A SUPERLINEARLY CONVERGENT SMOOTHING NEWTON CONTINUATION ALGORITHM FOR VARIATIONAL INEQUALITIES OVER DEFINABLE SETS*

C. B. CHUA† AND L. T. K. HIEN†

Abstract. In this paper, we use the concept of barrier-based smoothing approximations introduced by Chua and Li [*SIAM J. Optim.*, 23 (2013), pp. 745–769] to extend the smoothing Newton continuation algorithm of Hayashi, Yamashita, and Fukushima [*SIAM J. Optim.*, 15 (2005), pp. 593–615] to variational inequalities over general closed convex sets X. We prove that when the underlying barrier has a gradient map that is definable in some o-minimal structure, the iterates generated converge superlinearly to a solution of the variational inequality. We further prove that if X is proper and definable in the o-minimal structure $\mathbb{R}_{an}^{\mathbb{R}_{alg}}$, then the gradient map of its universal barrier is definable in the o-minimal expansion $\mathbb{R}_{an,exp}$. Finally, we consider the application of the algorithm to complementarity problems over epigraphs of matrix operator norm and nuclear norm and present preliminary numerical results.

Key words. variational inequalities, smoothing Newton continuation, superlinear convergence, barrier-based smoothing approximation

AMS subject classifications. 90C33, 65K05, 47J20

DOI. 10.1137/140957615

1. Introduction. Let $\mathbb{E}$ denote a finite-dimensional real vector space equipped with inner product $\langle \cdot, \cdot \rangle$, let X denote a closed convex subset of $\mathbb{E}$ with nonempty interior $\text{int}(X)$, let Ω denote a subset of $\mathbb{E}$ that contains X, and let $F : \Omega \to \mathbb{E}$ denote a continuous map that is differentiable in the interior $\text{int}(\Omega)$ of its domain Ω. The variational inequality $VI(X, F)$ is the problem of finding $x \in X$ such that

$$(1.1) \qquad \langle F(x), y - x \rangle \geq 0 \text{ for all } y \in X.$$

The variational inequality (1.1) can be solved directly by interior point methods (see, e.g., [19, 42]), or indirectly via a reformulation (see, e.g., [5, 15, 18, 21, 22]). For our approach, we use one of the most general nonsmooth reformulation equations of (1.1), the *natural map equation*

$$\begin{pmatrix} x - \Pi_X(x - y) \\ F(x) - y \end{pmatrix} = 0.$$

Here, Π_X denotes the Euclidean projector onto X; i.e.,

$$\Pi_X(z) = \arg\min_{x \in X} \frac{1}{2} \|x - z\|^2,$$

where $\|\cdot\|$ is the norm induced by the inner product $\langle \cdot, \cdot \rangle$. We denote by G^{nat} the *natural map*

$$(x, y) \in \Omega \times \mathbb{E} \mapsto (x - \Pi_X(x - y), F(x) - y).$$

Note that the domain of the natural map involves the domain Ω of F. This typically restricts the iterates in solution algorithms, unless the domain Ω is whole space $\mathbb{E}$. *To avoid this restriction, we shall assume that $\Omega = \mathbb{E}$.*

Since the Euclidean projector Π_X is generally nonsmooth, typical Newton-based methods do not apply to the natural map equation. One way to overcome this is to consider smoothing approximations. The study of smoothing approximations of Euclidean projectors is mostly limited to specific classes of convex sets, such as non-negative orthants, second-order cones, positive semidefinite cones, symmetric cones, and box-constrained sets. These include the combined smoothing and regularization method of Hayashi, Yamashita, and Fukushima [16], which was designed to solve second-order cone complementarity problems.

To the best of the authors' knowledge, there are only two known approaches to developing smoothing approximations of Euclidean projectors onto general convex sets; Qi and Sun [28] developed convolution-based smoothing approximations, while Chua and Li [7] developed barrier-based smoothing approximations. Although the former can be used to smooth any nonsmooth map, it is generally computed as a multivariate integral, and hence uncomputable in practice. In contrast, the latter applies only to the Euclidean projector onto any convex cone with nonempty interior and can be computed via proximal mappings of smooth maximal monotone maps, which is generally no more difficult than computing the projector itself.

In a subsequent work [27], Qi and Sun studied the use of convolution-based smoothing approximations in a smoothing merit function algorithm to solve the natural map equation; they proved that the algorithm converges globally under the assumption of a certain P-type property on F and the boundedness of the solution set of $VI(X, F)$. They further proved that the algorithm converges superlinearly to a solution x^* under nonsingularity of the Clarke generalized Jacobian of the smoothing approximation at $(x^*, 0)$, and the semismoothness of F at x^*, and of the projector at $x^* - F(x^*)$. Besides being applicable to all variational inequalities, the approach of Qi and Sun also avoids the need for the Jacobian consistency of the smoothing approximation (i.e., the convergence of the distance between Jacobians of the smoothing approximation and the Clarke generalized Jacobian to 0, when the smoothing parameter converges to 0), which plays a crucial role in almost all other superlinearly convergent smoothing algorithms.

In this paper, the barrier-based smoothing approach is extended to convex sets, and the barrier-based smoothing approximation is shown to be semismooth whenever the barrier used to define the smoothing approximation has a gradient map that is definable in some o-minimal structure. This result allows us to extend the local superlinear convergence of the smoothing Newton continuation algorithm of Hayashi, Yamashita, and Fukushima [16] to variational inequalities over convex sets using barriers with definable gradient maps, under assumptions of uniform nonsingularity (i.e., boundedness of the least singular value of the Jacobian away from 0) of the Newton system, and semismoothness of F. This is one of the main contributions of this paper. Furthermore, we prove that the universal barrier has a definable gradient map when the underlying set is definable in the o-minimal structure $\mathbb{R}_{\mathrm{an}}^{\mathbb{R}_{\mathrm{alg}}}$, hence showing the general applicability of the superlinear convergence result. Just as in the work of Qi and Sun, the Jacobian consistency of the smoothing approximation is not required in establishing superlinear convergence.

In [28], Qi and Sun also considered a different smoothing approximation of the Euclidean projector onto a convex set finitely generated by twice-differentiable convex functions. This smoothing approximation is in fact a barrier-based smoothing

approximation. Qi and Sun proved that when this smoothing approximation is used in a smoothing Newton continuation algorithm, the algorithm converges superlinearly to a solution x^* under the nonsingularity of the Clarke generalized Jacobian of the natural map at x^* and the linear independence constraint qualification at x^*. Our result with the barrier-based smoothing approximation will show that when the generating convex functions are definable in some o-minimal structure, the assumption of linear independence constraint qualification can be dropped.

Another main contribution of this paper is the application of the barrier-based smoothing approach to solve complementarity problems over epigraphs of matrix operator and nuclear norms. These complementarity problems can be used to approximate problems of finding low-rank (and sparse) matrices; see, e.g., [11, 13, 20, 30, 35]. As far as we can tell, there are no existing smoothing methods to solve these complementarity problems. Thus we present, for the first time, a practical smoothing Newton continuation method to solve nuclear norm complementarity problems, supported with preliminary numerical results. We show that under nondegeneracy, the uniform nonsingularity assumption on the Newton system is satisfied, and hence superlinear convergence is guaranteed as long as F is semismooth.

The paper is organized as follows. In the next section, some basic definitions and results on smoothing approximations, semismoothness, and o-minimal structures are given. In section 3, the barrier-based smoothing approximation is defined, and its semismoothness under the definability of the gradient of the barrier is established. We then present a proof of superlinear convergence of the smoothing Newton continuation algorithm of Hayashi, Yamashita, and Fukushima in section 4. In section 5, we demonstrate the definability of the gradient of the universal barrier when the set X is definable in the o-minimal structure $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$. Finally, in section 6, we consider the application of the smoothing Newton continuation algorithm of Hayashi et al. to complementarity problems over epigraphs of matrix operator and nuclear norms, and present preliminary numerical results.

1.1. Notation. We shall use $\mathbb{R}_+^m$ (respectively, $\mathbb{R}_{++}^m$) to denote the cone of nonnegative (respectively, positive) vectors in $\mathbb{R}^m$. For two vectors $x, y \in \mathbb{R}^m$, the notation $x \geq y$ (respectively, $x > y$) means $x - y \in \mathbb{R}_+^m$ (respectively, $x - y \in \mathbb{R}_{++}^m$). For any function $f : \mathbb{E} \to \mathbb{R}$ with $\lim_{x \to 0} f(x) = 0$ and $f(x) \neq 0$ near 0, we use "$g(x) = o(f(x))$ as $x \to 0$" to mean that g is a function with domain containing a neighborhood of 0 and satisfying

$$\lim_{x \to 0} \frac{g(x)}{f(x)} = 0.$$

For any function $f : \mathbb{E} \to \mathbb{R}$, we use "$h(x) = O(f(x))$ as $x \to 0$" to mean that h is a function with domain containing a neighborhood of 0, and there exists $C > 0$ such that

$$|h(x)| \leq C|f(x)|$$

for all x near 0. For any sequence $\{x_k\}$ of real numbers with $\lim_{k \to \infty} x_k = 0$ and $x_k \neq 0$ for all k, we use $y_k = o(x_k)$ to mean that $\{y_k\}$ is a sequence of real numbers satisfying

$$\lim_{k \to \infty} \frac{y_k}{x_k} = 0.$$

For any sequence $\{x_k\}$ of real numbers, we use $z_k = O(x_k)$ to mean that $\{z_k\}$ is a sequence of real numbers and there exists $C > 0$ such that

$$|z_k| \leq C|x_k|$$

for all k. For a Fréchet-differentiable function $f : \Omega \subseteq \mathbb{E} \to \mathbb{R}$, we use ∇f to denote the gradient of f. For a Fréchet-differentiable map $F : \Omega \subseteq \mathbb{E} \to \mathbb{E}'$, we use $\mathbf{J}F$ to denote the derivative of F. For a Fréchet-differentiable map

$$F : \Omega \times \Omega' \subseteq \mathbb{E} \times \mathbb{E}' \to \mathbb{E}'' : (x, y) \mapsto F(x, y),$$

we use $\mathbf{J}_x F$ and $\mathbf{J}_y F$ to denote the partial derivatives of F with respect to x and y, respectively.

2. Background. This section gives various basic definitions on smoothing approximations, semismoothness, and o-minimal structures, and establishes several basic results required in this paper.

2.1. Smoothing approximations. A *smoothing approximation* of a continuous map $G : \mathbb{E} \to \mathbb{E}'$ is a continuous map $H : \mathbb{E} \times \mathbb{R}^m_+ \to \mathbb{E}'$ such that $H(\cdot, 0) = G$, and for each $\mu \in \mathbb{R}^m_{++}$, $H(\cdot, \mu)$ is differentiable. The variable μ is called the (m-tuple of) *smoothing parameters*. When $H(\cdot, \mu)$ converges uniformly to G as $\mu \to 0$, we say that H is a *uniform smoothing approximation*. When $H(\cdot, \mu)$ is Lipschitz in the smoothing parameters (i.e., there exists $L > 0$ such that $\|H(x, \mu) - H(x, \nu)\| \leq L \|\mu - \nu\|$ for all $x \in \mathbb{E}$ and all $\mu, \nu \in \mathbb{R}^m_+$), we say that H is a *Lipschitzian smoothing approximation*.

For convenience, we shall extend the domain of a smoothing approximation H to include negative smoothing parameters, by defining

$$H(\cdot, \mu) := H(\cdot, \Pi_{\mathbb{R}^m_+}(\mu)) \text{ for any } \mu \in \mathbb{R}^m \setminus \mathbb{R}^m_+,$$

and call it an *extended smoothing approximation*. We note that the extended smoothing approximation remains continuous and retain any uniform convergence or Lipschitzian property.

An (extended) smoothing approximation $H : \mathbb{E} \times \mathbb{R}^m \to \mathbb{E}'$ of $G : \mathbb{E} \to \mathbb{E}'$ is said to *approximate superlinearly at* $\bar{x} \in \mathbb{E}$ if for any $(x, \mu) \to (\bar{x}, 0)$

$$\|H(x, \mu) - G(\bar{x}) - \mathbf{J}_x H(x, \mu)(x - \bar{x})\| = o(\|x - \bar{x}\|) + O(\|\mu\|);$$

see [28]. For any $\gamma > 0$, it is said to *approximate with order* $(1 + \gamma)$ *at* $\bar{x}$ if $o(\|x - \bar{x}\|)$ is replaced by $O(\|x - \bar{x}\|^{1+\gamma})$ in the above equation. An order-2 approximation is also called a *quadratic approximation*.

2.2. Semismoothness. A locally Lipschitz continuous map $F : \mathbb{E} \mapsto \mathbb{E}'$ is said to be *semismooth at* x if the limit

$$\lim_{\substack{V \in \partial F(x + th') \\ h' \to h, \ t \downarrow 0}} V h'$$

exists for every $h \in \mathbb{E}$, where $\partial F(x + th')$ denotes the generalized Jacobian of F at $x + th'$ as defined by Clarke [8, section 2.6]; see [29]. If a locally Lipschitz continuous map $F : \mathbb{E} \mapsto \mathbb{E}'$ is semismooth at x, then the directional derivative

$$F'(x; h) := \lim_{t \downarrow 0} \frac{F(x + th) - F(x)}{t}$$

exists and equals the limit

$$\lim_{\substack{V \in \partial F(x+th') \\ h' \to h, \ t \downarrow 0}} Vh';$$

see [29, Proposition 2.1].

It follows from [29, Theorem 2.3] and [12, Proposition 3.1.3] that a locally Lipschitz continuous map $F : \mathbb{E} \mapsto \mathbb{E}'$ is semismooth at x if and only if it is directionally differentiable at x and

$$\lim_{\substack{x+h \in D_F \\ h \to 0}} \frac{\|F(x+h) - F(x) - F'(x+h; h)\|}{\|h\|} = 0,$$

where D_F denotes the set of points at which F is Fréchet-differentiable; see also [33, Lemma 2.1].[1] This observation leads to the following notion of higher-order semismoothness: a locally Lipschitz continuous map $F : \mathbb{E} \mapsto \mathbb{E}'$ is said to be γ-*order semismooth at x* for some $\gamma \in (0, 1]$ if it is directionally differentiable at x and

$$\limsup_{\substack{x+h \in D_F \\ h \to 0}} \frac{\|F(x+h) - F(x) - F'(x+h; h)\|}{\|h\|^{1+\gamma}} < \infty;$$

see [34, Theorem 3.7]. A 1-order semismooth map is also said to be *strongly semismooth*. Finally, we note that a map is (γ-order) semismooth if its component functions are (γ-order) semismooth and that compositions of (γ-order) semismooth maps are (γ-order) semismooth. These follow from [29, Corollary 2.4], [23, Theorem 5], and [14, Theorem 19].

THEOREM 2.1. *If a locally Lipschitz continuous smoothing approximation $H :$ $\mathbb{E} \times \mathbb{R}^m \to \mathbb{E}'$ of $G : \mathbb{E} \to \mathbb{E}'$ is semismooth (respectively, γ-order semismooth) at $(\bar{x}, 0)$, then it approximates superlinearly (respectively, with order $1 + \gamma$) at $\bar{x}$.*

Proof. Since H is locally Lipschitz continuous at $(\bar{x}, 0)$, the Jacobian $\mathbf{J}_\mu H$ is locally bounded near $(\bar{x}, 0)$. Therefore for any $(x, \mu) \to (\bar{x}, 0)$,

$$\begin{aligned}
&\|H(x, \mu) - G(\bar{x}) - \mathbf{J}_x H(x, \mu)(x - \bar{x})\| \\
&\leq \|H(x, \mu) - H(\bar{x}, 0) - H'(x, \mu; x - \bar{x}, \mu)\| + \|\mathbf{J}_\mu H(x, \mu)\mu\| \\
&= o(\|(x - \bar{x}, \mu)\|) + O(\|\mu\|)
\end{aligned}$$

(respectively, $O(\|(x - \bar{x}, \mu)\|^{1+\gamma}) + O(\|\mu\|)$). Finally, $\|(x - \bar{x}, \mu)\| = O(\max\{\|x - \bar{x}\|, \|\mu\|\})$. ∎

2.3. O-minimal structures. An *o-minimal structure on the real ordered field* $\mathbb{R}$ is a sequence of Boolean algebras $\mathcal{O} = \{\mathcal{O}_n\}_{n=1}^\infty$ of subsets of $\mathbb{R}^n$ such that for each $n \geq 1$,

1. if $A \in \mathcal{O}_n$, then both $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{n+1}$;
2. $\mathcal{O}_n$ contains every algebraic subset of $\mathbb{R}^n$;
3. if $A \in \mathcal{O}_{n+1}$, then $\pi(A) \in \mathcal{O}_n$, where $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n : (x_1, \ldots, x_n, x_{n+1}) \mapsto (x_1, \ldots, x_n)$ is the projector on the first n coordinates; and
4. the sets in $\mathcal{O}_1$ are exactly the finite unions of intervals and points.

