CrossMark

A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points

A. Fischer[1] · M. Herrich[1] · A. F. Izmailov[2,3,4] iD **·
W. Scheck[1] · M. V. Solodov[5]**

Abstract The LP-Newton method for constrained equations, introduced some years ago, has powerful properties of local superlinear convergence, covering both possibly nonisolated solutions and possibly nonsmooth equation mappings. A related globally convergent algorithm, based on the LP-Newton subproblems and linesearch for the equation's infinity norm residual, has recently been developed. In the case of smooth

✉ A. F. Izmailov
izmaf@ccas.ru

A. Fischer
andreas.fischer@tu-dresden.de

M. Herrich
markus.herrich@tu-dresden.de

W. Scheck
wladimir.scheck@tu-dresden.de

M. V. Solodov
solodov@impa.br

[1] Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany

[2] OR Department, VMK Faculty, Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, Leninskiye Gory, Moscow, Russia 119991

[3] RUDN University, Miklukho-Maklaya Str. 6, Moscow, Russia 117198

[4] Department of Mathematic, Physics and Computer Sciences, Derzhavin Tambov State University, TSU, Internationalnaya 33, Tambov, Russia 392000

[5] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil

⌖ Springer

equations, global convergence of this algorithm to B-stationary points of the residual over the constraint set has been shown, which is a natural result: nothing better should generally be expected in variational settings. However, for the piecewise smooth case only a property weaker than B-stationarity could be guaranteed. In this paper, we develop a procedure for piecewise smooth equations that avoids undesirable accumulation points, thus achieving the intended property of B-stationarity.

Keywords Constrained equation · Piecewise smooth equation · LP-Newton method · Global convergence · Quadratic convergence

Mathematics Subject Classification 90C33 · 91A10 · 49M05 · 49M15

1 Introduction

The LP-Newton method (to be described shortly) was proposed in [7] for solving constrained equations of the form

$$F(z) = 0, \quad z \in \Omega, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a given mapping, and $\Omega \subset \mathbb{R}^n$ is a given nonempty polyhedral set. The LP-Newton method has very strong local convergence properties: the assumptions for its local quadratic convergence to a solution of (1.1) neither imply differentiability of F nor require the solutions to be isolated. We refer to [6,7,10] for detailed discussions, and also for comparisons with convergence assumptions of other Newton-type methods.

In this work we assume that F is a piecewise continuously differentiable (PC^1-) mapping. This means that F is everywhere continuous, and there exist continuously differentiable mappings $F^1, \ldots, F^q : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$F(z) \in \{F^1(z), \ldots, F^q(z)\}$$

holds for all $z \in \mathbb{R}^n$. The mappings $F^1, \ldots, F^q$ are called selection mappings. This setting covers various important problem classes. For example, consider the complementarity system

$$a(z) = 0, \quad b(z) \geq 0, \quad c(z) \geq 0, \quad d(z) \geq 0, \quad c(z)^\top d(z) = 0, \tag{1.2}$$

where the mappings $a : \mathbb{R}^n \to \mathbb{R}^l$, $b : \mathbb{R}^n \to \mathbb{R}^s$, $c : \mathbb{R}^n \to \mathbb{R}^r$, and $d : \mathbb{R}^n \to \mathbb{R}^r$ are smooth. If we set $m := l + r$ and define

$$F(z) := (a(z), \min\{c(z), d(z)\}), \tag{1.3}$$
$$\Omega := \{z \in \mathbb{R}^n \mid b(z) \geq 0, \ c(z) \geq 0, \ d(z) \geq 0\}, \tag{1.4}$$

then the system (1.2) is equivalent to the constrained equation (1.1), and F is piecewise smooth. The importance of the specific choice (1.4) for Ω is discussed in [7,10]

(there are other possible reformulations of (1.2), but (1.3) together with (1.4) ensures the strong local convergence properties mentioned above). Note also that introducing slack variables for inequalities and redefining F accordingly, one can always convert a given problem in such a form that in (1.1) the set Ω is polyhedral. Thus, our standing assumption that Ω is polyhedral is not restrictive. Complementarity systems include Karush–Kuhn–Tucker (KKT) systems arising from optimization, variational inequalities, and generalized Nash equilibrium problems (GNEPs), see [9, Chapter 1] and [15, Chapter 1].

To explain the goals of this paper, we first briefly describe the LP-Newton method [7], and its recent globalization in [11]. For this purpose, let the index set of selection mappings active at $z \in \mathbb{R}^n$ be defined as

$$\mathcal{A}(z) := \left\{ p \in \{1, \ldots, q\} \mid F(z) = F^p(z) \right\}.$$

Furthermore, let $G : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ be any matrix-valued mapping such that

$$G(z) \in \{(F^p)'(z) \mid p \in \mathcal{A}(z)\}$$

holds for all $z \in \mathbb{R}^n$. The basic LP-Newton method generates a sequence $\{z^k\} \subset \Omega$ as follows. For a current iterate $z^k \in \Omega$, the next iterate is $z^{k+1} := z^k + \zeta^k$, where (ζ^k, γ_k) is a solution of the following subproblem in the variables $(\zeta, \gamma) \in \mathbb{R}^n \times \mathbb{R}$:

$$\begin{aligned}
& \text{minimize } \gamma \\
& \text{subject to } \|F(z^k) + G(z^k)\zeta\| \leq \gamma \|F(z^k)\|^2, \\
& \qquad\qquad \|\zeta\| \leq \gamma \|F(z^k)\|, \\
& \qquad\qquad z^k + \zeta \in \Omega.
\end{aligned} \tag{1.5}$$

Here and throughout the paper, $\| \cdot \|$ stands for the infinity norm. Then, (1.5) is a linear program (LP), justifying the name of the method. As any Newtonian method, the one described by (1.5) ensures local convergence only, under suitable assumptions (as already commented, the particular appeal of (1.5) is that those assumptions are the best/weakest currently known for any Newton-type method). To obtain global convergence, at iteration k the algorithm in [11] performs backtracking linesearch for the natural merit function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(z) := \|F(z)\|,$$

from the current iterate z^k in the direction ζ^k obtained by solving the LP-Newton subproblem (1.5); see Algorithm 2.1 below for precise details. Note that the function f is everywhere locally Lipschitz-continuous and directionally differentiable in every direction, see [9, Lemma 4.6.1]. The algorithm in [11] retains local quadratic convergence properties of the original LP-Newton method. Moreover, in the case of a smooth F, it also has desirable global convergence guarantees. Specifically, every accumulation point $\bar{z}$ of any generated sequence is B-stationary for the optimization problem

$$\text{minimize } f(z) \text{ subject to } z \in \Omega, \tag{1.6}$$

which means that

$$f'(\bar{z}; \zeta) \geq 0 \quad \text{for all} \quad \zeta \in T_\Omega(\bar{z}), \tag{1.7}$$

where $f'(z; \zeta)$ is the usual directional derivative of f at z in a direction ζ, and $T_\Omega(z)$ denotes the usual tangent cone to Ω at z (which for a polyhedral Ω is the set of feasible directions). We emphasize that property (1.7) is the natural first-order necessary optimality condition for problem (1.6). Note that, in general, nothing better than first-order stationarity for a merit function can be expected when globalizing some Newton method in the variational (not optimization) setting, like our problem (1.1); see the related results in [9] and [15]. Observe also that due to Clarke regularity of this f when F is smooth, in the latter case B-stationarity of f on Ω is equivalent to the (generally weaker) C-stationarity property

$$0 \in \partial f(\bar{z}) + N_\Omega(\bar{z}), \tag{1.8}$$

where $\partial f(z)$ is the Clarke generalized gradient of f at z (see [2]), and $N_\Omega(z)$ denotes the normal cone to Ω at z.

However, if F is nonsmooth (as in the PC^1 case under consideration), B-stationarity (1.7) can be strictly stronger than C-stationarity (1.8). The following example exhibits that this can be so even if we restrict ourselves specifically to the reformulation (1.1) of a complementarity system (1.2), employing (1.3) and (1.4).

Example 1.1 Consider $F : \mathbb{R} \to \mathbb{R}$ with

$$F(z) := \min\{1 + z, \ 1 - z\} \quad \text{and} \quad \Omega := [-1, \ 1],$$

which corresponds to the complementarity system (1.2) with $n := 1, l := 0, s := 0$, $r := 1, c(z) := 1 + z, d(z) := 1 - z$.

For this F there are $q = 2$ smooth selection mappings: $F^1(z) := 1 + z$ and $F^2(z) := 1 - z$, both active at $\bar{z} = 0$ which belongs to the interior of Ω, but which is not a solution of (1.1). Taking into account $f(z) = F(z)$ for all $z \in [-1, 1]$, i.e., $f(z) = 1 + z$ for $z \in [-1, 0]$ and $f(z) = 1 - z$ for $z \in [0, 1]$, we have $0 \in \partial f(\bar{z}) = [-1, 1]$ yielding C-stationarity for problem (1.6). However, B-stationarity is evidently violated.

We next discuss global convergence results in [11] for the case when F is a PC^1-mapping. The analysis therein employs the additional assumption

$$f(z) \leq f_p(z) \quad \text{for all} \ p \in \{1, \ldots, q\} \text{ and all } z \in \Omega, \tag{1.9}$$

where $f_p : \mathbb{R}^n \to \mathbb{R}_+$ is given by

$$f_p(z) := \|F^p(z)\|.$$

It can be verified directly that (1.9) holds automatically for the reformulation (1.1) of a complementarity system (1.2) given by (1.3), (1.4). So this assumption is not too restrictive, or at least holds for a good number of important applications. Under (1.9),

every accumulation point $\bar{z}$ of any sequence generated by the algorithm from [11] satisfies

$$0 \in \partial f_p(\bar{z}) + N_\Omega(\bar{z}) \tag{1.10}$$

for at least one $p \in \mathcal{A}(\bar{z})$ [11, Theorem 4.1]. In other words, $\bar{z}$ is C-stationary for the problem of minimizing the norm of F^p over Ω for at least one selection mapping F^p active at $\bar{z}$. This is clearly weaker than the B-stationarity result obtained in [11] for the case of smooth F.

In fact, property (1.10) for some $p \in \mathcal{A}(\bar{z})$ does not necessarily imply even the C-stationarity property (1.8) for the problem of minimizing the norm of F on Ω [(and therefore, due to our discussion above, certainly does not imply the B-stationarity property (1.7)], even if (1.9) is satisfied. This is demonstrated by the following modification of Example 1.1.

