January 30, 1994

EQUIVALENCE OF COMPLEMENTARITY PROBLEMS TO DIFFERENTIABLE MINIMIZATION: A UNIFIED APPROACH[1]

by

Paul Tseng[2], Nobuo Yamashita[3] and Masao Fukushima[3]

ABSTRACT

We consider two merit functions for a generalized nonlinear complementarity problem (GNCP for short) based on quadratic regularization of the standard gap function linearized. The first extends Fukushima's merit function for variational inequality problems [Fuk92] and the second extends Mangasarian and Solodov's implicit Lagrangian for complementarity problems [MaS93]. We show, among other things, that the second merit function is in the order of the natural residual squared and we give conditions under which the stationary points of this function are the solutions to GNCP. These results extend those of Luo et al. [LMRS92] and of Yamashita and Fukushima [YaF93] on the properties of the implicit Lagrangian.

KEY WORDS. Complementarity problem, merit function, error bound, natural residual, quadratic regularization.

---

[1] The research of the first author is supported by the National Science Foundation, Grant No. CCR-9103804.

[2] Department of Mathematics, University of Washington, Seattle, WA 98195, U.S.A.

[3] Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara 630-01, Japan.

1 Introduction

There has recently been a number of studies on casting a nonlinear complementarity problem (NCP for short) and variational inequality problem equivalently as the minimization of a "nice" function over a simple set. Notable among these is a work of Fukushima [Fuk92] which cast the variational inequality problem equivalently as the minimization of a regularized gap function over the feasible set and designed feasible descent methods for solving this minimization problem. Closely related is an earlier work of Auchmuty [Auc89], whose function $\mathcal{G}$ (with f chosen a quadratic) coincides with the regularized gap function when both are specialized to NCP. Mangasarian and Solodov [MaS93] cast NCP equivalently as the unconstrained minimization of an implicit Lagrangian, which resembles a primal-dual version of the regularized gap function. Subsequently, Luo et al. [LMRS92] showed that, in the case where the function in NCP is affine (so NCP reduces to a linear complementarity problem or LCP for short), the implicit Lagrangian is in the order of the natural residual squared (so the square root of the implicit Lagrangian provides an error bound for LCP if and only if the natural residual does); Yamashita and Fukushima [YaF93] showed that, in the case where the function in NCP is strongly monotone, the implicit Lagrangian has compact level sets and a unique stationary point which is the solution to NCP. Both the regularized gap function and the implicit Lagrangian have the nice feature that they are differentiable if the function in the problem is and, under additional assumptions, their minimum may be computed using second-order methods (see [Kan93], [MaS93], [TFI93]).

The preceding studies raise a number of intriguing questions. What is the connection between the regularized gap function and the implicit Lagrangian? Is there a unified approach to generating these and other functions with similar features? Does the connection between the implicit Lagrangian and the natural residual go beyond LCP? Does the connection between the stationary points of the implicit Lagrangian and the solution to NCP extend beyond the strongly monotone case? In this paper, we answer these questions in the affirmative. And our argument is surprisingly simple, using only basic facts about projection onto a closed convex cone and its polar.

We take as our problem setting the following generalized nonlinear complementarity problem (GNCP for short):

$$\text{[GNCP]} \qquad \text{Find an } x \in \Re^m \text{ satisfying} \quad G(x) \in \mathcal{K}, \quad H(x) \in \mathcal{K}^\circ, \quad \langle G(x), H(x) \rangle = 0,$$

where $\mathcal{K}$ is a nonempty closed convex cone in $\Re^n$, $\mathcal{K}^\circ$ denotes the polar of $\mathcal{K}$, and G and H are continuous functions from $\Re^m$ to $\Re^n$. Here and throughout, we denote by $\langle \cdot, \cdot \rangle$ the Euclidean inner product on $\Re^n$ and by $\| \cdot \|$ the norm induced by this inner product. [Much of our analysis also extends to the Hilbert space setting, but for simplicity of exposition we will keep to the Euclidean space setting.] In the special case where $m = n$, $\mathcal{K} = \Re^n_+$, $G(x) = x - E(x)$ and $H(x) = -F(x)$

for all x, with E and F some functions from $\Re^m$ to $\Re^n$, GNCP reduces to the so-called implicit complementarity problem [DoM80], [Pan81] (also called quasi-complementarity problem [Noo88]); in the further special case where $E(x) = 0$ for all x, GNCP reduces to NCP:

$$\text{[NCP]} \qquad \text{Find an } x \in \Re^m \text{ satisfying} \quad x \geq 0, \quad F(x) \geq 0, \quad \langle x, F(x) \rangle = 0.$$

[See [Pan94] for recent survey of NCP.]

Our goal is to cast GNCP equivalently as the minimization of a merit function over a simple set. [We say a real-valued function on $\Re^m$ is a merit function for GNCP over some subset of $\Re^m$ if the global minima of this function over the subset coincide with the solutions to GNCP.] Moreover, we like this merit function to be differentiable whenever G and H are. Certainly GNCP is equivalent to the minimization of the so-called "gap" function

$$p(G(x), H(x))$$

over all $x \in \Re^m$ satisfying $G(x) \in \mathcal{K}$ and $H(x) \in \mathcal{K}^\circ$, where $p : \Re^n \times \Re^n \mapsto \Re$ is given by

$$p(u, v) = -\langle u, v \rangle,$$

which is differentiable and nonnegative for all $(u, v) \in \mathcal{K} \times \mathcal{K}^\circ$ and equals 0 if and only if in addition $\langle u, v \rangle = 0$. However, this minimization problem suffers the serious drawback that its feasible set is very complicated. We can circumvent this difficulty by borrowing an idea from (partial) proximal minimization, namely, we can release a subset of the constraints from the minimization problem (and thus simplifying the latter's feasible set) by adding a quadratic regularization term associated with these constraints to p. In particular, we can release the constraints $H(x) \in \mathcal{K}^\circ$ by adding a quadratic regularization term on v to p or, more precisely, GNCP is equivalent to the minimization of

$$\hat{p}_\alpha(G(x), H(x))$$

over all $x \in \Re^m$ satisfying $G(x) \in \mathcal{K}$, where $\hat{p}_\alpha : \Re^n \times \Re^n \mapsto \Re$ is given by

$$\hat{p}_\alpha(u, v) = \min_{\zeta \in \mathcal{K}^\circ} \left\{ p(u, \zeta) + \frac{1}{2\alpha} \|v - \zeta\|^2 \right\},$$

and $\alpha \in (0, \infty)$. It can be seen that $\hat{p}_\alpha(u, v)$ is differentiable and nonnegative for all $(u, v) \in \mathcal{K} \times \Re^n$ and equals 0 if and only if in addition $v \in \mathcal{K}^\circ$ and $\langle u, v \rangle = 0$. [We can alternatively release the constraints $G(x) \in \mathcal{K}$ by adding a quadratic regularization term on u to p. The choice depends on which of G and H is more complex.] Finally, we can release all the constraints by adding a quadratic regularization term on both u and v to p or, more precisely, GNCP is equivalent to the minimization of

$$\tilde{p}_\alpha(G(x), H(x))$$

over all $x \in \Re^m$, where $\tilde{p}_\alpha : \Re^n \times \Re^n \mapsto \Re$ is given by

$$\tilde{p}_\alpha(u,v) = \min_{\substack{\psi \in \mathcal{K} \\ \zeta \in \mathcal{K}^\circ}} \left\{ p(\psi,\zeta) + \frac{1}{2\alpha}(\|u-\psi\|^2 + \|v-\zeta\|^2) \right\},$$

and $\alpha \in (0,\infty)$. It can be seen that $\tilde{p}_\alpha(u,v)$ is differentiable and nonnegative for all $(u,v) \in \Re^n \times \Re^n$ and equals 0 if and only if in addition $(u,v) \in \mathcal{K} \times \mathcal{K}^\circ$ and $\langle u,v \rangle = 0$. Thus, both $\hat{p}_\alpha$ and $\tilde{p}_\alpha$ may be viewed as smooth exact penalty functions for the constraints $v \in \mathcal{K}^\circ$ and $(u,v) \in \mathcal{K} \times \mathcal{K}^\circ$, respectively. Although $\hat{p}_\alpha$ and $\tilde{p}_\alpha$ are nice in that they are differentiable, they suffer the drawback of being too difficult to compute. One way to simplify the computation is to replace $p(u,\zeta)$ in the definition of $\hat{p}_\alpha(u,v)$ and $p(\psi,\zeta)$ in the definition of $\tilde{p}_\alpha(u,v)$ by, respectively, the linear approximations $p(u,v) + \langle \nabla_v p(u,v), \zeta - v \rangle$ and $p(u,v) + \langle \nabla_u p(u,v), \psi - u \rangle + \langle \nabla_v p(u,v), \zeta - v \rangle$. Thus, we consider replacing $\hat{p}_\alpha$ and $\tilde{p}_\alpha$ by, respectively, the functions $\hat{\pi}_\alpha : \Re^n \times \Re^n \mapsto \Re$ and $\tilde{\pi}_\alpha : \Re^n \times \Re^n \mapsto \Re$ given by

