Solving Nonlinear Absolute Value Equations

Aris Daniilidis, Mounir Haddou, Trí Minh Lê, Olivier Ley

Abstract. In this work, we show that several problems, that are naturally represented as Nonlinear Absolute Value Equation problems (in short, NAVE), can be reformulated as Nonlinear Complementarity Problems (in short, NCP) and, under mild conditions, they can be efficiently solved using smoothing regularizing techniques. To the best of our knowledge, this is the first numerical approach that deals directly with Nonlinear Absolute Value Equations. We also identify a commonly used technical assumption in smoothing techniques and prove its equivalence to a classical Łojasiewicz inequality at infinity, confirming in particular that this assumption is not restrictive. The effectiveness of our approach is illustrated in several problems, including asymmetric ridge optimization and nonlinear ordinary differential equations.

Keywords: Nonlinear Absolute Value Equality, Complementarity problem, P_0-map, Numerical methods

AMS Classification: *Primary*: 90C33, 65K10 ; *Secondary*: 15B48, 65K15, 90C59.

1 Introduction

The last two decades, absolute value equation problems (in short, AVE problems) have been extensively studied in the literature. This interest is justified by the fact that this class of problems already covers a wide spectrum of applications: indeed, numerous problems stemming from real-life applications, as for instance all mixed integer linear programming problems, can be reformulated as AVE problems. It is also well-known, see [24, 32] *e.g.*, that AVE problems admit an equivalent description as Linear Complementarity problems (in short, LCP). The exact definitions of an AVE problem and a LCP are recalled below. Dealing efficiently with these problems is thus paramount.

The literature on this subject contains several theoretical results for existence as well as conditions guaranteeing uniqueness of the solution [22, 23, 34, 35]. Concurrently, there are also various numerical approaches to solve an AVE problem. Generally speaking, these methods can be divided into at least three categories [3] : iterative linear algebra based methods (also known as projective methods), semi-smooth Newton-like methods and smoothing methods. The aforementioned methods generally require some assumption on the matrix involved in the AVE problem. In particular, the classes of P_0-matrices and P-matrices (recalled below) turn out to be particularly relevant in this study [1].

In this work, we consider a natural generalization of (linear) AVE problems to nonlinear ones, known as Nonlinear Absolute Value Equations (in short, NAVE). This more general framework encompasses new applications including ridge regression models, bounded constrained nonlinear systems of equations, and stiff Ordinary Differential Equations (in short, stiff ODE). This approach to deal with the aforementioned problems, based on NAVE, is to the best of our knowledge, completely new in the literature.

Our main contribution is the following: we first show that similarly to the way that an AVE problem is linked to a LCP, nonlinear absolute value equations can also be associated to nonlinear complementarity problems (in short, NCP). Indeed, any NCP can be reformulated as NAVE. The converse is also true, but the association is generally given in an implicit way. Then, taking profit from the huge literature concerning existence, uniqueness and numerical resolution of NCP (see [9,23,35,37] *e.g.*), we propose a new method to solve a NAVE problem, based on the smoothing technique proposed in [10] and further developments in the follow-up work [30]. The proposed approach is explained in Section 2, while in Section 3 we discuss applications.

To ease the reading we start with some definitions and settings. The *Absolute Value Equality problem* (in short, AVE) is defined as follows:

$$\text{find } x \in \mathbb{R}^d : \qquad Ax - |x| = b, \tag{AVE}$$

where A is a $(d \times d)$-matrix and $b \in \mathbb{R}^d$. Throughout this work, given $x = (x^1, \ldots, x^d)^T \in \mathbb{R}^d$, we use the notation $|x| := (|x^1|, \ldots, |x^d|)^T$ componentwise to denote a vector in $\mathbb{R}^d_+$.

Denoting by I the identity matrix of $\mathbb{R}^d$ and assuming that either $A - I$ or $A + I$ is invertible, the (AVE) problem can be transformed to a *Linear Complementarity Problem* (in short, LCP). Indeed, setting (coordinate by coordinate)

$$\begin{cases} y = x^+ = \max\{\ \ x, 0\} \\ z = x^- = \max\{-x, 0\} \end{cases} \tag{1.1}$$

and performing the transformation $x = y - z$ and $|x| = y + z$ we obtain:

$$(A - I)\, y - (A + I)\, z = b.$$

Therefore for

$$\begin{cases} M := (A + I)^{-1}(A - I) \\ q := (A + I)^{-1}(-b) \end{cases} \quad \text{or respectively} \quad \begin{cases} \widetilde{M} := (A - I)^{-1}(A + I) \\ \widetilde{q} := (A - I)^{-1} b \end{cases} \tag{1.2}$$

we obtain

$$\begin{cases} z = My + q \\ 0 \le y \perp z \ge 0 \end{cases} \quad \text{or respectively} \quad \begin{cases} y = \widetilde{M} z + \widetilde{q} \\ 0 \le y \perp z \ge 0 \end{cases} \tag{LCP}$$

The above problem can be solved provided M (respectively $\widetilde{M}$) is a P-matrix (see below for details). It is important to notice that this property can be traced back to the matrix A; in particular, the property is ensured whenever the singular values of A are all greater than 1. Notice that this condition guarantees invertibility of both $A - I$ and $A + I$.

Solving (LCP) under the assumption that M (respectively $\widetilde{M}$) is a P-matrix has been treated in several works (see [4]). In this case, it can be shown that the problem has a unique solution $(\bar{y}, \bar{z})$ yielding that $\bar{x} := \bar{y} - \bar{z}$ is the (unique) solution of (AVE). Moreover, this solution can be obtained numerically, via smoothing regularization techniques (see [1,10,30] and Section 2.3 below).

In this work, we propose a new method of solving a nonlinear generalization of (AVE), namely the following *Nonlinear Absolute Value Equality* problem (in short, NAVE)

$$\text{Find } x \in \mathbb{R}^d : \qquad F(x) - |x| = 0 \quad \text{(coordinatewise)}, \tag{NAVE}$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a (nonlinear) mapping. By introducing new variables $y = x^+$ and $z = x^-$ (*cf.* (1.1)) so that $x = y - z$ and $|x| = y + z$, (NAVE) becomes

$$F(y - z) - (y + z) = 0.$$

By setting $z = H(y)$ (which is possible under regularity assumptions on F, see Lemma 2.5), we can transform (NAVE) to a *Nonlinear Complementarity Problem* (NCP):

$$\begin{cases} H(y) = F(y - H(y)) - y \\ 0 \le y \perp H(y) \ge 0. \end{cases} \tag{1.3}$$

As we shall see in Section 2.3, even though the function H is only defined implicitly, it is still possible to solve (1.3) numerically provided we are able to guarantee that H is a P_0-map (see Definition 2.2), a notion which generalizes P_0–matrices (*c.f.* Lemma 2.7).

This is the first work that directly addresses nonlinear absolute value equations, besides the fact that a variety of real problems and applications can naturally assume this form.

Another important contribution related to smoothing techniques for NCP is that we clearly characterize a technical assumption commonly employed in functions–smoothing and prove its equivalence with the classical Łojasiewicz inequality at infinity (see Theorem 2.8). This clearly reveals that the former assumption is not at all restrictive.

2 Setting of the problem and description of the method

2.1 Definitions and preliminaries

Given a $(d \times d)$ matrix A and $I \subset \{1, 2, \cdots, d\}$, we denote by A_{II} the submatrix made up of the rows and columns of I.

