# A GLOBALLY CONVERGENT LP–NEWTON METHOD [*]

ANDREAS FISCHER [†], MARKUS HERRICH [†], ALEXEY F. IZMAILOV[‡], AND
MIKHAIL V. SOLODOV[§]

**Abstract.** We develop a globally convergent algorithm based on the LP-Newton method which has been recently proposed for solving constrained equations, possibly nonsmooth and possibly with nonisolated solutions. The new algorithm makes use of linesearch for the natural merit function, and preserves the strong local convergence properties of the original LP-Newton scheme. We also present computational experiments on a set of generalized Nash equilibrium problems, and a comparison with the previous hybrid globalization employing the potential reduction method.

**Key words.** constrained equations; LP-Newton method; global convergence; quadratic convergence; generalized Nash equilibrium problems.

**AMS subject classifications.** 90C33, 91A10, 49M05, 49M15

**1. Introduction.** In this paper, we consider the constrained equation

$$(1.1) \qquad F(z) = 0, \quad z \in \Omega,$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping assumed to be at least piecewise differentiable, and $\Omega \subset \mathbb{R}^n$ is a given nonempty polyhedral set. Precise smoothness requirements on $F$ will be stated below, as needed.

We note that via appropriate reformulations, the setting of smooth or piecewise smooth constrained equations (1.1) includes a good number of important problems. For example, Karush–Kuhn–Tucker (KKT) systems arising from optimization, from variational inequalities, and from generalized Nash equilibrium problems (GNEPs); and, more generally, any systems of equalities and inequalities with complementarity relations. For some descriptions, we refer to [10, Chapter 1] and [18, Chapter 1]; see also Sections 4 and 5 below.

The LP-Newton method proposed in [8] generates a sequence $\{z^k\} \subset \Omega$ of approximate solutions of (1.1) as follows. Given a current iterate $z^k \in \Omega$, the next iterate is $z^{k+1} := z^k + \zeta^k$, where $(\zeta^k, \gamma_k)$ is a solution of the following subproblem in the variables $(\zeta, \gamma) \in \mathbb{R}^n \times \mathbb{R}$:

$$(1.2) \qquad \begin{aligned} \text{minimize} \quad & \gamma \\ \text{subject to} \quad & \|F(z^k) + G(z^k)\zeta\| \leq \gamma\|F(z^k)\|^2, \\ & \|\zeta\| \leq \gamma\|F(z^k)\|, \\ & z^k + \zeta \in \Omega. \end{aligned}$$

[†]Department of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany (andreas.fischer@tu-dresden.de, markus.herrich@tu-dresden.de).

[‡]Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia; RUDN University, Miklukho-Maklaya Str. 6, 117198 Moscow, Russia; and Department of Mathematic, Physics and Computer Sciences, Derzhavin Tambov State University, TSU, Internationalnaya 33, 392000 Tambov, Russia (izmaf@ccas.ru).

[§]IMPA – Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil (solodov@impa.br).

In (1.2), $G(z^k)$ is the Jacobian of $F$ at $z^k$ if $F$ is differentiable, or a suitable substitute (generalized derivative) otherwise. Formal requirements on the choice of the matrix-valued mapping $G : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ will be specified later. Throughout the paper, $\| \cdot \|$ denotes the infinity-norm. In particular, (1.2) is then a linear programming (LP) problem, which is precisely the reason for the "LP" part of the method's name. The "Newton" part is also clear, as the key ingredient to the construction is the linearization of the mapping $F$ in the first constraint of (1.2). We note, in passing, that in the original proposal in [8], the subproblem of the LP-Newton method has an extra constraint $\gamma \geq 0$. Clearly, omitting this constraint has no influence as long as $F(z^k) \neq 0$, and gives (1.2) above.

We emphasize that the LP-Newton method has very strong local convergence properties. In particular, the assumptions for local quadratic convergence to a solution of (1.1) neither imply differentiability of $F$ nor require the solutions to be isolated; see [8, 13]. This combination of difficulties is very challenging, and other methods require stronger assumptions; see the discussion in [8]. For example, even for the special case of KKT systems (or, in fact, for the even more special case of optimization), other Newton-type methods developed for degenerate problems of this specific structure, need assumptions that require the primal part of the solution to be isolated, e.g., [11, 12, 16, 17, 22]; see also [18, Chapter 7]. Not to mention classical methods such as sequential quadratic programming, for which both primal and dual solutions must be (locally) unique [18, Chapter 4]. By contrast, for the LP-Newton method applied to a KKT system, neither primal nor dual solutions have to be isolated.

However, building a reasonable globalization of the local scheme given by (1.2) has not been completely resolved as of yet. In fact, some globalizations of the LP-Newton method can be found in the literature for a very special problem class only, namely for constrained systems of equations arising from the reformulation of KKT-type conditions for GNEPs. In [5], a hybrid method for the solution of a smooth reformulation of the GNEP KKT-type system is described and analyzed. The global phase in [5] is the potential reduction algorithm proposed in [6] (see also [21]), and the local part is the LP-Newton method. A similar hybrid method is given also in [4], but for a nonsmooth reformulation of the GNEP KKT-type system.

In this paper, we develop a new globalization of the LP-Newton method, which is of the "direct" (rather than hybrid) kind. At each iteration $k$, a backtracking linesearch in the direction $\zeta^k$ obtained by solving the LP-Newton subproblem (1.2) is performed, for the natural merit function $\|F(\cdot)\|$. We shall first analyze our algorithm for the case when $F$ is smooth. We shall show that the algorithm is well-defined and that every accumulation point of the generated sequence satisfies the first-order necessary optimality condition for the optimization problem

$$(1.3) \qquad \begin{aligned} &\text{minimize} && \|F(z)\| \\ &\text{subject to} && z \in \Omega. \end{aligned}$$

A result along those lines is what one usually hopes to achieve for a globalization of some Newtonian method for a variational (not optimization) problem using a merit function; see [10, 18] (i.e., a global solution of (1.3) with zero optimal value cannot be theoretically guaranteed in general, and this is completely natural in the variational context). Moreover, we prove that the generated sequence converges quadratically to a solution of (1.1) whenever some iterate is close enough to a solution satisfying the assumptions in [8] for local quadratic convergence of the original LP-Newton method. Thus, unlike in hybrid approaches, the transition from global to fast local convergence occurs naturally. In addition, we shall extend our results to the case when $F$ is

piecewise continuously differentiable. This is possible if a certain additional condition is satisfied. The condition in question holds, for example, for (1.1) arising from a suitable reformulation of a complementarity system.

The rest of the paper is organized as follows. In Section 2 we describe our proposed globalization of the LP-Newton method. We consider the case when $F$ is a smooth mapping in Section 3, and analyze both global and local convergence properties of the algorithm. In Section 4 we generalize the approach for the case when $F$ is piecewise continuously differentiable. Finally, in Section 5, we present numerical results on a set of GNEPs, showing that the new algorithm is very promising.

We complete this section with a few words concerning our notation. We denote the solution set of (1.1) by $Z$. The usual directional derivative of a directionally differentiable function $f$ at $z$ in a direction $\zeta$ is denoted by $f'(z; \zeta)$. The Clarke directional derivative of $f$ at $z$ in a direction $\zeta$ is

$$f^\circ(z; \zeta) := \limsup_{y \to z,\, t \to 0+} (f(y + t\zeta) - f(y))/t.$$

The notation $\partial f(z)$ stands for the Clarke generalized gradient of a locally Lipschitz continuous function $f$ at $z$. The distance from a point $z \in \mathbb{R}^n$ to a nonempty set $U \subset \mathbb{R}^n$ is $\mathrm{dist}(z, U) := \inf\{\|z - \bar{z}\| \mid \bar{z} \in U\}$. For a set $U \subset \mathbb{R}^n$, we denote the convex hull of this set by $\mathrm{conv}\, U$, and the conic hull by $\mathrm{cone}\, U$. Finally, $T_\Omega(z)$ is the usual tangent cone to $\Omega$ at $z$, and $N_\Omega(z)$ is the normal cone to $\Omega$ at $z$.

**2. Globalized LP-Newton method.** Before formally stating the algorithm, we need some definitions. The following refers to the problem

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

i.e., problem (1.2) above with $z^k$ replaced by $z \in \Omega$, the point $z$ considered as a parameter. Clearly, problem (2.1) is always feasible. Let $\gamma(z)$ stand for the optimal value of (2.1). It is obvious that 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 in (2.1) is unbounded below on its feasible set, and thus $\gamma(z) = -\infty$.

We further define the natural merit function $f : \mathbb{R}^n \to \mathbb{R}$ for the original problem (1.1) by

(2.2)
$$f(z) := \|F(z)\|,$$

and the function $\Delta : \Omega \setminus Z \to \mathbb{R}$ by

(2.3)
$$\Delta(z) := -f(z)(1 - \gamma(z)f(z)).$$

Because of the use of the infinity norm, the function $f$ need not be differentiable even if $F$ is differentiable. As we shall show, the quantity given by $\Delta$ is a measure of directional descent for the merit function $f$ from the point $z \in \Omega$ in the LP-Newton direction $\zeta$. This leads to the (surprisingly simple-looking) Algorithm 1.

In the next section we analyze convergence properties of Algorithm 1 for the case when $F$ is smooth. The case of piecewise continuously differentiable problems follows in Section 4.

