OPERATIONS RESEARCH CENTER

Working Paper

On the Efficient Solution of Variational Inequalities; Complexity and Computational Efficiency

by

G. Perakis
M. Zaretsky

0R 360-02 **January 2002**

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

On the efficient solution of variational inequalities; complexity and computational efficiency *

G. Perakis † M. Zaretsky ‡

January, 2002

Abstract

In this paper we combine ideas from cutting plane and interior point methods in order to solve variational inequality problems efficiently. In particular, we introduce a general framework that incorporates nonlinear as well as linear "smarter" cuts. These cuts utilize second order information on the problem through the use of a gap function. We establish convergence as well as complexity results for this framework. Moreover, in order to devise more practical methods, we consider an affine scaling method as it applies to symmetric, monotone variational inequality problems and demonstrate its convergence. Finally, in order to further improve the computational efficiency of the methods in this paper, we combine the cutting plane approach with the affine scaling approach.

Key Words: Variational inequalities, Interior-point methods, Affine Scaling, Cutting Plane Methods
AMS Subject Classifications: Primary 90C06; Secondary 90C25

1 Introduction

Variational inequality problems (VIPs) arise frequently in a variety of applications that range from transportation and telecommunications to finance and economics. Moreover, variational inequalities provide a unifying framework for studying a number of important mathematical programming problems including equilibrium, minimax, saddle point, complementarity and optimization problems. As a result, variational inequalities have been the subject of extensive research over the past forty years. In particular, a variational inequality problem seeks a point

$$x^* \in K \text{ such that } f(x^*)'(x - x^*) \geq 0, \text{ for all } x \in K, \tag{1}$$

where $K \subseteq \mathbb{R}^n$ is the ground set and $f : K \to \mathbb{R}^n$ is the problem function.

*Acknowledgments: We would like to acknowledge the PECASE Award DMI-9984339 from the National Science Foundation, the Charles Reed Faculty Initiative Fund, the New England University Transportation Research Grant and the Singapore-MIT Alliance Program for supporting this research.
†Sloan School of Management and Operations Research Center, E53-359, MIT, Cambridge MA 02139, georgiap@mit.edu
‡Operations Research Center, MIT, Cambridge MA 02139, mzar@mit.edu

In this paper, we focus on devising methods for solving variational inequalities efficiently. To achieve this, we explore ideas from cutting plane methods as well as ideas from interior point methods.

Cutting plane methods have been used extensively in the literature for solving optimization problems. As applied to variational inequality problems, these methods include among others the ellipsoid method by Lüthi [6], the general geometric framework by Magnanti and Perakis [8], analytic center methods (see, for example, Goffin et. al. [4]), and barrier methods (see, for example, Nesterov and Nemirovskiy [11], and Nesterov and Vial [12]). Cutting plane methods incorporate several types of cuts: linear cuts (see, for example, Goffin et. al. [4]), quadratic cuts (see, for example, Lüthi and Büeler [7], Denault and Goffin [1]) and nonlinear cuts (see, for example, Nesterov and Nemirovskiy [12]). These methods converge to a solution when the problem function satisfies some form of a monotonicity condition. Moreover, for several of these methods, researchers have established complexity results (see, for example, [8], [12]). In this paper, we will consider an extension of the general geometric framework [8] that also incorporates nonlinear cuts. Moreover, this paper focuses on choices of "smarter" linear cuts. The motivation comes from the fact that this framework encapsulates several well-known methods for solving optimization problems such as the ellipsoid method, the volumetric center method, and the method of centers of gravity. Furthermore, an additional motivation comes from the fact that complexity bounds have been established for this framework.

Although one can establish complexity bounds for cutting plane methods, these methods are often computationally expensive in practice. This is due to the fact that the complexity bounds established are often tight in practice but also due to the fact that most cutting plane methods require the computation of a "nice" set and its "center" at each iteration. As a result, we also consider alternate methods for solving variational inequalities such as interior point methods. The motivation in considering this class of methods comes from the observation that it has been successful in solving linear optimization problems in practice. Moreover, for several methods in this class, researchers have established complexity bounds. These observations make these methods attractive for solving other problem classes as well. In particular, in this paper we consider versions of the affine scaling method. This method was originally developed for solving linear optimization problems by Dikin in 1967 [2]. Subsequently, Ye [16], [17] and more recently Tseng [14] extended this method for solving quadratic optimization problems, and Sun [13], for solving convex optimization problems. Our motivation in studying this method as it applies to variational inequalities, comes from its simplicity. Moreover, the version of the method that we introduce in this paper is motivated from the Frank-Wolfe method [3]. This method is widely used by transportation practitioners.

Our goal in this paper is twofold; (a) to develop results towards the efficient solution of a larger class of variational inequalities in theory (i.e., establish complexity bounds under weak conditions) as well as (b) develop results towards the efficient solution of variational inequalities in practice (i.e., develop methods that are computationally efficient). To achieve the first goal we extend the general geometric framework by Magnanti and Perakis [8], by considering linear cuts with "smarter" directions. To achieve the second goal, we introduce an affine scaling method for solving symmetric monotone VIPs (i.e., convex optimization problems) and establish its convergence. Moreover, we combine the two approaches and provide some computational results.

In summary, this paper contributes to the existing body of literature by

- proposing methods for solving a larger class of variational inequality problems efficiently;
- presenting polynomially convergent methods;
- proposing more practical versions of these methods that are easy to perform computationally.

The paper is organized as follows: in the remainder of this section we provide some background and describe some useful concepts. In Section 2, we introduce an extension of the general geometric framework by considering general nonlinear cuts. We prove its convergence under rather general conditions. In Section 3, we focus on linear cuts by considering "smarter" choices for the directions of the linear cuts. In Section 4, we establish complexity results. In Section 5, we focus on the solution of symmetric, monotone VIPs, by presenting an affine scaling method and providing convergence results. In Section 6, we introduce the cut ideas we proposed in the previous sections in the affine scaling method. This allows us to suggest schemes that are more tractable computationally. We also provide some preliminary computational results. In Section 8, we summarize our conclusions.

1.1 Preliminaries

In this section we review some basic definitions. Notice that in Appendix B we summarize the notation, in Appendix C we review all the basic definitions and finally, in Appendix D we summarize the assumptions we will use throughout this paper.

1.1.1 Weak Variational Inequalities

In this subsection we define the notion of a weak variational inequality problem and relate it to a variational inequality problem.

Definition 1 A point $\bar{x} \in K$ is a *weak VIP solution* if for all $x \in K$, $f(x)'(x - \bar{x}) \geq 0$. We will refer to the problem that seeks a weak VIP solution as a *weak variational inequality problem* ($WVIP$).

Definition 2

1. A function f is *quasimonotone* on K if $f(y)'(x-y) > 0$ implies that $f(x)'(x-y) \geq 0$, for all $x, y \in K$.
2. A function f is *pseudomonotone* on K if $f(y)'(x-y) \geq 0$ implies that $f(x)'(x-y) \geq 0$, for all $x, y \in K$.
3. A function f is *monotone* if $(f(x) - f(y))'(x-y) \geq 0$, for all $x, y \in K$.

Assumption 1 *K is a closed, bounded and convex set with a nonempty interior.*

In particular, through Assumption 1 we assume that there exist positive constants L and l such that the feasible region is contained in a ball of radius 2^L and, in turn, contains a ball of radius 2^{-l} (see [8] for a further discussion on how these constants can be explicitly defined for a polyhedral feasible region).

Lemma 1 *When the problem function f is continuous, a VIP is equivalent to a $WVIP$ if one of the following conditions holds:*

(a) The underlying problem function f is quasimonotone and for some $y \in K$, $f(x^*)'(y - x^*) > 0$;

(b) The underlying problem function f is pseudomonotone.

Proof. It is easy to see that when the problem function f is continuous, every solution x^* of a $WVIP$ is also a VIP solution. Next we assume that x^* is a VIP solution and show that it is also a $WVIP$ solution. Consider condition (a). Suppose that $f(\bar{x})'(\bar{x} - x^*) < 0$, for some $\bar{x} \in K$. The quasimonotonicity of problem function f implies that $f(x^*)'(x^* - \bar{x}) \geq 0$. Condition (a) then implies that for some $y \in K$, $f(x^*)'(y - \bar{x}) > 0$. Let $x_t = y + t(\bar{x} - y)$, since $f(x^*)'(y - \bar{x}) > 0$, then $f(x^*)'(x_t - \bar{x}) = (1-t) f(x^*)'(y - \bar{x}) > 0$, for all $t \in (0,1)$. This, in turn, implies that $f(x_t)'(x_t - x^*) \geq 0$, for all $t \in (0,1)$ (due to quasimonotonicity). Therefore, as $t \to 1$, the continuity of problem function f implies that $f(\bar{x})'(\bar{x} - x^*) \geq 0$. This contradicts our initial assumption.

In the case of condition (b), the pseudomonotonicity of problem function f implies that if $f(x^*)'(x - x^*) \geq 0$, for all $x \in K$, then $f(x)'(x - x^*) \geq 0$, for all $x \in K$. ■

1.1.2 Gap Functions

In what follows we notice that VIP and $WVIP$ solutions can be characterized through appropriate gap functions.

Definition 3 We define function $C_p(y) = \max_{z \in K} f(y)'(y - z)$ as the *primal gap function* and function $C_d(y) = \max_{z \in K} f(z)'(y - z)$ as the *dual gap function.*

It is well known that the VIP solution set X^* consists of points $\arg\min_{x \in K} C_p(x)$, i.e., $X^* = \{x \mid C_p(x) = 0\}$. Moreover, gap function C_d is a closed convex function, that is strictly positive outside the solution set. The solution set of a $WVIP$, coincides with the set $\arg\min_{x \in K} C_d(x) = \{x \mid C_d(x) = 0\}$. Finally, we note that when the conditions of Lemma 1 are satisfied, then the VIP solution set coincides with the $WVIP$ solution set.

On the Theoretical Complexity of VIPs

2 An Extension of the General Geometric Framework

2.1 The General Geometric Framework

In this section we present an extension of the general geometric framework (GGF) originally proposed in [8]. A key notion in the GGF is that of a "nice" set. These "nice" sets are constructed at each iteration so that they have the following properties: (a) an approximation of a "nice" set and its center can be computed efficiently, (b) a "nice" sets contains the VIP (or $WVIP$) solution set, (c) its volume strictly decreases at each iteration. At each iteration, the GGF first computes a cut through the center of the "nice" set, and subsequently constructs a new "nice" set of smaller volume that contains the remainder of the feasible region as well as the VIP (or $WVIP$) solutions.

More formally, at iteration k, the GGF maintains the following: a "nice" set P^k, a set K^k that contains the solution set and coincides with the set $P^k \bigcap K$, and, finally, the iterate x^k, which is the center of P^k. Initially set P^0 is chosen so that it contains the whole feasible region K. Furthermore, we assume that we can construct "nice" sets P^k so that their volumes strictly decrease, that is, $\text{Vol}(P^{k+1}) \leq b(n)\text{Vol}(P^k)$, for some constant $0 < b(n) < 1$. As a result, the volume $\text{Vol}(K^k)$ also decreases and converges to zero. In what follows, we outline a more general version of the geometric framework studied in [8].

An Extension of the General Geometric Framework

1. Start with an interior point x^0 - center of a "nice" set $P^0 \supseteq K^0 = K$.

2. Feasibility cuts. At iteration k, given P^k, K^k, and x^k (center of P^k):
   (i) Compute a surface $\bar{S}\left(x^k\right)$ that supports K. Moreover, we denote by $S\left(x^k\right)$ the set lying below this surface.
   (ii) Update P^{k+1}, so that $P^{k+1} \supseteq S(x^k) \cap P^k$, $K^{k+1} = K^k$.

3. Optimality cuts. At iteration k, given P^k, K^k, and x^k (center of P^k):
   (i) Compute a surface $\bar{S}\left(x^k\right)$ that cuts K^k through x^k. Moreover, the surface is such that the set $S\left(x^k\right)$ lying below this surface, contains all the VIP (or $WVIP$) solutions.
   (ii) Update $K^{k+1} = K^k \cap S(x^k)$, $P^{k+1} \supseteq S(x^k) \cap P^k$.
   (iii) Repeat steps 2 and 3 until a desirable precision is reached.

In Magnanti and Perakis [8], the surface $\bar{S}(x^k)$ is a hyperpane $H(x^k)$ determined through a linear cut with slope $f(x^k)$. In Section 3 of this paper, we will consider alternate choices for the slopes of these cuts.

