

Local Linear Convergence of an Outer Approximation Projection Method for Variational Inequalities

Shu Lu and Sudhanshu Singh

Communicated by M. Fukushima

Abstract: This paper considers an outer approximation projection method for variational inequalities, in which the projections are not performed on the original set that appears in the variational inequality, but on a polyhedral convex set defined by the linearized constraints. It shows that the method converges linearly, when the starting point is sufficiently close to the solution and the step lengths are sufficiently small.

Keywords: variational inequality, projection method, outer approximation, local convergence

AMS classification: 65K15, 90C33

The research reported here was sponsored by the National Science Foundation under Grant DMS-0807893.

Shu Lu (Corresponding author)
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: shulu@email.unc.edu

Sudhanshu Singh
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: sssingh@email.unc.edu

1 Introduction

In this paper, we consider a variational inequality defined by a continuously differentiable function from \mathbb{R}^n to \mathbb{R}^n , and a convex set in \mathbb{R}^n . The set is defined by finitely many constraints.

There are a host of effective algorithms on solving variational inequalities. The book [1] contains a comprehensive treatment on this subject. The focus of this paper is on the local linear convergence of an outer approximation projection method, in which the projections are not performed on the original set that defines the variational inequality, but on a polyhedral convex set defined by the linearized constraints.

Most existing projection methods require computing in each iteration projections onto the set that defines the variational inequality. Computation of the projection is easy, when this set has a simple structure, such as being the non-negative orthant. When the set is defined by nonlinear constraints, finding such projections amounts to solving a nonlinear convex program, and its computational complexity becomes an obstacle, that limits the applicability of projection methods. This motivates the development of outer approximation projection methods, studied by Fukushima and his coauthors in [2, 3, 4]. In [2], Fukushima described an algorithm for solving general convex programs. Each iteration of the algorithm solves a quadratic program, whose feasible set is defined by linearized constraints of the original problem. In [3], Fukushima proposed a relaxed projection method to solve variational inequalities, which in each iteration computes the projection of some point onto a halfspace that contains the set that defines the variational inequality. Fukushima and Taji [4] developed a merit function approach to solve variational inequalities, which in each iteration solves a quadratic program to find the search direction and then conducts an exact line search to minimize the merit function along this direction. Evaluation

of the merit function requires solving a quadratic program.

Recently, Censor and Gibali [5] proposed a scheme that finds a projection onto a halfspace which separates a ball around the current iterate from the set that defines the variational inequality, provided the current iterate be not in the interior of the set. This scheme generalizes the classical projection method and the relaxed projection method in [3]. Yang [6] proved the global convergence of the method in [3] under weakened assumptions. In [7], Cruz and Iusem proved the global convergence of a relaxed projection method under the assumption that the function defining the variational inequality be paramonotone. See references of these papers for more work on related methods.

The objective of this paper is to complement the results in [2, 3, 4] by conducting a local convergence analysis. The major assumptions we use are that the solution satisfies the linear independence constraint qualification, and that the Jacobian matrix of the function defining the variational inequality at the solution be positive definite. We do not require the function to be globally monotone.

In the algorithm, we use the affine set defined by all of the linearized constraints as an outer approximation of the original set that defines the variational inequality, as in [2, 4]. In each iteration, we compute the projection of some point onto the affine set. We show that the method converges linearly, when the starting point is sufficiently close to the solution and the step lengths are sufficiently small. We also give a formula for computing an upper bound for step lengths. Our numerical experiments show that the upper bound given by the formula tends to be too conservative; in practice, one could use step lengths bigger than that bound and still have fast convergence. The main purpose of that formula is to provide insights and guidance on step length selection.

In the next section, we introduce the algorithm and discuss its termination

condition. Section 3 analyzes the behavior of the (skewed) Euclidean projector onto the outer approximation set, and proves local convergence of the algorithm. Section 4 contains numerical examples and discussion.