---

**Algorithm 1** Globalized LP-Newton algorithm

---

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

1. If $F(z^k) = 0$, then stop.

2. Compute $(\zeta^k, \gamma_k)$ as a solution of the LP problem (1.2). If $\Delta(z^k) = 0$, stop.

3. Set $\alpha := 1$. If the inequality

$$(2.4) \qquad\qquad f(z^k + \alpha\zeta^k) \leq f(z^k) + \sigma\alpha\Delta(z^k)$$

is satisfied, set $\alpha_k := \alpha$. Otherwise, replace $\alpha$ by $\theta\alpha$, check the inequality (2.4) again, etc., until (2.4) becomes valid.

4. Set $z^{k+1} := z^k + \alpha_k\zeta^k$, increase $k$ by 1 and go to step 1.

---

**3. The case of smooth equations.** Throughout this section, the mapping $F$ is assumed to be continuously differentiable, and the matrix-valued mapping $G$ is given by $G := F'$. We first prove in Section 3.1 that Algorithm 1 is well-defined, and every accumulation point $\bar{z}$ of the generated sequence satisfies the first-order necessary optimality condition for problem (1.3), i.e.,

$$(3.1) \qquad\qquad 0 \in \partial f(\bar{z}) + N_\Omega(\bar{z}).$$

In Section 3.2 we discuss local convergence properties of the algorithm. In particular, we show that the method retains local quadratic convergence of the LP-Newton scheme.

Before starting the convergence analysis, we make some remarks on the structure of the directional derivative and of the Clarke subdifferential of the function $f$.

REMARK 3.1. The function $f$ is directionally differentiable at any $z \in \mathbb{R}^n$ in any direction $\zeta \in \mathbb{R}^n$, and

$$(3.2) \qquad f'(z; \zeta) = \max\left\{ \max_{i \in I_+(z)} \langle F_i'(z), \zeta \rangle, \max_{i \in I_-(z)} (-\langle F_i'(z), \zeta \rangle) \right\},$$

where

$$I_+(z) := \{i \in \{1, \ldots, m\} \mid F_i(z) = f(z)\},$$

$$I_-(z) := \{i \in \{1, \ldots, m\} \mid -F_i(z) = f(z)\}.$$

Let $\bar{z} \in \mathbb{R}^n$ be arbitrary but fixed. Since $F$ is continuously differentiable at $\bar{z}$, the function $f$ is Lipschitz-continuous in a neighborhood of $\bar{z}$, and hence, the Clarke subdifferential $\partial f(\bar{z})$ is well-defined; see [2]. Moreover, since for every $z \in \mathbb{R}^n$ close enough to $\bar{z}$ it holds that

$$f(z) = \max\left\{ \max_{i \in I_+(\bar{z})} F_i(z), \max_{i \in I_-(\bar{z})} (-F_i(z)) \right\},$$

by [2, Proposition 2.3.12] we have the equality

$$(3.3) \qquad \partial f(\bar{z}) = \text{conv}\left( \{F_i'(\bar{z}) \mid i \in I_+(\bar{z})\} \cup \{-F_i'(\bar{z}) \mid i \in I_-(\bar{z})\} \right).$$

**3.1. Global convergence.** We start with continuity properties of the subproblem's optimal value function $\gamma(\cdot)$.

LEMMA 3.1. *If $\bar{z} \in \Omega$ is such that $F(\bar{z}) \neq 0$, then $\gamma(\cdot)$ is continuous at $\bar{z}$ with respect to $\Omega$ (that is, $\gamma(z)$ tends to $\gamma(\bar{z})$ as $z \in \Omega$ tends to $\bar{z}$).*

*Proof.* The proof is by verification of the assumptions in [1, Proposition 4.4]. We consider $\Omega$ with the topology induced from $\mathbb{R}^n$ as the parameter space.

The objective function of problem (2.1) does not depend on the parameter $z$ and is obviously everywhere continuous, while the feasible set multifunction $\mathcal{F} : \mathbb{R}^n \to 2^{\mathbb{R}^n \times \mathbb{R}}$ with

$$\mathcal{F}(z) := \left\{ (\zeta, \gamma) \in \mathbb{R}^n \times \mathbb{R} \, \middle| \, \begin{array}{l} \|F(z) + F'(z)\zeta\| \leq \gamma \|F(z)\|^2, \\ \|\zeta\| \leq \gamma \|F(z)\|, \ z + \zeta \in \Omega \end{array} \right\}$$

is obviously closed (under the stated assumptions).

Furthermore, take any $c > 1/\|F(\bar{z})\|$. Then, for all $z \in \mathbb{R}^n$ close enough to $\bar{z}$, it holds that $(0, 1/\|F(z)\|)$ belongs to the level set

$$\{(\zeta, \gamma) \in \mathcal{F}(z) \mid \gamma \leq c\}$$

and moreover, all points in this set satisfy $\|\zeta\| \leq 2c\|F(\bar{z})\|$, $\gamma \leq c$. Hence, those level sets are uniformly bounded for all $z$ sufficiently close to $\bar{z}$. This verifies the inf-compactness condition required by [1, Proposition 4.4].

It remains to show that, for any solution $(\bar{\zeta}, \gamma(\bar{z}))$ of (2.1) with $z = \bar{z}$,

$$\text{dist}\left( (\bar{\zeta}, \gamma(\bar{z})), \ \mathcal{F}(z) \right) \to 0$$

as $z \to \bar{z}$. For all $z \in \mathbb{R}^n$ close enough to $\bar{z}$ it holds that $F(z) \neq 0$, and hence, we can define

$$\zeta := \bar{z} + \bar{\zeta} - z, \quad \gamma := \max \left\{ \frac{\|F(z) + F'(z)\zeta\|}{\|F(z)\|^2}, \ \frac{\|\zeta\|}{\|F(z)\|} \right\}.$$

Then, $(\zeta, \gamma) \in \mathcal{F}(z)$ obviously holds. Moreover, $\zeta$ tends to $\bar{\zeta}$ and $\gamma$ tends to

$$\bar{\gamma} := \max \left\{ \frac{\|F(\bar{z}) + F'(\bar{z})\bar{\zeta}\|}{\|F(\bar{z})\|^2}, \ \frac{\|\bar{\zeta}\|}{\|F(\bar{z})\|} \right\}$$

as $z \to \bar{z}$. Obviously, $\bar{\gamma}$ is the optimal value of (2.1) with $z = \bar{z}$, i.e., $\bar{\gamma} = \gamma(\bar{z})$ holds, which gives the needed property. □

Observe that Lemma 3.1 stays true if the infinity-norm is replaced by any other norm or if $\Omega$ is just a nonempty and closed set, not necessarily polyhedral or convex.

In the next lemma we prove descent properties of a direction which is the $\zeta$-part of the solution of the LP-Newton subproblem.

LEMMA 3.2. *For any $z^k \in \Omega$ and any solution $(\zeta^k, \gamma_k)$ of problem (1.2), it holds that*

$$f'(z^k; \zeta^k) \leq \Delta(z^k), \tag{3.4}$$

*where the functions $f$ and $\Delta(\cdot)$ are defined in (2.2) and (2.3), respectively.*

*Moreover, if*

$$0 \notin \partial f(\bar{z}) + N_\Omega(\bar{z}) \tag{3.5}$$

*is valid for some $\bar{z} \in \Omega$, then there exist $\delta > 0$ and $C > 0$ such that for every $z^k \in \Omega$ close enough to $\bar{z}$, and for every solution $(\zeta^k, \gamma_k)$ of problem (1.2), it holds that*

$$\Delta(z^k) \leq -\delta, \quad \|\zeta^k\| \leq C. \tag{3.6}$$

*Proof.* For every $i \in I_+(z^k)$, taking into account the first constraint in (1.2), we obtain that

$$
\begin{aligned}
\langle F_i'(z^k), \zeta^k \rangle &\leq -F_i(z^k) + |F_i(z^k) + \langle F_i'(z^k), \zeta^k \rangle| \\
&\leq -f(z^k) + \|F(z^k) + F'(z^k)\zeta^k\| \\
&\leq -f(z^k) + \gamma_k (f(z^k))^2 \\
&= \Delta(z^k).
\end{aligned}
$$

Similarly, for every $i \in I_-(z^k)$ we derive that

$$
-\langle F_i'(z^k), \zeta^k \rangle \leq \Delta(z^k).
$$

Hence, (3.2) immediately implies (3.4).

Now, assume that $\bar{z}$ is such that (3.5) holds, and let $(\bar{\zeta}, \bar{\gamma})$ be any solution of problem (2.1) with $z = \bar{z}$. If $F(\bar{z}) = 0$, then $\bar{z}$ is an unconstrained minimizer of $f$, and hence, $0 \in \partial f(\bar{z})$ by the Clarke necessary optimality condition [2, Proposition 2.3.2]. Obviously, this contradicts (3.5). Therefore, $F(\bar{z}) \neq 0$, and the pair $(\zeta, \gamma) = (0, 1/f(\bar{z}))$ is feasible for problem (2.1) with $z = \bar{z}$. This implies that

$$
\bar{\gamma} := \gamma(\bar{z}) \leq 1/f(\bar{z}),
$$

and moreover, this inequality is strict if the specified feasible point $(0, 1/f(\bar{z}))$ is not a solution of this instance of (2.1). We thus proceed with establishing the latter.