$$\begin{aligned}
\hat{\pi}_\alpha(u,v) &= \min_{\zeta \in \mathcal{K}^\circ} \left\{ p(u,v) - \langle u, \zeta - v \rangle + \frac{1}{2\alpha}\|v-\zeta\|^2 \right\} \\
&= -\langle u, [v+\alpha u]^+_{\mathcal{K}^\circ} \rangle + \frac{1}{2\alpha}\|v - [v+\alpha u]^+_{\mathcal{K}^\circ}\|^2, \\
\tilde{\pi}_\alpha(u,v) &= \min_{\substack{\psi \in \mathcal{K} \\ \zeta \in \mathcal{K}^\circ}} \left\{ p(u,v) - \langle v, \psi - u \rangle - \langle u, \zeta - v \rangle + \frac{1}{2\alpha}(\|u-\psi\|^2 + \|v-\zeta\|^2) \right\} \\
&= \langle u,v \rangle - \langle v, [u+\alpha v]^+_{\mathcal{K}} \rangle - \langle u, [v+\alpha u]^+_{\mathcal{K}^\circ} \rangle + \frac{1}{2\alpha}(\|u - [u+\alpha v]^+_{\mathcal{K}}\|^2 + \|v - [v+\alpha u]^+_{\mathcal{K}^\circ}\|^2)
\end{aligned}$$

with $\alpha \in (0,\infty)$. [Throughout, for any nonempty closed convex set C in $\Re^n$, we denote by $[\cdot]^+_C$ the orthogonal projection onto C.] This yields the functions $\hat{m}_\alpha : \Re^m \mapsto \Re$ and $\tilde{m}_\alpha : \Re^m \mapsto \Re$ given by

$$\begin{aligned}
\hat{m}_\alpha(x) &= \hat{\pi}_\alpha(G(x), H(x)), \\
\tilde{m}_\alpha(x) &= \tilde{\pi}_\alpha(G(x), H(x)),
\end{aligned}$$

with $\alpha \in (0,\infty)$. These two functions will be the focus of our study.

Our main contributions are as follows: First, we show that, for any $\alpha \in (0,\infty)$, $\hat{m}_\alpha$ is a merit function for GNCP on $G^{-1}(\mathcal{K})$, is at least in the order of the natural residual squared on $G^{-1}(\mathcal{K})$, and is differentiable whenever G and H are (see Theorem 3.1). Similarly, we show that, for any $\alpha \in (0,1)$ (respectively, $\alpha \in (1,\infty)$), $\tilde{m}_\alpha$ (respectively, $-\tilde{m}_\alpha$) is a merit function for GNCP on $\Re^m$, is in the order of the natural residual squared on $\Re^m$, and is differentiable whenever G and H are (see Theorem 3.2). Moreover, in the case where $G(x) = x$ for all x, the function $\hat{m}_\alpha$ reduces to the regularized gap function and that, in the case of NCP, the function $-\tilde{m}_\alpha$ reduces to the implicit Lagrangian. The preceding results on $\tilde{m}_\alpha$ extend [MaS93, Theorem 2.1] and [LMRS92, Theorems 3.2, 4.1] and, in particular, show that the implicit Lagrangian acts as a merit function for NCP on $\Re^m$ not only for $\alpha > 1$ but also for $\alpha < 1$, and that it is in the order of the natural residual squared

not only for LCP but for NCP in general. We also give, as a byproduct, an extension of [LMRS92, Theorem 3.1]. Second, we derive conditions on G and H under which the stationary points of $\tilde{m}_\alpha$ coincide with the solutions to GNCP (see Theorems 4.1 and 4.2). These results extend [YaF93, Theorem 2.2] which considers only NCP and further assumes $\alpha > 1$. Third, we show that, in the case $\mathcal{K} = \Re^n_+$, the global minima (respectively, global maxima) of $\tilde{m}_\alpha$, for $\alpha \in (0,1)$ (respectively, $\alpha \in (1,\infty)$), are isolated under a certain linear independence assumption on the columns of ∇G and ∇H (see Theorem 5.1). This result extends [MaS93, Theorem 2.3] which considers only NCP and further assumes $\alpha > 1$. Lastly, we consider the case $\mathcal{K} = \Re^n_+$ and compare $\tilde{m}_\alpha$ with the implicit Lagrangian applied to a reformulation of GNCP as an NCP.

A few words about notation. We denote by $\Re^n$ the n-dimensional real space and by $\Re^n_+$ the nonnegative orthant in $\Re^n$. For any function $F : \Re^m \mapsto \Re^n$, we denote by F_j the jth component of F, $j = 1, ..., n$, and by ∇F the $m \times n$ matrix whose jth column is the gradient function ∇F_j, $j = 1, ..., n$. For any real matrix A, we denote by A^* its transpose.

2 Properties of $\hat{\pi}_\alpha$ and of $\tilde{\pi}_\alpha$

It is well known and easily shown that an $(u, v) \in \Re^n \times \Re^n$ satisfies $(u, v) \in \mathcal{K} \times \mathcal{K}^\circ$ and $\langle u, v\rangle = 0$ if and only if they satisfy the fixed point equations $u = [u + v]^+_{\mathcal{K}}$ or, equivalently (by using the fact $u + v = [u + v]^+_{\mathcal{K}} + [u + v]^+_{\mathcal{K}^\circ})$ $v = [u + v]^+_{\mathcal{K}^\circ}$. Thus, the quantity

$$\|u - [u + v]^+_{\mathcal{K}}\| \tag{1}$$

measures how close (u, v) comes to satisfying the equations. In this section we analyze the differentiability of the functions $\hat{\pi}_\alpha$ and $\tilde{\pi}_\alpha$ and express their growth rate in terms of the quantity (1). Our analysis will frequently use the following fact about projection (see [GaB84, Lemma 1]):

$$\begin{array}{lll} \min\{1,\alpha\}\|u - [u + v]^+_{\mathcal{K}}\| & \le \|u - [u + \alpha v]^+_{\mathcal{K}}\| & \le \max\{1,\alpha\}\|u - [u + v]^+_{\mathcal{K}}\|, \\ \min\{1,\alpha\}\|u - [u + v]^+_{\mathcal{K}}\| & \le \|v - [v + \alpha u]^+_{\mathcal{K}^\circ}\| & \le \max\{1,\alpha\}\|u - [u + v]^+_{\mathcal{K}}\| \end{array} \tag{2}$$

for all $(u, v) \in \Re^n \times \Re^n$ and all $\alpha \in (0, \infty)$.

The following proposition states the key properties of $\hat{\pi}_\alpha$, showing that $\hat{\pi}_\alpha$ is differentiable and $\hat{\pi}_\alpha(u, v)$ is at least in the order of the quantity (1) squared on $\mathcal{K} \times \Re^n$.

Proposition 2.1 *For any $\alpha \in (0, \infty)$, the function $\hat{\pi}_\alpha$ is differentiable everywhere, with*

$$\nabla\hat{\pi}_\alpha(u, v) = \begin{bmatrix} -[v + \alpha u]^+_{\mathcal{K}^\circ} \\ \frac{1}{\alpha}(v - [v + \alpha u]^+_{\mathcal{K}^\circ}) \end{bmatrix},$$

and satisfies

$$\hat{\pi}_\alpha(u,v) \;\geq\; \frac{\min\{1,\alpha\}^2}{2\alpha}\|u-[u+v]^+_{\mathcal{K}}\|^2 \tag{3}$$

for all $(u,v)\in\mathcal{K}\times\Re^n$. *Moreover,* $\hat{\pi}_\alpha(u,v)$ *is nonnegative for all* $(u,v)\in\mathcal{K}\times\Re^n$ *and equals 0 if and only if in addition* $v\in\mathcal{K}^\circ$ *and* $\langle u,v\rangle=0$.