Definition 2.1 (P_0-matrix and P-matrix)**.** A matrix A is called a P_0–matrix (respectively, P–matrix) if one of the following equivalent properties holds

(i) for every $I \subset \{1, 2, \cdots, d\}$, $\det(A_{II}) \ge 0$ (respectively $\det(A_{II}) > 0$);

(ii) for every $x = (x_1, \cdots, x_d)^T \in \mathbb{R}^d$, $x \neq 0$,

$$\max_{1 \le j \le d} (Ax)^j x^j \ge 0 \qquad \left(\text{respectively } \max_{1 \le j \le d} (Ax)^j x^j > 0\right);$$

(iii) for every $I \subset \{1, 2, \cdots, d\}$, the real eigenvalues of A_{II} are nonnegative (respectively strictly positive).

The notion of P-matrix can be generalized to general nonlinear maps $H : \mathbb{R}^d \to \mathbb{R}^d$ as follows:

Definition 2.2 (P_0–map and P–map)**.** A mapping $H : \mathbb{R}^d \to \mathbb{R}^d$ is called P_0-map (respectively, P-map), if for every $x, y \in \mathbb{R}^d$, $x \neq y$, it holds:

$$\max_{j \in \{1,\ldots,d\}} \left(H(y)^j - H(x)^j\right)(y^j - x^j) \ge 0 \qquad \left(\text{respectively, } \max_{j \in \{1,\ldots,d\}} \left(H(y)^j - H(x)^j\right)(y^j - x^j) > 0\right);$$

If H is of the form $H(x) = Ax + b$ for some $(d \times d)$ matrix A and vector $b \in \mathbb{R}^d$, then it follows directly that H is a P_0–map (respectively, a P–map) if and only if A is a P_0–matrix (respectively, a P–matrix). More generally, we have the following result:

Lemma 2.3 ([27, Corollary 5.3, Theorem 5.8])**.** *Let $H : \mathbb{R}^d \to \mathbb{R}^d$ be C^1. Then H is a P_0-map if and only if, for every $x \in \mathbb{R}^d$, $\nabla H(x)$ is a P_0-matrix.*

We refer the reader to [8,27] for further results about P_0– and P–matrices and maps. We finish this section with the following useful lemma.

Lemma 2.4 (A characterization of P_0–matrices)**.** *Let A be a $(d \times d)$ matrix. Then A is a P_0-matrix if and only if, for every diagonal matrix Δ_1 with strictly positive entries and for every nonnegative diagonal matrix Δ_2, the matrix $\Delta_1 + \Delta_2 A$ is invertible.*

Proof. Let A be a P_0-matrix. Then for every diagonal matrix Δ_2 with nonnegative entries, the matrix $\Delta_2 A$ is also P_0, while for every diagonal matrix Δ_1 with strictly positive entries, the matrix $\Delta_1 + \Delta_2 A$ is a P-matrix, therefore, in particular, it is invertible.
Conversely, let us assume that A is not a P_0–matrix. Then there exists $v \in \mathbb{R}^d$, $v \neq 0$ such that

$$(Av)^i v^i < 0, \qquad \text{for every } i \in \{1, \cdots, d\}. \tag{2.1}$$

Let $\Delta_1 = \mathrm{diag}(\delta_1^1, \cdots, \delta_1^d)$ and $\Delta_2 = \mathrm{diag}(1, \cdots, 1)$. Then

$$(\Delta_1 + \Delta_2 A)v = (\delta_1^1 v^1 + (Av)^1, \cdots, \delta_1^d v^d + (Av)^d)^T,$$

and by setting, for every i, $\delta_1^i := -(Av)^i / v^i$ (which is well-defined and strictly positive thanks to (2.1)) we deduce that $(\Delta_1 + \Delta_2 A)v = 0$ and therefore $\Delta_1 + \Delta_2 A$ is not invertible. □

2.2 Transforming a (NAVE) problem to a (NCP) problem

Given a nonlinear mapping $F : \mathbb{R}^d \to \mathbb{R}^d$, we consider the (NAVE) problem

$$\text{Find } x \in \mathbb{R}^d : \qquad F(x) - |x| = 0.$$

By introducing new variables $y = x^+$ and $z = x^-$ so that $x = y - z$, $|x| = y + z$, $y \perp z$, the (NAVE) problem becomes

$$F(y - z) - (y + z) = 0. \tag{2.2}$$

The following lemma gives conditions under which (2.2) may be written as a (NCP) problem by setting $y = H(z)$ or $z = \widetilde{H}(y)$ for some suitable maps H or $\widetilde{H}$.

Lemma 2.5. *Assume that the mapping F is C^1 in a neighborhood of the point $x^* = y^* - z^* \in \mathbb{R}^d$ which is assumed to be solution of (2.2). Then it holds:*

(i). If $F - I$ is a P_0-map, then there exists a C^1 map $H : \mathbb{R}^d \to \mathbb{R}^d$ defined in a neighborhood of y^* such that $z^* = H(y^*)$ and y^* is a solution to the following (NCP) problem:

$$\begin{cases} H(y) = F(y - H(y)) - y \\ 0 \leq y \perp H(y) \geq 0. \end{cases} \tag{2.3}$$

(ii). If $-(F+I)$ is a P_0-map, then there exists a C^1 map $\widetilde{H} : \mathbb{R}^d \to \mathbb{R}^d$ defined in a neighborhood of z^* such that $y^* = \widetilde{H}(z^*)$ and z^* is a solution to the following (NCP) problem:

$$\begin{cases} \widetilde{H}(z) = F(\widetilde{H}(z) - z) - z \\ 0 \leq z \perp \widetilde{H}(z) \geq 0. \end{cases} \tag{2.4}$$

Proof. We consider the C^1 map $\mathcal{F} : \mathbb{R}^{2d} \to \mathbb{R}$ defined by

$$\mathcal{F}(y,z) = F(y-z) - (y+z).$$

We are going to apply the implicit function theorem at the point $(y^*, z^*) \in \mathbb{R}^{2d}$. Notice that

$$(\nabla_y \mathcal{F}(y^*, z^*), \nabla_z \mathcal{F}(y^*, z^*)) = (\nabla F(y^* - z^*) - I, -\nabla F(y^* - z^*) - I).$$

If $F - I$ is a P_0-map, then $\nabla F(x^*) - I$ is a P_0-matrix by Lemma 2.3. Applying Lemma 2.4, we obtain that $2I + \nabla F(x^*) - I = \nabla F(x^*) + I$ is invertible. Therefore $\nabla_z \mathcal{F}(y^*, z^*)$ is invertible and, by the implicit function theorem, we obtain a map H such that (i) holds. Similarly, if $-(F+I)$ is a P_0-map, then $2I - (\nabla F(z^*) + I) = I - \nabla F(z^*) = -\nabla_y \mathcal{F}(y^*, z^*)$ is invertible and we obtain a map $\widetilde{H}$ such that (ii) holds. □

Remark 2.6. (i) The condition $F - I$ (respectively, $-(F+I)$) being a P_0-map is actually quite natural since it is exactly the requested assumptions to solve the (NCP) problem, see Section 2.3.
(ii) (NAVE vs AVE). At this stage, the reader may have already noticed an analogy with the (LCP) reformulation of (AVE). Indeed, if $F(x) = Ax - b$, then (NAVE) coincides with (AVE), and if either $A - I$ or $-(A+I)$ is a P_0-matrix (which is automatically satisfied if, *e.g.*, the singular values of the matrix A are greater than 1), then the functions H and $\widetilde{H}$ are explicitly given by the formulae

$$H(y) = My + q \qquad \text{and} \qquad \tilde{H}(y) = \widetilde{M}z + \widetilde{q},$$

where $M, \widetilde{M}$, q and $\widetilde{q}$ appear in (1.2). Consequently, in this case it is possible to solve (AVE) as explained in the introduction.

