Minimax Theory and its Applications
Volume 9 (2024), No. 1, 55–84

Linear and Superlinear Convergence of a Potential Reduction Algorithm for Generalized Nash Equilibrium Problems

Axel Dreves
Dept. of Aerospace Engineering, Inst. for Applied Mathematics and Scientific Computing
Universität der Bundeswehr München, Neubiberg, Germany
axel.dreves@unibw.de

Received: February 18, 2022
Accepted: October 10, 2022

We present a detailed convergence analysis of the potential reduction algorithm for generalized Nash equilibrium problems (GNEPs), that is known to be a robust method for solving those problems. We prove Q-linear convergence of the merit function and R-linear convergence of the distance of the iterates to the set of KKT-points of the GNEP. Furthermore, we show that the stepsize is bounded from below, implying finite termination of the method for prescribed accuracy. Using a non-fixed linesearch parameter we prove superlinear convergence. Further, we give additional assumptions to transfer the convergence rates to an inexact potential reduction algorithm. By our analysis, we also discover indicators that could be used to estimate the active set at an accumulation point, and hence at a generalized Nash equilibrium, which might be exploited numerically.

Keywords: Generalized Nash equilibrium problem, potential reduction algorithm, linear convergence, superlinear convergence.

2010 Mathematics Subject Classification: 90C30, 90C51, 91A10, 65B99.

1. Introduction

In the last years, several algorithms for the solution of generalized Nash equilibrium problems (GNEPs) have been proposed. In [9, 15] one can find surveys for definitions and examples of GNEPs together with general numerical approaches. For Newton-type methods, see [5, 13], for penalty approaches [8, 16], for a multiplier-penalty method [18], for a variational inequality approach [10, 20], to mention just a few different numerical approaches.

The potential reduction algorithm for GNEPs, which is the content of this paper, has proven to be a robust method regarding global convergence to a KKT-point of the GNEP, as shown in [4]. The convergence speed will be discussed in this paper. In [1] a finite termination for a prescribed tolerance was proven, and it is based on showing boundedness of the step-size from below. Here, we will sharpen this result and further prove the linear rate of convergence. The algorithm in its original form has in general no faster local convergence. However, we will see that a sequence of linesearch parameters instead of a fixed one can be used to obtain superlinear convergence.

ISSN 2199-1413 (printed), ISSN 2199-1421 (electronic) / $ 2.50

Other methods that have a fast local convergence rate have been proposed, and in particular the LP-Newton method from [11, 12] is a promising approach. It was used to design a hybrid method with the potential reduction algorithm in [2, 3]. Further, a suitable globalization technique was developed in [14], and the method was successfully tested on GNEPs. The cost of each iteration is mainly the solution of one linear program, that has to be compared to the cheaper linear equation system in the potential reduction algorithm.

This paper is organized as follows. In Section 2, we recall the definition of a GNEP, and restate the potential reduction algorithm with its known convergence result. In Section 3, we present the notation for the analysis of the algorithm and the assumptions we will require. Section 4 contains the convergence analysis. The main result is the linear rate of convergence. In Section 5 we show, how a non-fixed linesearch parameter can be used to obtain superlinear convergence. In Section 6, the obtained convergence rates are extended to an inexact potential reduction algorithm for GNEPs. In Section 7 we present one numerical example that illustrates the linear and superlinear rates. Finally, we conclude the paper in Section 8.

2. Potential reduction algorithm for GNEPs

Here, we consider the exact version of the potential reduction algorithm for GNEPs, whose inexact version was introduced in [4]. The algorithm solves the KKT-system of a GNEP through an interior point algorithm, where the step-size is defined by an Armijo-rule for a potential function. We will use the notation from [1] and for convenience we recall all the basic definitions here.

We consider a GNEP with players $\nu = 1, \ldots, N$, where the ν-th player solves, for given strategies $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$ of all other players, the problem

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \quad \theta_\nu(x^\nu, x^{-\nu}) \qquad \text{s.t.} \qquad g^\nu(x^\nu, x^{-\nu}) \le 0,$$

with twice continuously differentiable functions $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$ and $g^\nu : \mathbb{R}^n \to \mathbb{R}^{m_\nu}$, whose second derivatives are locally Lipschitz continuous. A point $\bar{x} = (\bar{x}^1, \ldots, \bar{x}^N) \in \mathbb{R}^n$ with $g^\nu(\bar{x}) \le 0$ for all $\nu = 1, \ldots, N$ is called a *generalized Nash equilibrium*, if

$$\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \le \theta_\nu(x^\nu, \bar{x}^{-\nu}) \qquad \forall x^\nu \in \{x^\nu \in \mathbb{R}^{n_\nu} \mid g^\nu(x^\nu, \bar{x}^{-\nu}) \le 0\}$$

for all $\nu = 1, \ldots, N$. Defining

$$g(x) := \begin{pmatrix} g^1(x^1, x^{-1}) \\ \vdots \\ g^N(x^N, x^{-N}) \end{pmatrix}, \quad \lambda := \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^N \end{pmatrix}, \quad w := \begin{pmatrix} w^1 \\ \vdots \\ w^N \end{pmatrix},$$