Example 1.2 Consider $F : \mathbb{R} \to \mathbb{R}^2$ with

$$F(z) := \begin{pmatrix} 1-z \\ \min\{1+z, \ 1-z\} \end{pmatrix} \quad \text{and} \quad \Omega := [-1, \ 1],$$

which corresponds to the complementarity system (1.2) with $n := 1, l := 1, s := 0,$ $r := 1, a(z) := 1-z, c(z) := 1+z, d(z) := 1-z$. For this F, there are $q = 2$ smooth selection mappings:

$$F^1(z) := \begin{pmatrix} 1-z \\ 1+z \end{pmatrix} \quad \text{and} \quad F^2(z) := \begin{pmatrix} 1-z \\ 1-z \end{pmatrix},$$

both active at $\bar{z} = 0$ which belongs to the interior of Ω. It can be easily seen that $f_1(z) = \max\{1-z, \ 1+z\}, f_2(z) = 1-z,$ and $f(z) = 1-z = f_2(z) \le f_1(z)$ hold for all $z \in [-1, \ 1]$. Therefore, (1.9) is satisfied. Moreover, we have $0 \in \partial f_1(\bar{z}) = [-1, \ 1]$, i.e., (1.10) is valid for $p = 1$. But $\bar{z}$ does not satisfy (1.10) for $p = 2$ and, moreover, is not C-stationary for problem (1.6) because $0 \notin \partial f(\bar{z}) = \partial f_2(\bar{z}) = \{-1\}$.

The discussion above gives rise to the question of what one can do if a sequence $\{z^k\}$ generated by the algorithm from [11] gets stuck at or near some $\bar{z}$ satisfying (1.10) for some $p \in \mathcal{A}(\bar{z})$, but violating this condition for some other active selection. Is it possible to escape such situations and guarantee B-stationarity of the full merit function f at accumulation points of the algorithm?

In this paper, we give an affirmative answer to this question. Specifically, we present a modification of the algorithm from [11] and show that, under suitable conditions, every accumulation point $\bar{z}$ of any generated sequence satisfies (1.10) for all $p \in \mathcal{A}(\bar{z})$. Moreover, we establish that the latter implies the desired B-stationarity property (1.7).

Getting back for a moment to complementarity system (1.2), it should be noted that there certainly exist smooth constrained reformulations for it as well, e.g., employing the Hadamard product. For such reformulations, the difficulty with the lack of B-stationarity of accumulation points discussed above, does not arise. However, the numerical results for KKT-type systems of GNEPs, presented in [11], demonstrate that the globalized LP-Newton method applied to the nonsmooth reformulation can

be more efficient. It is interesting to note that, as simple examples show, B-stationarity properties for the merit functions associated to nonsmooth and smooth reformulations of complementarity systems (and even considering optimization KKT systems only) are incomparable: none implies the other, in general.

We next comment on some previous work on globally convergent Newton-type methods for general nonsmooth equations (i.e., those without any special structure such as given by reformulations of complementarity systems, etc.). The two relatively recent ones, [1] and [13], deal with unconstrained equations $F(z) = 0$ for F being at least locally Lipschitz-continuous and mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. The semismooth Newton method in [13] is globalized in a hybrid way. By this we mean that in some situations, instead of the Newton direction, certain safeguarding directions which are of descent for the residual are employed. It is further shown how such directions can be constructed for some applications (not for a generic equation). The local super-linear convergence of the semismooth Newton method relies on semismoothness and BD-regularity at some solution and on locally accepting semismooth Newton steps by the test on linear decrease of the residual. In [1], some kind of a path-search globaliza-tion is proposed for Kummer's inexact local Newton scheme (see [16]). For general nonlinear equations, both approaches mentioned above might be considered as frame-works rather than readily implementable algorithms (implementable algorithms can be developed on their basis afterwards, working out the necessary details for more specific applications). The algorithms considered in our paper are less general, as they are for piecewise smooth equations only, but on the other hand, they are fully implementable for this problem setting, as stated. Also, unlike the LP-Newton method, the methods in [1] and [13] are not intended for cases of nonisolated solutions: the assumptions needed for their local superlinear convergence imply that a solution is locally unique. Observe further that the globalization of the LP-Newton method con-sidered here and in [11] is "pure" (not hybrid), as it always employs the ζ-part of the solution of the LP-Newton subproblem (1.5) as a direction for linesearch, at every iteration.

The rest of the paper is organized as follows. In Sect. 2, we recall the globalized LP-Newton algorithm from [11] and formally state its global convergence and rate of convergence properties. We also elaborate on Example 1.2, demonstrating that the deficiencies of the global convergence theory of this algorithm, specified above, can indeed be observed in practice. The proposed modification of the algorithm is presented and analyzed in Sect. 4. It relies on an error bound result developed in Sect. 3, measuring the distance to a C-stationary point for a given selection, when the point is not a solution of (1.1).

2 The globalized LP-Newton algorithm and its convergence

In this section we briefly review the algorithm from [11] and its convergence properties. By means of an example, we then show that, in the nonsmooth case, this algorithm may indeed get stuck at a point which is not even C-stationary for the merit function f. Finally, we explain our idea on how to escape from the vicinity of such "bad" points.

Consider the LP-Newton subproblem (1.5) with z^k replaced by $z \in \Omega$, the point z playing the role of a parameter:

$$\begin{aligned}
&\text{minimize } \gamma \\
&\text{subject to } \|F(z) + G(z)\zeta\| \leq \gamma \|F(z)\|^2, \\
&\qquad\qquad\ \|\zeta\| \leq \gamma \|F(z)\|, \\
&\qquad\qquad\ z + \zeta \in \Omega.
\end{aligned} \tag{2.1}$$

Let $\gamma(z)$ stand for the optimal value of this problem. It is obvious that (2.1) is always feasible, and if z is not a solution of the original problem (1.1), then (2.1) has a solution and it holds that $\gamma(z) > 0$. If $F(z) = 0$, then the objective function of problem (2.1) is unbounded below on its feasible set, and thus $\gamma(z) = -\infty$. Let $Z := \{z \in \Omega \mid F(z) = 0\}$ denote the solution set of (1.1). Define the function $\Delta : \Omega \setminus Z \to \mathbb{R}$ by

$$\Delta(z) := -f(z)(1 - \gamma(z)f(z)). \tag{2.2}$$

The function Δ gives a measure of directional descent for f from some point $z \in \Omega$ in the LP-Newton direction ζ. The following algorithm and its convergence results are those from [11].

Algorithm 2.1 Choose $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $z^0 \in \Omega$ and set $k := 0$.

1. If $F(z^k) = 0$, stop.
2. Compute (ζ^k, γ_k) as a solution of subproblem (1.5). If $\Delta(z^k) = 0$, stop.
3. Set $\alpha := 1$. If the inequality

$$f(z^k + \alpha\zeta^k) \leq f(z^k) + \sigma\alpha\Delta(z^k) \tag{2.3}$$

   is satisfied, set $\alpha_k := \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (2.3) again, etc., until (2.3) becomes valid.
4. Set $z^{k+1} := z^k + \alpha_k\zeta^k$, increase k by 1 and go to step 1.

We next summarize the global and local convergence results from [11, Theorems 4.1 and 4.2, Corollary 4.1], in the PC^1 case.

Theorem 2.1 *Assume that (1.9) is satisfied. Then, the following assertions are valid:*

(a) *Algorithm 2.1 is well-defined and, for any starting point $z^0 \in \Omega$, it either terminates with some iterate $z^k \in \Omega$ satisfying*

$$0 \in \partial f_p(z^k) + N_\Omega(z^k) \tag{2.4}$$

for at least one $p \in \mathcal{A}(z^k)$, or it generates an infinite sequence $\{z^k\}$ such that every accumulation point $\bar{z}$ of this sequence satisfies (1.10) for at least one $p \in \mathcal{A}(\bar{z})$.

(b) *Let $\{z^k\} \subset \Omega$ be any sequence such that $\{f(z^k)\}$ is nonincreasing, and such that for every accumulation point $\bar{z}$ of $\{z^k\}$ there is a subsequence $\{z^{k_j}\}$ converging to $\bar{z}$ with z^{k_j+1} generated by an iteration of Algorithm 2.1 for all j. Then, each of these accumulation points satisfies (1.10) and $\Delta(z^{k_j}) \to 0$ as $j \to \infty$ holds for each of the corresponding subsequences.*

We note that assertion (b) of the theorem above slightly extends the corresponding part of [11, Theorem 4.1]. This extension easily follows observing the proof of [11, Theorem 4.1].

Let $\mathrm{dist}(z, U) := \inf\{\|z - u\| \mid u \in U\}$ denote the distance from a point z to a set U.

Theorem 2.2 *Assume that the selection mappings F^p, $p \in \mathcal{A}(\bar{z})$, have Lipschitz-continuous derivatives near $\bar{z} \in Z$. Assume further that Assumptions 2–3 in [7] are satisfied at $\bar{z}$, and that (1.9) holds.*

If Algorithm 2.1 generates an iterate which is close enough to $\bar{z}$, then the algorithm either terminates at some $z^k \in Z$ or it generates an infinite sequence $\{z^k\}$ converging to some $\hat{z} \in Z$, and the rate of convergence is Q-quadratic.

In particular, this assertion is valid if (1.9) is satisfied and if there exists $\omega > 0$ such that

$$\mathrm{dist}(z, Z_p) \le \omega \|F^p(z)\|$$

holds for all $z \in \Omega$ close enough to $\bar{z}$, and for all $p \in \mathcal{A}(\bar{z})$, where $Z_p := \{z \in \Omega \mid F^p(z) = 0\}$.

As already discussed in Sect. 1, Theorem 2.1 does not rule out the possibility that a sequence generated by Algorithm 2.1 might have an accumulation point $\bar{z}$ satisfying (1.10) for some $p \in \mathcal{A}(\bar{z})$ but violating this condition for another active selection, or that the algorithm even terminates at such a point (in the sense of finite convergence, i.e., hitting this point exactly). We next show that this situation may occur indeed.

Example 2.1 Let us consider again F and Ω defined in Example 1.2, i.e.,

$$F(z) := \begin{pmatrix} 1 - z \\ \min\{1 + z, \ 1 - z\} \end{pmatrix} \quad \text{and} \quad \Omega := [-1, \ 1].$$

The corresponding two smooth selection mappings F^1 and F^2 are also as defined in Example 1.2.

Then, the index set of active selection mappings and the corresponding mapping $G : \mathbb{R} \to \mathbb{R}^2$ are given by

$$\mathcal{A}(z) = \begin{cases} \{1\} & \text{for } z < 0, \\ \{1, 2\} & \text{for } z = 0, \\ \{2\} & \text{for } z > 0, \end{cases} \qquad G(z) = \begin{cases} (F^1)'(z) = (-1, \ 1) & \text{for } z \le 0, \\ (F^2)'(z) = (-1, \ -1) & \text{for } z > 0, \end{cases}$$

where for $z = 0$ also $G(0) = (-1, \ -1)$ could have been chosen.