2.3 Smoothing techniques to solve (NCP)

As already mentioned, even though the functions H and $\widetilde{H}$ are only implicitly defined, we can still solve (2.3)–(2.4) numerically (we shall do so below), whenever it is guaranteed that H, $\widetilde{H}$ are P_0-maps. This is the aim of the following lemma, yielding a criterium based on F.

Lemma 2.7 (Guaranteeing P_0-property for H, $\widetilde{H}$)**.**

(i). If $F - I$ is a P_0-map, then so is H.

(ii). If $-(F+I)$ is a P_0-map, then so is $\widetilde{H}$.

Proof. We now focus on the case of (2.3), the case of (2.4) can be adapted accordingly. Let $y_1, y_2 \in \mathbb{R}^d$, with $y_1 \neq y_2$. We infer from (2.3) that

$$2H(y_k) = F\left(y_k - H(y_k)\right) - \left(y_k - H(y_k)\right), \quad k \in \{1, 2\}.$$

Setting $t_k := y_k - H(y_k)$, it follows

$$2H(y_k) = (F - I)(t_k).$$

Using the fact that $F - I$ is a P_0-map, we deduce that for some coordinate $j = j(t_1, t_2) \in \{1, \ldots, d\}$, we have

$$2\left(H(y_1)^j - H(y_2)^j\right)\,(t_1^j - t_2^j) \geq 0,$$

from which we infer

$$\left(H(y_1)^j - H(y_2)^j\right)\,\left((y_1^j - y_2^j) - (H(y_1)^j - H(y_2)^j)\right) \geq 0,$$

yielding

$$\left(H(y_1)^j - H(y_2)^j\right)\,(y_1^j - y_2^j) \geq \left(H(y_1)^j - H(y_2)^j\right)^2 \geq 0.$$

This is the desired property for the map H. □

To solve (2.3), we will apply the smoothing approach proposed in [10] and more precisely the non-parametric technique introduced in [30]. The overall approach of [10] is based on functions θ satisfying the following properties:

- the function $\theta : \mathbb{R} \to (-\infty, 1)$ is concave, continuous and nondecreasing;
- $\theta(t) < 0$, for all $t \in (-\infty, 0)$, $\theta(0) = 0$ and $\lim_{t \to +\infty} \theta(t) = 1$.

These functions are used as *certificate of positivity*, that is, they "detect" whether $t = 0$ or $t > 0$ holds in a "continuous way", in the sense of the following characterization:

$$t > 0 \quad \Longleftrightarrow \quad \lim_{r \to 0} \theta\left(\frac{t}{r}\right) = 1.$$

The authors in [10] used these functions to regularize the (nonsmooth) (NCP)

$$0 \leq x \perp H(x) \geq 0, \tag{2.5}$$

by means of a sequence of smooth systems (indexed by $r > 0$) of the form

$$G_r(x, H(x)) = \left(G_r(x, H(x))^1, \cdots, G_r(x, H(x))^d\right)^T = (0, \cdots, 0)^T, \tag{2.6}$$

where

$$G_r(x, H(x))^i := r\psi^{-1}\left[\psi\left(\frac{x^i}{r}\right) + \psi\left(\frac{H(x)^i}{r}\right)\right] \qquad \text{with} \quad \psi := 1 - \theta.$$

Then they eventually take the limit as r tends to 0.

Several convergence results have been established under the assumption that the problem has at least one solution and H is a P_0–map. Although this approach is efficient numerically, it suffers from two drawbacks:

- There is no clear or optimal strategy to drive the parameter r to 0.

- The following ad hoc technical assumption on the function ψ has been used without rigorous explanation:

$$\text{there exist } a \in (0,1) \text{ and } R_a > 0 \text{ such that:} \quad \frac{\psi(t)}{2} \geq \psi\left(\frac{1}{a}t\right), \quad \text{for all } t \in (R_a, +\infty). \tag{2.7}$$

The first drawback has been addressed in [30] by considering a larger system of equations

$$\begin{cases} G_r(x, H(x)) = (0, \cdots, 0)^T, \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|H(x)^-\|^2 + r^2 + \varepsilon r = 0, \end{cases} \tag{2.8}$$

where $\varepsilon > 0$ is some positive parameter. The second drawback will be the subject of the following result which proves that this technical assumption (2.7) corresponds to a well-known property.

Theorem 2.8 (asymptotic behavior)**.** *Let $\psi : (0, \infty) \to (0, \infty)$ be a convex decreasing function satisfying*

$$\lim_{x\to\infty} \psi(x) = \inf \psi = 0.$$

The following assertions are equivalent:

(i). (Łojasiewicz inequality at infinity) There exists $c > 0$ such that

$$\liminf_{x\to\infty} \frac{x|\psi'(x)|}{\psi(x)} > c > 0.$$

(ii). There exist $m, n > 1$ and $R > 0$ such that:

$$\frac{\psi(x)}{m} \geq \psi(nx), \qquad \text{for all } x \in (R, +\infty) \tag{2.9}$$

(iii). For every $m > 1$ there exist $n > 1$ and $R > 0$ such that:

$$\frac{\psi(x)}{m} \geq \psi(nx), \qquad \text{for all } x \in (R, +\infty)$$

Notice that the technical assumption (2.7) corresponds to (ii). Therefore, the above result shows that it is equivalent to assume that ψ satisfies the Łojasiewicz inequality at infinity. This latter condition is always satisfied if the function ψ is semialgebraic: Indeed, in this case, the corresponding Hardy field (that is, the field of germs of real semialgebraic functions at infinity) has rank one, and consequently, for any non-ultimately zero semi-algebraic function ψ in the single variable x, the function $x \mapsto x\psi'(x)/\psi(x)$ has a non-zero limit as x goes to infinity (see [7, Remark 2.9]). The same argument applies also for the more general case of functions ψ that are definable in some polynomially bounded o-minimal structure (we refer to [6] for the corresponding definitions).

Proof. (i)$\Rightarrow$(ii). Let us assume that (ii) fails. We define inductively a sequence $\{y_n\}_n \subset [1, +\infty)$ such that

$$\lim_{n\to\infty} y_n = +\infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{y_n|\psi'(y_n)|}{\psi(y_n)} = 0.$$

To this end, we set $x_1 = y_1 = 1$. Since (ii) fails, for every $n \geq 2$, taking $m = 1 + \frac{1}{n}$ and $R = y_{n-1}$ we obtain the existence of some $x_n > R$ such that for $y_n := nx_n$ it holds

$$\frac{\psi(x_n)}{m} < \psi(y_n) \quad \text{yielding} \quad \frac{\psi(x_n)}{\psi(y_n)} - 1 < m - 1 = \frac{1}{n}. \tag{2.10}$$

Using convexity we also deduce that

$$\left|\psi'(y_n)\right| \leq \frac{\psi(x_n) - \psi(y_n)}{y_n - x_n} = \left(\frac{n}{n-1}\right)\left(\frac{\psi(x_n) - \psi(y_n)}{y_n}\right),$$

whence, from (2.10),

$$0 \leq \frac{y_n|\psi'(y_n)|}{\psi(y_n)} \leq \left(\frac{n}{n-1}\right)\left(\frac{\psi(x_n)}{\psi(y_n)} - 1\right) < \frac{1}{n-1}.$$

Taking the limit as $n \to \infty$ we conclude that (i) also fails to hold, which establishes the desired implication.