Proof. That $\hat{\pi}_\alpha$ is differentiable everywhere and its gradient has the given form follow from a result of Danskin (see [Aus76, Ch. 4, Theorem 1.7] or [BeT89, p. 649]). Since $u\in\mathcal{K}$, we have $\langle u,[v+\alpha u]^+_{\mathcal{K}^\circ}\rangle\leq 0$ and the formula for $\hat{\pi}_\alpha$ yields

$$\hat{\pi}_\alpha(u,v)\geq\frac{1}{2\alpha}\|v-[v+\alpha u]^+_{\mathcal{K}^\circ}\|^2.$$

Using (2) to bound below the righthand side yields (3). The remaining results readily follow from (3). □

On the other hand, except in the case where $\mathcal{K}$ is a subspace (so $\mathcal{K}$ and $\mathcal{K}^\circ$ are orthogonal complements), the function $\hat{\pi}_\alpha(u,v)$ always grows faster than the residual (1) squared (or raised to any power) on $\mathcal{K}\times\Re^n$. To see this, assume $\mathcal{K}$ is not a subspace, so there is some point z in the span of $\mathcal{K}$ (i.e., the smallest subspace containing $\mathcal{K}$) that is not in $\mathcal{K}\cup\mathcal{K}^\circ$. Let $w=z-[z]^+_{\mathcal{K}}$ (so w is an outer normal vector to $\mathcal{K}$ at $[z]^+_{\mathcal{K}}$ and, since $\mathcal{K}$ is a convex cone, $\langle[z]^+_{\mathcal{K}},w\rangle=0$) and let

$$u=[z]^+_{\mathcal{K}}-\epsilon w,$$

where $\epsilon\in(0,\infty)$ is chosen sufficiently small so that u is in $\mathcal{K}$. [Such an ϵ exists by our choice of z.] Since w is an outer normal vector to $\mathcal{K}$ at $[z]^+_{\mathcal{K}}$, we have that for any $\lambda\in(0,\infty)$, the point

$$v(\lambda)=(\lambda+\epsilon)w$$

satisfies $[u+v(\lambda)]^+_{\mathcal{K}}=[z]^+_{\mathcal{K}}$ and $[u+v(\lambda)]^+_{\mathcal{K}^\circ}=\lambda w$. The former yields $\|[u+v(\lambda)]^+_{\mathcal{K}}-u\|=\|[z]^+_{\mathcal{K}}-u\|=\epsilon\|w\|$ and the latter yields $-\langle u,[u+v(\lambda)]^+_{\mathcal{K}^\circ}\rangle=-\langle[z]^+_{\mathcal{K}}-\epsilon w,\lambda w\rangle=\epsilon\lambda\|w\|^2$ which, as $\lambda\to\infty$, grows faster than any power of $\epsilon\|w\|$.

Also, we note that, by using the following properties of projection onto a convex cone and its polar:

$$\langle[\frac{v}{\alpha}+u]^+_{\mathcal{K}},[v+\alpha u]^+_{\mathcal{K}^\circ}\rangle=0,\quad [v+\alpha u]^+_{\mathcal{K}^\circ}=v+\alpha u-[v+\alpha u]^+_{\mathcal{K}}=v+\alpha(u-[\frac{v}{\alpha}+u]^+_{\mathcal{K}}),$$

we may rewrite $\hat{\pi}_\alpha(u,v)$ as

$$\begin{aligned}\hat{\pi}_\alpha(u,v) &= -\langle u-[\frac{v}{\alpha}+u]^+_{\mathcal{K}},v+\alpha(u-[\frac{v}{\alpha}+u]^+_{\mathcal{K}})\rangle+\frac{\alpha}{2}\|u-[\frac{v}{\alpha}+u]^+_{\mathcal{K}}\|^2\\ &= -\langle u-[\frac{v}{\alpha}+u]^+_{\mathcal{K}},v\rangle-\frac{\alpha}{2}\|u-[\frac{v}{\alpha}+u]^+_{\mathcal{K}}\|^2\\ &= \max_{\psi\in\mathcal{K}}\left\{-\langle u-\psi,v\rangle-\frac{\alpha}{2}\|u-\psi\|^2\right\}.\end{aligned}\tag{4}$$

The alternative formula (4) for $\hat{\pi}_\alpha(u,v)$ will be used in the proof of Proposition 2.2 below and in relating $\hat{m}_\alpha$ to the regularized gap function.

The following proposition states the key properties of $\tilde{\pi}_\alpha$, showing that $\tilde{\pi}_\alpha$ is differentiable, has nice symmetry with respect to α, and $\tilde{\pi}_\alpha(u,v)$ is in the order of the quantity (1) squared on $\Re^n \times \Re^n$.

Proposition 2.2 *For any $\alpha \in (0,\infty)$, the function $\tilde{\pi}_\alpha$ is differentiable everywhere, with*

$$\nabla \tilde{\pi}_\alpha(u,v) = \begin{bmatrix} v - [v+\alpha u]^+_{\mathcal{K}^\circ} + \frac{1}{\alpha}(u - [u+\alpha v]^+_{\mathcal{K}}) \\ u - [u+\alpha v]^+_{\mathcal{K}} + \frac{1}{\alpha}(v - [v+\alpha u]^+_{\mathcal{K}^\circ}) \end{bmatrix},$$

and satisfies

$$(1-\alpha)\|u - [u+v]^+_{\mathcal{K}}\|^2 \;\leq\; \tilde{\pi}_\alpha(u,v) \;=\; -\tilde{\pi}_{1/\alpha}(u,v) \;\leq\; (1/\alpha - 1)\|u - [u+v]^+_{\mathcal{K}}\|^2 \qquad (5)$$

for all $(u,v) \in \Re^n \times \Re^n$. Moreover, for $\alpha \in (0,1)$ (respectively, $\alpha \in (1,\infty)$), we have that $\tilde{\pi}_\alpha(u,v)$ is nonnegative (respectively, nonpositive) for all $(u,v) \in \Re^n \times \Re^n$ and equals 0 if and only if $(u,v) \in \mathcal{K} \times \mathcal{K}^\circ$ and $\langle u, v\rangle = 0$. For $\alpha = 1$, we have $\tilde{\pi}_\alpha \equiv 0$.

Proof. That $\tilde{\pi}_\alpha$ is differentiable everywhere and its gradient has the given form follow from a result of Danskin (see [Aus76, Ch. 4, Theorem 1.7] or [BeT89, p. 649]). To prove (5), consider any $(u,v) \in \Re^n \times \Re^n$. We have from (4) that

$$-\langle u, [v+\alpha u]^+_{\mathcal{K}^\circ}\rangle + \frac{1}{2\alpha}\|v - [v+\alpha u]^+_{\mathcal{K}^\circ}\|^2 = \max_{\psi \in \mathcal{K}}\{-\langle u-\psi, v\rangle - \frac{\alpha}{2}\|u-\psi\|^2\},$$

and analogously (by switching the roles of u and v and of $\mathcal{K}$ and $\mathcal{K}^\circ$ in the above equation) that

$$-\langle v, [u+\alpha v]^+_{\mathcal{K}}\rangle + \frac{1}{2\alpha}\|u - [u+\alpha v]^+_{\mathcal{K}}\|^2 = \max_{\zeta \in \mathcal{K}^\circ}\{-\langle v-\zeta, u\rangle - \frac{\alpha}{2}\|v-\zeta\|^2\},$$

which when used in the formula for $\tilde{\pi}_\alpha(u,v)$ yields

$$\begin{aligned}
\tilde{\pi}_\alpha(u,v) &= \langle u, v\rangle + \max_{\psi \in \mathcal{K}}\{-\langle u-\psi, v\rangle - \frac{\alpha}{2}\|u-\psi\|^2\} + \max_{\zeta \in \mathcal{K}^\circ}\{-\langle v-\zeta, u\rangle - \frac{\alpha}{2}\|v-\zeta\|^2\} \\
&= -p(u,v) - \min_{\psi \in \mathcal{K}}\{\langle u-\psi, v\rangle + \frac{\alpha}{2}\|u-\psi\|^2\} - \min_{\zeta \in \mathcal{K}^\circ}\{\langle v-\zeta, u\rangle + \frac{\alpha}{2}\|v-\zeta\|^2\} \\
&= -\tilde{\pi}_{1/\alpha}(u,v),
\end{aligned}$$