The extension of the previous framework allows us to incorporate nonlinear cuts. For example, an obvious but perhaps not very practical choice of a cut, could be to introduce at each step k, a nonlinear cut $S(x^k) = \{x \mid f(x)'(x - x^k) \leq 0\}$. In some special cases of the GGF (such as the ellipsoid method [6], [8]), it is important that the sets K^k employed in the algorithm, preserve some properties (for example, convexity and/or connectivity). In these cases notice that if at each step of the GGF we use a cut determined through a surface $\bar{S}_g = \{x \mid g(x) = 0\}$, where g is a quasiconvex function on K, then the new set $K \cap S_g$, where $S_g = \{x \mid g(x) \leq 0\}$, remains convex. This leads us to conclude that quasiconvex cuts preserve the convexity as well as the connectivity of the feasible region. In particular, when the problem function f is strongly monotone, i.e., there exists for some $\alpha > 0$, such that $\left(f(x) - f(y)\right)'(x - y) \geq \alpha \|x - y\|^2$, for all x, y, we can consider quadratic cuts in the GGF. An example includes a cut where the set lying below the quadratic surface is determined by $S\left(x^k\right) = \left\{x \mid f\left(x^k\right)'(x - x^k) + \alpha\|x - x^k\|^2 \leq 0\right\}$. Lüthi and Büeler [7] have considered quadratic cuts of this type.

2.2 The Convergence of the GGF

The analysis of the framework in this paper relies on the observation that we can measure the "closeness" of a point $x \in K$ to a VIP (or a $WVIP$) solution using some function G. Examples of such functions include the primal or the dual gap function we described in Subsection 1.2.2. For a variational inequality problem with a symmetric Jacobian matrix ∇f, an alternate choice could be the corresponding objective function F (that is, when $\nabla F = f$). In what follows, we examine how the properties of the level sets of such a function G imply the convergence of the GGF. In particular, the following assumptions summarize the key properties.

Assumption 2 *Let X^* be the VIP (or, depending on the context, the $WVIP$) solution set. There exists a function $G : K \to \mathbb{R}^+$ such that $x^* \in \arg\min_{x \in K} G(x)$ if and only if $x^* \in X^*$.*

We also assume that the set $S(x)$ lying below the cutting surface at point x, contains some level set $L_\alpha = \{z \mid G(z) \leq \alpha\}$, for some $\alpha > 0$, i.e., for some $\alpha > 0$, $S(x) \supseteq L_\alpha$.

Examples of sets $S(x)$, include the half space $S(x) = H(x) = \{z \mid a(x)'z \leq b(x)\}$ or the set $S(x) = \{z \mid a(x)'z + z'Q(x)z \leq b(x)\}$, where $Q(x)$ is a positive semi-definite matrix. Notice that in both examples, set $S(x)$ is a convex set.

Assumption 3 *Given a point $y \in K$, and a small enough $\varepsilon > 0$, such that $\min_{x \in X^*} \|x - y\| \leq \varepsilon$, it follows that $G(y) \leq c(\varepsilon)$, where $\lim_{\varepsilon \to 0} c(\varepsilon) = 0$.*

Notice that the previous assumption seems to imply that function G is locally continuous close to the solution set.

In what follows, we establish the convergence of the GGF when the variational inequality problem satisfies Assumptions 1 - 3.

Theorem 1 *Consider the sequence $\{x^k\}$ induced by the GGF. At each iteration we introduce a cut through a set $S(x^k)$. Under Assumptions 1 - 3, every limit point of the sequence $\{x^k\}$ is a VIP solution.*

Proof. We first examine the properties of the sequences $\{x^k\}$ and $\{K^k\}$. Assumptions 1 - 3 imply that at the kth iteration, the framework adds a cut S^k, such that for some α^k, $\{z \in K \mid G(z) \leq \alpha^k\} \subseteq S^k$. We note that Assumption 2 implies that the solution set $X^* \subseteq \{z \in K \mid G(z) \leq \alpha^k\} \subseteq S^k$. Since we cut through x^k, it follows that $P^k \not\subseteq S^k$. Moreover, the description of the algorithm implies that $\text{Vol}(K^k) \to 0$, $x^* \in K^k$ and $\text{Vol}(K^k) > 0$, for all k. Let $\bar{K}$ denote the set $\lim_{k \to \infty} K^k$ and $\bar{K}^c$, its complement. Every limit point of the sequence $\{x^k\}$ belongs to the set $\bar{K}$.

Assumption 2 implies that $X^* \in \bar{K}$. Therefore we only need to show that every point $\bar{x} \in \bar{K}$ is a VIP solution. Since $\text{Vol}(K^k) \to 0$ and $\text{Vol}(K^k) > 0$, there exists a point in $\bar{K}^c \cap K$ that is arbitrarily close to some VIP solution x^*. According to Assumption 3, for small enough $\varepsilon > 0$, we can choose this point to be $y^\varepsilon \in \bar{K}^c$ such that $\|y^\varepsilon - x^*\| \leq \varepsilon$

and $G(y^\varepsilon) < c(\varepsilon)$, where $x^* \in X^*$. Since $y^\varepsilon \in \bar{K}^c$, it follows that $G(\bar{x}) < G(y^\varepsilon) \leq c(\varepsilon)$. By letting ε go to zero, we conclude that $G(\bar{x}) = 0$. Consequently, Assumption 2 implies that $\bar{x}$ is a VIP solution. This further implies that every point in the limiting set $\bar{K}$ is a VIP solution. Therefore, we conclude that $\bar{K}$ is the solution set of the VIP.
■

In the next section, we focus on linear cuts and consider particular choices for these cuts. We show that these cuts satisfy conditions discussed in this section for an appropriate function G. As a result we show that the GGF is convergent for these choices of cuts.

3 The GGF with Linear Cuts

3.1 Linear Cuts via an Exact Gap Function

In the previous section we considered an extension version of the GGF that incorporated cuts with general nonlinear surfaces $\bar{S}(x^k)$. Examples included linear and quadratic cuts. In what follows we will focus on linear cuts, that is, cuts of the form $H^k = \{z \mid {a^k}'(z - x^k) \leq 0\}$, where a^k is the slope of the cut. In this section we consider particular choices for the slopes of these cuts. We first introduce some notation. Let

$$y_x = \arg\max_{y \in K} f(y)'(x - y), \tag{2}$$

$y^k = \arg\max_{y \in K} f(y)'(x^k - y)$, and $Cut(y, x) = \{z \in K \mid f(y)'(z - x) \leq 0\}$ the half space through point x with slope $f(y)$.

In what follows we consider the GGF with cuts $S(x) = Cut(y_x, x)$. Notice that this modification of the GGF determines the direction of the cut using information from the dual gap function C_d at the current iterate.

Remark: Linear cuts of the form $Cut(x, x)$ are often considered in the literature (see, for example, [7] or [8]).

Assumption 4 *f is a bounded function. That is, for some $M > 0, \|f(x)\| \leq M$, for any $x \in K$.*

Theorem 2 *Suppose that a $WVIP$ satisfies Assumptions 1, 4. Let $\{x^k\}$ be the sequence induced by the GGF with cuts $Cut(y^k, x^k)$. Every limit point of $\{x^k\}$ is a $WVIP$ solution.*

Proof. The dual gap function $C_d(x)$ is the function we will employ in order to measure the closeness of point $x \in K$ to the solution set of the $WVIP$. We will show that for this function, Assumptions 2 and 3 hold. Then the result follows from Theorem 1.

In this theorem, we denote by X^* the $WVIP$ solution set. Since set X^* coincides with set $\{z \mid C_d(z) = 0\}$, in order to prove that Assumption 2 is valid for $G(x) = C_d(x)$, we need to show that the level sets of C_d are contained in half spaces. Wlog, it is sufficient to show that $\{z \in K \mid C_d(z) \leq C_d(x^k)\} \subseteq Cut(y^k, x^k)$. The following are consequences of the definition of the gap function C_d:

(a) $C_d(x) \leq C_d(x^k) \Leftrightarrow f(y_x)'(x - y_x) \leq f(y^k)'(x^k - y^k)$,

(b) $f(y_x)'(x - y_x) \geq f(y^k)'(x - y^k)$, since $y^k \in K$ and $y_x = \arg\max_{y \in K} f(y)'(x - y)$.

Therefore, if $x \in \{z \in K \mid C_d(z) \leq C_d(x^k)\}$, then

$$\begin{aligned} f(y^k)'(x - x^k) &= f(y^k)'(x - y^k) + f(y^k)'(y^k - x^k) \\ &\leq f(y_x)'(x - y_x) + f(y^k)'(y^k - x^k), \text{ (using (b))} \\ &\leq f(y^k)'(x^k - y^k) + f(y^k)'(y^k - x^k), \text{ (using (a))} \\ &= 0. \end{aligned}$$

This leads us to conclude that $x \in Cut(y^k, x^k)$, that is, Assumption 2 holds.

Assumption 3 holds as well. Indeed,

$$C_d(x) = f(y_x)'(x - y_x) = f(y_x)'(x^* - y_x) + f(y_x)'(x - x^*) \leq f(y_x)'(x - x^*). \blacksquare$$

Remarks:

1. This scheme induces a sequence $\{x^k\}$ whose limit points are $WVIP$ solutions. To prove this we required the feasible region to be a closed and convex set, with a nonempty interior and the problem function f to have a bounded norm. Notice that if in addition, the problem function satisfies some form of a quasimonotonicity condition, then these limit points are also VIP solutions.

2. Nevertheless, in this scheme determining the slope of a cut can be computationally expensive. This is due to

the nonlinearity and, perhaps, even the nonconvexity of the subproblem that generates point y^k. As a result, in the next section we consider schemes that compute approximations of the direction $f(y^k)$, yet generating sequences that converge to a $WVIP$ solution.

3.2 Approximation Schemes

3.2.1 Motivation

To motivate the schemes we introduce in this section, we first consider an approximation of the gap function C_d. In particular, given a point $y \in K$, if we apply the mean value theorem on function $f(\cdot)'(y-x)$ around point $x \in K$, it follows that $f(y)'(y-x) = f(x)'(y-x) + (y-x)'\nabla f(z)'(y-x)$, for some $z \in [x;y]$. As a result, subproblem (2) can be rewritten as $y_x = \arg\max_{y \in K} f(y)'(x-y) = \arg\max_{y \in K} \left(f(x)'(x-y) - (y-x)'\nabla f(z)(y-x)\right)$, for some $z \in [x;y]$. Motivated by this observation, we will consider cuts whose directions will be determined from this approximation. To develop this observation more formally, we first need to impose an additional assumption on the Jacobian matrix.

Definition 4 A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ has the property of *Jacobian similarity* if the Jacobian matrix is positive semi-definite and there exists some constant $\rho \geq 1$ such that the Jacobian matrix satisfies $d'\nabla f(y)d \leq \rho d'\nabla f(x)d$, for all $x, y \in K$, and $d \in \mathbb{R}^n$.

Assumption 5 *Problem function f satisfies the Jacobian similarity property.*

Remarks:

1. In the context of convex optimization, there is an analogous property called *Hessian similarity* (see [13]).

2. The property of Jacobian similarity holds for monotone functions with a bounded Jacobian matrix. This property is similar to the property of self-concordance as it applies to barrier functions. In the context of nonlinear optimization, Nemirovskiy and Nesterov [11] have shown that if the self-concordance property holds on a barrier function then the property of Hessian similarity also holds locally.

3. Moreover, when the problem function f is strictly monotone, one can choose the Jacobian similarity constant ρ as the ratio of the eigenvalues of the symmetrized Jacobian matrix $\frac{\nabla f^t + \nabla f}{2}$, that is, $\rho = \frac{\sup_{y \in K} \lambda_{\max}(y)}{\inf_{x \in K} \lambda_{\min}(x)}$. Finally, notice that when the problem function f is affine, then we can choose $\rho = 1$.

We denote by $H(x) = \frac{\nabla f(x) + \nabla f(x)'}{2}$ the symmetrized Jacobian matrix of f and let $\|w\|^2_{H(x)} = w'H(x)w$. Assumption 5 then implies that $f(y)'(y-x) = f(x)'(y-x) + (y-x)'\nabla f(z)'(y-x) \leq f(x)'(y-x) + \rho \|x-y\|^2_{H(x)}$, where $z \in [x; y]$. Therefore, $f(y)'(x-y) \geq f(x)'(x-y) - \rho \|x-y\|^2_{H(x)}$.

