Polyhedral Newton-min algorithms for complementarity problems

Jean-Pierre Dussault, Mathieu Frappier, Jean Charles Gilbert

Polyhedral Newton-min algorithms for complementarity problems

Jean-Pierre Dussault · **Mathieu Frappier** · **Jean Charles Gilbert**

March 15, 2024

Abstract The semismooth Newton method is a very efficient approach for computing a zero of a large class of nonsmooth equations. When the initial iterate is sufficiently close to a regular zero and the function is strongly semismooth, the generated sequence converges quadratically to that zero, while the iteration only requires to solve a linear system. If the first iterate is far away from a zero, however, it is difficult to force its convergence using linesearch or trust regions because a semismooth Newton direction may not be a descent direction of the associated least-square merit function, unlike when the function is differentiable. We explore this question in the particular case of a nonsmooth equation reformulation of the nonlinear complementarity problem, using the minimum function. We propose a globally convergent algorithm using a modification of a semismooth Newton direction that makes it a descent direction of the least-square function. Instead of requiring that the direction satisfies a linear system, it must be a feasible point of a convex polyhedron; hence, it can be computed in polynomial time. This polyhedron is defined by the often very few inequalities, obtained by linearizing pairs of functions that have close negative values at the current iterate; hence, somehow, the algorithm feels the proximity of a "negative kink" of the minimum function and acts accordingly. In order to avoid as often as possible the extra cost of having to find a feasible point of a polyhedron, a hybrid algorithm is also proposed, in which the Newton-min direction is accepted if a sufficient-descent-like criterion is satisfied, which is often the case in practice. Global and fast convergence to regular solutions is proved.

Keywords complementarity problem · global convergence · least-square merit function · linesearch · minimum function · nonsmooth reformulation · P-matrix · polyhedral Newton-min algorithm · quadratic convergence • semismooth Newton.

Mathematics Subject Classification (2020) 49M15 · 49M37 · 65K05 · 65K15 · 90C30 · 90C33 · 90C55.

Jean-Pierre Dussault
Département d'Informatique, Fac. des Sciences, Univ. de Sherbrooke, Québec, Canada
E-mail: Jean-Pierre.Dussault@Usherbrooke.ca, ORCID 0000-0001-7253-7462

Mathieu Frappier
Département de Mathématiques, Fac. des Sciences, Univ. de Sherbrooke, Québec, Canada
E-mail: Mathieu.Frappier@usherbrooke.ca,

Jean Charles Gilbert
Inria Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France
Département de Mathématiques, Fac. des Sciences, Univ. de Sherbrooke, Québec, Canada
E-mail: Jean-Charles.Gilbert@inria.fr, ORCID 0000-0002-0375-4663

1 Introduction

1.1 The complementarity problem

Let be given a positive integer n and two *smooth* functions $F : \Omega \to \mathbb{R}^n$ and $G : \Omega \to \mathbb{R}^n$ defined on an open subset Ω of $\mathbb{R}^n$. This paper considers, with an algorithmic point of view, the standard (nonlinear) complementarity problem. This problem consists in finding a vector $x \in \Omega$ such that

$$F(x) \geqslant 0, \qquad G(x) \geqslant 0, \qquad \text{and} \qquad F(x)^\mathsf{T} G(x) = 0, \tag{1.1a}$$

where vector inequalities must be taken in a componentwise fashion and $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto u^\mathsf{T} v = \sum_{i=1}^n u_i v_i$ is the Euclidean scalar product of $\mathbb{R}^n$ (the sign "$^\mathsf{T}$" is used to denote transposition of vectors and matrices). We denote by $[1:n] := \{1, \ldots, n\}$ the set of the first n positive integers. Below, the system (1.1a) is written compactly as follows:

$$0 \leqslant F(x) \perp G(x) \geqslant 0, \tag{1.1b}$$

where the sign "$\perp$" refers to the required orthogonality of the vectors $F(x)$ and $G(x)$. In many contributions [83], the map G is supposed to be the identity ; like in [39,40], we have preferred the *balanced* model (1.1), not only for its higher generality, but also because it presents the technical advantage of avoiding repeating reasoning, thanks to the possibility to switch F and G. The term "complementarity" comes from the fact that, due to the nonnegativity of $F(x)$ and $G(x)$ in (1.1), for all $i \in [1:n]$, either $F_i(x)$ or $G_i(x)$ must vanish and determining which of them is zero is part of the difficulty of the problem. The fact that these last conditions can be realized in 2^n different ways is at the origin of the complexity of the problem. It can be shown indeed that, even when the functions F and G are affine, finding a solution to (1.1) is NP-hard [24,68; 1989-1991]. The algorithms considered in this paper can be easily adapted to the *mixed nonlinear complementarity problem*, in which the number p of complementarity conditions is less than the number n of unknowns and there are $n - p$ additional nonlinear equality constraints. Less or more recent states of the art on the analysis of complementarity problems and numerical methods to solve them, in finite dimension, can be found in [79,61,83,43,28,29,63].

Occasionally, we shall make reference to the *linear complementarity problem* (LCP) in its standard form, which reads

$$0 \leqslant (Mx + q) \perp x \geqslant 0, \tag{1.2}$$

where the unknown is $x \in \mathbb{R}^n$, while $q \in \mathbb{R}^n$ and $M \in \mathbb{R}^{n \times n}$ are data. In that case, we shall consider that it corresponds to the nonlinear complementarity problem (1.1) with the affine map $F : x \mapsto Mx + q$ and the identity operator $G : x \mapsto x$.

Complementarity conditions arise spontaneously in the first order optimality conditions of an optimization problem with inequality constraints and these conditions can be written as a mixed nonlinear complementarity problem. The complementarity system (1.1) is also often used to model in part problems in which several systems of equations are, to some extend, in competition. The one that is active in a given place and at a given time, corresponding to a common index of $F(x)$ and $G(x)$, depends on threshold effects; if the threshold $F_i(x) = 0$ is not reached, i.e., $F_i(x) > 0$, then the equation $G_i(x) = 0$ is active, and vice versa. Examples include problems in nonsmooth mechanics and dynamics [5,1,18], the phase transition problem in multiphase flows [77,78,14,7,10,20,30,9], precipitation-dissolution problems in chemistry [19,70], portfolio management in finance [52], computer graphics [42], discrete Hamilton–Jacobi–Bellman equation solvers [94], meteorology simulation, economic equilibrium, to mention a few. Surveys on examples of applications of the complementarity problem can be found in [58,61,83,46,43].

1.2 A few linearization algorithms

Many techniques have been proposed to solve (1.1) since the problem was introduced by Cottle in his PhD thesis, dated 1964 [26,27]. It is beyond the scope of this paper to review all of them and we refer instead the interested reader to the recent monographs [43,63]. Below, we limit our account to the algorithms in close connection with the numerical methods proposed and analyzed in this paper. The motivation is to put in perspective the proposed algorithms, essentially within the Newton-min family of methods. On the way, we introduce notation and concepts used throughout the paper.

The adjacent numerical methods are related to the Newton algorithm to solve the nonsmooth system of equations

$$H(x) = 0, \tag{1.3a}$$

in which $H : \Omega \to \mathbb{R}^n$ is the function defined at $x \in \Omega$ by

$$H(x) := \min(F(x), G(x)), \tag{1.3b}$$

where the minimum is taken componentwise [2,82]. It is clear that problems (1.1) and (1.3) have the same solutions, since, for two real numbers a and b, $\min(a, b) = 0$ if and only if $a \geqslant 0$, $b \geqslant 0$ and $ab = 0$ (for other functions having that property, see [75,51,4] and the references therein). The term "Newton-min" was coined in [11,12,13] to name this solution strategy and we adopt it in this paper. The proposed methods are globalized by using the classical merit function associated with H [80, 36,16,17], which is the least-square function $\theta : \Omega \to \mathbb{R}$ defined at $x \in \Omega$ by

$$\theta(x) := \frac{1}{2}\,\|H(x)\|^2 = \frac{1}{2}\,\|\min(F(x), G(x))\|^2, \tag{1.4}$$

where $\|\cdot\|$ denotes the Euclidean norm. The goal of this paper is to focus on the reformulation (1.3) and its globalization, using linesearch on the natural merit function (1.4). More is said on the proposed approaches in section 1.3 below, after the presentation of some related linearization methods.

Many other equation reformulations of the complementarity problem have been proposed, see [75,32,66,65,91,21,33,48,44,85,60] and the references therein. Our choice of a reformulation by the minimum function is not only motivated by an intellectual curiosity (as we shall see, there are still holes in its implementation and its analysis), but also by its observed efficiency. This one is sometimes explained by the piecewise affine nature of the minimum function, which provides no additional nonlinearity besides its nondifferentiability. From a theoretical point of view, the required regularity at the solution to guarantee fast local convergence of a Newton-like algorithm on (1.3) is also less restrictive than with the Fischer reformulation [47], for instance; in addition, this algorithm has finite termination for the linear complementarity problem (1.2) [49], which cannot be expected when the reformulation is more nonlinear [43; § 9.2].

A first linearization method to solve (1.1) consists in applying Josephy-Newton (JN) iterations [64] on a functional inclusion reformulation of the problem [67] (see [43; § 7.3] for a reformulation using the normal map). This results in linearizing the functions in (1.1b) while keeping its complementarity problem structure: the new iterate $x+d$, following the current one x, is determined by taking for d an appropriate solution to the linear complementarity problem in d (if this solution exists)

$$0 \leqslant \big(F(x) + F'(x)d\big) \perp \big(G(x) + G'(x)d\big) \geqslant 0. \tag{1.5}$$

The SQP algorithm in nonlinear optimization can also be derived from this technique [64], so that the two methods have common features. The local quadratic convergence of this algorithm can be deduced from the one of the JN iterations for a functional inclusion (Josephy [64] assumes that the sought solution is strongly regular in the sense of Robinson [88], while Bonnans [15] only assumes the weaker so-called semistability and hemistability; see also [43; § 7.3] for related results). The globalization of this linearization approach for complementarity problems uses adapted merit functions (see [74] for an entry point). The JN approach has many attractive features, but, with respect to the methods proposed in section 1.3, the system (1.5) has the inconvenient of requiring the computation of a solution to a linear complementarity problem of dimension n at each iteration and we have already mentioned that such a problem is NP-hard. We also point out that this approach is not relevant in the case when the original problem (1.1) is a linear complementarity problem, since then (1.5) is exactly the same problem as the original one.

Another linearization approach to solve (1.1) consists in applying a Newton-like method to solve directly the equivalent nonsmooth system (1.3). Among these methods, one finds the B-Newton algorithm [81], which is adapted to B-differentiable maps [35,89,90]. For a locally Lipschitz function defined on a space of finite dimension, like H in (1.3b), the B-derivative is identical to the directional derivative [89,90], so that the direction d giving the new iterate $x+d$ in the B-Newton algorithm is taken as a solution (if any) to the (usually nonlinear) system

$$H(x) + H'(x;d) = 0, \tag{1.6}$$

where $H'(x;d) := \lim_{t\downarrow 0}[H(x+td) - H(x)]/t$ is the usual one-side directional derivative. It is easy to see that the function H given by (1.3b) is directionally differentiable (recall that F and G are supposed to be smooth) and that its directional derivative is given by

$$H_i'(x;d) = \begin{cases} \min(F_i'(x)d, G_i'(x)d) & \text{if } i \in \mathcal{E}(x), \\ F_i'(x)d & \text{if } i \in \mathcal{F}(x), \\ G_i'(x)d & \text{if } i \in \mathcal{G}(x), \end{cases} \tag{1.7}$$

where we have used the following mnemonic notation for index sets, which will be frequently encountered below:

$$\begin{aligned} \mathcal{E}(x) &:= \{i \in [1:n] : F_i(x) = G_i(x)\}, \\ \mathcal{F}(x) &:= \{i \in [1:n] : F_i(x) < G_i(x)\}, \\ \mathcal{G}(x) &:= \{i \in [1:n] : F_i(x) > G_i(x)\}. \end{aligned} \tag{1.8}$$

Combining (1.6), (1.3b) and (1.7), we see that the search direction d of the *B-Newton-min algorithm* is determined as a solution (if any) to the system

$$\begin{cases} (F(x) + F'(x)d)_{\mathcal{F}(x)} = 0, \\ (G(x) + G'(x)d)_{\mathcal{G}(x)} = 0, \\ 0 \leqslant (F(x) + F'(x)d)_{\mathcal{E}(x)} \perp (G(x) + G'(x)d)_{\mathcal{E}(x)} \geqslant 0. \end{cases} \tag{1.9}$$

Note that a solution to (1.5) may not be a solution to (1.9) (because $(1.9)_1$ and $(1.9)_2$ may not hold) and vice versa (because $(F(x) + F'(x)d)_{\mathcal{G}(x)} \geqslant 0$ and $(G(x) + G'(x)d)_{\mathcal{F}(x)} \geqslant 0$ may not hold). An interesting asset of the B-Newton-min approach, compared to the JN algorithm, is that the system (1.9) can be much easier to solve than (1.5), since its number $|\mathcal{E}(x)|$ of complementarity conditions is reduced to the number of indices i giving the equality $F_i(x) = G_i(x)$ at the current x and that this number can be very small. The convergence properties of this algorithm based on (1.9) derive from the one of the B-Newton algorithm (1.6) for solving the equation $H(x) = 0$, with a B-differentiable function H. According to [81; theorem 3], the algorithm converges when the first iterate is in some neighborhood of a zero x_* of H at which H is strongly Fréchet differentiable with a nonsingular $H'(x_*)$; this required smoothness assumption on H is awkward and rather restrictive when one aims at solving a nonsmooth system. Another interesting asset of the B-Newton direction d is that it is a descent direction of θ at x [81; lemma 1], which gives rise to a linesearch algorithm, generating sequences whose accumulation points x_* are solutions to (1.3a), provided H is strongly Fréchet differentiable at x_* and $H'(x_*)$ is injective [81; theorem 4(iii)]; these are again rather restrictive assumptions. In terms of the data of problem (1.1), when G is the identity, these conditions are guaranteed if the accumulation point x_* is regular in the sense of [81; definition 2] and $(x_*)_i = F_i(x_*) = 0$ for $i \in \mathcal{E}(x_*)$ [81; theorem 6]. Finally, we point out that the B-Newton-min is not appropriate to solve the linear complementarity problem (1.2), since (1.9) is identical to the original problem when $\mathcal{E}(x) = [1:n]$.

The B-Newton-min algorithm is modified in [82] in order to obtain convergence results with less demanding assumptions and the modification is shown in [57] to be part of a larger family of globally convergent algorithms for solving a nonsmooth system $H(x) = 0$. In the case of problem (1.1), the modified B-Newton-min algorithm consists in computing the new iterate $x + d$, from the current one x, by determining d as a solution (if any) to the nonlinear system [57; (4)]

$$H(x) + D(x, d) = 0, \tag{1.10}$$

where $D : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is no longer the directional derivative of H like in (1.6)-(1.7) but is defined by [57; (12)]

$$D_i(x, d) = \begin{cases} F_i'(x)d & \text{if } F_i(x) < G_i(x),\ G_i(x) \geqslant 0, \\ G_i'(x)d & \text{if } F_i(x) > G_i(x),\ F_i(x) \geqslant 0, \\ \min(F_i'(x)d, G_i'(x)d) & \text{otherwise.} \end{cases} \tag{1.11}$$

In comparison with (1.6), we see that some indices of $\mathcal{F}(x)$ and $\mathcal{G}(x)$ are now handled like those of $\mathcal{E}(x)$. Rewriting (1.10), with the form of H from (1.3b) and that of D from (1.11), we see that d has to solve the system

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } F_i(x) < G_i(x),\ G_i(x) \geqslant 0, \\ G_i(x) + G_i'(x)d = 0 & \text{if } F_i(x) > G_i(x),\ F_i(x) \geqslant 0, \\ 0 \leqslant (F_i(x) + F_i'(x)d) \perp (F_i(x) + G_i'(x)d) \geqslant 0 & \text{if } F_i(x) < G_i(x) < 0, \\ 0 \leqslant (G_i(x) + F_i'(x)d) \perp (G_i(x) + G_i'(x)d) \geqslant 0 & \text{if } 0 > F_i(x) > G_i(x), \\ 0 \leqslant (F_i(x) + F_i'(x)d) \perp (G_i(x) + G_i'(x)d) \geqslant 0 & \text{otherwise.} \end{cases} \tag{1.12}$$

This heterogeneous system has therefore more complementarity conditions than (1.9), but has also better convergence results. Conditions ensuring the existence and uniqueness of the solution to the mixed linear complementarity problem (1.12) can be obtained [82; § 5]. Furthermore, it can be shown that this direction d is a descent direction of θ at x, which gives rise to a linesearch algorithm whose global convergence (without the previously required smoothness of H) and the admissibility of the unit stepsize are studied in [82; §§ 6-8]. For the same reason as for the B-Newton-min algorithm, the present modification is not appropriate for linear complementarity problem (LCP), since the system (1.12) is identical to the original problem when $\mathcal{E}(x) = [1 : n]$.

A more drastic approach to solve a nonsmooth system $H(x) = 0$ is to use the semismooth Newton method [87,86], provided H is semismooth, which is the case of the function defined by (1.3b) when F and G are smooth. This method only requires to solve a linear system per iteration: one chooses a Jacobian J_x in the generalized Clarke differential $\partial_C H(x)$ of H at x [25] and defines the displacement d at x as a solution (if any) to

$$H(x) + J_x d = 0. \tag{1.13}$$

Despite its poor description of the function H at a point of nondifferentiability, this method has the remarkable property of having a superlinear speed of convergence (or quadratic, if H is strongly semismooth), when the first iterate is close enough to a *regular point* x_* of H, which means here that all the Jacobians of $\partial_C H(x_*)$ are nonsingular [43,63]. A drawback of this method is that it is often difficult to compute an element of $\partial_C H(x)$, for a particular function H, because this generalized Jacobian is not known or evaluating one of its elements is computationally expensive. Nevertheless, one can sometimes use a surrogate of the generalized Jacobian J_x in (1.13), while keeping the fast local convergence property of the pure approach (see [56,71] for the projection on a convex polyhedron) and for the function H given by (1.3), one can use the inexpensive central Jacobian of Xiang and Chen [93; theorem 2.2]. A drawback of the semismooth Newton direction, however, is that it is not necessarily a descent direction of the natural least-square merit function θ (see counter-example 2.4 below, for a linear complementarity problem), which explains why it is difficult to define a globally convergent algorithm based on this direction and the merit function (1.4).

A method inspired from the semismooth Newton algorithm or from [69], applied to (1.3) (algorithm 7.2.17 in [43]), computes the displacement d from x to the next iterate $x+d$ by solving (if possible) the linear system

$$\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{F}}(x), \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{G}}(x), \end{cases} \tag{1.14}$$

where the pair $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x))$ forms a partition of $[1:n]$ and satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$. This method differs from the semismooth Newton approach in that the matrix used in the system (1.14), namely

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}(x)}(x) \\ G'_{\tilde{\mathcal{G}}(x)}(x) \end{pmatrix},$$

may not be in the C-differential $\partial_C H(x) = \operatorname{co} \partial_B H(x)$ (or convex hull of the B-differential) of H at x [40]. This economical approach has the same drawback as the semismooth Newton direction (1.13), which is that its directions are not necessarily descent directions of the natural least-square merit function θ, because of an inappropriate choice of the indices of $\mathcal{E}(x)$ going into $\tilde{\mathcal{F}}(x)$ and $\tilde{\mathcal{G}}(x)$ (see again counter-example 2.4 below).