Now assume that Algorithm 2.1 is used for solving (1.1), and suppose that $z^k \in \Omega$ is some iterate with $F(z^k) \ne 0$. Then, the algorithm computes a solution (ζ^k, γ_k) of the LP-Newton subproblem (1.5) and the value $\Delta(z^k)$. If $\Delta(z^k)$ is negative, a stepsize α_k satisfying (2.3) is determined next, and it is easy to see that $\alpha_k = 1$. Indeed, taking into account that $f(z) = 1 - z$ for all $z \in \Omega$, (2.3) provides

$$1 - z^k - \alpha\zeta^k = f(z^k + \alpha\zeta^k) \le f(z^k) + \sigma\alpha\Delta(z^k) = 1 - z^k + \sigma\alpha\Delta(z^k),$$

Table 1 Behavior of Algorithm 2.1 on Example 1.2 for different current iterates $z^k \in [-1, \ 1)$

z^k	ζ^k	γ_k	$\Delta(z^k)$	Behavior of iteration
$-1 \leq z^k < \dfrac{1-\sqrt{5}}{2}$	$\dfrac{1-z^k}{2-z^k}$	$\dfrac{1}{2-z^k}$	< 0	$\dfrac{1-\sqrt{5}}{2} < z^{k+1} < 0 \quad z^{k+2} = 0$
$\dfrac{1-\sqrt{5}}{2} \leq z^k < 0$	$-z^k$	$\dfrac{1}{(1-z^k)^2}$	< 0	$z^{k+1} = 0$
$z^k = 0$	0	1	$= 0$	Algorithm 2.1 stops at z^k
$0 < z^k < 1$	$\dfrac{1-z^k}{2-z^k}$	$\dfrac{1}{2-z^k}$	< 0	$\underbrace{z^{k+1} = 1 - \dfrac{(1-z^k)^2}{2-z^k}}$
				quadratic convergence to $\hat{z} = 1$

which is equivalent to $-\alpha\zeta^k \leq \sigma\alpha\Delta(z^k)$. Since we know that, assuming $\Delta(z^k) < 0$, this inequality is satisfied with some $\alpha > 0$, it is also valid for $\alpha = 1$. Thus, the new iterate is $z^{k+1} := z^k + \zeta^k$. Now, elementary calculations show that subproblem (1.5) always has a unique solution. More precisely, the behavior of the iteration is shown in Table 1. In particular, we obtain that the sequence $\{z^k\}$ generated by Algorithm 2.1 converges quadratically to the only solution $\hat{z} = 1$ of (1.1) if the starting point z^0 belongs to $(0, \ 1)$. However, if $z^0 \in [-1, \ 0]$, then the algorithm terminates at $\bar{z} = 0$ after at most two steps. As already shown in Example 1.2, $\bar{z}$ in fact satisfies (1.10) for $p = 1 \in \mathcal{A}(\bar{z})$, but violates (1.10) for $p = 2 \in \mathcal{A}(\bar{z})$.

The observations in Example 2.1 lead to the following idea: if $\Delta(z^k) = 0$ holds at some iterate $z^k \in \Omega$ with $F(z^k) \neq 0$, then we can try to switch to another selection mapping active at z^k (if there is one), i.e., to solve the LP-Newton subproblem (1.5) again but with $G(z^k)$ replaced by the Jacobian of another active selection mapping. The motivation is that the new direction generated this way might allow to escape from the iterate z^k, which is not possible in Algorithm 2.1.

Furthermore, in practical computations one would not expect the algorithm to terminate with some iterate z^k satisfying $\Delta(z^k) = 0$ exactly. The much more plausible situation is generating a sequence with an accumulation point $\bar{z}$ satisfying $\Delta(\bar{z}) = 0$. Therefore, trying to switch to another selection mapping might already make sense if it happens that $\Delta(z^k)$ becomes close to zero while $\|F(z^k)\|$ is still relatively large. The new selection mapping should be such that we expect it to be active at a potential accumulation point $\bar{z}$. By modifying the algorithm along those lines, we would like to avoid (escape from) an accumulation point if it is not B-stationary for the merit function f.

The subsequent sections implement the idea described above. In Sect. 3, we prove the error bound result which can be used to estimate the set of smooth selection mappings active at a potential accumulation point of a sequence generated by Algorithm 2.1. In Sect. 4 this error bound is used to construct and analyze our modification of Algorithm 2.1.

3 Error bound results

We start with the following proposition, which extends the error bound result of [15, Proposition 1.64] from unconstrained to constrained equations. The notation is different from (1.1) because in the sequel Proposition 3.1 will actually be applied not to F but to Δ defined in (2.2).

Proposition 3.1 *Let $S \subset \mathbb{R}^n$ be locally closed and conical at $\bar{z} \in S$, i.e., there exists a neighborhood $\mathcal{O}$ of $\bar{z}$ such that*

$$S \cap \mathcal{O} = T_S(\bar{z}) \cap \mathcal{O} \tag{3.1}$$

(this is automatic if S is polyhedral). Let $\bar{z}$ be a solution of the equation

$$\Phi(z) = 0$$

for some $\Phi : S \to \mathbb{R}^r$ which is directionally differentiable at $\bar{z}$ in every direction $\zeta \in T_S(\bar{z})$.
   Then, the condition

$$\{\zeta \in T_S(\bar{z}) \mid \Phi'(\bar{z}; \zeta) = 0\} = \{0\} \tag{3.2}$$

is necessary for the constrained error bound

$$z - \bar{z} = O(\|\Phi(z)\|) \quad as \ z \in S \ tends \ to \ \bar{z}. \tag{3.3}$$

If Φ is Lipschitz-continuous near $\bar{z}$, then (3.2) is also sufficient for (3.3).

Proof In order to prove the necessity, consider any $\zeta \in T_S(\bar{z})$ such that $\Phi'(\bar{z}; \zeta) = 0$. From (3.1) and (3.3) it follows that

$$
\begin{aligned}
t\|\zeta\| &= \|\bar{z} + t\zeta - \bar{z}\| \\
&= O(\|\Phi(\bar{z} + t\zeta)\|) \\
&= O(\|\Phi(\bar{z} + t\zeta) - \Phi(\bar{z})\|) \\
&= O(t\|\Phi'(\bar{z}; \zeta)\|) + o(t) \\
&= o(t)
\end{aligned}
$$

as $t \to 0+$, which is possible only when $\zeta = 0$.
   To prove the sufficiency, suppose that there exists a sequence $\{z^k\} \subset S$ convergent to $\bar{z}$, with $z^k \neq \bar{z}$ for all k, and such that

$$\frac{\|\Phi(z^k)\|}{\|z^k - \bar{z}\|} \to 0 \ as \ k \to \infty. \tag{3.4}$$

For each k set $\zeta^k := (z^k - \bar{z})/\|z^k - \bar{z}\|$, $t_k := \|z^k - \bar{z}\|$. Passing to a subsequence, if necessary, we can assume that the sequence $\{\zeta^k\}$ converges to some $\zeta \in \mathbb{R}^n$ with

$\|\zeta\| = 1$. Further, by the definition of the tangent cone, $\zeta \in T_S(\bar{z})$. Therefore, taking into account (3.1), (3.4), and Lipschitz-continuity of Φ near $\bar{z}$, and observing that $\bar{z} + t_k \zeta^k = z^k \in S$, it holds that

$$\frac{\|\Phi(\bar{z} + t_k \zeta) - \Phi(\bar{z})\|}{t_k} \leq \frac{\|\Phi(\bar{z} + t_k \zeta^k)\|}{t_k} + \frac{\|\Phi(\bar{z} + t_k \zeta) - \Phi(\bar{z} + t_k \zeta^k)\|}{t_k}$$

$$= \frac{\|\Phi(z^k)\|}{\|z^k - \bar{z}\|} + O(\|\zeta^k - \zeta\|) \to 0 \text{ as } k \to \infty,$$

implying the equality $\Phi'(\bar{z}; \zeta) = 0$, which contradicts (3.2). □

All the subsequent considerations in this section are concerned with a fixed smooth selection mapping F^p; that is why we drop the index p, and write F instead of F^p. In particular, the values of $\Delta(\cdot)$ are computed using the Jacobian of this specific selection mapping F^p. It is assumed that F^p has a locally Lipschitz-continuous derivative.

The polyhedral nonempty set Ω that appears in (1.1) can be written as

$$\Omega = \{z \in \mathbb{R}^n \mid Cz \leq b\},$$

with some $C \in \mathbb{R}^{s \times n}$ and $b \in \mathbb{R}^s$. In this representation of Ω we assume, without loss of generality, that C has no zero rows. If it were not the case, as such constraints are redundant, they can simply be removed. Then, note that the assumption int $\Omega \neq \emptyset$ that will be employed in the sequel, is equivalent to saying that the constraints defining Ω satisfy the Slater condition (there exists z such that $Cz < b$). We note, in passing, that the Slater condition is automatic in our context, at least for some applications. For example it holds when Ω is formed by nonnegativity conditions on slack variables used to reformulate nonlinear inequalities as equalities.

With the above representation of Ω, the constraints of (2.1) can be written in the form

$$\mathcal{B}(z)u \leq \beta(z), \tag{3.5}$$

with $u \in \mathbb{R}^n \times \mathbb{R}$, $\mathcal{B}(z) \in \mathbb{R}^{M \times (n+1)}$, $\beta(z) \in \mathbb{R}^M$ given by

$$u := \begin{pmatrix} \zeta \\ \gamma \end{pmatrix}, \quad \mathcal{B}(z) := \begin{pmatrix} F'(z) & -f^2(z)e^m \\ -F'(z) & -f^2(z)e^m \\ I & -f(z)e^n \\ -I & -f(z)e^n \\ C & 0 \end{pmatrix}, \quad \beta(z) := \begin{pmatrix} -F(z) \\ F(z) \\ 0 \\ 0 \\ b - Cz \end{pmatrix}, \tag{3.6}$$

where $M := 2m + 2n + s$ and e^j stands for the column vector of ones in the space $\mathbb{R}^j$.

Remark 3.1 According to [11, Lemmas 3.1, 3.2], if $F(\bar{z}) \neq 0$ for some $\bar{z} \in \Omega$, then $\Delta(\cdot)$ is continuous at $\bar{z}$, and (1.8) holds if and only if $\Delta(\bar{z}) = 0$. By (2.2), $\Delta(\bar{z}) = 0$ holds if and only if $(\zeta, \gamma) = (0, 1/f(\bar{z}))$ is a solution of (2.1) for $z := \bar{z}$.