and the equality in (5) is proven. Next, we note that $\tilde{\pi}_\alpha(u,v)$ equals the minimum of the following quantity:

$$\begin{aligned}
& p(u,v) - \langle \psi - u, v\rangle - \langle u, \zeta - v\rangle + \frac{1}{2\alpha}(\|u-\psi\|^2 + \|v-\zeta\|^2) \\
=\;& \langle u, v\rangle - \langle \psi, v\rangle - \langle u, \zeta\rangle + \frac{1}{2\alpha}(\|u-\psi\|^2 + \|v-\zeta\|^2) \\
=\;& \langle \psi - u, \zeta - v\rangle - \langle \psi, \zeta\rangle + \frac{1}{2\alpha}(\|u-\psi\|^2 + \|v-\zeta\|^2) \\
=\;& \frac{1}{2}\|\psi - u + \zeta - v\|^2 - \langle \psi, \zeta\rangle + \frac{1}{2}\left(\frac{1}{\alpha} - 1\right)(\|u-\psi\|^2 + \|v-\zeta\|^2) \qquad (6)
\end{aligned}$$

over all (ψ, ζ) in $\mathcal{K} \times \mathcal{K}^\circ$, so $\tilde{\pi}_\alpha(u, v)$ is less than or equal to the value of this quantity at $\psi = [u+v]^+_{\mathcal{K}}$ and $\zeta = [u+v]^+_{\mathcal{K}^\circ}$. With this choice of ψ and ζ, the first two terms of (6) equal zero and $u-\psi = \zeta-v$, so the quantity (6) has value equal to the righthand side of (5) and the second inequality in (5) follows. To prove the first inequality in (5), we note that the second inequality in (5) still holds when α is replaced by $1/\alpha$, so that

$$\tilde{\pi}_\alpha(u, v) = -\tilde{\pi}_{1/\alpha}(u, v) \geq -(\alpha - 1)\|u - [u + v]^+_{\mathcal{K}}\|^2.$$

The remaining results readily follow from (5). □

An interesting question concerns whether $\tilde{\pi}_\alpha$ and Proposition 2.2 extend to other choices of the function p. This does not seem possible for the equality or the second inequality in (5) but is possible for the first inequality under very general setting. In particular, let $\mathcal{C}$ be any closed convex set in $\Re^l$ for some l and let p be any continuously differentiable function on $\Re^l$ that is nonnegative on $\mathcal{C}$ and has at most quadratic growth, i.e.,

$$p(\zeta) \leq p(z) + \langle \nabla p(z), \zeta - z \rangle + \frac{1}{2a}\|\zeta - z\|^2 \qquad \forall z,\ \zeta \in \Re^l,$$

for some scalar $a > 0$. Then, it is readily seen that $\tilde{\pi}_\alpha$ defined by

$$\tilde{\pi}_\alpha(z) = \min_{\zeta \in \mathcal{C}} \left\{ p(z) + \langle \nabla p(z), \zeta - z \rangle + \frac{1}{2\alpha}\|\zeta - z\|^2 \right\}$$

satisfies

$$\tilde{\pi}_\alpha(z) \geq \frac{1}{2}\left(\frac{1}{\alpha} - \frac{1}{a}\right) \|z - [z - \alpha \nabla p(z)]^+_{\mathcal{C}}\|^2$$

for all $z \in \Re^l$. Thus, the first inequality in (5) depends on nothing more than the fact that the bilinear function $(u, v) \mapsto -\langle u, v \rangle$ is nonnegative on $\mathcal{K} \times \mathcal{K}^\circ$ and has at most quadratic growth.

3 $\hat{\pi}_\alpha$ and $\tilde{\pi}_\alpha$: Merit Functions and Error Bounds for GNCP

In this section, we study how $\hat{m}_\alpha$ and $\tilde{m}_\alpha$ act as merit functions and provide error bounds for GNCP on, respectively, $G^{-1}(\mathcal{K})$ and $\Re^m$. We will also use the relationship between $\hat{p}_\alpha$ and $\tilde{p}_\alpha$ and the quantity (1) (see Propositions 2.1 and 2.2) to relate $\hat{m}_\alpha$ and $\tilde{m}_\alpha$ to the norm of the natural residual functions $R_\alpha : \Re^m \mapsto \Re^n$ and $S_\alpha : \Re^m \mapsto \Re^n$ given by

$$R_\alpha(x) = G(x) - [G(x) + \alpha H(x)]^+_{\mathcal{K}}, \qquad S_\alpha(x) = H(x) - [H(x) + \alpha G(x)]^+_{\mathcal{K}^\circ}$$

for all $x \in \Re^m$, with $\alpha \in (0, \infty)$. Notice that an $x \in \Re^m$ is a solution to GNCP if and only if $R_\alpha(x) = S_\alpha(x) = 0$, so $\|R_\alpha(x)\|$ (or $\|S_\alpha(x)\|$) raised to any positive power is a merit function for

GNCP on $\Re^m$. However, such a merit function has the serious drawback that it is not differentiable even when G and H are.

We have, as a consequence of Proposition 2.1, the first main result showing that $\hat{m}_\alpha$ is a merit function for GNCP on $G^{-1}(\mathcal{K})$, is at least in the order of $\|R_1\|$ squared, and is differentiable whenever G and H are.

Theorem 3.1 *For any $\alpha \in (0, \infty)$, the function $\hat{m}_\alpha$ satisfies*

$$\hat{m}_\alpha(x) \geq \frac{\min\{1, \alpha\}^2}{2\alpha} \|R_1(x)\|^2$$

for all $x \in \Re^m$ with $G(x) \in \mathcal{K}$. Moreover, $\hat{m}_\alpha(x)$ is nonnegative for all $x \in \Re^m$ with $G(x) \in \mathcal{K}$ and equals 0 if and only if x is a solution to GNCP. If in addition G and H are differentiable, then so is $\hat{m}_\alpha$ and

$$\nabla \hat{m}_\alpha(x) = \nabla G(x) \left(S_\alpha(x) - H(x)\right) + \frac{1}{\alpha} \nabla H(x) S_\alpha(x)$$

for all $x \in \Re^m$.

In the special case where $m = n$, $G(x) = x$ for all x and $H \equiv -F$ for some $F : \Re^m \mapsto \Re^n$, it is readily seen from the alternative formula (4) for $\hat{\pi}_\alpha$ that the function $\hat{m}_\alpha$ reduces to Fukushima's regularized gap function f defined in [Fuk92] (assuming, without loss of generality, that $G = I$ therein). On the other hand, the latter function is defined not only for convex cones but more generally for convex sets.

We have, as a consequence of Proposition 2.2, the second main result showing that $\tilde{m}_\alpha$ has nice symmetry with respect to α, is a merit function for GNCP on $\Re^m$, is in the order of $\|R_1\|$ squared, and is differentiable whenever G and H are.

Theorem 3.2 *For any $\alpha \in (0, \infty)$, the function $\tilde{m}_\alpha$ satisfies*

$$(1 - \alpha)\|R_1(x)\|^2 \leq \tilde{m}_\alpha(x) = -\tilde{m}_{1/\alpha}(x) \leq (1/\alpha - 1)\|R_1(x)\|^2$$

for all $x \in \Re^m$. Moreover, for $\alpha \in (0, 1)$ (respectively, $\alpha \in (1, \infty)$), we have that $\tilde{m}_\alpha(x)$ is nonnegative (respectively, nonpositive) for all $x \in \Re^m$ and equals 0 if and only if x is a solution to GNCP. If in addition G and H are differentiable, then so is $\tilde{m}_\alpha$ and

$$\nabla \tilde{m}_\alpha(x) = \nabla G(x) \left(S_\alpha(x) + \frac{1}{\alpha} R_\alpha(x)\right) + \nabla H(x) \left(R_\alpha(x) + \frac{1}{\alpha} S_\alpha(x)\right)$$

for all $x \in \Re^m$.