which are vectors of dimension $m := m_1 + \ldots + m_N$, and

$$F(x, \lambda) := \begin{pmatrix} \nabla_{x^1}\theta_1(x^1, x^{-1}) + \sum_{i=1}^{m_1} \lambda_i^1 \nabla_{x^1} g_i^1(x^1, x^{-1}) \\ \vdots \\ \nabla_{x^N}\theta_N(x^N, x^{-N}) + \sum_{i=1}^{m_N} \lambda_i^N \nabla_{x^N} g_i^N(x^N, x^{-N}) \end{pmatrix} \in \mathbb{R}^n,$$

we obtain all KKT points of the GNEP as solutions of the constrained equation

$$0 = H(z) := H(x, \lambda, w) := \begin{pmatrix} F(x, \lambda) \\ g(x) + w \\ \lambda \circ w \end{pmatrix} \quad \text{s.t.} \quad z \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^m_+, \tag{1}$$

where $\lambda \circ w$ is the Hadamard product. Let

$$S := \{z \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^m_+ \mid H(z) = 0\}$$

denote the set of all KKT-points.

We want to discuss a potential reduction algorithm, originally introduced in [21], and further analyzed under more general convergence conditions in [19]. An inexact version of this algorithm was adapted to GNEPs in [4]. Here, we consider the exact version, which is also used in [1, 3], and we recall some definitions from there. Define the set

$$Z_I := \{z = (x, \lambda, w) \in \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^m_{++} \mid g(x) + w > 0\}$$

and for a real number $\zeta > m$ the potential function $\psi : Z_I \to \mathbb{R}$,

$$\psi(z) := \zeta \ln(\|H(z)\|^2) - \sum_{i=1}^{2m} \ln(H_{n+i}(z)).$$

Further, set
$$a := (0_n^\top, 1_m^\top, 1_m^\top)^\top \in \mathbb{R}^{n+2m},$$

and for any $z^k \in Z_I$
$$\mu_k := \frac{a^\top H(z^k)}{\|a\|^2} = \frac{1}{2m} \sum_{i=1}^{2m} H_{n+i}(z^k),$$

where H_{n+i} denotes the $(n+i)$-th component of the vector H. Now, we restate the exact version of the potential reduction algorithm, c.f., [3], with a slight change in allowing non fixed constants $\eta_k \in (\underline{\eta}, 1)$ with $\underline{\eta} \in (0, 1)$ for the linesearch.

Algorithm 1: Potential Reduction Algorithm for GNEPs

`(S.0)`: Choose $z^0 \in Z_I$ and $\gamma \in (0, 1), \underline{\eta} \in (0, 1), \zeta \in (m, 2m)$. Set $k := 0$.

`(S.1)`: If $H(z^k) = 0$ then STOP.

`(S.2)`: Choose $\sigma_k \in (0, 1)$ and compute a solution d^k of the linear system

$$JH(z^k)d = -H(z^k) + \sigma_k \mu_k \, a. \tag{2}$$

`(S.3)`: Choose $\eta_k \in (\underline{\eta}, 1)$ and compute a stepsize $t_k := \max\{{\eta_k}^\ell \mid \ell = 0, 1, 2, \ldots\}$ such that $z^k + t_k d^k \in Z_I$ and

$$\psi(z^k + t_k d^k) \leq \psi(z^k) + \gamma t_k \nabla \psi(z^k)^\top d^k.$$

`(S.4)`: Set $z^{k+1} := z^k + t_k d^k$, $k := k + 1$, and go to `(S.1)`.

As a direct consequence of [4, Theorem 4.3], one has the following convergence theorem for fixed $\eta_k = \eta \in (0, 1)$ for all $k \in \mathbb{N}$, which was stated in [3, Theorem 2]. Allowing a sequence $\{\eta_k\}$ in the step-size condition instead of a fixed value η we obtain the same result, slightly adapting the proof. Rather than presenting the (lengthly) entire proof, we focus on the new part, regarding the stepsize condition.

Theorem 2.1. *Assume that $JH(z)$ is nonsingular for all $z \in Z_I$, and that the sequence $\{\sigma_k\}$ satisfies $\limsup_{k\to\infty} \sigma_k < 1$. Let $\{z^k\}_{k\in\mathbb{N}}$ be any sequence generated by Algorithm 1. Then the following assertions hold:*

(a) *The sequence $\{H(z^k)\}_{k\in\mathbb{N}}$ is bounded.*

(b) *Any accumulation point of $\{z^k\}$ is a solution of (1).*

Proof. We follow the proof of [4, Theorem 4.3]. The proof of part (a) holds without changes. For part (b) we assume that z^∞ is an accumulation point of the sequence $\{z^k\}$, and that $\{z^k\}_K$ is a corresponding subsequence converging to z^∞. For contradiction, we assume $H(z^\infty) \neq 0$. Then, we can follow the proof to obtain that $z^\infty \in Z_I$, $\lim_{k\in K} d^k = d^\infty$ for some vector $d^\infty \in \mathbb{R}^{n+2m}$, $\nabla\psi(z^\infty)^\top d^\infty < 0$ and that $\{\psi(z^k)\}_{k\in K}$ also converges.