We now consider $\bar{z} \in \mathbb{R}^n$ such that

$$\bar{z} \in \Omega, \quad F(\bar{z}) \neq 0, \quad \text{and} \quad \Delta(\bar{z}) = 0. \tag{3.7}$$

Then, $\bar{z} \in \Omega$ is such that the estimate $\Delta(\bar{z})$ of the directional derivative employed in the algorithm is zero, but the point $\bar{z}$ is *not* a solution of our problem (1.1). Hence, by Remark 3.1, (1.8) holds and the solution set of the LP problem (2.1) with $z := \bar{z}$ has the form

$$U(\bar{z}) = Z(\bar{z}) \times \{1/f(\bar{z})\}, \tag{3.8}$$

where the set $Z(\bar{z}) \subset \mathbb{R}^n$ consists of $\zeta \in \mathbb{R}^n$ satisfying

$$\|F(\bar{z}) + F'(\bar{z})\zeta\| \leq f(\bar{z}), \quad \|\zeta\| \leq 1, \quad C\zeta \leq b - C\bar{z}. \tag{3.9}$$

With our notation above, problem (2.1) with $z := \bar{z}$ can be written as

$$\text{minimize } \gamma \quad \text{subject to} \quad \bar{\mathcal{B}}u \leq \bar{\beta}, \tag{3.10}$$

where $\bar{\mathcal{B}} := \mathcal{B}(\bar{z})$ and $\bar{\beta} := \beta(\bar{z})$. To consider the LP that is dual to (3.10), let $w^{1+} \in \mathbb{R}^m$, $w^{1-} \in \mathbb{R}^m$, $w^{2+} \in \mathbb{R}^n$, $w^{2-} \in \mathbb{R}^n$, and $w^3 \in \mathbb{R}^s$ denote dual variables corresponding to the five blocks of constraints in (3.5). Then, with $w := (w^{1+}, w^{1-}, w^{2+}, w^{2-}, w^3) \in \mathbb{R}^M$, the dual of (3.10) can be written as

$$
\begin{aligned}
\text{maximize } \quad & F(\bar{z})^\top(w^{1-} - w^{1+}) + (b - C\bar{z})^\top w^3 \\
\text{subject to} \quad & F'(\bar{z})^\top(w^{1+} - w^{1-}) + (w^{2+} - w^{2-}) + C^\top w^3 = 0, \\
& f^2(\bar{z})(e^m)^\top(w^{1+} + w^{1-}) + f(\bar{z})(e^n)^\top(w^{2+} + w^{2-}) = 1, \\
& w \geq 0.
\end{aligned}
\tag{3.11}
$$

To describe the solution set $W(\bar{z})$ of this program let us recall that (3.7) and Remark 3.1 imply that $(\bar{\zeta}, \bar{\gamma}) := (0, 1/f(\bar{z}))$ belongs to the solution set $U(\bar{z})$ of the primal program (3.10). Therefore, the second constraint in (3.9) and the corresponding constraints $-e^n \leq \zeta \leq e^n$ in the primal program are not active at $\bar{\zeta} = 0$. Thus, by complementary slackness, the dual variables w^{2+} and w^{2-} associated to these constraints are zero for any solution in $W(\bar{z})$. This implies $f^2(\bar{z})(e^m)^\top(w^{1+} + w^{1-}) = 1$. Applying complementary slackness again, we obtain that $W(\bar{z})$ consists of all vectors $w \in \mathbb{R}^M$ satisfying

$$
\begin{aligned}
& (F'(\bar{z}))^\top(w^{1+} - w^{1-}) + C^\top w^3 = 0, \\
& (e^m)^\top(w^{1+} + w^{1-}) = 1/f^2(\bar{z}), \\
& w^{1+}_{I_+(\bar{z})} \geq 0, \quad w^{1+}_{\{1, \ldots, m\}\setminus I_+(\bar{z})} = 0, \\
& w^{1-}_{I_-(\bar{z})} \geq 0, \quad w^{1-}_{\{1, \ldots, m\}\setminus I_-(\bar{z})} = 0, \\
& w^{2+} = w^{2-} = 0, \\
& w^3_{J(\bar{z})} \geq 0, \quad w^3_{\{1, \ldots, s\}\setminus J(\bar{z})} = 0,
\end{aligned}
\tag{3.12}
$$

where

$$
\begin{aligned}
I_+(\bar{z}) &:= \{i \in \{1, \ldots, m\} \mid F_i(\bar{z}) = f(\bar{z})\}, \\
I_-(\bar{z}) &:= \{i \in \{1, \ldots, m\} \mid -F_i(\bar{z}) = f(\bar{z})\}, \\
J(\bar{z}) &:= \{j \in \{1, \ldots, s\} \mid (C\bar{z})_j = b_j\}.
\end{aligned}
$$

Lemma 3.1 *Let $\bar{z}$ satisfy* (3.7). *Then, the solution set $U(\bar{z})$ of the primal LP* (3.10) *is bounded. If* int $\Omega \neq \emptyset$, *then the solution set $W(\bar{z})$ of the dual LP* (3.11) *is also bounded.*

Proof Because of (3.8) and the second relation in (3.9), we have that the primal LP solution set $U(\bar{z})$ is bounded.

By the standard LP duality theory, the solution set $W(\bar{z})$ of the dual LP (3.11) is the Lagrange multipliers set of the primal LP (3.10). Next, by standard optimization theory, the multipliers set associated to any primal problem (not necessarily an LP) is bounded if and only if the Mangasarian–Fromovitz constraint qualification holds. In the convex case and when all constraints are inequalities, like in the primal LP (3.10), this constraint qualification is the same as the Slater condition (existence of a feasible point for which all the inequalities are strict). As we consider (without loss of generality) that C has no zero rows, our assumption int $\Omega \neq \emptyset$ means that there exists a (Slater) point $\hat{z}$ with $C\hat{z} < b$. To show that the constraints of the primal LP (3.10) satisfy the Slater condition, first recall that $\bar{\mathcal{B}}u \leq \bar{\beta}$ is the same as $\mathcal{B}(\bar{z})u \leq \beta(\bar{z})$. Taking into account (3.6) and $f(\bar{z}) > 0$, we easily get a Slater point for the constraints of (3.10) by just fixing ζ to $\hat{z} - \bar{z}$ and by taking γ sufficiently large. Thus, the Lagrange multipliers set of (3.10) is bounded. Since this set and the solution set of $W(\bar{z})$ of the dual LP (3.11) coincide, $W(\bar{z})$ is bounded as well. □

We next analyze the behavior of the optimal value function $(\mathcal{B}, \beta) \mapsto v(\mathcal{B}, \beta)$ of the parametric linear program

$$
\text{minimize } \gamma \quad \text{subject to} \quad \mathcal{B}u \leq \beta \tag{3.13}
$$

with $(\mathcal{B}, \beta) \in \mathbb{R}^{M \times (n+1)} \times \mathbb{R}^M$ in a neighborhood of $(\bar{\mathcal{B}}, \bar{\beta})$.

Lemma 3.2 *Let $\bar{z}$ satisfy* (3.7) *and suppose that* int $\Omega \neq \emptyset$. *Then, the optimal value function v of problem* (3.13) *is Lipschitz-continuous near $(\bar{\mathcal{B}}, \bar{\beta})$ and directionally differentiable at $(\bar{\mathcal{B}}, \bar{\beta})$ in any direction $(\mathcal{D}, \delta) \in \mathbb{R}^{M \times (n+1)} \times \mathbb{R}^M$. For this directional derivative, it holds that*

$$
v'(\bar{\mathcal{B}}, \bar{\beta}; (\mathcal{D}, \delta)) = \min_{u \in U(\bar{z})} \max_{w \in W(\bar{z})} w^\top (\mathcal{D}u - \delta) = \max_{w \in W(\bar{z})} \min_{u \in U(\bar{z})} w^\top (\mathcal{D}u - \delta).
$$

Proof We first show the local Lipschitz-continuity of the optimal value function v of the parametric problem (3.13) by applying Theorem 5.1 from [12]. To this end, let us first consider a modification of problem (3.13), namely

$$
\text{minimize } \gamma \quad \text{subject to} \quad \mathcal{B}u \leq \beta, \quad \gamma \leq 2/f(\bar{z}). \tag{3.14}
$$

Let $\mathcal{F}(\mathcal{B}, \beta)$ and $\mathcal{S}(\mathcal{B}, \beta)$ denote the feasible set of this problem and its solution set, respectively. Obviously, any $u = (\zeta, \gamma) \in \mathcal{F}(\bar{\mathcal{B}}, \bar{\beta})$ satisfies

$$-\gamma e^n \le \zeta \le \gamma e^n \quad \text{and} \quad \gamma \le 2/f(\bar{z}).$$

Therefore, if $\|\mathcal{B} - \bar{\mathcal{B}}\|$ and $\|\beta - \bar{\beta}\|$ are sufficiently small, we have that

$$\|u\| = \|(\zeta, \gamma)\| \le 4/f(\bar{z})$$

holds for all $u \in \mathcal{F}(\mathcal{B}, \beta)$. Thus, the point-to-set mapping $(\mathcal{B}, \beta) \to \mathcal{F}(\mathcal{B}, \beta)$ is uniformly compact near $(\bar{\mathcal{B}}, \bar{\beta})$ in the sense of [12, Definition 3.3].

It was shown in the proof of Lemma 3.1 that $\bar{\mathcal{B}}u \le \bar{\beta}$ satisfies the Slater condition. Thus, for $\|\mathcal{B} - \bar{\mathcal{B}}\|$ and $\|\beta - \bar{\beta}\|$ sufficiently small, $\mathcal{F}(\mathcal{B}, \beta) \ne \emptyset$ holds.

Finally, to apply [12, Theorem 5.1] the Mangasarian–Fromovitz constraint qualification (MFCQ) is needed at any $u \in \mathcal{S}(\bar{\mathcal{B}}, \bar{\beta})$. To verify this, we first note that the dual (3.11) of (3.10) has a bounded solution set $W(\bar{z})$ according to Lemma 3.1. By (3.8), the additional constraint $\gamma \le 2/f(\bar{z})$ in (3.14) cannot be active at a solution in $\mathcal{S}(\bar{\mathcal{B}}, \bar{\beta})$, so that the corresponding dual variable is 0 in the solution set of the dual problem associated to (3.14) for $(\mathcal{B}, \beta) := (\bar{\mathcal{B}}, \bar{\beta})$. Hence, this dual problem has a bounded solution set, which means that MFCQ holds at any u in the solution set $\mathcal{S}(\bar{\mathcal{B}}, \bar{\beta})$ of the primal problem.

Now, Theorem 5.1 in [12] yields the local Lipschitz-continuity of the optimal value function of problem (3.14) near $(\bar{\mathcal{B}}, \bar{\beta})$. As a consequence, the constraint $\gamma \le 2/f(\bar{z})$ cannot become active for any $(\mathcal{B}, \beta)$ in a sufficiently small neighborhood of $(\bar{\mathcal{B}}, \bar{\beta})$. Therefore, the optimal value function v associated to problem (3.13) is the same as the optimal value function of problem (3.14) and, hence, Lipschitz-continuous near $(\bar{\mathcal{B}}, \bar{\beta})$.

Recall again that the solution sets of the primal problem (3.10) and of its dual (3.11) are bounded according to Lemma 3.1. Thus, Theorems I and II in [17] give the remaining assertions regarding directional differentiability. □

Based on the previous lemma we will now prove a result on the directional derivative of the optimal value function of the LP-Newton subproblem (1.5).