Finally, we quote the algorithm of [84], which uses the merit function (1.4) and computes its directions by solving a piecewise quadratic convex function subject to linear constraints, for a complementarity problem of the form $0 \leqslant F(x) \perp x \geqslant 0$. Note also that there are other approaches, which use the least-square merit function and the Fischer complementarity function [45, 72, 33, 85].

1.3 A foretaste of the proposed algorithms

The methods proposed and analyzed in this paper are progressively introduced in section 2, but we can already give here a foretaste of their nature. They find their place in the panorama of linearization methods of the minimum function (1.3b) presented in the previous section, in the sense that their directions can be viewed as intermediates between the B-Newton direction d given by (1.9), or its modification given by (1.12), and the semismooth-like direction computed by (1.14), called the *plain Newton-min* direction in section 2.1. Their main advantage is to avoid the need of solving an LCP at each iteration, hence unlike in (1.9) or (1.12), and to guarantee global convergence, hence unlike (1.14).

Instead of having to solve an LCP, the direction must satisfy a system, made of affine equalities and (generally very few) *inequalities*, in order to guarantee the descent of the least-square merit function θ, defined in (1.4); see section 2.2. A least-norm displacement of this system can, for example, be obtained by solving a convex quadratic optimization problem, which can be done in polynomial time. An improvement of this direction is needed, however, to guarantee convergence in the sense and with the technique of proof presented in section 3.2: the set of inequalities defining the direction must be slightly enlarged when the iterate is near a "negative kink" of H (we call a *kink* a locus of points of nondifferentiability); see section 2.3.

Like any linearization algorithm with linesearch, convergence is restricted by a regularity assumption of the limit point. This notion of regularity depends on the computed direction. This issue is analyzed with care in section 3.1. Next, a global convergence result is given in section 3.2. Finally, to avoid these more expensive directions, due to the presence of inequalities in their definition, a hybrid algorithm is proposed in section 3.3, in which the plain Newton-min direction (1.14) is first tested for descent. This hybrid method is shown to possess both global and fast (superlinear or quadratic) local convergence. The paper ends with the conclusion section 4.

The design of the algorithms presented in this paper has been oriented by an intensive numerical exploration on LCPs, which has shown that the proposed method is competitive with other solvers on various applications, on some reference academic examples, and on randomly generated problems. These experiments are reported in [50] for the linear complementarity problem (1.2).

This paper is an abridged version of the more detailed report [38].

1.4 Notation and definition

We denote by $\|\cdot\|$ the Euclidean norm and by $|\!|\!|\cdot|\!|\!|$ an arbitrary norm, both on $\mathbb{R}^n$. The cardinality of a set S (i.e., its number of elements, which will be always finite) is denoted by $|S|$. The set of partitions of $[1:n]$ is denoted by $\mathcal{P}([1:n])$.

We say that a function is *continuously differentiable at* x if it is differentiable near (i.e., in a neighborhood of) x and its derivative is continuous at x.

2 Polyhedral Newton-min directions

This section introduces the directions of the proposed algorithms. It proceeds gradually, insisting on the motivation, which is to obtain descent directions of θ and to guarantee some global convergence property. We first observe that the plain Newton-min (NM) direction of section 2.1, already presented in (1.14) and obtained by solving a single linear system, is not necessarily a descent direction of θ (counter-example 2.4). We then examine in section 2.2 the reason of this descent property failure and propose a descent direction (proposition 2.5), which must satisfy a similar system as the one of the plain NM direction, but whose equations corresponding to the indices in $\{i \in [1:n] : F_i(x) = G_i(x) < 0\}$ are transformed into pairs of inequalities. This yields what we call a *polyhedral Newton-min* (PNM) direction since this one must be a feasible point of a certain polyhedron. This *plain PNM direction* is always a descent direction of θ. Nevertheless, it did not allowed us to prove the global convergence result of theorem 3.6 for a reason discussed in section 2.3. It seems important, indeed, that, when the current iterate is near "negative kinks" of H, the direction is built by picking information on the behavior of the function H on both sides of the kink. This leads us to propose in section 2.3.1 the *secure* PNM direction (2.13), whose definition depends on the proximity of the current iterate to these special kinks of H. Its descent property is viewed in section 2.3.2 as a consequence of proposition 2.7, which analyzes the potential descent property of a direction by averaging its effect on each term $H_i(x)^2$ defining the merit function θ. Section 2.3.2 also introduces the very permissive *inexact secure* PNM direction (2.22), for which descent property and global convergence hold, which is expensive to compute, but the inequalities in its definition can be used as stopping test during the computation of a secure PNM direction. We conclude with section 2.4, which presents a *generic PNM algorithm* (algorithm 2.10) and a possible implementation of the *PNM algorithm* (algorithm 2.11).

2.1 Plain Newton-min direction

The *plain Newton-min* (NM) *algorithm* is a semismooth Newton-like method on the reformulation (1.3) of the nonlinear complementarity problem (1.1), which uses the minimum function (algorithm 7.2.17 in [43]). It computes its direction d at $x \in \Omega$ by solving the linear system (1.14), which is reproduced here for the reader's convenience:

$$\boxed{\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{F}}(x), \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in \tilde{\mathcal{G}}(x). \end{cases}} \tag{2.1}$$

In this system, $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x)) \in \mathcal{P}([1:n])$ and satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$. By the "symmetry" in F and G of the complementarity problem (1.1), there is no natural reason to put all the indices of $\mathcal{E}(x)$ in $\tilde{\mathcal{F}}(x)$ or $\tilde{\mathcal{G}}(x)$, which motivates the flexibility admitted in the direction definition (2.1). We see that, at a point x on a possible kink of H, due to one of its components $i \in \mathcal{E}(x)$, a pseudo-derivative of H_i at x is chosen in $\{F_i'(x), G_i'(x)\}$.

To identify the points x at which the linear system (2.1) is guaranteed to have a solution, we introduce the notion of *NM-regularity*. This notion is linked to the plain NM algorithm (hence the prefix NM), like the nonsingularity of the Jacobian of a nonlinear system is a regularity assumption linked to Newton's method. In the following definition, we do not assume that the considered point x is a solution to the complementarity problem (1.1), which is motivated by the fact that this regularity assumption will be required at accumulation points that are not known a priori to be solutions to the problem (see the proof of theorem 3.9).

Definition 2.1 (NM-regularity) A point $x \in \mathbb{R}^n$ is said to be *NM-regular* (we also say that the complementarity problem (1.1) is *NM-regular* at $x \in \mathbb{R}^n$) if F and G are differentiable at x and if, for any $(\tilde{\mathcal{F}}, \tilde{\mathcal{G}}) \in \mathcal{P}([1:n])$ satisfying $\tilde{\mathcal{F}} \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}} \supseteq \mathcal{G}(x)$, the Jacobian

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}}(x) \\ G'_{\tilde{\mathcal{G}}}(x) \end{pmatrix} \tag{2.2}$$

is nonsingular. □

When G is the identity and x is nonnegative, this is a notion slightly weaker than the b-regularity of [84; definition 2] (any set $\tilde{\mathcal{F}}$ lying between $\{i \in \mathcal{F}(x) : G_i(x) \equiv x_i > 0\}$, which is smaller than $\mathcal{F}(x)$, and $\mathcal{G}(x)$ can be chosen in [84]); moreover, when x is also a solution to the complementarity problem (1.1), this is the notion of b-regularity of [43; definition 3.3.10]. The NM-regularity is stronger than the *strong BD-regularity* introduced in [86] to ensure the local and superlinear convergence of the semismooth Newton algorithm (when the Newton operator is taken in $\partial_B H(x)$), because the NM algorithm offers more freedom than the semismooth Newton algorithm in the choice of its Jacobians. Indeed, the NM-regularity amounts to requiring the nonsingularity of the Jacobians in the Cartesian product $\partial_B^\times H(x) := \partial_B H_1(x) \times \cdots \times \partial_B H_n(x)$ of the B-differentials of the components of H, while the strong BD-regularity means that the Jacobians of $\partial_B H(x)$ must be nonsingular; now use $\partial_B H(x) \subseteq \partial_B^\times H(x)$ to conclude. Note that the latter inclusion is strict only if the vectors $\{G'_i(x) - F'_i(x) : i \in \mathcal{E}(x),\ G'_i(x) \neq F'_i(x)\}$ are linearly dependent [40; proposition 4.2] [41] [93; corollary 2.1(i)].

The next proposition gives some consequences of the NM-regularity. The first property claims that the NM-regularity diffuses to the neighboring points. The second property will be useful for establishing the global convergence result of theorem 3.9 (see [84; lemma 3] for a similar property).

Proposition 2.2 (NM-regularity properties) *Suppose that F and G are continuously differentiable at $\bar{x} \in \mathbb{R}^n$ and that $\bar{x}$ is NM-regular. Then, there is a neighborhood V of $\bar{x}$ and a constant $C \geqslant 0$, such that, for all $x \in V$:*

1) *x is NM-regular,*
2) *the system (2.1) has a unique solution d and the norm of d is bounded by C.*

Proof By their differentiability property, F and G are continuous at $\bar{x}$. Furthermore, $\mathcal{F}(\bar{x})$ and $\mathcal{G}(\bar{x})$ are finite sets. Then, it follows that there is a neighborhood V_1 of $\bar{x}$ such that

$$\forall\, x \in V_1 : \qquad \mathcal{F}(\bar{x}) \subseteq \mathcal{F}(x) \quad \text{and} \quad \mathcal{G}(\bar{x}) \subseteq \mathcal{G}(x). \tag{2.3}$$

1) Let $(\tilde{\mathcal{F}}, \tilde{\mathcal{G}}) \in \mathcal{P}([1:n])$ satisfying $\tilde{\mathcal{F}} \supseteq \mathcal{F}(\bar{x})$ and $\tilde{\mathcal{G}} \supseteq \mathcal{G}(\bar{x})$. By the NM-regularity at $\bar{x}$,

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}}(\bar{x}) \\ G'_{\tilde{\mathcal{G}}}(\bar{x}) \end{pmatrix} \text{ is nonsingular.}$$

By the Banach perturbation lemma, there is a neighborhood $V_{\tilde{\mathcal{F}},\tilde{\mathcal{G}}} \subseteq V_1$ of $\bar{x}$ and a constant $C_{\tilde{\mathcal{F}},\tilde{\mathcal{G}}}$, such that for all $x \in V_{\tilde{\mathcal{F}},\tilde{\mathcal{G}}}$,

$$\begin{pmatrix} F'_{\tilde{\mathcal{F}}}(x) \\ G'_{\tilde{\mathcal{G}}}(x) \end{pmatrix} \text{ is nonsingular and } \left\| \begin{pmatrix} F'_{\tilde{\mathcal{F}}}(x) \\ G'_{\tilde{\mathcal{G}}}(x) \end{pmatrix}^{-1} \right\| \leqslant C_{\tilde{\mathcal{F}},\tilde{\mathcal{G}}}$$

Define

$$V_2 := \bigcap_{\substack{(\tilde{\mathcal{F}},\tilde{\mathcal{G}})\in\mathcal{P}([1:n])\\ \tilde{\mathcal{F}}\supseteq\mathcal{F}(\bar{x})\\ \tilde{\mathcal{G}}\supseteq\mathcal{G}(\bar{x})}} V_{\tilde{\mathcal{F}},\tilde{\mathcal{G}}} \subseteq V_1 \qquad\text{and}\qquad C_1 := \sup_{\substack{(\tilde{\mathcal{F}},\tilde{\mathcal{G}})\in\mathcal{P}([1:n])\\ \tilde{\mathcal{F}}\supseteq\mathcal{F}(\bar{x})\\ \tilde{\mathcal{G}}\supseteq\mathcal{G}(\bar{x})}} C_{\tilde{\mathcal{F}},\tilde{\mathcal{G}}}$$

Since the number of partitions $(\tilde{\mathcal{F}},\tilde{\mathcal{G}}) \in \mathcal{P}([1:n])$ is finite, V_2 is a neighborhood of $\bar{x}$ and $C_1 < \infty$. Therefore,

$$\begin{gathered}\forall\, x \in V_2,\ \forall\,(\tilde{\mathcal{F}},\tilde{\mathcal{G}}) \in \mathcal{P}([1:n]) \text{ satisfying } \tilde{\mathcal{F}} \supseteq \mathcal{F}(\bar{x}) \text{ and } \tilde{\mathcal{G}} \supseteq \mathcal{G}(\bar{x}) :\\ \begin{pmatrix} F'_{\tilde{\mathcal{F}}}(x)\\ G'_{\tilde{\mathcal{G}}}(x)\end{pmatrix} \text{ is nonsingular and } \left\| \begin{pmatrix} F'_{\tilde{\mathcal{F}}}(x)\\ G'_{\tilde{\mathcal{G}}}(x)\end{pmatrix}^{-1} \right\| \leqslant C_1.\end{gathered} \tag{2.4}$$

Suppose now that $x \in V_2$ and that $(\tilde{\mathcal{F}},\tilde{\mathcal{G}}) \in \mathcal{P}([1:n])$ satisfies $\tilde{\mathcal{F}} \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}} \supseteq \mathcal{G}(x)$. By (2.3) and $V_2 \subseteq V_1$, $(\tilde{\mathcal{F}},\tilde{\mathcal{G}})$ satisfies $\tilde{\mathcal{F}} \supseteq \mathcal{F}(\bar{x})$ and $\tilde{\mathcal{G}} \supseteq \mathcal{G}(\bar{x})$. By (2.4), the matrix (2.2) is nonsingular. Hence x is NM-regular.

2) By restricting the neighborhood V_2 to a neighborhood V of $\bar{x}$, in order to have $F(x)$ and $G(x)$ bounded in norm by C_2 on V (this is possible by the continuity of F and G at $\bar{x}$), we see that, using the bound C_1 on the matrix inverse in (2.4), for any $x \in V$, the direction d is uniquely defined by (2.1) and is also bounded by $C := C_1C_2$. □

In section 3.3, we will need the following result on the fast local convergence of the plain NM algorithm of section 2.1, which is transcribed from [43; theorem 7.2.18]. Recall also that a sequence $\{x_k\} \subseteq \mathbb{R}^n$ converges *superlinearly* to $\bar{x} \in \mathbb{R}^n$ if there exists a sequence of positive number ε_k converging to zero such that $\|x_{k+1} - \bar{x}\| \leqslant \varepsilon_k \|x_k - \bar{x}\|$ for all k, and that it converges *quadratically* to $\bar{x}$ if there exists a constant $C > 0$ such that $\|x_{k+1} - \bar{x}\| \leqslant C\|x_k - \bar{x}\|^2$ for all k. These last two properties do not depend on the choice of the norm $\|\cdot\|$ on $\mathbb{R}^n$. By *finite convergence* we mean convergence in a finite number of iterations.

Theorem 2.3 (fast local convergence of the NM algorithm) *Suppose that F and $G : \Omega \to \mathbb{R}^n$ are continuously differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Let $\bar{x} \in \Omega$ be a solution to (1.1) that is NM-regular. Then, there is an open neighborhood $V \subseteq \Omega$ of $\bar{x}$, such that, if the plain NM algorithm 2.1 starts at a point in V, the following properties hold.*

1) *A sequence is generated in V, which converges superlinearly to $\bar{x}$.*
2) *If F' and G' are Lipschitz on V, the convergence is quadratic.*
3) *If F and G are affine on V, the convergence is finite.*

The plain NM direction is very attractive since it can be computed by solving a single linear system and because it guarantees a local quadratic convergence (theorem 2.3). Unfortunately, this direction may not be a descent direction of the least-square merit function θ defined by (1.4), although this one is naturally associated with the system (1.3). Here is a counter-example of this phenomenon in the case of a linear complementarity problem (1.2) with a **P**-matrix M (recall that a square matrix M is a **P**-matrix if its principal minors are positive; a property that is denoted by "$M \in \mathbf{P}$"). This fact was already observed during the preparation of the PhD thesis of I. Ben Gharbia [8; example 5.8].

Counter-examples 2.4 (no descent direction from (2.1)) Consider the linear complementarity problem (1.2) in dimension $n = 2$ and the point x given by

$$M = \begin{pmatrix} 1 & \mu\\ 0 & 1\end{pmatrix}, \qquad q = \begin{pmatrix} -\mu\\ -2\end{pmatrix} \qquad\text{and}\qquad x = \begin{pmatrix} -2\\ 1\end{pmatrix}, \tag{2.5}$$

where $\mu > 0$ is a positive parameter. Note that $M \in \mathbf{P}$. Since $F(x) \equiv Mx+q = (-2,-1)$ and $G(x) \equiv x = (-2,1)$, the index sets (1.8) read $\mathcal{E}(x) = \{1\}$, $\mathcal{F}(x) = \{2\}$ and $\mathcal{G}(x) = \varnothing$. If one computes the

NM direction d by (2.1) with $\tilde{\mathcal{F}}(x) = \{1,2\}$ and $\tilde{\mathcal{G}}(x) = \varnothing$, one gets $d = -x - M^{-1}q = (2-\mu, 1)$. Then, for $t \geqslant 0$:

$$\theta(x+td) = \frac{(5-4\mu+\mu^2)t^2 + 2(2\mu-5)t + 5}{2} \qquad \text{and} \qquad \theta'(x;d) = 2\mu - 5,$$

which shows that the chosen NM direction d is an ascent direction of θ at x, provided $\mu > 5/2$. The figure 2.1 below gives the level curves of θ, which highlight the nonsmoothness and nonconvexity of

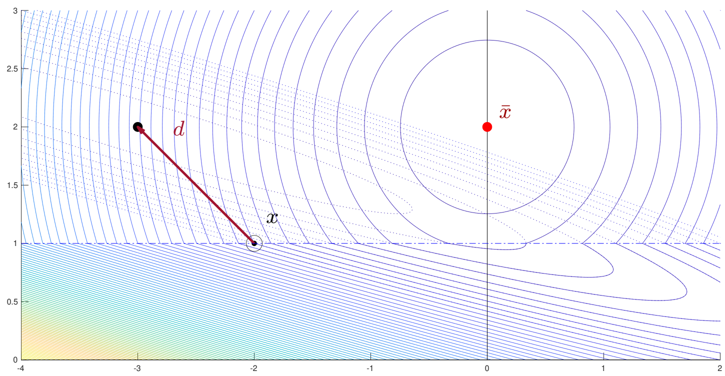

Fig. 2.1 The level curves of θ in counter-example 2.4 with $\mu = 3$ (and the dotted level curves of $x \mapsto \frac{1}{2}\|Mx+q\|^2$), the solution $\bar{x}$ to the LCP, the current point x and the unfortunate NM direction d, which is an ascent direction of θ.

the least-square merit function, as well as the chosen NM direction d, along which θ clearly increases. The increase of θ along the chosen NM direction is due to an unfortunate choice of $\tilde{\mathcal{F}}(x)$ and $\tilde{\mathcal{G}}(x)$. If one chooses the index sets $\tilde{\mathcal{F}}(x) = \{2\}$ and $\tilde{\mathcal{G}}(x) = \{1\}$, the solution to (2.1) becomes $d = (2,1)$, which is also the solution to the linear complementarity problem (1.9), and $x + d = (0,2)$ is the solution to the LCP.