Now, the new step-size rule has to be used, instead of the old one with fixed η. Since by definition $\{\psi(z^k)\}_{k\in\mathbb{N}}$ is monotonically decreasing, the whole sequence $\{\psi(z^k)\}_{k\in\mathbb{N}}$ converges. Further, we have

$$\psi(z^{k+1}) - \psi(z^k) \leq \gamma t_k \nabla\psi(z^k)^T d^k < 0$$

for all $k \in \mathbb{N}$. Since the left-hand side converges to zero, $\lim_{k\to\infty} t_k \nabla\psi(z^k)^T d^k = 0$ must hold. This, in turn, implies $\lim_{k\in K} t_k = 0$ since

$$\lim_{k\in K} \nabla\psi(z^k)^T d^k = \nabla\psi(z^\infty)^T d^\infty < 0.$$

Let $\ell_k \in \mathbb{N}_0$ be the unique index such that $t_k = \eta_k^{\ell_k}$ holds for all $k \in \mathbb{N}$. With $\lim_{k\in K} t_k = 0$, we must have $\ell_k > 0$ for all $k \in \mathbb{N}$ sufficiently large, and we also have $\lim_{k\in K} \frac{t_k}{\eta_k} = 0$, since η_k is bounded away from zero. Since the limit point z^∞ belongs to the open set Z_I, it therefore follows that the sequence $\{z^k + \frac{t_k}{\eta_k} d^k\}_{k\in K}$ also belongs to this set, at least for all sufficiently large $k \in K$. Consequently, for these $k \in K$, the line search test in (S.3) fails for the stepsize $\frac{t_k}{\eta_k} = \eta_k^{\ell_k - 1}$. Thus

$$\frac{\psi(z^k + \eta_k^{\ell_k-1} d^k) - \psi(z^k)}{\eta_k^{\ell_k-1}} > \gamma \nabla\psi(z^k)^T d^k$$

holds for all $k \in K$ sufficiently large. Taking the limit $k \to \infty$ on the subset K, the continuous differentiability of the potential function ψ on the set Z_I gives

$$\nabla\psi(z^\infty)^T d^\infty \geq \gamma \nabla\psi(z^\infty)^T d^\infty.$$

Since $\nabla\psi(z^\infty)^T d^\infty < 0$ and $\gamma \in (0,1)$, this is a contradiction. Consequently, we have $0 = H(z^\infty)$, and hence z^∞ is a solution of (1). □

3. Notations and assumptions

For our discussions on the convergence rate, we recall some definitions.

Definition 3.1. A sequence $\{z^k\}$ converges *Q-linearly* to $\bar{z}$ if there is some constant $c \in (0,1)$ such that for all $k \in \mathbb{N}$ sufficiently large we have

$$\|z^{k+1} - \bar{z}\| \leq c\,\|z^k - \bar{z}\|.$$

A sequence $\{z^k\}$ converges *R-linearly* to $\bar{z}$, if there is some sequence $\{\varepsilon_k\} \subset \mathbb{R}_{++}$ that converges Q-linearly to 0 and $\|z^k - \bar{z}\| \le \varepsilon_k$ for all $k \in \mathbb{N}$ sufficiently large. For two sequences $\{\alpha_k\}, \{\tilde{\alpha}_k\} \subset \mathbb{R}_{++}$ we use the notation

$$\begin{array}{lll} \alpha_k = o(\tilde{\alpha}_k) & \text{if} & \lim\limits_{k\to\infty} \dfrac{\alpha_k}{\tilde{\alpha}_k} = 0, \\ \alpha_k = \mathcal{O}(\tilde{\alpha}_k) & \text{if} & \dfrac{\alpha_k}{\tilde{\alpha}_k} \text{ is bounded.} \end{array} \qquad \Box$$

Further, let us recall some notation from [1]. For a set $\alpha \subseteq \{1, \dots, m\}$, define

$$\bar{\alpha} := \{1, \dots, m\} \setminus \alpha.$$

For $\bar{x} \in \mathbb{R}^n$ with $g(\bar{x}) \le 0$, define the set of active and inactive constraints by

$$I_=(\bar{x}) := \{j \mid g_j(\bar{x}) = 0\} \quad \text{and} \quad I_<(\bar{x}) := \{j \mid g_j(\bar{x}) < 0\},$$

respectively. Since in a GNEP several players may have the same constraint, let

$$p(\bar{x}) := \max\{|\alpha| \mid \alpha \subseteq I_=(\bar{x}), \quad g_i \not\equiv g_j \quad \forall i, j \in \alpha\}$$

denote the number of different active constraints at $\bar{x}$. Depending on $p(\bar{x})$ and $|I_=(\bar{x})|$ define the following sets of index sets:

$$\begin{array}{rcl} I_=^0(\bar{x}) & := & \left\{\alpha \subseteq I_=(\bar{x}) \big| |\alpha| = p(\bar{x}), \quad g_i \not\equiv g_j \quad \forall i, j \in \alpha\right\}; \\ I_=^{-k}(\bar{x}) & := & \left\{\alpha \subseteq I_=(\bar{x}) \big| \exists \beta_\alpha \in I_=^0(\bar{x})\, \exists i_1, \dots, i_k \in \beta_\alpha : \alpha = \beta_\alpha \setminus \{i_1, \dots, i_k\}\right\}, \\ & & \text{for } 1 \le k \le p(\bar{x}); \\ I_=^1(\bar{x}) & := & \left\{\alpha \subseteq I_=(\bar{x}) \big| \exists \beta_\alpha \in I_=^0(\bar{x})\, \exists i, j \in \alpha, i \ne j : g_i \equiv g_j,\, \alpha \setminus \{i\} \subseteq \beta_\alpha\right\}, \\ & & \text{if } |I_=(\bar{x})| > p(\bar{x}); \\ I_=^k(\bar{x}) & := & \left\{\alpha \subseteq I_=(\bar{x}) \big| \alpha \notin I_=^{k-1}(\bar{x}),\, \exists \beta_\alpha \in I_=^{k-1}(\bar{x})\, \exists i \in \alpha : \alpha \setminus \{i\} = \beta_\alpha\right\}, \\ & & \text{for } 2 \le k \le |I_=(\bar{x})| - p(\bar{x}). \end{array}$$

For more explanations on the definition of these sets, we refer to [1]. With these definitions, we have a disjoint partition of the power set $\mathcal{P}(I_=(\bar{x}))$ of $I_=(\bar{x})$, namely

$$\mathcal{P}(I_=(\bar{x})) = \bigcup_{-p(\bar{x}) \le k \le |I_=(\bar{x})| - p(\bar{x})} I_=^k(\bar{x}).$$

The assumptions we use for our convergence analysis are the ones from [1, Assumption 5.26, Assumption 5.33]. We have to exploit the structure of the matrix

$$JH(z) = \begin{pmatrix} J_x F(x, \lambda) & E(x) & 0 \\ J_x g(x) & 0 & I \\ 0 & W & \Lambda \end{pmatrix}$$

with $E(x) := \text{blockdiag}(\nabla_{x^1} g^1(x), \dots, \nabla_{x^N} g^N(x))$, $\Lambda := \text{diag}(\lambda)$ and $W := \text{diag}(w)$.

The matrix $\qquad M(x, \lambda) := J_x g(x)\, J_x F(x, \lambda)^{-1} E(x)$

will play a crucial role. We denote by $M_{\beta\beta}$ the submatrix of M that contains all columns and rows with indices in the set β.

Assumption 3.2. (A1) The sequence $\{z^k\}$ is bounded.

(A2) If there are identical constraints $g_i \equiv g_j$ with $i, j \in \{1, \dots, m\}, i \neq j$, then we have $w_i^0 = w_j^0$ for the starting vector w^0.

(A3) There is a constant $c > 0$ such that for all $k \in \mathbb{N}_0$

$$\frac{\sigma_k \mu_k}{\min\limits_{j=1,\dots,2m} H_{n+j}(z^k)} \leq c \|H(z^k)\|.$$

(A4) The matrix $J_x F(x, \lambda)$ is nonsingular for all $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+$.

(A5) At every solution $\bar{z} \in S$ there is a $\beta \in I^0_=(\bar{x})$, such that $M_{\beta\beta}(\bar{x}, \bar{\lambda})$ is nonsingular and $\bar{\lambda}_j > 0$ for all $j \in \beta$. Further, $\det(M_{\alpha\alpha}(\bar{x}, \bar{\lambda}))$ is zero or has the same sign for all $\alpha \in I^0_=(\bar{x})$.

(A6) The matrix $JH(z)$ is nonsingular for all $z \in Z_I$.

Remark 3.3. Let us discuss our assumptions:

(A1) This assumption guarantees the existence of an accumulation point and hence a solution under the conditions of Theorem 2.1. In [4, Theorem 4.10] sufficient conditions for (A1) were given, namely (A6), the coercitvity assumption

$$\lim_{\|x\| \to \infty} \| \max\{0, g(x)\} \| = +\infty,$$

and the extended Mangasarian-Fromowitz condition (EMFCQ) for each player, i.e., for all $\nu = 1, \dots, N$ and all $x \in \mathbb{R}^n$ there exists a $d^\nu \in \mathbb{R}^{n_\nu}$ such that

$$\nabla_{x^\nu} g_i^\nu(x)^\top d^\nu < 0 \quad \forall i \in \{j \in \{1, \dots, m_\nu\} \mid g_j^\nu(x) \geq 0\}.$$

Coercivity is a weak assumption, which for example holds for GNEPs with linear constraints and a compact feasible set. It can also be enforced by introducing additional box constraints. Coercivity is used to guarantee boundedness of the x and w part of the iterate $z^k = (x^k, \lambda^k, w^k)$ in the proof of [4, Theorem 4.10]. The EMFCQ is a standard constraint qualification in optimization, and here it is used to show boundedness of the multipliers for the active constraints. For the inactive ones, this follows without further assumptions. EMFCQ only depends on the constraints of each player. We will see in Corollary 4.3 that for repeated constraints, we can use equal starting values in the algorithm to obtain equal multipliers for repeated constraints at all iterates. Hence, if one of these multipliers stays bounded, this also holds for the others corresponding to repeated constraints. This allows us to even relax the EMFCQ condition for GNEPs with repeated constraints, namely the descent condition for a repeated active or violated constraint must only hold for one player and not for all.

(A2) This is a very mild assumption on the starting vector w^0 that can always be satisfied.

(A3) Here, we require that the sequence $\{\sigma_k\}$ vanishes fast enough, and in particular this implies the assumption

$$\limsup_{k \to \infty} \sigma_k < 1$$

from Theorem 2.1. (A3) can always be satisfied by an appropriate choice of the parameter sequence $\{\sigma_k\}$.