Lemma 3.3 *Let $\bar{z}$ satisfy (3.7) and assume that int $\Omega \ne \emptyset$. Moreover, suppose that F' is locally Lipschitz-continuous and that F is twice differentiable at $\bar{z}$. Then, the optimal value function $z \mapsto \gamma(z)$ associated to the LP-Newton subproblem (2.1) is Lipschitz-continuous near $\bar{z}$ and directionally differentiable at $\bar{z}$ in all directions $d \in T_\Omega(\bar{z})$, and*

$$\gamma'(\bar{z}; d) = \max_{w \in W(\bar{z})} \min_{\zeta \in Z(\bar{z})} (w^{1+} - w^{1-})^\top F''(\bar{z})[d, \zeta] - \frac{2}{f^2(\bar{z})} f'(\bar{z}; d).$$

Proof Let us note first that the mapping $z \mapsto \mathcal{B}(z)$ is Lipschitz-continuous near $\bar{z}$ and directionally differentiable at $\bar{z}$ in every direction $d \in \mathbb{R}^n$, while the mapping $z \mapsto \beta(z)$ is continuously differentiable near $\bar{z}$, and

$$
\mathcal{B}'(\bar{z};\, d) = \begin{pmatrix} F''(\bar{z})[d] & -2f(\bar{z})f'(\bar{z};\, d)e^m \\ -F''(\bar{z})[d] & -2f(\bar{z})f'(\bar{z};\, d)e^m \\ 0 & -f'(\bar{z};\, d)e^n \\ 0 & -f'(\bar{z};\, d)e^n \\ 0 & 0 \end{pmatrix}, \quad \beta'(\bar{z}) = \begin{pmatrix} -F'(\bar{z}) \\ F'(\bar{z}) \\ 0 \\ 0 \\ -C \end{pmatrix}.
$$

Combining this with the Lipschitz-continuity of the optimal value function v of (3.13) near $(\bar{\mathcal{B}},\ \bar{\beta})$ and the directional differentiability of v (both according to Lemma 3.2), and observing that by the polyhedrality of Ω for any $d \in T_\Omega(\bar{z})$ it holds that $\bar{z}+td \in \Omega$ for $t > 0$ sufficiently small, we derive that γ is Lipschitz-continuous near $\bar{z}$ and that

$$
\begin{aligned}
\gamma'(\bar{z};\, d) &= \lim_{t\to 0+} \frac{\gamma(\bar{z}+td)-\gamma(\bar{z})}{t} \\
&= \lim_{t\to 0+} \frac{v(\mathcal{B}(\bar{z}+td),\ \beta(\bar{z}+td)) - v(\mathcal{B}(\bar{z}),\ \beta(\bar{z}))}{t} \\
&= \lim_{t\to 0+} \frac{v(\mathcal{B}(\bar{z})+t\mathcal{B}'(\bar{z};\, d)+o(t),\ \beta(\bar{z})+t\beta'(\bar{z})d+o(t)) - v(\mathcal{B}(\bar{z}),\ \beta(\bar{z}))}{t} \\
&= \lim_{t\to 0+} \frac{v(\mathcal{B}(\bar{z})+t\mathcal{B}'(\bar{z};\, d),\ \beta(\bar{z})+t\beta'(\bar{z})d) - v(\mathcal{B}(\bar{z}),\ \beta(\bar{z}))}{t} \\
&= v'(\mathcal{B}(\bar{z}),\ \beta(\bar{z});\ (\mathcal{B}'(\bar{z};\, d),\ \beta'(\bar{z})d)).
\end{aligned}
$$

Thus, $\gamma'(\bar{z};\, d)$ exists for all $d \in T_\Omega(\bar{z})$. Using the formula for the directional derivative in Lemma 3.2 and $\gamma(\bar{z}) = 1/f(\bar{z})$, we further obtain that

$$
\begin{aligned}
\gamma'(\bar{z};\, d) &= \max_{w\in W(\bar{z})} \min_{u\in U(\bar{z})} w^\top \left(\mathcal{B}'(\bar{z};\, d)u - \beta'(\bar{z})d \right) \\
&= \max_{w\in W(\bar{z})} \min_{\zeta\in Z(\bar{z})} \Big\{ (w^{1+})^\top \left(F''(\bar{z})[d,\ \zeta] - 2f'(\bar{z};\, d)e^m + F'(\bar{z})d \right) \\
&\qquad\qquad\qquad +(w^{1-})^\top \left(-(F''(\bar{z})[d,\ \zeta] - 2f'(\bar{z};\, d)e^m - F'(\bar{z})d \right) \\
&\qquad\qquad\qquad -\frac{f'(\bar{z};\, d)}{f(\bar{z})}(w^{2+})^\top e^n - \frac{f'(\bar{z};\, d)}{f(\bar{z})}(w^{2-})^\top e^n + (w^3)^\top Cd \Big\} \\
&= \max_{w\in W(\bar{z})} \min_{\zeta\in Z(\bar{z})} \Big\{ (w^{1+}-w^{1-})^\top F''(\bar{z})[d,\ \zeta] \\
&\qquad\qquad\qquad +d^\top ((F'(\bar{z}))^\top (w^{1+}-w^{1-}) + C^\top w^3) \\
&\qquad\qquad\qquad -2f'(\bar{z};\, d)(w^{1+}+w^{1-})^\top e^m - \frac{f'(\bar{z};\, d)}{f(\bar{z})}(w^{2+}+w^{2-})^\top e^n \Big\}.
\end{aligned}
$$

Finally, exploiting that (3.12) is valid for any $w \in W(\bar{z})$, we get the desired expression for $\gamma'(\bar{z};\, d)$. □

We are now ready to establish our main error bound result, which will be the key to identifying all active selections at undesirable accumulation points, thus opening the possibility to switch to another path.

Theorem 3.1 *Let $\bar{z}$ satisfy (3.7) and assume that* int $\Omega \neq \emptyset$. *Moreover, suppose that F' is locally Lipschitz-continuous and that F is twice differentiable at $\bar{z}$.*

Assume further that

$$\text{for all } d \in Z_0(\bar{z}) \setminus \{0\} \text{ and all } w \in W(\bar{z}) \tag{3.15}$$
$$\text{there exists } \zeta \in Z_0(\bar{z}) \text{ such that } (w^{1+} - w^{1-})^\top F''(\bar{z})[d, \zeta] < 0,$$

where

$$Z_0(\bar{z}) := \left\{ \zeta \in T_\Omega(\bar{z}) \,\middle|\, F'_{I_+(\bar{z})}(\bar{z})\zeta \leq 0, \ F'_{I_-(\bar{z})}(\bar{z})\zeta \geq 0 \right\}.$$

Then

$$z - \bar{z} = O(|\Delta(z)|)$$

as $z \in \Omega$ tends to $\bar{z}$.

Proof To apply Proposition 3.1, let therein S stand for the polyhedral set Ω, and Φ be given by Δ. Then, $\Delta(\bar{z}) = 0$ follows by (3.7). Thus, it remains to show that (3.2) holds in our setting. To this end, we first note that $\Delta(z) = f^2(z)(\gamma(z) - 1/f(z))$ holds for all $z \in \Omega$ close enough to $\bar{z}$. Therefore, by Lemma 3.3, Δ is Lipschitz-continuous near $\bar{z}$ and directionally differentiable at $\bar{z}$ in every direction $d \in T_\Omega(\bar{z})$. Taking into account the equality $\gamma(\bar{z}) = 1/f(\bar{z})$ and Lemma 3.3, we obtain that

$$\Delta'(\bar{z}; d) = f^2(\bar{z}) \left(\gamma'(\bar{z}; d) + \frac{f'(\bar{z}; d)}{f^2(\bar{z})} \right)$$
$$= f^2(\bar{z}) \max_{w \in W(\bar{z})} \min_{\zeta \in Z(\bar{z})} (w^{1+} - w^{1-})^\top F''(\bar{z})[d, \zeta] - f'(\bar{z}; d), \tag{3.16}$$

where the first term in parentheses on the right-hand side cannot be positive since $0 \in Z(\bar{z})$ [see (3.9)]. Therefore, $\Delta'(\bar{z}; d) = 0$ implies $f'(\bar{z}; d) \leq 0$. By (3.7), Remark 3.1 shows that (1.8) is valid. Since $F = F^p$ is smooth, (1.8) implies that $\bar{z}$ is B-stationary, so that $f'(\bar{z}; d)$ cannot be negative. Hence, only the case of $f'(\bar{z}; d) = 0$ needs to be considered. In this case, we have that $d \in Z_0(\bar{z})$. Furthermore, for every $\zeta \in Z_0(\bar{z})$, from (3.9) we see that $t\zeta \in Z(\bar{z})$ for all $t > 0$ small enough. Therefore, from (3.15) and (3.16) we obtain that $\Delta'(\bar{z}; d) < 0$ for every $d \in T_\Omega(\bar{z}) \setminus \{0\}$. Thus, the equality $\Delta'(\bar{z}; d) = 0$ for some $d \in T_\Omega(\bar{z})$ implies $d = 0$. Hence, Proposition 3.1 yields the desired result. $\square$

Going back to Example 1.2, it holds there that $f'_1(\bar{z}; d) > 0$ for all $d \in \mathbb{R} \setminus \{0\}$, implying that $Z_0(\bar{z}) = \{0\}$, and hence, (3.15) trivially holds (for $p = 1$).

As another example, consider the case when $\bar{z} \in \text{int } \Omega$ and the equality $|F_i(\bar{z})| = f(\bar{z})$ is attained for a single index $i \in \{1, \ldots, m\}$. To be specific, let $I_+(\bar{z}) = \{i\}$, $I_-(\bar{z}) = \emptyset$. Then (3.12) implies that

$$w_i^{1+} = \frac{1}{(f(\bar{z}))^2}, \quad F'_i(\bar{z}) = 0,$$

while all other components of w are equal to zero. Therefore, $Z_0(\bar{z}) = \mathbb{R}^n$, and condition (3.15) takes the form

for all $d \in \mathbb{R}^n \setminus \{0\}$ there exists $\zeta \in \mathbb{R}^n$ such that $\langle F_i''(\bar{z})d, \zeta \rangle < 0$.

This always holds if $F_i''(\bar{z})$ is a nonsingular matrix (for example, if it is positive-definite, i.e., $\bar{z}$ satisfies the second-order sufficient optimality condition for minimizers of f, which coincides with F_i near $\bar{z}$).

4 Globally convergent LP-Newton method with an escape procedure from nonstationary points

We shall use the error bound established in Theorem 3.1 for identifying smooth selection mappings active at a potential "bad" accumulation point of a sequence generated by Algorithm 2.1. This idea is related to the technique for identifying active constraints proposed in [8], later used also in [3,14], for example. Once such selection mappings different from the one currently used are identified, Algorithm 4.1 below implements the possibility of switching to another selection.