---

[1]We thank an anonymous referee for bringing this lemma to our attention.

The sets in each $\mathcal{O}_n$ are said to be *definable in $\mathcal{O}$*. A map $F : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is said to be *definable in $\mathcal{O}$* if its graph is a definable set in $\mathcal{O}$; i.e., $\{(x, y) \in A \times \mathbb{R}^m : y = F(x)\} \in \mathcal{O}_{n+m}$.

EXAMPLE 2.2 (semialgebraic sets). *The smallest o-minimal structure on $\mathbb{R}$ is the class $\mathcal{SA}$ of all semialgebraic sets. A set is semialgebraic if it can be written as a finite union of sets of the form*

$$\{x \in \mathbb{R}^n : p_1(x) = \cdots = p_k(x) = 0, q_1(x) > 0, \ldots, q_l(x) > 0\},$$

where $p_1, \ldots, p_k, q_1, \ldots, q_l \in \mathbb{R}[X]$. This example appeals to the fact that the projection of a semialgebraic set is semialgebraic by the Tarski–Seidenberg principle; see, e.g., [3].

EXAMPLE 2.3 (globally subanalytic sets). *A set is said to be globally subanalytic if its image under the map*

$$(x_1, \ldots, x_n) \in \mathbb{R}^n \mapsto \left(\frac{x_1}{\sqrt{1 + x_1^2}}, \ldots, \frac{x_n}{\sqrt{1 + x_n^2}} \right)$$

is a subanalytic set; see, e.g., [36]. The collection of all globally subanalytic sets is an o-minimal structure $\mathbb{R}_{\mathrm{an}}$ on $\mathbb{R}$. It is the smallest o-minimal structure that contains sets of the form

$$\{(x, t) \in [-1, 1]^n \times \mathbb{R} : f(x) = t\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a restricted analytic function, i.e., a function such that $f|_{[-1,1]^n}$ is analytic and vanishes identically off $[-1, 1]^n$; see [40].

EXAMPLE 2.4 ($\mathbb{R}_{\mathrm{an,exp}}$). *The smallest o-minimal structure that contains $\mathbb{R}_{\mathrm{an}}$ and the set $\{(x, e^x) : x \in \mathbb{R}\}$ is denoted by $\mathbb{R}_{\mathrm{an,exp}}$. We say that $\mathbb{R}_{\mathrm{an,exp}}$ is the o-minimal expansion of $\mathbb{R}_{\mathrm{an}}$ by the exponential function; see, e.g., [38, 39].*

EXAMPLE 2.5 ($\mathbb{R}_{\mathrm{an}}^{\mathbb{R}}$ and $\mathbb{R}_{\mathrm{an}}^{\mathbb{R}_{\mathrm{alg}}}$). *We denote by $\mathbb{R}_{\mathrm{an}}^{\mathbb{R}}$ the o-minimal expansion of $\mathbb{R}_{\mathrm{an}}$ by all power functions*

$$x \mapsto \begin{cases} x^r & \text{if } x > 0, \\ 0 & \text{if } x \leq 0; \end{cases}$$

see, e.g., [24]. When we restrict the powers to real algebraic numbers, we get a smaller o-minimal expansion of $\mathbb{R}_{\mathrm{an}}$, which we denote by $\mathbb{R}_{\mathrm{an}}^{\mathbb{R}_{\mathrm{alg}}}$.

The o-minimal structures in these examples satisfy the following strict inclusions [40, Part 2.5]:

$$\mathcal{SA} \subsetneq \mathbb{R}_{\mathrm{an}} \subsetneq \mathbb{R}_{\mathrm{an}}^{\mathbb{R}_{\mathrm{alg}}} \subsetneq \mathbb{R}_{\mathrm{an}}^{\mathbb{R}} \subsetneq \mathbb{R}_{\mathrm{an,exp}}.$$

From the definition of an o-minimal structure, especially closure under projection, one can establish many stability results for definable sets and functions. We list some of these results here as they will be used in this paper. We refer the readers to [9, Theorem 1.13], [4], and [37] for their proofs.

PROPOSITION 2.6 (stability results). *Let $\mathcal{O}$ be an o-minimal structure on $\mathbb{R}$.*
1. *If $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^{n+m}$ are definable in $\mathcal{O}$, then the sets $\{x : \forall y \in A, (x, y) \in B\}$ and $\{x : \exists y \in A, (x, y) \in B\}$ are definable in $\mathcal{O}$.*
2. *The closure, interior, and product of definable sets in $\mathcal{O}$ are definable in $\mathcal{O}$.*

3. *If a map $G : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is definable in $\mathcal{O}$, then its derivative $\boldsymbol{J}G$ and its partial derivatives $\boldsymbol{J}_{x_i}G$ (if they exist) are definable in $\mathcal{O}$. If, in addition, G is injective, then its inverse map is definable in $\mathcal{O}$.*

4. *If the maps $G : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $F : V \supseteq G(U) \to \mathbb{R}^p$ are definable in $\mathcal{O}$, then the composition $F \circ G : U \to \mathbb{R}^p$ is definable in $\mathcal{O}$.*

5. *A vector-valued map is definable in $\mathcal{O}$ if and only if each of its component functions is definable in $\mathcal{O}$.*

3. Barrier-based smoothing approximations of closed convex sets. For each differentiable barrier $f : \mathrm{int}(X) \to \mathbb{R}$ on a closed convex set $X \in \mathbb{E}$ with nonempty interior (i.e., $f(x_k) \to \infty$ for any convergent sequence $\{x_k\}$ in $\mathrm{int}(X)$ with limit in the boundary of X), we define the map $p : \mathbb{E} \times \mathbb{R} \to \mathbb{E}$ via

(3.1)
$$\begin{cases} p(z, \mu) + \mu^2 \nabla f(p(z, \mu)) = z & \text{when } \mu > 0, \\ p(\cdot, \mu) = \Pi_X & \text{when } \mu \le 0. \end{cases}$$

We note that for each $\mu > 0$, the map $p(\cdot, \mu)$ is the proximal mapping of $x \mapsto \mu^2 \nabla f(x)$, which is maximal monotone by Löhne's characterization (cf. [7, Proposition 3.1]), and hence it is a bijection between $\mathrm{int}(X)$ and $\mathbb{E}$ by Minty's criterion. Thus p is well-defined.

THEOREM 3.1 (barrier-based smoothing approximation). *If f is a twice continuously differentiable barrier on $X \subset \mathbb{E}$, then the map $p : \mathbb{E} \times \mathbb{R} \to \mathbb{E}$ defined via (3.1) is an extended smoothing approximation of the Euclidean projector Π_X.*

Proof. By the implicit function theorem, the continuous differentiability of ∇f implies the differentiability of p over $\mathbb{E} \times \mathbb{R}_{++}$.

It remains to show that for each fixed $z \in \mathbb{E}$, $\{p(w_k, \mu_k)\}$ converges to $\Pi_X(z)$ for any sequence $\{(w_k, \mu_k)\}$ in $\mathbb{E} \times \mathbb{R}_{++}$ converging to $(z, 0)$. We first show that the sequence $\{p(w_k, \mu_k)\}$ is bounded. Pick an arbitrary but fixed $e \in \mathrm{int}(X)$. The sequence $\{e_k := e + \mu_k^2 \nabla f(e)\}$ is bounded. Moreover, $p(e_k, \mu_k) = e$ by definition. For each k, the nonexpansiveness of the proximal mapping $p(\cdot, \mu_k)$ implies

$$\|p(w_k, \mu_k)\| \le \|p(w_k, \mu_k) - p(e_k, \mu_k)\| + \|p(e_k, \mu_k)\| \le \|w_k - e_k\| + \|e\| ;$$

thus $\{p(w_k, \mu_k)\}$ is bounded. Finally, we note that since $\mu_k > 0$, it follows that $p(w_k, \mu_k)$ is the unique minimizer to the barrier problem $\min\{\frac{1}{2} \|x - w_k\|^2 + \mu_k^2 f(x)\}$, and every limit point of these minimizers must be the unique minimizer $\Pi_X(z)$ of the convex optimization problem $\min\{\frac{1}{2} \|x - z\|^2 : x \in X\}$.[2] ∎

DEFINITION 3.2 (barrier-based smoothing approximation). *Given a twice continuously differentiable barrier f on $S \subset \mathbb{E}$, the* (extended barrier-based) *smoothing approximation defined by f is the map $p : \mathbb{E} \times \mathbb{R} \to \mathbb{E}$ that satisfies* (3.1). *It is a smoothing approximation of the Euclidean projector Π_X.*

DEFINITION 3.3 (ϑ-barrier). *The* barrier parameter *of a twice continuously differentiable barrier f on $X \subset \mathbb{E}$ is*

$$\inf \left\{ \vartheta \ge 0 : \inf_{x \in \mathrm{int}(X), \ h \in \mathbb{E}} \vartheta \left\langle h, \nabla^2 f(x) h \right\rangle - \left\langle \nabla f(x), h \right\rangle^2 \ge 0 \right\}.$$

A ϑ-barrier is a twice continuously differentiable barrier with a finite barrier parameter ϑ.

THEOREM 3.4 (barrier-based uniform smoothing approximation). *The smoothing approximation p defined by a ϑ-barrier $f : \mathrm{int}(X) \to \mathbb{R}$ is $\sqrt{\vartheta}$-Lipschitzian, i.e.,*

---

[2]See, e.g., [2, Proposition 4.1.1].

Lipschitz continuous with modulus $\sqrt{\vartheta}$ in the smoothing parameter.

Consequently, in this case, p is a uniform smoothing approximation of the Euclidean projector Π_X.

Proof. By appealing to the continuity of the smoothing approximation, it suffices to only consider positive smoothing parameters. Since

$$
\begin{aligned}
&\vartheta(\mu - \nu)^2 - \|p(z, \mu) - p(z, \nu)\|^2 \\
&= \vartheta(\mu - \nu)^2 - \langle p(z, \mu) - p(z, \nu), p(z, \mu) - p(z, \nu) \rangle \\
&= \vartheta(\mu - \nu)^2 + \langle p(z, \mu) - p(z, \nu), \mu^2 \nabla f(p(z, \mu)) - \nu^2 \nabla f(p(z, \nu)) \rangle \\
&= \langle (p(z, \mu) - p(z, \nu), \mu - \nu), (\mu^2 \nabla f(p(z, \mu)) - \nu^2 \nabla f(p(z, \nu)), \vartheta\mu - \vartheta\nu) \rangle,
\end{aligned}
$$

it suffices to show that the map $(x, \mu) \in \operatorname{int}(X) \times \mathbb{R}_{++} \mapsto (\mu^2 \nabla f(x), \vartheta\mu)$ is monotone. To this end, we check that its Jacobian

$$
(h, \tau) \in \mathbb{E} \times \mathbb{R} \mapsto (\mu^2 \nabla^2 f(x)h + 2\mu \nabla f(x)\tau, \vartheta\tau)
$$

is a monotone linear map at each $(x, \mu) \in \operatorname{int}(X) \times \mathbb{R}_{++}$. Indeed, its symmetric part

$$
(h, \tau) \in \mathbb{E} \times \mathbb{R} \mapsto (\mu^2 \nabla^2 f(x)h + \mu \nabla f(x)\tau, \mu \langle \nabla f(x), h \rangle + \vartheta\tau)
$$

is positive semidefinite if and only if the Schur complement

$$
h \in \mathbb{E} \mapsto \frac{\mu^2}{\vartheta}(\vartheta \nabla^2 f(x)h - \langle \nabla f(x), h \rangle \nabla f(x))
$$

is positive semidefinite. ☐

Henceforth, we assume that the barrier f is a ϑ-barrier.

With the uniform smoothing approximation p defined by f, we can now define a smoothing approximation

$$
(3.2) \qquad H^{\mathrm{nat}} : (x, y, \mu, \varepsilon) \in \mathbb{E}^2 \times \mathbb{R}^2 \mapsto \begin{pmatrix} x - p(x - y, \mu) \\ F(x) + \Pi_{\mathbb{R}_+}(\varepsilon)x - y \end{pmatrix}
$$

of the natural map, which incorporates a regularization of the map F.

We note that the local Lipschitz continuity of F carries over to the smoothing approximation H^{nat} under the $\sqrt{\vartheta}$-Lipschitz continuity of p in the smoothing parameter.

PROPOSITION 3.5. *If f is a ϑ-barrier, and F is locally Lipschitz continuous, then the smoothing approximation H^{nat} is locally Lipschitz continuous.*

3.1. Definability and superlinear approximations. As shown by Bolte, Daniilidis, and Lewis [4], a locally Lipschitz definable function (more generally a locally Lipschitz tame function) is semismooth. Moreover, a locally Lipschitz function that is definable in a polynomially bounded o-minimal structure is γ-order semismooth for some $\gamma > 0$; see Remarks 3 and 4 of [4]. Examples of polynomially bounded o-minimal structures are substructures of $\mathbb{R}_{\mathrm{an}}^{\mathbb{R}}$; see, e.g., [32, p. 184]. Thus if the smoothing approximation p is definable in some o-minimal structure (respectively, polynomially bounded o-minimal structure), then it is semismooth (respectively, γ-order semismooth), and hence we may apply Theorem 2.1 to deduce that it approximates superlinearly (respectively, with order $1 + \gamma$). The following proposition gives a necessary and sufficient condition for p to be definable.

PROPOSITION 3.6. *The smoothing approximation p defined by f is definable in an o-minimal structure $\mathcal{O}$ if and only if ∇f is definable in $\mathcal{O}$.*

Consequently, p is semismooth when f is definable in $\mathcal{O}$. Moreover, p is γ-order semismooth for some $\gamma > 0$ when f is definable in a polynomially bounded o-minimal structure $\mathcal{O}$.

Proof. The graph of p is

$$\left\{ (z, \mu, x) : x \in \text{int}(K), \ \mu > 0, \ x + \mu^2 \nabla f(x) = z \right\}$$
$$\bigcup \left\{ (z, \mu, x) : \mu \leq 0, \ x \in K, \ \langle x, z - x \rangle = 0 \text{ and } \forall w \in K, \langle w, z - x \rangle \leq 0 \right\},$$

which is definable when ∇f and K are definable. Since K is the closure of the domain of ∇f, it is definable whenever ∇f is.

Conversely, the graph of ∇f is $\{(x, y) : x = p(x + y, 1)\}$, which is definable when p is definable.

The final statements then follow from Theorem 1 and Remarks 3 and 4 of [4]. □

As consequences of Propositions 3.5 and 3.6 and Theorem 2.1, we deduce sufficient conditions for the smoothing approximation of the natural map to approximate superlinearly (respectively, with order $1 + \gamma$).

PROPOSITION 3.7. *For any $x, y \in \mathbb{E}$, if f is a ϑ-barrier with a gradient map that is definable in an o-minimal structure (respectively, polynomially bounded o-minimal structure), and F is semismooth (respectively, γ'-order semismooth) at x, then the smoothing approximation H^{nat} defined in (3.2) is semismooth (respectively, γ-order semismooth for some $\gamma \in (0, \gamma']$) at $(x, y, 0)$. Consequently, it approximates superlinearly (respectively, with order $1 + \gamma$) at $(x, y, 0)$.*

We now give a few examples of twice continuously differentiable barriers with finite barrier parameters and with gradients that are definable in some o-minimal structures.