We claim that, in the special case of NCP (corresponding to GNCP with $m = n$, $\mathcal{K} = \Re^n_+$, $G(x) = x$ for all x, and $H \equiv -F$ for some $F : \Re^m \mapsto \Re^n$), the function $-\tilde{m}_\alpha$ reduces to Mangasarian and Solodov's implicit Lagrangian [MaS93]. This follows from the observation that, when $\mathcal{K} = \Re^n_+$, the function $\tilde{\pi}_\alpha$ simplifies to

$$\tilde{\pi}_\alpha(u,v) = \langle u, v\rangle + \frac{1}{2\alpha}\left(\|u\|^2 - \|[u+\alpha v]_+\|^2 + \|v\|^2 - \|[v+\alpha u]_+\|^2\right),$$

where $[\cdot]_+$ denotes the positive part of a vector. It was shown in [LMRS92] (see Theorem 4.1 therein) that, in the further case where F is affine (so NCP reduces to an LCP), the implicit Lagrangian with $\alpha \in (1,\infty)$ is between $(1-1/\alpha)\|x-[x-F(x)]_+\|^2$ and $(\alpha-1)\|x-[x-F(x)]_+\|^2$ for all $x \in \Re^m$. By using Theorem 3.2, we can further conclude that, regardless whether F is affine, the implicit Lagrangian with $\alpha \in (0,\infty)$ is between $(1-1/\alpha)\|x-[x-F(x)]_+\|^2$ and $(\alpha-1)\|x-[x-F(x)]_+\|^2$ for all $x \in \Re^m$.

It follows from Theorem 3.1 that, for any $\alpha \in (0,\infty)$, $\sqrt{\hat{m}_\alpha}$ provides an error bound for GNCP on $G^{-1}(\mathcal{K})$ (i.e., the Euclidean distance from any $x \in \Re^m$ satisfying $G(x) \in \mathcal{K}$ to the solution set of GNCP is bounded above by some constant times $\sqrt{\hat{m}_\alpha(x)}$) if $\|R_1\|$ does. Similarly, it follows from Theorem 3.2 that, for any $\alpha \in (0,\infty)$ with $\alpha \neq 1$, $\sqrt{|\tilde{m}_\alpha|}$ provides an error bound for GNCP on $\Re^m$ if and only if $\|R_1\|$ does. Moreover, $\hat{m}_\alpha$ has bounded level sets on $G^{-1}(\mathcal{K})$ if the solution set of GNCP is bounded and $\|R_1\|$ provides a global error bound for GNCP on $G^{-1}(\mathcal{K})$; $|\tilde{m}_\alpha|$ has bounded level sets on $\Re^m$ if the solution set of GNCP is bounded and $\|R_1\|$ provides a global error bound for GNCP on $\Re^m$. In the case $G(x) = x$ for all x, necessary and sufficient conditions for $\|R_1\|$ to provide a (local or global) error bound for GNCP have been much studied and are discussed in [LMRS92], [LuT92], [LuT93], [Man90], [Pan87], [MaP90]. For example, it is known (see [Pan87, Theorem 3.1]) that the solution set of GNCP is bounded (in fact, unique) and $\|R_1\|$ provides a global error bound for GNCP on $\Re^m$ if $G(x) = x$ for all x and $-H$ is strongly monotone and Lipschitz continuous on $\Re^m$. It follows that, in this case, $\hat{m}_\alpha$ has bounded level sets on $G^{-1}(\mathcal{K})$ and $|\tilde{m}_\alpha|$ has bounded level sets on $\Re^m$. The latter result improves on [YaF93, Theorem 2.4] which considers only the case $\mathcal{K} = \Re^n_+$ and further assumes α is sufficiently large.

Although $\tilde{m}_\alpha$ is more complicated than $\hat{m}_\alpha$, $\tilde{m}_\alpha$ has the advantage that it works on all of $\Re^m$, not only on $G^{-1}(\mathcal{K})$, and that it grows no faster than $\|R_1\|$ squared on $\Re^m$ (whereas $\hat{m}_\alpha$ can grow faster than $\|R_1\|$ raised to any power on $G^{-1}(\mathcal{K})$, as is suggested by the discussion following Proposition 2.1). Lastly, it can be seen from the formula for $\nabla\tilde{m}_\alpha$ given in Theorem 3.2 that, in the case where ∇G and ∇H are bounded on $\Re^m$, the norm of $\nabla\tilde{m}_\alpha$ is bounded above by a constant times $\|R_1\|$ on $\Re^m$. Thus, $\tilde{m}_\alpha$ has the additional nice property that its gradient does not grow too fast in norm.

Finally, it was shown in [LMRS92, Theorem 3.1] that, in the case of LCP, the square root of the partial derivative of 2α times the implicit Lagrangian (with respect to $\alpha \in (1,\infty)$) provides a

local error bound for LCP on $\Re^m$. We can extend this result to $\hat{m}_\alpha$ and $\tilde{m}_\alpha$ by using Danskin's result to compute the partial derivative of $\hat{m}_\alpha$ and $\tilde{m}_\alpha$ with respect to $\alpha \in (0,\infty)$, yielding

$$\begin{aligned}\frac{\partial}{\partial\alpha}\hat{m}_\alpha(x) &= -\frac{1}{2\alpha^2}\|S_\alpha(x)\|^2,\\ \frac{\partial}{\partial\alpha}\tilde{m}_\alpha(x) &= -\frac{1}{2\alpha^2}(\|R_\alpha(x)\|^2+\|S_\alpha(x)\|^2)\end{aligned}$$

for all $x \in \Re^m$ and all $\alpha \in (0,\infty)$. Using (2), it readily follows that $\sqrt{-\frac{\partial}{\partial\alpha}\hat{m}_\alpha}$ and $\sqrt{-\frac{\partial}{\partial\alpha}\tilde{m}_\alpha}$ provide local/global error bound for GNCP on $\Re^m$ (or any subset of $\Re^m$) if and only if $\|R_1\|$ does. By using Theorem 3.2 and the simple fact that

$$\frac{\partial}{\partial\alpha}[-2\alpha\tilde{m}_\alpha(x)] = -2\tilde{m}_\alpha(x) - 2\alpha\frac{\partial}{\partial\alpha}\tilde{m}_\alpha(x)$$

for all $x \in \Re^m$, we can further conclude that $\sqrt{\frac{\partial}{\partial\alpha}[-2\alpha\tilde{m}_\alpha]}$, with $\alpha \in (1,\infty)$, provides local/global error bound for GNCP on $\Re^m$ if and only if $\|R_1\|$ does. This result, when specialized to LCP for which it is known that $\|R_1\|$ provides local error bound on $\Re^m$ (see [LuT93]), yields [LMRS92, Theorem 3.1].

4 Equivalence of Stationary Points of $\tilde{m}_\alpha$ to Solutions to GNCP

It was shown in [Fuk92] that the regularized gap function has the nice property that, when the function in the variational inequality problem is strongly monotone, all its stationary points are solutions to the variational inequality problem. It was shown in [YaF93] that the implicit Lagrangian has the analogous property that, when the function F in NCP is strongly monotone, all its stationary points are solutions to NCP. Our third main result below extends [YaF93, Theorem 2.2] by giving conditions on G and H under which the stationary points of $\tilde{m}_\alpha$ are the solutions to GNCP. [It is desirable that the stationary points of $\tilde{m}_\alpha$ be solutions to GNCP since many methods for unconstrained minimization are guaranteed to find only stationary points.]

Theorem 4.1 *Assume G and H are differentiable. Then every solution to GNCP is a stationary point of $\tilde{m}_\alpha$, for $\alpha \in (0,\infty)$. Assume in addition that $m = n$ and G and H satisfy the following condition:*

A1: For every $x \in \Re^m$, the Jacobian $\nabla G(x)$ is nonsingular and there exists a real $n \times n$ matrix $\Lambda(x)$ such that $-\Lambda(x)[\nabla G(x)]^{-1}\nabla H(x)$ is positive definite and $\Lambda(x)$ has a block diagonal structure matching the Cartesian product structure of $\mathcal{K}$, with every block being some positive multiple of the identity matrix (see (9)).

Then, every stationary point of $\tilde{m}_\alpha$, for $\alpha \in (0,\infty)$ and $\alpha \neq 1$, is a solution to GNCP.