(ii)$\Rightarrow$(iii). Assume that (2.9) holds for some $m_0 > 1$, n_0 and $R_0 > 1$, that is, for all $x > R_0$ we have $\psi(x) \geq m_0\, \psi(n_0 x)$. Then since $n_0 x > x > R$ we also have:

$$\psi(n_0 x) \geq m_0\, \psi(n_0^2 x) \quad \text{yielding} \quad \psi(x) \geq m_0^2\, \psi(n_0^2 x).$$

We conclude that (2.9) also holds for $m_1 = m_0^2$ (under the choice of $n_1 = n_0^2$ and $R_1 = R_0$). Repeating this argument we deduce that (2.9) holds for all $m_k = m_0^k$, $k \geq 1$ (taking $n_k = n_0^k$ and $R_k = R_0$). Since $m_k \to \infty$, in order to establish (iii) it is sufficient to observe that if (2.9) holds for some $\bar{m} > 1$ (together with some $\bar{n} > 1$ and $\bar{R} > 0$) then it also holds for all $m \in (1, \bar{m}]$, since

$$\frac{\psi(x)}{m} \geq \frac{\psi(x)}{\bar{m}}.$$

(iii)$\Rightarrow$(i). Fix $m > 1$, $n > 1$ and $R > 0$ such that (2.9) holds and set

$$c := \left(\frac{m-1}{m}\right)\left(\frac{1}{n-1}\right) > 0.$$

Using convexity of ψ and (2.9), we deduce that for all $x > R$ we have:

$$\left|\psi'(x)\right| \geq \frac{\psi(x) - \psi(nx)}{nx - x} \implies \frac{x|\psi'(x)|}{\psi(x)} \geq \left(\frac{1}{n-1}\right)\left(1 - \frac{\psi(nx)}{\psi(x)}\right) \geq c.$$

This establishes (i) and finishes the proof. $\square$

Remark 2.9. As already mentioned, assertion (i) (Łojasiewicz inequality at infinity) holds true whenever the function ψ is semialgebraic (or more generally, definable in a polynomially bounded o-minimal structure). This already provides a broad assembly of examples of functions satisfying (i), together with straightforward criteria to detect easily whether the property holds, based on certificates of semialgebricity or o-minimality (see [6, Theorem 1.13] *e.g.*).

This being said, let us draw reader's attention to the fact that besides what is asserted in [7, Proposition 2.7], the assumption of *polynomial boundedness* is essential for the validity of (i). Indeed, as shown in [21, Remark 8], the convex function $\psi(x) = (\log(1+x))^{-1}$ is definable in the log-exp o-minimal structure but fails to satisfy (i).

2.4 Algorithm and numerical results

To solve the system of equation (2.8), we will apply the Newton-like method proposed in [30]. However, since H is defined implicitly, we first need to reformulate the problem as follows:

$$\begin{cases} z - F(y-z) + y = 0 & \\ r\psi^{-1}\left(\psi(\frac{y^i}{r}) + \psi(\frac{z^i}{r})\right) = 0 & i = 1\dots d, \\ \frac{1}{2}\|y^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r = 0, & \end{cases} \tag{2.11}$$

where the variable z plays the role of $H(y)$.

Remark 2.10. In this new system of equations we assume that case (i) of Lemma 2.5 holds. (One can proceed in a similar way if (ii) holds.)

In the definition of the following algorithm, we set $\mathbb{X} := (y, z, r)^T$ and

$$\mathbb{H}(\mathbb{X}) := \begin{cases} z - F(y-z) + y & \\ r\psi^{-1}\left(\psi(\frac{y^i}{r}) + \psi(\frac{z^i}{r})\right) & i = 1\dots d, \\ \frac{1}{2}\|y^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r & \end{cases} \tag{2.12}$$

so that (2.11) is reduced to $\mathbb{H}(\mathbb{X}) = 0$. This algorithm corresponds to a Newton method under a standard Armijo line search.

Algorithm

---

1. Chose $\mathbb{X}^0 = (\mathbf{X}^0, r^0)$, $\mathbf{X}^0 \in \Xi$, $r^0 = \langle y^0, z^0\rangle/n$, $\tau \in (0, 1/2)$, $\varrho \in (0, 1)$. Set $k = 0$.
2. If $\mathbb{H}(\mathbb{X}^k) = 0$, stop.
3. Find a direction $\mathbf{d}^k \in \mathbb{R}^{2n+1}$ such that
$$\mathbb{H}(\mathbb{X}^k) + \nabla_{\mathbb{X}}\mathbb{H}(\mathbb{X}^k)\mathbf{d}^k = 0.$$
4. Choose $\zeta^k = \varrho^{j_k} \in (0, 1)$, where $j_k \in \mathbb{N}$ is the smallest integer such that
$$\Theta(\mathbb{X}^k + \varrho^{j_k}\mathbf{d}^k) - \Theta(\mathbb{X}^k) \le \tau\varrho^{j_k}\ \nabla\Theta(\mathbb{X}^k)^T\mathbf{d}^k.$$
5. Set $\mathbb{X}^{k+1} = \mathbb{X}^k + \zeta^k\mathbf{d}^k$ and $k \leftarrow k + 1$. Go to step 2.

The merit function used in the line search corresponds to the square of the global error:

$$\Theta(\mathbb{X}) = \frac{1}{2}\|\mathbb{H}(\mathbb{X})\|^2.$$

To get a well defined algorithm, the initial point $(y^0, z^0)^T$ must be an interior point, and the initial value for r must be positive $r^0 > 0$.

3 Applications

In this section we show that several problems, which can be naturally restated as (NAVE) and can be solved efficiently thanks to the above transformation. We present in this section numerical experiments, in which the smoothing functions are restricted to two specific cases

$$\theta_1(t) := \begin{cases} \dfrac{t}{t+1}, & t \geq 0 \\ t, & t < 0 \end{cases} \qquad \text{and} \qquad \theta_2(t) := 1 - e^{-t}.$$

The numerical experiments are conducted in an ordinary computer. All program codes are written and executed in MATLAB R2023a. In Subsection 3.1 and 3.3, we employ a similar stopping criterion for every numerical method, by using a tolerance $Tol = 1e-10$ and fixing the maximum number of iterations to $N_{\max} = 2000$. Since the NAVE problems may have multiple solutions, in the following, the error will be computed by $Error = \|F(x_{\text{approximate}}) - |x_{\text{approximate}}|\|$ in Subsection 3.1 and respectively by $Error = \|F(x_{\text{approximate}}) - |x_{\text{approximate}}| - b\|$) in Subsection 3.3).

3.1 Ridge Regression

Ridge regression adds to the loss function $\mathcal{L}(x)$, $x \in \mathbb{R}^d$ a penalty term in order to avoid overfitting: historical development and the applications in data science of ridge regression can be found *e.g.* in [12, 14]. This penalty term usually consists of adding the squared magnitude of the coefficients (traditionally denoted by w).