EXAMPLE 3.8 (polyhedral sets). *A barrier of the polyhedral set $\{x : a_i^T x - b_i \leq 0 \ i = 1, \ldots, m\}$, where $a_1, \ldots, a_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$, is $x \mapsto -\sum_{i=1}^m \log(b_i - a_i^T x)$. Its barrier parameter is m, and its gradient is definable in the o-minimal structure $\mathcal{SA}$ of semialgebraic sets.*

EXAMPLE 3.9 (symmetric cones). *A symmetric cone K is the cone of squares of some Euclidean Jordan algebra $\mathfrak{J}$ and is thus definable in the o-minimal structure $\mathcal{SA}$. Its standard log-determinant barrier is $x \in \text{int}(K) \mapsto -\log \det(x)$, where the determinant $\det(\cdot)$ is a polynomial in x. The gradient of the log-determinant barrier is thus definable in the o-minimal structure $\mathcal{SA}$. Its barrier parameter is the rank of the symmetric cone.*

EXAMPLE 3.10 (homogeneous cones). *A homogeneous cone K is the cone associated with some T-algebra $\mathfrak{A}$; see, e.g., [6, Theorem 1]. It is thus definable in the o-minimal structure $\mathcal{SA}$. The only known optimal self-concordant barrier of K [6, section 3.1] has the form $x \mapsto -\sum_{i=1}^r \log \rho_i(u_x)^2$, where $x \mapsto \rho_i(u_x)^2$ are rational functions; see [41, section III.3]. Thus the gradient of this barrier is definable in the o-minimal structure $\mathcal{SA}$. Its barrier parameter is the rank of the homogeneous cone.*

EXAMPLE 3.11 (hyperbolicity cones). *By Proposition 18 and Theorem 20 of [31], the hyperbolicity cone defined by a hyperbolic polynomial q is definable in the o-minimal structure $\mathcal{SA}$. A barrier of this hyperbolicity cone is $x \mapsto -\log q(x)$. The barrier parameter of this barrier is the degree of the polynomial q. Since q is a polynomial, the gradient of this barrier is definable in the o-minimal structure $\mathcal{SA}$.*

EXAMPLE 3.12 (power cones). *A (high-dimensional) power cone is*

$$\left\{ (x_1, \ldots, x_n, z_1, \ldots, z_m) \in \mathbb{R}_+^n \times \mathbb{R}^m : \prod_{i=1}^n x_i^{\alpha_i} \geq \|z\| \right\},$$

where the exponents $\alpha_1, \ldots, \alpha_n \in (0, 1]$ sum to 1. A barrier of this cone is $-\log(\prod_{i=1}^{n} x_i^{2\alpha_i} - \|z\|^2)$, whose gradient is definable in the o-minimal structure $\mathbb{R}_{an}^{\mathbb{R}}$. It has barrier parameter 2.

EXAMPLE 3.13 (finitely generated convex sets). *Consider the convex set $X = \{x : g_i(x) \leq 0 \text{ for } i = 1, \ldots, N\}$ generated by a finite number of twice continuously differentiable convex functions $g_1, \ldots, g_N : \mathbb{E} \to \mathbb{R}$. When we assume that for $i = 1, \ldots, N$, $g_i(x) < 0$ for all $x \in \text{int}(X)$, and that $\text{int}(X) \neq \emptyset$, the function $f : x \in \text{int}(X) \mapsto -\sum_{i=1}^{N} \log(-g_i(x))$ is a barrier of X. This barrier has gradient and Hessian*

$$\nabla f : x \mapsto \sum_{i=1}^{N} \frac{-\nabla g_i(x)}{-g_i(x)}$$

and

$$\nabla^2 f(x) : h \mapsto \sum_{i=1}^{N} \frac{\nabla^2 g_i(x)h}{-g_i(x)} + \sum_{i=1}^{N} \frac{\langle \nabla g_i(x), h \rangle \nabla g_i(x)}{g_i(x)^2},$$

respectively. The gradient is definable in an o-minimal structure whenever the functions $g_1, \ldots, g_N$ are definable in the same o-minimal structure. Since $\nabla^2 g_i(x)$ is positive semidefinite for each i, we have that, for any $x \in \text{int}(X)$ and any $h \in \mathbb{E}$,

$$\langle \nabla f(x), h \rangle^2 = \left(\sum_{i=1}^{N} \frac{\langle -\nabla g_i(x), h \rangle}{-g_i(x)} \right)^2$$
$$\leq N \sum_{i=1}^{N} \left(\frac{\langle -\nabla g_i(x), h \rangle}{-g_i(x)} \right)^2$$
$$\leq N \langle h, \nabla^2 f(x)h \rangle ;$$

i.e., f has a finite barrier parameter $\vartheta \leq N$. The uniform smoothing approximation defined by f coincides with the one defined in [28, section 4].

4. Smoothing Newton continuation algorithm. Given $G : \mathbb{E} \to \mathbb{E}$ and a smoothing approximation $H : \mathbb{E} \times \mathbb{R}^m \to \mathbb{E}$ of G, consider an algorithm that generates an infinite sequence $\{(w_k, \mu_k)\}$ in $\mathbb{E} \times \mathbb{R}_+^m$ such that $\mu_k \to 0$ and $\|H(w_k, \mu_k)\| \to 0$. The continuity of H then implies that any accumulation point of $\{w_k\}$ is a zero of G. We shall now consider the local superlinear convergence of such algorithms.

LEMMA 4.1. *Suppose $\{w_k\}$ is a convergent sequence in $\mathbb{E}$ converging to a zero w^* of G. If, in addition, H is Lipschitz continuous near $(w^*, 0)$ and approximates superlinearly (respectively, with order $1 + \gamma$) at $(w^*, 0)$, and $\{\mu_k\}$ and $\{\nu_k\}$ are sequences in $\mathbb{R}_{++}^m$ such that*

1. *$\{\boldsymbol{J}_w H(w_k, \mu_k)\}$ is uniformly nonsingular (i.e., each term is nonsingular with the sequence of inverses having uniformly bounded norms), and*

2. *$\|\mu_k\|, \|\nu_k\| = o(\|H(w_k, 0)\|)$ (respectively, $\|\mu_k\|, \|\nu_k\| = O(\|H(w_k, 0)\|^{1+\gamma})$), then the sequence of solutions $\{d_k\}$ to $H(w_k, \nu_k) + \boldsymbol{J}_w H(w_k, \mu_k)d_k = 0$ satisfies*

$$\|w_k + d_k - w^*\| = o(\|w_k - w^*\|)$$

(respectively, $\|w_k + d_k - w^*\| = O(\|w_k - w^*\|^{1+\gamma})$).

Moreover, for any sequences $\{\tilde{\mu}_k\}, \{\tilde{\nu}_k\}, \{\hat{\nu}_k\}$ in $\mathbb{R}^m$ such that $\|\tilde{\mu}_k\|, \|\tilde{\nu}_k\|, \|\hat{\nu}_k\| = o(\|H(w_k, 0)\|)$,

(i) $\left\|H(w_k + d_k, \tilde{\mu}_k)\right\|^2 - \left\|H(w_k, \tilde{\mu}_k)\right\|^2 + \left\|H(w_k, \tilde{\nu}_k)\right\|^2 = o(\left\|H(w_k, \tilde{\nu}_k)\right\|^2)$, and

(ii) $\left\|H(w_k + d_k, \hat{\nu}_k)\right\| = o(\left\|H(w_k, \hat{\nu}_{k-1})\right\| + \left\|\hat{\nu}_k - \hat{\nu}_{k-1}\right\|)$.

Proof. We first note that the Lipschitz continuity of H near $(w^*, 0)$ implies that

(4.1)
$$o(\|H(w_k, 0)\|) = o(\|H(w_k, 0) - H(w^*, 0)\|) = o(\|w_k - w^*\|) \quad \text{and}$$
$$O(\|H(w_k, 0)\|^{1+\gamma}) = O(\|H(w_k, 0) - H(w^*, 0)\|^{1+\gamma}) = O(\|w_k - w^*\|^{1+\gamma}).$$

Under the hypothesis of the boundedness of $\{\mathbf{J}_w H(w_k, \mu_k)^{-1}\}$,

$$\begin{aligned}
\left\|w_k + d_k - w^*\right\| &= \left\|w_k - \mathbf{J}_w H(w_k, \mu_k)^{-1} H(w_k, \nu_k) - w^*\right\| \\
&= O(\left\|\mathbf{J}_w H(w_k, \mu_k)(w_k - w^*) - H(w_k, \nu_k)\right\|) \\
&= O(\left\|\mathbf{J}_w H(w_k, \mu_k)(w_k - w^*) - H(w_k, \mu_k) + H(w^*, 0)\right\|) \\
&\quad + O(\left\|H(w_k, \mu_k) - H(w_k, \nu_k)\right\|).
\end{aligned}$$

The local Lipschitz continuity of H near $(w^*, 0)$ bounds the second term by $O(\|\mu_k - \nu_k\|)$. The first term is $o(\|w_k - w^*\|) + O(\|\mu_k\|)$ (respectively, $O(\|w_k - w^*\|^{1+\gamma}) + O(\|\mu_k\|)$) under the hypothesis that H is a superlinear (respectively, order-$(1+\gamma)$) approximation at $(w^*, 0)$. Subsequently, we deduce from the hypothesis $\|\mu_k\|, \|\nu_k\| = o(\|H(w_k, 0)\|)$ (respectively, $\|\mu_k\|, \|\nu_k\| = O(\|H(w_k, 0)\|^{1+\gamma})$) and (4.1) that

(4.2)
$$\|w_k + d_k - w^*\| = o(\|w_k - w^*\|)$$

(respectively, $\|w_k + d_k - w^*\| = O(\|w_k - w^*\|^{1+\gamma})$).

(i) We now show $\left\|H(w_k + d_k, \tilde{\mu}_k)\right\|^2 - \left\|H(w_k, \tilde{\mu}_k)\right\|^2 + \left\|H(w_k, \tilde{\nu}_k)\right\|^2 = o(\left\|H(w_k, \tilde{\nu}_k)\right\|^2)$. The last two terms on the left can be bounded by

$$\begin{aligned}
&\left\|H(w_k, \tilde{\nu}_k)\right\|^2 - \left\|H(w_k, \tilde{\mu}_k)\right\|^2 \\
&= (\left\|H(w_k, \tilde{\nu}_k)\right\| - \left\|H(w_k, \tilde{\mu}_k)\right\|)(\left\|H(w_k, \tilde{\nu}_k)\right\| + \left\|H(w_k, \tilde{\mu}_k)\right\|) \\
&\leq \left\|H(w_k, \tilde{\nu}_k) - H(w_k, \tilde{\mu}_k)\right\| (2\left\|H(w_k, \tilde{\nu}_k)\right\| + \left\|H(w_k, \tilde{\mu}_k) - H(w_k, \tilde{\nu}_k)\right\|) \\
&= O(\left\|\tilde{\mu}_k - \tilde{\nu}_k\right\| (\left\|H(w_k, \tilde{\nu}_k)\right\| + \left\|\tilde{\mu}_k - \tilde{\nu}_k\right\|)) \\
&= o(\left\|w_k - w^*\right\| (\left\|H(w_k, \tilde{\nu}_k)\right\| + \left\|w_k - w^*\right\|))
\end{aligned}$$

via the Lipschitz continuity of H near $(w^*, 0)$, (4.1), and the hypothesis $\|\tilde{\mu}_k\|, \|\tilde{\nu}_k\| = o(\|H(w_k, 0)\|)$. To bound the first term on the left, we note from (4.2) that $w_k + d_k \to w^*$, whence we may apply the Lipschitz continuity of H near $(w^*, 0)$ to deduce

(4.3)
$$\begin{aligned}
\left\|H(w_k + d_k, \tilde{\mu}_k)\right\| &= \left\|H(w_k + d_k, \tilde{\mu}_k) - H(w^*, 0)\right\| \\
&= O(\left\|w_k + d_k - w^*\right\| + \left\|\tilde{\mu}_k\right\|) \\
&= o(\left\|w_k - w^*\right\|),
\end{aligned}$$

where we have used the hypothesis $\|\tilde{\mu}_k\| = o(\|H(w_k, 0)\|)$ and (4.1) and (4.2) in the last equality. It remains to show that $\|w_k - w^*\| = O(\|H(w_k, \tilde{\nu}_k)\|)$. This follows from

(4.4)
$$\begin{aligned}
\left\|w_k - w^*\right\| &= \left\|w_k + d_k - w^* - \mathbf{J}_w H(w_k, \mu_k)^{-1} H(w_k, \nu_k)\right\| \\
&= o(\left\|w_k - w^*\right\|) + O(\left\|H(w_k, \nu_k)\right\|) \\
&= o(\left\|w_k - w^*\right\|) + O(\left\|H(w_k, \tilde{\nu}_k)\right\| + \left\|\nu_k - \tilde{\nu}_k\right\|) \\
&= o(\left\|w_k - w^*\right\|) + O(\left\|H(w_k, \tilde{\nu}_k)\right\|),
\end{aligned}$$

ALGORITHM 4.2 (Algorithm 2 of [16]).

Inputs: Initial data $w_0 = (x_0, y_0) \in \mathbb{E} \times \mathbb{E}$ and parameters $\beta > 0$, $\alpha, \eta \in (0, 1)$, $\bar{\eta} \in (0, \eta]$, $\sigma \in (0, 1/2)$, and $\kappa > 0$.

Set $k = 0$ and $\mu_0 = \varepsilon_0 = \left\| G^{\mathrm{nat}}(w_0) \right\|$ and repeat the following steps until $\left\| G^{\mathrm{nat}}(w_k) \right\| = 0$.

Step 1. Set $j = 0$ and $v_{k0} = w_k$.

    Step 1a. Find $d_{kj} \in \mathbb{E}^2$ such that

$$H^{\mathrm{nat}}(v_{kj}, \mu_k, \varepsilon_k) + \mathbf{J}_w H^{\mathrm{nat}}(v_{kj}, \mu_k, \varepsilon_k) d_{kj} = 0.$$

    Step 1b. If $\left\| H^{\mathrm{nat}}(v_{kj} + d_{kj}, \mu_k, \varepsilon_k) \right\| \leq \beta \eta^k$, set $w_{k+1} = v_{kj} + d_{kj}$ and proceed to Step 2.

    Step 1c. Otherwise, find the largest $\lambda_{kj} \in \{1, \alpha, \alpha^2, \dots\}$ such that

$$\left\| H^{\mathrm{nat}}(v_{kj} + \lambda_{kj} d_{kj}, \mu_k, \varepsilon_k) \right\|^2 - \left\| H^{\mathrm{nat}}(v_{kj}, \mu_k, \varepsilon_k) \right\|^2$$
$$\leq -2\sigma \lambda_{kj} \left\| H^{\mathrm{nat}}(v_{kj}, \mu_k, \varepsilon_k) \right\|^2$$

    and set $v_{k,j+1} = v_{kj} + \lambda_{kj} d_{kj}$.

    Step 1d. If $\left\| H^{\mathrm{nat}}(v_{k,j+1}, \mu_k, \varepsilon_k) \right\| \leq \beta \eta^k$, set $w_{k+1} = v_{k,j+1}$ and proceed to Step 2.

    Otherwise, update $j = j + 1$ and return to Step 1a.

Step 2. Set $\mu_{k+1} = \min\{\kappa \left\| G^{\mathrm{nat}}(w_{k+1}) \right\|^2, \mu_0 \bar{\eta}^{k+1}\}$

    and $\varepsilon_{k+1} = \min\{\kappa \left\| G^{\mathrm{nat}}(w_{k+1}) \right\|^2, \varepsilon_0 \bar{\eta}^{k+1}\}$ and update $k = k + 1$.

where we have used (4.2), the boundedness of $\{\mathbf{J}_w H(w_k, \mu_k)^{-1}\}$, the Lipschitz continuity of H near $(w^*, 0)$, the hypothesis $\|\nu_k\|, \|\tilde{\nu}_k\| = o(\|H(w_k, 0)\|)$, and (4.1).

(ii) Since the equations in (4.3) and (4.4) hold when, respectively, $\tilde{\mu}_k$ and $\tilde{\nu}_k$ are replaced by $\hat{\nu}_k$, we can further deduce that

$$\|H(w_k + d_k, \hat{\nu}_k)\| = o(\|w_k - w^*\|)$$
$$= o(\|H(w_k, \hat{\nu}_k)\|)$$
$$= o(\|H(w_k, \hat{\nu}_k) - H(w_k, \hat{\nu}_{k-1})\| + \|H(w_k, \hat{\nu}_{k-1})\|)$$
$$= o(\|\hat{\nu}_k - \hat{\nu}_{k-1}\| + \|H(w_k, \hat{\nu}_{k-1})\|),$$

where we have used the Lipschitz continuity of H near $(w^*, 0)$. $\quad\square$

We now consider a specific algorithm for solving the natural map equation and prove its local superlinear convergence. The algorithm, Algorithm 4.2, was analyzed in [16] for its global and local superlinear convergence. This algorithm, although only stated and analyzed for second-order cone complementarity problems in [16], has been shown to be globally convergent when solving general natural map equations.

Regularization is incorporated to ensure the boundedness of the level sets of $H^{\mathrm{nat}}(\cdot, \mu, \varepsilon)$ for all $\mu, \varepsilon > 0$. We note the slight difference in the update of μ: the additional condition in the update formula in [16] was imposed to exploit the Jacobian consistency of the smoothing approximation and has no effect on global convergence. We drop this additional condition here because we no longer assume the Jacobian consistency of the smoothing approximation.

THEOREM 4.3. *If*

1. $F: \mathbb{E} \to \mathbb{E}$ *is locally Lipschitz continuous and monotone, and*

2. *the solution set of $VI(X,F)$ is nonempty and bounded,*

then Algorithm 4.2 generates a bounded sequence $\{w_k = (x_k, y_k)\}$ with an accumulation point $w^* = (x^*, y^*)$ that is a zero of G^{nat}.