Proof. Since G and H are differentiable, by Theorem 3.2, $\tilde{m}_\alpha$ is differentiable and an $x \in \Re^m$ is a stationary point of $\tilde{m}_\alpha$ if and only if

$$0 = \nabla G(x) \left(S_\alpha(x) + \frac{1}{\alpha} R_\alpha(x) \right) + \nabla H(x) \left(R_\alpha(x) + \frac{1}{\alpha} S_\alpha(x) \right). \tag{7}$$

Thus, if x is a solution to GNCP so $R_\alpha(x) = S_\alpha(x) = 0$, then x clearly satisfies (7) and hence is a stationary point of $\tilde{m}_\alpha$. Now, assume in addition that $m = n$ and G and H satisfy condition A1. If x satisfies (7), taking the inner product of both sides of (7) with $(\Lambda(x)[\nabla G(x)]^{-1})^*(R_\alpha(x) + \frac{1}{\alpha} S_\alpha(x))$ yields (for brevity, we drop the "x" argument in what follows)

$$0 = \langle R_\alpha + \frac{1}{\alpha} S_\alpha, \Lambda(S_\alpha + \frac{1}{\alpha} R_\alpha) \rangle + \langle R_\alpha + \frac{1}{\alpha} S_\alpha, \Lambda[\nabla G]^{-1} \nabla H \left(R_\alpha + \frac{1}{\alpha} S_\alpha \right) \rangle. \tag{8}$$

First, we show that the first righthand term in (8) is nonpositive. We have that $G + \alpha H = [G + \alpha H]^+_{\mathcal{K}} + [G + \alpha H]^+_{\mathcal{K}^\circ}$ and $H + \alpha G = [H + \alpha G]^+_{\mathcal{K}} + [H + \alpha G]^+_{\mathcal{K}^\circ}$. We also have, by the assumption on Λ, that (by rearranging the coordinates if necessary)

$$\Lambda = \begin{bmatrix} \lambda_1 I & & \\ & \ddots & \\ & & \lambda_l I \end{bmatrix}, \qquad \mathcal{K} = \mathcal{K}_1 \times \cdots \times \mathcal{K}_l, \tag{9}$$

for some positive integer l, some positive scalars $\lambda_1, ..., \lambda_l$, and some closed convex cones $\mathcal{K}_1, ..., \mathcal{K}_l$. Using these two facts, we see that

$$\begin{aligned}
& \langle R_\alpha + \frac{1}{\alpha} S_\alpha, \Lambda(S_\alpha + \frac{1}{\alpha} R_\alpha) \rangle \\
= \; & \langle -[G + \alpha H]^+_{\mathcal{K}} + \frac{1}{\alpha}[H + \alpha G]^+_{\mathcal{K}}, \Lambda(-[H + \alpha G]^+_{\mathcal{K}^\circ} + \frac{1}{\alpha}[G + \alpha H]^+_{\mathcal{K}^\circ}) \rangle \\
= \; & \sum_{j=1}^{l} \lambda_j \left[\langle -[G_j + \alpha H_j]^+_{\mathcal{K}_j} + \frac{1}{\alpha}[H_j + \alpha G_j]^+_{\mathcal{K}_j}, -[H_j + \alpha G_j]^+_{(\mathcal{K}_j)^\circ} + \frac{1}{\alpha}[G_j + \alpha H_j]^+_{(\mathcal{K}_j)^\circ} \rangle \right] \\
= \; & \sum_{j=1}^{l} \lambda_j \left[\langle [G_j + \alpha H_j]^+_{\mathcal{K}_j}, [H_j + \alpha G_j]^+_{(\mathcal{K}_j)^\circ} \rangle + \frac{1}{\alpha^2} \langle [H_j + \alpha G_j]^+_{\mathcal{K}_j}, [G_j + \alpha H_j]^+_{(\mathcal{K}_j)^\circ} \rangle \right] \\
\leq \; & 0,
\end{aligned}$$

where the second equality follows from (9) with $G_1, ..., G_l$ and $H_1, ..., H_l$ denoting the subvectors of G and H, respectively, compatible with the Cartesian product structure of $\mathcal{K}$; the third equality and the inequality follow from properties of projection onto a convex cone and its polar. Second, we have from the positive definiteness assumption on $-\Lambda[\nabla G]^{-1}\nabla H$ that the second righthand term in (8) is less than or equal to some negative scalar times $\|R_\alpha + \frac{1}{\alpha} S_\alpha\|^2$. Thus (8) implies $R_\alpha + \frac{1}{\alpha} S_\alpha = 0$ which together with (7) and the nonsingularity of ∇G yields that $S_\alpha + \frac{1}{\alpha} R_\alpha = 0$. These two equations, together with the assumption $\alpha \neq 1$, show that $R_\alpha = S_\alpha = 0$. □

Condition A1, very roughly speaking, requires that one of ∇G and ∇H be positive definite and the other be negative definite. Notice that condition A1 is unchanged in nature when the terms ∇G

and ∇H are interchanged in its statement. Also, the assumption that $m = n$ is not essential and can be removed by replacing "$\nabla G(x)$" and "$\nabla H(x)$" in condition A1 with, respectively, '$\nabla G(x)^* \nabla G(x)$" and "$\nabla G(x)^* \nabla H(x)$." Our fourth main result below relates condition A1 to some alternative conditions:

Theorem 4.2 *Assume G and H are differentiable and $m = n$. Condition A1 is equivalent to the following condition:*

A2: For every $x \in \Re^m$, the Jacobians $\nabla G(x)$ and $\nabla H(x)$ are nonsingular and there exist real $n \times n$ matrices $U(x)$ and $V(x)$ such that both $U(x)\nabla G(x)$ and $V(x)\nabla H(x)$ are positive semidefinite, with at least one of them being positive definite, and $V(x)\nabla G(x) + \nabla H(x)^* U(x)^*$ has a block diagonal structure matching the Cartesian product structure of $\mathcal{K}$, with every block being some negative multiple of the identity matrix.

Also, condition A1 is implied by any of the following three conditions:

B1: $\mathcal{K} = \Re^n_+$ and, for every $x \in \Re^m$, the Jacobians $\nabla G(x)$ and $-\nabla H(x)$ are positive definite and there exist scalars $\psi(x) \geq 0$ and $\zeta(x) \leq 0$ such that $\zeta(x)\nabla G(x) + \psi(x)\nabla H(x)^*$ is negative definite diagonal.

B2: For every $x \in \Re^m$, the Jacobian $\nabla G(x)$ is nonsingular and $-[\nabla G(x)]^{-1}\nabla H(x)$ is positive definite.

B3: For every $x \in \Re^m$, the Jacobians $\nabla G(x)$ and $-\nabla H(x)$ are positive definite symmetric.

Proof. If A1 holds, clearly A2 holds with $U(x) = 0$ and $V(x) = -\Lambda(x)[\nabla G(x)]^{-1}$. If A2 holds, we claim that A1 holds with

$$\Lambda(x) = -V(x)\nabla G(x) - \nabla H(x)^* U(x)^*.$$

Certainly $\Lambda(x)$ has a block diagonal structure matching the Cartesian product structure of $\mathcal{K}$, with every block being some positive multiple of the identity matrix. Thus, it suffices to show that $-\Lambda(x)[\nabla G(x)]^{-1}\nabla H(x)$ is positive definite. To show this, we multiply both sides of the preceding equation on the left by $-[\nabla H(x)^*]^{-1}$ and on the right by $[\nabla G(x)]^{-1}$, yielding

$$-[\nabla H(x)^*]^{-1}\Lambda(x)[\nabla G(x)]^{-1} = [\nabla H(x)^*]^{-1}V(x) + U(x)^*[\nabla G(x)]^{-1}.$$

For any two square matrices A and B with B nonsingular, we have that AB is positive semidefinite (respectively, positive definite) if and only if $[B^*]^{-1}A$ is positive semidefinite (respectively, positive definite). Thus, the two matrices on the righthand side are both positive semidefinite with at least one of them being positive definite. This implies the matrix on the lefthand side is positive definite and, by the above fact again, the matrix $-\Lambda(x)[\nabla G(x)]^{-1}\nabla H(x)$ is positive definite.

If B1 holds, clearly A2 holds with $U(x) = \psi(x)I$ and $V(x) = \zeta(x)I$.

If B2 holds, clearly A1 holds with $\Lambda(x) = I$.