We hereby consider an asymmetric ridge regression of the form:

$$\min_{x \in \mathbb{R}^d} \left\{ \mathcal{L}(x) + \sum_{j=1}^{d} \left(\lambda_j \max\{x^j, 0\}^2 + \mu_j \max\{-x^j, 0\}^2 \right) \right\}, \tag{3.1}$$

where the penalization parameters λ_j and μ_j satisfy $\lambda_j - \mu_j \neq 0$ for all $j \in \{1, \cdots, d\}$. The case $\lambda_j = \mu_j = \lambda$ for every j corresponds to the classical ridge regression, which will not be considered here. On the other hand, the case $\lambda_j = 0$ for all j and $\mu_j > 0$, corresponds to a penalization of the negativity of the coefficients, promoting solutions with positive coefficients. The necessary condition for optimality reads as follows:

$$\nabla \mathcal{L}(x) + 2\lambda \max\{x, 0\} - 2\mu \max\{-x, 0\} = 0,$$

where the two vectors $\lambda \max\{x, 0\}$ and $\mu \max\{-x, 0\}$ are to be understood componentwise. Noticing that $2\max\{x, 0\} = |x| + x$ and $2\max\{-x, 0\} = |x| - x$, we end up with the following (NAVE) problem

$$F(x) - |x| = 0 \qquad \text{with} \qquad F(x) = \left(\frac{1}{\mu - \lambda}\right) \nabla \mathcal{L}(x) + \left(\frac{\mu + \lambda}{\mu - \lambda}\right) x \quad \text{(coordinatewise)}$$

Therefore, one can solve the previous problem if either $F - I$ or $-(F + I)$ is a P_0–map, that is

$$\text{either} \quad (\mu - \lambda)^{-1} (\nabla \mathcal{L} + 2\lambda I) \quad \text{or} \quad -(\mu - \lambda)^{-1} (\nabla \mathcal{L} + 2\mu I) \quad \text{is a } P_0\text{–map.}$$

To illustrate for asymmetric ridge regression, we consider the loss function

$$\mathcal{L}(x) = \frac{1}{2}\|Ax - b\|^2, \qquad \text{where } A \in \mathbb{R}^{m \times d} \text{ and } b \in \mathbb{R}^d. \tag{3.2}$$

We performed numerical experiments, fixing $\lambda_j = \bar{\lambda}$ and $\mu_j = \bar{\mu}$ for every $j \in \{1, \cdots, d\}$. These parameters, matrix $A \in \mathbb{R}^{m\times d}$ and vector $b \in \mathbb{R}^d$ were randomly generated with values in $[-5, 5]$.

As shown in Table 1, considering the average number of iterations with similar tolerance, using the function θ_2 is better, while in an exceptional case $m = 20 > d = 10$ and $(\bar{\lambda}, \bar{\mu}) = (0, 100)$, θ_2–smoothing performs worse. On the other hand, while the parameters $\bar{\lambda}$ and $\bar{\mu}$ become greater, which can be compared to the ascent of the (classical) ridge parameter, θ_2–smoothing performs within a better tolerance in a small number of iterations.

Table 1: Comparing (asymmetric) ridge regression with different smoothing functions

		Error		Iterations		Running time($\times e-2$(s))	
$(\bar{\lambda}, \bar{\mu})$	(m, d)	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
$(0, 100)$	$(3, 10)$	$1.9e-11$	$7.3e-15$	18	21	4.81	2.96
	$(5, 10)$	$5.8e-11$	$1.3e-16$	17	71	4.78	6.26
	$(10, 10)$	$5.4e-11$	$1.9e-16$	18	24	5.78	3.45
	$(20, 10)$	$3.8e-11$	$3.5e-3$	17	2000	4.94	819
$(200, 1000)$	$(3, 10)$	$8.9e-11$	$1.4e-17$	18	23	5.62	2.99
	$(5, 10)$	$2.7e-11$	$3.6e-18$	19	22	5.12	2.9
	$(10, 10)$	$6e-11$	$2.7e-17$	18	33	4.93	3.87
	$(20, 10)$	$4.18e-11$	$1.2e-10$	18	40	5.41	4.17

To end this part, we give a heuristic observation on a sparse optimization problem (see *e.g.* [13,36]). Let us consider the following problem

$$\min_{x\in\mathbb{R}^d} \mathcal{L}(x) + \lambda\|x\|_1. \tag{3.3}$$

The first order optimality condition for (3.3) has the form

$$0 \in \nabla\mathcal{L}(x) \, + \, \lambda\, \partial\|\cdot\|_1(x), \tag{3.4}$$

where the subdifferential of ℓ_1 norm can be written explicitly as

$$q \in \partial\|\cdot\|_1(x) \text{ if and only if } \begin{cases} q_i = \operatorname{sign}(x_i), & \text{if } x_i \neq 0, \\ |q_i| \leq 1, & \text{if } x_i = 0. \end{cases}$$

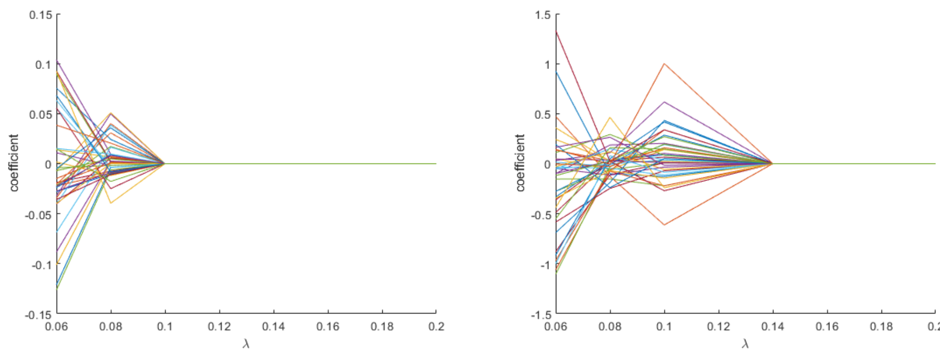
Using the fact that $\alpha \operatorname{sign}(\alpha) = |\alpha|$ for every $\alpha \in \mathbb{R}$, the inclusion (3.4) can be transformed into

$$x\nabla\mathcal{L}(x) + \lambda|x| = 0 \;\; \text{(coordinatewise)}.$$

The above equation just provides a necessary condition for optimal solution. In the following, we give a short numerical observation to guarantee its potential utility in sparse optimization. In order to apply results from previous sections, it is necessary to ensure that one of the following maps is a P_0-map

$$-\frac{1}{\lambda}\, x\, \nabla\mathcal{L}(x) - I \qquad \text{and} \qquad \frac{1}{\lambda}\, x\, \nabla\mathcal{L}(x) - I.$$

In the following figures, we use the same quadratic loss function (3.2), where matrix A and vector b are randomly generated ranging from -1 to 1 and $-0,05$ to 0, respectively. Figure 1 shows the behavior of each coefficient while increasing the tuning parameter $\lambda > 0$.



(a) θ_1–smoothing (b) θ_2–smoothing.

Figure 1: Problem in dimension $m = 20$ and $d = 40$.

3.2 Nonlinear ordinary differential equations

A NAVE problem also naturally arises when we deal with a discretization of a nonlinear ordinary differential equation (ODE, for short) involving rough velocity, for example $\dot{\gamma}(t) = \sqrt{|\gamma(t)|}$ as well as an ODE of the form

$$\Phi(X^{(2k)}, X^{(2k-1)}, \ldots, \dot{X}) = |X|$$

In this subsection we provide two examples (one being a stiff ODE) to illustrate the effectiveness of smoothing techniques when using finite difference schemes for ODEs.