Note also that the matrices used in (2.1) to compute the directions $d = (2-\mu, 1)$ and $d = (2,1)$ above, namely M and I, are both in the B-differential $\partial_B H(x)$ of H at x; see [38,40]. Therefore, the belonging of the Jacobian (2.2) to $\partial_B H(x)$ is not a guarantee to get the descent property. To put it more synthetically, a semismooth Newton direction of the form (1.13) may not be a descent direction of the least-square merit function θ.

To conclude, note that the semismooth Newton direction (1.13) with the Xiang-Chen central Jacobian [93; (2.4)], which is in the C-differential of H at x and reads here

$$J_{1/2} := \frac{1}{2}(M+I) = \begin{pmatrix} 1 & \mu/2 \\ 0 & 1 \end{pmatrix},$$

is also not a descent direction of θ, when $\mu > 5$; see [38]. □

To the best of our knowledge, this intrinsic difficulty of the plain NM algorithm has not been considered with full attention (we quote, however, algorithm 9.2.2 in [43], which requires to solve a convex piecewise quadratic optimization problem at each iteration with n bound constraints and is therefore more expensive than the algorithms proposed below). In sections 2.2 and 2.3, we propose to overcome the difficulty by imposing the direction to be a feasible point of a particular convex polyhedron, defined by a very small number of linear inequalities, instead of being the solution to a linear system. The computation of these directions is therefore more expensive than for the plain NM direction, but remains polynomial. In addition, in section 3.3, a heuristics is proposed to avoid as much as possible the need to find a point in a polyhedron.

2.2 Plain polyhedral Newton-min direction

The direction proposed in this section is based on the following computation, which highlights the reason why a plain NM direction may not be a descent direction of the least-square merit function θ defined in (1.4).

First, observe that the map θ is directionally differentiable as a composition of H, which is directionally differentiable, and $\frac{1}{2}\|\cdot\|^2$ which is locally Lipschitz continuous and smooth. In this case, the chain rule applies (see [17; lemma 11.1] for example):

$$\theta'(x;d) = H(x)^{\mathsf{T}} H'(x;d).$$

From (1.3b) and (1.7), one gets

$$\begin{aligned}\theta'(x;d) = {} & F_{\mathcal{F}(x)}(x)^{\mathsf{T}} F'_{\mathcal{F}(x)}(x)d + G_{\mathcal{G}(x)}(x)^{\mathsf{T}} G'_{\mathcal{G}(x)}(x)d \\ & + F_{\mathcal{E}(x)}(x)^{\mathsf{T}} \min(F'_{\mathcal{E}(x)}(x)d, G'_{\mathcal{E}(x)}(x)d).\end{aligned} \tag{2.6}$$

Since, for $\tilde{x}$ near x, $H_{\mathcal{F}(x)}(\tilde{x}) \equiv F_{\mathcal{F}(x)}(\tilde{x})$ and $H_{\mathcal{G}(x)}(\tilde{x}) \equiv G_{\mathcal{G}(x)}(\tilde{x})$, it is natural to impose to a Newton-like direction d to verify

$$(F(x) + F'(x)d)_{\mathcal{F}(x)} = 0 \qquad \text{and} \qquad (G(x) + G'(x)d)_{\mathcal{G}(x)} = 0. \tag{2.7}$$

Note, however, that it will be necessary to infringe this rule below, in order to approach the "negative kinks" of H with caution. Using (2.6), (2.7), and $F_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x)$, the directional derivative $\theta'(x;d)$ becomes

$$\begin{aligned}\theta'(x;d) &= -\|F_{\mathcal{F}(x)}(x)\|^2 - \|G_{\mathcal{G}(x)}(x)\|^2 - \|F_{\mathcal{E}(x)}(x)\|^2 \\ &\quad + F_{\mathcal{E}(x)}(x)^{\mathsf{T}} \min\Big(F_{\mathcal{E}(x)}(x) + F'_{\mathcal{E}(x)}(x)d, G_{\mathcal{E}(x)}(x) + G'_{\mathcal{E}(x)}(x)d\Big) \\ &= -2\,\theta(x) + F_{\mathcal{E}(x)}(x)^{\mathsf{T}} \min\Big(F_{\mathcal{E}(x)}(x) + F'_{\mathcal{E}(x)}(x)d, G_{\mathcal{E}(x)}(x) + G'_{\mathcal{E}(x)}(x)d\Big).\end{aligned}$$

The first term in the right-hand side is satisfactory since it corresponds to the formula of the directional derivative of the least-square function when H is smooth, while the second term is at the origin of the positive directional derivative observed in counter-example 2.4. Let us dissect this last term in order to see what conditions the direction must verify to make it nonpositive (we take up again an observation already made during the preparation of the PhD thesis of I. Ben Gharbia [8; 2012] for the LCP (1.2)). For this, we introduce the following partition $(\mathcal{E}^-(x), \mathcal{E}^0(x), \mathcal{E}^+(x))$ of $\mathcal{E}(x)$, as well as the index set $\mathcal{E}^{0+}(x) := \mathcal{E}^0(x) \cup \mathcal{E}^+(x)$:

$$\begin{aligned}\mathcal{E}^-(x) &:= \{i \in [1:n] : F_i(x) = G_i(x) < 0\}, \\ \mathcal{E}^0(x) &:= \{i \in [1:n] : F_i(x) = G_i(x) = 0\}, \\ \mathcal{E}^+(x) &:= \{i \in [1:n] : F_i(x) = G_i(x) > 0\}, \\ \mathcal{E}^{0+}(x) &:= \{i \in [1:n] : F_i(x) = G_i(x) \geqslant 0\}.\end{aligned} \tag{2.8}$$

Let $i \in \mathcal{E}(x) = \mathcal{E}^{0+}(x) \cup \mathcal{E}^-(x)$.

- If $i \in \mathcal{E}^{0+}(x)$, then $F_i(x) \geqslant 0$. If one of the linearized functions $F_i(x) + F'_i(x)d$ or $G_i(x) + G'_i(x)d$ vanishes, their minimum is nonpositive, yielding $F_i(x) \min(F_i(x) + F'_i(x)d, G_i(x) + G'_i(x)d) \leqslant 0$.
- If $i \in \mathcal{E}^-(x)$, then $F_i(x) < 0$. To get $F_i(x) \min(F_i(x) + F'_i(x)d, G_i(x) + G'_i(x)d) \leqslant 0$, it is now necessary to have $\min(F_i(x) + F'_i(x)d, G_i(x) + G'_i(x)d) \geqslant 0$, meaning that the following inequalities must hold:

$$F_i(x) + F'_i(x)d \geqslant 0 \quad \text{and} \quad G_i(x) + G'_i(x)d \geqslant 0. \tag{2.9}$$

Therefore, the decrease of θ is ensured along a direction d if this one satisfies (2.7), either $F_i(x)+F_i'(x)d=0$ or $G_i(x)+G_i'(x)d=0$ when $i\in\mathcal{E}^{0+}(x)$, and both inequalities in (2.9) for $i\in\mathcal{E}^-(x)$.

The above discussion leads us to the definition of the following direction. Let us denote by $(\mathcal{E}^{0+}_{\mathcal{F}}(x),\mathcal{E}^{0+}_{\mathcal{G}}(x))$ an arbitrary partition of $\mathcal{E}^{0+}(x)$, meaning that

$$\mathcal{E}^{0+}(x)=\mathcal{E}^{0+}_{\mathcal{F}}(x)\cup\mathcal{E}^{0+}_{\mathcal{G}}(x) \qquad\text{and}\qquad \mathcal{E}^{0+}_{\mathcal{F}}(x)\cap\mathcal{E}^{0+}_{\mathcal{G}}(x)=\varnothing. \tag{2.10}$$

A *plain polyhedral Newton-min* (PNM) *direction* is a direction d that satisfies the following system

$$\boxed{\begin{cases} F_i(x)+F_i'(x)d=0 \ \text{ if } i\in\mathcal{F}(x)\cup\mathcal{E}^{0+}_{\mathcal{F}}(x)\\ G_i(x)+G_i'(x)d=0 \ \text{if } i\in\mathcal{G}(x)\cup\mathcal{E}^{0+}_{\mathcal{G}}(x)\\ F_i(x)+F_i'(x)d\geqslant 0 \ \text{ if } i\in\mathcal{E}^-(x)\\ G_i(x)+G_i'(x)d\geqslant 0 \ \text{if } i\in\mathcal{E}^-(x).\end{cases}} \tag{2.11}$$

Therefore, we have imposed inequality constraints on the linearized functions $F_i(x)+F_i'(x)d$ and $G_i(x)+G_i'(x)d$ for the indices in $i\in\mathcal{E}^-(x)$, like suggested by (2.9), rather than arbitrarily forcing one of them to vanish, like in the plain NM algorithm (2.1).

The computation of a plain PNM direction is more expensive than the computation of the plain NM direction (2.1), since a feasible point of a convex polyhedron must be found instead of the solution to a linear system. Nevertheless, a direction satisfying (2.11) can be computed in polynomial time using linear or quadratic optimization (see [34,22] and the references therein) or other approaches (see [23] for a polynomially convergent algorithm and [3,54,55,31] for the linearly convergent relaxation method). Such an extra cost is acceptable, even when one solves a linear complementarity problem. In the next section, we continue to explore this vein and in section 3.3, we introduce a way of reducing the cost of the direction computation that is very successful in practice.

We summarize the discussion of this section in the following proposition.

Proposition 2.5 (descent property with (2.11)) *For any direction d satisfying (2.11), one has $\theta'(x;d)\leqslant -2\,\theta(x)$.*

2.3 Secure polyhedral Newton-min direction

Although a vector d satisfying (2.11) is a descent direction of θ, we were not able to get a global convergence result like those of section 3.2 below with that direction. Our first attempt in [50] to get global convergence involved adding all indices from $\{i\in[1\!:\!n] : F_i(x)<G_i(x)<0\}$ and $\{i\in[1\!:\!n] : G_i(x)<F_i(x)<0\}$ in the inequalities in (2.11), which corresponds to the index set $\mathcal{E}^-_\infty(x)$ in (2.12c) below. This is clearly very costly, which motivates us to develop the more economic strengthening which we present now under the name *secure PNM direction*.

In the approach followed in the proof of theorem 3.6, on which theorems 3.7, 3.9 and 3.11 rest, a difficulty may arise with a limit point $\bar{x}$ for which $\mathcal{E}^-(\bar{x})\neq\varnothing$, which is likely to be on a kink of H, then called a *negative kink*. When an iterate x_k is close to such an $\bar{x}$ and $i\in\mathcal{F}(x_k)$ say (by symmetry, the reasoning is the same if $i\in\mathcal{G}(x_k)$), the system (2.11) gives an information on the variation of F_i at x_k along d_k (through the equation $F_i(x_k)+F_i'(x_k)d_k=0$) but nothing is said on the variation of G_i along the same direction (since $G_i(x_k)+G_i'(x_k)d_k$ may take any value), while an information on $G_i'(x_k)d_k$ may also be necessary when the linesearch at x_k explores the two sides of the kink. It happens, actually, that relaxing the equality $F_i(x_k)+F_i'(x_k)d_k=0$ into the inequality $F_i(x_k)+F_i'(x_k)d_k\geqslant 0$ and adding the inequality $G_i(x_k)+G_i'(x_k)d_k\geqslant 0$ suffice to complete the proof (see its point 4.1.2), while keeping the descent property (see corollary 2.8).

We first present in section 2.3.1 the *exact* version (2.13) of the direction described in the previous paragraph and discuss its links with directions proposed in other contributions. Next, we analyze its descent property in section 2.3.2 and exhibit the *inexact* version (2.22) of the direction, which also enjoys the descent property.

2.3.1 Direction

Based on the previous discussion, we introduce a device that is able to measure the proximity to a point $\bar{x}$ with a nonempty index set $\mathcal{E}^-(\bar{x})$. Let $\tau \in (0, \infty)$ be the *kink tolerance*, used to detect such a proximity (numerically τ should be taken small, but we want to be rather general at this stage of the presentation) and define the index set

$$\mathcal{E}_\tau^-(x) := \{i \in [1:n] : F_i(x) < 0,\ G_i(x) < 0,\ |F_i(x) - G_i(x)| < \tau\}. \tag{2.12a}$$

We also define

$$\mathcal{E}_0^-(x) := \cap_{\tau>0}\mathcal{E}_\tau^-(x) = \{i \in [1:n] : F_i(x) = G_i(x) < 0\} = \mathcal{E}^-(x), \tag{2.12b}$$

$$\mathcal{E}_\infty^-(x) := \cup_{\tau>0}\mathcal{E}_\tau^-(x) = \{i \in [1:n] : F_i(x) < 0,\ G_i(x) < 0\}. \tag{2.12c}$$

Note that the set $\mathcal{E}_\tau^-(x)$ is expanding with τ, meaning that $\mathcal{E}_{\tau_1}^-(x) \subseteq \mathcal{E}_{\tau_2}^-(x)$ when $0 \leqslant \tau_1 \leqslant \tau_2 \leqslant \infty$.

A direction d is said to be a *secure PNM direction* if it satisfies

$$\boxed{\begin{cases} F_i(x) + F_i'(x)d = 0 & \text{if } i \in E_F(x) \\ G_i(x) + G_i'(x)d = 0 & \text{if } i \in E_G(x) \\ F_i(x) + F_i'(x)d \geqslant 0 & \text{if } i \in I(x) \\ G_i(x) + G_i'(x)d \geqslant 0 & \text{if } i \in I(x), \end{cases}} \tag{2.13}$$

where we have used the following index sets:

$$E_F(x) := \left[\mathcal{F}(x) \setminus \mathcal{E}_\tau^-(x)\right] \cup \mathcal{E}_{\mathcal{F}}^{0+}(x), \tag{2.14a}$$

$$E_G(x) := \left[\mathcal{G}(x) \setminus \mathcal{E}_\tau^-(x)\right] \cup \mathcal{E}_{\mathcal{G}}^{0+}(x), \tag{2.14b}$$

$$I(x) := \mathcal{E}_\tau^-(x), \tag{2.14c}$$

in which $\tau \in [0, \infty]$ and $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ is a partition of $\mathcal{E}^{0+}(x)$. In some proofs and discussions below, we have found convenient to detect the index set to which a particular index $i \in [1:n]$ belongs by looking at the position of $(F_i(x), G_i(x))$ in the graph of figure 2.2. The index sets $E_F(x)$, $E_G(x)$,

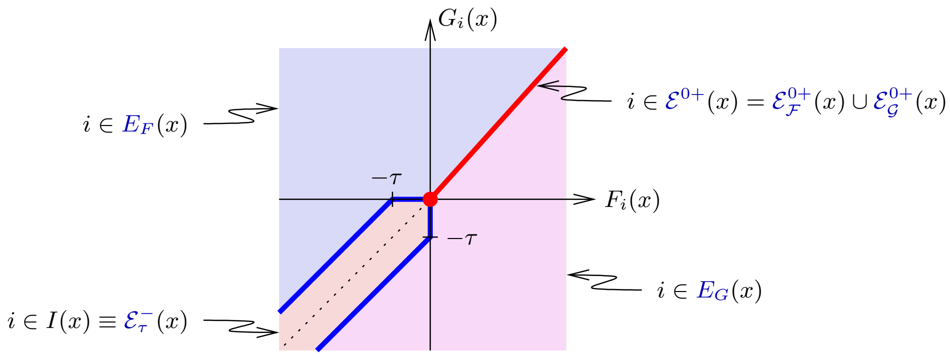

Fig. 2.2 The pair $(F_i(x), G_i(x))$ determines which index sets $E_F(x)$, $E_G(x)$, or $I(x)$, i belongs to; see (2.14) for the definition of these index sets. The nondifferentiability of H_i can only occur on the main diagonal, at points x for which $F_i(x) = G_i(x)$. Nevertheless the *secure PNM direction* (2.13) carefully deals with points that are *near* an x such that $F_i(x) = G_i(x) < 0$, those with a pair $(F_i(x), G_i(x))$ inside the fork in the left-bottom part of the picture (it then replaces one equality defining the plain NM direction (2.1) by a pair of inequalities). The width of this fork is controlled by the *kink tolerance* $\tau > 0$, which can be taken very small in practice, in order to avoid having too many inequalities in (2.13).

and $I(x)$ will be continually used in the sequel and it is important to observe that they form a partition of $[1:n]$, which is claimed in the next lemma.

Lemma 2.6 ((E_F, E_G, I) partition) *One has* $\big(E_F(x),\ E_G(x),\ I(x)\big) \in \mathcal{P}([1:n])$.

Proof Observe first that the triplet $(E_F(x), E_G(x), I(x))$ covers $[1:n]$:

$$\begin{aligned}
&E_F(x) \cup E_G(x) \cup I(x)\\
&\quad = \big(\mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x)\big) \cup \big(\mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)\big) \cup \mathcal{E}^{-}_{\tau}(x) \qquad [(2.14)]\\
&\quad \supseteq \mathcal{F}(x) \cup \mathcal{G}(x) \cup \mathcal{E}^{0+}(x) \cup \mathcal{E}^{-}(x) \qquad [\mathcal{E}^{0+}_{\mathcal{F}}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) = \mathcal{E}^{0+}(x),\ \mathcal{E}^{-}_{\tau}(x) \supseteq \mathcal{E}^{-}(x)]\\
&\quad = [1:n] \qquad [\mathcal{E}^{0+}(x) \cup \mathcal{E}^{-}(x) = \mathcal{E}(x) \text{ and } \mathcal{E}(x) \cup \mathcal{F}(x) \cup \mathcal{G}(x) = [1:n]].
\end{aligned}$$

To conclude, it suffices to show that the sets of the triplet are two by two disjoint:

- $E_F(x) \cap E_G(x) = \varnothing$, since $E_F(x) \subseteq \mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x)$, $E_G(x) \subseteq \mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)$ and $(\mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x)) \cap (\mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)) = \varnothing$;
- $E_F(x) \cap I(x) = \varnothing$, since $(\mathcal{F}(x) \setminus \mathcal{E}^{-}_{\tau}(x)) \cap \mathcal{E}^{-}_{\tau}(x) = \varnothing$ and $\mathcal{E}^{0+}_{\mathcal{G}}(x) \cap \mathcal{E}^{-}_{\tau}(x) = \varnothing$;
- $E_G(x) \cap I(x) = \varnothing$ for a similar reason as in the previous case (switch F and G). □

As a consequence of this lemma, the system (2.13) has $|E_F(x)| + |E_G(x)| = n - |I(x)|$ equalities and $2|I(x)|$ inequalities.

By taking a value of τ close to zero, the number of inequalities in (2.13) should be small and the computation of the direction should be inexpensive. Our proof of global convergence (theorems 3.6, 3.7, 3.9 and 3.11) requires to have $\tau > 0$, however. Then, the set $\mathcal{E}^{-}_{\tau}(x)$ is stable with respect to (or unchanged by) a small perturbation of x, which makes it adapted to floating point calculation.

