following statements hold :*

1. $M_\alpha(x) \geq 0$ *for all* $x \in \Re^n$.
2. $x \in \Re^n$ *solves NCP(F) if and only if* $M_\alpha(x) = 0$.
3. $\alpha^{-1}(\alpha-1)\|r(x)\|^2 \leq M_\alpha(x) \leq (\alpha-1)\|r(x)\|^2$ *for all* $x \in \Re^n$.

In particular, $M_\alpha(\cdot)$ is nonnegative on $\Re^n$ and assumes the value of zero precisely at the solutions of NCP(F). Thus one way to solve NCP(F) is via solving the unconstrained minimization problem

$$\min_{x\in\Re^n} M_\alpha(x).$$

Note that $M_\alpha(\cdot)$ is continuously differentiable whenever $F(\cdot)$ is continuously differentiable.

In [42], it was shown that when $F(\cdot)$ is strongly monotone and continuously differentiable, that is for some $\mu > 0$

$$\langle F(x) - F(y), x - y\rangle \geq \mu\|x-y\|^2 \quad \forall\, x, y \in \Re^n,$$

or, equivalently,

$$\langle \nabla F(x) y, y\rangle \geq \mu\|y\|^2,$$

then the direction

$$d(x) := (\beta - \alpha)\,(x - (x - \alpha F(x))_+) + (1 - \alpha\beta)\,(F(x) - (F(x) - \alpha x)_+) \tag{3}$$

is a descent direction for $M_\alpha(\cdot)$ at $x \in \Re^n$, provided $\beta > 0$ is sufficiently small and x is not the solution of NCP(F). A descent method was proposed which uses an Armijo-type search along this direction, and it was shown that this method converges to the solution of NCP(F). An attractive feature of this algorithm is that it is derivative-free, i.e. no derivatives of $F(\cdot)$ need to be computed. This makes the method suitable for large-scale problems, as well as applications where the derivatives of $F(\cdot)$ are not available or are costly to compute.

In this paper, we establish linear rate of convergence for a method based on the direction given by (3) but with a stepsize rule somewhat different from that in [42]. Note that this result is not obvious because the implicit Lagrangian $M_\alpha(\cdot)$ is not known to be strongly convex (even locally, in the neighborhood of the solution). Even convexity can be established only for the strongly monotone *linear* complementarity problems and under certain restrictions on the parameter α [33, 21]. Moreover, our algorithm does not make use of the gradient of $M_\alpha(\cdot)$.

We briefly mention some other approaches to solving NCPs. Various unconstrained reformulations based on appropriate merit functions have been investigated in [18, 14, 24, 21, 27, 40, 41, 5]. Nonnegatively constrained reformulations are considered in [11, 26, 39, 9, 37, 10, 34]. A recent survey of merit functions for NCPs and related issues can be found in [12]. Among other approaches we mention nonsmooth and semismooth methods [7, 8, 22, 3, 32, 16, 2, 19] and smooth equation-based methods [25, 38, 17].

The following *error bound* result is central in the convergence rate analysis.

THEOREM 2 *(See [30]) Let $F(\cdot)$ be strongly monotone with modulus $\mu > 0$ and Lipschitz continuous with constant $L > 0$ on some set X containing x^*, the solution of NCP(F). Then*

$$\|x - x^*\| \leq \frac{L+1}{\mu}\|r(x)\| \quad \forall x \in X.$$

A few words about our notation. For a real-valued matrix A of any dimension, $A^\top$ denotes its transpose. For a differentiable function $f : \Re^n \to \Re$, ∇f will denote the n-dimensional vector of partial derivatives with respect to x. For a differentiable vector function $F : \Re^n \to \Re^n$, ∇F will denote the $n \times n$ Jacobian matrix whose rows are the gradients of the components of F. By Q-linear and R-linear convergence we mean linear convergence in the quotient sense and in the root sense, respectively, as defined in [28].