Example 3.1. We consider a stiff ODE with initial value as follows

$$\begin{cases} \ddot{x} + 1001\dot{x} - 1000|x| = 0, \; t > 0 \\ x(0) = x_0 < 0, \; \dot{x}(0) = 0, \end{cases} \tag{3.5}$$

whose exact solution is

$$x_{\text{exact}}(t) = x_0 \left(-\frac{1}{999} e^{-1000t} + \frac{1000}{999} e^{-t} \right) \approx x_0 e^{-t}.$$

Let us consider problem (3.5) in time domain $I = [0, T]$. We use a uniform mesh $\boldsymbol{t} = (t_i)$, where $t_i = ih$ for $i \in \{0, \cdots, N\}$ and $h = T/N$, and the approximation solution will be $\boldsymbol{x} = (x_i)$ where $x_i \approx x(t_i)$. For the first and the second derivative, we use the 2nd–order approximation

$$\ddot{x}(t_i) \approx \frac{x_{i-2} - 2x_{i-1} + x_i}{h^2} \text{ and } \dot{x}(t_i) \approx \frac{x_{i+1} - x_{i-1}}{2h}.$$

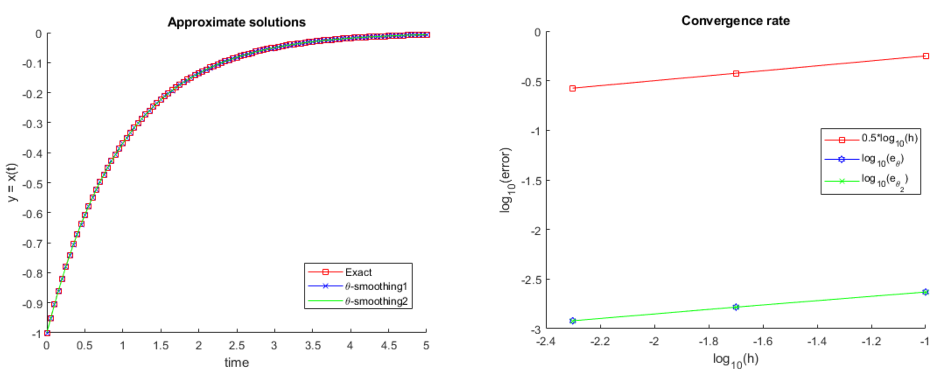
Remarkably, at the final time, the first derivative $\dot{x}(t_N)$ will be computed via the 2nd–order backward formula $\dot{x}(t_N) \approx (x_{N-2} - 4x_{N-1} + 3x_N)/2h$. Since the initial velocity is zero, using 1st–order backward approximation, we note that $x_{-1} = x_0$. The discretization of (3.5) can be written as

$$\frac{1}{1000}\boldsymbol{A}\boldsymbol{x} + \frac{1001}{1000}\boldsymbol{B}\boldsymbol{x} - |\boldsymbol{x}| = \boldsymbol{b},$$

where $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{N\times N}$ is determined by

$$\boldsymbol{A} = \frac{1}{h^2}\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -2 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix} \text{ and } \boldsymbol{B} = \frac{1}{2h}\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -4 & 3 \end{pmatrix} \tag{3.6}$$

Here, vector $\boldsymbol{b} = (b_i) \in \mathbb{R}^N$ is defined by $b_i = 0$ for $i \geq 2$, $b_1 = x_0(1/(1000h^2) + 1001/(2000h))$ and $b_2 = -x_0/(1000h^2)$.
In Figure 2.(a), we approximate the solution of equation (3.5) with initial condition $x_0 = -1$ and time interval $I = [0, 5]$. The finite difference scheme was computed with mesh size $h = 0.05$ and the error is $9.22e - 4$ when applying θ_1 and θ_2 smoothing funtions. To get convergence rate in Figure 2.(b), we apply difference mesh sizes in the same time interval $I = [0, 1]$ and intial condition $x_0 = -2$.



(a) Approximate solutions. (b) Convergence rate $O(h^{1/2})$.

Figure 2: Solving equation 3.5

Let us now make some comments on the utility of NAVE for boundary value problems: we consider a boundary value problem related to (3.5)

$$\begin{cases} \ddot{x} + 1001\dot{x} - 1000|x| = 0,\ t \in (0, T), \\ x(0) = x_0 < 0,\ x(T) = y_0 \in \mathbb{R}. \end{cases} \tag{3.7}$$

In order to illustrate this case, we consider the time interval $I = [0, 2]$ and exact solution is determined by x_{exact}. Using similar time mesh as above, the first and second derivatives are approximate as follows

$$\ddot{x}(t_i) \approx \frac{x_{i-1} - 2x_i + x_{i+1}}{h^2} \text{ and } \dot{x}(t_i) \approx \frac{x_i - x_{i-1}}{h}.$$

Figure 3 shows the convergence rate for the boundary value problem (3.7). The smoothing technique used in this problem presents a better accuracy compared to the above initial value problem, which seems to be natural because of the stiffness of the problem (3.5). It is noteworthy that Figure 3 also depicts an expected convergence rate since we have used a first order approximation for $\dot{x}$.

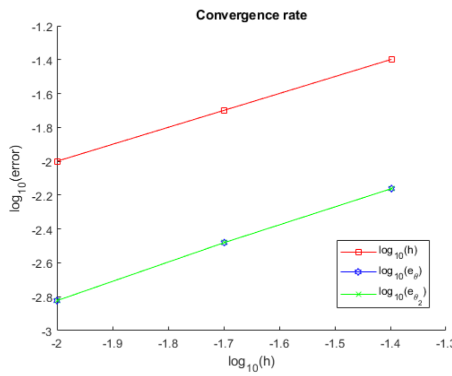


Figure 3: Convergence rate $O(h)$ for a boundary value problem

Example 3.2. For a continuous function $f : [0, +\infty) \to \mathbb{R}$, we consider an ODE

$$\begin{cases} \ddot{x} + \arctan(x) - |x| = f(t),\ t > 0, \\ x(0) = x_0 \in \mathbb{R},\ \dot{x}(0) = 0. \end{cases} \tag{3.8}$$

Using a similar discretization as in Example 3.1, the unknown variable $\boldsymbol{x} \approx x(\boldsymbol{t})$ solves a NAVE problem as follows

$$\boldsymbol{A}\boldsymbol{x} + \arctan(\boldsymbol{x}) - |\boldsymbol{x}| = \boldsymbol{b}, \tag{3.9}$$