2 The Algorithm

To define the algorithm, we need to introduce some notation. Let F be a C^1 function from \mathbb{R}^n to \mathbb{R}^n , and the set S be defined by m constraints:

$$S := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\}. \quad (1)$$

Throughout the paper we assume each g_i be a convex C^2 function from \mathbb{R}^n to \mathbb{R} . The latter assumption ensures that S is convex. We also assume S be nonempty. We use the notation $N_S(x)$ to denote the normal cone to S at x . The variational inequality problem is to find a point $x \in S$, such that

$$0 \in F(x) + N_S(x). \quad (2)$$

Given any point $x' \in \mathbb{R}^n$, we define a polyhedral convex set

$$S(x') := \{x \in \mathbb{R}^n \mid g_i(x') + \langle \nabla g_i(x'), x - x' \rangle \leq 0, i = 1, \dots, m\}.$$

Note that the inclusion

$$S \subset S(x')$$

holds for every $x' \in \mathbb{R}^n$ under the convexity assumption. Moreover, $x' \in S(x')$ if and only if $x' \in S$. Finally, if $x' \in S$ and a constraint qualification holds at

x' , then the tangent cone to S at x' , denoted by $T_S(x')$, can be written as

$$T_S(x') := \{v \in \mathbb{R}^n \mid \langle \nabla g_i(x'), v \rangle \leq 0, i \in I(x')\},$$

where $I(x')$ denotes the set of active constraints at x' . Note that, for each such x' , the set $x' + T_S(x')$ locally coincides with $S(x')$ around x' .

Let D be an $n \times n$ symmetric positive definite matrix, and C be a convex set in \mathbb{R}^n . We use the notation $\Pi_{C,D}$ to denote the skewed projector onto the set C defined by the matrix D ; that is, for every point $x \in \mathbb{R}^n$, $\Pi_{C,D}(x)$ is the solution of the following convex program in the variable y :

$$\min_{y \in C} \frac{1}{2}(y-x)^T D(y-x). \quad (3)$$

We use $\|\cdot\|_D$ to denote the D -norm on \mathbb{R}^n ; that is,

$$\|x\|_D := (x^T D x)^{1/2} \text{ for each } x \in \mathbb{R}^n.$$

It is well known that the skewed projector is nonexpansive under the D -norm; see [1, Exercise 1.8.16].

Algorithm.

Step 0. Choose x_0 , t_0 and a symmetric positive definite matrix D , and let

$$k := 0.$$

Step 1. Let

$$x_{k+1} = \Pi_{S(x_k), D}(x_k - t_k D^{-1} F(x_k)). \quad (4)$$

Step 2. If $x_{k+1} = x_k$ then stop. Otherwise, choose $t_{k+1} > 0$, let $k := k + 1$, and go to Step 1.

Clearly, the algorithm terminates at a point x^* if it satisfies

$$x^* = \Pi_{S(x^*), D}(x^* - t^* D^{-1} F(x^*)) \quad (5)$$

with t^* being the step length at that iteration. It is not hard to see that (5) holds if and only if

$$-F(x^*) \in N_{S(x^*)}(x^*).$$

The following lemma further relates (5) to the variational inequality (2), and shows that the algorithm terminates if and only if it finds a solution of (2).

Lemma 2.1. *If x^* is a solution to the variational inequality (2) and a constraint qualification holds at x^* , then (5) holds for any $t^* > 0$. Conversely, if a point $x^* \in \mathbb{R}^n$ satisfies (5) with some $t^* > 0$, then it is a solution to (2).*

Proof. First, suppose x^* be a solution to (2) at which a constraint qualification holds. By the remarks at the beginning of this section, $x^* + T_S(x^*)$ locally coincides with $S(x^*)$. It follows that the tangent cone to $S(x^*)$ at x^* is exactly $T_S(x^*)$, and that the normal cone to $S(x^*)$ at x^* is $N_S(x^*)$. Since x^* solves (2), the vector $-F(x^*)$ belongs to $N_S(x^*)$, so it also belongs to the normal cone to $S(x^*)$ at x^* . This proves (5) in view of the remarks right before the Lemma 2.1.