The polyhedral set $\Omega$ can always be represented in the following form:

$$
\Omega = \{z \in \mathbb{R}^n \mid \langle a^r, z \rangle \leq b_r, \ r = 1, \ldots, s\},
$$

with some $a^r \in \mathbb{R}^n$, $b_r \in \mathbb{R}$, $r = 1, \ldots, s$. The constraints of the LP problem in question can then be written in the form

(3.7)
$$
\begin{aligned}
\langle F_i'(\bar{z}), \zeta \rangle - (f(\bar{z}))^2 \gamma &\leq -F_i(\bar{z}), & i &= 1, \ldots, m, \\
-\langle F_i'(\bar{z}), \zeta \rangle - (f(\bar{z}))^2 \gamma &\leq F_i(\bar{z}), & i &= 1, \ldots, m, \\
\zeta_j - f(\bar{z})\gamma &\leq 0, & j &= 1, \ldots, n, \\
-\zeta_j - f(\bar{z})\gamma &\leq 0, & j &= 1, \ldots, n, \\
\langle a^r, \zeta \rangle &\leq b_r - \langle a^r, \bar{z} \rangle, & r &= 1, \ldots, s.
\end{aligned}
$$

The constraints of the dual of this LP problem are then

(3.8)
$$
\sum_{i=1}^m (w_i^{1+} - w_i^{1-})F_i'(\bar{z}) + w^{2+} - w^{2-} + \sum_{r=1}^s \omega_r a^r = 0,
$$

$$
(f(\bar{z}))^2 \sum_{i=1}^m (w_i^{1+} + w_i^{1-}) + f(\bar{z}) \sum_{j=1}^n (w_j^{2+} + w_j^{2-}) = 1,
$$

$$
w^{1+} \geq 0, \ w^{1-} \geq 0, \ w^{2+} \geq 0, \ w^{2-} \geq 0, \ \omega \geq 0,
$$

where $w^{1+} \in \mathbb{R}^m$, $w^{1-} \in \mathbb{R}^m$, $w^{2+} \in \mathbb{R}^n$, $w^{2-} \in \mathbb{R}^n$, and $\omega \in \mathbb{R}^s$ are dual variables corresponding to the five groups of constraints in (3.7). Employing now the complementarity conditions for the primal problem at $(\zeta, \gamma) = (0, 1/f(\bar{z}))$, we conclude that it is a solution if and only if there exist $w^{1+} \in \mathbb{R}^m$, $w^{1-} \in \mathbb{R}^m$, and $\omega \in \mathbb{R}^s$, satisfying

$$
\sum_{i \in I_+(\bar{z})} w_i^{1+} F_i'(\bar{z}) - \sum_{i \in I_-(\bar{z})} w_i^{1-} F_i'(\bar{z}) + \sum_{r \in A(\bar{z})} \omega_r a^r = 0,
$$

$$
(f(\bar{z}))^2 \left( \sum_{i \in I_+(\bar{z})} w_i^{1+} + \sum_{i \in I_-(\bar{z})} w_i^{1-} \right) = 1,
$$

$$
w_{I_+(\bar{z})}^{1+} \geq 0, \ w_{I_-(\bar{z})}^{1-} \geq 0, \ \omega_{A(\bar{z})} \geq 0,
$$

where $A(\bar{z}) := \{r \in \{1, \ldots, s\} \mid \langle a^r, \bar{z} \rangle = b_r\}$. Since

$$N_\Omega(\bar{z}) = \text{cone}\{a^r \mid r \in A(\bar{z})\},$$

this evidently implies that the right-hand side of (3.3) has a nonempty intersection with $-N_\Omega(\bar{z})$. But then (3.5) cannot hold, which gives a contradiction.

We have thus established that

$$\bar{\gamma} = \gamma(\bar{z}) < 1/f(\bar{z}),$$

implying that

$$\Delta(\bar{z}) < 0,$$

where the function $\Delta(\cdot)$ is defined according to (2.3). Recall that by Lemma 3.1, $\gamma(\cdot)$ is continuous at $\bar{z}$ with respect to $\Omega$. Therefore,

$$(3.9) \qquad \gamma_k = \gamma(z^k) \to \bar{\gamma} = \gamma(\bar{z}) \text{ as } z^k \to \bar{z}.$$

Fix any $\delta \in (0, \Delta(\bar{z}))$. By continuity of the function $\Delta(\cdot)$ defined in (2.3) at $\bar{z}$ with respect to $\Omega$, we then have that

$$\Delta(z^k) \leq -\delta$$

holds for every $z^k$ close enough to $\bar{z}$. This proves the first relation in (3.6).

The second relation in (3.6) follows from the second constraint in (1.2), from (3.9), and from continuity of $F$ at $\bar{z}$. □

Assuming that $F(z^k) \neq 0$, Lemma 3.2 implies, in particular, that $\Delta(z^k) = 0$ if and only if

$$(3.10) \qquad 0 \in \partial f(z^k) + N_\Omega(z^k)$$

holds or, equivalently, if and only if $(\zeta, \gamma) = (0, 1/f(z^k))$ is a solution of (1.2). Otherwise, any direction $\zeta^k$ obtained by solving (1.2) is a direction of descent for $f$ at $z^k$.

We next give our main global convergence result.

THEOREM 3.1. *Algorithm 1 is well-defined and, for any starting point $z^0 \in \Omega$, it either terminates with some iterate $z^k$ satisfying (3.10) or generates an infinite sequence $\{z^k\}$ such that any accumulation point $\bar{z}$ of this sequence satisfies (3.1). Moreover, in the latter case, for every subsequence $\{z^{k_j}\}$ convergent to $\bar{z}$, it holds that $\Delta(z^{k_j}) \to 0$ as $j \to \infty$.*

*Proof.* As already commented and easy to see, the subproblem (1.2) always has solutions, unless $F(z^k) = 0$. If $F(z^k) = 0$ then the method stops, and (3.10) is obvious in that case. The other case of finite termination is given by $\Delta(z^k) = 0$, and (3.10) also holds, as explained above.

Furthermore, from (3.4) and from the first relation in (3.6) of Lemma 3.2, the standard directional descent argument shows that if (3.10) does not hold, then the linesearch procedure in step 3 of Algorithm 1 is well-defined, i.e., it generates some stepsize value $\alpha_k > 0$ after a finite number of backtrackings. Therefore, Algorithm 1 is well-defined.

Suppose that (3.10) never holds, and in particular, the algorithm generates an infinite sequence $\{z^k\}$. By (2.4) and the first relation in (3.6) of Lemma 3.2, the

sequence $\{f(z^k)\}$ is monotonically decreasing. Since this sequence is bounded below (by zero), it converges. Then (2.4) implies that

$$(3.11) \qquad\qquad \alpha_k \Delta(z^k) \to 0$$

as $k \to \infty$.

Let $\bar{z} \in \Omega$ be an accumulation point of the sequence $\{z^k\}$, violating (3.1), and let $\{z^{k_j}\}$ be a subsequence convergent to $\bar{z}$ as $j \to \infty$. If $\{\alpha_{k_j}\}$ is bounded away from zero by some positive constant, then (3.11) implies that $\Delta(z^{k_j}) \to 0$, and according to Lemma 3.2, this would only be possible if (3.1) holds. Therefore, it remains to consider the case when $\liminf_{j\to\infty} \alpha_{k_j} = 0$, and to demonstrate that such a situation is not possible.

Passing to a further subsequence if necessary, we may assume that $\alpha_{k_j} \to 0$ as $j \to \infty$. Also, by the second relation in (3.6) of Lemma 3.2, we may assume that $\{\zeta^{k_j}\}$ converges to some $\bar{\zeta}$. Then, for all $j$ large enough, it holds that $\alpha_{k_j} < 1$, which implies that $\alpha = \alpha_{k_j}/\theta$ does not satisfy (2.4), i.e., it holds that

$$(3.12) \qquad f(z^{k_j} + (\alpha_{k_j}/\theta)\zeta^{k_j}) > f(z^{k_j}) + \sigma(\alpha_{k_j}/\theta)\Delta(z^{k_j}).$$

Observe that the function $f$ is everywhere Clarke regular, as a composition of a convex function and a continuously differentiable mapping (see [2, Proposition 2.3.6 (b), Theorem 2.3.10]). Then, by the definition of Clarke directional derivative, using also the Lipschitz-continuity of $f$ near $\bar{z}$ and the continuity of $\Delta(\cdot)$ at $\bar{z}$ with respect to $\Omega$, as well as (3.12), we obtain that

$$
\begin{aligned}
f'(\bar{z}; \bar{\zeta}) = f^\circ(\bar{z}; \bar{\zeta}) \\
&\geq \limsup_{j\to\infty} \frac{f(z^{k_j} + (\alpha_{k_j}/\theta)\bar{\zeta}) - f(z^{k_j})}{\alpha_{k_j}/\theta} \\
&= \limsup_{j\to\infty} \frac{f(z^{k_j} + (\alpha_{k_j}/\theta)\zeta^{k_j}) - f(z^{k_j})}{\alpha_{k_j}/\theta} \\
&\geq \sigma \lim_{j\to\infty} \Delta(z^{k_j}) \\
(3.13) \qquad &= \sigma\Delta(\bar{z}).
\end{aligned}
$$

Note that the continuity of $\Delta(\cdot)$ at $\bar{z}$ with respect to $\Omega$ follows from the continuity of $\gamma(\cdot)$ at $\bar{z}$ with respect to $\Omega$. The latter was proved in Lemma 3.1. Using the continuity of $\gamma(\cdot)$ at $\bar{z}$ with respect to $\Omega$ again, we conclude that $(\bar{\zeta}, \gamma(\bar{z}))$ is feasible for problem (1.2) with $z^k := \bar{z}$. Hence, this point is a solution of this problem. Then, by (3.4) and by the first relation in (3.6) of Lemma 3.2, we obtain that

$$f'(\bar{z}; \bar{\zeta}) \leq \Delta(\bar{z}) < 0,$$

which contradicts (3.13) because $\sigma \in (0,1)$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Recall that according to [2, Corollary 2.4.3], (3.1) is a necessary optimality condition for the problem

$$(3.14) \qquad \begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & z \in \Omega. \end{array}$$

For general locally Lipschitz-continuous functions, this condition does not imply B-stationarity of $\bar{z}$, which consists of saying that

$$(3.15) \qquad\qquad f'(\bar{z}; \zeta) \geq 0 \qquad \text{for all } \zeta \in T_\Omega(\bar{z}).$$

Indeed, (3.1) holds at $\bar{z} := 0$ for $f : \mathbb{R} \to \mathbb{R}$, $f(z) := -|z|$, $\Omega := \mathbb{R}$, while (3.15) is evidently violated.