If, in addition,

1. H^{nat} is based on the smoothing approximation defined by a ϑ-barrier with a gradient map that is definable in an o-minimal structure (respectively, polynomially bounded o-minimal structure),
2. F is semismooth (respectively, γ'-order semismooth) at x^*, and
3. for any subsequence $\{w_{k_l}\}$ converging to w^*, $\{\boldsymbol{J}_w H^{\mathrm{nat}}(w_{k_l}, \mu_{k_l}, \varepsilon_{k_l})\}$ is uniformly nonsingular,

then the full Newton step is eventually always taken (i.e., $w_{k+1} = w_k + d_{k0}$ for all k sufficiently large), and the x-component of the sequence $\{w_k\}$ converges superlinearly (respectively, with order $1 + \gamma$ for some $\gamma \in (0, \gamma']$) to the solution x^* of $VI(X,F)$.

Proof. The global convergence of the algorithm is proved in Theorem 4.3 of [16].

To prove superlinear convergence, construct a subsequence $\{w_{k_l}\}$ by taking $k_0 = 0$ and recursively taking $k_{l+1} \geq k_l$ to be the least index satisfying $\|w_{k_{l+1}} - w^*\| \leq \frac{1}{2}\|w_{k_l} - w^*\|$. This subsequence is well-defined since w^* is an accumulation point of $\{w_k\}$. Since $\|(\mu_{k_l}, \varepsilon_{k_l})\| \leq \kappa\sqrt{2}\|H^{\mathrm{nat}}(w_{k_l}, 0)\|^2 = O(\|H^{\mathrm{nat}}(w_{k_l}, 0)\|^2) = o(\|H^{\mathrm{nat}}(w_{k_l}, 0)\|)$, we can apply Lemma 4.1, together with Proposition 3.7, on this subsequence by taking $\{\mu_k\}, \{\nu_k\}, \{\tilde{\mu}_k\}, \{\tilde{\nu}_k\}, \{\hat{\nu}_k\}$ in the lemma to be $\{(\mu_{k_l}, \varepsilon_{k_l})\}$ to get

1. $\|w_{k_l} + d_{k_l 0} - w^*\| = o(\|w_{k_l} - w^*\|)$
   (respectively, $\|w_{k_l} + d_{k_l 0} - w^*\| = O(\|w_{k_l} - w^*\|^{1+\gamma})$),
2. $\|H^{\mathrm{nat}}(w_{k_l} + d_{k_l 0}, \mu_{k_l}, \varepsilon_{k_l})\|^2 = o(\|H^{\mathrm{nat}}(w_{k_l}, \mu_{k_l}, \varepsilon_{k_l})\|^2)$, and
3. $H^{\mathrm{nat}}(w_{k_l} + d_{k_l 0}, \mu_{k_l}, \varepsilon_{k_l}) = o(\|H^{\mathrm{nat}}(w_{k_l}, \mu_{k_l-1}, \varepsilon_{k_l-1})\| + \|(\mu_{k_l} - \mu_{k_l-1}, \varepsilon_{k_l} - \varepsilon_{k_l-1})\|)$,

where $d_{k_l 0}$ is the search direction determined in Step 1a for $j = 0$. The third conclusion, together with $\|H^{\mathrm{nat}}(w_{k_l}, \mu_{k_l-1}, \varepsilon_{k_l-1})\| \leq \beta\eta^{k_l-1}$ and

$$(|\mu_{k_l} - \mu_{k_l-1}|, |\varepsilon_{k_l} - \varepsilon_{k_l-1}|) \leq (\mu_{k_l-1}, \varepsilon_{k_l-1}) \leq \bar{\eta}^{k_l-1}(\mu_0, \varepsilon_0) \leq \eta^{k_l-1}(\mu_0, \varepsilon_0),$$

implies that $\|H^{\mathrm{nat}}(w_{k_l} + d_{k_l 0}, \mu_{k_l}, \varepsilon_{k_l})\| \leq \beta\eta^{k_l}$ eventually always holds. Thus the full Newton step is eventually always taken at Step 1b. It then follows from the first conclusion that for all sufficiently large l,

$$\|w_{k_{l+1}} - w^*\| = \|w_{k_l} + d_{k_l 0} - w^*\| \leq \frac{1}{2}\|w_{k_l} - w^*\|,$$

whence $k_{l+1} = k_l + 1$. This means $w_{k+1} = w_k + d_{k0}$ for all k sufficiently large, and w_k converges superlinearly (respectively, with order $1 + \gamma$) to w^*. $\square$

We end this section with a sufficient condition for $\{\boldsymbol{J}_w H^{\mathrm{nat}}(w_k, \mu_k, \varepsilon_k)\}$ to be uniformly nonsingular. We will show in section 6 that this sufficient condition is implied by nondegeneracy when X is the epigraph of the matrix operator norm, or the nuclear norm.

THEOREM 4.4. *If $F : \mathbb{E} \to \mathbb{E}$ is monotone and $\{(x_k, y_k, \mu_k, \varepsilon_k)\}$ is a convergent sequence generated by Algorithm 4.2 converging to $(x^*, y^*, 0, 0)$, then in order for the sequence $\{\boldsymbol{J}_{(x,y)} H^{\mathrm{nat}}(x_k, y_k, \mu_k, \varepsilon_k)\}$ to be uniformly nonsingular, it is sufficient that for any limit point J of $\{\boldsymbol{J}_z p(x_k - y_k, \mu_k)\}$, the linear map $\boldsymbol{J}F(x^*)J + I - J$ is nonsingular.*

Moreover the nonsingularity of $\boldsymbol{J}F(x^*)J + I - J$ is equivalent to

$$L^\perp \cap (\mathrm{nullspace}(J) \times \mathrm{nullspace}(I - J)) = \{0\},$$

where L denotes the linear subspace $\{(h,k) : \mathbf{J}F(x^*)h = k\}$.

Proof. From the definition (3.2) of H^{nat}, we get

$$\mathbf{J}_{(x,y)}H^{\mathrm{nat}}(x,y,\mu,\varepsilon) : (h,k) \in \mathbb{E}^2 \mapsto \begin{pmatrix} h - \mathbf{J}_z p(x-y,\mu)(h-k) \\ \mathbf{J}F(x)h + \varepsilon h - k \end{pmatrix}$$

for any $(x,y) \in \mathbb{E}^2$ and any $\mu, \varepsilon > 0$. It suffices to show that under the stated sufficient condition, the linear map $(h,k) \in \mathbb{E}^2 \mapsto (h - J(h-k), \mathbf{J}F(x^*)h - k)$ is nonsingular for any limit point J of $\{\mathbf{J}_z p(x_k - y_k, \mu_k)\}$. Note that

$$(h - J(h-k), \mathbf{J}F(x)h - k) = (0,0)$$
$$\Longleftrightarrow h = J(h-k) \quad \text{and} \quad (\mathbf{J}F(x^*)J + I - J)(h-k) = 0.$$

Thus this linear map is nonsingular if $\mathbf{J}F(x^*)J + I - J$ is nonsingular.

For the moreover part, we shall consider, equivalently, the nonsingularity of $(MJ + I - J)^* = JM^* + I - J$, where M denotes $\mathbf{J}F(x^*)$ and M^* denotes the adjoint map of M.

For simplicity of notation, we denote by $\mathbb{E}_0$, $\mathbb{E}_1$, and $\mathbb{E}_\rho$ the subspaces $\mathrm{nullspace}(J)$, $\mathrm{nullspace}(I - J)$, and $(\mathrm{nullspace}(J) \oplus \mathrm{nullspace}(I - J))^\perp$, respectively, and for $v \in \mathbb{E}$, we denote by v_a the projection $\Pi_{\mathbb{E}_a}(v)$ for $a \in \{0,1,\rho\}$. Define the linear maps

$$M_{ab}^* : x \in \mathbb{E}_b \mapsto (M^*x)_a \quad \text{for } a,b \in \{0,1,\rho\}.$$

The equation $(JM^* + I - J)v = 0$ is equivalent to

$$v_0 = 0,$$
$$M_{10}^*v_0 + M_{11}^*v_1 + \qquad\qquad M_{1\rho}^*v_\rho = 0, \quad \text{and}$$
$$M_{\rho0}^*v_0 + M_{\rho1}^*v_1 + (M_{\rho\rho}^* + J_\rho^{-1} - I)v_\rho = 0,$$

where J_ρ denotes the restriction of J to $\mathbb{E}_\rho$. Note that J_ρ has eigenvalues strictly between 0 and 1, and we have used this fact in scaling both sides of the last equation by J_ρ^{-1}. Observe that

$$\langle M_{11}^*v_1 + M_{\rho1}^*v_1 + M_{1\rho}^*v_\rho + (M_{\rho\rho}^* + J_\rho^{-1} - I)v_\rho, v_1 + v_\rho \rangle$$
$$= \langle M_{11}^*v_1 + M_{\rho1}^*v_1 + M_{1\rho}^*v_\rho + M_{\rho\rho}^*v_\rho, v_1 + v_\rho \rangle + \langle (J_\rho^{-1} - I)v_\rho, v_1 + v_\rho \rangle$$
$$= \langle M^*(v_1 + v_\rho), v_1 + v_\rho \rangle + \langle (J_\rho^{-1} - I)v_\rho, v_\rho \rangle$$

is the sum of two nonnegative terms since M is monotone, and $u_\rho \in \mathbb{E}_\rho \mapsto (J_\rho^{-1} - I)u_\rho$ is positive definite. Therefore if

$$M_{11}^*v_1 + \qquad\qquad M_{1\rho}^*v_\rho = 0 \quad \text{and}$$
$$M_{\rho1}^*v_1 + (M_{\rho\rho}^* + J_\rho^{-1} - I)v_\rho = 0,$$

then $\langle M_{11}^*v_1 + M_{\rho1}^*v_1 + M_{1\rho}^*v_\rho + (M_{\rho\rho}^* + J_\rho^{-1} - I)v_\rho, v_1 + v_\rho \rangle = 0$, and hence $v_\rho = 0$.

Thus $(JM^* + I - J)v = 0$ is equivalent to $v_0 = 0$, $v_\rho = 0$, and

$$M_{11}^*v_1 = 0 \quad \text{and} \quad M_{\rho1}^*v_1 = 0.$$

Hence it remains to show that $(M_{11}^*v_1 = 0 \wedge M_{\rho1}^*v_1 = 0)$ has a unique solution if and only if $L^\perp \cap \mathrm{nullspace}(J) \times \mathrm{nullspace}(I - J) = \{0\}$. This follows from

$$(M_{11}^*v_1 = 0 \wedge M_{\rho1}^*v_1 = 0) \iff (M^*v_1, v_1) \in \mathbb{E}_0 \times \mathbb{E}_1$$
$$\iff (M^*v_1, -v_1) \in \{(M^*w, -w) : w \in \mathbb{E}\} \cap \mathbb{E}_0 \times \mathbb{E}_1$$
$$\iff (M^*v_1, -v_1) \in L^\perp \cap \mathrm{nullspace}(J) \times \mathrm{nullspace}(I - J). \qquad \square$$

5. Barriers with definable gradients. A twice-differentiable barrier $f : \text{int}(X)$ $\to \mathbb{R}$ with a finite barrier parameter and a definable gradient map is necessary to employ Theorem 4.3 to deduce the superlinear convergence of Algorithm 4.2 in solving the natural map equation. For every closed convex set X that is proper (i.e., has nonempty interior and does not contain any affine subspace), it is well known that the universal barrier is a twice-differentiable barrier with a finite barrier parameter; see [25, Theorem 2.5.1]. The *universal barrier* of X is

$$f : x \in \text{int}(X) \mapsto \log(vol(X^\sharp(x))),$$

where

$$X^\sharp(x) = \{y \in \mathbb{E} : \forall z \in X \, \langle y, z - x \rangle \leq 1\}$$

is the polar set of X at x, and $vol(\cdot)$ denotes the Lebesgue measure on $\mathbb{E}$.

It remains to check the definability of the gradient of the universal barrier. We shall use a recent result of Kaiser [17]—which states that the Lebesgue measure is $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$-compatible—to prove the definability of ∇f when f is the universal barrier of a proper closed convex set X definable in $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$.

Given an o-minimal structure $\mathcal{O}$, a Borel measure λ on $\mathbb{R}^n$ is said to be $\mathcal{O}$-*compatible* if there exists an o-minimal expansion $\mathcal{O}^*$ of $\mathcal{O}$ such that for every $A \subseteq \mathbb{R}^m \times \mathbb{R}^n$ definable in $\mathcal{O}$, the set

$$\{(x, \lambda(A_x)) : \lambda(A_x) < \infty\}$$

is definable in $\mathcal{O}^*$, where A_x denotes the set $\{y \in \mathbb{R}^n : (x, y) \in A\}$. We call $\mathcal{O}^*$ an $\mathcal{O}$-*measuring o-minimal structure of* λ. The measure λ is said to be *strongly $\mathcal{O}$-compatible* if the projection of the above set on the x-component is definable in $\mathcal{O}$.

THEOREM 5.1 (Theorem 1.9 of [17]). *The Lebesgue measure vol on $\mathbb{R}^n$ is strongly $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$-compatible, and $\mathbb{R}_{\text{an,exp}}$ is an $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$-measuring o-minimal structure of vol.*

THEOREM 5.2. *The universal barrier f of a proper closed convex set $X \in \mathbb{R}^m$ that is definable in the o-minimal structure $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$ has a gradient map ∇f that is definable in the o-minimal expansion $\mathbb{R}_{\text{an,exp}}$.*

Proof. Let A be the set $\{(x, y) : x \in \text{int}(X), \, y \in X^\sharp(x)\} \cup \{(x, y) : x \notin \text{int}(X)\}$. Since X is definable in $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$, so are $X^\sharp(x)$ and A. Moreover, for every $x \in \mathbb{R}^m$,

$$\{y \in \mathbb{R}^n : (x, y) \in A\} = \begin{cases} X^\sharp(x) & \text{if } x \in \text{int}(X), \\ \mathbb{R}^n & \text{otherwise.} \end{cases}$$

Therefore, the function $g : x \in \text{int}(X) \mapsto vol(X^\sharp(x))$, whose graph is

$$\{(x, vol(A_x)) : vol(A_x) < \infty\},$$

is definable in $\mathbb{R}_{\text{an,exp}}$ by Theorem 5.1. Thus the gradient map $\nabla f = \frac{\nabla g}{g}$ of the universal barrier of X is definable in $\mathbb{R}_{\text{an,exp}}$.   □

We conclude this section by mentioning a conjecture of Kaiser [17]:

The Lebesgue measure in any arity is $\mathcal{O}$-compatible for every o-minimal structure $\mathcal{O}$.

If this conjecture is true, then we can trivially extend the above proof to deduce that the universal barrier of any definable proper closed convex set always has a gradient map definable in some o-minimal expansion. This will, of course, allow us to apply the main results in this paper to variational inequalities on any definable proper closed convex set.

6. Application to epigraphs of matrix operator and nuclear norms. In this section, we consider the variational inequality $VI(K, F)$, where $K \subseteq \mathbb{E}$ is a closed convex cone. This is equivalent to the complementarity problem of finding $x \in K$ such that $F(x) \in K^\sharp$ and $\langle F(x), x \rangle = 0$.

We shall focus on two cones, the epigraphs of the matrix operator norm and the nuclear norm. These cones are duals of each other. We show how the smoothing approximations and their Jacobians for these cones can be computed efficiently, and we establish that nondegeneracy is sufficient for the uniform nonsingularity of the Newton system in Algorithm 4.2. We end the section with some preliminary numerical results.

6.1. Operator norm. Consider $\mathbb{E} = \mathbb{R} \times \mathbb{R}^{m \times n}$ with inner product $\langle (x_0, x), (y_0, y) \rangle = x_0 y_0 + \mathrm{trace}(x^T y)$, and let K be the cone

$$K_{m,n} := \{(x_0, x) \in \mathbb{E} : x_0 \geq \|x\|\},$$

where $\|x\|$ denotes the operator norm of $x \in \mathbb{R}^{m \times n}$. We assume, without loss of generality, that $m \leq n$.

We use the log-determinant barrier

$$f_{m,n} : (x_0, x) \in \mathrm{int}(K_{m,n}) \mapsto -\log \det \begin{pmatrix} x_0 I_m & x \\ x^T & x_0 I_n \end{pmatrix},$$

where I_m denotes the m-by-m identity matrix.