For the converse direction, suppose x^* satisfies (5) with some $t^* > 0$. This implies that $x^* \in S(x^*)$, and that $-F(x^*)$ belongs to $N_{S(x^*)}(x^*)$. The fact $x^* \in S(x^*)$ implies $x^* \in S$. The fact that $S \subset S(x^*)$ implies that $N_S(x^*)$ contains $N_{S(x^*)}(x^*)$, so $-F(x^*)$ belongs to $N_S(x^*)$. \square

3 Local Convergence

To prepare for the convergence analysis of the algorithm, we first analyze the behavior of the skewed Euclidean projector onto the set $S(x')$ as x' varies. To

this end, let D be a symmetric positive definite matrix, and consider the problem of finding $\Pi_{S(x'), D}(y)$, namely,

$$\min_{x \in S(x')} \frac{1}{2} (y - x)^T D (y - x). \quad (6)$$

Since $S(x')$ is a nonempty, closed and convex set for each $x' \in \mathbb{R}^n$, the problem (6) has a unique solution, denoted by $x(y, x')$ for each x' and y in \mathbb{R}^n . We will show that the function $x(y, x')$ is B-differentiable at certain points, and will give a formula for its B-derivatives. The following definition of B-differentiability follows from [8].

Definition 3.1. *A function $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is B-differentiable at a point $z \in \mathbb{R}^n$, if there exists a positively homogeneous function $dG(z) : \mathbb{R}^n \rightarrow \mathbb{R}^m$, such that*

$$\lim_{v \rightarrow 0} \frac{G(z + v) - G(z) - dG(z)(v)}{\|v\|} = 0.$$

Because $S(x')$ is defined by linear constraints, the problem (6) is equivalent to its first-order necessary conditions:

$$\begin{aligned} Dx - Dy + \sum_{i=1, \dots, m} \nabla_x g_i(x') \lambda_i &= 0, \\ 0 \leq \lambda_i \perp g_i(x') + \langle \nabla g_i(x'), x - x' \rangle &\leq 0, i = 1, \dots, m. \end{aligned} \quad (7)$$

If we define a function $f : \mathbb{R}^{3n+m} \rightarrow \mathbb{R}^{n+m}$ by

$$f(x, \lambda, y, x') = \begin{bmatrix} Dx - Dy + \sum_{i=1, \dots, m} \nabla g_i(x') \lambda_i \\ -g(x') - \nabla g(x')(x - x') \end{bmatrix},$$

where $\nabla g(x')$ denotes the Jacobian matrix of g at x' , then (7) can be equivalently written as a variational inequality with (x, λ) being the variable, and (y, x')

being the parameter:

$$0 \in f(x, \lambda, y, x') + N_{\mathbb{R}^n \times \mathbb{R}_+^m}(x, \lambda). \quad (8)$$

The equivalence between (7) and (8) follows from the fact that the normal cone $N_{\mathbb{R}^n \times \mathbb{R}_+^m}(x, \lambda)$ is exactly the collection of vectors $(0, w) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $0 \leq \lambda_i \perp w_i \leq 0$ for each $i = 1, \dots, m$.

Now suppose that x^* be a solution to (2), which satisfies the linear independence constraint qualification (abbreviated as LICQ). In the following theorem, we apply the sensitivity analysis technique in [9] to (8), to give a formula for the B-derivative of the function x at $(x^* - t^* D^{-1} F(x^*), x^*)$ for each $t^* \geq 0$. By assumption, x^* solves (2) and satisfies the LICQ, so it follows from Lemma 2.1 that

$$x^* = x(x^* - t^* D^{-1} F(x^*), x^*).$$

Moreover, there exists a unique vector $\lambda(t^*)$ in \mathbb{R}^m such that $(x, \lambda, y, x') = (x^*, \lambda(t^*), x^* - t^* D^{-1} F(x^*), x^*)$ satisfies (7). The function $\lambda(t^*)$ is positively homogeneous; if we write $\lambda_1^* = \lambda(1)$ then $\lambda(t^*) = t^* \lambda_1^*$.