By setting $\tau = 0$, at the left bound of the interval $[0, \infty]$, one has $\mathcal{E}^{-}_{0}(x) = \mathcal{E}^{-}(x)$, $\mathcal{F}(x) \setminus \mathcal{E}^{-}_{0}(x) = \mathcal{F}(x)$, and $\mathcal{G}(x) \setminus \mathcal{E}^{-}_{0}(x) = \mathcal{G}(x)$, so that the system (2.13) becomes

$$\begin{cases}
F_i(x) + F'_i(x)d = 0 \ \text{ if } i \in \mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x)\\
G_i(x) + G'_i(x)d = 0 \text{ if } i \in \mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)\\
F_i(x) + F'_i(x)d \geqslant 0 \ \text{ if } i \in \mathcal{E}^{-}(x)\\
G_i(x) + G'_i(x)d \geqslant 0 \text{ if } i \in \mathcal{E}^{-}(x).
\end{cases} \tag{2.15}$$

This is the system (2.11) defining the plain PNM direction.

By setting $\tau = \infty$, at the right bound of the interval $[0, \infty]$, one has $\mathcal{F}(x) \setminus \mathcal{E}^{-}_{\infty}(x) = \{i : F_i(x) < G_i(x),\ G_i(x) \geqslant 0\}$, $\mathcal{G}(x) \setminus \mathcal{E}^{-}_{\infty}(x) = \{i : G_i(x) < F_i(x),\ F_i(x) \geqslant 0\}$, so that the system (2.13) becomes

$$\begin{cases}
F_i(x) + F'_i(x)d = 0 \ \text{ if } [F_i(x) < G_i(x) \text{ and } G_i(x) \geqslant 0] \text{ or } i \in \mathcal{E}^{0+}_{\mathcal{F}}(x)\\
G_i(x) + G'_i(x)d = 0 \text{ if } [G_i(x) < F_i(x) \text{ and } F_i(x) \geqslant 0] \text{ or } i \in \mathcal{E}^{0+}_{\mathcal{G}}(x)\\
F_i(x) + F'_i(x)d \geqslant 0 \ \text{ if } F_i(x) < 0 \text{ and } G_i(x) < 0\\
G_i(x) + G'_i(x)d \geqslant 0 \text{ if } F_i(x) < 0 \text{ and } G_i(x) < 0.
\end{cases} \tag{2.16}$$

This system can be viewed as a relaxation of the following mixed LCP

$$\begin{cases}
F_i(x) + F'_i(x)d = 0, \quad \text{if } [F_i(x) < G_i(x) \text{ and } G_i(x) \geqslant 0] \text{ or } i \in \mathcal{E}^{0+}_{\mathcal{F}}(x)\\
G_i(x) + G'_i(x)d = 0, \quad \text{if } [G_i(x) < F_i(x) \text{ and } F_i(x) \geqslant 0] \text{ or } i \in \mathcal{E}^{0+}_{\mathcal{G}}(x)\\
0 \leqslant (F(x) + F'(x)d)_{\mathcal{E}^{-}_{\infty}(x)} \perp (G(x) + G'(x)d)_{\mathcal{E}^{-}_{\infty}(x)} \geqslant 0,
\end{cases}$$

which has an orthogonality condition that is not present in (2.16). This last system has some similarities with the system (1.12), obtained in [82] using other considerations.

2.3.2 Descent property

The computation of a secure PNM direction satisfying (2.13), can be more time consuming than solving the linear system (2.1) defining the plain Newton-min direction. This is due to the presence of inequalities in the system (2.13). It is therefore tempting to see whether it is possible to design a criterion allowing an algorithm to take as often as possible a less expensive direction, for example one that only requires to solve a linear system, like the plain NM direction. This is the idea supporting the hybrid algorithm defined in section 3.3 (algorithm 3.8) and the first steps towards that algorithm are done in the present section: we focus on the design of such a criterion and on its validation.

Around a solution, the plain NM direction is known to be appropriate because it yields fast convergence (theorem 2.3), while this might not be the case far from a solution because it may fail to be a descent direction of the least-square merit function θ defined in (1.4) (see counter-example 2.4). This observation speaks for a criterion based on the directional derivative of θ. Taking some safeguard, there is a temptation to accept a direction d when it satisfies the inequality

$$\theta'(x;d) \leqslant -2(1-\eta)\,\theta(x) \tag{2.17}$$

where η is some constant in $[0,1)$. This inequality is natural since it is satisfied with $\eta = 0$ when d is the Newton direction on a smooth function H and θ is the map $x \mapsto \frac{1}{2}\|H(x)\|^2$. We have not been able to prove a global convergence result in the style of theorem 3.9 below with such a simple criterion and for an arbitrary direction d satisfying it, so that we design below a more robust criterion.

For an arbitrary direction $d \in \mathbb{R}^n$, proposition 2.7 below will show that

$$\theta'(x;d) \leqslant -\sum_{i\in[1:n]} (1-\rho_i(x,d))\, H_i(x)^2, \tag{2.18}$$

provided the $\rho_i(x,d)$'s are the *arbitrarily signed* values defined by formula (2.20) below (note that these values depend on τ and the partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, through the index sets $E_F(x)$, $E_G(x)$, and $I(x)$, but the validity of (2.18) does not depend on them, as shown by proposition 2.7 below). We shall see in corollary 2.8 that $\rho_i(x,d) \leqslant 0$ for the secure PNM direction (2.13), so that the inequality (2.17) with $\eta = 0$ follows from (2.18) for that direction. As a result, the secure PNM direction is a descent direction of θ at x (corollary 2.8).

The criterion for accepting an arbitrary direction d in the linesearch will be that the right-hand side of (2.18) is less than the right-hand side of (2.17), namely: for some τ and some partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, it is required to have

$$-\sum_{i\in[1:n]} (1-\rho_i(x,d))\, H_i(x)^2 \leqslant -2(1-\eta)\,\theta(x), \tag{2.19a}$$

where η is some constant in $[0,1)$. From the expression (1.4) of θ, we see that this criterion simplifies into

$$\boxed{\frac{1}{2}\sum_{i\in[1:n]} \rho_i(x,d)\, H_i(x)^2 \leqslant \eta\,\theta(x),} \tag{2.19b}$$

where, again, the $\rho_i(x,d)$'s are the *arbitrarily signed* values defined by formula (2.20) below. The acceptation criterion (2.19) is more demanding than (2.17) since, thanks to (2.18), it implies (2.17). We see that the contributions of the terms in the sum in the right-hand side of (2.18) can be compensated by each other: the negativity of the directional derivative $\theta'(x;d)$ can be obtained by some negative terms in this sum, despite the positivity of other terms. This flexibility will allow the hybrid algorithm of section 2.4 to accept very often the plain NM direction (as observed in our experiments [50]). The important point is that the criterion (2.19) happens to be sufficient to get the global convergence of theorem 3.9, because it is the left-hand side of the inequality (2.19a) that appears in its proof (see the one of theorem 3.6).

In the rest of this section, we focus on the proof of the inequality (2.18) and on its ability to detect descent directions. First, let us define the quantities $\rho_i(x,d)$ appearing in the right-hand side of (2.18). Let $x \in \mathbb{R}^n$ be an arbitrary point and $d \in \mathbb{R}^n$ be an arbitrary direction. We define $\rho_i(x,d)$ by

$$\rho_i(x,d) := \begin{cases} \frac{F_i(x)+F_i'(x)d}{F_i(x)} & \text{if } i \in E_F(x) \text{ and } F_i(x) \neq 0 \\ 0 & \text{if } i \in E_F(x) \text{ and } F_i(x) = 0 \\ \frac{G_i(x)+G_i'(x)d}{G_i(x)} & \text{if } i \in E_G(x) \text{ and } G_i(x) \neq 0 \\ 0 & \text{if } i \in E_G(x) \text{ and } G_i(x) = 0 \\ \max\left(\frac{F_i(x)+F_i'(x)d}{F_i(x)}, \frac{G_i(x)+G_i'(x)d}{G_i(x)}\right) & \text{if } i \in I(x), \end{cases} \tag{2.20}$$

where the partition $(E_F(x), E_G(x), I(x))$ of $[1\,{:}\,n]$ has been defined in (2.14) (hence, the five groups of indices in (2.20) also form a partition of $[1\,{:}\,n]$). The zero value given to $\rho_i(x,d)$ when $F_i(x) = 0$ or $G_i(x) = 0$ allows us to simplify the statement of corollary 2.8 below but, as we shall see, an arbitrary value could have been given instead, since this one does not occur in the calculations that follow. Note that the $\rho_i(x,d)$'s depend on τ through the index sets $E_F(x)$, $E_G(x)$, and $I(x)$.

Let us stress the fact that the $\rho_i(x,d)$'s given by (2.20) are not necessarily less than one and such a restriction on d is not imposed in the next proposition. Hence, the formula (2.18) does not give an upper bound of $\theta'(x;d)$ as a sum of nonpositive terms and does not imply the negativity of that directional derivative. This is utterly normal, since d is arbitrary in the next proposition.

Proposition 2.7 (overestimation of $\theta'(x;d)$) *Let $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$, H be the function defined by (1.3b), and the $\rho_i(x,d)$'s be defined by (2.20) with any $\tau \geqslant 0$ and any partition $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ of $\mathcal{E}^{0+}(x)$. Then, (2.18) holds.*

Proof Let us first show that

$$\forall\, i \in \mathcal{F}(x): \quad F_i(x)F_i'(x)d \leqslant -(1-\rho_i)F_i(x)^2, \tag{2.21a}$$
$$\forall\, i \in \mathcal{G}(x): \quad G_i(x)G_i'(x)d \leqslant -(1-\rho_i)G_i(x)^2, \tag{2.21b}$$

where we set $\rho_i \equiv \rho_i(x,d)$ to alleviate notation. Consider first (2.21a). Observe that $i \in \mathcal{F}(x)$ if and only if $(i \in \mathcal{F}(x) \setminus \mathcal{E}_\tau^-(x)$ and $F_i(x) \neq 0)$ or $(i \in \mathcal{F}(x)$ and $F_i(x) = 0)$ or $(i \in \mathcal{F}(x) \cap \mathcal{E}_\tau^-(x))$.

- If $i \in \mathcal{F}(x) \setminus \mathcal{E}_\tau^-(x) \subseteq E_F(x)$ and $F_i(x) \neq 0$, $(2.20)_1$ gives $F_i'(x)d = -(1-\rho_i)F_i(x)$, hence (2.21a) with equality follows by multiplying both sides of this identity by $F_i(x)$.
- If $i \in \mathcal{F}(x)$ and $F_i(x) = 0$, (2.21a) is clearly satisfied with equality.
- If $i \in \mathcal{F}(x) \cap \mathcal{E}_\tau^-(x) \subseteq I(x)$, $(2.20)_5$ gives $-F_i'(x)d \leqslant (1-\rho_i)F_i(x)$, hence (2.21a) follows by multiplying both sides of this inequality by $-F_i(x) > 0$.

To get (2.21b), use the same arguments, with G instead of F and with $(2.20)_3$ instead of $(2.20)_1$.

Now using (2.6), (2.21), and $F_i(x) = G_i(x)$ for $i \in \mathcal{E}(x)$, we get

$$\begin{aligned} \theta'(x;d) \leqslant &- \sum_{i\in\mathcal{F}(x)} (1-\rho_i)F_i(x)^2 - \sum_{i\in\mathcal{G}(x)} (1-\rho_i)G_i(x)^2 - \sum_{i\in\mathcal{E}(x)} (1-\rho_i)F_i(x)^2 \\ &+ \sum_{i\in\mathcal{E}(x)} F_i(x)\min\big((1-\rho_i)F_i(x) + F_i'(x)d, (1-\rho_i)G_i(x) + G_i'(x)d\big). \end{aligned}$$

Therefore, to get (2.18), it suffices to show that the last term in the right-hand side of the previous inequality is nonpositive. For this, we consider the partition $(\mathcal{E}^0(x), \mathcal{E}^+(x), \mathcal{E}^-(x))$ of $\mathcal{E}(x)$ defined by (2.8).

- If $i \in \mathcal{E}^0(x)$, then $F_i(x) = G_i(x) = 0$ and the corresponding term vanishes.

- If $i \in \mathcal{E}^+(x) = (\mathcal{E}^{0+}_{\mathcal{F}}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)) \setminus \mathcal{E}^0(x) \subseteq E_F(x) \cup E_G(x)$, then one of the arguments of the minimum vanish by the definition of ρ_i in $(2.20)_1$ and $(2.20)_3$, so that the minimum if nonpositive. Since $F_i(x) > 0$, the term of the sum corresponding to the considered $i \in \mathcal{E}^+(x)$ is also nonpositive.
- If $i \in \mathcal{E}^-(x) = \mathcal{E}^-_0(x) \subseteq \mathcal{E}^-_\tau(x) = I(x)$, then, by $(2.20)_5$ and $F_i(x) = G_i(x) < 0$, the minimum vanishes. □

Corollary 2.8 (descent secure PNM direction) *Suppose that d is a secure PNM direction, hence satisfying (2.13) for some $\tau \in [0, \infty]$ and some partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$. Then, the $\rho_i(x,d)$'s defined by (2.20) are nonpositive and, consequently, (2.17) and (2.19) hold with $\eta = 0$. In particular, if $\theta(x) \neq 0$, d is a descent direction of θ at x.*

Proof Suppose that d satisfies (2.13) at $x \in \mathbb{R}^n$, for some $\tau \in [0, \infty]$ and some partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$. For $i \in E_F(x)$, $(2.13)_1$ shows that $F_i(x) + F'_i(x)d = 0$, so that $\rho_i(x,d) = 0$ by $(2.20)_1$ and $(2.20)_2$. Similarly, $\rho_i(x,d) = 0$ for $i \in E_G(x)$. For $i \in I(x)$, $(2.13)_3$ and $(2.13)_4$ show that $\rho_i(x,d) \leqslant 0$ by $(2.20)_5$. We have shown that the $\rho_i(x,d)$'s defined by (2.20) are nonpositive.

Now, the inequality $\theta'(x;d) \leqslant -2\,\theta(x)$, which is (2.17) with $\eta = 0$, follows immediately from (2.18), which holds by proposition 2.7, since the terms with $\rho_i(x,d)$ in factor in the right-hand side are nonpositive and can be discarded. For the same reason, (2.19) holds with $\eta = 0$.

Finally, if $\theta(x) \neq 0$, the inequality $\theta'(x;d) \leqslant -2\,\theta(x)$ yields $\theta'(x;d) < 0$, showing that d is a descent direction of θ at x. □

As another illustration of the usefulness of proposition 2.7, consider an *inexact secure PNM direction* d, which, by definition, verifies, for some $\eta \geqslant 0$, the following inequalities:

$$\boxed{\begin{array}{rll} & F_i(x) + F'_i(x)d \leqslant \eta F_i(x), & \forall\, i \in \mathcal{F}^{0+}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x), \\ \eta F_i(x) \leqslant & F_i(x) + F'_i(x)d, & \forall\, i \in \mathcal{F}^{-}(x) \cup \mathcal{E}^{-}_{\tau}(x), \\ & G_i(x) + G'_i(x)d \leqslant \eta G_i(x), & \forall\, i \in \mathcal{G}^{0+}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x), \\ \eta G_i(x) \leqslant & G_i(x) + G'_i(x)d, & \forall\, i \in \mathcal{G}^{-}(x) \cup \mathcal{E}^{-}_{\tau}(x), \end{array}} \tag{2.22}$$

where the index sets $\mathcal{F}^-(x)$, $\mathcal{F}^{0+}(x)$, $\mathcal{G}^-(x)$, and $\mathcal{G}^{0+}(x)$ are defined by

$$\mathcal{F}^-(x) := \{i \in \mathcal{F}(x) : F_i(x) < 0\}, \qquad \mathcal{F}^{0+}(x) := \{i \in \mathcal{F}(x) : F_i(x) \geqslant 0\}, \tag{2.23a}$$

$$\mathcal{G}^-(x) := \{i \in \mathcal{G}(x) : G_i(x) < 0\}, \qquad \mathcal{G}^{0+}(x) := \{i \in \mathcal{G}(x) : G_i(x) \geqslant 0\}. \tag{2.23b}$$

The system (2.22) brings much flexibility regarding the realization of the equalities and inequalities in the previous system (2.13), since a direction verifying (2.13) satisfies (2.22) with $\eta = 0$.

The following result can be shown for the direction (2.22). See [38] for a proof.

Corollary 2.9 (descent inexact secure PNM direction) *Suppose that d is an inexact secure PNM direction, hence satisfying (2.22) for some $\tau \in [0, \infty]$ and $\eta \geqslant 0$. Then, the $\rho_i(x,d)$'s defined by (2.20) do not exceed η and, consequently, (2.17) and (2.19) hold with the given η. In particular, if $\theta(x) \neq 0$ and $\eta \in [0,1)$, d is a descent direction of θ at x.*

2.4 Statement of the algorithms

The convergent algorithms 2.11 and 3.8, presented below, are variations of the following generic algorithm. It is the global convergence of this generic algorithm that will be analyzed in section 3.2, and more particularly in theorem 3.6, in which an additional assumption is made on the computed directions (their boundedness). In this algorithm, the term “constant” means “independent of the iteration”.

Algorithm 2.10 (generic NM algorithm) Let x be the current iterate. Let $\eta \in [0,1)$ be the constant appearing in the acceptation criterion (2.19), let $\tau \in (0,\infty]$ be the constant kink tolerance used in the definition of the index sets $E_F(x)$, $E_G(x)$, and $I(x)$ by (2.14), and let $\omega \in (0,\frac{1}{2})$ and $\beta \in (0,1)$ be the two constants used in the linesearch of step 4 below. The next iterate $x_+ \in \mathbb{R}^n$ is computed as follows.

1. *Stopping criterion.* If $\theta(x) = 0$, stop (then, x is a solution to (1.1)).
2. *Index sets.* Choose some partition $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ of $\mathcal{E}^{0+}(x)$ and compute the index sets $E_F(x)$, $E_G(x)$, and $I(x)$ defined by (2.14).
3. *Direction.* Compute a direction $d \in \mathbb{R}^n$ satisfying (2.19) for the $\rho_i(x,d)$'s defined in (2.20),
4. *Stepsize.* Set $\alpha := \beta^i$, where i is the smallest nonnegative integer such that

$$\theta(x + \alpha d) \leqslant (1-2\omega\alpha(1-\eta))\, \theta(x). \tag{2.24}$$

5. *New iterate.* $x_+ := x + \alpha d$.

The well-posedness of this algorithm is discussed below, after having presented two of its instances, which only differ by their way of computing the direction d in step 3.

A first instance of the generic NM algorithm is the one that computes the direction d as the minimum norm solution to (2.13).

Algorithm 2.11 (PNM algorithm) It is the instance of the generic algorithm 2.10, in which the direction d in step 3 is computed as a solution to the following problem

$$\min\, \{|\!|\!| d |\!|\!| : d \text{ satisfies } (2.13)\}, \tag{2.25}$$

where $|\!|\!| \cdot |\!|\!|$ is an arbitrary norm.

The norm $|\!|\!| \cdot |\!|\!|$ in (2.25) may be arbitrary with regard to the convergence of the algorithm; if the Euclidean norm is used, (2.25) is a standard strictly convex quadratic optimization problem (with the squared norm), which has a unique solution and can be solved in polynomial time. Since the solution d to (2.25) satisfies (2.13), it also satisfies (2.19) with $\eta = 0$ (corollary 2.8), which shows that algorithm 2.11 is indeed an instance of algorithm 2.10.