Proposition 4.1 *Let the assumptions of Theorem 3.1 hold with F replaced by F^π for some $\pi \in \mathcal{A}(\bar{z})$. Let $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ be any function satisfying $t = o(\rho(t))$ as $t \to 0+$. For every $z \in \Omega$, define the index set*

$$\mathcal{A}_\rho(z) := \left\{ p \in \{1, \ldots, q\} \mid \|F^p(z) - F(z)\| \leq \rho(|\Delta(z)|) \right\},$$

where $\Delta(z)$ is computed using $G(z) = (F^\pi)'(z)$. Then, for all $z \in \Omega$ close enough to $\bar{z}$, it holds that

$$\mathcal{A}_\rho(z) = \mathcal{A}(\bar{z}).$$

Proof Suppose first that $p \in \mathcal{A}(\bar{z})$. Then, by Theorem 3.1, for all $z \in \Omega$ close enough to $\bar{z}$ it holds that

$$\begin{aligned}
\|F^p(z) - F(z)\| &= \|(F^p(z) - F(z)) - (F^p(\bar{z}) - F(\bar{z}))\| \\
&= O(\|z - \bar{z}\|) \\
&= O(|\Delta(z)|) \\
&\leq \rho(|\Delta(z)|).
\end{aligned}$$

Hence, $p \in \mathcal{A}_\rho(z)$ follows.

Suppose now that $p \notin \mathcal{A}(\bar{z})$, which means that $\|F^p(\bar{z}) - F(\bar{z})\| > 0$. According to Remark 3.1, $\Delta(z) \to 0$ as $z \to \bar{z}$, and hence, $p \notin \mathcal{A}_\rho(z)$ for all $z \in \Omega$ close enough to $\bar{z}$. $\qquad \square$

Now we are in the position to describe our modification of Algorithm 2.1. In order to estimate whether some iterate z^k might be close to a point $\bar{z}$ which is not a solution of problem (1.1) but which satisfies (1.10) for some $p \in \mathcal{A}(\bar{z})$, we choose some $\delta_0 > 0$ and verify, after computing $\Delta(z^k)$, whether the ratio $\Delta(z^k)/f(z^k)$ is greater than or equal to $-\delta_0$. If it so, the algorithm tries to switch to another selection mapping.

Algorithm 4.1 Choose $\delta_0 > 0$, $\delta_1 > 0$, $\sigma \in (0, 1)$, and $\theta \in (0, 1)$. Fix a function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $t = o(\rho(t))$ as $t \to 0+$. Choose $z^0 \in \Omega$ and set $k := 0$.

1. If $F(z^k) = 0$, stop.
2. Let $\pi \in \mathcal{A}(z^k)$ be such that $G(z^k) = (F^\pi)'(z^k)$.

   Compute $(\zeta^{k,\pi}, \gamma_{k,\pi})$ as a solution of the LP-Newton subproblem (1.5).
   If $\Delta(z^k)/f(z^k) \geq -\delta_0$ and $\mathcal{A}_\rho(z^k) \setminus \{\pi\} \neq \emptyset$, go to step 3.
   Otherwise, if $\Delta(z^k) < 0$, set $(\zeta^k, \gamma_k) := (\zeta^{k,\pi}, \gamma_{k,\pi})$ and go to step 5.
   Otherwise, stop.
3. (a) Choose any $p \in \mathcal{A}_\rho(z^k) \setminus \{\pi\}$.
       If $F^p(z^k) = 0$, set $(\zeta^{k,p}, \gamma_{k,p}) := (0, 0)$ and $\Delta_p(z^k) := 0$.
       Otherwise, compute $(\zeta^{k,p}, \gamma_{k,p})$ as a solution of the LP-Newton subproblem

$$
\begin{aligned}
&\text{minimize } \gamma \\
&\text{subject to } \|F^p(z^k) + (F^p)'(z^k)\zeta\| \leq \gamma\|F^p(z^k)\|^2, \\
&\qquad\qquad \|\zeta\| \leq \gamma\|F^p(z^k)\|, \\
&\qquad\qquad z^k + \zeta \in \Omega,
\end{aligned}
\tag{4.1}
$$

   and set $\Delta_p(z^k) := -f_p(z^k)(1 - \gamma_{k,p} f_p(z^k))$.
   If $\Delta_p(z^k) > -\delta_1$, repeat the procedure for the next $p \in \mathcal{A}_\rho(z^k) \setminus \{\pi\}$, etc.,
   until some p will be found for which $\Delta_p(z^k) \leq -\delta_1$ holds, or until the set
   $\mathcal{A}_\rho(z^k) \setminus \{\pi\}$ is exhausted. In the latter case, take $p \in \mathcal{A}_\rho(z^k) \setminus \{\pi\}$ with the
   smallest associated $\Delta_p(z^k)$.
   (b) If $\Delta_p(z^k) < \Delta(z^k)$, go to step 4.
       Otherwise, if $\Delta(z^k) < 0$, set $(\zeta^k, \gamma_k) := (\zeta^{k,\pi}, \gamma_{k,\pi})$ and go to step 5.
       Otherwise, stop.
4. (a) Set $\alpha := 1$.
       If the inequality

$$
f_p(z^k + \alpha\zeta^{k,p}) \leq f_p(z^k) + \sigma\alpha\Delta_p(z^k)
\tag{4.2}
$$

   is satisfied, set $\alpha_{k,p} := \alpha$.
   Otherwise, replace α by $\theta\alpha$, check inequality (4.2) again, etc., until (4.2)
   becomes valid.
   (b) If

$$
f(z^k + \alpha_{k,p}\zeta^{k,p}) < f(z^k),
\tag{4.3}
$$

   set $(\zeta^k, \gamma_k) := (\zeta^{k,p}, \gamma_{k,p})$, $\alpha_k := \alpha_{k,p}$, and go to step 6.
   Otherwise, set $(\zeta^k, \gamma_k) := (\zeta^{k,\pi}, \gamma_{k,\pi})$ and go to step 5.
5. Set $\alpha := 1$.
   If inequality (2.3) is satisfied, set $\alpha_k := \alpha$.
   Otherwise, replace α by $\theta\alpha$, check inequality (2.3) again, etc., until (2.3) becomes
   valid.
6. Set $z^{k+1} := z^k + \alpha_k\zeta^k$, increase k by 1 and go to step 1.

   Some comments concerning various possibilities in Algorithm 4.1 are in order.

Remark 4.1 (i) If Algorithm 4.1 terminates with some iterate z^k after a finite number of iterations, then z^k is either a solution of (1.1), or $\Delta_p(z^k) = 0$ holds for all $p \in \mathcal{A}(z^k)$. The latter is equivalent to saying that z^k satisfies (2.4) for all $p \in \mathcal{A}(z^k)$.

(ii) In step 3(a) of Algorithm 4.1, we explicitly check if the current iterate z^k is a zero of the selection mapping F^p which is next to be considered. If it is a zero, then z^k is a minimizer of f_p, and so using this particular selection mapping would not lead to any progress. This is the reason why $\Delta_p(z^k)$ is set equal to zero in that case. Note, however, that $F^p(z^k) = 0$ but $F(z^k) \neq 0$ is not possible if the condition (1.9) is satisfied. In particular, this situation cannot occur if (1.1) arises from a complementarity system (1.2) employing (1.3), (1.4).

(iii) When the number q of selection mappings is large, the set $\mathcal{A}_\rho(z^k)$ can also be quite large at some iterates z^k. Then some rule for choosing $p \in \mathcal{A}_\rho(z^k)$ in step 3 should be adopted. An obvious idea would be to first consider a selection mapping for which $\|F^p(z^k) - F(z^k)\|$ is the smallest. If there is more than one selection mapping for which $\|F^p(z^k) - F(z^k)\|$ is (nearly) the smallest, it makes sense to choose the one which differs from the currently used selection mapping by a component where the infinity norm of $F(z^k)$ is attained. This is motivated by the observations in Example 4.2 and Proposition 4.2 below.

(iv) The parameters δ_0 and δ_1 are assumed to be positive and fixed. The value of δ_0 should be within the interval $(0, 1)$ in order to guarantee that high convergence rate is retained; see Theorem 4.3 below. Apart from that, one could also think of some dynamic rules for δ_0 and δ_1. The global convergence properties stated in Theorem 4.1 below and the quadratic convergence rate established in Theorem 4.3, remain valid as long as these quantities are not allowed to go below some fixed positive thresholds.

Global convergence properties of Algorithm 4.1 are characterized by the following theorem.

Theorem 4.1 *Assume that* (1.9) *holds, and that* int $\Omega \neq \emptyset$. *Then, Algorithm 4.1 is well-defined and, for any starting point* $z^0 \in \Omega$, *it either terminates with some iterate* $z^k \in \Omega$ *satisfying* (2.4) *for all* $p \in \mathcal{A}(z^k)$, *or generates an infinite sequence* $\{z^k\}$. *In the latter case, if* $\bar{z}$ *is an accumulation point of this sequence, and the selection mappings* F^p, $p \in \mathcal{A}(\bar{z})$, *have Lipschitz-continuous derivatives near* $\bar{z}$ *and are twice differentiable at* $\bar{z}$, *then either* $\bar{z}$ *violates* (3.15) *with* F *replaced by* F^p *for some* $p \in \mathcal{A}(\bar{z})$, *or satisfies*

$$0 \in \partial f_p(\bar{z}) + N_\Omega(\bar{z}) \quad \text{for all } p \in \mathcal{A}(\bar{z}). \tag{4.4}$$

Proof Similarly to the proof of Theorem 2.1 above, provided in [11], we come to the following conclusions. As explained in Remark 4.1(i), Algorithm 4.1 terminates for some iteration index k if either z^k is a solution of (1.1), or $\Delta_p(z^k) = 0$ for all $p \in \mathcal{A}(z^k) \subset \mathcal{A}_\rho(z^k)$. Further, by Remark 3.1, in both cases (2.4) must hold for all $p \in \mathcal{A}(z^k)$.

If the algorithm does not terminate, it generates an infinite sequence $\{z^k\}$. Let $\bar{z}$ be an accumulation point of $\{z^k\}$, and suppose that there exists $p \in \mathcal{A}(\bar{z})$ such that (1.10) does not hold. The latter implies that $f(\bar{z}) \neq 0$.

If $\{z^{k_j}\}$ denotes a subsequence of $\{z^k\}$ convergent to $\bar{z}$ such that for all j

$$\frac{\Delta(z^{k_j})}{f(z^{k_j})} < -\delta_0, \tag{4.5}$$

then the iterate z^{k_j+1} is generated by Algorithm 2.1 (without any modifications), and hence, by Theorem 2.1, $\Delta(z^{k_j}) \to 0$ as $j \to \infty$, giving a contradiction with (4.5). Therefore, for any subsequence $\{z^{k_j}\}$ convergent to $\bar{z}$, the modification of the algorithm is initiated at iteration k_j (i.e., $\Delta(z^{k_j})/f(z^{k_j}) \geq -\delta_0$) for all j large enough.