Theorem 3.1. *Let x^* be a solution to (2) which satisfies the LICQ, D be a symmetric positive definite $n \times n$ matrix, and $t^* \geq 0$. Define*

$$y^* = x^* - t^* D^{-1} F(x^*),$$

and write $\lambda^ = t^* \lambda_1^*$, with λ_1^* being as defined above. Partition the index set $\{1, \dots, m\}$ into three subsets,*

$$I_1 = \{i : g_i(x^*) < 0\}, \quad I_{00} = \{i : g_i(x^*) = 0, \lambda_i^* = 0\},$$

$$I_{01} = \{i : g_i(x^*) = 0, \lambda_i^* > 0\},$$

and define

$$K_x = \left\{ v \in \mathbb{R}^n \left| \begin{array}{l} \langle \nabla g_i(x^*), v \rangle = 0 \text{ if } i \in I_{01} \\ \langle \nabla g_i(x^*), v \rangle \leq 0 \text{ if } i \in I_{00} \end{array} \right. \right\}. \quad (9)$$

The function x is B -differentiable at (y^*, x^*) , and its B -derivative for the direction (u, v') is given by

$$dx(y^*, x^*)(u, v') = \Pi_{K_x, D}(u - \sum_{i=1, \dots, m} \lambda_i^* D^{-1} \nabla_{xx}^2 g_i(x^*) v'). \quad (10)$$

Proof. Let L be the Jacobian matrix of f with respect to (x, λ) at $(x^*, \lambda^*, y^*, x^*)$,

$$L = \begin{bmatrix} D & \nabla g(x^*)^T \\ -\nabla g(x^*) & 0 \end{bmatrix},$$

and let K be the critical cone to $\mathbb{R}^n \times \mathbb{R}_+^m$ at (x^*, λ^*) associated with the vector $-f(x^*, \lambda^*, y^*, x^*)$. Then

$$\begin{aligned} K &= \{(v, w) \in T_{\mathbb{R}^n \times \mathbb{R}_+^m}(x^*, \lambda^*) \mid \langle f(x^*, \lambda^*, y^*, x^*), (v, w) \rangle = 0\} \\ &= \mathbb{R}^n \times \{w \in T_{\mathbb{R}_+^m}(\lambda^*) \mid \langle g(x^*), w \rangle = 0\} \\ &= \mathbb{R}^n \times \left\{ w \in \mathbb{R}^m \left| \begin{array}{l} w_i = 0 \text{ if } g_i(x^*) < 0 \\ w_i \in \mathbb{R} \text{ if } g_i(x^*) = 0 \text{ and } \lambda_i^* > 0 \\ w_i \geq 0 \text{ if } g_i(x^*) = 0 \text{ and } \lambda_i^* = 0 \end{array} \right. \right\}. \end{aligned}$$

It follows that

$$K = \mathbb{R}^n \times \left\{ w \in \mathbb{R}^m \left| \begin{array}{l} w_i = 0 \text{ if } i \in I_1 \\ w_i \in \mathbb{R} \text{ if } i \in I_{01} \\ w_i \geq 0 \text{ if } i \in I_{00} \end{array} \right. \right\}.$$

In order to apply the technique in [9], let us define the normal map induced

by the linear map L on K as

$$L_K(v', w') := L(\Pi_K(v', w')) + (v', w') - \Pi_K(v', w').$$

The LICQ assumption implies that the normal map L_K is a homeomorphism from \mathbb{R}^{n+m} to \mathbb{R}^{n+m} (see [10, Theorem 3.1]). Applying [9, Theorem 3] to (8), we find neighborhoods X of x^* , Λ of λ^* , Y of y^* and X' of x^* , such that for each $y \in Y$ and $x' \in X'$, the variational inequality (8) has a unique solution in $X \times \Lambda$, which we denote by $h(y, x')$. Note that the x -component of $h(y, x')$ is exactly $x(y, x')$. Accordingly, we can write $h(y, x') = (x(y, x'), \lambda(y, x'))$. By [9, Theorem 4], the function $h(y, x')$ is B-differentiable at (y^*, x^*) , with the B-derivative $dh(y^*, x^*)$ being given by

$$\begin{aligned} dh(y^*, x^*)(u, v') &= \Pi_K(L_K)^{-1}(-d_{y,x'}f(x^*, \lambda^*, y^*, x^*)(u, v')) \\ &= \Pi_K(L_K)^{-1} \begin{bmatrix} Du - \sum_{i=1, \dots, m} \lambda_i^* \nabla_{xx}^2 g_i(x^*) v' \\ 0 \end{bmatrix}. \end{aligned} \quad (11)$$