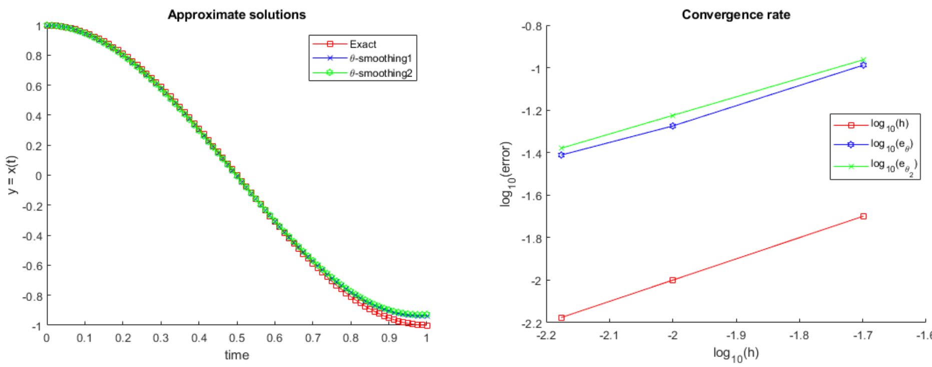
where the matrix $\boldsymbol{A}$ is determined as in Example 3.1 and the vector $\boldsymbol{b} \in \mathbb{R}^N$ is defined by $b_i = f(t_i)$ for $i \geq 2$ and

$$b_1 = f(t_1) + \frac{x_0}{h^2} \text{ and } b_2 = f(t_2) - \frac{x_0}{h^2}.$$

To illustrate for this example, we consider problem (3.8) with source term

$$f(t) = \arctan(\cos(\pi t)) - |\cos(\pi t)| - \pi^2 \cos(\pi t),$$

whose exact solution is $x_{exact}(t) = \cos(\pi t)$. Figure 4.(a) shows the approximate solution on the time interval $I = [0, 1]$ with mesh size $h = 0.0125$. The error between θ_1 (resp. θ_2) approximation and exact solution is 0.06 (resp. 0.0725). Besides, Figure 4.(b) displays convergence rate of the combination of finite difference scheme and the θ–smoothing applying for the associated NAVE problem.



(a) Approximate solutions. (b) Convergence rate $O(h)$.

Figure 4: Solving equation 3.8.

3.3 Comparison of methods for NAVE

Instead of smoothing procedure considered in Section 2, one can solve a NCP via other numerical methods. In this subsection we give examples to compare the efficiency of four methods

- Newton–like method with smoothing functions θ_1 and θ_2;
- approximating by Soft–Max function, in which the main idea is to approximate the complementarity condition via the limit

$$\max_{i\in\{1,\cdots,d\}} x_i = \lim_{r\searrow 0} r\log\left(\sum_{i=1}^{d} e^{x_i/r}\right).$$

  which have been widely used in many optimization problems, for example [31, Example 1.30], [19, 20, 28];
- using interior point method, for example, one can find the use of interior point method for complementarity problems in [11, 15, 17, 29].

Now, in the following examples, we solve the system $\tilde{F}(x) - |x| = b$, especially, Example 3.4 and 3.5 can be found in [2, 18].

Example 3.3. We consider $\tilde{F}(x) = Ax$, where

$$A = \text{tridiag}(-1, 4, -1) \in \mathbb{R}^{d\times d},\ x^* \in \mathbb{R}^d,\ b = Ax^* - |x^*|. \tag{3.10}$$

Example 3.4. $\tilde{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is defined by

$$\tilde{F}(x) := \begin{pmatrix} 2x_1 - 2 \\ 2x_2 + x_2^3 - x_3 + 3 \\ x_2 + 2x_3 + 2x_3^3 - 3 \end{pmatrix}.$$

Example 3.5. $\tilde{F} : \mathbb{R}^4 \to \mathbb{R}^4$ is defined by

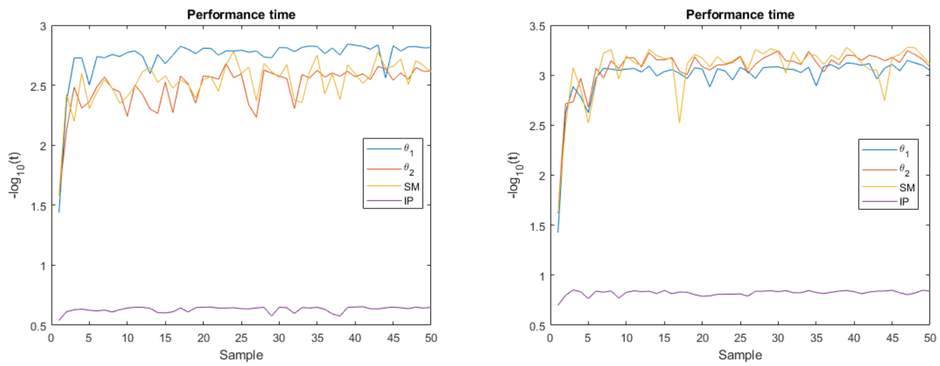
$$\tilde{F}(x) := \begin{pmatrix} 3x_1^2 + x_1 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 \\ 2x_1^2 + x_1 + x_2^2 + x_2 + 10x_3 + 2x_4 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 3x_3 + 9x_4 \\ x_1^2 + 3x_2^2 + 2x_3 + 4x_4 \end{pmatrix}.$$

Table 2 compares the four methods: smoothing method with θ_1 and θ_2, Soft Max (denoted SM) and Interior Point (denoted IP) method on the NAVE problem associated to Example 3.3–3.5. In the first example, the vector b is randomly generated with values in $[-5, 5]$ and the problem is considered in dimensions $d = 10, 50, 200$. In Example 3.4 and 3.5, we respectively consider $b_1 = (-1, -5, 10)^T$, $b_2 = (9, -100, 10)^T$, $b_3 = (200, 0, 900)^T$, $b_1^* = (10, 10, -12, 0)^T$, $b_2^* = (20, -100, -12, 1)^T$ and $b_3^* = (200, 10, -5, -5)^T$. We observe that the smoothing method (especially with θ_2–smoothing function) is the most robust among the considered methods. In connection with convergence speed, the interior point method performs much less competitively than the others, while it only reaches $1e-2$ after $N = 2000$ iterations. Another point that can be recognized from Table 2 is that the Soft Max method could only solve problems with small size, for example for problems in dimension $N = 50$ and $N = 200$ the singularities appear after less than 100 iterations.

Table 2: Several methods to solve NAVE

Example	Vector b	Error				Iterations			
		θ_1	θ_2	SM	IP	θ_1	θ_2	SM	IP
3.3	$d = 10$	$9.3e-11$	$1.5e-11$	$5.5e-12$	$1.6e-2$	20	13	173	2000
	$d = 50$	$1.4e-11$	$5.3e-15$	NaN	$8.96e-3$	29	41	41	2000
	$d = 200$	$8.39e-11$	$1.32e-14$	NaN	$9.55e-3$	45	76	96	2000
3.4	b_1	$4.7e-11$	$1.7e-11$	$7e-14$	$5.5e-2$	14	9	8	2000
	b_2	$1.7e-11$	$3.6e-11$	$1.8e-12$	$4e-1$	22	16	15	2000
	b_3	$9.3e-11$	$1.4e-13$	$1.4e-13$	$2e+2$	211	205	205	2000
3.5	b_1^*	$5.5e-11$	$1.3e-14$	$2e-14$	$7e-2$	16	12	12	2000
	b_2^*	$5.5e-11$	$2.2e-14$	$4.7e-14$	$9.1e-1$	26	22	16	2000
	b_3^*	$6.1e-11$	$1.4e-10$	$3e+1$	$3.8e-0$	50	43	2000	2000

Figure 5(a) and 5(b) display the performance time between different methods for Example 3.3 with the size $n = 20$ and Example 3.4, respectively. We did the observation with 50 samples and the vector b is randomly generated with values in $[-10, 10]$. At a first sight, the interior point method appears to be the slowest one in comparison with the other three methods. As shown in Figure 5(a), the θ_1–smoothing performs the best choice among all the methods. If we look carefully, in lower dimension as Example 3.4, the Soft Max and θ_2–smoothing performs slightly better than θ_1–smoothing method.



(a) Example 3.1 in dimension $n = 20$.　　(b) Example 3.2.

Figure 5: Performance time.