(A4) The nonsingularity of $J_xF(x,\lambda)$ is a commonly used condition, and it is also one part of the sufficient conditions for nonsingularity of $JH(z^k)$ given in [4, Theorem 4.6]. (A4) is an assumption on the functions defining the GNEP, and it can be easily verified for GNEPs with quadratic cost and affine linear constraint functions (LQGNEPs), because then we have only a single matrix independent of x and λ. However, it restricts the number of GNEPs our theory can deal with. For example LGNEPs, where cost and constraint functions are all affine linear, violate (A4) since $J_xF(x,\lambda) = 0$. For the potential reduction algorithm in the context of LGNEPs, we refer to [7].

(A5) The first part of (A5) is a strong condition that was also used in slightly different variants in [2, 3, 17] in order to prove a local error bound for GNEPs. For GNEPs with shared constraints it is typical to have a singular matrix $JH(\bar{z})$ at a solution $\bar{z}$. Therefore, we cannot assume its nonsingularity at the solution. By dropping rows from repeated constraints except one, and the corresponding columns, we may obtain nonsingularity of the remaining sub-matrix $\begin{pmatrix} J_xF(x,\lambda) & E_\beta(x) \\ J_xg_\beta(x) & 0 \end{pmatrix}$. This, together with (A4), is by definition of $M_{\beta\beta}(x,\lambda) = J_xg_\beta(x)J_xF(x,\lambda)^{-1}E_\beta(x)$ the nonsingularity assumption in (A5).
Further, the requirement of positive multipliers is a relaxation of the strict complementarity condition. For repeated constraints we only require that one player has a positive multiplier and not all the players. This condition, however, is usually hard to be verified without knowing the solutions and the corresponding multipliers.
The second part of (A5) is satisfied, if the sufficient conditions for the nonsingularity of $JH(z)$ from [4, Theorem 4.6] are satisfied. Those are (A4) and that $M(x,\lambda)$ is a P_0-matrix, i.e., $\det M_{\alpha\alpha}(x,\lambda) \geq 0$ for all $\alpha \subseteq \{1,\ldots,m\}$. Note, that all conditions on $M(x,\lambda)$ can be verified for LQGNEPs, since we only have a single matrix in that case. Further, (A5) is actually only required at those solutions that are accumulation points of the algorithm. Thus, if we use equal starting values for repeated constraints, we will have, in view of Corollary 4.3, only to consider solutions with equal multipliers.

(A6) As already mentioned, sufficient conditions for the nonsingularity of $JH(z)$ are (A4) and the P_0 property of $M(x,\lambda)$, as shown in [4, Theorem 4.6]. In [6] it was (in the context of quasi-variational inequalities) shown that for GNEPs with linear constraints, where (A4) holds, the P_0-property is also necessary for (A6). The P_0 property can be verified in the context of LQGNEPs, but it might be computationally expensive.
(A6) is, at least at all iterates z^k, necessary for the direction d^k to be uniquely defined in our algorithm. Furthermore, inspecting the contradiction-based proof of [4, Theorem 4.3], it is necessary at any accumulation point of the sequence that lies in the open set Z_I (but not at its boundary, as the solutions do). Hence, it would be enough that $JH(z)$ is nonsingular for all $z \in Z_I$ that lie in the neighborhood of the sequence $\{z^k\}$. □

Now, let us have a closer look on the equation system (2).

Partitioning the vector $d^k := \left((d_x^k)^\top, (d_\lambda^k)^\top, (d_w^k)^\top\right)^\top \in \mathbb{R}^{n+m+m}$, equation (2) becomes

$$J_x F(x^k,\lambda^k) d_x^k + E(x^k) d_\lambda^k = -F(x^k,\lambda^k), \tag{3}$$

$$J_x g(x^k) d_x^k + d_w^k = -g(x^k) - w^k + \sigma_k \mu_k 1_m, \tag{4}$$

$$W^k d_\lambda^k + \Lambda^k d_w^k = -\Lambda^k W^k 1_m + \sigma_k \mu_k 1_m. \tag{5}$$

As shown in [1, Equation 5.33] under assumption (A4), this yields

$$\left(M^k \Lambda^k + W^k\right) \frac{d_\lambda^k}{\lambda^k} = r^k, \tag{6}$$

with $M^k := J_x g(x^k)\, J_x F(x^k,\lambda^k)^{-1}\, E(x^k),$

$r^k := -J_x g(x^k)\, J_x F(x^k,\lambda^k)^{-1} F(x^k,\lambda^k) + g(x^k) + \sigma_k \mu_k ((\Lambda^k)^{-1} - I_m) 1_m.$

Using Cramer's rule we obtain for each component $i \in \{1,\dots,m\}$

$$\frac{d_{\lambda,i}^k}{\lambda_i^k} = \frac{\det(\tilde{M}_i^k \tilde{\Lambda}_i^k + \tilde{W}_i^k)}{\det(M^k \Lambda^k + W^k)}, \tag{7}$$

where we used $\tilde{\Lambda}_i^k := \operatorname{diag}\left(\lambda_1^k,\dots,\lambda_{i-1}^k, 1, \lambda_{i+1}^k, \dots \lambda_m^k\right),$

$\tilde{W}_i^k := \operatorname{diag}\left(w_1^k,\dots,w_{i-1}^k, 0, w_{i+1}^k, \dots w_m^k\right),$

$\tilde{M}_i^k := \left(M_1^k,\dots,M_{i-1}^k, r^k, M_{i+1}^k, \dots M_m^k\right).$

The behavior of these fractions is the crucial part in the upcoming analysis.