To prove (10), consider the following nonlinear program with u being the parameter and v being the variable:

$$\min \frac{1}{2}(u - v)^T D(u - v), \quad \text{s.t. } v \in K_x, \quad (12)$$

where K_x is defined in (9). Because K_x is a polyhedral convex set, v solves (12) if and only if there exist multipliers $\lambda_i, i \in I_{01} \cup I_{00}$, such that the following

first-order conditions hold:

$$\begin{aligned}
Dv - Du + \sum_{i \in I_{01} \cup I_{00}} \lambda_i \nabla g_i(x^*) &= 0, \\
\langle \nabla g_i(x^*), v \rangle &= 0, i \in I_{01}, \\
\lambda_i &\geq 0 \perp \langle \nabla g_i(x^*), v \rangle \leq 0, i \in I_{00}.
\end{aligned} \tag{13}$$

Define additionally $\lambda_i = 0$ for $i \in I_1$. Using the expression of K at the beginning of this proof, we may rewrite (13) as

$$\begin{bmatrix} -Dv + Du - \sum_{i=1, \dots, m} \lambda_i \nabla g_i(x^*) \\ \nabla g(x^*)v \end{bmatrix} \in N_K(v, \lambda). \tag{14}$$

Now, define an affine map $\hat{L} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^{n+m}$ by

$$\hat{L}(v, \lambda) = \begin{bmatrix} Dv - Du + \sum_{i=1, \dots, m} \lambda_i \nabla g_i(x^*) \\ -\nabla g(x^*)v \end{bmatrix},$$

to further rewrite (14) as

$$0 \in \hat{L}(v, \lambda) + N_K(v, \lambda).$$

This is equivalent to

$$\begin{bmatrix} v \\ \lambda \end{bmatrix} = \Pi_K(\hat{L}_K)^{-1}(0).$$

Note that $\hat{L}(v, \lambda) = L(v, \lambda) + (-Du, 0)$. It then follows from the definition of normal maps that $\hat{L}_K(v, \lambda) = L_K(v, \lambda) + (-Du, 0)$. Accordingly,

$$(\hat{L}_K)^{-1}(0) = (L_K)^{-1}(Du, 0).$$

We have so far shown that the solution v to (12), together with the associated multiplier λ , can be equivalently expressed as

$$\begin{bmatrix} v \\ \lambda \end{bmatrix} = \Pi_K(L_K)^{-1}(Du, 0).$$

Recall the formula for $dh(y^*, x^*)$ in (11). Write

$$dh(y^*, x^*) = (dx(y^*, x^*), d\lambda(y^*, x^*)).$$

We see that $dx(y^*, x^*)(u, v')$ is the solution of (12) with u replaced by $u - \sum_{i=1, \dots, m} \lambda_i^* D^{-1} \nabla_{xx}^2 g_i(x^*) v'$. This proves (10). \square

Next, define a function $\hat{x} : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ by

$$\hat{x}(x', t) = \Pi_{S(x'), D}(x' - tD^{-1}F(x')). \quad (15)$$

We have

$$\hat{x}(x', t) = x(x' - tD^{-1}F(x'), x').$$

By the chain rule, the function \hat{x} is B-differentiable at (x^*, t^*) for any $t^* \geq 0$, with the B-derivative given by

$$\begin{aligned} d\hat{x}(x^*, t^*)(v, t) = \\ \Pi_{K_x, D}(v - t^* D^{-1} \nabla F(x^*) v - t D^{-1} F(x^*) - \sum_{i=1, \dots, m} t^* (\lambda_1^*)_i D^{-1} \nabla_{xx}^2 g_i(x^*) v). \end{aligned} \quad (16)$$

The equation above will play an important role in the local convergence proof. Before we present that proof, we give the following basic lemma. Recall the definition of the D -norm on \mathbb{R}^n in Section 2.

Lemma 3.1. *Let M be a positive definite $n \times n$ matrix, I_n be the $n \times n$ identity matrix, and D be a symmetric positive definite $n \times n$ matrix. Let $\|D^{-1}M\|_D$ denote the D -norm of $D^{-1}M$, that is,*

$$\|D^{-1}M\|_D = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|D^{-1}Mx\|_D}{\|x\|_D}.$$

Define a scalar

$$\mu := \min_{x \in \mathbb{R}^n, x \neq 0} \frac{x^T Mx}{\|x\|_D^2}.$$

Then $\mu > 0$, and for each scalar t one has

$$\|I_n - tD^{-1}M\|_D \leq (1 + t^2\|D^{-1}M\|_D^2 - 2t\mu)^{1/2},$$

which is less than 1 when $t < 2\mu/\|D^{-1}M\|_D^2$.