The first two derivatives of the barrier $f_{m,n}$ are

$$(6.1) \qquad Df_{m,n}(x_0, x)(h_0, h) = -\mathrm{trace} \begin{pmatrix} x_0 I_m & x \\ x^T & x_0 I_n \end{pmatrix}^{-1} \begin{pmatrix} h_0 I_m & h \\ h^T & h_0 I_n \end{pmatrix}$$

and

$$\begin{aligned}
(6.2) \qquad & D^2 f_{m,n}(x_0, x)(h_0, h; k_0, k) \\
& = \mathrm{trace} \begin{pmatrix} x_0 I_m & x \\ x^T & x_0 I_n \end{pmatrix}^{-1} \begin{pmatrix} h_0 I_m & h \\ h^T & h_0 I_n \end{pmatrix} \begin{pmatrix} x_0 I_m & x \\ x^T & x_0 I_n \end{pmatrix}^{-1} \begin{pmatrix} k_0 I_m & k \\ k^T & k_0 I_n \end{pmatrix}.
\end{aligned}$$

By decomposing $x = u \sigma v^T$ into its singular values, where σ is an m-by-m diagonal matrix and u and v are, respectively, m-by-m and n-by-m matrices with orthonormal columns, we can write

$$\begin{pmatrix} x_0 I_m & x \\ x^T & x_0 I_n \end{pmatrix} = \begin{pmatrix} u & 0 & 0 \\ 0 & v & \tilde{v} \end{pmatrix} \begin{pmatrix} x_0 I_m & \sigma & 0 \\ \sigma & x_0 I_m & 0 \\ 0 & 0 & x_0 I_{n-m} \end{pmatrix} \begin{pmatrix} u^T & 0 \\ 0 & v^T \\ 0 & \tilde{v}^T \end{pmatrix},$$

where $\tilde{v}$ is any n-by-$(n-m)$ matrix such that $[v|\tilde{v}]$ is an orthogonal matrix. This gives

$$\begin{aligned}
& \begin{pmatrix} x_0 I_m & x \\ x^T & x_0 I_n \end{pmatrix}^{-1} \\
& = \begin{pmatrix} u & 0 & 0 \\ 0 & v & \tilde{v} \end{pmatrix} \begin{pmatrix} x_0(x_0^2 I_m - \sigma^2)^{-1} & -\sigma(x_0^2 I_m - \sigma^2)^{-1} & 0 \\ -\sigma(x_0^2 I_m - \sigma^2)^{-1} & x_0(x_0^2 I_m - \sigma^2)^{-1} & 0 \\ 0 & 0 & \frac{1}{x_0} I_{n-m} \end{pmatrix} \begin{pmatrix} u^T & 0 \\ 0 & v^T \\ 0 & \tilde{v}^T \end{pmatrix}.
\end{aligned}$$

We can then determine from, respectively, (6.1) and (6.2) that the gradient of $f_{m,n}$ is

$$\nabla f_{m,n} : (x_0, x) \in \text{int}(K_{m,n}) \mapsto \left(-2x_0 \sum_{i=1}^{m} \frac{1}{x_0^2 - \sigma_i^2} - \frac{n-m}{x_0}, -2u\sigma(x_0^2 I_m - \sigma^2)^{-1} v^T \right)$$

and its Hessian at $(x_0, x) \in \text{int}(K_{m,n})$ is

$$\nabla^2 f_{m,n}(x_0, x) : (h_0, h) \in \mathbb{E}$$

$$\mapsto \left(\begin{array}{c} h_0 \left(2 \sum_{i=1}^{m} \frac{x_0^2 + \sigma_i^2}{(x_0^2 - \sigma_i^2)^2} + \frac{n-m}{x_0^2} \right) - 4x_0 \, \text{trace}(v\sigma(x_0^2 I_m - \sigma^2)^{-2} u^T h), \\ -4h_0 x_0 u \sigma (x_0^2 I_m - \sigma^2)^{-2} v^T + 2x_0^2 u (x_0^2 I_m - \sigma^2)^{-1} u^T h v (x_0^2 I_m - \sigma^2)^{-1} v^T \\ +2u\sigma(x_0^2 I_m - \sigma^2)^{-1} v^T h^T u \sigma (x_0^2 I_m - \sigma^2)^{-1} v^T + 2u(x_0^2 I_m - \sigma^2)^{-1} u^T h \tilde{v} \tilde{v}^T \end{array} \right),$$

where $\sigma_1, \ldots, \sigma_m$ are the diagonal entries of σ. We remark that the gradient of the barrier $f_{m,n}$ is in the o-minimal structure $\mathcal{SA}$.

We denote by $p_{m,n}$ the smoothing approximation determined by $f_{m,n}$ and call it the *standard smoothing approximation for* $K_{m,n}$. From the definition of the smoothing approximation, for each $\mu > 0$, $p_{m,n}(z_0, z; \mu) = (x_0, x)$ if and only if $|\sigma_1|, \ldots, |\sigma_m| < x_0$ and $u^T z v$ is an m-by-m diagonal matrix containing diagonal entries $\zeta_1, \ldots, \zeta_m$ satisfying

$$(6.3) \qquad \begin{aligned} x_0 - 2\mu^2 x_0 \sum_{i=1}^{m} \frac{1}{x_0^2 - \sigma_i^2} - \mu^2 \frac{n-m}{x_0} &= z_0 \quad \text{and} \\ \sigma_i + 2\mu^2 \frac{\sigma_i}{x_0^2 - \sigma_i^2} &= \zeta_i \quad \text{for } i = 1, \ldots, m. \end{aligned}$$

Thus given any $(z_0, z) \in \mathbb{E}$ and any $\mu > 0$, we can determine $(x_0, x) = p_{m,n}(z_0, z; \mu)$ by decomposing $z = u\zeta v^T$ into its singular values and solving (6.3) to get a unique solution $(x_0, \sigma_1, \ldots, \sigma_m)$ satisfying $|\sigma_1|, \ldots, |\sigma_m| < x_0$, followed by forming $x = u\sigma v^T$ where σ is the diagonal matrix $\text{Diag}(\sigma_1, \ldots, \sigma_m)$.

To solve (6.3), we first observe that for each fixed $x_0 > 0$ and ζ_i, we can uniquely determine σ_i as the second largest root of a cubic polynomial with three distinct real roots $\gamma_1, \gamma_2, \gamma_3$ satisfying $\gamma_1 < -x_0 < \gamma_2 < x_0 < \gamma_3$. Thus for any given z, the last m equations of (6.3) determine each σ_i as an analytic function of x_0 over $(0, \infty)$. Moreover these analytic functions $\sigma_i(x_0)$ are strictly increasing since $(x_0, \sigma_i) \mapsto \sigma_i + 2\mu^2 \frac{\sigma_i}{x_0^2 - \sigma_i^2}$ is increasing in σ_i but decreasing in x_0. Therefore

$$x_0 \mapsto x_0 - 2\mu^2 x_0 \sum_{i=1}^{m} \frac{1}{x_0^2 - \sigma_i(x_0)^2} - \mu^2 \frac{n-m}{x_0}$$

$$= x_0 + x_0 \sum_{i=1}^{m} \frac{\sigma_i(x_0) - \zeta_i}{\sigma_i(x_0)} - \mu^2 \frac{n-m}{x_0}$$

is strictly increasing. Consequently we can determine x_0 with the bisection method; all we need is an upper bound on x_0. This upper bound can be obtained by adding

all $m + 1$ equations in (6.3) to get

$$z_0 + \sum_{i=1}^{m} |\zeta_i| = x_0 + \sum_{i=1}^{m} |\sigma_i| - 2\mu^2 \sum_{i=1}^{m} \frac{x_0 - |\sigma_i|}{x_0^2 - \sigma_i^2} - \mu^2 \frac{n - m}{x_0}$$

$$= x_0 + \sum_{i=1}^{m} |\sigma_i| - 2\mu^2 \sum_{i=1}^{m} \frac{1}{x_0 + |\sigma_i|} - \mu^2 \frac{n - m}{x_0}$$

$$\geq x_0 - 2\mu^2 \sum_{i=1}^{m} \frac{1}{x_0} - \mu^2 \frac{n - m}{x_0} = x_0 - \mu^2 \frac{n + m}{x_0},$$

which gives $\frac{1}{2}\left(\sqrt{(z_0 + \sum_{i=1}^{m} |\zeta_i|)^2 + 4(n + m)\mu^2} + z_0 + \sum_{i=1}^{m} |\zeta_i|\right)$ as an upper bound.

Once $(x_0, x) = p_{m,n}(z_0, z; \mu)$ is determined, the Jacobian $\mathbf{J}_{(z_0,z)}p_{m,n}(z_0, z; \mu)$ can then be computed as $\left(I + \mu^2 \nabla^2 f_{m,n}(x_0, x)\right)^{-1}$, where I denotes the identity map. To simplify computations, we introduce the unitary transformation

(6.4) $$\Theta_{u,v,\tilde{v}} : (h_0, h) \in \mathbb{E} \mapsto (h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h}) \in \mathbb{R}^{1+mn},$$

where

$$\grave{h} = (\grave{h}_1, \ldots, \grave{h}_m), \qquad\qquad \acute{h} = (\acute{h}_{12}, \acute{h}_{13}, \acute{h}_{23}, \ldots, \acute{h}_{1m}, \ldots, \acute{h}_{m-1,m}),$$

$$\hat{h} = (\hat{h}_{12}, \hat{h}_{13}, \hat{h}_{23}, \ldots, \hat{h}_{1m}, \ldots, \hat{h}_{m-1,m}), \quad \tilde{h} = (\tilde{h}_{11}, \ldots, \tilde{h}_{m1}, \ldots, \tilde{h}_{1,n-m}, \ldots, \tilde{h}_{m,n-m})$$

with

$$\grave{h}_i = u_i^T h v_i, \quad \acute{h}_{ij} = \frac{u_i^T h v_j + v_i^T h u_j}{\sqrt{2}}, \quad \hat{h}_{ij} = \frac{u_i^T h v_j - v_i^T h u_j}{\sqrt{2}}, \quad \text{and} \quad \tilde{h}_{ij} = u_i^T h \tilde{v}_j.$$

Here u_i, v_i, and $\tilde{v}_i$ denote the ith column of, respectively, u, v, and $\tilde{v}$, and $x = u\sigma v^T$ is any singular value decomposition with σ an m-by-m diagonal matrix containing the singular values $\sigma_1, \ldots, \sigma_m$, and $\tilde{v}$ is any matrix such that $[v|\tilde{v}]$ is orthogonal. It is straightforward to verify that its adjoint map is

$$\Theta_{u,v,\tilde{v}}^* = \Theta_{u,v,\tilde{v}}^{-1} : (h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h}) \in \mathbb{R}^{1+mn}$$

$$\mapsto \left(h_0, u \begin{pmatrix} \grave{h}_1 & & \frac{\acute{h}_{ij}+\hat{h}_{ij}}{\sqrt{2}} \\ & \ddots & \\ \frac{\acute{h}_{ij}-\hat{h}_{ij}}{\sqrt{2}} & & \grave{h}_m \end{pmatrix} v^T + u \begin{pmatrix} \tilde{h}_{11} & \cdots & \tilde{h}_{1,n-m} \\ \vdots & \ddots & \vdots \\ \tilde{h}_{m1} & \cdots & \tilde{h}_{m,n-m} \end{pmatrix} \tilde{v}^T \right).$$

We can then write $\nabla^2 f_{m,n}(x_0, x) = \Theta_{u,v,\tilde{v}}^{-1} H_{x_0,\sigma_1,\ldots,\sigma_m} \Theta_{u,v,\tilde{v}}$, where $H_{x_0,\sigma_1,\ldots,\sigma_m}$ maps a vector $(h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h}) \in \mathbb{R}^{1+mn}$ to the vector $(k_0, \grave{k}, \acute{k}, \hat{k}, \tilde{k})$ with

$$k_0 = h_0 \left(\sum_{i=1}^{m} \frac{2(x_0^2 + \sigma_i^2)}{(x_0^2 - \sigma_i^2)^2} + \frac{n - m}{x_0^2}\right) + \sum_{i=1}^{m} \frac{-4x_0\sigma_i \grave{h}_i}{(x_0^2 - \sigma_i^2)^2},$$

$$\grave{k}_i = \frac{-4x_0\sigma_i h_0}{(x_0^2 - \sigma_i^2)^2} + \frac{2(x_0^2 + \sigma_i^2)\grave{h}_i}{(x_0^2 - \sigma_i^2)^2}, \quad \acute{k}_{ij} = \frac{2(x_0^2 + \sigma_i\sigma_j)\acute{h}_{ij}}{(x_0^2 - \sigma_i^2)(x_0^2 - \sigma_j^2)},$$

$$\hat{k}_{ij} = \frac{2(x_0^2 - \sigma_i\sigma_j)\hat{h}_{ij}}{(x_0^2 - \sigma_i^2)(x_0^2 - \sigma_j^2)}, \qquad\qquad \tilde{k}_{ij} = \frac{2\tilde{h}_{ij}}{x_0^2 - \sigma_i^2}.$$

The Jacobian of the smoothing approximation is thus

$$(6.5) \qquad \mathbf{J}_{(z_0,z)}p_{m,n}(z_0, z; \mu) = \Theta_{u,v,\tilde{v}}^{-1} \left(I + \mu^2 H_{x_0, \sigma_1, \dots, \sigma_m} \right)^{-1} \Theta_{u,v,\tilde{v}},$$

where $\left(I + \mu^2 H_{x_0, \sigma_1, \dots, \sigma_m} \right)^{-1}$ takes a vector $(h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h})$ to the vector $(k_0, \grave{k}, \acute{k}, \hat{k}, \tilde{k})$ with

$$k_0 = \frac{h_0 - b^T C^{-1} \grave{h}}{a - b^T C^{-1} b}, \qquad\qquad \grave{k} = C^{-1} \grave{h} - \frac{h_0 - b^T C^{-1} \grave{h}}{a - b^T C^{-1} b} C^{-1} b,$$

$$\acute{k}_{ij} = \frac{(x_0^2 - \sigma_i^2)(x_0^2 - \sigma_j^2) \acute{h}_{ij}}{(x_0^2 - \sigma_i^2)(x_0^2 - \sigma_j^2) + 2\mu^2(x_0^2 + \sigma_i \sigma_j)}, \quad \hat{k}_{ij} = \frac{(x_0^2 - \sigma_i^2)(x_0^2 - \sigma_j^2) \hat{h}_{ij}}{(x_0^2 - \sigma_i^2)(x_0^2 - \sigma_j^2) + 2\mu^2(x_0^2 - \sigma_i \sigma_j)},$$

$$\tilde{k}_{ij} = \frac{(x_0^2 - \sigma_i^2) \tilde{h}_{ij}}{x_0^2 - \sigma_i^2 + 2\mu^2},$$

where

$$a = 1 + \sum_{i=1}^{m} \frac{2\mu^2(x_0^2 + \sigma_i^2)}{(x_0^2 - \sigma_i^2)^2} + \mu^2 \frac{n-m}{x_0^2}, \quad b = \left(\frac{-4\mu^2 x_0 \sigma_1}{(x_0^2 - \sigma_1^2)^2}, \dots, \frac{-4\mu^2 x_0 \sigma_m}{(x_0^2 - \sigma_m^2)^2} \right),$$

and

$$C^{-1} = \mathrm{Diag}\left(\left\{ 1 + \frac{2\mu^2(x_0^2 + \sigma_i^2)}{(x_0^2 - \sigma_i^2)^2} \right\}_{i=1}^{m} \right)^{-1} = \mathrm{Diag}\left(\left\{ \frac{(x_0^2 - \sigma_i^2)^2}{(x_0^2 - \sigma_i^2)^2 + 2\mu^2(x_0^2 + \sigma_i^2)} \right\}_{i=1}^{m} \right).$$

6.2. Nuclear norm. Consider again $\mathbb{E} = \mathbb{R} \times \mathbb{R}^{m \times n}$ with inner product $\langle (x_0, x),$ $(y_0, y) \rangle = x_0 y_0 + \mathrm{trace}(x^T y)$, and let K be the cone

$$K_{m,n}^{\sharp} := \{ (x_0, x) \in \mathbb{E} : x_0 \geq \|x\|_* \},$$

where $\|x\|_*$ denotes the nuclear norm of $x \in \mathbb{R}^{m \times n}$, i.e., the sum of all singular values of x. We assume, without loss of generality, that $m \leq n$. The cone $K_{m,n}^{\sharp}$ is the dual cone of $K_{m,n}$.