The algorithms 2.10 and 2.11 are rather standard in their structure. Only the computation of the direction in step 3, whose conception has been progressively introduced above, makes exception. Let us give some more comments.

1. There is an implicit assumption in algorithm 2.11, which will have to be clarified in the results on this algorithm. Indeed, the algorithm assumes that problem (2.25) has a solution at each iteration, which may not be the case if the affine system (2.13) is infeasible. A rather weak condition guaranteeing the feasibility of (2.13), for x near a limit point $\bar{x}$, is introduced and discussed in section 3.1.
2. If not empty, the polyhedron defined by (2.13) may be unbounded, which raises some difficulty in the convergence proof of section 3.2. For this reason, in (2.25), we take the option of taking a minimum norm direction in that polyhedron, but any other technique guaranteeing the boundedness of the directions computed at a converging sequence of x's would be appropriate.
3. The directions computed in step 3 of algorithm 2.10, if any, are necessarily descent directions of θ at x. This is because they satisfy (2.19) with $\eta < 1$, hence (2.17) with the same $\eta < 1$, implying that $\theta'(x; d) < 0$ when x is not a solution to the complementarity problem (1.1) (this is guaranteed by step 1). As a result, in that case, the linesearch in step 4 is able to compute a stepsize $\alpha > 0$ in a finite number of trials [17,53].

4. Condition (2.24) derives from the standard Armijo inequality [6,36,17]

$$\theta(x + \alpha d) \leqslant \theta(x) + \omega\alpha\theta'(x; d),$$

in which the negative upper bound $-2(1-\eta)\,\theta(x)$ of $\theta'(x;d)$ given by (2.17) has replaced the directional derivative.

3 Algorithm analysis

This section starts with giving a *regularity condition* at a point $\bar{x} \in \mathbb{R}^n$ that ensures several properties (section 3.1). First, it implies that the system (2.13), defining the secure PNM direction d, has a solution when x is near $\bar{x}$ (proposition 3.2). Next, it also certifies that the algorithm can choose a solution to (2.13) that has a continuity property (proposition 3.4). Finally, the continuity property guarantees that the chosen directions are bounded for x near $\bar{x}$ (corollary 3.5). This boundedness property is useful for establishing global convergence results (section 3.2).

3.1 Regularity

3.1.1 Regularity at a point

Let $\bar{x} \in \mathbb{R}^n$ be a point that is not necessarily a solution to the complementarity problem (1.1), an assumption that is important in the proof of the global convergence result (theorem 3.6), since there the accumulation points of the generated sequence have a priori no particular properties. Our vehicle for highlighting conditions ensuring the solvability of the affine system (2.13), when x is near $\bar{x}$, is the Mangasarian-Fromovitz constraint qualification (MFCQ) [76], which reads

$$\begin{array}{l}\sum_{i\in E_F(x)} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(x)} \beta_i \nabla G_i(x) + \sum_{i\in I(x)} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0 \\ \text{and } (\alpha_{I(x)}, \beta_{I(x)}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0.\end{array} \tag{3.1a}$$

Another equivalent version reads

$$\begin{array}{l}\text{For all } (a, a', b, b') \in \mathbb{R}^{|E_F(x)|} \times \mathbb{R}^{|E_G(x)|} \times \mathbb{R}^{|I(x)|} \times \mathbb{R}^{|I(x)|}, \text{ there is a direction } d \in \mathbb{R}^n \\ \text{such that } F'_{E_F(x)}(x)d = a,\ G'_{E_G(x)}(x)d = a',\ F'_{I(x)}(x)d \geqslant b \text{ and } G'_{I(x)}(x)d \geqslant b'.\end{array} \tag{3.1b}$$

Clearly, the system (2.13) has a solution d when (3.1b) holds at x (and this MFCQ condition is almost necessary, since the independent terms in (2.13), deduced from $F(x)$ and $G(x)$, can be arbitrary).

It is not sufficient to require the satisfaction of the MFCQ condition (3.1) at $x = \bar{x}$ to get it at x near $\bar{x}$, like above. The reason comes from the change in the index sets $E_F(x)$, $E_G(x)$ and $I(x)$ with x. Suppose indeed that only (3.1a) holds at $x = \bar{x}$. It is well known that the implication in (3.1a) is insensitive to small perturbations in the gradients $\nabla F_i(x)$ and $\nabla G_i(x)$ in its premise (see [53; exercise 4.16] for instance). Therefore, if we assume the continuity of the derivatives F' and G' at $\bar{x}$ and if x is near $\bar{x}$, it follows from (3.1a) at $x = \bar{x}$ that

$$\begin{array}{l}\sum_{i\in E_F(\bar{x})} \alpha_i \nabla F_i(x) + \sum_{i\in E_G(\bar{x})} \beta_i \nabla G_i(x) + \sum_{i\in I(\bar{x})} \big[\alpha_i \nabla F_i(x) + \beta_i \nabla G_i(x)\big] = 0 \\ \text{and } (\alpha_{I(\bar{x})}, \beta_{I(\bar{x})}) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0,\end{array}$$

where the gradients are evaluated at x, while the index sets are evaluated in $\bar{x}$. Here, however, none of these sets $E_F(\bar{x})$, $E_G(\bar{x})$ and $I(\bar{x})$ are guaranteed to be invariant when $\bar{x}$ is slightly modified. Therefore, (3.1) at $x = \bar{x}$ may not imply that (3.1) holds at x near $\bar{x}$. For this reason, we adopt a stronger regularity condition.

The set of partitions (E_F, E_G, I) of $[1:n]$, such that $(E_F, E_G, I) = (E_F(x), E_G(x), I(x))$ for some x in a neighborhood V of $\bar{x}$ and some partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, decreases when V

gets smaller. Since the number of partition of $[1:n]$ is finite, one can find a neighborhood V_{PNM} of $\bar{x}$ for which these partitions (E_F, E_G, I) are minimal. Then, denote by

$$\mathcal{P}_{\text{PNM}} := \{(E_F(x), E_G(x), I(x)) : x \in V_{\text{PNM}},\ (\mathcal{E}_{\mathcal{F}}^{0+}(x),\, \mathcal{E}_{\mathcal{G}}^{0+}(x)) \text{ is a partition of } \mathcal{E}^{0+}(x)\}$$

this smallest set of partitions.

Definition 3.1 (PNM-regularity) Let be given $\tau \in (0, \infty]$. A point $\bar{x} \in \mathbb{R}^n$ is said to be *PNM-regular* (we also say that *PNM-regularity* holds at $\bar{x} \in \mathbb{R}^n$) if, for all $x \in V_{\text{PNM}}$, F and G are differentiable at x and for all $(E_F, E_G, I) \in \mathcal{P}_{\text{PNM}}$, one has

$$\begin{array}{l} \sum_{i \in E_F} \alpha_i \nabla F_i(\bar{x}) + \sum_{i \in E_G} \beta_i \nabla G_i(\bar{x}) + \sum_{i \in I} \big[\alpha_i \nabla F_i(\bar{x}) + \beta_i \nabla G_i(\bar{x})\big] = 0 \\ \text{and } (\alpha_I, \beta_I) \geqslant 0 \text{ imply that } (\alpha, \beta) = 0, \end{array} \tag{3.2a}$$

or, equivalently,

$$\begin{array}{l} \text{for all } (a, a', b, b') \in \mathbb{R}^{|E_F|} \times \mathbb{R}^{|E_G|} \times \mathbb{R}^{|I|} \times \mathbb{R}^{|I|}, \text{ there is a direction } d \in \mathbb{R}^n \text{ such} \\ \text{that } F'_{E_F}(\bar{x})d = a,\ G'_{E_G}(\bar{x})d = a',\ F'_I(\bar{x})d \geqslant b \text{ and } G'_I(\bar{x})d \geqslant b'. \end{array} \tag{3.2b}$$

□

Proposition 3.2 ((3.1) near a PNM-regular point) *Suppose that the PNM-regularity condition 3.1 holds at $\bar{x}$. Then, there is a neighborhood $V'_{\text{PNM}} \subseteq V_{\text{PNM}}$ of $\bar{x}$ such that for all $x \in V'_{\text{PNM}}$, (3.1) holds at x.*

Proof By the PNM-regularity definition 3.1, for all $(E_F, E_G, I) \in \mathcal{P}_{\text{PNM}}$, (3.2b) holds. This implies that there is a neighborhood V of $\bar{x}$ such that, for all $x \in V$:

$$\begin{array}{l} \text{for all } (a, a', b, b') \in \mathbb{R}^{|E_F|} \times \mathbb{R}^{|E_G|} \times \mathbb{R}^{|I|} \times \mathbb{R}^{|I|}, \text{ there is a direction } d \in \mathbb{R}^n \text{ such} \\ \text{that } F'_{E_F}(x)d = a,\ G'_{E_G}(x)d = a',\ F'_I(x)d \geqslant b \text{ and } G'_I(x)d \geqslant b'. \end{array} \tag{3.3}$$

Since $\mathcal{P}_{\text{PNM}}$ is finite, there is a neighborhood $V'_{\text{PNM}} \subseteq V_{\text{PNM}}$ of $\bar{x}$ such that for all $x \in V'_{\text{PNM}}$ and all $(E_F, E_G, I) \in \mathcal{P}_{\text{PNM}}$, (3.3) holds. By definition of $\mathcal{P}_{\text{PNM}}$, for any $x \in V_{\text{PNM}}$ (hence $x \in V'_{\text{PNM}}$) and any partition $(\mathcal{E}_{\mathcal{F}}^{0+}(x), \mathcal{E}_{\mathcal{G}}^{0+}(x))$ of $\mathcal{E}^{0+}(x)$, $(E_F(x), E_G(x), I(x)) \in \mathcal{P}_{\text{PNM}}$. It follows that, for any $x \in V'_{\text{PNM}}$, one has

$$\begin{array}{l} \text{for all } (a, a', b, b') \in \mathbb{R}^{|E_F(x)|} \times \mathbb{R}^{|E_G(x)|} \times \mathbb{R}^{|I(x)|} \times \mathbb{R}^{|I(x)|}, \text{ there is a direction } d \in \mathbb{R}^n \\ \text{such that } F'_{E_F(x)}(x)d = a,\ G'_{E_G(x)}(x)d = a',\ F'_{I(x)}(x)d \geqslant b \text{ and } G'_{I(x)}(x)d \geqslant b'. \end{array}$$

This is (3.1) at x. □

We conclude this section by giving a counter-example showing that the PNM-regularity of definition 3.1 does not imply the NM-regularity of definition 2.1. Now, with the pair of inequalities that must be satisfied for the indices in $I(x)$ in (2.13), a priori, the NM-regularity may not imply the PNM-regularity. Therefore, the two concepts of regularity cannot really be compared.

Counter-examples 3.3 (PNM-regularity $\not\Rightarrow$ NM-regularity) Consider the LCP (1.2), in which

$$n = 2, \qquad M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad q = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Let us make the correspondence between the LCP (1.2) and the general complementarity problem (1.1) by defining F and G at x by $F(x) = Mx + q$ and $G(x) = x$. Then, at $\bar{x} = (-1, -2)$, one has $\mathcal{F}(\bar{x}) = \mathcal{G}(\bar{x}) = \varnothing$ and $\mathcal{E}(\bar{x}) = \{1, 2\}$. Taking $(\tilde{\mathcal{F}}, \tilde{\mathcal{G}}) = (\{2\}, \{1\})$ as partition of $\{1, 2\}$ satisfying $\tilde{\mathcal{F}} \supseteq \mathcal{F}(\bar{x})$ and $\tilde{\mathcal{G}} \supseteq \mathcal{G}(\bar{x})$, the Jacobian of the system (2.1) reads

$$\begin{pmatrix} F'_2(\bar{x}) \\ G'_1(\bar{x}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

This one is singular, showing the $\bar{x}$ is not NM-regular in the sense of definition 2.1 (and in the present case, the system (2.1) has no solution). However, the PNM-regularity in the sense of definition 3.1 holds at $\bar{x}$, since, for x near $\bar{x}$, $E_F(x) = E_G(x) = \varnothing$ and $I(x) = \{1, 2\}$, so that

$$\mathcal{P}_{\text{PNM}} = \{(\varnothing, \varnothing, \{1, 2\})\}.$$

The premise in (3.2a) reads

$$\alpha_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 \qquad \text{and} \qquad (\alpha, \beta) \geqslant 0.$$

This one clearly implies that $\alpha = \beta = 0$, showing the PNM-regularity holds at $\bar{x}$. As desired and proved in proposition 3.4(2) below, for x near $\bar{x}$, the system (2.13), namely

$$\begin{pmatrix} x_2 + 1 \\ x_1 - 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} d \geqslant 0 \qquad \text{and} \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} d \geqslant 0$$

has a solution, since it consists in the system of inequalities $d_1 \geqslant \max(1 - x_1, -x_1) = 1 - x_1$ and $d_2 \geqslant \max(-1 - x_2, -x_2) = -x_2$, which presents no compatibility problem. □

3.1.2 Continuity of selected directions

We now consider the question of whether a solution d to (2.13) at x can be *chosen* in such a way that these directions are bounded when x is near a given arbitrary point $\bar{x}$. This will be a consequence of the continuity property stated in the next proposition, which is guaranteed when the PNM-regularity condition 3.1 holds at $\bar{x}$. The boundedness property is useful for establishing the global convergence result of theorems 3.7 and 3.9 below.

We say that a function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is *locally radially Lipschitz continuous* at $\bar{x} \in \mathbb{R}^n$ for the Euclidean norm $\|\cdot\|$ if there is a neighborhood V of $\bar{x}$ in $\mathbb{R}^n$ and a constant $L \geqslant 0$, such that for all $x \in V$, $\|\varphi(x) - \varphi(\bar{x})\| \leqslant L\|x - \bar{x}\|$.

Proposition 3.4 (continuity of the selected directions) *Suppose that F and G are continuously differentiable at $\bar{x} \in \mathbb{R}^n$, that $\tau \in (0, \infty]$ and that the PNM-regularity condition 3.1 holds at $\bar{x}$. Then, the following properties hold.*

1) *For any $(E_F, E_F, I) \in \mathcal{P}_{\text{PNM}}$, the system*

$$\begin{aligned} F_{E_F}(\bar{x}) + F'_{E_F}(\bar{x})d = 0,\ G_{E_G}(\bar{x}) + G'_{E_G}(\bar{x})d = 0, \\ F_I(\bar{x}) + F'_I(\bar{x})d \geqslant 0,\ G_I(\bar{x}) + G'_I(\bar{x})d \geqslant 0 \end{aligned} \tag{3.4}$$

has a solution $\bar{d}$. Denote by $\bar{D}$ the finite set of these selected $\bar{d}$'s, each of them being associated with one $(E_F, E_F, I) \in \mathcal{P}_{\text{PNM}}$.

2) *For any $\delta > 0$, there is a neighborhood V of $\bar{x}$ such that, for any $x \in V$ and any partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, the system (2.13) has a solution $d(x)$ that satisfies*

$$\min_{\bar{d} \in \bar{D}} \|d(x) - \bar{d}\| < \delta.$$

3) *If, in addition, F' and G' are locally radially Lipschitz continuous at $\bar{x}$, then, there is a neighborhood V' of $\bar{x}$ and a constant $L \geqslant 0$ such that, for any $x \in V'$ and any partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, the system (2.13) has a solution $d(x)$ that satisfies*

$$\min_{\bar{d} \in \bar{D}} \|d(x) - \bar{d}\| \leqslant L\|x - \bar{x}\|.$$

Proof 1) Let $(E_F, E_F, I) \in \mathcal{P}_{\text{PNM}}$ be one of the partitions of $[1:n]$ considered in the PNM-regularity condition 3.1. By (3.2b), the system (3.4) has a solution $\bar{d}$. Since $\mathcal{P}_{\text{PNM}}$ is finite, the set $\bar{D}$ of these selected $\bar{d}$'s is finite.

2) Let V'_{PNM} be the neighborhood of $\bar{x}$ given by proposition 3.2, which assumes the PNM-regularity condition 3.1 at $\bar{x}$. This proposition tells us that, for any $x \in V'_{\text{PNM}}$ and any partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, the MFCQ condition (3.1b) holds at x, so that the convex polyhedron

$$\begin{aligned}\mathfrak{P}(x) := \{d \in \mathbb{R}^n :\ & F_{E_F(x)}(x) + F'_{E_F(x)}(x)d = 0,\ G_{E_G(x)}(x) + G'_{E_G(x)}(x)d = 0,\\ & F_{I(x)}(x) + F'_{I(x)}(x)d \geqslant 0,\ G_{I(x)}(x) + G'_{I(x)}(x)d \geqslant 0\},\end{aligned}$$

is nonempty (recall that the partition $(E_F(x), E_G(x), I(x))$ of $[1:n]$ is given by (2.14)).

For each $x \in V'_{\text{PNM}}$ and each partition $(\mathcal{E}^{0+}_{\mathcal{F}}(x), \mathcal{E}^{0+}_{\mathcal{G}}(x))$ of $\mathcal{E}^{0+}(x)$, one determines an element $d(x)$ of $\mathfrak{P}(x)$ as follows. Let $\bar{d}(x)$ be the direction of $\bar{D}$ associated with $(E_F, E_F, I) = (E_F(x), E_F(x), I(x)) \in \mathcal{P}_{\text{PNM}}$. Then, define $d(x)$ as the Euclidean projection of $\bar{d}(x)$ on $\mathfrak{P}(x)$, which is written

$$d(x) := \mathrm{P}_{\mathfrak{P}(x)}(\bar{d}(x)).$$

By Hoffman's error bound for polyhedron [59; 1952], one has the following upper bound on the distance from $\bar{d}(x)$ to $\mathfrak{P}(x)$:

$$\|d(x) - \bar{d}(x)\| \leqslant \mathfrak{h}(x) \left\| \begin{pmatrix} F_{E_F}(x) + F'_{E_F}(x)\bar{d}(x) \\ G_{E_G}(x) + G'_{E_G}(x)\bar{d}(x) \\ [F_I(x) + F'_I(x)\bar{d}(x)]^- \\ [G_I(x) + G'_I(x)\bar{d}(x)]^- \end{pmatrix} \right\|, \tag{3.5a}$$

where $\|\cdot\|$ denotes the Euclidean norm (for example), the *Hoffman factor* $\mathfrak{h}(x)$ only depends on $F'(x)$ and $G'(x)$, $t^- := \max(0, -t)$ for $t \in \mathbb{R}$, and v^- is defined componentwise when v is a vector. Now, by (3.4) and the definition of $\bar{d}(x)$, one has

$$\begin{pmatrix} F_{E_F}(\bar{x}) + F'_{E_F}(\bar{x})\bar{d}(x) \\ G_{E_G}(\bar{x}) + G'_{E_G}(\bar{x})\bar{d}(x) \\ [F_I(\bar{x}) + F'_I(\bar{x})\bar{d}(x)]^- \\ [G_I(\bar{x}) + G'_I(\bar{x})\bar{d}(x)]^- \end{pmatrix} = 0, \tag{3.5b}$$

so that (3.5a) becomes

$$\|d(x) - \bar{d}(x)\| \leqslant \mathfrak{h}(x) \left\| \begin{pmatrix} F_{E_F}(x) + F'_{E_F}(x)\bar{d}(x) - [F_{E_F}(\bar{x}) + F'_{E_F}(\bar{x})\bar{d}(x)] \\ G_{E_G}(x) + G'_{E_G}(x)\bar{d}(x) - [G_{E_G}(\bar{x}) + G'_{E_G}(\bar{x})\bar{d}(x)] \\ [F_I(x) + F'_I(x)\bar{d}(x)]^- - [F_I(\bar{x}) + F'_I(\bar{x})\bar{d}(x)]^- \\ [G_I(x) + G'_I(x)\bar{d}(x)]^- - [G_I(\bar{x}) + G'_I(\bar{x})\bar{d}(x)]^- \end{pmatrix} \right\|. \tag{3.5c}$$

Suppose that $\mathfrak{h}(x)$ is bounded for x near $\bar{x}$ (this will be proven below). Then, using the 1-Lipschitz continuity of t^- (which means that $|t_2^- - t_1^-| \leqslant |t_2 - t_1|$ for all t_1 and $t_2 \in \mathbb{R}$), the continuity of F, G, F' and G' at $\bar{x}$ and the fact that $\bar{d}(x)$ is bounded (it belongs to the finite set $\bar{D}$), we see that for any $\delta > 0$ and for x sufficiently close to $\bar{x}$, one can find $d(x) \in \mathfrak{P}(x)$ and $\bar{d}(x) \in \bar{D}$ such that $\|d(x) - \bar{d}(x)\| < \delta$. The inequality in conclusion of point 2 follows.