4. Convergence analysis

Let us first formally state that our assumptions imply convergence of $\{H(z^k)\}_{k\in\mathbb{N}}$ to zero.

Lemma 4.1. *Assume* (A1), (A3) *and* (A6). *Then, the entire sequence* $\{H(z^k)\}_{k\in\mathbb{N}}$ *converges to zero.*

Proof. The proof is by contradiction. Assume, that $\{H(z^k)\}$ does not converge to zero. Then, we can find a subsequence of $\{z^k\}$, where $\{\|H(z^k)\|\}$ is bounded away from zero. The boundedness assumption (A1) implies that this subsequence has an accumulation point $\bar{z}$. By (A3) and (A6) we can apply Theorem 2.1. Thus, $\bar{z}$ must be a solution satisfying $H(\bar{z}) = 0$, contradicting that $\{\|H(z^k)\|\}$ is bounded away from zero. □

Now, let us recall [1, Lemma 5.27] together with its simple proof.

Lemma 4.2. *Let* (A2) *be satisfied. Then, if there are indices* $i,j \in \{1,\dots,m\}$ *with* $g_i \equiv g_j$, *we have* $w_i^k = w_j^k$ *for all* $k \in \mathbb{N}_0$. *Furthermore,*

$$\left(\prod_{i\in\bar{\alpha}_1\setminus I_<(\bar{x})} w_i^k\right) = \left(\prod_{i\in\bar{\alpha}_2\setminus I_<(\bar{x})} w_i^k\right)$$

holds for all $\alpha_1, \alpha_2 \in I_=^0(\bar{x})$ *and all* $k \in \mathbb{N}_0$, *that is the product is independent of the choice of* $\alpha \in I_=^0(\bar{x})$.

Proof. If we have indices $i, j \in \{1, \dots, m\}$ with $g_i \equiv g_j$ and $w_i^k = w_j^k$ for some $k \in \mathbb{N}_0$, equation (4) yields

$$\begin{aligned} d_{w,i}^k &= -w_i^k - g_i(x^k) - \nabla g_i(x^k)^T d_x^k + \sigma_k \mu_k \\ &= -w_j^k - g_j(x^k) - \nabla g_j(x^k)^T d_x^k + \sigma_k \mu_k = d_{w,j}^k, \end{aligned}$$

and hence $$w_i^{k+1} = w_i^k + t_k d_{w,i}^k = w_j^k + t_k d_{w,j}^k = w_j^{k+1}.$$

With $w_i^0 = w_j^0$ by (A2) this implies $w_i^k = w_j^k$ for all $k \in \mathbb{N}_0$. Now, let $\alpha_1, \alpha_2 \in I_=^0(\bar{x})$ be given. By definition $\bar{\alpha}_1 \setminus I_<(\bar{x})$ and $\bar{\alpha}_2 \setminus I_<(\bar{x})$ contain only active constraints that are joined by several players, and they contain all but one of each joined active constraint. Thus, (A2) and the first assertion imply

$$\Big(\prod_{i \in \bar{\alpha}_1 \setminus I_<(\bar{x})} w_i^k \Big) = \Big(\prod_{i \in \bar{\alpha}_2 \setminus I_<(\bar{x})} w_i^k \Big). \qquad \square$$

This Lemma has a nice consequence if we further choose the starting values of the λ^0 variables equal.

Corollary 4.3. *Assume* (A2) *and that if there are identical constraints $g_i \equiv g_j$ with $i, j \in \{1, \dots, m\}, i \neq j$, then we have $\lambda_i^0 = \lambda_j^0$ for the starting vector λ^0. Then, we have $\lambda_i^k = \lambda_j^k$ for all $k \in \mathbb{N}_0$. In particular, any accumulation point of a sequence $\{z^k\}$ generated by Algorithm 1 has equal multipliers for repeated constraints.*

Proof. Let identical constraints $g_i \equiv g_j$ with $i, j \in \{1, \dots, m\}, i \neq j$ be given. Then, (A2) and Lemma 4.2 imply $w_i^k = w_j^k$ and also $d_{w,i}^k = d_{w,j}^k$ for all $k \in \mathbb{N}_0$. Assume $\lambda_i^k = \lambda_j^k$ for some $k \in \mathbb{N}_0$. Then, equation (5) yields

$$d_{\lambda,i}^k = \frac{-\lambda_i^k w_i^k + \sigma_k \mu_k - \lambda_i^k d_{w,i}^k}{w_i^k} = \frac{-\lambda_j^k w_j^k + \sigma_k \mu_k - \lambda_j^k d_{w,j}^k}{w_j^k} = d_{\lambda,j}^k.$$

With $\lambda_i^0 = \lambda_j^0$ this implies $\lambda_i^k = \lambda_j^k$ for all $k \in \mathbb{N}_0$. Hence, we have equal multipliers for identical constraints at every iteration, and this also transfers to any accumulation point. $\square$

Note, that by Corollary 4.3 we can use Algorithm 1 for GNEPs with shared constraints with equal starting values in w^0 and λ^0 for the repeated constraints in order to compute normalized Nash equilibria.