We use the (modified) Fenchel conjugate barrier

$$f_{m,n}^{\sharp} : (s_0, s) \in \mathrm{int}(K_{m,n}^{\sharp}) \mapsto -\inf\{ s_0 x_0 + \mathrm{trace}(s^T x) + f_{m,n}(x_0, x) : (x_0, x) \in K_{m,n} \}.$$

Let $p_{m,n}^{\sharp}$ denote the smoothing approximation defined by $f_{m,n}^{\sharp}$ and call it the *standard smoothing approximation for* $K_{m,n}^{\sharp}$. To compute $p_{m,n}^{\sharp}$, we note that

$$\mu^2 \nabla f_{m,n}(x_0, x) = -(s_0, s) \iff \mu^2 \nabla f_{m,n}^{\sharp}(s_0, s) = -(x_0, x),$$

and hence
(6.6)

$$p_{m,n}^{\sharp}(z_0, z; \mu) = (s_0, s)$$

$$\iff (s_0, s) + \mu^2 \nabla f_{m,n}^{\sharp}(s_0, s) = (z_0, z)$$

$$\iff \exists (x_0, x), \ (s_0, s) - (x_0, x) = (z_0, z) \text{ and } (x_0, x) = -\mu^2 \nabla f_{m,n}^{\sharp}(s_0, s)$$

$$\iff \exists (x_0, x), \ (s_0, s) - (x_0, x) = (z_0, z) \text{ and } (s_0, s) = -\mu^2 \nabla f_{m,n}(x_0, x)$$

$$\iff \exists (x_0, x), \ (s_0, s) = (z_0, z) + (x_0, x) \text{ and } \mu^2 \nabla f_{m,n}(x_0, x) + (x_0, x) = -(z_0, z)$$

$$\iff \exists (x_0, x), \ (s_0, s) = (z_0, z) + (x_0, x) \text{ and } p_{m,n}(-z_0, -z; \mu) = (x_0, x)$$

$$\iff (s_0, s) = (z_0, z) + p_{m,n}(-z_0, -z; \mu).$$

To compute the Jacobian $\mathbf{J}_{(z_0,z)}p_{m,n}^\sharp(z_0,z;\mu)$, we differentiate $p_{m,n}^\sharp(z_0,z;\mu) = (z_0,z) + p_{m,n}(-z_0,-z;\mu)$ with respect to (z_0,z) to get

$$\mathbf{J}_{(z_0,z)}p_{m,n}^\sharp(z_0,z;\mu) = I - \mathbf{J}_{(z_0,z)}p_{m,n}(-z_0,-z;\mu).$$

6.3. Uniform nonsingularity of Newton system. For closed convex cone K, we denote by $K^\sharp$ its dual cone, and for any face F of K, we denote by $F^\triangle$ the *complementary face* $\{v \in K^\sharp : \forall u \in F, \langle v,u \rangle = 0\}$ of F.

LEMMA 6.1. *If $\{z_k = (z_{0,k}, \bar{z}_k)\}$ is a convergent sequence in $\mathbb{E}$ and μ_k is a positive sequence converging to 0, then every limit point J of the sequence $\{\mathbf{J}_z p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k)\}$ satisfies*

$$nullspace(J) \subseteq span(F_{x^*}^\triangle) \quad and \quad nullspace(I - J) \subseteq span(F_{y^*}^\triangle),$$

where F_{x^*} and F_{y^*} are, respectively, the smallest faces of $K_{m,n}$ and $K_{m,n}^\sharp$ containing the limits

$$x^* := (x_0^*, \bar{x}^*) := \lim_{k \to \infty} p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k)$$

and

$$y^* := (y_0^*, \bar{y}^*) := \lim_{k \to \infty} p_{m,n}^\sharp(-z_{0,k}, -\bar{z}_k; \mu_k).$$

Proof. By taking a subsequence if necessary, we may assume $\{\mathbf{J}_z p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k)\}$ converges to J.

If F_{x^*} is the trivial face $\{0\}$, then nullspace$(J) \subseteq$ span$(F_{x^*}^\triangle) = \mathbb{E}$ trivially holds. Similarly, nullspace$(I - J) \subseteq$ span$(F_{y^*}^\triangle) = \mathbb{E}$ when F_{y^*} is the trivial face.

If F_{x^*} is the cone $K_{m,n}$, then $x^* \in \text{int}(K_{m,n})$, whence z_k converges to a point z^* in the interior of $K_{m,n}$. Therefore $\nabla^2 f_{m,n}(z_{0,k}, \bar{z}_k)$ converges to the positive definite map $\nabla^2 f_{m,n}(z^*)$, and thus $\mathbf{J}_z p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k) = (I + \mu^2 \nabla^2 f_{m,n}(z_{0,k}, \bar{z}_k))^{-1}$ converges to the identity map. So $\{0\} = $ nullspace$(J) \subseteq$ span$(F_{x^*}^\triangle)$ trivially holds. Similarly, $\{0\} = $ nullspace$(I - J) \subseteq$ span$(F_{y^*}^\triangle)$ when F_{y^*} is the cone $K_{m,n}^\sharp$.

Henceforth in this proof, we assume that F_{x^*} and F_{y^*} are nontrivial proper faces of, respectively, $K_{m,n}$ and $K_{m,n}^\sharp$; i.e., x^* and y^* are nonzero vectors on the boundary of their respective cones. For clarity, the remainder of the proof is presented in four parts.

(1) For each k, let $\bar{z}_k = u_k \zeta_k v_k^T$ be a singular value decomposition with ζ_k an m-by-m diagonal matrix containing the singular values $\zeta_{1,k} \geq \cdots \geq \zeta_{m,k} \geq 0$. Recall from (6.3) that for $(x_{0,k}, \bar{x}_k) := p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k)$ and $(y_{0,k}, \bar{y}_k) := p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k) - (z_{0,k}, \bar{z}_k)$, which, by (6.6), is equivalent to $p_{m,n}^\sharp(-z_{0,k}, -\bar{z}_k; \mu_k)$, the products $u_k^T \bar{x}_k v_k$ and $u_k^T \bar{y}_k v_k$ are diagonal matrices σ_k and τ_k with diagonal entries $\sigma_{1,k}, \ldots, \sigma_{m,k}$ and $\tau_{1,k}, \ldots, \tau_{m,k}$, respectively, satisfying $\sigma_{1,k} \geq \cdots \geq \sigma_{m,k} \geq 0$, $\tau_{1,k} \leq \cdots \leq \tau_{m,k} \leq 0$, and

$$(6.7) \qquad \tau_{i,k} = -2\mu_k^2 \frac{\sigma_{i,k}}{x_{0,k}^2 - \sigma_{i,k}^2} \quad \text{for } i = 1, \ldots, m.$$

In terms of the unitary transformation $\Theta_{u_k,v_k,\tilde{v}_k}$ defined in (6.4), where $\tilde{v}_k$ is any matrix such that $[v_k | \tilde{v}_k]$ is orthogonal, this means $(x_{0,k}, \bar{x}_k) = \Theta_{u_k,v_k,\tilde{v}_k}^{-1}(x_{0,k}, \grave{x}_k, 0, 0, 0)$ with $\grave{x}_k = (\sigma_{1,k}, \ldots, \sigma_{m,k})$, and $(y_{0,k}, \bar{y}_k) = \Theta_{u_k,v_k,\tilde{v}_k}^{-1}(y_{0,k}, \grave{y}_k, 0, 0, 0)$ with $\grave{y}_k = (\tau_{1,k}, \ldots, \tau_{m,k})$. By taking a subsequence if necessary, we may assume, without loss

of generality, that $\{(u_k, v_k, \tilde{v}_k)\}$ converges to, say, $(u^*, v^*, \tilde{v}^*)$. Then x^* and y^* are given by

$$
\begin{aligned}
(6.8) \quad & (x_0^*, \bar{x}^*) = \Theta_{u^*,v^*,\tilde{v}^*}^{-1}(x_0^*, \dot{x}^*, 0, 0, 0) \text{ with } \dot{x}^* = (\sigma_1^*, \ldots, \sigma_m^*) \\
& \text{and} \quad (y_0^*, \bar{y}^*) = \Theta_{u^*,v^*,\tilde{v}^*}^{-1}(y_0^*, \dot{y}^*, 0, 0, 0) \text{ with } \dot{y}^* = (\tau_1^*, \ldots, \tau_m^*),
\end{aligned}
$$

where $\sigma^* = \lim_{k\to\infty} \sigma_k$ and $\tau^* = \lim_{k\to\infty} \tau_k$ have diagonal entries satisfying

$$
\begin{aligned}
(6.9) \quad & \sigma_1^* = \cdots = \sigma_r^* > \sigma_{r+1}^* \geq \cdots \geq \sigma_{m+1}^* := 0 \\
& \text{and} \quad \tau_1^* \leq \cdots \leq \tau_{r^\sharp}^* < \tau_{r^\sharp+1}^* = \cdots = \tau_{m+1}^* := 0
\end{aligned}
$$

for some $r, r^\sharp \in \{1, \ldots, m\}$. Moreover $r \geq r^\sharp$ because $(x_0^*, \bar{x}^*)$ and $(y_0^*, \bar{y}^*)$ are orthogonal.

(2) Based on the above characterizations of x^* and y^*, the smallest faces of $K_{m,n}$ and $K_{m,n}^\sharp$ containing, respectively, $x^* = (x_0^*, \bar{x}^*)$ and $y^* = (y_0^*, \bar{y}^*)$ are

$$
F_{x^*} = \left\{ (x_0, \bar{x}) : \bar{x} = u^* \begin{pmatrix} x_0 I_r & 0 \\ 0 & M \end{pmatrix} (v^*)^T, \ M \in \mathbb{R}^{(m-r)\times(n-r)}, \ \|M\| \leq x_0 \right\}
$$

and

$$
F_{y^*} = \left\{ (y_0, \bar{y}) : \bar{y} = u^* \begin{pmatrix} -N & 0 \\ 0 & 0 \end{pmatrix} (v^*)^T, \ N \in \mathbb{S}_+^{r^\sharp}, \ \text{trace}(N) = y_0 \right\},
$$

where $\mathbb{S}_+^{r^\sharp}$ denotes the cone of $r^\sharp$-by-$r^\sharp$ real symmetric positive semidefinite matrices; see Examples 5.6 and 5.7 of [10]. The respective complementary faces are thus

$$
F_{x^*}^\triangle = \left\{ (y_0, \bar{y}) : \bar{y} = u^* \begin{pmatrix} -N & A \\ C & 0 \end{pmatrix} (v^*)^T, \ N \in \mathbb{R}^{r\times r}, \ \text{trace}(N) = y_0 \geq \left\| \begin{pmatrix} N & -A \\ -C & 0 \end{pmatrix} \right\|_* \right\}
$$

and

$$
F_{y^*}^\triangle = \left\{ (x_0, \bar{x}) : \bar{x} = u^* \begin{pmatrix} x_0 I_{r^\sharp} & A \\ C & M \end{pmatrix} (v^*)^T, \ M \in \mathbb{R}^{(m-r^\sharp)\times(n-r^\sharp)}, \ \left\| \begin{pmatrix} x_0 I_{r^\sharp} & A \\ C & M \end{pmatrix} \right\| \leq x_0 \right\}.
$$

Observe that by von Neumann's trace inequality, $\text{trace}(N) = y_0 \geq \left\| \begin{pmatrix} N & -A \\ -C & 0 \end{pmatrix} \right\|_*$ implies that $\text{trace}(N) \geq \left\| \begin{pmatrix} N & -A \\ -C & 0 \end{pmatrix} \right\|_* \geq \text{trace} \begin{pmatrix} N & -A \\ -C & 0 \end{pmatrix} = \text{trace}(N)$, whence the fact that equality holds here further implies that $\begin{pmatrix} N & -A \\ -C & 0 \end{pmatrix}$ is of the form $[H|0]$, where H is symmetric and positive semidefinite (see, e.g., [1, p. 29]). Hence both A and C are zero matrices, which simplifies the complementary face $F_{x^*}^\triangle$ to give

$$
(6.10) \quad \text{span}(F_{x^*}^\triangle) = \Theta_{u^*,v^*,\tilde{v}^*}^{-1} \{ (y_0, \dot{y}, \acute{y}, 0, 0) : \dot{y}_j, \acute{y}_{ij} = 0 \, \forall j > r, \ y_0 + \dot{y}_1 + \cdots + \dot{y}_r = 0 \}.
$$

Also observe that the operator norm of a matrix is at least the norm of each of its columns/rows, and hence $x_0 \geq \left\| \begin{pmatrix} x_0 I_{r^\sharp} & A \\ C & M \end{pmatrix} \right\|$ implies that both A and C are zero matrices. This simplifies the complementary face $F_{y^*}^\triangle$ to give
(6.11)
$$
\text{span}(F_{y^*}^\triangle) = \Theta_{u^*,v^*,\tilde{v}^*}^{-1} \{ (x_0, \dot{x}, \acute{x}, \hat{x}, \tilde{x}) : \acute{x}_{ij}, \hat{x}_{ij}, \tilde{x}_{ij} = 0 \, \forall i \leq r^\sharp, \ \dot{x}_i = x_0 \, \forall i \leq r^\sharp \}.
$$

(3) It follows from (6.5) that the inner product $\left\langle \mathbf{J}_z p_{m,n}(z_{0,k}, \bar{z}_k; \mu_k)(h_0, \bar{h}), (h_0, \bar{h}) \right\rangle$ is

(6.12)

$$\frac{(h_0 - \mathrm{trace}(b_k^T C_k^{-1} \grave{h}))^2}{a_k - \mathrm{trace}(b_k^T C_k^{-1} b_k)} + \sum_{i=1}^{m} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)^2 \grave{h}_i^2}{(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)}$$

$$+ \sum_{1 \le i < j \le m} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)\acute{h}_{ij}^2}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}\sigma_{j,k})}$$

$$+ \sum_{1 \le i < j \le m} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)\hat{h}_{ij}^2}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 - \sigma_{i,k}\sigma_{j,k})} + \sum_{i=1}^{m} \sum_{j=1}^{n-m} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)\tilde{h}_{ij}^2}{x_{0,k}^2 - \sigma_{i,k}^2 + 2\mu_k^2},$$

where $(h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h}) = \Theta_{u_k, v_k, \tilde{v}_k}^{-1}(h_0, \bar{h})$,

$$a_k = 1 + \sum_{i=1}^{m} \frac{2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)^2} + \mu_k^2 \frac{n-m}{x_{0,k}^2}, \quad b_k = \left(\frac{-4\mu_k^2 x_{0,k} \sigma_{1,k}}{(x_{0,k}^2 - \sigma_{1,k}^2)^2}, \dots, \frac{-4\mu_k^2 x_{0,k} \sigma_{m,k}}{(x_{0,k}^2 - \sigma_{m,k}^2)^2} \right),$$

and

$$C_k^{-1} = \mathrm{Diag}\left(\left\{ \frac{(x_{0,k}^2 - \sigma_{i,k}^2)^2}{(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)} \right\}_{i=1}^{m} \right).$$

By taking limit as $k \to \infty$ and using (6.7) and (6.9), we get $\left\langle J(h_0, \bar{h}), (h_0, \bar{h}) \right\rangle$ to be the sum of squares