Proof. The fact $\mu > 0$ follows directly from the positive definiteness of M . For any $x \in \mathbb{R}^n$, we have

$$\begin{aligned} \|x - tD^{-1}Mx\|_D^2 &= \|x\|_D^2 - 2tx^T Mx + t^2\|D^{-1}Mx\|_D^2 \\ &\leq (1 + t^2\|D^{-1}M\|_D^2 - 2t\mu)\|x\|_D^2. \end{aligned}$$

The lemma follows. □

The following theorem is the main result of this paper.

Theorem 3.2. *Let x^* be a solution to (2) which satisfies the LICQ, and let D be a symmetric positive definite $n \times n$ matrix. Let λ_1^* be as defined before Theorem 3.1. Suppose that $\nabla F(x^*)$ be positive definite. Define*

$$M := \nabla F(x^*) + \sum_{i=1, \dots, m} (\lambda_1^*)_i \nabla_{xx}^2 g_i(x^*),$$

$$\mu := \min_{x \in \mathbb{R}^n, x \neq 0} \frac{x^T M x}{\|x\|_D^2},$$

and let t^* be a positive scalar satisfying $t^* < 2\mu/\|D^{-1}M\|_D^2$. Then there exists a neighborhood X_0 of x^* , such that the sequence $\{x_k\}$ will remain in X_0 and converge Q -linearly to x^* , whenever $x_0 \in X_0$ and $t_n = t^*$ for all n .

Proof. Let t^* be as given in the statement of the present theorem, and write $y^* = x^* - t^*D^{-1}F(x^*)$. Let ϵ be a positive scalar satisfying

$$\epsilon < 1 - (1 + (t^*)^2\|D^{-1}M\|_D^2 - 2t^*\mu)^{1/2}.$$

Because the function \hat{x} is B-differentiable at (x^*, t^*) with its B-derivative given in (16), there exist neighborhoods X_0 of x^* and T_0 of t^* such that

$$\|\hat{x}(x, t) - \hat{x}(x^*, t^*)\|_D \leq \|d\hat{x}(x^*, t^*)(x - x^*, t - t^*)\|_D + \epsilon\|x - x^*\|_D + \epsilon\|t - t^*\|,$$

for each $x \in X_0$ and $t \in T_0$. Without loss of generality, we may assume that X_0 be a ball around x^* under the D -norm.

Now, assume that $x_k \in X_0$. We have

$$\begin{aligned} \|x_{k+1} - x^*\|_D &= \|\hat{x}(x_k, t^*) - \hat{x}(x^*, t^*)\|_D \\ &\leq \|d\hat{x}(x^*, t^*)(x_k - x^*, 0)\|_D + \epsilon\|x_k - x^*\|_D \\ &= \|\Pi_{K_x, D}(x_k - x^* - t^*D^{-1}M(x_k - x^*))\|_D + \epsilon\|x_k - x^*\|_D \\ &\leq \|x_k - x^* - t^*D^{-1}M(x_k - x^*)\|_D + \epsilon\|x_k - x^*\|_D \\ &\leq ((1 + (t^*)^2\|D^{-1}M\|_D^2 - 2t^*\mu)^{1/2} + \epsilon)\|x_k - x^*\|_D. \end{aligned}$$

This proves that x_{k+1} remains in X_0 and that the sequence $\{x_k\}$ converges Q -linearly to x^* with rate $(1 + (t^*)^2\|M\|^2 - 2t^*\mu)^{1/2} + \epsilon$. \square

Theorem 3.2 suggests, that for fast convergence, one shall choose t^* to min-

imize

$$(1 + (t^*)^2 \|D^{-1}M\|_D^2 - 2t^*\mu)^{1/2}.$$

Clearly, the minimizer of the quantity above is

$$t^* = \mu / \|D^{-1}M\|_D^2, \tag{17}$$

and the minimum value achieved is

$$(1 - \mu^2 / \|D^{-1}M\|_D^2)^{1/2}.$$

Thus, the algorithm converges fast locally near a solution, if $\mu / \|D^{-1}M\|_D$ is close to 1, for example, when $D = I_n$ and M is well-conditioned.