Now, we start with the technical analysis of the convergence speed. The following Lemma is closely related to [1, Lemma 5.29], but with an explicit computation of the term hidden in the constant there.

Lemma 4.4. *Let Assumption* 3.2 *hold. Let $\{(x^k, \lambda^k, w^k)\}$ be a subsequence, generated by Algorithm* 1*, that converges to $(\bar{x}, \bar{\lambda}, \bar{w})$. Then, with $\beta \in I_=^0(\bar{x})$ from* (A5)*, we obtain*

$$\begin{aligned} &\det(M^k \Lambda^k + W^k) \qquad (8) \\ &= \Big(\prod_{j \in \bar{\beta} \setminus I_<(\bar{x})} w_j^k \Big) \sum_{\alpha \in I_=^0(\bar{x})} \Big(\prod_{j \in I_<(\bar{x})} w_j^k \Big) \Big(\prod_{j \in \alpha} \lambda_j^k \Big) \det(M_{\alpha\alpha}^k) + o\Big(\prod_{j \in \bar{\beta} \setminus I_<(\bar{x})} w_j^k \Big). \end{aligned}$$

Proof. By (A6) Algorithm 1 is well defined, and by the boundedness assumption (A1) we have a convergent subsequence. Its limit $(\bar{x}, \bar{\lambda}, \bar{w})$ is by Theorem 2.1 a KKT point. By (A4) all matrices M^k are well defined.

Using $\det(M_{\emptyset\emptyset}) = 1$ and exploiting the diagonal structure of Λ^k and W^k, we get

$$\det(M^k\Lambda^k + W^k) = \sum_{\alpha\subseteq\{1,\dots,m\}} \det(W^k_{\bar{\alpha}\bar{\alpha}}) \det\left((M^k\Lambda^k)_{\alpha\alpha}\right)$$
$$= \sum_{\alpha\subseteq\{1,\dots,m\}} \det(W^k_{\bar{\alpha}\bar{\alpha}}) \det(M^k_{\alpha\alpha}) \det(\Lambda^k_{\alpha\alpha}) = \sum_{\alpha\subseteq\{1,\dots,m\}} \left(\prod_{j\in\bar{\alpha}} w_j^k\right)\left(\prod_{j\in\alpha} \lambda_j^k\right) \det(M^k_{\alpha\alpha}).$$

If the set α contains more than one repeated constraint, in particular if $\alpha\setminus I_<(\bar{x}) \in I_=^\ell$ with $\ell \geq 1$, the matrix $M^k_{\alpha\alpha}$ has two identical rows, and its determinant vanishes. Hence, to get non-zero summands the set α must particularly contain for each repeated constraint being active at $\bar{x}$ no more than one copy, i.e., $\alpha \setminus I_<(\bar{x}) \in I_=^\ell(\bar{x})$ with $\ell \leq 0$. Then, $\bar{\alpha}$ contains for every non-zero summand at least all but one of the active repeated constraints, i. e., we must have an index set $\beta_\alpha \in I_=^0(\bar{x})$ such that $\bar{\beta}_\alpha \setminus I_<(\bar{x}) \subseteq \bar{\alpha}$. Using (A2) and Lemma 4.2, we get with the index set $\beta \in I_=^0(\bar{x})$ from (A5)

$$\left(\prod_{j\in\bar{\beta}_\alpha\setminus I_<(\bar{x})} w_j^k\right) = \left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w_j^k\right).$$

Thus, we can write

$$\det(M^k\Lambda^k + W^k)$$
$$= \left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w_j^k\right) \sum_{\substack{\alpha\subseteq\{1,\dots,m\},\\ \alpha\setminus I_<(\bar{x})\in I_=^\ell(\bar{x}),\ell\leq 0}} \left(\prod_{j\in\bar{\alpha}\setminus(\bar{\beta}_\alpha\setminus I_<(\bar{x}))} w_j^k\right)\left(\prod_{j\in\alpha} \lambda_j^k\right) \det(M^k_{\alpha\alpha}).$$

If $\alpha \setminus I_<(\bar{x}) \in I_=^\ell(\bar{x})$ for some $\ell < 0$, we must then have by definition an index $j \in \bar{\alpha} \setminus (\bar{\beta}_\alpha \setminus I_<(\bar{x}))$ with $w_j^k \to 0$. Since by (A1) and (A4) all iterates and $\det(M^k_{\alpha\alpha})$ are bounded, this implies

$$\sum_{\substack{\alpha\subseteq\{1,\dots,m\},\\ \alpha\setminus I_<(\bar{x})\in I_=^\ell(\bar{x}),\ell<0}} \left(\prod_{j\in\bar{\alpha}\setminus(\bar{\beta}_\alpha\setminus I_<(\bar{x}))} w_j^k\right)\left(\prod_{j\in\alpha} \lambda_j^k\right) \det(M^k_{\alpha\alpha}) = o(1).$$