2. Descent Method and Its Convergence

First, let us re-write the implicit Lagrangian function in the following form :

$$M_\alpha(x) = \sum_{i=1}^{n} \psi_\alpha(x_i, F_i(x)), \tag{4}$$

where $\psi_\alpha : \Re^2 \to \Re$ is given by

$$\psi_\alpha(a,b) := ab + \frac{1}{2\alpha}\left((a-\alpha b)_+^2 - a^2 + (b-\alpha a)_+^2 - b^2\right). \tag{5}$$

It is easy to verify that

$$\frac{\partial\psi_\alpha}{\partial a}(a,b) = b + \frac{1}{\alpha}\left((a-\alpha b)_+ - a - \alpha(b-\alpha a)_+\right)$$

and

$$\frac{\partial\psi_\alpha}{\partial b}(a,b) = a + \frac{1}{\alpha}\left(-\alpha(a-\alpha b)_+ + (b-\alpha a)_+ - b\right).$$

We will adopt the following notation :

$$\frac{\partial\psi_\alpha}{\partial a}(x,F(x)) := \begin{pmatrix} \frac{\partial\psi_\alpha}{\partial a}(x_1,F_1(x)) \\ \vdots \\ \frac{\partial\psi_\alpha}{\partial a}(x_n,F_n(x)) \end{pmatrix}, \quad \frac{\partial\psi_\alpha}{\partial b}(x,F(x)) := \begin{pmatrix} \frac{\partial\psi_\alpha}{\partial b}(x_1,F_1(x)) \\ \vdots \\ \frac{\partial\psi_\alpha}{\partial b}(x_n,F_n(x)) \end{pmatrix}.$$

With this notation, the gradient of $M_\alpha(\cdot)$ is given by

$$\nabla M_\alpha(x) = \frac{\partial\psi_\alpha}{\partial a}(x,F(x)) + \nabla F(x)^\top \frac{\partial\psi_\alpha}{\partial b}(x,F(x)).$$

We start with some preliminary results. The first assertion of Lemma 1 has been proven in [18, 42] but we include its proof for completeness. Note that the second assertion is new, and it is the key for obtaining the rate of convergence result.

LEMMA 1 *Let $\alpha > 1$. The following properties hold for all $x \in \Re^n$:*

1. $\langle \frac{\partial\psi_\alpha}{\partial a}(x,F(x)), \frac{\partial\psi_\alpha}{\partial b}(x,F(x))\rangle \geq 0.$
2. $\|\frac{\partial\psi_\alpha}{\partial a}(x,F(x)) + \frac{\partial\psi_\alpha}{\partial b}(x,F(x))\| \geq \alpha^{-1}(\alpha-1)\|r(x)\|.$

Proof: We establish the above inequalities component-wise. Let $a := x_i$ and $b := F_i(x)$, where $i \in \{1,\dots,n\}$. Note that

$$\frac{\partial\psi_\alpha}{\partial a}(a,b) + \frac{\partial\psi_\alpha}{\partial a}(a,b) = (1-\frac{1}{\alpha})\left(a - (a-\alpha b)_+ + b - (b-\alpha a)_+\right).$$

We consider the following four possible cases.
Case 1 : $a - \alpha b \geq 0$ and $b - \alpha a \geq 0$.
In that case,

$$\frac{\partial\psi_\alpha}{\partial a}(a,b) = \alpha a - b \leq 0 \quad \text{and} \quad \frac{\partial\psi_\alpha}{\partial b}(a,b) = \alpha b - a \leq 0.$$

Hence

$$\frac{\partial\psi_\alpha}{\partial a}(a,b)\frac{\partial\psi_\alpha}{\partial b}(a,b) \geq 0.$$

Since $a \geq \alpha b \geq \alpha^2 a$ and $\alpha > 1$, it follows that $a \leq 0$ and $b \leq 0$. Therefore

$$\begin{aligned}
\left(\frac{\partial\psi_\alpha}{\partial a}(a,b) + \frac{\partial\psi_\alpha}{\partial b}(a,b)\right)^2 &= ((\alpha-1)(a+b))^2 \\
&\geq (\alpha-1)^2 b^2 \\
&= \left(\frac{\alpha-1}{\alpha}\right)^2 (a-(a-\alpha b)_+)^2 \\
&\geq \left(\frac{\alpha-1}{\alpha}\right)^2 (a-(a-b)_+)^2,
\end{aligned}$$

where the last inequality follows from [13, Lemma 1] and the fact that $\alpha > 1$.