4 Conclusion and discussion

In this work, we applied smoothing techniques commonly used for Nonlinear Complementarity Problems (NCP) to Nonlinear Absolute Value Equations (NAVE). We first showed that a NAVE can be formulated as a NCP with an implicitly known corresponding mapping. We then established that a NAVE can be effectively addressed under a mild direct assumption on the NAVE function. Additionally, we clarified a technical assumption used in some smoothing approaches, proving its equivalence to a Łojasiewicz inequality, thus showing its broad applicability. Last but not least, we provided illustrative examples and applications that through numerical verification reveal effectiveness and potential of this approach.

In a future work, we aim to establish error bounds or estimations and study the complexity of the proposed method in favorable situations, such as when the implicitly corresponding mapping in the NCP formulation is monotone or strongly monotone.

Acknowledgement. This work was initiated during a research stay of Aris Daniilidis and Trí Minh Lê to INSA Rennes (February 2023). These authors thank their hosts for hospitality. The first author acknowledges support from the Austrian Science Fund (FWF, P–36344-N).

References

[1] L. Abdallah, M. Haddou, T. Migot, Solving absolute value equation using complementarity and smoothing functions. *J. Comput. Appl. Math.* **327** (2018), 196–207.

[2] J. H. Alcantara, J-S. Chen. A new class of neural networks for NCPs using smooth perturbations of the natural residual function, *J. Comput. Appl. Math.* **407** (2022), Paper No. 114092, 22 pp.

[3] J. H. Alcantara, J.-S. Chen, M. K. Tam, Method of alternating projections for the general absolute value equation. *J. Fixed Point Theory Appl.* **25** (2023), Paper No. 39, 38 pp.

[4] I. Ben Gharbia J.C. Gilbert, Nonconvergence of the plain Newton-min algorithm for linear complementarity problems with a P-matrix, *Mathematical Programming* **134** (2012), 349–364.

[5] J. Y. Bello Cruz, O. P. Ferreira, L. F. Prudente, On the global convergence of the inexact semi-smooth Newton method for absolute value equation, *Comput. Optim. Appl.* **65** (2016), 93–108.

[6] M. Coste, *An Introduction to o-minimal Geometry*, Instituti Editoriali e Poligrafici Internazionali, Pisa (2000).

[7] D. Dacunto, V. Grandjean, A gradient inequality at infinity for tame functions, *Rev. Mat. Complut.* **18** (2005), 493–501.

[8] M. Fiedler, V. Pták, On matrices with nonpositive off-digagonal elements and positive principal minors, *Czech. Math. J.* **12** (1962), 382–400.

[9] S.-L. Hu and Z.-H. Huang, A note on absolute value equations, *Optim. Lett.* **4** (2010), 417–424.

[10] M. Haddou, P. Maheux, Smoothing methods for nonlinear complementarity problems, *J. Optim. Theory Appl.* **160** (2014), 711–729.

[11] M. Haddou, T. Migot, J. Omer, A generalized direction in interior point method for monotone linear complementarity problems, *Optim. Lett.* **13** (2019), 35–53.

[12] T. Hastie, Ridge regularization: An essential concept in data science, *Technometrics* **62** (2020), 426–433.

[13] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning. Data mining, inference and prediction*, Springer (New York, 2001).

[14] R. W. Hoerl, Ridge regression: A historical context, *Technometrics* **62** (2020), 420–425.

[15] A. N. Iusem, An interior point method for the nonlinear complementarity problem, *Appl. Numer. Math.* **24** (1997), 469–482.

[16] C. Kanzow, A new approach to continuation methods for complementarity problems with uniform P-functions, *Oper. Res. Lett.* **20** (1997), 85–92.

[17] M. Kojima, N. Megiddo, T. Noma, A. Yoshise, *A unified approach to interior point algorithms for linear complementarity problems*, Lecture Notes in Comput. Sci. **538**, Springer (Berlin, 1991).

[18] M. Kojimam S. Shindo, Extension of Newton and quasi-Newton methods to systems of PC^1 equations, *J. Oper. Res. Soc. Jpn.* **29** (1986), 352–374.

[19] X. Li, An entropy–based aggregate method for minimax optimization, *Eng. Opt.* **18** (1992), 277–285.

[20] Y. Li, T. Tan, X. Li, A log-exponential smoothing method for mathematical programs with complementarity constraints, *Appl. Math. Comput.* **218**(2012), 5900—5909.

[21] T. L. Loi, Łojasiewicz inequalities in o-minimal structures, *Manuscripta Math.* **150** (2016), 59–72.

[22] T. Lotfi and H. Veiseh, A note on unique solvability of the absolute value equation, 2013.

[23] O. L. Mangasarian, RR. Meyer, Absolute value equations, *Linear Algebra Appl.* **419** (2006), 359–367.

[24] O. L. Mangasarian, Linear complementarity as absolute value equation solution, *Optim. Lett.* **8** (2014), 1529–1534.

[25] MATLAB R2023A, Natick, Massachusetts: The MathWorks Inc., 2010.

[26] J.-J. Moré, Global methods for nonlinear complementarity problems, *Math. Oper. Res.* **21** (1996), 589–614.

[27] J.-J. Moré , W.C. Rheinboldt, On P- and S- functions and related classes of n- dimensional nonlinear mappings, *Linear Algebra Appl.* **6** (1973), 45–68.

[28] Y. Nesterov, Smooth minimization of nonsmooth functions, *Math. Program. Ser. A* **103** (2005), 127–152.

[29] F. A. Potra, Y. Ye, Interior-point methods for nonlinear complementarity problems, *J. Optim. Theory Appl.* **88**(1996), 617–642.

[30] E. H. Osmani, M. Haddou, N. Bensalem, L. Abdallah, A new smoothing method for nonlinear complementarity problems involving P_0-function. *Stat. Optim. Inf. Comput.* **10** (2022), 1267–1292.

[31] R. T. Rockafellar, J.-B. R. Wets, *Variational analysis*, Grundlehren Math. Wiss. **317**, Springer (Berlin, 1998).

[32] O. Prokopyev, On equivalent reformulations for absolute value equations, *Comput. Optim. Appl.* **44** (2009), 363–372.

[33] J. Rohn, A theorem of the alternatives for the equation $ax + b|x| = b$, *Linear Mult. Algebra* **52** (2004), 421–426.

[34] J. Rohn, On unique solvability of the absolute value equation, *Optim. Lett.* **3** (2009), 603–606.

[35] J. Rohn, V. Hooshyarbakhsh, and R. Farhadsefat, An iterative method for solving absolute value equations and sufficient conditions for unique solvability, *Optim. Lett.* **8** (2014), 35–44.

[36] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B* **58** (1996), 267–288.

[37] S.-L. Wu and C.-X. Li, A note on unique solvability of the absolute value equation, *Optim. Lett.* **14** (2020), 1957–1960.

Aris Daniilidis, Trí Minh Lê
Institut für Stochastik und Wirtschaftsmathematik, VADOR E105-04
TU Wien, Wiedner Hauptstraße 8, A-1040 Wien
E-mail: {aris.daniilidis, minh.le}@tuwien.ac.at
https://www.arisdaniilidis.at/
Research supported by the grants:
Austrian Science Fund (FWF P-36344N) (Austria)

Mounir Haddou, Olivier Ley
Univ Rennes, INSA, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France
E-mail: {mounir.haddou, olivier.ley}@insa-rennes.fr
http://{haddou, ley}.perso.math.cnrs.fr/
Research supported by the Centre Henri Lebesgue ANR-11-LABX-0020-01.