However, in our specific setting, B-stationarity does hold, as we show next.

PROPOSITION 3.2. *If* (3.1) *holds, then* $\bar{z}$ *is a B-stationary point of problem* (3.14), *i.e.,* (3.15) *holds.*

*Proof.* Recall again that by [2, Proposition 2.3.6 (b), Theorem 2.3.10], $f$ is Clarke-regular at $\bar{z}$ (which is in fact the only property $f$ must have for the assertion of this proposition to be valid). From [2, Proposition 2.1.2] we then derive that for every $\zeta \in \mathbb{R}^n$,

$$(3.16) \qquad f'(\bar{z}; \zeta) = f^\circ(\bar{z}; \zeta) = \max\{\langle g, \zeta \rangle \mid g \in \partial f(\bar{z})\}.$$

At the same time, condition (3.1) is equivalent to the existence of $g \in \partial f(\bar{z})$ such that

$$\langle g, \zeta \rangle \geq 0 \qquad \text{for all } \zeta \in T_\Omega(\bar{z}).$$

Therefore, the right-hand side of (3.16) is nonnegative for $\zeta \in T_\Omega(\bar{z})$. This gives (3.15). □

**3.2. Local quadratic convergence.** We next establish conditions under which the unit stepsize is accepted by the linesearch procedure in step 3 of Algorithm 1. Naturally, this leads directly to the quadratic convergence rate of the algorithm.

We note that the next lemma is valid for any norm.

LEMMA 3.3. *Let* $\bar{z}$ *be a solution of problem* (1.1), *and assume that* $\gamma(\cdot)$ *is bounded from above on* $\Omega$ *near* $\bar{z}$.

*Then for every* $z^k \in \Omega \setminus Z$ *close enough to* $\bar{z}$, *it holds that* $\Delta(z^k) < 0$, *and for every solution* $(\zeta^k, \gamma_k)$ *of problem* (1.2),

$$\|F(z^k + \zeta^k)\| \leq \|F(z^k)\| + \sigma\Delta(z^k),$$

*where* $\Delta(\cdot)$ *is defined by* (2.3).

*Proof.* By the mean-value theorem, and by the constraints in (1.2), we derive that

$$
\begin{aligned}
\|F(z^k + \zeta^k)\| &\leq \|F(z^k + \zeta^k) - F(z^k) - F'(z^k)\zeta^k\| + \|F(z^k) + F'(z^k)\zeta^k\| \\
&\leq \sup\{\|F'(z^k + t\zeta^k) - F'(z^k)\| \mid t \in [0, 1]\}\|\zeta^k\| + \gamma_k\|F(z^k)\|^2 \\
&\leq \sup\{\|F'(z^k + t\zeta^k) - F'(z^k)\| \mid t \in [0, 1]\}\gamma_k\|F(z^k)\| + \gamma_k\|F(z^k)\|^2 \\
(3.17) \qquad &= o(\|F(z^k)\|)
\end{aligned}
$$

as $z^k \to \bar{z}$. Moreover, by the definition of $\Delta(\cdot)$ in (2.3) we have

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

Hence, $\Delta(z^k) < 0$ and

$$\|F(z^k + \zeta^k)\| - \|F(z^k)\| - \sigma\Delta(z^k) = -(1 - \sigma)\|F(z^k)\| + o(\|F(z^k)\|) \leq 0$$

follow, provided $z^k$ is close enough to $\bar{z}$. □

Now, local convergence results of Algorithm 1 can be derived using those for the original LP-Newton method in [8]; see also [13].

THEOREM 3.3. *Let the derivative of $F$ be Lipschitz-continuous near $\bar{z} \in Z$. Assume that there exists $\omega > 0$ such that the error bound*

$$(3.18) \qquad\qquad \mathrm{dist}(z,\, Z) \le \omega \|F(z)\|$$

*holds for all $z \in \Omega$ close enough to $\bar{z}$.*

*If Algorithm 1 generates an iterate close enough to $\bar{z}$, then either the algorithm terminates at some $z^k \in Z$, or generates an infinite sequence $\{z^k\}$ convergent to some $\hat{z} \in Z$, and the rate of convergence is Q-quadratic.*

*Proof.* First note that according to [8, Corollary 1 and Proposition 1], the requirements stated above guarantee that Assumptions 1–4 in [8] are satisfied at $\bar{z}$. Then, from [8, Theorem 1], employing also some additional facts from its proof, we have that for every $\varepsilon > 0$ there exists $\rho > 0$ with the following property: if some iterate $z^k$ of a local counterpart of Algorithm 1, employing $\alpha_k = 1$ for all $k$, satisfies $\|z^k - \bar{z}\| \le \rho$, then for all subsequent iterates it holds that $\|z^k - \bar{z}\| \le \varepsilon$, and either the algorithm terminates at some iteration, or it generates an infinite sequence converging Q-quadratically to some $\hat{z} \in Z$ .

Assumption 3 in [8] consists of saying that $\gamma(\cdot)$ is bounded from above on $\Omega$ near $\bar{z}$. By the definition of the mapping $\Delta(\cdot)$ and by Lemma 3.3 it then follows that there exists $\varepsilon > 0$ such that for every $z^k \in \Omega$ satisfying $\|z^k - \bar{z}\| \le \varepsilon$ and $F(z^k) \ne 0$, it holds that $\Delta(z^k) \ne 0$ (and in particular, the algorithm cannot terminate at such point) and the line search procedure in step 3 of Algorithm 1 accepts $\alpha_k = 1$. For a corresponding $\rho > 0$, we then have that once some iterate $z^k$ of Algorithm 1 satisfies $\|z^k - \bar{z}\| \le \rho$, the algorithm either terminates at some point in $Z$ or generates an infinite sequence converging Q-quadratically to some $\hat{z} \in Z$. $\qquad\square$

**4. The case of piecewise smooth equations.** In this section we extend the results developed above to the case when $F$ belongs to a certain class of piecewise continuously differentiable (PC$^1$-) mappings. The latter 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

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

for all $z \in \mathbb{R}^n$. The mappings $F^1, \ldots, F^q$ are called selection mappings.

An important class of problems that gives rise to piecewise smooth equations are systems that involve complementarity conditions. For example, consider the problem of finding a point $z \in \mathbb{R}^n$ such that

$$(4.1) \qquad a(z) = 0, \quad b(z) \ge 0, \quad c(z) \ge 0, \quad d(z) \ge 0, \quad \langle c(z),\, d(z) \rangle = 0,$$

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

$$(4.2) \qquad\qquad F(z) := (a(z),\, \min\{c(z),\, d(z)\}),$$

and define

$$(4.3) \qquad\qquad \Omega := \{z \in \mathbb{R}^n \mid b(z) \ge 0,\, c(z) \ge 0,\, d(z) \ge 0\},$$

then (4.1) is equivalent to the constrained equation (1.1) with piecewise smooth $F$. Convergence conditions for the associated LP-Newton method are analyzed in detail in [13]. The importance of this specific choice of $\Omega$ is discussed in [8, 13]. Moreover,

introducing slack variables, we can reformulate the original problem into a constrained equation with the set $\Omega$ being polyhedral (if it was not polyhedral in the original formulation). One important example of a complementarity system (4.1) itself, to be considered in Section 5 below, is given by KKT-type conditions for GNEPs; see, e.g., [5, 9, 8, 13, 19].

For $z \in \mathbb{R}^n$, we denote by $\mathcal{A}(z)$ the index set of all selection mappings which are active at $z$, i.e.,

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

Throughout this section, $G : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ is any matrix-valued mapping such that

$$(4.4) \qquad G(z) \in \{(F^p)'(z) \mid p \in \mathcal{A}(z)\}$$

holds for all $z \in \mathbb{R}^n$. For every $p = 1, \ldots, q$, let $f_p : \mathbb{R}^n \to \mathbb{R}_+$ be given by

$$(4.5) \qquad f_p(z) := \|F^p(z)\|.$$

As in the previous section, we establish global convergence properties of Algorithm 1 first, in Section 4.1. Conditions for local fast convergence are analyzed in Section 4.2.

**4.1. Global convergence.** The following is a counterpart of Lemma 3.2.

LEMMA 4.1. *For any $z^k \in \Omega$ and any solution $(\zeta^k, \gamma_k)$ of problem (1.2) with $G(z^k) = (F^p)'(z^k)$ for some $p \in \mathcal{A}(z^k)$, it holds that*

$$(4.6) \qquad f_p'(z^k; \zeta^k) \leq \Delta(z^k),$$

*where $f_p$ is defined in (4.5), and $\Delta(\cdot)$ is defined in (2.3).*

*Moreover, for any $\bar{z} \in \Omega$ such that*

$$(4.7) \qquad 0 \notin \partial f_p(\bar{z}) + N_\Omega(\bar{z}) \qquad \text{for all } p \in \mathcal{A}(\bar{z}),$$

*there exist $\delta > 0$ and $C > 0$ such that (3.6) holds for every $z^k \in \Omega$ close enough to $\bar{z}$ and every solution $(\zeta^k, \gamma_k)$ of problem (1.2) with $G(z^k) = (F^p)'(z^k)$ for any $p \in \mathcal{A}(z^k)$.*

The proof repeats the one of Lemma 3.2 with evident modifications.