If B3 holds, we claim that B2 holds (and hence A1 holds). Since $[\nabla G(x)]^{-1}$ is positive definite symmetric, to show this, it suffices to show that the product of two positive definite symmetric matrices A and B is positive definite. Since B is positive definite symmetric, B may be expressed as $B = B^{\frac{1}{2}} B^{\frac{1}{2}}$, for some positive definite symmetric matrix $B^{\frac{1}{2}}$. Since the eigenvalues of AB are the same as those of $B^{\frac{1}{2}} A B^{\frac{1}{2}}$, it suffices to show that $B^{\frac{1}{2}} A B^{\frac{1}{2}}$ is positive definite. To show this, notice that for any $u \in \Re^n$ such that $u \neq 0$, the symmetry of $B^{\frac{1}{2}}$ implies that

$$\langle u, B^{\frac{1}{2}} A B^{\frac{1}{2}} u \rangle = \langle B^{\frac{1}{2}} u, A B^{\frac{1}{2}} u \rangle > 0,$$

where the strict inequality follows from A being positive definite and $B^{\frac{1}{2}}$ being nonsingular. □

It follows from Theorems 4.1 and 4.2 that if G and H satisfy any of the conditions A1 or A2 or B1 or B2 or B3, then GNCP is equivalent to finding a stationary point of $\tilde{m}_\alpha$, for any $\alpha \in (0, \infty)$ with $\alpha \neq 1$. This result extends [YaF93, Theorem 2.2] which considers the special case of NCP and further assumes that $\alpha > 1$.

We note that condition B3 cannot be replaced by the weaker condition that $\nabla G(x)$ and $-\nabla H(x)$ are positive definite and either $\nabla G(x)$ or $\nabla H(x)$ is symmetric. To see this, consider the following example with $m = n = 2$, where

$$\mathcal{K} = \Re^2_+, \qquad G(x_1, x_2) = \begin{bmatrix} 2x_1 - x_2 + \frac{1}{3} \\ -x_1 + x_2 \end{bmatrix}, \qquad H(x_1, x_2) = \begin{bmatrix} -x_1 - 2x_2 \\ 2x_1 - x_2 - \frac{1}{3} \end{bmatrix}.$$

It can be seen that, for every $x \in \Re^2$, $\nabla G(x)$ and $-\nabla H(x)$ are positive definite and $\nabla G(x)$ is symmetric. However, we have $\nabla \tilde{m}_2(x) = 0$ when $x = (1, 2)$, but $(1, 2)$ is not a solution to GNCP.

5 Isolatedness of Global Minima/Maxima of $\tilde{m}_\alpha$ when $\mathcal{K} = \Re^n_+$

Our fifth main result is a fairly direct extension of [MaS93, Theorem 2.3] and it states that, in the case $m = n$ and $\mathcal{K} = \Re^n_+$, the global minima (respectively, global maxima) of $\tilde{m}_\alpha$, for $\alpha \in (0, 1)$ (respectively, $\alpha \in (1, \infty)$), are isolated under a certain linear independence assumption on the columns of ∇G and ∇H. Isolatedness of global minima/maxima is of interest since many methods for unconstrained minimization/maximization attain fast local convergence only when the minima/maxima are isolated.

Theorem 5.1 *Assume $m = n$ and $\mathcal{K} = \Re^n_+$. If $\bar{x}$ is a solution to GNCP such that G and H are twice differentiable at $\bar{x}$ and satisfy the strict complementarity condition*

$$G_j(\bar{x}) = 0 \ \forall j \in J, \qquad G_j(\bar{x}) > 0 \ \forall j \in K,$$

$$H_j(\bar{x}) < 0 \ \forall j \in J, \qquad H_j(\bar{x}) = 0 \ \forall j \in K,$$

for some partition J and K of $\{1, ..., n\}$ (i.e., $J \cap K = \emptyset$ and $J \cup K = \{1, ..., n\}$), and if the gradient vectors $\nabla G_j(\bar{x})$, $j \in J$, and $\nabla H_k(\bar{x})$, $k \in K$, are linearly independent, then $\bar{x}$ is an isolated global minimum (respectively, global maximum) of $\tilde{m}_\alpha$ for any $\alpha \in (0, 1)$ (respectively, $\alpha \in (1, \infty)$).

Proof. To prove the result, it suffices to show that $\nabla^2 \tilde{m}_\alpha(\bar{x})$ is positive definite for $\alpha \in (0, 1)$ and is negative definite for $\alpha \in (1, \infty)$. We have from Theorem 3.2 that

$$\alpha \nabla \tilde{m}_\alpha(x) = (\alpha \nabla G(x) + \nabla H(x)) S_\alpha(x) + (\alpha \nabla H(x) + \nabla G(x)) R_\alpha(x)$$

for all $x \in \Re^m$. Since $R_\alpha(\bar{x}) = S_\alpha(\bar{x}) = 0$, this yields

$$\alpha \nabla^2 \tilde{m}_\alpha(\bar{x}) \ = \ (\alpha \nabla G(\bar{x}) + \nabla H(\bar{x})) \nabla S_\alpha(\bar{x})^* + (\alpha \nabla H(\bar{x}) + \nabla G(\bar{x})) \nabla R_\alpha(\bar{x})^*,$$

We also have from the strict complementarity condition (and rearranging coordinates if necessary) that

$$\begin{aligned} \nabla R_\alpha(\bar{x}) &= \nabla G(\bar{x}) - \left[\begin{array}{cc} 0_J & \nabla G_K(\bar{x}) + \alpha \nabla H_K(\bar{x}) \end{array} \right] \\ \nabla S_\alpha(\bar{x}) &= \nabla H(\bar{x}) - \left[\begin{array}{cc} \nabla H_J(\bar{x}) + \alpha \nabla G_J(\bar{x}) & 0_K \end{array} \right], \end{aligned}$$

where subscript J (respectively, K) denotes the columns whose indexes are in J (respectively, K). Thus,

$$\begin{aligned} \alpha \nabla^2 \tilde{m}_\alpha(\bar{x}) &= \alpha \nabla G(\bar{x}) \nabla H(\bar{x})^* + \nabla H(\bar{x}) \nabla H(\bar{x})^* \\ &- (\alpha \nabla G_J(\bar{x}) + \nabla H_J(\bar{x}))(\alpha \nabla G_J(\bar{x})^* + \nabla H_J(\bar{x})^*) \\ &+ \alpha \nabla H(\bar{x}) \nabla G(\bar{x})^* + \nabla G(\bar{x}) \nabla G(\bar{x})^* \\ &- (\alpha \nabla H_K(\bar{x}) + \nabla G_K(\bar{x}))(\alpha \nabla H_K(\bar{x})^* + \nabla G_K(\bar{x})^*) \\ &= \alpha \nabla G(\bar{x}) \nabla H(\bar{x})^* + \nabla H(\bar{x}) \nabla H(\bar{x})^* - \alpha^2 \nabla G_J(\bar{x}) \nabla G_J(\bar{x})^* \\ &- \alpha \nabla G_J(\bar{x}) \nabla H_J(\bar{x})^* - \alpha \nabla H_J(\bar{x}) \nabla G_J(\bar{x})^* - \nabla H_J(\bar{x}) \nabla H_J(\bar{x})^* \\ &+ \alpha \nabla H(\bar{x}) \nabla G(\bar{x})^* + \nabla G(\bar{x}) \nabla G(\bar{x})^* - \alpha^2 \nabla H_K(\bar{x}) \nabla H_K(\bar{x})^* \\ &- \alpha \nabla H_K(\bar{x}) \nabla G_K(\bar{x})^* - \alpha \nabla G_K(\bar{x}) \nabla H_K(\bar{x})^* - \nabla G_K(\bar{x}) \nabla G_K(\bar{x})^* \\ &= (1 - \alpha^2)(\nabla G_J(\bar{x}) \nabla G_J(\bar{x})^* + \nabla H_K(\bar{x}) \nabla H_K(\bar{x})^*) \\ &= (1 - \alpha^2)(\nabla G_J(\bar{x}) \ \nabla H_K(\bar{x})) \left(\begin{array}{c} \nabla G_J(\bar{x})^* \\ \nabla H_K(\bar{x})^* \end{array} \right). \end{aligned}$$

Since $m = n$ and the columns of $\nabla G_J(\bar{x})$ and $\nabla H_K(\bar{x})$ are linearly independent, it follows that $\nabla^2 \tilde{m}_\alpha(\bar{x})$ is positive definite for $\alpha \in (0, 1)$ and is negative definite for $\alpha \in (1, \infty)$. □

6 Reformulation of GNCP with $\mathcal{K} = \Re^n_+$ as NCP

In the special case of GNCP in which $m = n$, $\mathcal{K} = \Re^n_+$ and G has an inverse, it is tempting to make the substitution $z = G(x)$ so to reduce GNCP to an NCP:

$$\text{[RNCP]} \qquad \text{Find } z \in \Re^n \text{ satisfying} \quad \langle z, Q(z)\rangle = 0,\ z \geq 0,\ Q(z) \geq 0,$$

where we let

$$Q(z) = -H(G^{-1}(z))$$

for all $z \in \Re^n$. Then, we can apply the implicit Lagrangian to RNCP and obtain as a merit function for RNCP on $\Re^n$ the function $h_\alpha : \Re^n \to \Re$ given by

$$h_\alpha(z) = \langle z, Q(z)\rangle + \frac{1}{2\alpha}\left(\|[z - \alpha Q(z)]_+\|^2 - \|z\|^2 + \|[Q(z) - \alpha z]_+\|^2 - \|Q(z)\|^2\right)$$

where $\alpha \in (1, \infty)$. By [YaF93, Theorem 2.1] (or, more generally, by Theorem 4.1), if Q is differentiable and ∇Q is positive definite everywhere, the stationary points of h_α coincide with the solutions to RNCP.