3.2.2 Approximation Scheme 1

The previous discussion motivates us to consider the GGF with cuts whose directions are determined by solving a quadratic approximation of subproblem (2). In particular, we approximate the gap function $C_d(x)$ with its quadratic approximation

$$C_d^1(x) = \max_{y \in K} \left(f(x)'(x-y) - \rho \|x-y\|^2_{H(x)} \right).$$

Let point y_x^1 be the maximizer in this quadratic approximation of $C_d(x)$. At each iteration, Approximation Scheme 1 introduces a cut defined by $Cut\left(y^{k,1}, x^k\right) = \left\{ z \mid f\left(y^{k,1}\right)'\left(z - x^k\right) \leq 0 \right\}$, where point x^k is the current iterate of the framework, while point $y^{k,1} = \arg\max_{y \in K} C_d^1(x^k)$. In order to illustrate the convergence of this modified GGF, we first need to prove the following two propositions.

Proposition 1 *Under Assumption 5,* $\left\{ x \in K \mid C_d(x) \leq C_d^1\left(x^k\right) \right\} \subseteq \left\{ x \in K \mid f(y^{k,1})'(x - x^k) \leq 0 \right\}$.

Proof. Notice that inequality $C_d(x) \leq C_d^1\left(x^k\right)$ holds if and only if for all $y \in K$,

$$f(y)'(x-y) \leq f(x^k)'(x^k - y^{k,1}) - \rho \left\|x^k - y^{k,1}\right\|^2_{H(x^k)}. \tag{3}$$

An application of the mean value theorem together with Assumption 5 (i.e., the Jacobian similarity property) imply that there exists a point $z^k \in \left[x^k; y^{k,1}\right]$ such that

$$f(x^k)'(x^k - y^{k,1}) - \left\|x^k - y^{k,1}\right\|^2_{H(z^k)} = f(y^{k,1})'(x^k - y^{k,1}). \tag{4}$$

Therefore, if x is such that $C_d(x) \leq C_d^1(x)$, then

$$\begin{aligned} f(y^{k,1})'(x-x^k) &= f(y^{k,1})'(x-y^{k,1}) + f(y^{k,1})'(y^{k,1}-x^k) \\ &\leq f(x^k)'(x^k-y^{k,1}) - \rho \left\|x^k - y^{k,1}\right\|^2_{H(x^k)} + f(y^{k,1})'(y^{k,1}-x^k), \text{ (using (3))} \\ &\leq f(x^k)'(x^k-y^{k,1}) - \left\|x^k - y^{k,1}\right\|^2_{H(z^k)} + f(y^{k,1})'(y^{k,1}-x^k), \text{ (using Assumption 5)} \\ &= f(y^{k,1})'(x^k-y^{k,1}) + f(y^{k,1})'(y^{k,1}-x^k), \text{ (using (4))} \\ &= 0. \blacksquare \end{aligned}$$

Proposition 2 *Under Assumptions 1 and 5, the WVIP solution set X^* coincides with the set of minimizers of $C_d^1(x)$.*

Proof. Let x^* be a *WVIP* solution and y_x, y_x^1 and $y^{k,1}$ be the points defined above. Observe that $C_d^1(x) \geq 0$, therefore $f(x^k)'(x^k - y^{k,1}) \geq \rho \left\|x^k - y^{k,1}\right\|^2_{H(x^k)}$. Since $C_d(x^*) = 0 \leq C_d^1(x)$, for all $x \in K$, Proposition 1 implies that $x^* \in \left\{x \in K \mid f(y^{k,1})'(x - x^k) \leq 0\right\}$.

It what follows we show that $C_d(x) \geq C_d^1(x)$. From the definitions of points y_x and y_x^1, it follows that $C_d(x) = f(y_x)'(x - y_x)$ and $C_d^1(x) = f(x)'(x - y_x^1) - \rho \left\|x - y_x^1\right\|^2_{H(x)}$ respectively. Therefore,

$$\begin{aligned} C_d(x) &= f(y_x)'(x-y_x) = \arg\max_{y \in K} f(y)'(x-y) \geq f(y_x^1)'(x-y_x^1) \\ &= f(x)'(x-y_x^1) - \left\|x-y_x^1\right\|^2_{H(z)}, \text{ (for some } z \in [x; y_x^1]) \\ &\geq f(x)'(x-y_x^1) - \rho \left\|x-y_x^1\right\|^2_{H(x)} = C_d^1(x). \end{aligned}$$

Since $C_d^1(x^*) \geq 0 = C_d(x^*)$, the previous inequality implies that $C_d^1(x^*) = 0$.

So far we have shown that any *WVIP* solution minimizes function $C_d^1(x)$. Next we show that every $\bar{x} \in \arg\min_{x \in K} C_d^1(x)$ is also a *WVIP* solution. Suppose, the opposite is true, that is, for every such $\bar{x}$, there exists some y such that $f(y)'(y - \bar{x}) < 0$. Applying the mean value theorem and the Jacobian similarity property, it follows that $f(\bar{x})'(\bar{x}-y) = f(y)'(\bar{x}-y) + \|y-\bar{x}\|^2_{H(\tilde{z})} \geq f(y)'(\bar{x}-y) + \frac{1}{\rho}\|y-\bar{x}\|^2_{H(\bar{x})}$, for some $\tilde{z} \in [\bar{x}; y]$. Consider $\alpha = 1/\rho^2 \in (0, 1]$.

Then for $y_\alpha = \bar{x} + \alpha(y - \bar{x}) \in K$, it follows that

$$f(\bar{x})'(\bar{x} - y_\alpha) - \rho \, ||\bar{x} - y_\alpha||^2_{H(\bar{x})} = \alpha f(y)'(\bar{x} - y) - \frac{\alpha}{\rho}(\rho^2\alpha - 1) \, ||\bar{x} - y||^2_{H(\bar{x}))} > 0.$$

However, we argued above that $C_d^1(\bar{x}) = 0$, for all $\bar{x} \in \arg\min_{x \in K} C_d^1(x)$. This leads to a contradiction. ■

We are now ready to prove the convergence of the sequence induced by Approximation Scheme 1.

Theorem 3 *Consider a $WVIP$ satisfying Assumptions 1 and 5. Let $\{x^k\}$ be the sequence induced by the GGF with cuts $Cut(y^{k,1}, x^k)$. Then every limit point of the sequence $\{x^k\}$ is a $WVIP$ solution.*

Proof. The proof is similar to that of Theorem 2. Moreover, notice that in Propositions 1 and 2 we have shown that Assumption 2 holds. Finally, in Proposition 2 we proved that $C_d^1(x) \leq C_d(x)$. Therefore, Assumption 3 also holds. ■

Remark: Under the Jacobian similarity property, the set of minimizers of gap function C_d^1 also coincides with the VIP solution set. This follows from the observation that for any VIP solution x^*, $f(x^*)'(x^* - y) - \rho \, ||x^* - y||^2_{H(x^*)} \leq 0$. This in turn implies that $C_d^1(x^*) = 0$. Just as in the proof of Proposition 2, it follows that every minimizer of C_d^1 is a VIP solution. In other words, under the Jacobian similarity property, the $WVIP$ solution set coincides with the VIP solution set. Therefore Approximation Scheme 1 also computes a VIP solution.

3.2.3 Approximation Scheme 2

Approximation Scheme 1 considered a quadratic approximation of the dual gap function C_d. Although this approximation simplified the objective function in the dual gap function computation, it did not concern itself with the structure of the feasible region K. In what follows we consider a polyhedral feasible region K of the form $\{x \mid Ax = b, \; x \geq 0\}$.

Assumption 6 *Matrix A has a full row rank.*

Assumption 7 *Matrix AX^2A' is invertible for all $x \in K$.*

In what follows, we consider an approximation of the gap function C_d (similar to C_d^1) by also restricting the maximization problem over a Dikin ellipsoid rather than maximizing over the entire feasible region K. We denote a

Dikin ellipsoid by $D(x) = \{y \in K \mid \|X^{-1}(y-x)\| \leq r\}$, where matrix $X = diag(x)$ and constant $r \in (0,1)$. We then define

$$C_d^2(x) = \max_y \left\{ f(x)'(x-y) - \rho \|x-y\|_{H(x)}^2 \mid Ay = b, \| X^{-1}(y-x)\| \leq r \right\}.$$

Notice that under Assumptions 6 and 7, the function $C_d^2(x)$ is well defined for every x in the polyhedron K. Moreover, the computation of point y_x^2 in the definition of $C_d^2(x)$, when x lies in the the interior of K, can be performed in polynomial time (see [14], [17]).

Let $y^{k,2}$ be a maximizer in the definition of $C_d^2(x^k)$, then Approximation Scheme 2 introduces at each iteration k, cuts $Cut(y^{k,2}, x^k)$ in the GGF. The convergence proof for Approximation Scheme 2 is completely analogous to the proof for Approximation Scheme 1. Therefore, for the sake of brevity, in what follows we outline the key properties, omitting the proofs.

Theorem 4 *Suppose that a WVIP satisfies Assumptions 1, 5-7. Let $\{x^k\}$ be the sequence induced by the GGF with cuts $Cut(y^{k,2}, x^k)$. Then every limit point of the sequence $\{x^k\}$ is a WVIP (and VIP) solution.*

Proof.

1. $\{x \in K \mid C_d(x) \leq C_d^2(x^k)\} \subseteq \{x \in K \mid f(y^{k,2})'(x-x^k) \leq 0\}$.
2. $C_d^2(x) \geq 0$; $C_d(x) \geq C_d^2(x)$.
3. $C_d^2(x^*) = 0$; x^* is a $WVIP$ solution if and only if $C_d^2(x^*) = 0$.

The proof that every limit point of $\{x^k\}$ is a $WVIP$ solution follows, as before, from Theorem 1. Moreover, the Jacobian similarity property (that is, Assumption 5) implies that every limit point is also a VIP solution. ■

4 Complexity Analysis

So far we have presented an extension of the GGF for solving $WVIP$s by introducing at each step linear cuts with slopes that exploited second order information of the problem function. We established the convergence of these methods to $WVIP$ solutions under rather weak assumptions. Moreover, we established the convergence of these methods to VIP solutions under a quasimonotonicity type of condition on the problem function. In what follows, we provide complexity results for the previous schemes.

4.1 Preliminaries and Key Properties

We start the analysis by describing the notion of an approximate solution in the context of a variational inequality as well as a weak variational inequality problem. The definitions we introduce use the gap function concepts described in Subsection 1.1.2.

Definition 5 For any $\varepsilon > 0$, a point $x^I \in K$ is an *ε-approximate* VIP solution if $C_p(x^I) \leq \varepsilon$, where C_p is the primal gap function.

Definition 6 For any $\varepsilon > 0$, a point $x^{II} \in K$ is an *ε-approximate* $WVIP$ solution if $C_d\left(x^{II}\right) \leq \varepsilon$, where C_d is the dual gap function.

Remark: Notice that we can also state scale-invariant versions of Definitions 5 and 6. For example, a point $x^I \in K$ is an *ε-approximate* VIP solution, if for any $\varepsilon > 0$, $C_p(x^I) \leq 2\varepsilon 2^{-l} M$. Constant M is defined in Assumption 4, and l is defined in Subsection 1.1. We can adjust the proofs in this section to be also applicable to this scale-invariant definition (see [8] for a more detailed discussion of these definitions).

At this point it is natural to ask when Definitions 5 and 6 become equivalent. First we notice that in Section 1 we introduced assumptions under which variational inequality and weak variational inequality problems have the same solutions. However, these results do not directly translate into the equivalence of the respective approximate solutions as we defined them above. As a result, in what follows we examine the relationship between approximate VIP and $WVIP$ solutions.

Definition 7 A function f is *Lipschitz continuous* with a Lipschitz constant $\lambda > 0$ if $d'\nabla f(x)d \leq \lambda d'd$, for all $d \in \mathbb{R}^n$.

Assumption 8 *The problem function f is Lipschitz continuous with Lipschitz constant λ.*

The next two propositions establish a connection between approximate VIP and $WVIP$ solutions.