We note that the global convergence result of Theorem 3.1 does not extend to an arbitrary $PC^1$-mapping. However, in the next theorem we specify an important class of $PC^1$-mappings for which an extension is possible. In particular, it can be easily seen that condition (4.8) on smooth selections of $F$ in Theorem 4.1 below holds for the reformulation (1.1) of the complementarity system (4.1) given by (4.2), (4.3).

THEOREM 4.1. *Assume that*

$$(4.8) \qquad f(z) \leq f_p(z) \qquad \text{for all } p \in \{1, \ldots, q\} \text{ and all } z \in \Omega.$$

*Then, Algorithm 1 is well-defined and, for any starting point $z^0 \in \Omega$, it either terminates with some iterate $z^k \in \Omega$ satisfying*

$$(4.9) \qquad 0 \in \partial f_p(z^k) + N_\Omega(z^k)$$

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

$$(4.10) \qquad 0 \in \partial f_p(\bar{z}) + N_\Omega(\bar{z})$$

*for at least one $p \in \mathcal{A}(\bar{z})$. Moreover, in the latter case, for every subsequence $\{z^{k_j}\}$ convergent to $\bar{z}$, it holds that $\Delta(z^{k_j}) \to 0$ as $j \to \infty$.*

*Proof.* The well-definedness of Algorithm 1 can be shown similar as in the proof of Theorem 3.1, now by means of Lemma 4.1 (instead of using Lemma 3.2). However, there is one exception. Given some $z^k$ with $F(z^k) \neq 0$ and $\Delta(z^k) \neq 0$, we obtain (as in the proof of Theorem 3.1 with $f$ replaced by $f_p$) that step 3 of the algorithm determines some $\alpha_k > 0$ so that

$$(4.11) \qquad f_p(z^k + \alpha_k \zeta^k) \leq f_p(z^k) + \sigma \alpha_k \Delta(z^k)$$

holds for $p \in \mathcal{A}(z^k)$ with $G(z^k) = (F^p)'(z^k)$. Now, the additional condition (4.8) comes into play. Using this condition, $f_p(z^k) = f(z^k)$ for all $p \in \mathcal{A}(z^k)$, and (4.11), we conclude that (2.4) holds with $\alpha := \alpha_k$, and hence, the algorithm successfully generates $z^{k+1}$.

Algorithm 1 may only terminate for some iteration index $k$ if either $z^k$ is a solution of (1.1) or $\Delta(z^k) = 0$. According to Lemma 4.1, in both cases (4.9) must hold for some $p \in \mathcal{A}(z^k)$ (if $F(z^k) = 0$, then $z^k$ is an unconstrained minimizer of $f_p$ for every $p \in \mathcal{A}(z^k)$, implying that $0 \in \partial f_p(z^k)$ by [2, Proposition 2.3.2]).

From now on, suppose that (4.9) never holds for any $p \in \mathcal{A}(z^k)$, and the algorithm generates an infinite sequence $\{z^k\}$. The rest of the proof essentially repeats the one for Theorem 3.1, with only one modification: a subsequence $\{z^{k_j}\}$ convergent to $\bar{z}$ should be selected in such a way that $G(z^{k_j}) = (F^p)'(z^{k_j})$ for all $j$, for some fixed $p \in \mathcal{A}(\bar{z})$. Such a subsequence always exists since $\mathcal{A}(z) \subset \mathcal{A}(\bar{z})$ for all $z \in \mathbb{R}^n$ close enough to $\bar{z}$, where $\mathcal{A}(\bar{z})$ is a finite set, and due to (4.4). In the remaining part of the proof, one should only replace $f$ by the corresponding $f_p$. $\square$

Observe that property (4.10) for some $p \in \mathcal{A}(\bar{z})$ does not necessarily imply that the first-order necessary condition (3.1) is satisfied, even if (4.8) holds and even if we restrict ourselves to reformulation (1.1) of complementarity system (4.1), employing (4.2), (4.3). This raises the following question. What can be done if a sequence gets stuck near some $\bar{z}$ satisfying (4.10) for some $p \in \mathcal{A}(\bar{z})$ but violating (3.1): is it possible to escape from such points and to guarantee B-stationarity of accumulation points for problem (3.14)? We leave this question for future research. This would definitely require a great deal of further theoretical developments and numerical testings.

**4.2. Local quadratic convergence.** We proceed with local convergence properties.

THEOREM 4.2. *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 [8] are satisfied at $\bar{z}$, and that (4.8) holds.*

*If Algorithm 1 generates an iterate which is close enough to $\bar{z}$, then either the algorithm terminates at some $z^k \in Z$ or it generates an infinite sequence $\{z^k\}$ convergent to some $\hat{z} \in Z$, and the rate of convergence is Q-quadratic.*

*Proof.* After observing that Assumption 1 in [8] is automatic for $PC^1$-mappings and that (4.8) implies Assumption 4 in [8], the proof repeats almost literally that of Theorem 3.3, but with Lemma 3.3 applied to a selection mapping $F^p$, $p \in \mathcal{A}(z^k)$, such that $G(z^k) = (F^p)'(z^k)$, and taking into account (4.8). $\square$

In [13] a thorough discussion of Assumptions 2–4 from [8] for the case when $F$ is a $PC^1$-function can be found. The results in [13] lead to the following.

COROLLARY 4.1. *Assume that the selection mappings $F^1$, ..., $F^q$ have Lipschitz-continuous derivatives near $\bar{z} \in Z$. Assume further that* (4.8) *is satisfied and that there exists $\omega > 0$ such that*

$$(4.12) \qquad \qquad \operatorname{dist}(z, Z_p) \leq \omega \|F^p(z)\|$$

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

*Then, the assertion of Theorem 4.2 is valid.*

*Proof.* Observe that (4.12) exactly corresponds to Condition 2 in [13], while Condition 3 in [13] follows immediately from (4.8). Therefore, by [13, Proposition 3.6, Theorem 3.8], Assumptions 2–4 in [8] are always satisfied under the assumptions of this corollary, and the needed result follows from Theorem 4.2. $\qquad \square$

We emphasize that Theorem 3.3 and Corollary 4.1 ensure local quadratic convergence by exploiting the LP-Newton subproblems (1.2), under assumptions that do not imply the local uniqueness of the solution $\bar{z}$ of the original problem (1.1). Of course, solving the linear programming subproblems in the LP-Newton method might be more costly than solving systems of linear equations, but this cost is certainly still reasonable. A larger class of reasonable (convex quadratic programming) subproblems can be found in [7]. The Newton framework in [20, Chapters 10, 11] covers several other subproblems coming from linear or nonlinear approximations of a locally Lipschitzian nonlinear equation. However, the assumptions for convergence used in the latter framework imply that the solution in question is isolated.

**5. Computational results.** We start with discussing some details of implementation, as well as useful modifications of Algorithm 1, which do not affect the presented convergence theory.

Observe that intuitively, far from solutions ($\|F(z^k)\|$ large), the constant in the right-hand side of the second constraint in (1.2), i.e., of

$$(5.1) \qquad \qquad \|\zeta\| \leq \gamma \|F(z^k)\|,$$

can be too small relative to the right-hand side of the first constraint in (1.2), namely

$$\|F(z^k) + G(z^k)\zeta\| \leq \gamma \|F(z^k)\|^2.$$

Or, from a related viewpoint, the constraint (5.1) can be "too tight", forcing rather short directions $\zeta^k$ (at least, shorter than desirable, in some sense). To see this, consider the following simple example (already employed in [8]). Let $n := m := 1$, $F(z) := z$, $\Omega := \mathbb{R}$. Then, for any $z^k \in \mathbb{R} \setminus \{0\}$, the unique solution of (1.2) is

$$(\zeta^k, \gamma_k) = \left( -\frac{z^k}{1 + |z^k|}, \ \frac{1}{1 + |z^k|} \right),$$

and the length of $\zeta^k$ is close to 1 when $|z^k|$ is large. Therefore, even full steps in such directions would be a bit short compared to the distance to the solution, which is $|z^k|$. Clearly, longer directions are desirable.

Fortunately, the following modification of (5.1) does the trick, while not affecting our convergence theory. In our computational experiments, we replaced (5.1) by

$$(5.2) \qquad \qquad \|\zeta\| \leq \gamma \max\{\|F(z^k)\|, \ \tau_k \|F(z^k)\|^2\},$$

where $\tau_k$ is some parameter chosen from an interval $[\tau_{\min}, \tau_{\max}]$ with given $\tau_{\max} > \tau_{\min} > 0$ (our precise rule for $\tau_k$ will be given later). Note that when $z^k$ is close to some solution (so that $\|F(z^k)\| \leq 1/\tau_k$, which automatically holds when $\|F(z^k)\| \leq 1/\tau_{\max}$), (5.2) becomes exactly the same as (5.1). Hence, the proposed modification does not interfere with the local convergence theory. Moreover, it can be easily verified directly that replacing constraint (5.1) within subproblem (1.2) by (5.2) does not affect our results on the global convergence of Algorithm 1 in Theorems 3.1 and 4.1. Roughly speaking, using (5.2) instead of (5.1) within subproblem (1.2) only leads to a possibly larger feasible set of this subproblem without affecting its solvability, essential parts of optimality conditions, etc. This is enough for obtaining global convergence.