Suppose now that the stated smoothness requirements are satisfied for $\bar{z}$, and that (3.15) holds with F replaced by F^p, for all $p \in \mathcal{A}(\bar{z})$. Since, by continuity, $\mathcal{A}(z^{k_j}) \subset \mathcal{A}(\bar{z})$ for all j large enough, from Proposition 4.1 we have that $\mathcal{A}_\rho(z^{k_j}) = \mathcal{A}(\bar{z})$ for all j large enough. From [11, Lemma 3.2] and from the existence of $p \in \mathcal{A}(\bar{z})$ violating (1.10), it then follows by construction of the algorithm that, perhaps after passing to a further subsequence, there exist $\hat{\delta} \in (0, \delta_1]$ and $\hat{p} \in \mathcal{A}(\bar{z})$ such that, for all j, Algorithm 4.1 computes $(\zeta^{k_j}, \gamma_{k_j})$ and the corresponding α_{k_j} using $G(z^{k_j}) = (F^{\hat{p}})'(z^{k_j})$, with $\Delta_{\hat{p}}(z^{k_j}) \leq -\hat{\delta}$. If $z^{k_j+1} = z^{k_j} + \alpha_{k_j}\zeta^{k_j}$ is accepted by the algorithm for infinitely many indices j, then passing to a further subsequence, and employing (1.9), and (2.3) or (4.2), we obtain that

$$f(z^{k_j+1}) \leq f_{\hat{p}}(z^{k_j}) - \sigma\alpha_{k_j}\hat{\delta} \tag{4.6}$$

holds for all j. Since by the construction of the algorithm the sequence $\{f(z^k)\}$ is non-increasing and bounded below, it converges. Then, for any accumulation point $\bar{z}$ of $\{z^k\}$, the sequence $\{f(z^k)\}$ (and hence, its subsequence $\{f(z^{k_j+1})\}$) converges to $f(\bar{z})$, by continuity of f. But since $\{z^{k_j}\}$ is convergent to $\bar{z}$, it follows by continuity of $f_{\hat{p}}$ that $\{f_{\hat{p}}(z^{k_j})\}$ converges to $f_{\hat{p}}(\bar{z}) = f(\bar{z})$ (the last equality holds because $\hat{p} \in \mathcal{A}(\bar{z})$). Therefore, (4.6) implies that $\alpha_{k_j} \to 0$ as $j \to \infty$, and the rest of the proof repeats the corresponding part of the proof of [11, Theorem 3.1], but for f replaced by $f_{\hat{p}}$, eventually giving a contradiction.

It remains to consider the case when the modified step is not accepted by test (4.3) for $k = k_j$ and all j large enough. Employing again (1.9) and (4.2), we then obtain

$$f_{\hat{p}}(z^{k_j}) - \sigma\alpha_{k_j}\hat{\delta} \geq f_{\hat{p}}(z^{k_j} + \alpha_{k_j}\zeta^{k_j}) \geq f(z^{k_j} + \alpha_{k_j}\zeta^{k_j}) \geq f(z^{k_j}).$$

By convergence of $\{z^{k_j}\}$ to $\bar{z}$, and by continuity of f and $f_{\hat{p}}$, this again implies that $\alpha_{k_j} \to 0$ as $j \to \infty$, leading to a contradiction as above. □

Our next result completes the job, proving that (4.4) implies B-stationarity of $\bar{z}$ for the optimization problem (1.6), i.e., that of minimizing the residual. Thus, unlike Algorithm 2.1, its modification Algorithm 4.1 either terminates at some point being B-stationary for (1.6) after a finite number of steps, or generates an infinite sequence with its accumulation points being B-stationary for (1.6).

Theorem 4.2 *If (4.4) holds at some $\bar{z} \in \Omega$, then $\bar{z}$ is a B-stationary point of problem* (1.6), *i.e.,* (1.7) *is satisfied.*

Conversely, if (1.9) *holds, but* (1.10) *is violated for some* $p \in \mathcal{A}(\bar{z})$, *then* $\bar{z}$ *is not a B-stationary point of problem* (1.6).

Proof To prove the first assertion, assume that there exists some $\zeta \in T_\Omega(\bar{z})$ such that $f'(\bar{z}; \zeta) < 0$. Evidently, there exist $p \in \mathcal{A}(\bar{z})$ and a sequence of reals $\{t_k\} \to 0+$, such that $f(\bar{z} + t_k\zeta) = f_p(\bar{z} + t_k\zeta)$ for all k. Then

$$0 > f'(\bar{z}; \zeta) = \lim_{k\to\infty} \frac{f(\bar{z} + t_k\zeta) - f(\bar{z})}{t_k} = \lim_{k\to\infty} \frac{f_p(\bar{z} + t_k\zeta) - f_p(\bar{z})}{t_k} = f_p'(\bar{z}; \zeta).$$

By [11, Proposition 3.2] applied to f_p, this contradicts (4.4).

The second assertion readily follows from [11, Lemma 3.2] applied to f_p, and from (1.9). □

Note that B-stationarity cannot be replaced by C-stationarity in the second assertion of Theorem 4.2; this is demonstrated by Example 1.1.

Finally, we prove that the convergence rate properties of Algorithm 2.1 stated in Theorem 2.2 remain valid for Algorithm 4.1 if $\delta_0 \in (0, 1)$.

Theorem 4.3 *The assertion of Theorem* 2.2 *remains valid with Algorithm* 2.1 *replaced by Algorithm* 4.1, *assuming that in the latter* $\delta_0 \in (0, 1)$.

Proof Assumption 3 in [7] consists of saying that $\gamma(\cdot)$ is bounded from above on Ω near $\bar{z} \in Z$. Therefore, by the definition of the function Δ in (2.2) it holds that

$$\Delta(z^k) = -\|F(z^k)\| + O(\|F(z^k)\|^2)$$

as $z^k \to \bar{z}$. Since $\delta_0 < 1$, this implies that

$$\frac{\Delta(z^k)}{\|F(z^k)\|} < -\delta_0,$$

if $F(z^k) \neq 0$ and z^k is close enough to $\bar{z}$. Therefore, z^{k+1} is computed by an iteration of Algorithm 2.1 (without any modifications).

The assertion is then that of Theorem 2.2. □

We complete this section by illustrating Algorithm 4.1 on some examples. We shall also give some further insight into step 3(a) of the algorithm.

Example 1.2 demonstrates the advantages of Algorithm 4.1 as compared to Algorithm 2.1, in the nonsmooth setting.

Example 4.1 Consider again problem (1.1) with F and Ω defined in Example 1.2, i.e.,

$$F(z) := \begin{pmatrix} 1 - z \\ \min\{1 + z, \ 1 - z\} \end{pmatrix} \quad \text{and} \quad \Omega := [-1, 1].$$

The smooth selection mappings F^1 and F^2 are as defined in Example 1.2, while the index set $\mathcal{A}(z)$ of selection mappings active at z and $G(z)$ are given in Example 2.1.

Let Algorithm 4.1 be applied to (1.1), where we take $\rho(t) := \sqrt{t}$. Obviously, this function satisfies $t = o(\rho(t))$ as $t \to 0+$. Before analyzing in detail how Algorithm 4.1 works and how its behavior differs from that of Algorithm 2.1, we start with determining the index set $\mathcal{A}_\rho(z)$ for $z \in \Omega$. By some elementary calculations, it can be shown that

$$\mathcal{A}_\rho(z) = \begin{cases} \{1\} = \mathcal{A}(z) \text{ for } z \in [-1, -1/4), \\ \{2\} = \mathcal{A}(z) \text{ for } z \in (\tilde{z}, 1], \\ \{1, 2\} \qquad \text{ for } z \in [-1/4, \tilde{z}], \end{cases}$$

where $\tilde{z}$ denotes the smallest positive zero of the polynomial $4z^3 - 8z^2 - z + 1$, i.e., $\tilde{z} \approx 0.3183$.

Now suppose that $z^k \in \Omega$ is some iterate with $F(z^k) \neq 0$. Let $\pi \in \mathcal{A}(z^k)$ be the index such that $G(z^k) = (F^\pi)'(z^k)$. Then, Algorithm 4.1 computes a solution $(\zeta^{k,\pi}, \gamma_{k,\pi})$ of the LP subproblem (1.5), the value $\Delta(z^k)$, and the ratio $\Delta(z^k)/f(z^k)$. As observed above, $\mathcal{A}_\rho(z^k)$ is a singleton if z^k is less than $-1/4$ or greater than $\tilde{z}$. Therefore, for such z^k, other selection mappings are not considered, and the next iterate generated by Algorithm 4.1 coincides with the iterate obtained by Algorithm 2.1. More precisely, using our analysis in Example 2.1, we have

$$\begin{aligned} z^{k+1} &\in ((1 - \sqrt{5})/2, 0) &&\text{for } z^k \in [-1, (1 - \sqrt{5})/2), \\ z^{k+1} &= 0 &&\text{for } z^k \in [(1 - \sqrt{5})/2, -1/4), \\ z^{k+1} &= 1 - (1 - z^k)^2/(2 - z^k) &&\text{for } z^k \in (\tilde{z}, 1). \end{aligned}$$

For $z^k \in [-1/4, \tilde{z}]$, Algorithms 2.1 and 4.1 might generate different next iterates. To analyze this, suppose first that $z^k \in [-1/4, 0]$. Then we have $\pi = 1$, i.e., the selection mapping F^1 and its Jacobian are used to compute $(\zeta^{k,\pi}, \gamma_{k,\pi})$. From the analysis in Example 2.1 it follows that $\Delta(z^k)/f(z^k) = z^k/(1 - z^k)$. Therefore, depending on the value of δ_0 and the precise value of z^k, the inequality $\Delta(z^k)/f(z^k) \geq -\delta_0$ might be valid (for $z^k = 0$ it is valid for any $\delta_0 > 0$). Suppose that this inequality holds. Then, since $2 \in \mathcal{A}_\rho(z^k)$, a solution $(\zeta^{k,2}, \gamma_{k,2})$ of the LP problem (4.1) with $p = 2$ is computed. It is not difficult to see that

$$(\zeta^{k,2}, \gamma_{k,2}) := \left(\frac{1 - z^k}{2 - z^k}, \frac{1}{2 - z^k} \right)$$

is the unique solution of that problem, so that

$$\Delta_2(z^k) = -\frac{1 - z^k}{2 - z^k} < \Delta(z^k) = z^k \leq 0.$$

Therefore, regardless whether the condition $\Delta_2(z^k) > -\delta_1$ is satisfied or not, a step size $\alpha_{k,2}$ is computed according to the rule in step 4 of Algorithm 4.1 for $p = 2$. Taking into account that $f_2(z) = 1 - z$ for all $z \in \Omega$, and that $\Delta_2(z^k) < 0$, it can be easily shown that $\alpha_{k,2} = 1$ holds; see the corresponding discussion in Example 2.1. Moreover, the inequality (4.3) is satisfied for $p = 2$. Thus, $\zeta^k := \zeta^{k,2}$ is taken as the

new direction, and the new iterate is $z^{k+1} := z^k + \zeta^k$. By elementary calculations it can be seen that $z^{k+1} > 0$ holds.