Proposition 3 *If the VIP problem function f is monotone, then an ε-approximate VIP solution x^I is also an ε-approximate $WVIP$ solution.*

Proof. This result follows from the definitions of approximate solutions and monotone functions. ∎

Proposition 4 *Suppose that Assumptions 1 and 8 hold and L is the constant defined in Subsection 1.1. Then an ε-approximate $WVIP$ solution is also a $\sqrt{2^{2L+2}\lambda\varepsilon}$-approximate VIP solution.*

Proof. Suppose x^{II} is an ε-approximate $WVIP$ solution, i.e., for all $z \in K$, $f(z)'(x^{II}-z) \leq \varepsilon$. Then an application of the mean value theorem implies that for some $y \in [x;z]$, $f(x^{II})'(z-x^{II}) = f(z)'(z-x^{II}) - \left\|z-x^{II}\right\|^2_{H(y)} \geq -\varepsilon - \lambda\left\|z-x^{II}\right\|^2$ (*). For $\alpha \in (0,1]$, we now define point $z_\alpha = x^{II} + \alpha\left(x-x^{II}\right)$, for any $x \in K$. Then the convexity of set K, implies that point z_α lies in set K. It follows that $f(x^{II})'(z_\alpha - x^{II}) = \alpha f(x^{II})'(x-x^{II})$. Therefore, $f(x^{II})'(x-x^{II}) = \frac{1}{\alpha}f(x^{II})'(z_\alpha - x^{II})$. If we apply (*) for a choice of $z = z_\alpha$, it follows that $f(x^{II})'(x-x^{II}) \geq \frac{1}{\alpha}\left(-\varepsilon - \lambda\alpha^2\left\|x-x^{II}\right\|^2\right) \geq -\frac{1}{\alpha}\varepsilon - \lambda\alpha 2^{2L}$. Furthermore, notice that this inequality is true for any choice of $x \in K$.

Setting $\bar{\varepsilon}(\alpha) = \frac{1}{\alpha}\varepsilon + \lambda\alpha 2^{2L}$, we conclude that x^{II} is an $\bar{\varepsilon}(\alpha)-$approximate VIP solution. In particular, the choice of $\alpha = \sqrt{\frac{\varepsilon}{\lambda 2^{2L}}}$ minimizes $\bar{\varepsilon}(\alpha)$. Such an α gives rise to $\bar{\varepsilon} = \sqrt{2^{2L+2}\lambda\varepsilon}$. Therefore, point x^{II} is a $\sqrt{2^{2L+2}\lambda\varepsilon}$-approximate VIP solution. Also notice that, as $\frac{\varepsilon}{\alpha} \to 0$ and $\alpha \to 0$, it follows that $f(x^{II})'(x-x^{II}) \geq 0$, for all $x \in K$. ■

The next lemma, relates the approximate VIP solutions with the approximate $WVIP$ solutions of Approximation Schemes 1 and 2.

Lemma 2 *If Assumption 5 holds and $C_d^1(x) \leq \varepsilon$ (or $C_d^2(x) \leq \varepsilon$ and Assumption 6 holds), then $C_d(x) \leq \rho^2\varepsilon$, i.e., x is a $\rho^2\varepsilon$-approximate $WVIP$ solution.*

Proof. Suppose, on the contrary, that for some x, $C_d^1(x) \leq \varepsilon$, while $C_d(x) > \rho^2\varepsilon$. It follows that for some $\bar{y}$, $f(\bar{y})'(x^k - \bar{y}) > \rho^2\varepsilon$. Consider a point $y_\alpha = (1-\alpha)x + \alpha\bar{y}$, for some $\alpha \in [0,1]$. We define function $g(y_\alpha) \triangleq f(x)'(x-y_\alpha) - \rho\left\|x-y_\alpha\right\|^2_{H(x)}$. Then for some $z \in [x;\bar{y}]$,

$$\begin{aligned} g(y_\alpha) &= \alpha f(x)'(x-\bar{y}) - \alpha^2\rho\left\|x-\bar{y}\right\|^2_{H(x)} = \alpha\left(f(\bar{y})'(x-\bar{y}) + \left\|\bar{y}-x\right\|^2_{H(z)}\right) - \alpha^2\rho\left\|x-\bar{y}\right\|^2_{H(x)} \\ &> \alpha\left(\rho^2\varepsilon + \left\|\bar{y}-x\right\|^2_{H(z)}\right) - \alpha^2\rho\left\|x-\bar{y}\right\|^2_{H(x)} = \alpha\rho^2\varepsilon + \alpha\left\|\bar{y}-x\right\|^2_{H(z)} - \alpha^2\rho\left\|x-\bar{y}\right\|^2_{H(x)} \\ &\geq \alpha\rho^2\varepsilon - \frac{\alpha}{\rho}\left(1-\alpha\rho^2\right)\left\|x-\bar{y}\right\|^2_{H(x)}. \end{aligned}$$

Letting $\alpha = \frac{1}{\rho^2}$, we observe that $g(y_\alpha) > \varepsilon$. Hence, $C_d^1(x) = \max_{y\in K} g(y) > \varepsilon$. This contradicts the assumption that $C_d^1(x) \leq \varepsilon$. Similarly we can show that $C_d^2(x) \leq \varepsilon$ implies that $C_d(x) \leq \rho^2\varepsilon$. Notice that if the property of Jacobian similarity holds for some constant ρ, then it also holds for all $\varrho \geq \rho$. Therefore, if we assume that $C_d(x) > \rho^2\varepsilon$, then $g(y_\alpha) > \varepsilon$, for every $\alpha \leq \frac{1}{\rho^2}$ (where y_α is defined above). Since for a sufficiently small α, point y_α lies inside the Dikin ellipsoid $D(x)$, it also holds that $C_d^2(x) > \varepsilon$, (i.e $C_d(x) > \rho^2\varepsilon$ implies that $C_d^2(x) > \varepsilon$). ■

4.2 Complexity Bounds

In this subsection we establish complexity bounds for the schemes we introduced so far.

Assumption 9 $2l \geq \log n + 1; L_1 = L + 3l + \log(\frac{M}{\varepsilon})$.

We introduce a contraction map $T : K \to K$ defined as $T(x) = x^* + 2^{-L_1+l+\log n}(x - x^*)$, where x^* is some solution of the VIP. This is a one-to-one map between sets K and $T(K)$. Therefore, we can retrieve a point x from its image $y = T(x)$, as $x = x^* + 2^{L_1-l-\log n}(y - x^*)$.

Proposition 5 *Let $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$. Assumptions 1, 4, 9 imply that there is a point $y \in T(K) \cap P^{\bar{k}^c}$, where $P^{\bar{k}^c}$ is the complement of $P^{\bar{k}}$.*

Proof. Observe that the volume of set P^k is at most 2^{-nL_1} in $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ iterations. Moreover, as we discussed in Subsection 1.1, Assumption 1 implies that there is a point $z \in K$ such that the ball $S(z, 2^{-l})$ of radius 2^{-l}, centered at point z , is contained in the feasible region K and $S = S(z, 2^{-l}) \neq K$. Therefore $T(S) \subseteq T(K)$ and

$$\text{Vol}(T(K)) > \text{Vol}(T(S)) = \text{Vol}(S(T(z), 2^{-L_1+\log n})) \geq \frac{(2^{-L_1+\log n})^n}{n^n} \geq \text{Vol}(P^{\bar{k}}).$$

Since Vol$(T(K))$ >Vol$(P^{\bar{k}})$, it follows that $T(K) \cap P^{\bar{k}^c} \neq \emptyset$. ■

Theorem 5 *Suppose Assumptions 1, 4- 9 are satisfied, then the GGF computes an ε-approximate $WVIP$ solution $\bar{x}$ in $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ iterations. Moreover, $C_d^1(\bar{x}) \leq \varepsilon$ (or $C_d^2(\bar{x}) \leq \varepsilon$) in the same number of iterations.*

Proof. As in the preceding proposition we first note that since the volume of set P^k is at most 2^{-nL_1} in $\bar{k} = O\left(-\frac{nL_1}{\log b(n)}\right)$ iterations, the GGF reaches a feasible solution in at most $\bar{k}$ steps. Furthermore, notice that Proposition

5 implies that there exists a point y that lies in $T(K) \cap P^{\bar{k}^c}$. Then it follows that

$$\begin{aligned}
C_d(\bar{x}) &\leq C_d(y), \left(\text{since } y \in T(K) \cap P^{\bar{k}^c}\right) \\
&= f(y_y)'(y - y_y), \text{ (using the definition of } C_d(y)) \\
&= f(y_y)'(y - x^*) + f(y_y)'(x^* - y_y) \\
&\leq f(y_y)'(y - x^*), \text{ (since } x^* \text{ is a } WVIP \text{ solution)} \\
&= 2^{-L_1 + l + \log n} f(y_y)'(z - x^*), \left(\text{setting } z = T^{-1}(y)\right) \\
&\leq 2^{-L_1 + l + \log n} \cdot M \cdot 2^{L+1}, \text{ (using Assumption 4)} \\
&\leq \varepsilon, \text{ (using Assumption 9)}.
\end{aligned}$$

Since for any $x \in K$, $C_d^i(x) \leq C_d(x)$, for $i = 1, 2$ (see Proposition 2 and Theorem 4), it follows that in $\bar{k}$ iterations, Approximation Scheme i satisfies $C_d^i(\bar{x}) \leq \varepsilon$. ■

The previous theorem developed a complexity result for the weak variational inequality problem. We are now ready to prove a complexity result for the variational inequality problem.

Theorem 6 *Let $L_2 = L + 3l + 2\log(\frac{(M\lambda)^{0.5}\rho 2^{L+1}}{\varepsilon})$. Consider a VIP satisfying Assumptions 1, 4 - 9. In $O\left(-\frac{nL_2}{\log b(n)}\right)$ iterations, the schemes we considered in Section 3 compute an ε-approximate VIP solution.*

Proof. Theorem 5 implies that in $O\left(-\frac{nL_2}{\log b(n)}\right)$ iterations each of the schemes computes a point x at which the corresponding dual gap function approximation (i.e., C_d, C_d^1 and C_d^2 respectively) does not exceed $\frac{\varepsilon^2}{\rho^2 \lambda 2^{2L+2}}$. From Lemma 2, it follows that this point is also an $\frac{\varepsilon^2}{\lambda 2^{2L+2}}$-approximate $WVIP$ solution. Moreover, Proposition 4 implies that such a point is also an ε-approximate VIP solution. ■

Example: (see [8]). All the methods that are special cases of the GGF can also be modified to incorporate the cuts we introduced in this paper. Below we list some of these methods together with the respective descriptions of the volume reduction constants as well as the complexity bounds.

Method of centers of gravity: $b(n) = \frac{e-1}{e}$, $O(nL_2)$.

Ellipsoid Method: $b(n) = 2^{O(\frac{1}{n})}$, $O(n^2 L_2)$.

Method of inscribed ellipsoids: $b(n) = 0.843$, $O(nL_2)$.

Volumetric Center Method: $b(n) = const$, $O(nL_2)$.

On the Computational Solution of VIPs

5 An Affine Scaling Method

So far in this paper, we have introduced a cutting plane framework for solving variational inequalities. This framework extended the general geometric framework in [8] by considering nonlinear cuts as well as "smarter" linear cuts. Even though we showed polynomial complexity for the methods in this framework, the task of finding a "nice" set P^k and its center x^k, at each iteration k, might be computationally difficult. Moreover, the complexity bounds for the methods in this framework tend to be tight in practice. For this reason, in the remainder of this paper, we will consider methods that are simpler to perform and, perhaps, as a result computationally more efficient. Our motivation in this part of the paper comes from the success of (i) interior point methods for solving linear optimization problems in practice, but also (ii) the Frank-Wolfe method for solving traffic equilibrium problems. This latter method is widely used by transportation practitioners in the context of traffic equilibrium. However, it often exhibits pathological behavior when close to a solution. In this section, we will introduce a method that closely relates to the Frank-Wolfe method and is a variation of the affine scaling method for solving linear optimization. Moreover, we will incorporate in the affine scaling method (AS method) the cut ideas from the first part of the paper (Section 3). Combining these ideas will also allow us to propose more practical versions of the schemes we considered in the previous sections.

In the remainder of this section we assume that the feasible region is a polyhedron and that Assumptions 1 and 6 hold. We will also assume the following:

Assumption 10 *The Jacobian matrix of the problem function f is symmetric and positive semi-definite.*

Under this assumption, we can represent the problem function f as the gradient of a convex objective function F (i.e., $f = \nabla F$). Then the variational inequality problem becomes equivalent to the convex optimization problem

$$\min\{F(x) \mid Ax = b,\ x \geq 0\}. \tag{5}$$

Under this assumption, the results in the remainder of this section apply to problem (5). To motivate the approach we take, we first describe the Frank-Wolfe method.