As mentioned above, far from solutions, constraint (5.2) is usually less tight than its counterpart (5.1), and thus allows for longer directions $\zeta^k$ generated as solution of subproblem (1.2). To demonstrate the latter, let us go back to the example considered in the beginning of this section. There, the modified subproblem (1.2) with constraint (5.2) instead of (5.1) has the unique solution

$$(\zeta^k, \gamma_k) = \left( -\frac{\tau_k z^k}{1 + \tau_k}, \ \frac{1}{(1 + \tau_k)|z^k|} \right).$$

In particular, $|\zeta^k|$ grows as $\tau_k$, thus allowing longer directions than of order 1 in the original scheme (approaching the "ideal" direction $\zeta^k = -z^k$ as $\tau_k \to +\infty$).

Another useful observation is the following. The proof of Lemma 3.2 suggests to use the pair $(\zeta, \gamma) = (0, 1/f(z^k))$ as the starting point to initialize the LP solver at step 2 of Algorithm 1, at least when $z^k$ is far from solutions (i.e., when $f(z^k)$ is not too small).

We tested our proposed globalization of the LP-Newton method on a library of GNEPs used in [6] and [5]. We refer to [6] for information about the dimensions, i.e., the numbers of players, variables, and inequality constraints. The references where these GNEPs are described in detail can also be found in [6]. We note that the choice of GNEPs is very appropriate for our numerical assessment, because this is a difficult problem class, in particular with naturally nonisolated solutions. Thus, strong convergence properties of the LP-Newton method are especially relevant in this context.

Before reporting the results of our numerical tests, let us briefly recall the GNEP setting, as well as suitable reformulations of the corresponding KKT-type system as a constrained equation. In a GNEP, $N$ players compete with each other. A player indexed by $\nu$ controls $n_\nu$ variables. The vector of the $\nu$-th player's variables is denoted by $x^\nu \in \mathbb{R}^{n_\nu}$. The aim of each player is to minimize his individual objective function $\varphi_\nu : \mathbb{R}^{n_x} \to \mathbb{R}$, where $n_x := \sum_{\nu=1}^N n_\nu$ indicates the total number of variables (of all the players combined). Each objective function $\varphi_\nu$ may depend on both the variables of the $\nu$-th player and the variables $x^{-\nu}$ of the rival players. GNEPs in our test library have inequality constraints, i.e., the optimization problem of the $\nu$-th player is given by

$$(5.3) \qquad \begin{array}{ll} \underset{x^\nu}{\text{minimize}} & \varphi_\nu(x^\nu, x^{-\nu}) \\ \text{subject to} & g^\nu(x^\nu, x^{-\nu}) \leq 0. \end{array}$$

The functions $\varphi_\nu : \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_x - n_\nu} \to \mathbb{R}$ and the mappings $g^\nu : \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_x - n_\nu} \to \mathbb{R}^{m_\nu}$, $\nu = 1, \ldots, N$, are twice differentiable with locally Lipschitz-continuous second derivatives for each problem in the library, at least in some neighborhood of the solution

set. Problems A10a–A10e in the test set all contain one affine equality-constraint. To comply with the structure of problem (5.3) this constraint was transformed into two inequalities (as it was also done in [5], for instance).

The (parametric) Lagrangian $L_\nu : \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_x - n_\nu} \times \mathbb{R}^{m_\nu} \to \mathbb{R}$ of (5.3) is given by

$$L_\nu(x^\nu,\, x^{-\nu},\, \lambda^\nu) := \varphi_\nu(x^\nu,\, x^{-\nu}) + \langle \lambda^\nu,\, g^\nu(x^\nu,\, x^{-\nu}) \rangle.$$

By means of $L_\nu$, we can write down the KKT system of the $\nu$-th player's optimization problem:

$$\frac{\partial L_\nu}{\partial x^\nu}(x^\nu,\, x^{-\nu},\, \lambda^\nu) = 0, \quad \lambda^\nu \geq 0, \quad g^\nu(x^\nu,\, x^{-\nu}) \leq 0, \quad \langle \lambda^\nu,\, g^\nu(x^\nu,\, x^{-\nu}) \rangle = 0.$$

Concatenating the KKT systems of all players yields the KKT-type system of the GNEP:

$$(5.4) \qquad \mathcal{L}(x, \lambda) = 0, \quad \lambda \geq 0, \quad g(x) \leq 0, \quad \langle \lambda^\nu,\, g^\nu(x) \rangle = 0,\ \nu = 1,\, \ldots,\, N,$$

with

$$\lambda := \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^N \end{pmatrix},\ g(x) := \begin{pmatrix} g^1(x) \\ \vdots \\ g^N(x) \end{pmatrix},\ \text{and} \quad \mathcal{L}(x, \lambda) := \begin{pmatrix} \dfrac{\partial L_1}{\partial x^1}(x^1,\, x^{-1},\, \lambda^1) \\ \vdots \\ \dfrac{\partial L_N}{\partial x^N}(x^N,\, x^{-N},\, \lambda^N) \end{pmatrix}.$$

Thus, the GNEP KKT-type system (5.4) is an instance of the complementarity system (4.1). It can be cast in the form of a constrained equation (1.1), using both smooth and piecewise smooth mappings. For our numerical tests, we considered two reformulations of (5.4).

The first one employs the smooth mapping $F$ given by

$$(5.5) \qquad\qquad F(z) := \begin{pmatrix} \mathcal{L}(x, \lambda) \\ g(x) + u \\ \lambda \circ u \end{pmatrix}$$

and the constraint set

$$(5.6) \qquad\qquad \Omega := \mathbb{R}^{n_x} \times \mathbb{R}_+^{m_g} \times \mathbb{R}_+^{m_g},$$

where $z := (x, \lambda, u)$, $m_g := \sum_{\nu=1}^N m_\nu$ denotes the total number of inequality constraints and slack variables $u = (u^1,\, \ldots,\, u^N) \in \mathbb{R}^{m_g}$, and $\lambda \circ u := (\lambda_1 u_1, \ldots, \lambda_{m_g} u_{m_g})$ is the Hadamard product. Slack variables have been introduced in order to obtain a polyhedral set $\Omega$.

The second reformulation employs the nonsmooth mapping

$$(5.7) \qquad\qquad F(z) := \begin{pmatrix} \mathcal{L}(x, \lambda) \\ g(x) + u \\ \min\{\lambda,\, u\} \end{pmatrix},$$

where the min-operation is understood componentwise, and the same $\Omega$ defined in (5.6). Note that the equivalence of (5.4) and (1.1) with $F$ defined in (5.7) is valid even for $\Omega = \mathbb{R}^{n_x} \times \mathbb{R}^{m_g} \times \mathbb{R}^{m_g}$. However, the assumptions in [8] implying local

quadratic convergence of the LP-Newton method (and hence, of our Algorithm 1) may not hold without including the bound constraints $\lambda \geq 0$ and $u \geq 0$ in the definition of $\Omega$. Moreover, these constraints are also needed to satisfy condition (4.8), required for the global convergence analysis in Theorem 4.1.

For each of the 35 GNEPs in the test library we used five different methods to solve the corresponding KKT-type system. The first method is the modification of Algorithm 1 corresponding to (5.2) and explained in the beginning of the current section, applied to the smooth reformulation with $F$ defined in (5.5) and $\Omega$ defined in (5.6). Since $F$ is smooth, we set $G := F'$. In the following, this first method is referred to as "LP-N smooth".

The second method is the same modification of Algorithm 1, but employing nonsmooth $F$ defined in (5.7). Since this $F$ is a $PC^1$-mapping, we use the matrix-valued mapping $G$ such that for any $z$, matrix $G(z)$ is the derivative of one of the selection mappings active at $z$ (as described in Section 4). More precisely, we use the following rule: if, for some $z = (x, \lambda, u)$, it holds that $\lambda_i = u_i$ for some $i \in \{1, \ldots, m_g\}$, then the selection mapping was used whose $(n_x + m_g + i)$-th component is equal to $\lambda_i$. We refer to the second method as "LP-N nonsmooth".

The third method is a modification of the first one, with the (monotone) Armijo linesearch rule (2.4) replaced by a nonmonotone linesearch. Specifically, on the $k$-th iteration, a stepsize $\alpha$ is chosen to satisfy the condition

$$(5.8) \qquad f(z^k + \alpha\zeta^k) \leq R_k + \sigma\alpha\Delta(z^k),$$

with the reference value

$$R_k := \max\{f(z^\ell) \mid \max\{0, k - 10\} \leq \ell \leq k\}.$$

Analogously, the fourth method is the same as the second, but using the nonmonotone linesearch (5.8) instead of the Armijo rule (2.4). Subsequently, the third method and the fourth method are referred to as "LP-N nonmonotone smooth" and "LP-N nonmonotone nonsmooth", respectively.

As a benchmark for comparison, we used the hybrid method proposed in [5] for (1.1) with $F$ given by (5.5) and $\Omega$ defined in (5.6). The global phase of this hybrid method is the potential reduction algorithm proposed in [6] (see also [21]), whereas the local part is the LP-Newton method. We will refer to the hybrid method as "PRA + LP-N". Note that this hybrid method is in fact the only other currently available globalization of the LP-Newton method, with one exception: a similar hybrid method analyzed in [4] where the local part (the LP-Newton method) is applied to the nonsmooth reformulation of the KKT system.

All algorithms were implemented in MATLAB®, version 8.5.0.197613 (R2015a). For solving LPs, we used the solver "cplexlp" from the optimization toolbox CPLEX®, version 12.6.0, with standard settings except that the tolerances used for stopping the solver were set to $10^{-9}$.