Suppose now that $z^k \in (0, \tilde{z}]$. Then we have $\pi = 2$, i.e., the selection mapping F^2 and its Jacobian are used to compute $(\zeta^{k,\pi}, \gamma_{k,\pi})$. From the analysis in Example 2.1 it follows that $\Delta(z^k)/f(z^k) = -1/(2 - z^k)$ holds. Depending on the value of δ_0, the inequality $\Delta(z^k)/f(z^k) \geq -\delta_0$ might be valid, and if this is the case, since $1 \in \mathcal{A}_\rho(z^k)$, a solution $(\zeta^{k,1}, \gamma_{k,1})$ of the LP subproblem (4.1) for $p = 1$ is computed. It is not difficult to see that

$$(\zeta^{k,1}, \gamma_{k,1}) := \left(-z^k, \frac{1}{(1 + z^k)^2}\right)$$

is the unique solution of that problem, so that $\Delta_1(z^k) = -z^k$ holds. However, it can be shown that

$$-z^k = \Delta_1(z^k) > \Delta(z^k) = -\frac{1 - z^k}{2 - z^k} \quad \text{for any } z^k \in (0, \tilde{z}].$$

Therefore, regardless whether the condition $\Delta_1(z^k) > -\delta_1$ is satisfied or not, the line-search is performed in the direction $\zeta^k := \zeta^{k,\pi}$ according to step 5 of Algorithm 4.1. Therefore, the new iterate coincides with the iterate generated by Algorithm 2.1, i.e., $z^{k+1} = 1 - (1 - z^k)^2/(2 - z^k)$.

Let us summarize the main conclusions for this example. Similarly to Algorithm 2.1, the modified Algorithm 4.1 generates a sequence which converges quadratically to the unique solution $z = 1$ of (1.1) for any starting point $z^0 \in (0, 1)$. But, unlike Algorithm 2.1, the modified algorithm also converges quadratically to $z = 1$ if the starting point is chosen within $[-1, 0]$. In particular, if Algorithm 4.1 exactly hits $z = 0$ at some iteration, it escapes this point.

It might happen that Algorithm 4.1 hits some point $\bar{z}$ where (1.10) is satisfied for at least one $p \in \mathcal{A}(\bar{z})$, and more than two selection mappings are active at $\bar{z}$. Then, there is the question which active selection mapping should be considered first in step 3 of the algorithm. A related question is whether it is a priori clear that using some $p \in \mathcal{A}(\bar{z})$ cannot lead to any progress because (1.10) is satisfied for those indices as well. We discuss these issues next, starting with the following example.

Example 4.2 Consider the problem (1.1) with

$$F(z) := \begin{pmatrix} 1 - z \\ \min\{1 + z, 1 + z + z^2, 1 - z\} \\ \min\{z, z(z - 1)\} \end{pmatrix} \quad \text{and} \quad \Omega := [-1, 1].$$

There are $q = 6$ smooth selection mappings:

$$
\begin{array}{ll}
F^1(z) := (1 - z, 1 + z, z), & F^2(z) := (1 - z, 1 + z, z(z - 1)), \\
F^3(z) := (1 - z, 1 + z + z^2, z), & F^4(z) := (1 - z, 1 + z + z^2, z(z - 1)), \\
F^5(z) := (1 - z, 1 - z, z), & F^6(z) := (1 - z, 1 - z, z(z - 1)),
\end{array}
$$

all active at $\bar{z} := 0$. It is not difficult to see that $\mathcal{A}(z) = \{1\}$ for $z \in [-1, 0)$, and $\mathcal{A}(z) = \{6\}$ for $z \in (0, 1]$. Consequently, we have that $G(z) = (F^1)'(z)$ for $z \in [-1, 0)$, and $G(z) = (F^6)'(z)$ for $z \in (0, 1]$. At the origin we set $G(0) := (F^1)'(0) = (-1, 1, 1)$.

As in Example 1.2, $\bar{z}$ satisfies (1.10) for $p = 1$ so that, due to our definition of $G(0)$, Algorithm 2.1 terminates if it hits $\bar{z}$. It can be observed that this actually happens after (at most) two iterations of Algorithm 2.1, if the starting point is chosen within $[-1, 0)$; see the related discussion for Example 1.2 in Sect. 2.

Consider now the case when $\bar{z}$ is the exact iteration produced by Algorithm 4.1. Then, since $\Delta(\bar{z}) = \Delta_1(\bar{z}) = 0$, another selection mapping active at $\bar{z}$ must be tried. So the LP subproblem (4.1) for some other p is solved, and the related value $\Delta_p(\bar{z})$ is computed. It turns out that $\Delta_2(\bar{z}) = \Delta_3(\bar{z}) = \Delta_4(\bar{z}) = 0$ holds, whereas $\Delta_5(\bar{z}) = \Delta_6(\bar{z}) = -1/2$. This means that only switching to the selection mappings F^5 or F^6 may lead to progress. Elementary calculations show that in both cases $(\zeta, \gamma) = (1/2, 1/2)$ is the unique solution of the corresponding LP subproblem. Moreover, $z = 1/2$ is accepted by the test in step 4(b) of the algorithm, and it can be seen that from this point on, two more iterations are needed to exactly hit the only solution $\bar{z} = 1$ of problem (1.1).

It is not surprising that $\Delta_3(\bar{z})$ is equal to zero because we already know that $\Delta_1(\bar{z}) = 0$ holds and, moreover, not only the values but also the Jacobians of F^1 and F^3 coincide at $\bar{z} = 0$. Obviously, switching to another active selection mapping with the same Jacobian cannot give any progress, as the LP subproblem would be the same.

At the same time, it is not obvious that using F^2 or F^4 instead of F^1 cannot lead to progress. Note, however, that F^2 differs from F^1 by the third component only. This is a component whose absolute value at $\bar{z} = 0$ is less than the infinity norm of $F^1(\bar{z})$ (and of $F^2(\bar{z})$). The next proposition demonstrates that switching to an active selection mapping with this property cannot lead to any progress.

Proposition 4.2 Let $\bar{z} \in \Omega$ be such that $F(\bar{z}) \neq 0$, and assume that $p_1, p_2 \in \mathcal{A}(\bar{z})$ are such that $(F_i^{p_1})'(\bar{z}) = (F_i^{p_2})'(\bar{z})$ for all $i \in \{1, \ldots, m\}$ satisfying $|F_i(\bar{z})| = \|F(\bar{z})\|$. Then (1.10) holds for $p = p_1$ if and only if it holds for $p = p_2$.

Proof According to [11, Remark 3.1], for every $p \in \mathcal{A}(\bar{z})$, the Clarke generalized gradient of f_p at $\bar{z}$ is defined only by the gradients $(F_i^p)'(\bar{z})$ for $i \in \{1, \ldots, m\}$ such that $|F_i^p(\bar{z})| = \|F(\bar{z})\|$. Therefore, under the assumptions of this proposition, $\partial f_{p_1}(\bar{z}) = \partial f_{p_2}(\bar{z})$ holds, implying the needed assertion.   □

This proposition suggests that when trying to switch to another active selection, one should only consider selections with a different gradient for at least one component on which the infinity norm is attained.

To conclude, we comment on our numerical experience with Algorithms 2.1 and 4.1 on the library of GNEPs from [5] and [4], and on why we do not report it here. Algorithm 2.1 had already been tested in [11], with detailed results reported there. During the preparation of the current paper, we have also tested Algorithm 4.1, using the same testing environment with the same settings. The results obtained turned out to be quite similar to those for Algorithm 2.1, and that is why we do not report the details. The explanation for the observed numerical behavior is that in all the cases of convergence of Algorithm 2.1 to points which were not solutions of the

constrained equation, these points were nevertheless local minimizers of the merit function over the feasible set, and hence, B-stationary. Therefore, there was simply no room for Algorithm 4.1 to improve on Algorithm 2.1 *on this* test collection. It might be interesting to figure out whether this behavior is related to some special features of GNEPs and/or their reformulations. At the same time, we emphasize that the use of the escape procedure did not harm in any way, which is the best result one might expect from Algorithm 4.1 in the situations when improvements are not possible. We stress that the developed escape procedure provides a theoretical guarantee of better global convergence properties, while not harming computational performance. That it might be necessary and might improve convergence indeed is demonstrated by the examples presented above, while the experiments with GNEPs show that it does not degrade the performance in other cases.

References

1. Bütikofer, S., Klatte, D.: A nonsmooth Newton method with path search and its use in solving $C^{1,1}$ programs and semi-infinite problems. SIAM J. Optim. **20**, 2381–2412 (2010)
2. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
3. Daryina, A.N., Izmailov, A.F., Solodov, M.V.: A class of active-set Newton methods for mixed complementarity problems. SIAM J. Optim. **15**, 409–429 (2004/2005)
4. Dreves, A., Facchinei, F., Fischer, A., Herrich, M.: A new error bound result for generalized Nash equilibrium problems and its algorithmic application. Comput. Optim. Appl. **59**, 63–84 (2014)
5. Dreves, A., Facchinei, F., Kanzow, C., Sagratella, S.: On the solution of the KKT conditions of generalized Nash equilibrium problems. SIAM J. Optim. **21**, 1082–1108 (2011)
6. Facchinei, F., Fischer, A., Herrich, M.: A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. Math. Method Oper. Res. **77**, 433–443 (2013)
7. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. **146**, 1–36 (2014)
8. Facchinei, F., Fischer, A., Kanzow, C.: On the accurate identification of active constraints. SIAM J. Optim. **9**, 14–32 (1999)
9. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
10. Fischer, A., Herrich, M., Izmailov, A.F., Solodov, M.V.: Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. Comput. Optim. Appl. **63**, 425–459 (2016)
11. Fischer, A., Herrich, M., Izmailov, A.F., Solodov, M.V.: A globally convergent LP-Newton method. SIAM J. Optim. **26**, 2012–2033 (2016)
12. Gauvin, J., Dubeau, F.: Differential properties of the marginal function in mathematical programming. Math. Program. Stud. **19**, 101–119 (1982)
13. Ito, K., Kunisch, K.: On a semi-smooth Newton method and its globalization. Math. Program. **118**, 347–370 (2009)
14. Izmailov, A.F., Solodov, M.V.: An active set Newton method for mathematical programs with complementarity constraints. SIAM J. Optim. **19**, 1003–1027 (2008)
15. Izmailov, A.F., Solodov, M.V.: Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Switzerland (2014)
16. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. Kluwer, Dordrecht (2002)
17. Williams, A.C.: Marginal values in linear programming. J. Soc. Ind. Appl. Math. **11**, 82–94 (1963)