The Frank-Wolfe Method.

1. Start with a feasible point x_0, tolerance $\varepsilon > 0$.

2. At step k:

   $d_k = \arg\min_d \left\{ f\left(x^k\right)' d \mid Ad = 0,\ x^k + d \geq 0 \right\}$

   $\alpha_k = \arg\min \left\{ F(x^k + \alpha d^k) \mid \alpha \in [0,1] \right\}$

   $x^{k+1} \leftarrow x^k + \alpha^k d^k$

3. Stop when $|f(x^k)' d^k| < \varepsilon$.

The approach below modifies Step 2 by restricting the direction finding subproblem to a Dikin ellipsoid.

An Affine Scaling Method.

1. Start with a strictly feasible point x_0, tolerance $\varepsilon > 0$, and constant $r \in (0,1)$.

2. At step k:

   $d_k = \arg\min \left\{ f\left(x^k\right)' d \mid Ad = 0,\ \|(X^k)^{-1} d\| \leq r \right\}$

   $\alpha^k = \arg\min_{\alpha \in [0,1]} F(x^k + \alpha^k d^k)$

   $x^{k+1} \leftarrow x^k + \alpha^k d^k$

3. Stop when $|f(x^k)' d^k| < \varepsilon$.

Remark: In this section we consider a short step version of the affine scaling method. Nevertheless, the results in this section can be easily modified to apply to the long step version of the method.

This method differs from other affine scaling methods in the literature for solving nonlinear optimization problems, (see [13], [16], [17]). In particular, notice that the direction finding subproblem in this paper optimizes a linear objective. The literature (for example, [13] and [16]) often considers a quadratic objective.

The necessary conditions of optimality for problem (5) are

$$Ax^* = b,\quad x^* \geq 0,\quad f\left(x^*\right) + A'y^* - s^* = 0,\quad x^{*\prime} s^* = 0,\quad s^* \geq 0. \tag{6}$$

For a feasible point x, the affine scaling algorithm finds a direction of descent by solving

$$\min \{f(x)' d \mid Ad = 0,\ \|X^{-1}d\| \leq r\}. \tag{7}$$

The KKT conditions for this problem are

$$f(x) + A'y + 2\mu X^{-2}d = 0,\ \ \mu(d'X^{-2}d - r^2) = 0,\ \ \mu \geq 0,\ \ Ad = 0,\ \ \|X^{-1}d\| \leq r. \tag{8}$$

If $\mu = 0$, then the current iterate is a global optimum. If $\mu \neq 0$, the solution to problem (8) is

$$\mu = \frac{\|PXf(x)\|}{2r},\ \ d = -r\frac{XPXf(x)}{\|PXf(x)\|},\ \ y = -(AX^2A')^{-1}AX^2f(x),$$

where $P = I - XA'(AX^2A')^{-1}AX$ is a projection matrix on the null space of matrix AX. We also introduce a variable related to the dual variable s^* in (6) by letting $s(x) = -2\mu X^{-2}d = X^{-1}PXf(x)$.

Notice that d is a direction of descent, that is $f(x)'d = -2\mu r^2 < 0$. Moreover, iterate $x^{k+1} = x^k + \alpha^k d^k$ is strictly feasible since constant $r < 1$.

Before proving the convergence of this method, we prove two key lemmas. These lemmas demonstrate that the limit points of the affine scaling method satisfy the complementary slackness and dual feasibility properties.

Lemma 3 *Consider the sequence $\{x^k\}$ generated by the affine scaling algorithm and suppose that $f(x^k)'d^k \to 0$, as $k \to \infty$. Then it follows that $\|X^k s(x^k)\| \to 0$.*

Proof. Conditions (8) imply that $f(x^k)d^k = -2\mu^k r^2$. Therefore, from the definition of $s(x^k)$, it follows that $\|X^k s(x^k)\| = \left\|2\mu^k X^{k^{-1}} d^k\right\| = 2\mu^k r = -\frac{f(x^k)'d^k}{r}$. ■

Lemma 4 *Suppose that the sequences $\{x^k\}$ and $\{s^k\}$ generated by the affine scaling method converge to points $\bar{x}$ and $\bar{s}$ respectively. Then $\bar{s} \geq 0$.*

Proof. Let us assume, by contradiction, that $\bar{s} < 0$. That is, there exist integer indices i and K_0, such that for all

$k \geq K_0$, $s_i^k \leq -\varepsilon < 0$. We rewrite x^{k+1} as $x^{k+1} = x^k + \alpha^k d^k = x^k - \alpha^k \frac{r}{\|P^k X^k f(x^k)\|}(X^k)^2 s^k$, then

$$x_i^{k+1} = x_i^k \left(1 - \alpha^k r \frac{x_i^k s_i^k}{\|P^k X^k f(x^k)\|}\right) \geq x_i^k \left(1 + \alpha^k r \frac{x_i^k \varepsilon}{\|P^k X^k f(x^k)\|}\right) > x_i^k,$$

since $x_i^k, \alpha^k r, \varepsilon > 0$. Therefore, the limit of the sequence x_i^k is strictly positive while $\lim_{k\to\infty} x_i^k s_i^k < 0$. However, since sequence $\{x^k\}$ converges, $\lim_{k\to\infty} f(x^k)' d^k = 0$. Lemma 3 leads us to a contradiction. It follows that $\lim_{k\to\infty} s^k = \bar{s} \geq 0$. ∎

In Subsections 5.1 and 5.2, we will describe two different convergence proofs for the AS Method under different assumptions.

5.1 The Convergence of the AS Method under Strict Complementarity

In order to describe the first convergence approach of the AS method, we first need to describe how we choose the step sizes α^k at each iteration in the method. As we described in the beginning of Section 5, we let $x^k(\alpha) = x^k + \alpha d^k$, with $\alpha \in [0,1]$. We notice that the line search procedure we described is equivalent to finding a step size $\alpha^{k+1} \in [0,1]$ such that $f(x(\alpha^{k+1}))'(x(\alpha) - x(\alpha^{k+1})) \geq 0$, for all $\alpha \in [0,1]$. We set as the next iterate the point $x^{k+1} = x(\alpha^{k+1})$. Notice that this point is well-defined since f is a monotone function. Furthermore, notice that if $f(x^k + d^k)' d^k > 0$, then using the mean value theorem, we can express the step size α^k as $\alpha^k = -\frac{f(x^k)' d^k}{\|d^k\|_{G_k}^2}$, where $\|d^k\|_{G_k}^2 = d^{k\prime} \nabla f(z^k) d^k$, for some $z^k \in [x^k; x^k + d^k]$. On the other hand, if $f(x^k + d^k)' d^k \leq 0$, then we set $\alpha^k = 1$ and as a result, $x^{k+1} = x^k + d^k$. Moreover, notice that unless we have reached a solution, step size $\alpha^k > 0$. As a result, wlog, we set $\alpha^k = \min\left\{-\frac{f(x^k)' d^k}{\|d^k\|_{G_k}^2}, 1\right\}$. When the step sizes are chosen in this fashion, sequence $\{F(x^k)\}$ is nonincreasing. This follows using mean value theorem since

$$F(x^k) = F(x^{k+1}) + f(x^{k+1})'(x^k - x^{k+1}) + \|x^k - x^{k+1}\|_{H(z^k)}^2, \tag{9}$$

for some $z^k \in [x^k; x^{k+1}]$. The line search procedure we described above implies that $f(x^{k+1})'(x^k - x^{k+1}) \geq 0$. This line search procedure together with relation (9) and Assumption 10 imply that $F(x^k) \geq F(x^{k+1})$.

In what follows we will establish a convergence result for the AS method with step sizes as defined above. To achieve this we first need to impose two additional assumptions. These assumptions ensure primal nondegeneracy (that is, Assumption 7 from Subsection 3.2.3) and strict complementarity for the VIP solutions.

Definition 8 A limit point $(\bar{x}, \bar{s})$ satisfies the property of *strict complementarity*, if $\bar{x}_i \bar{s}_i = 0$ and $\bar{x}_i + \bar{s}_i \neq 0$, $\forall i$.

Assumption 11 *Every limit point of the sequence $\{x^k, s(x^k)\}$ has the property of strict complementarity.*

Theorem 7 *Under Assumptions 1, 6, 7, 10, 11 the AS method converges to an optimal solution.*

Before proceeding with the proof of this theorem, we establish several intermediate results.

Lemma 5 *Suppose that Assumptions 1 and 10 hold. Consider the sequence $\{x^k\}$ generated by the AS method with step sizes as described above. Then $\|X^k s(x^k)\| \to 0$ as $k \to \infty$.*

Proof. We will first show that $f(x^k)'d^k \to 0$ as $k \to \infty$. Notice that sequence $\{F(x^k)\}$ is nonincreasing and bounded and, therefore, it has a limit point. We denote this limit point by $\bar{F}$. For $\forall k$,

$$F(x^k) - \bar{F} \geq -\alpha^k f(x^k)'d^k + o(\|d^k\|^2) = -\min\left\{-\frac{f(x^k)'d^k}{\|d^k\|^2_{G_k}}, 1\right\} f(x^k)'d^k + o(\|d^k\|^2).$$

It follows that $f(x^k)'d^k \to 0$, whenever $F(x^k) \to \bar{F}$, i.e., as $k \to \infty$. Lemma 3 implies the result. ■

Proposition 6 *Suppose that Assumptions 1, 7 and 11 are satisfied. Then the sequence $\{x^k\}$ generated by the AS method converges.*

Proof. For some point $\bar{s}$, we denote by $N = \{i \mid \bar{s}_i \neq 0\}$, $B = \{i \mid \bar{s}_i = 0\}$, $S^* = \{x \mid Xs(x) = 0,\ Ax = b,\ x \geq 0\}$, and $C_\delta = \{x \mid x_i \in [0, \delta)\ \forall i \in N\}$.

First notice that from Assumption 7 and the continuity of matrix AX^2A', it follows that $s(x)$ is a continuous variable. Hence for a small enough $\delta > 0$, when $x \in C_\delta$ and $j \in N$, it holds that $|s_j| \geq \frac{1}{2}|\bar{s}_j| > 0$.

We next show that Assumption 11 implies that the set of limit points is discrete. Since the sequence $\{(x^k, s^k)\}$ is bounded, it has some limit point, say $(\bar{x}, \bar{s})$. Notice that every limit point $\bar{x}$ of the sequence $\{x^k\}$ induced by the AS method belongs to the set S^*. This set also contains all $x \in K$ satisfying the strict complementarity property. We will show that for some $\delta > 0$ there exists a neighborhood C_δ such that $C_\delta \bigcap S^* = \bar{x}$. Suppose, on the contrary, that $\forall \delta > 0$, $\exists x^\delta \in S^* \bigcap C_\delta$ such that $x^\delta \neq \bar{x}$. Then, since point $\bar{x}$ is a vector of finite dimension, for some $j \in N$, there exists a sequence $\{x^{\delta_n}\}$ with $x^{\delta_n} \in S^* \bigcap C_{\delta_n}$, $x^{\delta_n} \neq \bar{x}$, $x_j^{\delta_n} \to 0$, $x_j^{\delta_n} > 0$, and $|x_j^{\delta_n} - \bar{x}_j| < \frac{1}{n}$. Notice that the strict complementarity property implies that $s_j^{\delta_n} = 0$ whereas $\bar{s}_j \neq 0$. By continuity, however, $s_j^{\delta_n} \to \bar{s}_j$, which is a contradiction. Therefore, for any limit point of the sequence $\{x^k\}$, there exists a neighborhood whose intersection with the set of all limit points is a singleton.

At this point, we notice that conditions analogous to those in [17] are satisfied and the convergence of the sequence $\{x^k\}$ can be shown by contradiction. Suppose that the entire sequence $\{x^k\}$ does not converge to $\bar{x}$. Then for some $\delta > 0$, $x^k \in C_\delta^c$ infinitely often, whereas $|s_j(x)| \geq \frac{1}{2}|\bar{s}_j| > 0$, when $x \in C_\delta, j \in N$.