As for the parameters, we set $\theta := 0.5$ and $\sigma := 0.001$ for all variants of Algorithm 1. The following updating rule for $\tau_k$ in the modified constraint (5.2) was used:

$$\tau_0 := 1,$$
$$\tau_{k+1} := \begin{cases} \min\{10\tau_k, \tau_{\max}\} & \text{if } \|\zeta^k\| \geq \gamma_k \max\{\|F(z^k)\|, \tau_k\|F(z^k)\|^2\} - 10^{-8}, \\ \max\{0.1\tau_k, \tau_{\min}\} & \text{if } \|\zeta^k\| < \gamma_k \max\{\|F(z^k)\|, \tau_k\|F(z^k)\|^2\} - 10^{-8}, \end{cases}$$

where we set $\tau_{\min} := 1$ and $\tau_{\max} := 10^8$. Roughly speaking, we increase $\tau_k$ if the constraint $\|\zeta\| \leq \gamma \max\{\|F(z^k)\|, \tau_k\|F(z^k)\|^2\}$ was active at $(\zeta^k, \gamma_k)$ (up to round-off errors), and decrease it otherwise.

For the hybrid method PRA + LP-N, all the parameters are the same used in [5], with one exception: the parameter in the Armijo linesearch (called $\eta$ in [5]) was set to 0.001 instead of 0.01. We also used certain modifications suggested in [5, Section 3.4].

In all the algorithms, if the LP solver applied to (1.2) reports an error on some iteration $k$, then we try to solve the following LP instead:

$$
(5.9) \quad
\begin{aligned}
\text{minimize} \quad & \tilde{\gamma} \\
\text{subject to} \quad & \|F(z^k) + G(z^k)\zeta\| \leq \tilde{\gamma}\|F(z^k)\|, \\
& \|\zeta\| \leq \tilde{\gamma}, \\
& z^k + \zeta \in \Omega,
\end{aligned}
$$

with respect to $(\zeta, \gamma)$. It can easily be verified that a point $(\zeta^k, \gamma_k)$ is a solution of (1.2) if and only if $(\zeta^k, \tilde{\gamma}_k)$ with $\tilde{\gamma}_k := \gamma_k\|F(z^k)\|$ solves (5.9). The motivation to consider (5.9) instead of (1.2) is that the former might be better scaled when $\|F(z^k)\|$ is very small.

We terminate the algorithms once one of the following conditions is satisfied:

- $\|F(z^k)\| \leq 10^{-8}$ with $F$ defined in (5.7);
- the LP solver reports an error on some iteration (even after trying to solve (5.9) instead of (1.2));
- the number of iterations exceeds $k_{\max} := 500$;
- the trial value for stepsize $\alpha$ goes below $\alpha_{\min} := 10^{-13}$, i.e., the descent condition ((2.4) or (5.8), depending on the method) is not satisfied for $\alpha \geq \alpha_{\min}$;
- $|\Delta(z^k)| \leq 10^{-12}$, where $\Delta(z^k)$ is defined by (2.3) and uses $\gamma(z^k) := \gamma_k$ with $(\zeta^k, \gamma_k)$ obtained from solving the subproblem with constraint (5.1) replaced by (5.2) (this stopping criterion is only relevant for the four variants of Algorithm 1).

A run terminated according to the first criterion is regarded successful; all other outcomes are failures.

For each test problem and each method, we performed 20 runs with different starting points (the same for all methods). We next describe how any given starting point $z^0 = (x^0, \lambda^0, u^0)$ was generated. First, the components $x_j^0$ of $x^0$ were independently drawn from the interval $[l, 20]$, $j = 1, \ldots, n_x$, according to the uniform distribution. For most of the test problems, we set $l := -20$. However, we increased the value for some problems in order to guarantee that all problem functions of the GNEP are defined at $x^0$. More precisely, we used a higher value of $l$ for the following problems: A1, A2, A14, A16a–A16d, and Lob, where we set $l := 0.0001$ in each case; A9a and A9b, where we set $l := 0$ in each case; and Heu, where we set $l := 0.1$ for the components of the variables $x^A$ and $x^B$ (see [15, Section 1.4] for a description of the problem and an explanation of the variables). The $\lambda$- and $u$-components of $z^0$ were generated according to the rule used in [4]–[6]. Specifically, all components of $\lambda^0$ were set equal to 10, while the components $u_i^0$ of $u^0$ were set equal to $\max\{10,\, 5 - g_i(x^0)\}$, $i = 1, \ldots, m_g$.

We next discuss the most important conclusions of our numerical tests. In Figure 1, we present the results in the form of performance profiles [3], demonstrating both relative robustness and efficiency. Our use of performance profiles is basically standard here, i.e., they are as described in [3]. It should only be commented that, for each of the 35 test problems, we make runs from 20 different starting points, and the profiles in Figure 1 are based on all all those runs (i.e., a test example here is a
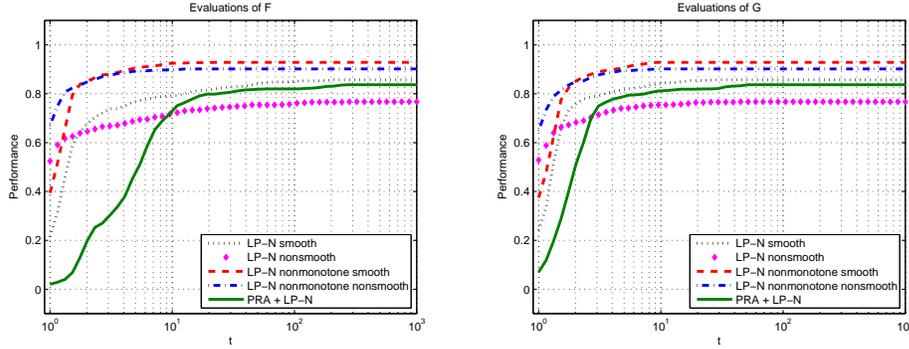
FIG. 1. *Performance profiles. All runs which did not terminate at an approximate solution are counted as failures.*

combination of a problem and a starting point, resulting in 20 instances out of each problem). On the left is the performance profile for the number of evaluations of $F$, and on the right for the number of evaluations of the matrix-valued function $G$. The numbers of these evaluations are considered here as a measure of efficiency of the algorithms. Note also that the number of evaluations of $G$ is also the number of iterations for all the variants of Algorithm 1. For the hybrid method, this number might be slightly smaller than the iteration count, because $G$ is not evaluated again on those iterations where the method makes a switch from the LP-Newton method back to the potential reduction.

The performance profiles in Figure 1 show that the LP-N smooth method (with monotone linesearch) is already at least at the same level of robustness as the hybrid method PRA + LP-N. Furthermore, both LP-N smooth and especially LP-N nonsmooth methods seriously outperform PRA + LP-N in respect to the specified measure of efficiency. Moreover, the variants of Algorithm 1 with the nonmonotone linesearch technique improve further – they are superior to the other methods both in efficiency and robustness. In particular, in terms of the specified measure of efficiency, they are quite significantly better than the hybrid method using potential reduction, which is the only globalization of the LP-Newton scheme previously available.

The bar diagram in Figure 2 shows, for each method and each problem, for how many starting points (out of 20 used) the generated sequence converged to a solution. Roughly speaking, this is a measure of robustness, but reported here for each problem separately (unlike the summary information in the performance profiles).

To get an additional insight, after termination of each run we checked whether the last computed iterate $z^k$ satisfies (3.10) (for the smooth reformulation), or (4.9) with $p \in \mathcal{A}(z^k)$ such that $G(z^k) = (F^p)'(z^k)$ (for the nonsmooth reformulation).

We next describe how condition (3.10) was verified ((4.9) was verified similarly). We used the tolerance of $10^{-8}$ to decide whether some component $F_i(z^k)$ was equal to $\|F(z^k)\|$ or $-\|F(z^k)\|$, or whether some constraint $z_j \geq 0$ was active at $z^k$. Specifically, we define the index sets

$$\tilde{I}_+(z^k) := \{i \mid F_i(z^k) \geq \|F(z^k)\| - 10^{-8}\},$$
$$\tilde{I}_-(z^k) := \{i \mid F_i(z^k) \leq -\|F(z^k)\| + 10^{-8}\},$$
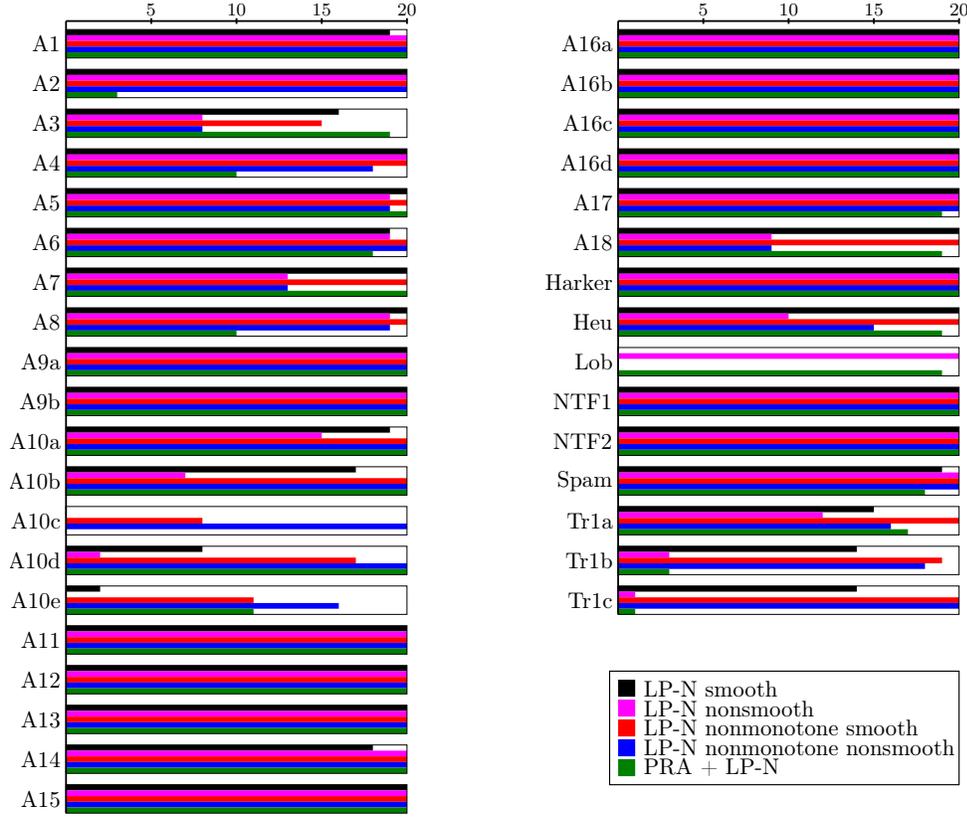$$\tilde{A}(z^k) := \{j \in \{n_x + 1, \ldots, n\} \mid z_j^k \leq 10^{-8}\},$$