Case 2 : $a - \alpha b \geq 0$ and $b - \alpha a < 0$.
In that case,

$$\frac{\partial\psi_\alpha}{\partial a}(a,b) = 0 \quad \text{and} \quad \frac{\partial\psi_\alpha}{\partial b}(a,b) = \frac{\alpha^2-1}{\alpha}b.$$

Clearly,

$$\frac{\partial\psi_\alpha}{\partial a}(a,b)\frac{\partial\psi_\alpha}{\partial b}(a,b) \geq 0.$$

Furthermore, we obtain

$$\begin{aligned}
\left(\frac{\partial\psi_\alpha}{\partial a}(a,b) + \frac{\partial\psi_\alpha}{\partial b}(a,b)\right)^2 &= \left(\frac{\alpha^2-1}{\alpha}b\right)^2 \\
&= \left(\frac{\alpha^2-1}{\alpha^2}\right)^2 (a-(a-\alpha b)_+)^2 \\
&\geq \left(\frac{\alpha-1}{\alpha}\right)^2 (a-(a-b)_+)^2,
\end{aligned}$$

where the last step follows from [13, Lemma 1] and the fact that, for $\alpha > 1$, $(\alpha^2-1)/\alpha^2 > (\alpha-1)/\alpha$.

Case 3 : $a - \alpha b < 0$ and $b - \alpha a \geq 0$.
In that case,

$$\frac{\partial\psi_\alpha}{\partial b}(a,b) = 0 \quad \text{and} \quad \frac{\partial\psi_\alpha}{\partial a}(a,b) = \frac{\alpha^2-1}{\alpha}a.$$

Clearly,

$$\frac{\partial\psi_\alpha}{\partial a}(a,b)\frac{\partial\psi_\alpha}{\partial b}(a,b) \geq 0.$$

We also have

$$\begin{aligned}\left(\frac{\partial\psi_\alpha}{\partial a}(a,b)+\frac{\partial\psi_\alpha}{\partial b}(a,b)\right)^2 &= \left(\frac{\alpha^2-1}{\alpha}a\right)^2\\ &= \left(\frac{\alpha^2-1}{\alpha}\right)^2(a-(a-\alpha b)_+)^2\\ &\geq \left(\frac{\alpha-1}{\alpha}\right)^2(a-(a-b)_+)^2,\end{aligned}$$

where the last step follows from [13, Lemma 1] and the fact that, for $\alpha>1$, $\alpha^2>\alpha$.

Case 4 : $a-\alpha b<0$ and $b-\alpha a<0$.
In that case,

$$\frac{\partial\psi_\alpha}{\partial a}(a,b)=\frac{\alpha b-a}{\alpha}>0 \quad\text{and}\quad \frac{\partial\psi_\alpha}{\partial b}(a,b)=\frac{\alpha a-b}{\alpha}>0$$

Thus

$$\frac{\partial\psi_\alpha}{\partial a}(a,b)\frac{\partial\psi_\alpha}{\partial b}(a,b)\geq 0.$$

Note that $a<\alpha b<\alpha^2 a$ and $\alpha>1$ imply that $a>0$ and $b>0$. Therefore

$$\begin{aligned}\left(\frac{\partial\psi_\alpha}{\partial a}(a,b)+\frac{\partial\psi_\alpha}{\partial b}(a,b)\right)^2 &= ((1-1/\alpha)(a+b))^2\\ &> \left(\frac{\alpha-1}{\alpha}a\right)^2\\ &= \left(\frac{\alpha-1}{\alpha}\right)^2(a-(a-\alpha b)_+)^2\\ &\geq \left(\frac{\alpha-1}{\alpha}\right)^2(a-(a-b)_+)^2,\end{aligned}$$

where the last step follows from [13, Lemma 1].