(6.13)

$$\frac{\left(h_0 + \sum_{i=1}^{r^\sharp} \grave{h}_i + \sum_{i=r^\sharp+1}^{r} \alpha_{ii} \grave{h}_i \right)^2}{1 + r^\sharp + \sum_{i=1}^{r} \alpha_{ii}} + \sum_{i=r^\sharp+1}^{r} (1 - \alpha_{ii}) \grave{h}_i^2 + \sum_{i=r+1}^{m} \grave{h}_i^2$$

$$+ \sum_{\substack{(i \le r^\sharp \le r < j) \\ \vee (r^\sharp < i < j \le r)}} (1 - \alpha_{ij}) \acute{h}_{ij}^2 + \sum_{r^\sharp < i, r < j} \acute{h}_{ij}^2$$

$$+ \sum_{(i \le r^\sharp) \wedge (i < j)} \beta_{ij} \hat{h}_{ij}^2 + \sum_{r^\sharp < i < j} \hat{h}_{ij}^2 + \sum_{j=1}^{n-m} \left(\sum_{i=1}^{r^\sharp} \gamma_i \tilde{h}_{ij}^2 + \sum_{i=r^\sharp+1}^{m} \tilde{h}_{ij}^2 \right)$$

for some α_{ij}, β_{ij}, and γ_i satisfying

$$\alpha_{ij} \begin{cases} \in (0,1) & \text{when } i \le r^\sharp \le r < j, \\ \in [0,1] & \text{when } r^\sharp < i \le j \le r, \end{cases} \quad \text{and} \quad \beta_{ij}, \gamma_i \in (0,1) \text{ when } i \le r^\sharp,$$

where $(h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h}) = \Theta_{u^*, v^*, \tilde{v}^*}^{-1}(h_0, \bar{h})$; see the details in Appendix A. Using the fact that $\Theta_{u^*, v^*, \tilde{v}^*}$ is unitary, we then get

$$\left\langle (I - J)(h_0, \bar{h}), (h_0, \bar{h}) \right\rangle = \left\| (h_0, \grave{h}, \acute{h}, \hat{h}, \tilde{h}) \right\|^2 - \left\langle J(h_0, \bar{h}), (h_0, \bar{h}) \right\rangle$$

to be the sum of squares

$$
\begin{aligned}
&\frac{\sum_{i=1}^{r^\sharp}(h_0-\grave h_i)^2}{1+r^\sharp+\sum_{i=1}^r\alpha_{ii}}+\frac{\sum_{i=r^\sharp+1}^r\alpha_{ii}(h_0-\grave h_i)^2}{1+r^\sharp+\sum_{i=1}^r\alpha_{ii}}+\frac{\sum_{i<j\le r^\sharp}(\grave h_i-\grave h_j)^2}{1+r^\sharp+\sum_{i=1}^r\alpha_{ii}}\\
&+\frac{\sum_{i\le r^\sharp<j\le r}\alpha_{jj}(\grave h_i-\grave h_j)^2}{1+r^\sharp+\sum_{i=1}^r\alpha_{ii}}+\frac{\sum_{r^\sharp<i<j\le r}\alpha_{ii}\alpha_{jj}(\grave h_i-\grave h_j)^2}{1+r^\sharp+\sum_{i=1}^r\alpha_{ii}}\\
&+\sum_{\substack{(i\le r^\sharp\le r<j)\\ \vee(r^\sharp<i<j\le r)}}\alpha_{ij}\acute h_{ij}^2+\sum_{(i\le r^\sharp)\wedge(i<j\le r)}\acute h_{ij}^2\\
&+\sum_{(i\le r^\sharp)\wedge(i<j)}(1-\beta_{ij})\hat h_{ij}^2+\sum_{j=1}^{n-m}\sum_{i=1}^{r^\sharp}(1-\gamma_i)\tilde h_{ij}^2.
\end{aligned}
$$
(6.14)

(4) For any $(h_0,\bar h)\in\text{nullspace}(J)$, the inner product $\langle J(h_0,\bar h),(h_0,\bar h)\rangle$ vanishes. In this case, we deduce from (6.13) that $(h_0,\grave h,\acute h,\hat h,\tilde h)=\Theta_{u^*,v^*,\tilde v^*}(h_0,\bar h)$ satisfies

$$
h_0+\sum_{i=1}^{r^\sharp}\grave h_i+\sum_{i=r^\sharp+1}^r\alpha_{ii}\grave h_i=0,\qquad \grave h_i=0\quad\text{for }i\in\{r^\sharp+1,\dots,r\}\text{ with }\alpha_{ii}<1,
$$

$$
\grave h_i=0\quad\text{for }r<i,\qquad \acute h_{ij}=0\quad\text{for }r<j,\qquad \hat h=0,\qquad \tilde h=0,
$$

which, by (6.10), shows $(h_0,\bar h)\in\text{span}(F_{y^*}^\triangle)$. Similarly, for any $(h_0,\bar h)\in\text{nullspace}(I-J)$, the inner product $\langle(I-J)(h_0,\bar h),(h_0,\bar h)\rangle$ vanishes. In this case, we deduce from (6.14) that $(h_0,\grave h,\acute h,\hat h,\tilde h)=\Theta_{u^*,v^*,\tilde v^*}(h_0,\bar h)$ satisfies

$$
h_0=\grave h_1=\cdots=\grave h_{r^\sharp},\qquad \acute h_{ij},\hat h_{ij},\tilde h_{ij}=0\quad\text{for }i\le r^\sharp,
$$

which, by (6.11), shows that $(h_0,\bar h)\in\text{span}(F_{x^*}^\triangle)$. □

REMARK 6.2. Since $\mathbf J_z p_{m,n}^\sharp=I-\mathbf J_z p_{m,n}$, the statement of Lemma 6.1 holds when $p_{m,n}$ is replaced by $p_{m,n}^\sharp$.

Given a closed convex cone $K\subseteq\mathbb E$, its dual cone $K^\sharp$, and a linear subspace $L\subseteq\mathbb E$, a vector $x\in K$ is said to be *nondegenerate for L* if the smallest face F of K containing x satisfies $L^\perp\cap\text{span}(F^\triangle)=\{0\}$; see, e.g., [26].

THEOREM 6.3. *Let $K\subseteq\mathbb E$ be either $K_{m,n}$ or $K_{m,n}^\sharp$, and let p be its standard smoothing approximations. For any monotone nonlinear map $F:\mathbb E\to\mathbb E$, if the sequence $\{(x_k,y_k,\mu_k,\varepsilon_k)\}$ generated by Algorithm 4.2 converges to $(x^*,y^*,0,0)$ with $(x^*,y^*)\in K\times K^\sharp$ nondegenerate for $L=\{(h,k):\mathbf JF(x^*)h=k\}$, then $\{\mathbf J_{(x,y)}H^{\mathrm{nat}}(x_k,y_k,\mu_k,\varepsilon_k)\}$ is uniformly nonsingular.*

Proof. By Theorem 4.4, it suffices to prove that for any limit point J of $\{\mathbf J_z p(x_k-y_k,\mu_k)\}$,

$$
L^\perp\cap(\text{nullspace}(J)\times\text{nullspace}(I-J))=\{0\},
$$

where $L=\{(h,k):\mathbf JF(x^*)h=k\}$. By Lemma 6.1, the remark immediately following it, and the fact that

$$
\lim_{k\to\infty}G^{\mathrm{nat}}(x_k,y_k)=0\implies\lim_{k\to\infty}x_k-p(x_k-y_k,\mu_k)=0\implies x^*=\Pi_K(x^*-y^*),
$$

we conclude that

$$
L^\perp\cap(\text{nullspace}(J)\times\text{nullspace}(I-J))\subseteq L^\perp\cap\text{span}(F_{y^*}^\triangle)\times\text{span}(F_{x^*}^\triangle).
$$

Thus if (x^*, y^*) is nondegenerate for L, then

$$\{0\} = L^\perp \cap \text{span}((F_{y^*} \times F_{x^*})^\triangle) = L^\perp \cap (\text{span}(F_{y^*}^\triangle) \times \text{span}(F_{x^*}^\triangle))$$
$$\supseteq L^\perp \cap (\text{nullspace}(J) \times \text{nullspace}(I - J))$$

proves the theorem.   □

6.4. Numerical tests. We tested Algorithm 4.2 on randomly generated instances of operator norm and nuclear norm linear complementarity problems. In each problem instance, F is the affine map $x \in \mathbb{R}^N \mapsto Mx + q$, where M is an N-by-N random matrix of a certain rank r and q is a random N-vector.

The matrix M is obtained by taking the product BB^T, where B is a random n-by-r matrix, where each entry is a standard normal random variable. The vector q is obtained by taking the difference $Mx - y$, where x is a random n-vector with the standard normal distribution, and y is a random n-vector in the interior of the underlying dual cone. This ensures that Slater's condition is satisfied, and hence there is at least one solution. The random vector y is obtained by first generating a random matrix with the standard normal distribution for the $\bar{y}$-component, followed by adding the absolute value of a standard normal random variable to the maximum or sum of the singular values of the $\bar{y}$-component.

We run the MATLAB implementation of Algorithm 4.2 in MATLAB version 7.13.0.564 on a machine with Intel Q6600 CPU at 2.40 GHz, with 3GB of RAM, and running Windows 7 Enterprise (64-bit). We use the following values for the parameters: $w_0 = (0, 0)$, $\alpha = 0.5$, $\eta = \bar{\eta} = 0.25$, $\sigma = 0.4$, $\kappa = 0.01$, $\beta = \|H^{\text{nat}}(w_0)\|$. The Newton system in Step 1a is solved directly with the MATLAB backslash operator. We terminate the algorithm when

$$\text{relative error} := \max\left\{\frac{\|x_k - \Pi_K(x_k - y_k)\|}{\|x_k\| + 1}, \frac{\|F(x_k) - y_k\|}{\|y_k\| + 1}\right\} \leq 10^{-6}.$$

Table 1 shows the results for $K = K_{m,n}$ and $K = K_{m,n}^\sharp$ for the choices $(m, n) = (20, 25), (25, 30), (30, 35)$ with corresponding $N = 501, 751, 1051$. For each choice of K, we generate problems of 9 difference ranks r, evenly distributed between about 10% and about 90% of N. For each size and rank, we generated 100 random instances and report the averages over these 100 random instances.

We see from Table 1 that on average, the algorithm takes no more than 15 iterations and less than 24 subiterations (i.e., total number of times that the Newton system is solved). There is no observable difference in the average number of main and subiterations among the different ranks for $K = K_{m,n}^\sharp$; but there is a noticeable increase in the average number of both main and subiterations with rank for $K = K_{m,n}$. From our experiments with other choices of parameters, we find that this phenomenon depends not only on the structure of the cone, but also on the choice of the parameters, such as η and $\bar{\eta}$.

Table 2 shows the results for mixed cones $K = K_{m,n} \times K_{m,n}^\sharp$ for the choices $(m, n) = (15, 15), (15, 25), (20, 25)$ with corresponding $N = 452, 752, 1002$. As before, for each choice of K, we generate problems of 9 difference ranks r, evenly distributed between about 10% and about 90% of N, and report, for each size and rank, the averages over 100 random instances.

We see from Table 2 that on average, the algorithm takes no more than 15 iterations and fewer than 29 subiterations. Just like the case $K = K_{m,n}$, there is a noticeable increase in the average number of both main and subiterations with rank.

TABLE 1
Average relative errors, numbers of iterations, and running times for various cones and ranks.

$K = K_{20,25}$ $(N = 501)$

r	Relative error	# of iter. main	# of iter. sub	Time (sec)
50	8.51E−8	12.86	18.78	2.2
100	3.43E−8	12.99	19.35	2.3
150	6.04E−8	13.00	19.09	2.3
200	1.60E−7	13.04	18.66	2.2
250	2.59E−7	13.13	19.14	2.3
300	2.80E−7	13.40	19.88	2.4
350	1.79E−7	13.77	20.98	2.5
400	2.34E−8	13.97	21.55	2.6
450	2.90E−8	13.98	22.45	2.7

$K = K_{25,30}$ $(N = 751)$

r	Relative error	# of iter. main	# of iter. sub	Time (sec)
75	2.49E−7	13.36	19.92	5.8
150	5.99E−8	13.88	20.37	5.9
225	1.44E−9	14.00	20.77	6.0
300	4.49E−9	14.00	20.64	6.0
375	1.01E−8	14.00	20.56	6.0
450	3.84E−8	14.00	20.94	6.1
525	1.60E−7	14.02	21.28	6.2
600	3.83E−7	14.04	21.67	6.3
675	4.06E−7	14.31	22.87	6.6

$K = K_{30,35}$ $(N = 1051)$

r	Relative error	# of iter. main	# of iter. sub	Time (sec)
105	2.23E−8	14.01	21.13	13.6
210	9.95E−8	14.01	21.73	14.1
315	2.60E−7	14.02	21.01	13.6
420	3.96E−7	14.08	20.40	13.2
525	4.19E−7	14.40	22.13	14.2
630	1.06E−7	14.87	22.80	14.7
735	9.92E−9	14.99	23.59	15.2
840	6.31E−9	15.00	23.77	15.3
945	7.60E−8	15.00	24.49	15.7

$K = K_{20,25}^{\sharp}$ $(N = 501)$

r	Relative error	# of iter. main	# of iter. sub	Time (sec)
50	1.38E−7	13.36	20.82	2.5
100	2.04E−7	12.21	18.37	2.2
150	1.14E−7	12.00	17.73	2.1
200	6.91E−8	12.02	17.99	2.2
250	9.47E−8	11.93	18.02	2.2
300	8.60E−8	11.97	17.97	2.1
350	7.30E−8	12.01	17.70	2.1
400	1.16E−7	11.99	17.81	2.1
450	8.88E−8	11.96	18.14	2.2

$K = K_{25,30}^{\sharp}$ $(N = 751)$

r	Relative error	# of iter. main	# of iter. sub	Time (sec)
75	1.69E−7	13.77	21.46	6.2
150	1.76E−7	12.70	19.28	5.6
225	3.18E−7	12.27	19.18	5.6
300	2.51E−7	12.16	18.79	5.5
375	2.03E−7	12.11	18.87	5.5
450	2.11E−7	12.15	19.08	5.5
525	2.13E−7	12.13	19.24	5.6
600	1.94E−7	12.11	18.63	5.4
675	1.93E−7	12.19	19.08	5.6

$K = K_{30,35}^{\sharp}$ $(N = 1051)$

r	Relative error	# of iter. main	# of iter. sub	Time (sec)
105	1.25E−7	13.88	21.52	13.8
210	1.17E−7	13.02	20.66	13.2
315	5.41E−8	12.99	20.27	13.1
420	1.32E−7	12.85	19.88	12.8
525	1.10E−7	12.91	19.81	12.8
630	1.84E−7	12.78	20.78	13.5
735	1.41E−7	12.81	21.96	14.1
840	1.49E−7	12.82	22.23	14.3
945	1.26E−7	12.86	21.91	14.0

Moreover, when compared to the single cone problems of similar size N, the algorithm requires less time to solve these problems, even when more subiterations are required. This is possibly explained by the shorter time needed to compute the Jacobians of the smoothing approximations for cones of smaller size.

Table 3 shows the results for various cones of larger size. This time, due to much longer running times, we generate and report, for each type of cone, the averages over 50 random instances of various randomly selected ranks between about 10% and about 90% of N.

We see from Table 3 that on average, the algorithm takes about 13 to 16 iterations and fewer than 25 subiterations for the single cones, but more subiterations for the mixed cones. This is consistent with our experiments reported in Tables 1 and 2. As before, when comparing the mixed cone problems to the single cone problems of

TABLE 2
Average relative errors, numbers of iterations, and running times for various mixed cones and ranks.

$K = K_{15,15} \times K_{15,15}^{\sharp}$ $(N = 452)$

	Relative	# of iter.		Time
r	error	main	sub	(sec)
45	1.33E−7	12.67	19.93	1.2
90	3.74E−8	12.99	19.41	1.1
135	1.29E−7	13.03	19.37	1.1
180	2.25E−7	13.63	20.85	1.2
225	5.18E−8	14.00	23.25	1.4
270	2.37E−7	13.14	21.72	1.3
315	6.15E−8	13.00	21.90	1.3
360	2.78E−8	13.06	22.25	1.3
405	2.84E−8	13.04	22.68	1.4

$K = K_{15,25} \times K_{15,25}^{\sharp}$ $(N = 752)$

	Relative	# of iter.		Time
r	error	main	sub	(sec)
75	1.77E−7	13.36	20.40	3.4
150	1.76E−8	13.97	20.85	3.5
225	2.87E−8	14.00	22.56	3.8
300	3.71E−7	14.24	22.25	3.7
375	7.14E−8	14.97	25.96	4.4
450	1.08E−7	14.01	24.40	4.1
525	4.79E−8	13.97	24.67	4.2
600	1.83E−7	13.76	25.47	4.3
675	3.91E−7	13.44	26.84	4.6

$K = K_{20,25} \times K_{20,25}^{\sharp}$ $(N = 1002)$

	Relative	# of iter.		Time
r	error	main	sub	(sec)
100	2.06E−8	14.00	21.91	7.5
200	1.11E−7	14.01	22.11	7.6
300	3.09E−7	14.54	23.60	8.1
400	2.00E−8	15.00	24.27	8.3
500	2.89E−7	15.29	28.64	9.8
600	3.67E−7	14.29	27.05	9.3
700	1.27E−7	14.02	27.30	9.4
800	8.33E−8	14.06	28.43	9.8
900	6.50E−8	14.05	27.99	9.6

TABLE 3
Average relative errors, numbers of iterations, and running times for various larger cones.

		Relative	# of iter.		Time
K	N	error	main	sub	(sec)
$K_{40,50}$	2001	1.58E−7	15.42	23.88	71.7
$K_{50,60}$	3001	1.87E−7	16.04	24.34	205.2
$K_{40,50}^{\sharp}$	2001	3.75E−7	13.30	22.68	67.2
$K_{50,60}^{\sharp}$	3001	1.12E−7	14.00	24.88	207.8
$K_{30,35} \times K_{30,35}^{\sharp}$	2102	2.13E−7	15.44	31.50	64.4
$K_{35,45} \times K_{35,45}^{\sharp}$	3152	2.13E−7	16.16	32.86	183.0

similar size N, the algorithm requires less time to solve the mixed cone problems, even when more subiterations are required.