Fig. 2. *Number of runs which led to a solution.*

and then solve the following LP problem:

$$
\begin{aligned}
\text{minimize} \quad & \varepsilon \\
\text{subject to} \quad & \sum_{i \in \tilde{I}_+(z^k)} w_i^+ F_i'(z^k) - \sum_{i \in \tilde{I}_-(z^k)} w_i^- F_i'(z^k) - \sum_{j \in \tilde{A}(z^k)} \omega_j e^j \leq \varepsilon e, \\
& \sum_{i \in \tilde{I}_+(z^k)} w_i^+ F_i'(z^k) - \sum_{i \in \tilde{I}_-(z^k)} w_i^- F_i'(z^k) - \sum_{j \in \tilde{A}(z^k)} \omega_j e^j \geq -\varepsilon e, \\
& \sum_{i \in \tilde{I}_+(z^k)} w_i^+ + \sum_{i \in \tilde{I}_-(z^k)} w_i^- = 1, \\
& w^+ \geq 0, \; w^- \geq 0, \; \omega \geq 0,
\end{aligned}
$$

(5.10)

with respect to $(\varepsilon, w^+, w^-, \omega)$, where $e^j$ denotes the $j$-th vector of the canonical basis in $\mathbb{R}^n$, and $e$ is the vector of ones in $\mathbb{R}^n$. According to the argument in the proof of Lemma 3.2 and the comments following this proof, if the optimal value of this LP is zero, then (3.10) holds.

Figure 3 contains the same kind of performance profiles as in Figure 1, but with the following difference: if the optimal value of LP (5.10), computed after termination of Algorithm 1, was not greater than $10^{-8}$, then the run was also counted as successful. Note that in Lemma 3.2 it has been proved that the optimal value of (5.10) is equal to zero if and only if $\Delta(z^k) = 0$ holds. Let us briefly explain why we did not simply count the runs being terminated because of $|\Delta(z^k)| \leq 10^{-12}$, which was one of our
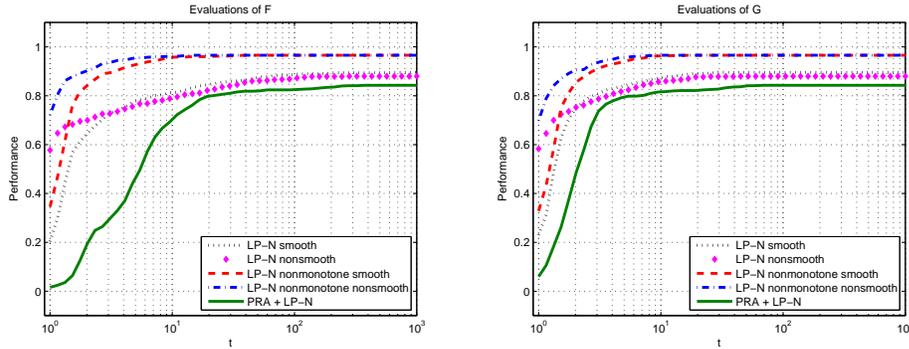
FIG. 3. *Performance profiles with runs for which the optimal values of* (5.10) *were not greater than* $10^{-8}$ *counted as successful.*

stopping criteria. There were a few runs terminated because of stepsize becoming too small, where the optimal value of (5.10) computed after termination was equal to zero (up to round-off errors), although $|\Delta(z^k)|$ was still greater than $10^{-12}$. Moreover, there is no stopping criterion similar to $|\Delta(z^k)| \leq 10^{-12}$ in the hybrid method. At the same time, some runs of the hybrid method terminated at points for which the optimal value of (5.10) was less than $10^{-8}$.

Naturally, robustness of all algorithms in Figure 3 is higher than in Figure 1. The difference is particularly apparent for the methods using the nonsmooth reformulation (5.7), i.e., LP-N nonsmooth and LP-N nonmonotone nonsmooth. Therefore, the variants of the algorithm employing nonsmooth reformulation, while being faster than those employing smooth reformulation, are more attracted by stationary points of problem (1.3) which are not solutions of (1.1). However, the overall picture in Figure 3 is similar to that in Figure 1, and thus, the conclusions concerning relative performance of the methods in question remain the same. For more detailed results on all the test problems we refer to [14], where also computer times are provided. Regarding the latter, we note that for the three largest problems in our test library, the hybrid method PRA + LP-N is clearly faster than our current implementation of the other algorithms being tested. This is due to the fact that the global phase of the hybrid method requires solving linear systems of equations, which is generally faster than solving LP problems. More experience on larger problem libraries with a variety of large problems might be needed for reliable conclusions regarding computer time. Moreover, changing some default settings of CPLEX (e.g., forcing it to always use the dual simplex method as an LP-solver [14]) may seriously influence the time record for our algorithms, especially on large problems.

Finally, we mention that for an absolute majority of successful runs of our algorithms, the unit stepsize has been eventually accepted, and superlinear convergence rate has been detected. The only exceptions seem to be some runs for the test problems A10c and Spam, where the unit stepsize is eventually accepted, but the convergence rate looks more like linear. This can be due to the lack of error bound near some solutions of A10c, and due to round-off errors, especially for Spam, which is a quite large problem.

## REFERENCES

[1] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[2] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983.

[3] E. D. DOLAN AND J. J. MORE, *Benchmarking optimization software with performance profiles*, Math. Program., 91 (2002), pp. 201–213.

[4] A. DREVES, *Improved error bound and a hybrid method for generalized Nash equilibrium problems*, Comput. Optim. Appl., DOI 10.1007/s10589-014-9699-z, 2014.

[5] A. DREVES, F. FACCHINEI, A. FISCHER, AND M. HERRICH, *A new error bound result for Generalized Nash Equilibrium Problems and its algorithmic application*, Comput. Optim. Appl., 59 (2014), pp. 63–84.

[6] A. DREVES, F. FACCHINEI, C. KANZOW, AND S. SAGRATELLA, *On the solution of the KKT conditions of generalized Nash equilibrium problems*, SIAM J. Optim., 21 (2011), pp. 1082–1108.

[7] F. FACCHINEI, A. FISCHER, AND M. HERRICH, *A family of Newton methods for nonsmooth constrained systems with nonisolated solutions*, Math. Methods Oper. Res., 77 (2013), pp. 433–443.

[8] F. FACCHINEI, A. FISCHER, AND M. HERRICH, *An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions*, Math. Program., 146 (2014), pp. 1–36.

[9] F. FACCHINEI, A. FISCHER, AND V. PICCIALLI, *Generalized Nash equilibrium problems and Newton methods*, Math. Program., 117 (2009), pp. 163–194.

[10] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[11] D. FERNÁNDEZ AND M. SOLODOV, *Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems*, Math. Program., 125 (2010), pp. 47–73.

[12] A. FISCHER, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Math. Oper. Res., 24 (1999), pp. 699–727.

[13] A. FISCHER, M. HERRICH, A. F. IZMAILOV, AND M. V. SOLODOV, *Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions*, Comput. Optim. Appl., 63 (2016), pp. 425–459.

[14] A. FISCHER, M. HERRICH, A. F. IZMAILOV, AND M. V. SOLODOV, *Detailed numerical results for several variants of a globally convergent LP-Newton method*, http://www.math.tu-dresden. de/~fischer/research/publications/TR-DNR-2016.pdf, 2016.

[15] A. VON HEUSINGER, *Numerical Methods for the Solution of the Generalized Nash Equilibrium Problem* Ph.D. thesis, Institute of Mathematics, University of Würzburg, Würzburg, 2009.

[16] A. F. IZMAILOV AND M. V. SOLODOV, *Karush–Kuhn–Tucker systems: regularity conditions, error bounds and a class of Newton-type methods*, Math. Program., 95 (2003), pp. 631–650.

[17] A. F. IZMAILOV AND M. V. SOLODOV, *Newton-type methods for optimization problems without constraint qualifications*, SIAM J. Optim., 15 (2004), pp. 210–228.

[18] A. F. IZMAILOV AND M. V. SOLODOV, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, Springer International Publishing Switzerland, Cham, 2014.

[19] A. F. IZMAILOV AND M. V. SOLODOV, *On error bounds and Newton-type methods for generalized Nash equilibrium problems*, Comput. Optim. Appl., 59 (2014), pp. 201–218.

[20] D. KLATTE AND B. KUMMER, *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications*, Kluwer, Dordrecht, 2002.

[21] R. D. C. MONTEIRO AND J.-S. PANG, *A potential reduction Newton method for constrained equations*, SIAM J. Optim., 9 (1999), pp. 729–754.

[22] S. J. WRIGHT, *An algorithm for degenerate nonlinear programming with rapid local convergence*, SIAM J. Optim., 15 (2005), pp. 673–696.