Summing up for all $i=1,\dots,n$ completes the proof. ∎

We are now ready to state our derivative-free descent algorithm for minimizing the implicit Lagrangian merit function $M_\alpha(\cdot)$, and thus solving NCP(F).

ALGORITHM 1 *Choose* $x^0\in\Re^n$, $\gamma\in(0,1)$, *and a sufficiently small constant* $\beta>0$.
Having x^i, *let*

$$d^i := -\frac{\partial\psi_\alpha}{\partial b}(x^i,F(x^i))-\beta\frac{\partial\psi_\alpha}{\partial a}(x^i,F(x^i)), \tag{6}$$

Compute

$$x^{i+1}=x^i+\eta_i d^i, \tag{7}$$

where $\eta_i = \gamma^{k_i}$ with k_i being the smallest nonnegative integer k satisfying

$$M_\alpha(x^i) - M_\alpha(x^i + \gamma^k d^i) \geq \gamma^{2k} \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)) + \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i)) \|^2. \qquad (8)$$

Some remarks are in order on the choice of the constant β involved in computing the search direction in Algorithm 1 (we note that the same constant appears, for example, in the methods of [42, 43]). When the modulus of strong monotonicity μ, and the bound ν for $\|\nabla F\|$ over the (bounded) level set $\{x \mid M_\alpha(x) \leq M_\alpha(x^0)\}$ are available, the proof of Theorem 3 shows that we can choose any $\beta \leq 2\mu/\nu^2$. When those constants are not readily available, often some estimates of them can still be obtained (without computing the derivatives). However, sometimes those estimates can be too costly to obtain or they are too conservative. As a practical matter, we can start Algorithm 1 with some reasonably small value of β and adapt iteratively, by decreasing it if the linesearch step fails or the algorithm does not appear to make *predicted* progress, i.e. the (computable) sequence $\{M_\alpha(x^i)\}$ does not appear to decrease linearly to zero. Note that Theorem 2.1 establishes *global* linear convergence. Thus by monitoring the ratio $M_\alpha(x^{i+1})/M_\alpha(x^i)$ over the first few iterations, it should be possible to decrease β to an acceptable value, if the initial value is not adequate. Such adaptive approach would seem reasonable and fully implementable.

Note that no derivatives of $F(\cdot)$ are needed for computing the search direction or the stepsize in Algorithm 1. It can be checked that this algorithm employs the same search direction as the method described in [42] (which is given by (3)). However, the stepsize is computed according to a different rule.

Let x^0 be any starting point for Algorithm 1. Note that under our assumption of strong monotonicity, the level set

$$\mathcal{L}(M_\alpha, x^0) := \{x \mid M_\alpha(x) \leq M_\alpha(x^0)\}$$

is bounded ([15, Proposition 9]). By continuity of $F(\cdot)$, it further follows that the quantity

$$D(x^0) := \sup\{\|d(x)\| \mid x \in \mathcal{L}(M_\alpha, x^0)\}$$

is finite, where $d(x)$ is the search direction in Algorithm 1 computed at the point x. Hence, the set

$$\mathcal{B}(x^0) := \mathcal{L}(M_\alpha, x^0) + \{x \mid \|x\| \leq D(x^0)\}$$

is also bounded. We are now ready to state and prove our convergence result.