To prove the boundedness for x near $\bar{x}$ of the Hoffman factor $\mathfrak{h}(x)$, appearing in (3.5a), we trust the perturbation property in [73; theorem 5.5]. This property claims that if the MFCQ holds for a system "$Ad = a$ and $Bd \leqslant b$ in d" (A and B are matrices and a and b are vectors of appropriate dimensions), then the Hoffman constant is bounded for any convex polyhedron $\{d \in \mathbb{R}^n : \tilde{A}d = \tilde{a}, \tilde{B}d \leqslant \tilde{b}\}$ with arbitrary $(\tilde{a}, \tilde{b})$ and with $(\tilde{A}, \tilde{B})$ close enough to (A, B) (the reciprocal is also true).

The Hoffman factor $\mathfrak{h}(x)$ was associated in (3.5a) with the convex polyhedron $\mathfrak{P}(x)$ or, with $(E_F, E_F, I) := (E_F(x), E_F(x), I(x))$,

$$\{d \in \mathbb{R}^n : F_{E_F}(x) + F'_{E_F}(x)d = 0,\ G_{E_G}(x) + G'_{E_G}(x)d = 0,\\ F_I(x) + F'_I(x)d \geqslant 0,\ G_I(x) + G'_I(x)d \geqslant 0\}.$$

With the fixed partition (E_F, E_F, I) of $[1:n]$ in $\mathcal{P}_{\text{PNM}}$, this one can be viewed as a perturbation of the convex polyhedron

$$\{d \in \mathbb{R}^n : F_{E_F}(\bar{x}) + F'_{E_F}(\bar{x})d = 0,\ G_{E_G}(\bar{x}) + G'_{E_G}(\bar{x})d = 0,\\ F_I(\bar{x}) + F'_I(\bar{x})d \geqslant 0,\ G_I(\bar{x}) + G'_I(\bar{x})d \geqslant 0\}.$$

By (3.2b), MFCQ holds for this polyhedron. Therefore, by [73; theorem 5.5], the Hoffman factor is constant for x near $\bar{x}$ and the chosen partition (E_F, E_F, I). Now, $\mathcal{P}_{\text{PNM}}$ is finite, so that the Hoffman factor $\mathfrak{h}(x)$ appearing in (3.5a) is bounded for x near $\bar{x}$.

3) The reasoning is identical to the one presented in point 2. But now, one can use the local radial Lipschitz property of F, G, F' and G' at $\bar{x}$ to deduce from (3.5c) the existence of a neighborhood $V' \subseteq V_{\text{PNM}}$ and a constant $L \geqslant 0$, such that, for $x \in V'$, one can find $d(x) \in \mathfrak{P}(x)$ and $\bar{d}(x) \in \bar{D}$ such that $\|d(x) - \bar{d}(x)\| \leqslant L\,\|x - \bar{x}\|$. The inequality in point 3 follows. □

The next property will be useful for establishing the global convergence result of theorems 3.7 and 3.9.

Corollary 3.5 (local boundedness of the directions) *Suppose that F and G are continuously differentiable at $\bar{x} \in \mathbb{R}^n$, that $\tau \in (0, \infty]$ and that the PNM-regularity condition 3.1 holds at $\bar{x}$. Then, there is a constant C, such that, for x near $\bar{x}$, the system (2.13) has a solution d that satisfies $\|d\| \leqslant C$.*

Proof It is a consequence of proposition 3.4(2), since $\bar{D}$ is bounded by its finite cardinality. □

3.2 Global convergence

The global convergence results of this section accept directions d such that the right-hand side of (2.18) is sufficiently negative in the sense of (2.19a), an inequality that we reproduce here for the reader's convenience:

$$-\sum_{i \in [1:n]} (1 - \rho_i(x, d)) H_i(x)^2 \leqslant -2(1 - \eta)\,\theta(x), \tag{3.6}$$

where $\rho_i(x, d)$ is defined by (2.20), H is the function defined by (1.3b) and η is a constant (independent of k) such that $\eta < 1$. By proposition 2.7, this inequality implies that d is a descent direction of θ at x, since then

$$\theta'(x; d) \leqslant -2(1 - \eta)\,\theta(x), \tag{3.7}$$

and the right-hand side is negative when $\theta(x) \neq 0$, that is when x is not a solution to the NCP (1.1). It would have been less restrictive to impose the satisfaction of (3.7), instead of that of (3.6), but the technique used in the proof of theorem 3.6 below would have then required to have a reverse inequality in (2.18) in order to recover (3.6), since it is (3.6) that is required in the adopted proof; the reverse inequality in (2.18) looks problematic to us. Recall that inequality (3.6) simplifies into (2.19b).

We start the global convergence analysis with theorem 3.6, which assumes that the generic algorithm 2.10 generates a sequence $\{x_k\}$, hence is well-posed, and the boundedness of the direction subsequence $\{d_k\}_{k\in\mathcal{K}}$ when the subsequence $\{x_k\}_{k\in\mathcal{K}}$ of $\{x_k\}$ converges to some point $\bar{x}$. Conditions ensuring the convergence of the algorithms 2.11 and 3.8 will be examined in theorems 3.7 and 3.9, respectively. The proof of theorem 3.6 contains the main arguments. We have preferred presenting the convergence result in two stages (theorem 3.6 and theorems 3.7 and 3.9), since the boundedness assumption may be due to the structure of the problem, making the theorem useful in that circumstance. In theorems 3.7 and 3.9, which can also be viewed as corollaries of theorem 3.6, it is the assumed regularity of the limit point $\bar{x}$ that ensures the boundedness of $\{d_k\}_{k\in\mathcal{K}}$ and therefore the global convergence of the algorithm. These global convergence results of theorems 3.7 and 3.9 are rather weak since they assume that the generated sequence has a limit point (this will be certainly the case when this sequence is bounded) and that the limit point is regular in a certain sense (a typical assumption of linesearch methods). It may occur, however, that the generated sequence $\{x_k\}$ has no regular limit points, in which case the theorem provides no information. Nevertheless, it acts as a filter that the algorithms must pass, which was very useful to us for the design of an acceptation test (2.19)-(3.6) for the hybrid algorithm 3.8.

As a last remark on the assumptions, let us stress the fact that claiming that the algorithms generate a sequence $\{x_k\}$ implicitly assumes that the algorithms are not stuck at an iterate, for example because the system (2.13) has no solution in the case of the PNM and HNM algorithms 2.11 and 3.8. If this last event will not occur close to a point $\bar{x}$ satisfying the PNM-regularity 3.1 (corollary 3.5), this is not guaranteed far from such a point. Therefore, the global nature of the obtained convergence must be put into perspective.

Theorem 3.6 (global convergence of the generic NM algorithm 2.10) *Let F and $G : \Omega \to \mathbb{R}^n$ be differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the generic algorithm 2.10 generates a sequence $\{x_k\} \subseteq \Omega$. If $\bar{x} \in \Omega$ is an accumulation point of $\{x_k\}$, at which F' and G' are continuous, and if the subsequence $\{d_k : x_k \text{ is near } \bar{x}\}$ is bounded, then all the sequence $\{\theta(x_k)\}_{k\geqslant 1}$ converges to zero and $\bar{x}$ is a solution to (1.1).*

Proof By the Armijo inequality (2.24), the sequence $\{\theta(x_k)\}$ decreases; since this sequence is also bounded below (by zero), it converges. By the Armijo inequality (2.24) again and the fact that $\eta < 1$, it follows that

$$\lim_{k\to\infty} \alpha_k\, \theta(x_k) = 0. \tag{3.8}$$

Let us examine two complementary cases.

If $\limsup_{k\to\infty} \alpha_k > 0$ (or, equivalently, $\alpha_k \not\to 0$), there is a subsequence $\mathcal{K}' \subseteq \mathbb{N}$ such that $\{\alpha_k\}_{k\in\mathcal{K}'}$ is bounded away from zero. Then, (3.8) implies that $\lim_{k\to\infty,\, k\in\mathcal{K}'} \theta(x_k) = 0$ and actually $\lim_{k\to\infty} \theta(x_k) = 0$, since the sequence $\{\theta(x_k)\}$ decreases. By the continuity of θ, any accumulation point $\bar{x}$ of $\{x_k\}$ satisfies $\theta(\bar{x}) = 0$, which means that $\bar{x}$ solves (1.1). We have shown the conclusions of the theorem in that case.

We now consider the more difficult case when $\limsup_{k\to\infty} \alpha_k = 0$ (or, equivalently, $\alpha_k \to 0$). Let us first sketch the proof, which is inspired from that in [57]; see also [82]. Let $\{x_k\}_{k\in\mathcal{K}}$ be a subsequence converging to the accumulation point $\bar{x}$ ($k \to \infty$ in some infinite subset $\mathcal{K}$ of $\mathbb{N}$). With no loss of generality, one can assume that $\alpha_k < 1$, which implies that the stepsize $\hat{\alpha}_k := \alpha_k/\beta$ is rejected by the Armijo rule (2.24). Of course $\hat{\alpha}_k \to 0$. Let

$$\hat{x}_k := x_k + \hat{\alpha}_k d_k$$

be the corresponding rejected point. Then, $\theta(\hat{x}_k) > \theta(x_k) - 2\omega\hat{\alpha}_k(1-\eta)\,\theta(x_k)$ or

$$4\omega\hat{\alpha}_k(1-\eta)\,\theta(x_k) > 2[\theta(x_k) - \theta(\hat{x}_k)]. \tag{3.9}$$

The tactic of the proof consists in writing the right-hand side of this inequality as follows

$$2[\theta(x_k) - \theta(\hat{x}_k)] = \sum_{i=1}^{n} \Big[\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2\Big] \tag{3.10}$$

and to find a lower bound of each term of the sum in the right-hand side of the previous identity. More specifically, we shall show that, since $\{d_k\}_{k\in\mathcal{K}}$ is assumed to be bounded, for any $i \in [1:n]$ and any iterate x_k sufficiently close to $\bar{x}$, the following inequality holds

$$\begin{aligned} &\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\ &\qquad \geqslant 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k), \end{aligned} \tag{3.11}$$

where $\rho_{k,i}$ is an abbreviation for $\rho_i(x_k, d_k)$ and the term $o(\hat{\alpha}_k)$ means that $o(\hat{\alpha}_k)/\hat{\alpha}_k \to 0$ when $k \to \infty$ in $\mathcal{K}$. Then, the inequality (3.9), with its right-hand side bounded below thanks to the identity (3.10) and the inequalities (3.11), yields

$$\begin{aligned} &4\omega\hat{\alpha}_k(1-\eta)\,\theta(x_k) \\ &\quad \geqslant 2\hat{\alpha}_k \sum_{i\in[1:n]} (1-\rho_{k,i}) \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [(3.9), (3.10), (3.11)] \\ &\quad \geqslant 4\hat{\alpha}_k(1-\eta)\,\theta(x_k) + o(\hat{\alpha}_k). \qquad [(2.19)] \end{aligned}$$

After division by $4\hat{\alpha}_k(1-\eta)$, we get

$$\omega\,\theta(x_k) \geqslant \theta(x_k) + \frac{o(\hat{\alpha}_k)}{\hat{\alpha}_k}. \tag{3.12}$$

Taking the limit when $k \to \infty$ in $\mathcal{K}$ shows that $\omega\,\theta(\bar{x}) \geqslant \theta(\bar{x})$. Since $\omega \in (0,1)$ and $\theta(\bar{x}) \geqslant 0$, this implies that $\theta(\bar{x}) = 0$. Therefore, all the sequence $\{\theta(x_k)\}$ tends to zero and $\bar{x}$ solves (1.1). We have also shown the conclusions of the theorem in that case.

Therefore, to conclude the proof, we only have to show (3.11)*, for all* $i \in [1:n]$ *and* x_k *sufficiently close to* $\bar{x}$.

Since $\{d_k\}_{k\in\mathcal{K}}$ is bounded by assumption and $\alpha_k \to 0$, it follows that $\hat{x}_k \to \bar{x}$ when $k \to \infty$ in $\mathcal{K}$. Now, for $i \in [1:n]$, the differentiability of F_i and the mean value theorem provide

$$|F_i(\hat{x}_k) - F_i(x_k) - F_i'(x_k)(\hat{x}_k - x_k)| \leqslant \left(\sup_{z\in(x_k,\hat{x}_k)} \|F_i'(z) - F_i'(x_k)\|\right) \|\hat{x}_k - x_k\|,$$

where $(x_k, \hat{x}_k)$ is the open segment $\{(1-t)x_k + t\hat{x}_k : t \in (0,1)\}$. A similar estimation holds for G_i. By the continuity of F' at $\bar{x}$, the factor in parenthesis in the right-hand side tends to zero when $k \to \infty$ in $\mathcal{K}$. Using $\hat{x}_k - x_k = \hat{\alpha}_k d_k$ and the boundedness of $\{d_k\}$, we get

$$\begin{aligned} F_i(\hat{x}_k) &= F_i(x_k) + \hat{\alpha}_k F_i'(x_k) d_k + o(\hat{\alpha}_k), \\ G_i(\hat{x}_k) &= G_i(x_k) + \hat{\alpha}_k G_i'(x_k) d_k + o(\hat{\alpha}_k). \end{aligned}$$

Below, we shall need to give a lower bound on $F_i(x_k)^2 - F_i(\hat{x}_k)^2$ and $G_i(x_k)^2 - G_i(\hat{x}_k)^2$. By the previous estimates, we have

$$F_i(x_k)^2 - F_i(\hat{x}_k)^2 = -2\hat{\alpha}_k F_i(x_k) F_i'(x_k) d_k + o(\hat{\alpha}_k), \tag{3.13a}$$

$$G_i(x_k)^2 - G_i(\hat{x}_k)^2 = -2\hat{\alpha}_k G_i(x_k) G_i'(x_k) d_k + o(\hat{\alpha}_k). \tag{3.13b}$$

Let us now examine each term of the sum in (3.10) for the indices i in the following partition of $[1:n]$:

$$\Big(\mathcal{F}(\bar{x}),\;\; \mathcal{G}(\bar{x}),\;\; \mathcal{E}^+(\bar{x}),\;\; \mathcal{E}^-(\bar{x}),\;\; \mathcal{E}^0(\bar{x})\Big).$$

Note that τ does not intervene in that partition.

1. $i \in \mathcal{F}(\bar{x})$.

   By the strict inequality $F_i(\bar{x}) < G_i(\bar{x})$ defining $\mathcal{F}(\bar{x})$ in (1.8), the continuity of F and G at $\bar{x}$, and the fact that x_k is close to $\bar{x}$ when k is large in $\mathcal{K}$, we have $F_i(x_k) < G_i(x_k)$ or $i \in \mathcal{F}(x_k)$ for large k in $\mathcal{K}$. Let us show that

$$-F_i(x_k)F_i'(x_k)d_k \geqslant (1-\rho_{k,i})F_i(x_k)^2. \tag{3.14}$$

   One of the following three complementary cases must occurs.

   - If $F_i(x_k) = 0$, (3.14) is clearly verified with equality.
   - If $i \in \mathcal{F}(x_k) \setminus \mathcal{E}_\tau^-(x_k) \subseteq E_F(x_k)$ and $F_i(x_k) \neq 0$, $(2.20)_1$ gives $F_i'(x_k)d_k = -(1-\rho_{k,i})F_i(x_k)$. Multiplying both sides of this equality by $-F_i(x_k)$ yields (3.14) with equality.
   - If $i \in \mathcal{F}^-(x_k) \cap \mathcal{E}_\tau^-(x_k) \subseteq I(x_k)$ (in which case $F_i(x_k) < 0$), $(2.20)_5$ gives $F_i'(x_k)d_k \geqslant -(1-\rho_{k,i})F_i(x_k)$. Multiplying both sides of this inequality by $-F_i(x_k) > 0$ yields (3.14).

   Next, since $\hat{x}_k \to \bar{x}$ when $k \to \infty$ in $\mathcal{K}$ and since $F_i(\bar{x}) < G_i(\bar{x})$ when $i \in \mathcal{F}(\bar{x})$, one also has $F_i(\hat{x}_k) < G_i(\hat{x}_k)$. Therefore,

$$\begin{aligned}
&\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\
&= F_i(x_k)^2 - F_i(\hat{x}_k)^2 \qquad [F_i(x_k) < G_i(x_k) \text{ and } F_i(\hat{x}_k) < G_i(\hat{x}_k)] \\
&= -2\hat{\alpha}_k F_i(x_k)F_i'(x_k)d_k + o(\hat{\alpha}_k) \qquad [(3.13a)] \\
&\geqslant 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.14)] \\
&= 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) < G_i(x_k)].
\end{aligned}$$

   We have obtained the desired inequality (3.11).

2. $i \in \mathcal{G}(\bar{x})$.

   One can proceed like in case 1, by switching the roles of F and G [38].

3. $i \in \mathcal{E}^+(\bar{x})$.

   In this case, $F_i(x_k)$, $G_i(x_k)$, $F_i(\hat{x}_k)$ and $G_i(\hat{x}_k)$ are positive for k large in $\mathcal{K}$, which implies that i is in one of the sets $\mathcal{F}^+(x_k) \cup \mathcal{E}_\mathcal{F}^{0+}(x_k)$ or $\mathcal{G}^+(x_k) \cup \mathcal{E}_\mathcal{G}^{0+}(x_k)$, where

$$\mathcal{F}^+(x) := \{i \in \mathcal{F}(x) : F_i(x) > 0\} \qquad \text{and} \qquad \mathcal{G}^+(x) := \{i \in \mathcal{G}(x) : G_i(x) > 0\}.$$

   We now consider these sets one after the other.

   3.1. $i \in \mathcal{F}^+(x_k) \cup \mathcal{E}_\mathcal{F}^{0+}(x_k)$.