Consider a subsequence k_p such that $x^{k_p} \in C_\delta$ but $x^{k_p+1} \in C_\delta^c$. Since x is a vector of finite dimension, it follows that for some index $j \in N$, there exists a subsequence $\{k_l\} \subseteq \{k_p\}$ such that $x_j^{k_l+1} > \delta$. The properties of the AS method imply that $x_j^{k_l} > \frac{\delta}{2}$. On the other hand, notice that $|s_j^{k_l}| \geq \frac{1}{2}|\bar{s}_j| > 0$, since $x^{k_l} \in C_\delta$. Therefore,

$$|x_j^{k_l} s_j^{k_l}| \geq \frac{\delta}{2}|\bar{s}_j| > 0, \text{ for all } k_l.$$

This contradicts the fact that $||X^k s(x^k)|| \to 0$ as $k \to \infty$ (see Lemma 5). We conclude that the sequence $\{x^k\}$ converges to a point $\bar{x}$. ■

We are now ready to prove Theorem 7 which is the main convergence result of the AS method.

Proof of Theorem 7. Proposition 6 shows that $\{x^k\}$ is a convergent sequence. Moreover, since function $s(x)$ is continuous in x, sequence $\{s^k\}$ is also convergent. Hence, Lemmas 4 and 5 imply that the sequence $\{s^k\}$ converges to a nonnegative point and $\lim_{k\to\infty} ||X^k s^k|| = 0$. Therefore, the limit point of the sequence $\{x^k\}$ satisfies the necessary conditions of optimality. Assumption 10 then implies the result. ■

5.2 The Convergence of the AS Method under the Jacobian Similarity Property

In the subsection we introduce an alternate convergence proof for the AS method. In this proof we do not assume the property of strict complementarity (Assumption 11). Instead we assume that the problem function f satisfies the Jacobian similarity property (Assumption 5). An essential element behind this convergence approach is the modification of the line search procedure in the AS method. In particular, we set the next iterate as $x^{k+1} = x^k + \alpha^k d^k$, where the direction d^k is chosen as in Subsection 5.1 and the "optimal" step size α^k is determined through a line search. That is, if for all $\alpha \in [0, 1]$, we define $x^k(\alpha) = x^k + \alpha\rho d^k$, where constant ρ is the Jacobian similarity constant, then we choose

$$\alpha^k \in [0,1] \text{ satisfying } (f(x^k(\alpha^k))'d^k).(\alpha - \alpha^k) \geq 0, \ \forall \alpha \in [0,1].$$

Notice that step size α^k equals the one in Subsection 5.1 scaled by $\frac{1}{\rho}$, where ρ is the Jacobian similarity constant.

In this subsection, we assume that the step sizes α^k are bounded away from zero. In particular,

Assumption 12 *For some $\alpha > 0$, $\alpha^k \geq \alpha$, for all k.*

Remark: Indeed some of the examples we consider in Section 6, induce step sizes that satisfy this condition. As a result, this assumption allows us to relax the assumption of strict complementarity that we imposed in the previous subsection.

Theorem 8 *Under Assumptions 5, 6, 10 and 12, the AS method converges to an optimal solution.*

Before proving this theorem we show some intermediate results.

Proposition 7 *Consider the sequence of step sizes $\{\alpha^k\}$ as described above. For some $\bar{F}$ the following relation holds* $F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha^k}{2\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right)$.

Proof. We define an auxiliary point $y^k = x^k + \rho\alpha^k d^k$, and use a Taylor expansion to obtain

$$\begin{aligned} F(x^k) &= F(y^k) + f(y^k)'(x^k - y^k) + \frac{1}{2}(x^k - y^k)'\nabla f(z^k)(x^k - y^k), \text{ (for some } z^k \in [x^k; y^k]) \\ &\geq F(y^k). \text{ (This follows from the definition of } y^k \text{ and the monotonicity of problem function } f.) \end{aligned}$$

Moreover, the convexity of the objective function F and the previous inequality imply that

$$F(x^{k+1}) = F\left(\left(1 - \frac{1}{\rho}\right)x^k + \frac{1}{\rho}y^k\right) \leq \left(1 - \frac{1}{\rho}\right)F(x^k) + \frac{1}{\rho}F(y^k) \leq F(x^k).$$

Therefore, it follows that the sequence $\{F(x^k)\}$, with step sizes as defined above, is monotonically nonincreasing. Since this sequence is also bounded, it has a limit point. We denote this limit by $\bar{F}$ and let set $S = K \cap \{x \mid F(x) \leq \bar{F}\}$, where K is the feasible region. Notice that $S \neq \emptyset$, since every cluster point of the sequence $\{x^k\}$ belongs to this set. Using similar arguments as in [13], we can show that for sufficiently large k,

$\min_{z\in S}\left\|(X^k)^{-1}\left(z-x^k\right)\right\| \leq \sqrt{n}$. In what follows, we use this result to devise a feasible point in the direction finding subproblem. In particular, suppose that point $z^k \in S$ satisfies $\left\|(X^k)^{-1}\left(z^k-x^k\right)\right\| \leq \sqrt{n}$, then point $x = x^k + \frac{a}{\sqrt{n}}(z^k - x^k)$, with $0 \leq a \leq 1$, is feasible for the affine scaling direction finding subproblem. Therefore,

$$f(x^k)'d^k \leq \frac{a}{\sqrt{n}} f(x^k)'\left(z^k - x^k\right), \text{ for all } 0 \leq a \leq 1. \tag{10}$$

An application of the mean value theorem implies that

$$f(y^k)'\left(y^k - x^k\right) = \alpha^k \rho f(x^k)'d^k + (\alpha^k\rho)^2 d^{k\prime}\nabla f(\hat{z})d^k, \text{ for some } \hat{z} \in \left[x^k; y^k\right].$$

Therefore, the definition of point y^k implies that

$$\alpha^k \rho d^{k\prime}\nabla f(\hat{z})d^k \leq -f(x^k)'d^k. \tag{11}$$

Combining these results we obtain that

$$\begin{aligned}
F(x^{k+1}) &= F\left(x^k\right) + \alpha^k f(x^k)'d^k + \frac{1}{2}(\alpha^k)^2 d^{k\prime}\nabla f(\bar{z})d^k, (\text{ for some } \bar{z} \in [x^k; x^{k+1}]) \\
&\leq F\left(x^k\right) + \alpha^k f(x^k)'d^k + \frac{\rho}{2}{\alpha^k}^2 d^{k\prime}\nabla f(\hat{z})d^k, \text{ (using the property of Jacobian similarity)} \\
&\leq F\left(x^k\right) + \alpha^k f(x^k)'d^k - \frac{1}{2}\alpha^k f(x^k)'d^k, \text{ (using (11))} \\
&\leq F\left(x^k\right) + \frac{\alpha^k}{2}\frac{1}{\sqrt{n}} f(x^k)'\left(z^k - x^k\right), \text{ (using (10))} \\
&\leq F\left(x^k\right) + \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\left(F\left(z^k\right) - F\left(x^k\right)\right), \text{ (using the convexity of the objective function } F) \\
&\leq \left(1 - \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\right) F\left(x^k\right) + \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\bar{F}. \\
&\text{Hence, } F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha^k}{2}\frac{1}{\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right). \blacksquare
\end{aligned}$$

This result also allows us to obtain a rate of convergence for the AS method in this paper. Next we illustrate that the sequence that this method induces is indeed convergent. This result will also allow us to provide a convergence result for the AS method (that is, prove Theorem 8).

Theorem 9 *Under Assumptions 5, 6, 10 and 12, the sequence $\{x^k\}$ generated by the AS method converges.*

Proof. First we observe that analogously to [13] (see also Appendix A), Assumptions 5 and 12 imply that there exists some constant $c > 0$, such that $F(x^k) - F(x^{k+1}) \geq c{d^k}'d^k$. Therefore,

$$\|x^k - x^s\| \leq \sum_{s+1}^{k} \|x^i - x^{i-1}\| = O\left(\sum_{s+1}^{k} \left|F\left(x^i\right) - F\left(x^{i-1}\right)\right|^{1/2}\right).$$

Assumption 12 and Proposition 7 imply that $F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha}{2\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right)$. Hence,

$$\begin{aligned} F\left(x^{k+1}\right) - F\left(x^k\right) &\leq F\left(x^{k+1}\right) - \bar{F} \leq \left(1 - \frac{\alpha}{2\sqrt{n}}\right)\left(F\left(x^k\right) - \bar{F}\right), \text{ and as a result,} \\ \|x^k - x^s\| &\leq O\left(\sum_{i=1}^{k-s}\left(1 - \frac{\alpha}{2\sqrt{n}}\right)^{i-1} \left|F\left(x^s\right) - \bar{F}\right|^{1/2}\right) \\ &= \frac{2\sqrt{n}}{\alpha}\left(1 - \frac{\alpha}{2\sqrt{n}} - \left(1 - \frac{\alpha}{2\sqrt{n}}\right)^{k-s}\right) O\left(\left|F\left(x^s\right) - \bar{F}\right|^{1/2}\right). \end{aligned}$$

Since the sequence $F\left(x^s\right)$ converges to $\bar{F}$, it follows that as $k, s \to \infty$, $\|x^k - x^s\| \to 0$. We conclude that $\{x^k\}$ is a Cauchy sequence and, therefore, it is a convergent sequence. ■

We are now ready to prove the general convergence result of this subsection.

Proof of Theorem 8. Theorem 9 shows that $\{x^k\}$ is a convergent sequence. Moreover, since function $s(x)$ is a continuous function in x, sequence $\{s^k\}$ is also convergent. Therefore, from Lemma 4, it follows that $\bar{s} = \lim_{k\to\infty} s^k \geq 0$. Since $\alpha^k > \alpha$, $F(x^k) - \bar{F} \geq -\alpha f(x^k)'d^k + o(\|d^k\|^2)$. Hence, as $k \to \infty$, $f(x^k)'d^k \to 0$, whenever $F(x^k) \to \bar{F}$. Then Lemma 3 implies that $\|X^k s^k\| \to 0$. It follows that the limit of sequence $\{x^k\}$ satisfies the necessary conditions of optimality. The convexity of the objective function F implies the result. ■

6 Variations of the AS Method and Computational Results

Our goal in this section is to examine computationally the performance of the methods we introduced in this paper. For this reason, we consider the affine scaling method as well as several variations that incorporate the cut ideas from Section 3. In order to test the methods in this paper, we chose several randomly generated instances of symmetric, affine variational inequality problems. In particular, we considered variational inequality problems of the following format,

$$\text{Find } \; x^* \in P \text{ such that } \; f(x^*)'(x - x^*) \geq 0, \quad \forall x \in P,$$

with (i) problem function $f(x) = Mx - c$, where M is a symmetric, positive semi-definite $n \times n$ matrix, vector $c \in \mathbb{R}^n$, and (ii) polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, where A is an $m \times n$ matrix, vector $b \in \mathbb{R}^m$. We implemented all the methods using MATLAB version 6.1, on a personal computer with a Dual Xeon processor, 1.5GHz speed and 1GB RAM memory. Finally, the CPU time was computed using MATLAB's build-in function. In the stopping criterion we used as tolerance level $\varepsilon = 10^{-4}$.

In Table 1, we compare the performance of the affine scaling method introduced in this paper with the performance of the Frank-Wolfe method. We observe that in all the examples we studied, the affine scaling method in this paper fixes the zigzagging behavior of the Frank-Wolfe method. Moreover, in most of the examples, the affine scaling method computes a solution in shorter CPU time. Nevertheless, in one of the examples, the Frank-Wolfe method computed the solution faster (in terms of CPU time) than the affine scaling method. This was an example where the Frank-Wolfe method did not zigzag but rather the solution of the variational inequality problem lied at a corner point of the polyhedral feasible region. The Frank-Wolfe method computed the solution in one step (that is, by solving a single linear optimization subproblem). It is worth noting that a possible reason for the faster convergence of the Frank-Wolfe method in this example, is that as it solved a single linear optimization subproblem using Matlab's built-in optimization solver, it relied on the speed of this implementation. On the other hand, in the same example, the affine scaling method applied a sequence of much simpler steps. Our belief is that a better implementation of the affine scaling method will yield a similar performance even in this example. In conclusion, we observed that in most of the examples we generated, the affine scaling method outperformed the Frank-Wolfe method.