THEOREM 3 *Suppose $F(\cdot)$ is continuously differentiable and strongly monotone with modulus $\mu > 0$. Let $x^0 \in \Re^n$ be any given starting point, and suppose that $F(\cdot)$ and $\nabla F(\cdot)$ are Lipschitz continuous with some constant $L > 0$ on the set $\mathcal{B}(x^0)$. Then for the sequence $\{x^i\}$ generated by Algorithm 1, it holds that the sequence $\{M_\alpha(x^i)\}$ converges to zero Q-linearly, and $\{x^i\}$ converges R-linearly to the solution of NCP(F).*

Proof: Because $F(\cdot)$ and $\nabla F(\cdot)$ are Lipschitz continuous on $\mathcal{B}(x^0)$, it is clear that $\nabla M_\alpha(\cdot)$ is Lipschitz continuous on this bounded set. In particular, there exists $L_1 > 0$ such that

$$\|\nabla M_\alpha(x) - \nabla M_\alpha(y)\| \le L_1 \|x - y\| \quad \forall\, x, y \in \mathcal{B}(x^0). \tag{9}$$

Also, for some $\nu > 0$, we have

$$\|\nabla F(x)\| \le \nu \quad \forall\, x \in \mathcal{L}(M_\alpha, x^0). \tag{10}$$

By the construction of the algorithm (see (8)) the sequence $\{M_\alpha(x^i)\}$ is nonincreasing. Therefore the sequence $\{x^i\}$ is contained in $\mathcal{L}(M_\alpha, x^0)$.

For any $\theta \in [0,1]$ we have that $x^i, x^i + \theta d^i \in \mathcal{B}(x^0)$ and

$$\begin{aligned}
M_\alpha(x^i + \theta d^i) - M_\alpha(x^i) &= \int_0^\theta \langle \nabla M_\alpha(x^i + td^i), d^i\rangle dt \\
&= \theta \langle \nabla M_\alpha(x^i), d^i\rangle \\
&\quad + \int_0^\theta \langle \nabla M_\alpha(x^i + td^i) - \nabla M_\alpha(x^i), d^i\rangle dt \\
&\le \theta \langle \nabla M_\alpha(x^i), d^i\rangle + L_1 \int_0^\theta t\|d^i\|^2 dt \\
&= \theta \left(\langle \nabla M_\alpha(x^i), d^i\rangle + \frac{L_1 \theta}{2}\|d^i\|^2 \right),
\end{aligned} \tag{11}$$

where the inequality follows from (9) and the Cauchy-Schwarz inequality.

We further obtain

$$\begin{aligned}
-\|d^i\|^2 &= -\|\frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2 - 2\beta \langle \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)), \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\rangle \\
&\quad -\beta^2 \|\frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i))\|^2 \\
&\ge -\|\frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2 - 2\langle \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)), \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\rangle \\
&\quad -\|\frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i))\|^2 \\
&= -\|\frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)) + \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2
\end{aligned} \tag{12}$$

where the inequality follows from Lemma 1 if we take $\beta \in (0,1)$.

Furthermore,

$$
\begin{aligned}
-\langle \nabla M_\alpha(x^i), d^i\rangle &= -\langle \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)) + \nabla F(x^i)^\top \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i)), d^i\rangle \\
&= \langle \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)), \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\rangle + \beta \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i))\|^2 \\
&\quad + \langle \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i)), \nabla F(x^i) \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\rangle \\
&\quad + \beta \langle \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)), \nabla F(x^i)^\top \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\rangle \\
&\geq \beta \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i))\|^2 + \mu \| \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2 \\
&\quad -\beta\nu \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i))\| \| \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|, \qquad (13)
\end{aligned}
$$

where the inequality follows from Lemma 1, strong monotonicity of $F(\cdot)$, the Cauchy-Schwarz inequality, and (10).