   In this case, $0 < F_i(x_k) \leqslant G_i(x_k)$. Because $i \in [\mathcal{F}(x_k) \setminus \mathcal{E}_{\tau_k}^-(x_k)] \cup \mathcal{E}_\mathcal{F}^{0+}(x_k) = E_F(x_k)$ and $F_i(x_k) \neq 0$, $(2.20)_1$ tells us that $F_i'(x_k)d_k = -(1-\rho_{k,i})F_i(x_k)$ and finally

$$-F_i(x_k)F_i'(x_k)d_k = (1-\rho_{k,i})F_i(x_k)^2. \tag{3.15}$$

   Therefore, for k large in $\mathcal{K}$:

$$\begin{aligned}
&\underbrace{\min(F_i(x_k), G_i(x_k))}_{=F_i(x_k)}{}^2 - \underbrace{\min(F_i(\hat{x}_k), G_i(\hat{x}_k))}_{0 \leqslant \cdot \leqslant F_i(\hat{x}_k)}{}^2 \\
&\geqslant F_i(x_k)^2 - F_i(\hat{x}_k)^2 \\
&= -2\hat{\alpha}_k F_i(x_k)F_i'(x_k)d_k + o(\hat{\alpha}_k) \qquad [(3.13a)] \\
&= 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.15)] \\
&= 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) \leqslant G_i(x_k)].
\end{aligned}$$

   We have obtained the desired inequality (3.11).

3.2. $i \in \mathcal{G}^+(x_k) \cup \mathcal{E}_{\mathcal{G}}^{0+}(x_k)$.

One can proceed like in case 3.1, by switching the roles of F and G [38].

4. $i \in \mathcal{E}^-(\bar{x})$.

In this case, for k large in $\mathcal{K}$, $F_i(x_k)$, $G_i(x_k)$, $F_i(\hat{x}_k)$ and $G_i(\hat{x}_k)$ are negative and $|F_i(x_k) - G_i(x_k)| < \tau$, so that $i \in \mathcal{E}_\tau^-(x_k) = I(x_k)$. Then, by (2.20)$_5$,

$$F_i'(x_k)d_k \geqslant -(1-\rho_{k,i})F_i(x_k), \tag{3.16a}$$

$$G_i'(x_k)d_k \geqslant -(1-\rho_{k,i})G_i(x_k), \tag{3.16b}$$

so that

$$-F_i(x_k)F_i'(x_k)d_k \geqslant (1-\rho_{k,i})F_i(x_k)^2, \tag{3.17a}$$

$$-G_i(x_k)G_i'(x_k)d_k \geqslant (1-\rho_{k,i})G_i(x_k)^2. \tag{3.17b}$$

Now, one (or both) of the following two cases must occur.

4.1. $F_i(x_k) \leqslant G_i(x_k)$, which are both negative. We divide the analysis of this case into two complementary subcases.

4.1.1. $F_i(\hat{x}_k) \leqslant G_i(\hat{x}_k)$, which are both negative.

For k large in $\mathcal{K}$, the following holds

$$\begin{aligned}
&\min(F_i(x_k), G_i(x_k))^2 - \min(F_i(\hat{x}_k), G_i(\hat{x}_k))^2 \\
&\quad = F_i(x_k)^2 - F_i(\hat{x}_k)^2 \qquad [F_i(x_k) \leqslant G_i(x_k) \text{ and } F_i(\hat{x}_k) \leqslant G_i(\hat{x}_k)] \\
&\quad = -2\hat{\alpha}_k F_i(x_k) F_i'(x_k) d_k + o(\hat{\alpha}_k) \qquad [(3.13a)] \\
&\quad \geqslant 2(1-\rho_{k,i})\hat{\alpha}_k F_i(x_k)^2 + o(\hat{\alpha}_k) \qquad [(3.17a)] \\
&\quad = 2(1-\rho_{k,i})\hat{\alpha}_k \min(F_i(x_k), G_i(x_k))^2 + o(\hat{\alpha}_k) \qquad [F_i(x_k) \leqslant G_i(x_k)].
\end{aligned}$$

We have obtained the desired inequality (3.11).

4.1.2. $G_i(\hat{x}_k) < F_i(\hat{x}_k)$, which are both negative.

Let us show that

$$2(1-\rho_{k,i})\hat{\alpha}_k \leqslant 1, \quad \text{for } k \text{ large in } \mathcal{K}. \tag{3.18}$$

This is certainly the case when $\rho_{k,i} \geqslant 0$, since then, $2(1-\rho_{k,i})\hat{\alpha}_k \leqslant 2\hat{\alpha}_k \leqslant 1$ because $\hat{\alpha}_k \to 0$ for $k \to \infty$ in $\mathcal{K}$. When $\rho_{k,i} < 0$, we use (3.16a), which also reads

$$\rho_{k,i}\, F_i(x_k) \leqslant F_i(x_k) + F_i'(x_k)d_k.$$

Hence, for k large enough in $\mathcal{K}$:

$$0 < \frac{1}{2}\,\rho_{k,i}\, F_i(\bar{x}) \leqslant \rho_{k,i}\, F_i(x_k) \leqslant F_i(x_k) + F_i'(x_k)d_k \leqslant C,$$

where the constant $C > 0$ comes for the fact that $x_k \to \bar{x}$ for $k \to \infty$ in $\mathcal{K}$, from the assumed continuity of F' at $\bar{x}$, and from the assumed boundedness of $\{d_k\}$. This shows that $\rho_{k,i}$ is bounded below, so that (3.18) also holds when $\rho_{k,i} < 0$. Then, the following holds

$$\begin{aligned}
&\min(F_i(x_k),G_i(x_k))^2-\min(F_i(\hat{x}_k),G_i(\hat{x}_k))^2\\
&=F_i(x_k)^2-G_i(\hat{x}_k)^2 \qquad [F_i(x_k)\leqslant G_i(x_k) \text{ and } G_i(\hat{x}_k)<F_i(\hat{x}_k)]\\
&=G_i(x_k)^2-G_i(\hat{x}_k)^2+F_i(x_k)^2-G_i(x_k)^2\\
&=-2\hat{\alpha}_kG_i(x_k)G_i'(x_k)d_k+F_i(x_k)^2-G_i(x_k)^2+o(\hat{\alpha}_k) \qquad [(3.13b)]\\
&\geqslant 2(1-\rho_{k,i})\hat{\alpha}_kG_i(x_k)^2+F_i(x_k)^2-G_i(x_k)^2+o(\hat{\alpha}_k) \qquad [(3.17b)]\\
&=2(1-\rho_{k,i})\hat{\alpha}_kF_i(x_k)^2+(1-2(1-\rho_{k,i})\hat{\alpha}_k)(F_i(x_k)^2-G_i(x_k)^2)+o(\hat{\alpha}_k)\\
&\geqslant 2(1-\rho_{k,i})\hat{\alpha}_kF_i(x_k)^2+o(\hat{\alpha}_k) \qquad [(3.18) \text{ and } F_i(x_k)^2\geqslant G_i(x_k)^2]\\
&=2(1-\rho_{k,i})\hat{\alpha}_k\min(F_i(x_k),G_i(x_k))^2+o(\hat{\alpha}_k) \qquad [F_i(x_k)\leqslant G_i(x_k)].
\end{aligned}$$

We have obtained the desired inequality (3.11).

4.2. $G_i(x_k)\leqslant F_i(x_k)$, which are both negative. One can proceed like in case 4.1, by switching the roles of F and G [38].

5. $i\in\mathcal{E}^0(\bar{x})$.

In this case, we write

$$\begin{aligned}
&\min(F_i(x_k),G_i(x_k))^2-\min(F_i(\hat{x}_k),G_i(\hat{x}_k))^2\\
&=\Big(\min(F_i(x_k),G_i(x_k))-\min(F_i(\hat{x}_k),G_i(\hat{x}_k))\Big)\\
&\quad\times\Big(\min(F_i(x_k),G_i(x_k))+\min(F_i(\hat{x}_k),G_i(\hat{x}_k))\Big).
\end{aligned}$$

Since $x\mapsto\min(F(x),G(x))$ is Lipschitz continuous near $\bar{x}$, the first factor in the right-hand side is bounded by a constant times $\|\hat{x}_k-x_k\|$, which is an $O(\hat{\alpha}_k)$ by the boundedness of $\{d_k\}$, while the second factor in the right-hand side converges to zero (since in this case $F_i(\bar{x})=G_i(\bar{x})=0$). Thus the whole term is $o(\hat{\alpha}_k)$. This is enough to get (3.11), since the first term in the right-hand side of (3.11) is also an $o(\hat{\alpha}_k)$, so that the left-hand side of (3.11) minus the first term in its right-hand side is indeed (larger than) an $o(\hat{\alpha}_k)$. □

Theorem 3.7 (global convergence of the PNM algorithm 2.11) *Let F and $G:\Omega\to\mathbb{R}^n$ be differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the PNM algorithm 2.11 generates a sequence $\{x_k\}\subseteq\Omega$. If $\bar{x}\in\Omega$ is an accumulation point of $\{x_k\}$ that is PNM-regular in the sense of definition 3.1 and if F' and G' are continuous at $\bar{x}$, then all the sequence $\{\theta(x_k)\}_{k\geqslant1}$ converges to zero and $\bar{x}$ is a solution to (1.1).*

Proof According to theorem 3.6, we just have to prove that the subsequence $\{d_k : x_k \text{ is near } \bar{x}\}$ is bounded. Since the directions d_k are computed by (2.25), this property is given by corollary 3.5, which rests on the additional assumption on the PNM-regularity at $\bar{x}$, in the sense of definition 3.1. □

3.3 A hybrid Newton-min algorithm

This section presents a hybrid Newton-min (HNM) algorithm for solving the complementarity problem (1.1), which benefits from both a global convergence in the style of theorem 3.6 and a fast local convergence in the style of theorem 2.3. Around an NM-regular point, by its ability to provide superlinear/quadratic convergence (recalled in theorem 2.3), the displacement d^{NM} computed by (2.1) reduces significantly the least-square merit function θ, in the sense (3.19d) below. For this reason, the HNM algorithm stated below proposes to compute an NM direction $d^{\rm NM}$ at each iteration and to check its relevance first on the decrease of θ (test 3.2 of the HNM algorithm) and next, if necessary, on the descent condition (2.19) (test 3.3 of the HNM algorithm). This rule gives two chances to accept the cheap NM direction $d^{\rm NM}$.

The extra cost coming from the computation of an NM direction (a single linear system to solve) should be modest with respect to the cost of the computation of a secure PNM direction by (2.25) (solving a convex quadratic optimization problem) and has the advantage of sometimes avoiding the computation of this latter direction (we have observed this trend in numerical experiments on LCPs [50]). Here is the algorithm. It is in the style of algorithm 9.2.3 in [43].

Algorithm 3.8 (HNM algorithm) It is the instance of the generic algorithm 2.10, in which the direction d in step 3 is computed as follows.

3.1. For some partition $(\tilde{\mathcal{F}}(x), \tilde{\mathcal{G}}(x)) \in \mathcal{P}([1:n])$ that satisfies $\tilde{\mathcal{F}}(x) \supseteq \mathcal{F}(x)$ and $\tilde{\mathcal{G}}(x) \supseteq \mathcal{G}(x)$, compute a plain NM direction $d^{\rm NM} \in \mathbb{R}^n$ as a solution d to (2.1).
3.2. If (2.24) holds with $d = d^{\rm NM}$ and $\alpha = 1$, go to step 5 with $\alpha = 1$.
3.3. 3.3.1. If (2.19) holds with $d = d^{\rm NM}$, set $d = d^{\rm NM}$.
3.3.2. Else, compute d as a solution to (2.25).

In step 3.1, it is implicitely assumed that the linear system (2.1) has a solution. Note that, instead of $d^{\rm NM}$, one could compute any direction by an algorithm satisfying the conclusion of theorem 2.3. The convergence assumptions of theorem 3.9 below must then be adapted to those ensuring the conclusions of theorem 2.3 for that direction.

In the proof of the next theorem, one uses the notation $\{u_k\} \sim \{v_k\}$, for two sequences $\{u_k\}$ and $\{v_k\}$ in a normed space $(\mathbb{E}, \|\cdot\|)$, to mean that $\|u_k\| = O(\|v_k\|)$ and $\|v_k\| = O(\|u_k\|)$.

Theorem 3.9 (global and fast local convergence of the HNM algorithm 3.8) *Suppose that F and $G : \Omega \to \mathbb{R}^n$ are continuously differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the HNM algorithm 3.8 generates a sequence $\{x_k\} \subseteq \Omega$ and that $\bar{x} \in \Omega$ is an accumulation point of $\{x_k\}$ that is NM and PNM-regular, in the sense of definitions 2.1 and 3.1. Then,*

1) *$\{\theta(x_k)\}_{k\geqslant 1}$ converges monotonically to zero and $\bar{x}$ is a solution to (1.1),*
2) *$\{x_k\}$ converges superlinearly to $\bar{x}$,*
3) *if F' and G' are Lipschitz near $\bar{x}$, the convergence is quadratic,*
4) *if F and G are affine near $\bar{x}$, the convergence is finite.*

Proof [*Property* 1*: global convergence*] If the plain NM direction $d_k^{\rm NM}$ computed in step 3.1 of algorithm 3.8 satisfies (2.24) with $\alpha = 1$ (i.e., the test in step 3.2 is passed) infinitely often, then $\theta(x_k) \downarrow 0$. Indeed, for the iterations k at which $d_k = d_k^{\rm NM}$ passes the test in step 3.2, one has $x_{k+1} = x_k + d_k^{\rm NM}$ and $\theta(x_{k+1}) \leqslant (1-2\omega(1-\eta))\theta(x_k)$, where $1-2\omega(1-\eta)$ is a constant in $(0,1)$. For the other iterations, one has $\theta(x_{k+1}) \leqslant \theta(x_k)$. The monotone convergence of $\theta(x_k)$ to zero follows. By the continuity of θ and the convergence of a subsequence of $\{x_k\}$ to $\bar{x}$, one deduces that $\theta(\bar{x}) = 0$, i.e., $\bar{x}$ is a solution to (1.1). Property 1 is proved in that case.

Otherwise, after a finite number of iterations, a direction satisfying (2.19) is computed (it can be either the plain NM direction $d_k^{\rm NM}$ computed in step 3.1 or a secure PNM direction that solves (2.25), the latter satisfying (2.19) by corollary 2.8) and a linesearch is performed in step 4 to get (2.24). Since the accumulation point $\bar{x}$ is assumed to be NM and PNM-regular, the computed directions d_k are bounded for x_k near $\bar{x}$, thanks to proposition 2.2 and corollary 3.5. Property 1 now follows from theorem 3.6, since both directions satisfy (2.19).

[*Properties* 2, 3 *and* 4*: fast local convergence*] Our goal is to apply theorem 2.3, whose main assumptions hold. Let us first prove that $x_{k+1} = x_k + d_k^{\rm NM}$ when x_k is in a neighborhood of $\bar{x}$.

Let $V \subseteq \Omega$ be the open neighborhood of $\bar{x}$ given by theorem 2.3. Since $\bar{x}$ is an accumulation point of the generated sequence $\{x_k\}$, there is a subsequence $\{x_k\}_{k\in\mathcal{K}} \subseteq V$, with $\mathcal{K} \subseteq \mathbb{N}$, that converges to $\bar{x}$. By theorem 2.3,

$$\|x_k + d_k^{\mathrm{NM}} - \bar{x}\| = o(\|x_k - \bar{x}\|), \quad \text{for } k \in \mathcal{K}. \tag{3.19a}$$

It follows that ([17; lemme 13.5] for instance)

$$\{d_k^{\mathrm{NM}}\}_{k\in\mathcal{K}} \sim \{x_k - \bar{x}\}_{k\in\mathcal{K}}. \tag{3.19b}$$

Since d_k^{NM} satisfies (2.1) and since the matrices of the linear system are bounded and have bounded inverses, thanks to the NM-regularity, one also has

$$\{d_k^{\mathrm{NM}}\}_{k\in\mathcal{K}'} \sim \{H(x_k)\}_{k\in\mathcal{K}'}, \tag{3.19c}$$

for some subsequence $\mathcal{K}' \subseteq \mathcal{K}$. Combining (3.19b) and (3.19c), one gets $\{H(x_k)\}_{k\in\mathcal{K}'} \sim \{x_k - \bar{x}\}_{k\in\mathcal{K}'}$ and therefore, using (3.19a), $\|H(x_k + d_k^{\mathrm{NM}})\| = o\|H(x_k)\|$, for $k \in \mathcal{K}'$, and finally

$$\theta(x_k + d_k^{\mathrm{NM}}) = o(\theta(x_k)), \quad \text{for } k \in \mathcal{K}'. \tag{3.19d}$$

This implies that there is an index $k_1 \in \mathcal{K}'$ such that d_k^{NM} is accepted in step 3.2 of the HNM algorithm and $x_{k+1} = x_k + d_k^{\mathrm{NM}}$, far all $k \geqslant k_1$ in $\mathbb{N}$. Therefore, from that index k_1, theorem 2.3 applies. Since the assumptions of theorem 2.3 hold, this one yields the properties 2, 3 and 4 in the statement of the present theorem. □

According to the first part of the proof of the previous theorem, the NM-regularity intervenes in the *global* convergence for the following reason. It may happen that step 3.2 of algorithm 3.8 is successful only finitely often but that the NM direction d^{NM} is accepted in step 3.3.1 infinitely often. Then, to be able to apply theorem 3.6, it must be shown that d^{NM} remains bounded, which is guaranteed by the NM-regularity. This discussion shows that it is easy to modify algorithm 3.8 so that the NM-regularity is no longer necessary in obtaining *global* convergence (it is still useful for fast *local* convergence, however): one accepts d^{NM} only in step 3.2, not in step 3.3. We believe that this option is less efficient numerically, since then the more expensive PNM direction is more often computed by (2.25). Nevertheless, to weaken the condition of global convergence, we present such a variation of algorithm 3.8 as algorithm 3.10 below.

Algorithm 3.10 (HNM' algorithm) It is the instance of algorithm 3.8, in which step 3.3 is replaced as follows.

3.3. Compute the direction d as a solution to (2.25).

Theorem 3.9 then takes the following form.

Theorem 3.11 (global and fast local convergence of the HNM' algorithm 3.10) *Suppose that F and $G : \Omega \to \mathbb{R}^n$ are continuously differentiable functions defined on an open set Ω of $\mathbb{R}^n$. Suppose that the HNM' algorithm 3.10 generates a sequence $\{x_k\} \subseteq \Omega$ and that $\bar{x} \in \Omega$ is an accumulation point of $\{x_k\}$ that is PNM-regular, in the sense of definition 3.1. Then,*

1) *$\{\theta(x_k)\}_{k\geqslant 1}$ converges monotonically to zero and $\bar{x}$ is a solution to (1.1).*

If the NM-regularity also holds at $\bar{x}$, then

2) *$\{x_k\}$ converges superlinearly to $\bar{x}$,*
3) *if F' and G' are Lipschitz near $\bar{x}$, the convergence is quadratic,*
4) *if F and G are affine near $\bar{x}$, the convergence is finite.*

Proof Only the second paragraph of the proof of theorem 3.9, which uses the NM-regularity to get claim 1, must be adapted. This paragraph changes since, with algorihm 3.10, at this point of the proof, the sequence $\{x_k\}$ is generated by linesearch along secure PNM directions computed by (2.25). These directions satisfy (2.19) (corollary 2.8) and are bounded near $\bar{x}$ by the PNM-regularity and the minimal property resulting from the optimization problem in (2.25) (corollary 3.5). Then, theorem 3.6 can be applied and this one yields claim 1. □

4 Conclusion

This paper presents algorithms for solving the complementarity problem (1.1), based on semismooth-like iterations on the nonsmooth equation (1.3), reformulating the problem with the minimum function. In practice, this solution strategy is often more efficient than with other reformulations but it is difficult to implement up to completeness, because the associated least-square merit function may not decrease along the semismooth direction. The paper proposes to overcome the difficulty by slightly modifying this direction in the neighborhood of the negative kinks of the minimum function. A global convergence result can be established for the resulting algorithm 2.11, provided some specific regularity condition holds at the accumulation points of the generated sequence (the PNM regularity).