Moreover, in Table 2, we compare the performance of the long step affine scaling method we introduced in this paper (LAS) with that of two other affine scaling methods: i) QLAS, a quadratic approximation long step affine scaling method (see [17]), ii) DLAS, a long step affine scaling method that considers a quadratic approximation of the objective further using a diagonal approximation of the Jacobian matrix. We chose the long step versions of these methods since we noticed that within the family of affine scaling methods the long step versions perform the best computationally. The two quadratic approximation methods (QLAS and DLAS) perform similarly computationally. Furthermore, in most of the examples we generated, the affine scaling method introduced in this paper outperformed

m	n	FW		LAS	
		iterations	time	iterations	time
8	7	2	0.0628	10	0.028
10	6	>3000	106.6596	11	0.0315
31	30	18	0.8967	137	1.1397
51	30	>3000	>171	46	0.7505
43	40	2	0.0691	21	0.3962
101	100	>3000	171.244	221	22.0946
171	100	>300	>300	803	252.9649
111	110	269	139.112	801	99.2196

FW Frank-Wolfe Algorithm
LAS long step Affine Scaling Algorithm

Table 1: LAS method vs. FW method

both of these methods in CPU time. We attribute this partly to the simplicity of each iteration. Nevertheless, in two examples the affine scaling method (LAS) performed worse. Even in these two examples, we drastically improved the performance of the method (LAS) when we incorporated cuts (see Table 3). This new method significantly outperformed the two quadratic approximation methods. In what follows, we will discuss this in further detail.

m	n	LAS		QLAS		DLAS	
		iterations	time	iterations	time	iterations	time
8	7	10	0.028	10	0.0685	9	0.0667
11	10	9	0.0468	10	0.0859	10	0.0904
31	30	137	1.1397	12	1.3045	11	1.2324
51	30	46	0.7505	15	1.4003	12	1.2455
43	40	21	0.3962	16	3.2849	16	3.4243
166	60	153	41.5669	100	92.3103	100	95.1253
154	77	100	22.3608	95	189.2372	100	201.0876
101	100	221	22.0946	20	87.9	24	101.2705
171	100	803	252.9649	17	79.537	49	234.6041
111	110	801	99.2196	20	97.3761	15	69.733

LAS long step Affine Scaling Algorithm
QLAS LAS with a quadratic approximation of the objective in the subproblem
DLAS LAS with a diagonalized approximation of the objective in the subproblem

Table 2: LAS method vs. QLAS and DLAS methods

In order to further improve the computational results in this paper, we incorporate the cut ideas we discussed in Section 3 into the affine scaling method. The motivation in this comes from the observation that methods utilizing cuts often provide better complexity results in theory as well as in practice.

The AS Method with Cuts

1. Start with a strictly feasible point x_0, feasible region $K^0 = K$, tolerance $\varepsilon > 0$, and constant $r \in (0, 1)$.

2. At iteration k:

   (a) Find $y^k \in K$ such that $F(y^k) \leq F(x^k)$.

   (b) $d^k = \arg\min \{f(y^k)'d \mid Ad = 0, \ \|(Y^k)^{-1}d\| \leq r\}$.

   (c) Choose step size α^k such that $(\alpha - \alpha^k) f(y^k + \alpha^k d^k)'d^k \geq 0$, for all $\alpha \in (0, \alpha^k_{\max})$, where $\alpha^k_{\max} \geq 1$.

   (d) $x^{k+1} \leftarrow y^k + \alpha^k d^k$.

   (e) Update $K^{k+1} = K^k \cap Cut(y^k, x^k)$.

3. Stop when $|f(x^k)'d^k| \leq \varepsilon$.

In what follows, we compare the long step affine scaling method (LAS) from Section 5 with several special cases of the AS method with cuts we just described. These special cases include the following:

(1) **Simple cuts.** Set point $y^k = x^k$ and introduce cut $Cut(x^k, x^k) = \left\{x \mid f\left(x^k\right)' \left(x - x^k\right) \leq 0\right\}$ (LASC).

(2) **Cuts based on the dual gap function.** Suppose that Assumption 5 holds. We introduce cut $Cut(y^k, x^k) = \left\{x \mid f\left(y^k\right)' \left(x - x^k\right) \leq 0\right\}$ with

   (a) $y^k = \arg\max_{y \in K, y \geq x^k} \left\{f(x^k)'(x^k - y) - \rho \left\|x^k - y\right\|^2_{H(x^k)} \mid Ay = b, \| \ (X^k)^{-1}(y - x^k)\| \leq r\right\}$ (LASGs).

   (b) $y^k = \arg\max_{y \in K} \left\{f(x^k)'(x^k - y) - \rho \left\|x^k - y\right\|^2_{H(x^k)} \mid Ay = b, \| \ (X^k)^{-1}(y - x^k)\| \leq r\right\}$ (LASGD).

m	n	LAS iterations	LAS time	LASC iterations	LASC time	LASGs iterations	LASGs time	LASGD iterations	LASGD time
8	7	10	0.028	10	0.0325	6	0.0586	8	0.052
11	10	9	0.0468	9	0.0302	6	0.066	7	0.0678
31	30	137	1.1397	24	0.2985	8	0.9622	17	1.9883
51	30	46	0.7505	24	0.5143	9	0.7311	13	1.2917
43	40	21	0.3962	21	0.6716	10	0.8234	18	2.2375
166	60	153	41.5669	161	33.2601	12	4.1987	181	61.9024
154	77	100	22.3608	104	25.89	13	4.6085	133	104.8659
101	100	221	22.0946	25	2.9223	12	13.4612	89	392.7285
171	100	803	252.9649	40	15.8761	10	13.4899	50	241.7931
111	110	801	99.2196	33	4.659	13	12.9911	26	108.0049

LAS	long step Affine Scaling Algorithm
LASC	LAS with cuts Cut(x,x)
LASGs	LAS with cuts Cut(y,x), where y is found within { z | Az >=a Ax , with 0<a<=1}
LASGD	LAS with cuts Cut(y,x), where y is found within Dikin ellipsoid

Table 3: AS method vs. AS methods with cuts

Table 3 summarizes the computations that compare the various versions of the affine scaling method of this paper. We notice that the two best versions in terms of CPU time are the method that uses simple cuts (LASC) as well as the method that uses cuts determined via a gap function where the direction of the cut is found within a restricted feasible region (LASGs). Moreover, Table 3 demonstrates that these two versions compute a solution in a comparable or even less number of iterations than the quadratic approximation affine scaling methods. Nevertheless, in terms of CPU time, both methods with cuts are faster. Among the versions of the method with cuts determined via a gap function, the method where the direction of the cut is found within a restricted feasible region (LASGs) has consistently the least number of iterations. In colclusion, both LASC and LASGs outperform considerably the affine scaling method without cuts, both in terms of number of iterations and in terms of CPU time.

Below we summarize our learnings from the computational experiments we performed.

1. All the versions of the affine scaling method we introduced in this paper fix the zigzagging behavior of the Frank-Wolfe method.

2. In one example the Frank-Wolfe method performed better than the affine scaling method. This was an example where the Frank-Wolfe method did not zigzag, but rather the Frank-Wolfe method found the solution in one step through the solution of a linear optimization subproblem. We attribute this to the quality of the Matlab built-in linear optimization solver. We believe that a better implementation of the affine scaling method will also yield comparable results in terms of CPU time, even in this example.

3. In theory, the affine scaling method will perform in the worst case similarly to the Frank-Wolfe method. Nevertheless, in most cases in practice, we believe that the affine scaling will perform better.

4. After comparing all the versions of the affine scaling method we considered in this paper, we conclude that the LASC and LASGs methods perform better in terms of CPU time. Moreover, in terms of number of iterations, the LASC method performs a similar number of iterations while the LASGs method performs fewer iterations than the quadratic approximation affine scaling methods.

5. The LASC and LASGs versions of the affine scaling method outperform the affine scaling method (LAS) without cuts both in terms of number of iterations and in terms of CPU time.

6. In particular, the LASGs method performs fewer iterations than all the other versions of the affine scaling method we considered in this paper. Furthermore, it has the best or second best CPU time.

7. As the dimension of the problem grows, the LASC and LASGs versions of the affine scaling method consistently outperform the Frank-Wolfe method, the quadratic approximation affine scaling methods we considered, and finally, the affine scaling method without cuts.

We would like to note that the theoretical convergence properties of the general affine scaling method with cuts are similar to those of the affine scaling method without cuts. For the sake of brevity we do not include this discussion in the paper. Finally, we would also like to add that although the theoretical results we established for the affine scaling method (see Section 5) apply to the symmetric variational inequality problem, computationally this method seems to also work well for asymmetric problems.

7 Conclusions

In this paper, we have introduced an extension of the general geometric framework for solving variational inequalities by incorporating both linear and nonlinear cuts. In particular, in this framework we considered as special cases "smarter" linear cuts. The directions of these cuts were based on the dual gap function associated with the variational inequality problem. Furthermore, we established complexity results for the methods using these cuts. To make our results computationally efficient in practice, we introduced an affine scaling method that computed at each step a direction by solving a subproblem with a linear objective within a Dikin ellipsoid. We proved the convergence of this method in the case of monotone and symmetric variational inequality problems. The theoretical convergence of the method in the asymmetric case remains an open question. Finally, we also incorporated cutting plane ideas into the affine scaling method. In our computational experiments, we observed that even the affine scaling method without cuts outperformed in terms of CPU time both the Frank-Wolfe algorithm as well as variations of the affine scaling method that use a quadratic approximation of the objective in the direction finding subproblem. Furthermore, the affine scaling method with cuts reduced considerably the number of iterations as well as the CPU time for larger dimensional examples as compared to other methods. Although more extensive computational testing is needed, preliminary testing seems to indicate that these methods may perform well in practice.

References

[1] M. Denault and J.L. Goffin, "Variational inequalities with quadratic cuts," Preprint, 1998.

[2] I.I. Dikin, "Iterative solution of linear and quadratic programming", Soviet Mathematics Doklady, 8, 674-675, 1967.

[3] M. Frank and P. Wolfe, "An algorithm for quadratic programming," Naval Res. Logist. Quart. 3: 95-110, 1956.

[4] J.L. Goffin, P. Marcotte, and D. Zhu, "An analytic center cutting plane method for pseudomonotone variational inequalities," Operations Research Letters 20: 1-6, 1997.

[5] J.L. Goffin and J.-P. Vial, "Convex nondifferentiable optimization: a survey focused on the analytic center cutting plane method," Logilab Technical Report, 1999.

[6] H.J. Lüthi, "On the solution of variational inequalities by the ellipsoid method," Mathematics of Operations Research 10: 515-522, 1985.

[7] H.J. Lüthi and B. Büeler, "The analytic center quadratic cut method (ACQCM) for strongly monotone variational inequality problems," SIAM Journal of Optimization 2: 415-426, 2000.

[8] T. Magnanti and G. Perakis, "A unifying geometric solution framework and complexity analysis for variational inequalities," Mathematical Programming 71: 327-351, 1995.

[9] T. Magnanti and G. Perakis, "From Frank-Wolfe to steepest descent; a descent framework for solving asymmetric variational inequalities," Working Paper, ORC, MIT, 1998.

[10] F. S. Mokhtarian and J.-L. Goffin, "An analytic center quadratic cut method for convex quadratic feasibility program," Preprint, 2000.

[11] Y. Nesterov and A. Nemirovskii, "Interior Point Polynomial Algorithms in Convex Programming," SIAM Stud. Appl. Math. 13, Philadelphia, 1994.

[12] Y. Nesterov and J.-P. Vial, "Homogeneous analytic cneter cutting plane methods for convex problems and variational inequalities," SIAM Journal of Optimization 3: 707-728, 1999.

[13] J. Sun, "A convergence analysis for a convex version of Dikin algorithm," Annals of Operations Research 62: 357-374, 1996.

[14] P. Tseng, "Convergence properties of Dikin's affine scaling algorithm for nonconvex quadratic minimization," submitted for publication to *Journal for Global Optimization*, 2001.

[15] R. J. Vanderbei and J.C. Lagarias, "I.I. Dikin's convergence result for the affine-scaling algorithm," Contemprorary Mathematics, Vol. 114: 109-119, 1990.

[16] Y. Ye, "An extension of Karmakar's algorithm and the trust region method for quadratic programming", ed. N. Megiddo, *Progress in Mathematical Programming*, Springer, NY, 1989.

[17] Y. Ye, "On affine scaling algorithms for nonconvex quadratic-programming", Mathematical Programming 56: 285-300, 1992.

A

Proposition 8 *Suppose Assumptions 5 and 12 hold. Then* $F(x^k) - \bar{F} \geq c||d^k||^2$ *for some* $c > 0$.