We shall now consider the right-hand-side of (13). Denote $u := \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i))$ and $v := \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))$. We shall show that for a suitable choice of β, there exists a constant $\lambda > 0$ such that

$$\beta \|u\|^2 + \mu \|v\|^2 - \beta\nu \|u\| \|v\| \geq \lambda \|u + v\|^2. \qquad (14)$$

By the Cauchy-Schwarz inequality, it is sufficient to show that

$$(\mu - \lambda)\|v\|^2 + (\beta - \lambda)\|u\|^2 - (\beta\nu + 2\lambda)\|u\|\|v\| \geq 0. \qquad (15)$$

If $\mu > \lambda$, $\beta > \lambda$ and

$$2\sqrt{(\mu - \lambda)(\beta - \lambda)} \geq \beta\nu + 2\lambda \qquad (16)$$

then the left-hand-side of (15) is greater or equal than $(\sqrt{\mu - \lambda}\|v\| - \sqrt{\beta - \lambda}\|u\|)^2$ which is nonnegative. Hence it is enough to check (16) or, equivalently, the inequality

$$\mu\beta - \frac{1}{4}\beta^2\nu^2 \geq \lambda\beta + \lambda\mu + \beta\lambda\nu.$$

Take any $\beta \leq 2\mu/\nu^2$. Then for the inequality above to hold, it is enough to guarantee that

$$\frac{1}{2}\mu\beta \geq \lambda\beta(1 + \nu) + \lambda\mu.$$

It is easy to check that the latter inequality is satisfied for

$$\lambda \leq \min\left\{ \frac{\mu}{4(1+\nu)}, \frac{\beta}{4} \right\}.$$

It follows that for this choice of β and λ, (14) holds.

Combining (13) and (14), we obtain

$$-\langle \nabla M_\alpha(x^i), d^i\rangle \geq \lambda \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)) + \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2.$$

The latter inequality, (11) and (12) yield

$$M_\alpha(x^i) - M_\alpha(x^i + \theta d^i) \geq \theta \left(\lambda - \frac{L_1 \theta}{2}\right) \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)) + \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2.$$

It follows from the latter inequality that (8) is satisfied whenever

$$\lambda - L_1 \gamma^k/2 \geq \gamma^k,$$

that is

$$\gamma^k \leq \lambda(1 + L_1/2)^{-1}.$$

In particular, (8) is certainly satisfied for

$$k \geq \left\lceil \log_{1/\gamma}\left(\frac{1 + L_1/2}{\lambda}\right)\right\rceil =: K.$$

It follows that

$$\eta_i \geq \gamma^K =: \eta > 0.$$

Now from (8) and Lemma 1, we obtain

$$\begin{aligned} M_\alpha(x^i) - M_\alpha(x^{i+1}) &\geq \sigma\eta^2 \| \frac{\partial \psi_\alpha}{\partial a}(x^i, F(x^i)) + \frac{\partial \psi_\alpha}{\partial b}(x^i, F(x^i))\|^2 \\ &\geq \sigma\eta^2 \left(\frac{\alpha - 1}{\alpha}\right)^2 \|r(x^i)\|^2. \end{aligned}$$

From Theorem 1, we further deduce that

$$M_\alpha(x^i) - M_\alpha(x^{i+1}) \geq \sigma\eta^2 \frac{\alpha - 1}{\alpha^2} M_\alpha(x^i)$$

and, rearranging terms, we obtain

$$\left(1 - \sigma\eta^2 \frac{\alpha - 1}{\alpha^2}\right) M_\alpha(x^i) \geq M_\alpha(x^{i+1}) \geq 0$$

which means that $\{M_\alpha(x^i)\}$ converges Q-linearly to zero.

Furthermore, from Theorems 1 and 2, we have

$$\|x^i - x^*\|^2 \leq \frac{\alpha(L+1)^2}{\mu^2(\alpha - 1)} M_\alpha(x^i).$$

Hence, the fact that $\{M_\alpha(x^i)\}$ converges Q-linearly to zero implies that $\{x^i\}$ converges to x^* with R-linear rate. ■

It can be checked that under the assumption of local Lipschitz continuity of F and ∇F in some neighborhood of the solution of the problem, the sequence generated by Algorithm 1 achieves an *asymptotic* linear rate of convergence. The result of Theorem 3 gives a global linear rate under global assumption.