Note that one does not know the values of μ and $\|D^{-1}M\|_D$ before finding the solution x^* . Hence, in reality, one would not use the formula (17) to determine precisely an upper bound for the parameter t^* . The purpose of that formula is to serve as a general guideline for the choice of t^* .

4 Numerical Examples

We use the following examples to illustrate the algorithm. For all the examples, we use identity matrices of appropriate sizes as the matrix D . Starting from a point close enough to the solution, we use the t^* values computed by (17). For comparison purposes, we also test some other choices of step lengths. We stop the algorithm when $\|x_{k+1} - x_k\| \leq 0.001$.

Example 4.1. *Let*

$$F(x) = \begin{pmatrix} 3x_1 + 4x_2 + 5 \\ 2x_1 + 5x_2 - 4 \end{pmatrix},$$

and let the set S be

$$S = \left\{ x \in \mathbb{R}^2 \left| \begin{array}{l} x_1^2 + 4x_2^2 \leq 4 \\ 2x_1^2 + x_2^2 \leq 6 \\ 2x_1 + x_2 \geq -1 \end{array} \right. \right\}.$$

The solution to the problem is $x^* = (-0.9412, 0.8824)$. The first and third constraints are binding at the solution. The multipliers for the solution are $(\lambda_1^*)_1 = 0.5404$, $(\lambda_1^*)_2 = 0$ and $(\lambda_1^*)_3 = 2.344$. The formula (17) gives $t^* = 0.0296$. Starting from the initial point $(-0.8, 0.8)$ and letting $t_k = t^*$ for all k , the algorithm converges to the solution in 3 iterations. The choice $t_k = 1/k$ also gives convergence in three iterations from the same starting point. Indeed, one can choose t_k to be any constant value in the set $\{0.001, 0.002, \dots, 0.999, 1\}$, and still find convergence in three iterations.

Example 4.2. This example is adapted from the example in [3]. Let

$$F(x) = \begin{pmatrix} 2x_1 + 0.2x_1^3 - 0.5x_2 + 0.1x_3 - 4 \\ -0.5x_1 + x_2 + 0.1x_2^3 + 0.5 \\ 0.5x_1 - 0.2x_2 + 2x_3 - 0.5 \end{pmatrix},$$

and let the set S be

$$S = \left\{ x \in \mathbb{R}^3 \left| \begin{array}{l} x_1^2 + 0.4x_2^2 + 0.6x_3^2 \leq 1 \\ 0.6x_1^2 + 0.4x_2^2 + x_3^2 \leq 1 \\ x_1 + x_2 + x_3 \geq \sqrt{3} \end{array} \right. \right\}.$$

The solution to the problem is $x^* = (0.9168, 0.4850, 0.3303)$. The first and third constraints are binding at the solution. The multipliers for the solution are $(\lambda_1^*)_1 = 1.9091$, $(\lambda_1^*)_2 = 0$ and $(\lambda_1^*)_3 = 1.2787$. The formula (17) gives $t^* =$

0.0629. Starting from the initial point $(0.9, 0.48, 0.33)$, the algorithm converges to the solution in 2 iterations. The choice $t_k = 1/k$ gives convergence in 4 iterations.

Example 4.3. This example is modified from the Suzuki Rosen problem in [11]. We changed the function F in that problem so that ∇F is positive definite. Let

$$F(x) = Ax - b,$$

where

$$A = \begin{pmatrix} 3.0006 & 0.0212 & 0.0141 & 0.0215 & 0.0088 \\ 0.0212 & 3.7093 & 0.4708 & 0.7193 & 0.2930 \\ 0.0141 & 0.4708 & 4.3125 & 0.4775 & 0.1945 \\ 0.0215 & 0.7193 & 0.4775 & 3.7295 & 0.2971 \\ 0.0088 & 0.2930 & 0.1945 & 0.2971 & 3.1210 \end{pmatrix},$$