7. Conclusion. In this paper, we extend the combined smoothing and regularization method of Hayashi, Yamashita, and Fukushima [16] via a barrier-based smoothing approach to solve general variational inequalities over convex sets using barriers with definable gradient maps, under assumptions of uniform nonsingularity of the Newton system, and semismoothness of F. We further demonstrate the general applicability of the extension by showing that the universal barrier has a definable gradient map when the underlying set is definable in the o-minimal structure $\mathbb{R}_{\text{an}}^{\mathbb{R}_{\text{alg}}}$. On the practical side, we study the application of the barrier-based smoothing approach to complementarity problems over epigraphs of matrix operator and nuclear norms. We prove that the uniform nonsingularity assumption is satisfied when the

solution is nondegenerate. We also provide preliminary numerical results as evidence of the practical efficiency of the approach.

We would like to end with some open questions.

1. A conjecture by Kaiser [17] on the compatibility of the Lebesgue measure for o-minimal structures in general, if proven true, would imply the general applicability of our approach to any definable convex sets.

2. The sufficient condition for uniform nonsingularity of the Newton system stated in Theorem 4.4 is a consequence for nondegeneracy when the underlying set belongs to various classes of cones, including symmetric cones and the epigraphs of matrix operator norm and nuclear norm. However, we think that nondegeneracy is not sufficient in general. An open question is to give a sufficient condition that is independent of the choice of the barrier used in the smoothing approximation.

Appendix A. Verification of (6.13) in Lemma 6.1. We first verify the coefficients of the $\acute{h}_{ij}$'s in (6.13). For any $1 \leq i \leq j \leq m$, it follows from (6.7) and (6.9) that as $k \to \infty$,

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}\sigma_{j,k})} \to \frac{((x_0^*)^2 - (\sigma_i^*)^2)((x_0^*)^2 - (\sigma_j^*)^2)}{((x_0^*)^2 - (\sigma_i^*)^2)((x_0^*)^2 - (\sigma_j^*)^2) + 0} = 1$$

if $r < i \leq j$,

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}\sigma_{j,k})} = \frac{\sigma_{i,k}(x_{0,k}^2 - \sigma_{j,k}^2)}{\sigma_{i,k}(x_{0,k}^2 - \sigma_{j,k}^2) - \tau_{i,k}(x_{0,k}^2 + \sigma_{i,k}\sigma_{j,k})}$$

$$\to \begin{cases} \frac{x_0^*((x_0^*)^2 - (\sigma_j^*)^2)}{x_0^*((x_0^*)^2 - (\sigma_j^*)^2) + 0} = 1 & \text{if } r^\sharp < i \leq r < j, \\ \frac{x_0^*((x_0^*)^2 - (\sigma_j^*)^2)}{x_0^*((x_0^*)^2 - (\sigma_j^*)^2) - \tau_i^* x_0^*(x_0^* + \sigma_j^*)} =: 1 - \alpha_{ij} \in (0,1) & \text{if } i \leq r^\sharp \leq r < j, \\ \frac{x_0^* \times 0}{x_0^* \times 0 - 2\tau_i^*(x_0^*)^2} = 0 & \text{if } i \leq r^\sharp \text{ and } j \leq r, \end{cases}$$

and

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}\sigma_{j,k})} = \frac{\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{2\mu_k^2}}{\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{2\mu_k^2} + (x_{0,k}^2 + \sigma_{i,k}\sigma_{j,k})}$$

$$\to \frac{\lim_{k\to\infty} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{2\mu_k^2}}{\lim_{k\to\infty} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{2\mu_k^2} + 2(x_0^*)^2} =: 1 - \alpha_{ij} \in [0,1]$$

if $r^\sharp < i \leq j \leq r$. Here we have assumed, without loss of generality, that the sequence $\{\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{2\mu_k^2}\}$ converges to a limit in $[0, \infty]$.

Similarly in verifying the coefficients of the $\hat{h}_{ij}$'s, for every $1 \leq i \leq j \leq m$,

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 - \sigma_{i,k}\sigma_{j,k})} \to 1$$

if $r < i \leq j$,

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 - \sigma_{i,k}\sigma_{j,k})} = \frac{\sigma_{i,k}(x_{0,k}^2 - \sigma_{j,k}^2)}{\sigma_{i,k}(x_{0,k}^2 - \sigma_{j,k}^2) - \tau_{i,k}(x_{0,k}^2 - \sigma_{i,k}\sigma_{j,k})}$$

$$\to \frac{x_0^*((x_0^*)^2 - (\sigma_j^*)^2)}{x_0^*((x_0^*)^2 - (\sigma_j^*)^2) - \tau_i^* x_0^*(x_0^* - \sigma_j^*)} = \frac{x_0^* + \sigma_j^*}{x_0^* + \sigma_j^* - \tau_i^*} \begin{cases} = 1 & \text{if } r^\sharp < i \leq r < j, \\ =: \beta_{ij} \in (0,1) & \text{if } i \leq r^\sharp \leq r < j, \end{cases}$$

and

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + 2\mu_k^2(x_{0,k}^2 - \sigma_{i,k}\sigma_{j,k})}$$

$$= \frac{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{j,k}^2) + \mu_k^2(x_{0,k} - \sigma_{i,k})(x_{0,k} + \sigma_{j,k}) + \mu_k^2(x_{0,k} + \sigma_{i,k})(x_{0,k} - \sigma_{j,k})}$$

$$= \left(1 + \frac{-\tau_{j,k}(x_{0,k} + \sigma_{j,k})}{2\sigma_{j,k}(x_{0,k} + \sigma_{i,k})} + \frac{-\tau_{i,k}(x_{0,k} + \sigma_{i,k})}{2\sigma_{i,k}(x_{0,k} + \sigma_{j,k})}\right)^{-1}$$

$$\to \left(1 + \frac{-2(\tau_i^* + \tau_j^*)x_0^*}{4(x_0^*)^2}\right)^{-1} = \frac{2x_0^*}{2x_0^* - \tau_i^* - \tau_j^*} \begin{cases} = 1 & \text{if } r^\sharp < i \le j \le r, \\ =: \beta_{ij} \in (0,1) & \text{if } i \le r^\sharp \text{ and } j \le r. \end{cases}$$

Next we verify the coefficients of the $\tilde{h}_{ij}$'s. For $1 \le i \le m$,

$$\frac{(x_{0,k}^2 - \sigma_{i,k}^2)}{x_{0,k}^2 - \sigma_{i,k}^2 + 2\mu_k^2} \to \begin{cases} \frac{(x_0^*)^2 - (\sigma_i^*)^2}{(x_0^*)^2 - (\sigma_i^*)^2 + 0} = 1 & \text{when } r < i, \\ \lim_{k\to\infty} \frac{\sigma_{i,k}}{\sigma_{i,k} - \tau_{i,k}} = \frac{x_0^*}{x_0^* - \tau_i^*} = 1 & \text{when } r^\sharp < i \le r, \\ \lim_{k\to\infty} \frac{\sigma_{i,k}}{\sigma_{i,k} - \tau_{i,k}} = \frac{x_0^*}{x_0^* - \tau_i^*} =: \gamma_i \in (0,1) & \text{when } i \le r^\sharp. \end{cases}$$

Similarly to the coefficients of the $\acute{h}_{ij}$'s, the coefficient of each $\grave{h}_i$ in the second and third terms of (6.13) is

$$\begin{cases} 1 & \text{if } r < i, \\ 1 - \alpha_{ii} \in [0,1] & \text{if } r^\sharp < i \le r, \\ 0 & \text{if } i \le r^\sharp. \end{cases}$$

Finally we verify the coefficients of h_0 and the $\grave{h}_i$'s in the first term of (6.13). First consider the limit of $C_k^{-1}b_k$. The ith entry of $-C_k^{-1}b_k$ is

$$\frac{4\mu_k^2 x_{0,k}\sigma_{i,k}}{(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)} = \frac{-2\tau_{i,k}x_{0,k}}{(x_{0,k}^2 - \sigma_{i,k}^2) - \frac{\tau_{i,k}}{\sigma_{i,k}}(x_{0,k}^2 + \sigma_{i,k}^2)} \to \frac{-2\tau_i^* x_0^*}{0 - \frac{\tau_i^*}{x_0^*} \times 2(x_0^*)^2} = 1$$

if $i \le r^\sharp$,

$$\frac{4\mu_k^2 x_{0,k}\sigma_{i,k}}{(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)} = \frac{2x_{0,k}\sigma_{i,k}}{\frac{(x_{0,k}^2 - \sigma_{i,k}^2)^2}{2\mu_k^2} + (x_{0,k}^2 + \sigma_{i,k}^2)}$$

$$\to \frac{2(x_0^*)^2}{\lim_{k\to\infty} \frac{(x_{0,k}^2 - \sigma_{i,k}^2)^2}{2\mu_k^2} + 2(x_0^*)^2} = \alpha_{ii} \in [0,1]$$

if $r^\sharp < i \le r$, and

$$\frac{4\mu_k^2 x_{0,k}\sigma_{i,k}}{(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)} \to \frac{0}{((x_0^*)^2 - (\sigma_i^*)^2)^2 + 0} = 0$$

if $r < i$. Thus $-C_k^{-1}b_k \to (1,\ldots,1,\alpha_{r^\sharp+1},\ldots,\alpha_r,0,\ldots,0)$, where there are $r^\sharp$ ones

and $(m - r)$ zeros. Next we consider the limit of

$$
a_k - b_k^T C_k^{-1} b_k
$$
$$
= 1 + 2\mu_k^2 \sum_{i=1}^{m} \frac{(x_{0,k}^2 + \sigma_{i,k}^2)((x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)) - 8\mu_k^2 x_{0,k}^2 \sigma_{i,k}^2}{(x_{0,k}^2 - \sigma_{i,k}^2)^2((x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2))} + \mu_k^2 \frac{n - m}{x_{0,k}^2}
$$
$$
= 1 + 2\mu_k^2 \sum_{i=1}^{m} \frac{(x_{0,k}^2 + \sigma_{i,k}^2)(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 - \sigma_{i,k}^2)^2}{(x_{0,k}^2 - \sigma_{i,k}^2)^2((x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2))} + \mu_k^2 \frac{n - m}{x_{0,k}^2}
$$
$$
= 1 + \sum_{i=1}^{m} \frac{2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2 + 2\mu_k^2)}{(x_{0,k}^2 - \sigma_{i,k}^2)^2 + 2\mu_k^2(x_{0,k}^2 + \sigma_{i,k}^2)} + \mu_k^2 \frac{n - m}{x_{0,k}^2}.
$$

In a similar fashion as before, it follows from (6.7) and (6.9) that

$$
\lim_{k \to \infty} a_k - b_k^T C_k^{-1} b_k = 1 + \sum_{i=1}^{r^\sharp} 1 + \sum_{i=r^\sharp+1}^{r} \alpha_{ii} + \sum_{i=r+1}^{m} 0 = 1 + r^\sharp + \sum_{i=r^\sharp+1}^{r} \alpha_{ii}.
$$

REFERENCES

[1] H. H. BAUSCHKE, O. GÜLER, A. S. LEWIS, AND H. S. SENDOV, *Hyperbolic Polynomials and Convex Analysis*, Research Report CORR 98-29, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, 1998.
[2] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.
[3] J. BOCHNAK, M. COSTE, AND M.-F. ROY, *Real Algebraic Geometry*, Springer, Berlin, Heidelberg, 1998.
[4] J. BOLTE, A. DANIILIDIS, AND A. LEWIS, *Tame functions are semismooth*, Math. Program., 117 (2009), pp. 5–19.
[5] X. D. CHEN, D. SUN, AND J. SUN, *Complementarity functions and numerical experiments on some smoothing Newton methods for second-order-cone complementarity problems*, Comput. Optim. Appl., 25 (2003), pp. 39–56.
[6] C. B. CHUA, *A T-algebraic approach to primal-dual interior-point algorithms*, SIAM J. Optim., 20 (2009), pp. 503–523.
[7] C. B. CHUA AND Z. LI, *A barrier-based smoothing proximal point algorithm for NCPs over closed convex cones*, SIAM J. Optim., 23 (2013), pp. 745–769.
[8] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[9] M. COSTÉ, *An Introduction to O-Minimal Geometry*, RAAG Notes, Institut de Recherche Mathématiques de Rennes, Rennes, France, 1999.
[10] E. M. DE SÁ, *Faces of the unit ball of a unitarily invariant norm*, Linear Algebra Appl., 197/198 (1994), pp. 451–493.
[11] C. DING, D. SUN, AND K.-C. TOH, *An introduction to a class of matrix cone programming*, Math. Program., 144 (2014), pp. 141–179.
[12] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume* I, Springer-Verlag, New York, Berlin, Heidelberg, 2003.
[13] M. FĂZEL, H. HINDI, AND S. P. BOYD, *A rank minimization heuristic with application to minimum order system approximation*, in Proceedings of the American Control Conference, IEEE, Washington, DC, 2001, pp. 4734–4739.
[14] A. FISCHER, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Program., 76 (1997), pp. 513–532.
[15] M. FUKUSHIMA, L. QI, AND D. DUN, *A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J.-S. Pang, eds., Proceedings in Applied Mathematics, SIAM, Philadelphia, 1997, pp. 452–473.

[16] S. Hayashi, N. Yamashita, and M. Fukushima, *A combined smoothing and regularization method for monotone second-order cone complementarity problems*, SIAM J. Optim., 15 (2005), pp. 593–615.

[17] T. Kaiser, *First order tameness of measures*, Ann. Pure Appl. Logic, 163 (2012), pp. 1903–1927.

[18] C. Kanzow and H. Pieper, *Jacobian smoothing methods for nonlinear complementarity problems*, SIAM J. Optim., 9 (1999), pp. 342–373.

[19] M. Kojima, Y. Kurita, and S. Mizuno, *Large-step interior point algorithms for linear complementarity problems*, SIAM J. Optim., 3 (1993), pp. 398–412.

[20] Z. Liu and L. Vandenberghe, *Interior-point method for nuclear norm approximation with application to system identification*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1235–1256.

[21] Z. Q. Lou and P. Tseng, *A new class of merit functions for the nonlinear complementarity problem*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J.-S. Pang, eds., Proceedings in Applied Mathematics, SIAM, Philadelphia, 1997, pp. 204–225.

[22] O. L. Mangasarian, *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Appl. Math., 31 (1976), pp. 89–92.

[23] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.

[24] C. Miller, *Expansions of the real field with power functions*, Ann. Pure Appl. Logic, 68 (1994), pp. 79–94.

[25] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, 1994.

[26] G. Pataki and L. Tunçel, *On the generic properties of convex optimization problems in conic form*, Math. Program., 89 (2001), pp. 449–457.

[27] L. Qi and D. Sun, *Solving variational inequality problems via smoothing-nonsmooth reformulations*, J. Comput. Appl. Math., 129 (2001), pp. 37–62.

[28] L. Qi and D. Sun, *Smoothing functions and smoothing Newton method for complementarity and variational inequality problems*, J. Optim. Theory Appl., 113 (2002), pp. 121–147.

[29] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[30] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.

[31] J. Renegar, *Hyperbolic programs, and their derivative relaxations*, Found. Comput. Math., 6 (2006), pp. 59–79.

[32] C. Steinhorn, *Tame topology and o-minimal structures*, in Computational Aspects of General Equilibrium Theory, Lecture Notes in Econom. and Math. Systems 604, Springer, Berlin, Heidelberg, 2008, pp. 165–191.

[33] D. Sun, *A further result on an implicit function theorem for locally Lipschitz functions*, Oper. Res. Lett., 28 (2001), pp. 193–198.

[34] D. Sun and J. Sun, *Semismooth matrix-valued functions*, Math. Oper. Res., 27 (2002), pp. 150–169.

[35] M. Tao and X. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57–81.

[36] L. van den Dries, *A generalization of the Tarski-Seidenberg theorem, and some nondefinability results*, Bull. Amer. Math. Soc. (N.S.), 15 (1986), pp. 189–193.

[37] L. van den Dries, *Tame Topology and O-minimal Structures*, Cambridge University Press, Cambridge, UK, 1998.

[38] L. van den Dries, A. Macintyre, and D. Marker, *The elementary theory of restricted analytic functions with exponentiation*, Ann. of Math. (2), 140 (1994), pp. 183–205.

[39] L. van den Dries and C. Miller, *On the real exponential field with restricted analytic functions*, Israel J. Math., 85 (1994), pp. 19–56.

[40] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J., 84 (1996), pp. 497–540.

[41] È. B. Vinberg, *The theory of convex homogeneous cones*, Trans. Moscow Math. Soc., 12 (1963), pp. 340–403.

[42] S. J. Wright, *An infeasible-interior-point algorithm for linear complementarity problems*, Math. Program., 67 (1994), pp. 29–51.