Finally, an interesting open question is whether the linear convergence result still holds when the strong monotonicity assumption on $F(\cdot)$ is replaced by the uniform P-property.

3. Concluding Remarks

A descent algorithm for solving strongly monotone nonlinear complementarity problems was considered. The algorithm is based on the implicit Lagrangian merit function and does not require any derivatives information. It was shown that the iterates generated by this method converge linearly to the solution of the problem. This appears to be the first rate of convergence result for the class of derivative-free descent methods for solving nonlinear complementarity problems. Whether (and under what conditions) derivative-free descent methods based on other merit functions converge linearly, can be a subject of future research.

Acknowledgments

We are grateful to the two anonymous referees for their constructive review and useful comments.

References

1. R.W. Cottle, J.-S. Pang, and R.E. Stone. *The Linear Complementarity Problem.* Academic Press, New York, 1992.
2. S.P. Dirkse and M.C. Ferris. The PATH solver : A nonmonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5:123–156, 1995.
3. F. Facchinei, A. Fischer, and C. Kanzow. Inexact Newton methods for semismooth equations with applications to variational inequality problems. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 125–139. Plenum Press, 1996.
4. F. Facchinei and C. Kanzow. On unconstrained and constrained stationary points of the implicit Lagrangian. *Journal of Optimization Theory and Applications*, 92:99–115, 1997.
5. F. Facchinei and J. Soares. A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM Journal on Optimization*, 7:225–247, 1997.
6. M.C. Ferris and J.-S. Pang (editors). *Complementarity and variational problems : State of the Art.* SIAM Publications, 1997.
7. M.C. Ferris and D. Ralph. Projected gradient methods for nonlinear complementarity problems via normal maps. In D.-Z. Du, L. Qi, and R. Womersley, editors, *Recent Advances in Nonsmooth Optimization*, pages 1–29. World Scientific Publishers, 1994.
8. A. Fischer. An NCP function and its use for the solution of complementarity problems. In D.-Z. Du, L. Qi, and R. Womersley, editors, *Recent Advances in Nonsmooth Optimization*, pages 88–105. World Scientific Publishers, Singapore, 1995.