A hybrid algorithm is also proposed, which tries to use the inexpensive plain Newton-min direction at each iteration. If the sequence generated by the resulting algorithm 3.8 has an accumulation point that is regular (NM and PNM regularities), this point is necessarily a solution to the complementarity problem and all the sequence converges to it. A similar result is obtained with algorithm 3.10, with only the PNM regularity of the accumulation point. The convergence is also fast (superlinear or quadratic, depending on the smoothness of the functions defining the problem).

Algorithms 2.11 and 3.8 can be used to solve linear complementarity problems.

A number of issues still need to be considered to improve the robustness of the proposed algorithms, to finalize their analysis, to estimate their complexity and to enhance their attractiveness. Some of them are explored in [92,37,39,40] and others will be considered in subsequent contributions.

Acknowledgments

We thank the referees for their remarks and recommendations, which have helped us improve the readability and quality of the paper, in particular with the addition of the asymptotic analysis of section 3.3.

References

1. Vincent Acary and Bernard Brogliato. *Numerical Methods for Nonsmooth Dynamical Systems - Applications in Mechanics and Electronics.* Number 35 in Lecture Notes in Applied and Computational Mechanics. Springer, 2008. [doi]. 2
2. Muhamed Aganagić. Newton's method for linear complementarity problems. *Mathematical Programming*, 28:349–362, 1984. [doi]. 3
3. Shmuel Agmon. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:382–392, 1954. 12
4. Jan Harold Alcantara, Chen-Han Lee, Chieu Thanh Nguyen, Yu-Lin Chang, and Jein-Shan Chen. On construction of new NCP functions. *Operations Research Letters*, 48(2):115–121, 2010. [doi]. 3
5. Mihai Anitescu and Florian Alexandru Potra. Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. *Nonlinear Dynamics*, 14(3):231–247, 1997. [doi]. 2
6. Larry Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966. [doi]. 19

7. Laurence Beaude, Konstantin Brenner, Simon Lopez, Roland Masson, and Farid Smai. Non-isothermal compositional liquid gas Darcy flow: formulation, soil-atmosphere boundary condition and application to high-energy geothermal simulations. *Computational Geosciences*, 23(3):443–470, 2019. [doi]. 2
8. Ibtihel Ben Gharbia. *Résolution de Problèmes de Complémentarité – Application à un Écoulement Diphasique Dans un Milieu Poreux*. PhD thesis, Université Paris-Dauphine, 2012. 9, 11
9. Ibtihel Ben Gharbia, Jad Dabaghi, Vincent Martin, and Martin Vohralík. A posteriori error estimates for a compositional two-phase flow with nonlinear complementarity constraints. *Computational Geosciences*, 24(3):1031–1055, 2020. [doi]. 2
10. Ibtihel Ben Gharbia and Éric Flauraud. Study of compositional multiphase flow formulation using complementarity conditions. *Oil & Gas Sciences and Technology*, 74:1–15, 2019. [doi]. 2
11. Ibtihel Ben Gharbia and Jean Charles Gilbert. Nonconvergence of the plain Newton-min algorithm for linear complementarity problems with a P-matrix. *Mathematical Programming*, 134:349–364, 2012. [doi]. 3
12. Ibtihel Ben Gharbia and Jean Charles Gilbert. An algorithmic characterization of P-matricity. *SIAM Journal on Matrix Analysis and Applications*, 34(3):904–916, 2013. [doi]. 3
13. Ibtihel Ben Gharbia and Jean Charles Gilbert. An algorithmic characterization of P-matricity II: adjustments, refinements, and validation. *SIAM Journal on Matrix Analysis and Applications*, 40(2):800–813, 2019. [doi]. 3
14. Ibtihel Ben Gharbia and Jérôme Jaffré. Gas phase appearance and disappearance as a problem with complementarity constraints. *Mathematics and Computers in Simulation*, 99:28–36, 2014. [doi]. 2
15. J. Frédéric Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994. [doi]. 3
16. J. Frédéric Bonnans, Jean Charles Gilbert, Claude Lemaréchal, and Claudia Sagastizábal. *Optimisation Numérique – Aspects théoriques et pratiques*. Number 27 in Mathématiques et Applications. Springer Verlag, Berlin, 1997. [editor]. 3
17. J. Frédéric Bonnans, Jean Charles Gilbert, Claude Lemaréchal, and Claudia Sagastizábal. *Numerical Optimization – Theoretical and Practical Aspects* (second edition). Universitext. Springer Verlag, Berlin, 2006. [authors] [editor] [doi]. 3, 11, 18, 19, 30
18. Bernard Brogliato. *Nonsmooth Mechanics - Models, Dynamics and Control* (third edition). Springer, 2016. [doi]. 2
19. Hannes Buchholzer, Christian Kanzow, Peter Knabner, and Serge Kräutle. The semismooth Newton method for the solution of reactive transport problems including mineral precipitation-dissolution reactions. *Computational Optimization and Applications*, 50(2):193–221, 2011. [doi]. 2
20. Quan M. Bui and Howard C. Elman. Semi-smooth Newton methods for nonlinear complementarity formulation of compositional two-phase flow in porous media. *Journal of Computational Physics*, 407:109163, 2020. [doi]. 2
21. Bintong Chen, Xiaojun Chen, and Christian Kanzow. A penalized Fischer-Burmeister NCP-function. *Mathematical Programming*, 88:211–216, 2000. [doi]. 3
22. Alice Chiche and Jean Charles Gilbert. How the augmented Lagrangian algorithm can deal with an infeasible convex quadratic optimization problem. *Journal of Convex Analysis*, 23(2):425–459, 2016. [pdf] [editor]. 12
23. Sergei Chubanov. A polynomial projection algorithm for linear feasibility problems. *Mathematical Programming*, 153:687–713, 2015. [doi]. 12
24. Sung Jin Chung. NP-completeness of the linear complementarity problem. *Journal of Optimization Theory and Applications*, 60:393–399, 1989. [doi]. 2
25. Frank H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983. Reprinted in 1990 by SIAM, Classics in Applied Mathematics 5 [doi]. 5
26. Richard Warren Cottle. *Nonlinear Programs with Positively Bounded Jacobians*. PhD thesis, University of California, Berkeley, USA, 1964. 3
27. Richard Warren Cottle. Nonlinear programs with positively bounded jacobians. *SIAM Journal on Applied Mathematics*, 14:147–158, 1966. [doi]. 3
28. Richard Warren Cottle. Linear complementarity since 1978. In *Variational analysis and applications*, number 79 in Nonconvex Optimization and Its Applications, pages 239–257. Springer, New York, 2005. [doi]. 2
29. Richard Warren Cottle, Jong-Shi Pang, and Richard E. Stone. *The Linear Complementarity Problem*. Number 60 in Classics in Applied Mathematics. SIAM, Philadelphia, PA, USA, 2009. [doi]. 2
30. Jad Dabaghi, Vincent Martin, and Martin Vohralík. Adaptive inexact semismooth Newton methods for the contact problem between two membranes. *Journal of Scientific Computing*, 84(2), 2020. [doi]. 2
31. Jesús A. De Loera, Jamie Haddock, and Deanna Needell. A sampling Kaczmarz-Motzkin algorithm for linear feasibility. *SIAM Journal on Scientific Computing*, 39(5):566–587, 2017. [doi]. 12
32. Tecla De Luca, Francisco Facchinei, and Christian Kanzow. A semismooth equation approach to the solution of nonlinear complementarity problems. *Mathematical Programming*, 75:407–439, 1996. [doi]. 3
33. Tecla De Luca, Francisco Facchinei, and Christian Kanzow. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Computational Optimization and Applications*, 16:173–205, 2000. [doi]. 3, 6

34. Frédéric Delbos and Jean Charles Gilbert. Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. *Journal of Convex Analysis*, 12(1):45–69, 2005. [preprint] [editor]. 12
35. Vladimir Fedorovich Demyanov and Alexander M. Rubinov. On quasidifferentiable mappings. *Optimization*, 14(1):3–21, 1983. [doi]. 4
36. John E. Dennis and Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Number 16 in Classics in Applied Mathematics. SIAM, Philadelphia, PA, USA, 2009. [doi]. 3, 19
37. Jean-Pierre Dussault, Mathieu Frappier, and Jean Charles Gilbert. A lower bound on the iterative complexity of the Harker and Pang globalization technique of the Newton-min algorithm for solving the linear complementarity problem. *EURO Journal on Computational Optimization*, 7(4):359–380, 2019. [doi]. 31
38. Jean-Pierre Dussault, Mathieu Frappier, and Jean Charles Gilbert. Polyhedral Newton-min algorithms for complementarity problems – The full report. Research Report, 2023. [hal-04097624]. 6, 10, 17, 26, 27, 28
39. Jean-Pierre Dussault and Jean Charles Gilbert. Exact computation of an error bound for the balanced linear complementarity problem with unique solution. *Mathematical Programming*, 199(1-2):1221–1238, 2023. [doi]. 2, 31
40. Jean-Pierre Dussault, Jean Charles Gilbert, and Baptiste Plaquevent-Jourdain. On the B-differential of the componentwise minimum of two affine vector functions. *Mathematical Programming Computation* (in revision), 2023. [hal-03872711]. 2, 6, 8, 10, 31
41. Jean-Pierre Dussault, Jean Charles Gilbert, and Baptiste Plaquevent-Jourdain. Partial description of the B-differential of the componentwise minimum of two vector functions by linearization. Research Report (in preparation), 2024. 8
42. Kenny Erleben. Numerical methods for linear complementarity problems in physics-based animation. In *ACM SIGGRAPH 2013 Courses*, SIGGRAPH '13, pages 8:1–8:42, New York, NY, USA, 2013. ACM. [doi]. 2
43. Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems* (two volumes). Springer Series in Operations Research. Springer, 2003. 2, 3, 5, 6, 7, 8, 9, 10, 29
44. Michael C. Ferris and Christian Kanzow. Complementarity and related problems. In P.M. Pardalos and M.G.C. Resende, editors, *Handbook of Applied Optimization*. Oxford University Press, New York, NY, USA, 2002. 3
45. Michael C. Ferris, Christian Kanzow, and Todd Steven Munson. Feasible descent algorithms for mixed complementarity problems. *Mathematical Programming*, 86:475–497, 1999. [doi]. 6
46. Michael C. Ferris and Jong-Shi Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997. [doi]. 2
47. Andreas Fischer. A special Newton-type optimization method. *Optimization*, 24:269–284, 1992. [doi]. 3
48. Andreas Fischer and Houyuan Jiang. Merit functions for complementarity and related problems: a survey. *Computational Optimization and Applications*, 17:159–182, 2000. [doi]. 3
49. Andreas Fischer and Christian Kanzow. On finite termination of an iterative method for linear complementarity problems. *Mathematical Programming*, 74:279–292, 1996. [doi]. 3
50. Mathieu Frappier. Reformulation semi-lisse appliquée au problème de complémentarité. Master's thesis, Département de Mathématiques, Faculté des Sciences, Université de Sherbrooke, Canada, 2019. [internet]. 6, 12, 15, 29
51. Aurél Galántai. Properties and construction of NCP functions. *Computational Optimization and Applications*, 52(3):805–824, 2012. [doi]. 3
52. Yu Gao, Haiming Song, Xiaoshen Wang, and Kai Zhang. Primal-dual active set method for pricing American better-of option on two assets. *Communications in Nonlinear Science and Numerical Simulation*, 80, 2020. [doi]. 2
53. Jean Charles Gilbert. *Fragments d'Optimisation Différentiable – Théorie et Algorithmes*. Lecture Notes (in French) of courses given at ENSTA and at Paris-Saclay University, Saclay, France, 2021. [hal-03347060, pdf]. 18, 19
54. Jean-Louis Goffin. On the finite convergence of the relaxation method for solving systems of inequalitites. Operations Research Center Report ORC 71-36, Univ. of California at Berkeley, 1971. 12
55. Jean-Louis Goffin. The relaxation method for solving systems of linear inequalities. *Mathematics of Operations Research*, 5:388–414, 1980. [doi]. 12
56. Ji Ye Han and De Feng Sun. Newton and quasi-Newton methods for normal maps with polyhedral sets. *Journal of Optimization Theory and Applications*, 94:659–676, 1997. [doi]. 5
57. Shih-Ping Han, Jong-Shi Pang, and Narayan Rangaraj. Globally convergent Newton methods for nonsmooth equations. *Mathematics of Operations Research*, 17:586–607, 1992. [doi]. 5, 24
58. Patrick T. Harker and Jong-Shi Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990. [doi]. 2

59. Alan J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standard*, 49:263–265, 1952. 22
60. Tim Hoheisel, Christian Kanzow, Boris S. Mordukhovich, and Hung M. Phan. Generalized Newton's method based on graphical derivatives. *Nonlinear Analysis*, 75(3):1324–1340, 2012. [doi]. 3
61. George Isac. *Complementarity Problems.* Number 1528 in Lecture Notes in Mathematics. Springer, Berlin, 1992. [doi]. 2
62. Kazufumi Ito and Karl K. Kunisch. On a semi-smooth Newton method and its globalization. *Mathematical Programming*, 118:347–370, 2009. [doi].
63. Alexey F. Izmailov and Mikhail V. Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering. Springer, 2014. [doi]. 2, 3, 5
64. Norman H. Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI, USA, 1979. 3
65. Christian Kanzow and Helmut Kleinmichel. A new class of semismooth Newton-type methods for nonlinear complementarity problems. *Computational Optimization and Applications*, 3:227–251, 1998. [doi]. 3
66. Christian Kanzow, Nobuo Yamashita, and Masao Fukushima. New NCP-functions and their properties. *Journal of Optimization Theory and Applications*, 94(1):115–135, 1997. [doi]. 3
67. S. Karamardian. Generalized complementarity problem. *Journal of Optimization Theory and Applications*, 8(3):161–168, 1971. [doi]. 3
68. Masakazu Kojima, Nimrod Megiddo, Toshihito Noma, and Akiko Yoshise. *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems.* Number 538 in Lecture Notes in Computer Science. Springer, Berlin, 1991. [doi]. 2
69. Masakazu Kojima and Susumu Shindo. Extension of Newton and quasi-Newton methods to systems of PC^1 equations. *Journal of Operations Research Society of Japan*, 29:352–375, 1986. [doi]. 6
70. Serge Kräutle. The semismooth Newton method for multicomponent reactive transport with minerals. *Advances in Water Resources*, 34(1):137–151, 2011. [doi]. 2
71. Xudong Li, Defeng Sun, and Kim-Chuan Toh. On the efficient computation of a generalized Jacobian of the projector over the Birkhoff polytope. *Mathematical Programming*, 179:419–446, 2020. [doi]. 5
72. Vera Lúcia Rocha Lopes, José Mario Martínez, and Rosana Pérez. On the local convergence of quasi-Newton methods for nonlinear complementarity problems. *Applied Numerical Mathematics*, 30:3–22, 1999. [doi]. 6
73. Zhi-Quan Luo and Paul Tseng. Perturbation analysis of a condition number for linear systems. *SIAM Journal on Matrix Analysis and Applications*, 15(2):636–660, 1994. [doi]. 22, 23
74. Mend-Amar Majig and Masao Fukushima. Restricted-step Josephy-Newton method for general variational inequalities with polyhedral constraints. *Pacific Journal of Optimization*, 6(2):375–390, 2010. 4
75. Olvi Leon Mangasarian. Equivalence of the complementarity to a system of nonlinear equations. *SIAM Journal on Applied Mathematics*, 31(1):89–92, 1976. [doi]. 3
76. Olvi Leon Mangasarian and Stanley Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967. [doi]. 19
77. Estelle Marchand, Torsten Müller, and Peter Knabner. Fully coupled generalised hybrid-mixed finite element approximation of two-phase two-component flow in porous media. Part II: numerical scheme and numerical results. *Computational Geosciences*, 16(3):691–708, 2012. [doi]. 2
78. Estelle Marchand, Torsten Müller, and Peter Knabner. Fully coupled generalised hybrid-mixed finite element approximation of two-phase two-component flow in porous media. Part I: formulation and properties of the mathematical model. *Computational Geosciences*, 17(2):431–442, 2013. [doi]. 2
79. Katta G. Murty. *Linear Complementarity, Linear and Nonlinear Programming* (Internet edition, prepared by Feng-Tien Yu, 1997). Heldermann Verlag, Berlin, 1988. 2
80. James M. Ortega and Werner C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970. Reprinted in 2000 by SIAM, Classics in Applied Mathematics 30, [doi]. 3
81. Jong-Shi Pang. Newton's method for B-differentiable equations. *Mathematics of Operations Research*, 15:311–341, 1990. [doi]. 4
82. Jong-Shi Pang. A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Mathematical Programming*, 51(1-3):101–131, 1991. [doi]. 3, 5, 14, 24
83. Jong-Shi Pang. Complementarity problems. In R. Horst and P.M. Pardalos, editors, *Handbook of Global Optimization*, volume 2 of *Nonconvex Optimization and Its Applications*, pages 271–338. Kluwer, Dordrecht, 1995. [doi]. 2
84. Jong-Shi Pang and Steven A. Gabriel. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60:295–337, 1993. [doi]. 6, 8
85. Sandra Pieraccini, Maria Grazia Gasparo, and Aldo Pasquali. Global Newton-type methods and semismooth reformulations for NCP. *Applied Numerical Mathematics*, 44:367–384, 2003. [doi]. 3, 6

86. Liqun Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18:227–244, 1993. [doi]. 5, 8
87. Liqun Qi and Jie Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58:353–367, 1993. [doi]. 5
88. Stephen M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980. [doi]. 3
89. Stephen M. Robinson. Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity. *Mathematical Programming Study*, 30:45–66, 1987. [doi]. 4
90. Alexander Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66:477–487, 1990. [doi]. 4
91. Defeng Sun and Liqun Qi. On NCP-functions. *Computational Optimization and Applications*, 13(1-3):201–220, 1999. [doi]. 3
92. Kenji Ueda and Nobuo Yamashita. Global complexity bound analysis of the Levenberg-Marquardt method for nonsmooth equations and its application to the nonlinear complementarity problem. *Journal of Optimization Theory and Applications*, 152:450–467, 2012. [doi]. 31
93. Shuhuang Xiang and Xiaojun Chen. Computation of generalized differentials in nonlinear complementarity problems. *Computational Optimization and Applications*, 50:403–423, 2011. [doi]. 5, 8, 10
94. Shu Zi Zhou and Zhan Yong Zou. A new iterative method for discrete HJB equations. *Numerische Mathematik*, 111(1):159–167, 2008. [doi]. 2