and $b = (-1.5849, 15.8236, 13.1763, 12.0172, 138.7089)$. The set S is defined as

$$S = \left\{ x \in \mathbb{R}^5 \left| \begin{array}{l} x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 + x_5 \leq 0 \\ 4x_1^2 + 4x_2^2 + 5x_3^2 + 4x_4^2 - 2x_1 - 8x_2 - 18x_3 + 4x_4 + x_5 - 24 \leq 0 \\ 4x_1^2 + 7x_2^2 + 5x_3^2 + 7x_4^2 - 8x_1 - 5x_2 - 21x_3 + 4x_4 + x_5 - 30 \leq 0 \\ 7x_1^2 + 4x_2^2 + 5x_3^2 + x_4^2 + x_1 - 8x_2 - 21x_3 + 4x_4 + x_5 - 15 \leq 0 \\ -100 \leq x_i \leq 0, \quad i = 1, \dots, 5 \end{array} \right. \right\}.$$

The solution to the problem is $x^* = (0, 1, 2, -1, 44)$. The first, second and third constraints are binding at the solution. The multipliers for the solution are $(\lambda_1^*)_1 = \frac{1}{3}$, $(\lambda_1^*)_2 = \frac{1}{3}$ and $(\lambda_1^*)_3 = \frac{1}{3}$, and the remaining components of (λ_1^*) are zero. The formula (17) gives $t^* = 0.0317$. Starting from the initial solution $(0, 1.2, 2, -1, 44.2)$, the algorithm converges to the solution in 4 iterations. The sequence $t_k = 1/k$ gives convergence in 8 iterations.

5 Conclusions

This paper proves the local linear convergence of an outer approximation projection method for solving variational inequalities. Such convergence is guaranteed, if the function that defines the variational inequality has a positive definite Jacobian matrix at the solution, the LICQ holds there, and the step lengths are sufficiently small. Most existing projection based algorithms for solving variational inequalities use step lengths converging to zero. This leads to very small steps when the iterates are close to the solution. As our results show, the step lengths need not converge to zero. Using longer step lengths makes it possible to converge in fewer iterations.

In practice, one does not know in advance if the constraint qualification holds at the solution, or if the iterate is close enough to the solution. As a general guideline, when two iterates are close, and the Jacobian matrix of the function defining the variational inequality is positive definite at the current iterate, we can use the formula (17) with information given by the current iterate, to estimate an upper bound for step lengths in the remaining iterations.

References

1. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Series in Operations Research. Springer-Verlag, New York. Published in two volumes, paginated continuously (2003)
2. Fukushima, M.: An outer approximation algorithm for solving general convex programs. *Operations Research* 31, 101–113 (1983)

3. Fukushima, M.: A relaxed projection method for variational inequalities. *Mathematical Programming* 35, 58–70 (1986)
4. Taji, K., Fukushima, M.: A new merit function and a successive quadratic programming algorithm for variational inequality problems. *SIAM Journal on Optimization* 6, 704–713 (1996)
5. Censor, Y., Gibali, A.: Projections onto super-half-spaces for monotone variational inequality problems in finite-dimensional space. *Journal of Non-linear and Convex Analysis* 9, 461–475 (2008)
6. Yang, Q.: On variable-step relaxed projection algorithm for variational inequalities. *Journal of Mathematical Analysis and Applications* 302, 166–179 (2005)
7. Cruz, Y., Iusem, A.: Convergence of direct methods for paramonotone variational inequalities. *Computational Optimization and Applications* 46, 243–263 (2010)
8. Robinson, S.M.: An implicit-function theorem for a class of nonsmooth functions. *Mathematics of Operations Research* 16, 292–309 (1991)
9. Robinson, S.M.: Sensitivity analysis of variational inequalities by normal-map techniques. In: F. Giannessi, A. Maugeri (eds.) *Variational Inequalities and Network Equilibrium Problems*, 257–269. Plenum Press, New York (1995)
10. Robinson, S.M.: Strongly regular generalized equations. *Mathematics of Operations Research* 5, 43–62 (1980)
11. Rosen, J., Suzuki, S.: Construction of non-linear programming test problems. *Communications of the ACM* 8, 113 (1965)