9. A. Fischer. A new constrained optimization reformulation for complementarity problems. Technical Report MATH-NM-10-1995, Institute for Numerical Mathematics, Technical University of Dresden, D-01062 Dresden, Germany, 1995.
10. A. Friedlander, J.M. Martínez, and S.A. Santos. Solution of linear complementarity problems using minimization with simple bounds. *Journal of Global Optimization*, 6:253–267, 1995.
11. M. Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, 53:99–110, 1992.
12. M. Fukushima. Merit functions for variational inequality and complementarity problems. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 155–170. Plenum Publishing Corporation, New York, New York, 1996.
13. E.M. Gafni and D.P. Bertsekas. Two–metric projection methods for constrained optimization. *SIAM Journal on Control and Optimization*, 22:936–964, 1984.
14. C. Geiger and C. Kanzow. On the resolution of monotone complementarity problems. *Computational Optimization and Applications*, 5:155–173, 1996.
15. H. Jiang. Unconstrained minimization approaches to nonlinear complementarity problems. *Journal of Global Optimization*, 9:169–181, 1996.
16. H. Jiang and L. Qi. A new nonsmooth equations approach to nonlinear complementarity problems. *SIAM Journal on Control and Optimization*, 35:178–193, 1997.
17. C. Kanzow. Some equation-based methods for the nonlinear complementarity problem. *Optimization Methods and Software*, 3:327–340, 1994.
18. C. Kanzow. Nonlinear complementarity as unconstrained optimization. *Journal of Optimization Theory and Applications*, 88:139–155, 1996.
19. C. Kanzow and H. Kleinmichel. A new class of semismooth Newton-type methods for nonlinear complementarity problems, 1997. *Computational Optimization and Applications*, to appear.
20. C. Kanzow, N. Yamashita, and M. Fukushima. New NCP-functions and their properties. *Journal of Optimization Theory and Applications*, 94:115–135, 1997.
21. W. Li. A merit function and a Newton-type method for symmetric linear complementarity problems. In M.C. Ferris and J.-S. Pang, editors, *Complementarity and variational problems : State of the art*. SIAM Publications, 1997.
22. T. De Luca, F. Facchinei, and C. Kanzow. A semismooth equation approach to the solution of nonlinear complementarity problems. *Mathematical Programming*, 75:407–439, 1996.
23. Z.-Q. Luo, O.L. Mangasarian, J. Ren, and M.V. Solodov. New error bounds for the linear complementarity problem. *Mathematics of Operations Research*, 19:880–892, 1994.
24. Z.-Q. Luo and P. Tseng. A new class of merit functions for the nonlinear complementarity problem. In M.C. Ferris and J.-S. Pang, editors, *Complementarity and variational problems : State of the art*. SIAM Publications, 1997.
25. O.L. Mangasarian. Equivalence of the complementarity problem to a system of nonlinear equations. *SIAM Journal of Applied Mathematics*, 31:89–92, 1976.
26. O.L. Mangasarian and M.V. Solodov. Nonlinear complementarity as unconstrained and constrained minimization. *Mathematical Programming*, 62:277–297, 1993.
27. J. Moré. Global methods for nonlinear complementarity problems. *Mathematics of Operations Research*, 21:589–614, 1996.
28. J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, 1970.
29. J.-S. Pang. Inexact Newton methods for the nonlinear complementarity problem. *Mathematical Programming*, 36:54–71, 1986.
30. J.-S. Pang. A posteriori error bounds for the linearly–constrained variational inequality problem. *Mathematics of Operations Research*, 12:474–484, 1987.
31. J.-S. Pang. Complementarity problems. In R. Horst and P. Pardalos, editors, *Handbook of Global Optimization*, pages 271–338. Kluwer Academic Publishers, Boston, Massachusetts, 1995.
32. J.-S. Pang and S.A. Gabriel. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60:295–337, 1993.
33. J.-M. Peng. Convexity of the implicit Lagrangian. *Journal of Optimization Theory and Applications*, 92:331–341, 1996.

34. J.-M. Peng. A global method for monotone variational inequality problems with inequality constraints. *Journal of Optimization Theory and Applications*, 95:419–430, 1997.
35. M.V. Solodov. Implicit Lagrangian. An invited article for the *Encyclopedia of Optmimization*, C. Floudas and P. Pardalos (editors), Kluwer Academic Publishers, 1999.
36. M.V. Solodov. Some optimization reformulations of the extended linear complementarity problems, 1997. *Computational Optimization and Applications*, to appear.
37. M.V. Solodov. Stationary points of bound constrained reformulations of complementarity problems. *Journal of Optimization Theory and Applications*, 94:449–467, 1997.
38. P.K. Subramanian. Gauss-Newton methods for the complementarity problem. *Journal of Optimization Theory and Applications*, 77:467–482, 1993.
39. K. Taji and M. Fukushima. Optimization based globally convergent methods for the nonlinear complementarity problems. *Journal of the Operations Research Society of Japan*, 37:310–331, 1994.
40. P. Tseng. Growth behavior of a class of merit functions for the nonlinear complementarity problem. *Journal of Optimization Theory and Applications*, 89:17–37, 1996.
41. P. Tseng, N. Yamashita, and M. Fukushima. Equivalence of complementarity problems to differentiable minimization : A unified approach. *SIAM Journal on Optimization*, 6:446–460, 1996.
42. N. Yamashita and M. Fukushima. On stationary points of the implicit Lagrangian for nonlinear complementarity problems. *Journal of Optimization Theory and Applications*, 84:653–663, 1995.
43. N. Yamashita, K. Taji, and M. Fukushima. Unconstrained optimization reformulations of variational inequality problems. *Journal of Optimization Theory and Applications*, 92:439–456, 1997.