Unfortunately, the preceding reduction really does not buy us anything new. First, it is easily seen using the connection between the implicit Lagrangian and $\tilde{m}_\alpha$ (see the discussion following Theorem 3.2) that

$$h_\alpha(G(x)) = -\tilde{m}_\alpha(x)$$

for all $x \in \Re^m$. Second, it is straightforward to check that if G and H are differentiable, we have

$$\nabla Q(G(x)) = -[\nabla G(x)]^{-1}\nabla H(x)$$

for all $x \in \Re^m$, so the preceding sufficient condition for the stationary points of h_α to coincide with the solutions to RNCP (i.e., ∇Q be positive definite everywhere) is really no better than the sufficient condition for the stationary points of $\tilde{m}_\alpha$ to coincide with the solutions to GNCP (see condition B2 in Theorem 4.2).

7 Conclusion and Extensions

We have shown that quadratic regularization of the gap function linearized provides a unified (and simplified) approach to generating merit functions for GNCP that are differentiable whenever G and H are. We have also studied the growth properties of these merit functions and have given conditions under which the stationary points of one such merit function coincide with the solutions to GNCP and, in the case $m = n$ and $\mathcal{K} = \Re^n_+$, conditions under which the global minima/maxima

of this merit function are isolated. Much of our results and analysis also extend to a Hilbert space setting.

There are a number of directions in which our results and analysis can be extended. For one, we can study merit functions for GNCP based on non-quadratic regularization. For another, we can study conditions under which the natural residual $\|R_1\|$ provides local/global error bound for GNCP on $\Re^m$. We can also seek to derive an analog of Theorem 4.1 for $\hat{m}_\alpha$. Lastly, we can study descent methods for minimizing (respectively, maximizing) $\tilde{m}_\alpha$ for $\alpha \in (0,1)$ (respectively, for $\alpha \in (1,\infty)$), as is done in [YaF93, Section 3] for the case of NCP. For example, it follows from the proof of Theorem 4.1 that, in the case where G and H satisfy condition A1, the vector

$$d(x) = (\Lambda(x)[\nabla G(x)]^{-1})^*(R_\alpha(x) + \frac{1}{\alpha}S_\alpha(x))$$

is a descent direction for $\tilde{m}_\alpha$ at x. Then, by using $d(x)$, we can develop a descent method for minimizing $\tilde{m}_\alpha$ and study the convergence properties of this method.

An interesting question, raised by Professor Mangasarian, concerns whether $\tilde{m}_\alpha$ may alternatively be derived using the approach he and Solodov used in [MaS93] to derive the implicit Lagrangian, namely, taking the augmented Lagrangian associated with the "natural" nonlinear programming formulation of the complementarity problem and replacing the Lagrange multipliers by the complementary quantities. As we show below, the answer is indeed a "yes." Consider the following nonlinear programming formulation of GNCP:

$$\min\{\ -\langle G(x), H(x)\rangle \mid G(x) \in \mathcal{K},\ H(x) \in \mathcal{K}^\circ\ \},$$

whose associated augmented Lagrangian, with v and u denoting the Lagrange multiplier vectors attached to the constraints $G(x) \in \mathcal{K}$ and $H(x) \in \mathcal{K}^\circ$ respectively, is

$$L_\alpha(x,u,v) = -\langle G(x), H(x)\rangle + \min_{\substack{\psi\in\mathcal{K}\\ \zeta\in\mathcal{K}^\circ}} \{\langle G(x)-\psi, v\rangle + \langle H(x)-\zeta, u\rangle + \frac{\alpha}{2}\|G(x)-\psi\|^2 + \frac{\alpha}{2}\|H(x)-\zeta\|^2\},$$

with $\alpha \in (0,\infty)$. It is readily verified using the formula for $\tilde{\pi}_{1/\alpha}$ that

$$L_\alpha(x, G(x), H(x)) = \tilde{\pi}_{1/\alpha}(G(x), H(x)) = \tilde{m}_{1/\alpha}(x) = -\tilde{m}_\alpha(x)$$

for all $x \in \Re^m$, where the last equality follows from Theorem 3.2.

References

[Auc89] G. Auchmuty, Variational Principles for Variational Inequalities, *Numerical Functional Analysis and Optimization*, 10 (1989), 863–874.

[Aus76] A. Auslender, *Optimisation: Méthodes Numériques*, Masson, Paris, 1976.

[BeT89] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, 1989.

[DoM80] I. C. Dolcetta and U. Mosco, Implicit Complementarity Problems and Quasi-Variational Inequalities, in *Variational Inequalities and Complementarity Problems: Theory and Applications*, Edited by R. W. Cottle, F. Giannessi, J.-L. Lions, Wiley, New York, 1980, 75–87.

[Fuk92] M. Fukushima, Equivalent Differentiable Optimization Problems and Descent Methods for Asymmetric Variational Inequality Problems, *Mathematical Programming*, 53 (1992), 99–110.

[GaB84] E. M. Gafni, and D. P. Bertsekas, Two–Metric Projection Methods for Constrained Optimization, *SIAM Journal on Control and Optimization*, 22 (1984), 936–964.

[Kan93] C. Kanzow, Nonlinear Complementarity as Unconstrained Optimization, Preprint 67, Institut für Angewandten Mathematik, Universität Hamburg, Hamburg, Germany (March 1993).

[LuT92] Z.–Q. Luo and P. Tseng, On a Global Error Bound for a Class of Monotone Affine Variational Inequality Problems, *Operations Research Letters*, 11 (1992), 159–165.

[LuT93] Z.–Q. Luo and P. Tseng, Error Bounds and Convergence Analysis of Feasible Descent Methods: A General Approach, *Annals of Operations Research*, 32 (1993), to appear.

[LMRS92] Z.–Q. Luo, O. L. Mangasarian, J. Ren, and M. V. Solodov, New Error Bounds for the Linear Complementarity Problem, *Mathematics of Operations Research*, to appear.

[Man90] O. L. Mangasarian, Error Bounds for Nondegenerate Monotone Linear Complementarity Problems, *Mathematical Programming*, 48 (1990), 437–445.

[MaS93] O. L. Mangasarian and M. V. Solodov, Nonlinear Complementarity as Unconstrained and Constrained Minimization, *Mathematical Programming*, 62 (1993), 277–297.

[MaP90] R. Mathias and J.–S. Pang, Error Bounds for the Linear Complementarity Problem with a P–Matrix, *Linear Algebra and its Applications*, 132 (1990), 123–136.

[Noo88] M. A. Noor, Quasi-Complementarity Problem, *Journal of Mathematical Analysis and Applications*, 130 (1988), 344–353.

[Pan81] J.-S. Pang, The Implicit Complementarity Problem, in *Nonlinear Programming 4*, Edited by O. L. Mangasarian, R. R. Meyer and S. M. Robinson, Academic Press, New York, 1981, 487–518.

[Pan87] J.-S. Pang, A Posteriori Error Bounds for the Linearly–Constrained Variational Inequality Problem, *Mathematics of Operations Research*, 12 (1987), 474–484.

[Pan94] J.-S. Pang, Complementarity Problems, in *Handbook on Global Optimization*, Edited by R. Horst and P. Pardalos, Kluwer Academic Publishers, 1994.

[TFI93] K. Taji, M. Fukushima, and T. Ibaraki, A Globally Convergent Newton Method for Solving Strongly Monotone Variational Inequalities, *Mathematical Programming*, 58 (1993), 369–383.

[YaF93] N. Yamashita and M. Fukushima, On Stationary Points of the Implicit Lagrangian for Nonlinear Complementarity Problems, Information Science Technical Report, Nara Institute of Science and Technology, Nara (August 1993); to appear in *Journal of Optimization Theory and Applications*.