Proof. Paper [13] has shown that when the problem function f has the property of Jacobian similarity, then for some positive integer $n_1 \leq n$, there is an orthogonal matrix M, such that for any $x \in K$, it holds that $M'HM = \begin{pmatrix} P_x & 0 \\ 0 & 0 \end{pmatrix}$, where $H = \nabla f(x)$ and matrix P_x is an $n_1 \times n_1$ bounded, positive definite matrix. Therefore, if we let $H^k = \nabla f(x^k)$, then

$$M'H^kM = \begin{pmatrix} P^k & 0 \\ 0 & 0 \end{pmatrix}.$$

We can express $d^k = M\begin{pmatrix} y^k \\ z^k \end{pmatrix}$ or $d^k = s^k + w^k$, where $s^k = M\begin{pmatrix} y^k \\ 0 \end{pmatrix}$ and $w^k = M\begin{pmatrix} 0 \\ z^k \end{pmatrix}$. From the proof of Proposition 7 it follows that

$$F(x^k) - F(x^{k+1}) \geq (\alpha^k)^2 \frac{\rho}{2} d^k \nabla f(\hat{z}) d^k,$$

therefore

$$F(x^k) - F(x^{k+1}) \geq \frac{(\alpha^k)^2}{2} y^k P^0 y^k.$$

When $\alpha^k \geq \alpha$,

$$\|s^k\| = \|y^k\| \leq \frac{\mu_1}{\alpha}\left(F(x^k) - F(x^{k+1})\right)^{1/2}.$$

Next we show that $\|w^k\| \leq \frac{\mu_2}{\alpha}\left(F(x^k) - F(x^{k+1})\right)^{1/2}$, for some constant μ_2. Suppose this is not true. Then, since $\{w^k\}$ is a bounded sequence, there exists a subsequence S and a subset J of $\{1, ..., n\}$ so that

$$\left\{\frac{|F(x^{k+1}) - F(x^k)|^{1/2}}{\|w^k\|}\right\}_S \to 0, \quad \lim_{k\to\infty, k\in S} \frac{w_j^k}{\|w^k\|} > 0, \ \forall j \in J, \quad \lim_{k\to\infty, k\in S} \frac{w_j^k}{\|w^k\|} = 0, \ \forall j \in J^c.$$

Then from the properties we established above for $\|s^k\|$, it follows that

$$\left\{\frac{\|s^k\|}{\|w_j^k\|}\right\}_S \to 0, \ \forall j \in J.$$

Consider the following system of linear equations

$$\begin{cases} A(s^k + w) = 0 \\ f(x^0)'w = f(x^0)'w^k \\ w_j = w_j^k \ \forall j \in J \\ (M'w)_j = 0 \ \forall j = 1, ..., n_1. \end{cases}$$

Next we will show that there exists a solution of the above system bounded by $|F(x^k) - F(x^{k+1})|^{1/2}$. To achieve this it suffices to show that in the system above the norm of the left-hand-side is bounded by $O(|F(x^k) - F(x^{k+1})|^{1/2})$, and the matrix on the right-hand-side is bounded. Paper [13] showed that for any system $Ax \leq b$ that has a feasible solution, there exists some solution whose norm is bounded by $\lambda\|b\|$, where λ is a constant dependent only on matrix A. Notice that

$$\begin{aligned} F(x^k) - F(x^{k+1}) &= \alpha^k f(x^k)'d^k + {\alpha^k}^2/2\, d^{k\prime} \nabla f(z^k) d^k \\ &= \alpha^k f(x^0)'(w^k + s^k) + \alpha^k (x^k - x^0)' \nabla f(z_0^k) d^k + {\alpha^k}^2/2\, d^{k\prime} \nabla f(z^k) d^k. \end{aligned}$$

From this relation it follows that

$$\begin{aligned} -f(x^0)'w^k &= f(x^0)'s^k + \frac{1}{\alpha^k}\left(F(x^k) - F(x^{k+1})\right) \\ &\quad + (x^k - x^0)'M\begin{pmatrix} Z_0^k & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} y^k \\ z^k \end{pmatrix} + \alpha^k y^{k\prime} Z^k y^k, \end{aligned}$$

where $\|Z_0^k\|$, $\|Z^k\|$, $\|x^k - x^0\|$ are bounded on K, and $\alpha \leq \alpha_k \leq 1$. Therefore, since there exists a solution of this system, namely w^k, there exists a solution $\bar{w}^k$ that is of order $O\left(|F(x^k) - F(x^{k+1})|\right)$. Therefore $\{\|\bar{w}^k\|/\|w^k\|\}_S \to 0$.

Let $t^k = w^k - \bar{w}^k$. Then it follows that

$$\begin{cases} [M't^k]_i = 0, \ i = 1, ..., n_1 \\ At^k = 0 \\ f(x^k)'t^k = f(x^0)'t^k + (x^k - x^0)'M \begin{pmatrix} Z^t & 0 \\ 0 & 0 \end{pmatrix} M't^k = 0 \\ H_k t_k = M \begin{pmatrix} P^k & 0 \\ 0 & 0 \end{pmatrix} M't_k = 0. \end{cases}$$

Observe that for large $k \in S$, $d^k - t^k$ is an optimal solution of (7) lying in the interior of $D(x^k)$. Notice that $f(x^k)'(d^k - t^k) = f(x^k)'d^k$, $A(d^k - t^k) = 0$ and

$$||X^{k^{-1}}(d^k - t^k)|| = \sum_1^n \left(1 - \frac{\bar{w}_i^k}{w_i^k}\right)^2 \left(\frac{w_i^k}{x_i^k}\right)^2 \leq \sum_1^n \left(1 - \frac{\bar{w}_i^k}{w_i^k}\right)^2 \left(\frac{d_i^k}{x_i^k}\right)^2 \leq \sum_1^n \left(\frac{d_i^k}{x_i^k}\right)^2.$$

However our assumptions imply that $\{||\bar{w}^k||\}/\{||w^k||\}_{k \in S} \to 0$, as $k \to \infty$. This is a contradiction. ■

B Notation

VIP	variational inequality problem
$WVIP$	weak variational inequality problem
GGF	general geometric framework
FW	Frank-Wolfe method
AS	affine scaling method
f	VIP problem function
$K \subseteq \mathbb{R}^n$	feasible region
x^*, X^*	solution, solution set of a VIP
L_α	level set
$H(x)$	in GGF hyperplane $\{z \in \mathbb{R}^n \mid a'(z-x) \leq 0\}$ for vector $a \in \mathbb{R}^n$
$H(x)$	symmetrized Hessian of f
$Cut(y,x)$	$\{z \in K \mid f(y)'(z-x) \leq 0\}$
A, b	feasible region is later specified as $K = \{x \vert Ax = b, x \geq 0\}$
$S(x,R)$	ball centered in x of radius R
$M > 0$	$\Vert f(x) \Vert \leq M$
$L > 0$	$K \subseteq S(2^L)$
$l > 0$	$K \supseteq S(2^{-l})$
$\rho \geq 1$	Jacobian similarity constant
$\lambda > 0$	Lipschitz constant
L_1	$L_1 = L + 3l + log_2(M/\varepsilon)$
L_2	$L_2 = L + 3l + 2log_2\left(\frac{(M\lambda)^{0.5}\rho 2^{L+1}}{\varepsilon}\right)$
$C_p(x)$	primal gap function
$C_d(x)$	dual gap function
$C_d^1(x)$	approximation of C_d
$C_d^2(x)$	approximation of C_d restricted to Dikin ellipsoid
$g(y)$	$g(y) = f(x)'(x-y) - \rho \Vert x-y \Vert^2_{H(x)}$
P^k	"nice" set as defined in GGF
K^k	remaining part of a feasible region in GGF
$b(n) > 0$	$\mathrm{Vol}(P^{k+1}) \leq b(n)\mathrm{Vol}(P^k)$
$F(x)$	objective function such that $\nabla F = f$
$G(x)$	measure of the distance to the solution set
X	$\mathrm{diag}(x)$
P	$P = I - XA'(AX^2A')^{-1}AX$
$s(x)$	$s(x) = X^{-1}PXf(x)$
$D(x)$	$\{y \in K \mid \Vert X^{-1}d \Vert \leq r\}$ Dikin ellipsoid around x
y_x	$\arg\max_{y\in K} f(y)'(x-y)$
y_x^1	$\arg\max_{y\in K} f(x)'(x-y) - \rho\Vert x-y\Vert^2_{H(x)}$
y_x^2	$\arg\max_y \{f(x)'(x-y) - \rho\Vert x-y\Vert^2_{H(x)} \mid Ay = b, y \in D(x)\}$
x^I	an ε approximate VIP solution
x^{II}	an ε approximate $WVIP$ solution
N	$N = \{i \mid \bar{s}_i \neq 0\}$
B	$B = \{i \mid \bar{s}_i = 0\}$
S^*	$S^* = \{x \mid Xs(x) = 0,\ Ax = b,\ x \geq 0\}$
C_δ	$C_\delta = \{x \mid x_i \in [0,\delta)\ \forall i \in N\}$

C Definitions

- f is *monotone* if $(f(x) - f(y))'(x-y) \geq 0$ for all $x, y \in K$.
- f is *pseudomonotone* if $f(y)'(x-y) \geq 0$ implies that $f(x)'(x-y) \geq 0$, for all $x, y \in K$.
- f is *quasimonotone* if $f(y)'(x-y) > 0$ implies that $f(x)'(x-y) \geq 0$, for all $x, y \in K$.

- f is *strongly-f-monotone* if there exists $\alpha > 0$ such that $(f(y) - f(x)'(y-x) \geq \alpha \|f(x) - f(y)\|^2$ for all $x, y \in K$.
- f is *Lipschitz continuous* if for some constant $\lambda > 0$ it holds that $d'\nabla f(x)d \leq \lambda \|d\|^2$.
- f satisfies the *Jacobian similarity property* if for some constant $\rho \geq 1$, it holds that $d'\nabla f(x)d \leq \rho d'\nabla f(y)d$ for all $x, y \in K$, $d \in \mathbb{R}^n$.
- F is *convex* if $F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y)$ for all $x, y \in K$, $\alpha \in [0,1]$.
- F is *α-self-concordant* if $(F'''(x)[d,d,d])^{2/3} \leq \alpha d' F''(x)d$, for every $x \in K$, $d \in \mathbb{R}^n$.
- A point $(\bar{x}, \bar{s})$ satisfies the property of *strict complementarity* if $\bar{x}_i \bar{s}_i = 0$ and $\bar{x}_i + \bar{s}_i \neq 0$ for every i.
- A point $\bar{x}$ is a *weak* variational inequality solution if $f(y)'(y - x) \geq 0$ for all $y \in K$.
- A *primal gap function* $C_p(x) = \max_{y \in K} f(x)'(x-y)$.
- A *dual gap function* $C_d(x) = \max_{y \in K} f(y)'(x-y)$.
- A point x is an ε-approximate VIP solution if $C_p(x) \leq \varepsilon$.
- A point x is an ε-approximate $WVIP$ solution if $C_d(x) \leq \varepsilon$.

D Assumptions

1. K is a closed, bounded and convex set with a nonempty interior.
2. There exists a function $G : K \to R^+$ such that $x^* \in \arg\min_{x \in K} G(x)$ if and only if x^* is a solution of the VIP. Each of the level sets of G lies in a nonlinear convex set S as follows:

$$L_\alpha = \{z \mid G(z) \leq \alpha\} \subseteq S.$$

3. Given a point $y \in K$, and a small enough $\varepsilon > 0$, such that $\min_{x \in X^*} \|x - y\| \leq \varepsilon$, it follows that $G(y) \leq \min_{x \in X^*}[a(y)'(y-x) - c(y-x)'Q(y)(y-x)]$, where vector $a(y)$ has a bounded norm and matrix $Q(y)$ has a bounded operator norm.
4. f is a bounded function. That is, for some $M > 0$, $\|f(x)\| \leq M$, for any $x \in K$.
5. Problem function f exhibits the Jacobian similarity property.
6. A has full row rank.
7. AX^2A' is invertible for all $x \in K$.
8. The problem function f is Lipschitz continuous with Lipschitz constant λ.
9. $2l \geq \log n + 1; L_1 = L + 3l + \log(\frac{M}{\varepsilon})$.
10. The problem function f has a symmetric and positive semi-definite Jacobian matrix.
11. Every limit point $(\bar{x}, \bar{s})$ of the sequence $\{x^k, s(x^k)\}$ has the property of *strict complementarity*.
12. For some $\alpha > 0$, for all k, $\alpha^k \geq \alpha$.