If $\alpha \cap I_<(\bar{x}) \neq \emptyset$ we have an index $j \in \alpha$ with $\lambda_j^k \to 0$. This again yields

$$\sum_{\substack{\alpha\subseteq\{1,\dots,m\},\\ \alpha\cap I_<(\bar{x})\neq\emptyset}} \left(\prod_{j\in\bar{\alpha}\setminus(\bar{\beta}_\alpha\setminus I_<(\bar{x}))} w_j^k\right)\left(\prod_{j\in\alpha} \lambda_j^k\right) \det(M^k_{\alpha\alpha}) = o(1).$$

This means that we can only have index sets $\alpha \in I_=^0(\bar{x})$ that lead to non-vanishing summands in the limit. For these we have $\bar{\alpha} \setminus (\bar{\beta}_\alpha \setminus I_<(\bar{x})) = I_<(\bar{x})$. Altogether, we have shown (8). □

In the next lemma we use the technique from the proof of [1, Lemma 5.30].

Lemma 4.5. *Let Assumption* 3.2 *hold. Let* $\{(x^k, \lambda^k, w^k)\}$ *be a subsequence, generated by Algorithm* 1*, that converges to* $(\bar{x}, \bar{\lambda}, \bar{w})$*. Then, there is some constant* $C > 0$ *such that, with* $\beta \in I_=^0(\bar{x})$ *from* (A5)*, we have:*

$$\textit{if } i \in I_<(\bar{x}): \quad \det(\tilde{M}_i^k \tilde{\Lambda}_i^k + \tilde{W}_i^k) \tag{9}$$

$$= -\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w_j^k\Big) \sum_{\alpha\in I_=^0(\bar{x})} \Big(\prod_{j\in I_<(\bar{x})} w_j^k\Big)\Big(\prod_{j\in\alpha}\lambda_j^k\Big)\det(M_{\alpha\alpha}^k) + o\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w_j^k\Big)$$

$$\textit{if } i \in I_=(\bar{x}): \quad \Big|\det(\tilde{M}_i^k \tilde{\Lambda}_i^k + \tilde{W}_i^k)\Big| \le C\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w_j^k\Big)\|H(z^k)\| \tag{10}$$

Proof. Using the diagonal structure of $\tilde{\Lambda}_i^k$ and $\tilde{W}_i^k$ we get

$$\begin{aligned}
\det(\tilde{M}_i^k \tilde{\Lambda}_i^k + \tilde{W}_i^k) &= \sum_{\alpha\subseteq\{1,\dots,m\}} \det((\tilde{W}_i^k)_{\bar{\alpha}\bar{\alpha}})\det((\tilde{\Lambda}_i^k)_{\alpha\alpha})\det((\tilde{M}_i^k)_{\alpha\alpha}) \\
&= \sum_{\substack{\alpha\subseteq\{1,\dots,m\},\\ i\in\alpha}} \Big(\prod_{j\in\bar{\alpha}} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}}\lambda_j^k\Big)\det((\tilde{M}_i^k)_{\alpha\alpha}) \qquad (11)
\end{aligned}$$

where the last equation exploited $(\tilde{\Lambda}_i^k)_{ii} = 1$ and $(\tilde{W}_i^k)_{ii} = 0$. Developing $\det((\tilde{M}_i^k)_{\alpha\alpha})$ by the i-th column corresponding to r_α^k, we obtain

$$\det((\tilde{M}_i^k)_{\alpha\alpha}) = \sum_{s=1}^{|\alpha|}(-1)^{i+s}\, r_{\alpha_s}^k \,\det\left(M_{\alpha\setminus\{\alpha_s\},\alpha\setminus\{i\}}^k\right). \tag{12}$$

Step 1: Let us first take an arbitrary $i \in I_=(\bar{x})$. Now we distinguish four cases

(a) $\alpha \setminus I_<(\bar{x}) \in I_=^\ell(\bar{x})$ with $\ell \ge 2$:

There are two possibilities:

Either we have at least three different indices $l_1, l_2, l_3 \in \alpha\setminus I_<(\bar{x})$ with $g_{l_1} \equiv g_{l_2} \equiv g_{l_3}$. Then the corresponding rows of $(\tilde{M}_i^k)_{\alpha\alpha}$ can only differ in the single column i. Hence, the rows are linearly dependent and $\det(\tilde{M}_i^k)_{\alpha\alpha} = 0$.

Or there are four indices $j_1, j_2, l_1, l_2 \in \alpha \setminus I_<(\bar{x})$ with $j_1 \ne j_2$, $l_1 \ne l_2$ and $g_{j_1} \equiv g_{j_2} \not\equiv g_{l_1} \equiv g_{l_2}$. Then, the differences of the two corresponding pairs of rows are linearly dependent, since they both can only have a nonzero entry in column $\{i\}$. Hence, the four rows of $(\tilde{M}_i^k)_{\alpha\alpha}$ are linearly dependent and again $\det(\tilde{M}_i^k)_{\alpha\alpha} = 0$. Thus, the corresponding summands in (11) are zero.

(b) $\alpha \setminus I_<(\bar{x}) \in I_=^1(\bar{x})$:

This means that there are exactly two indices $s_1, s_2 \in \alpha$ that belong to the same active shared constraint. The definition of $I_=^1(\bar{x})$ implies the existence of an index set $\beta_\alpha \in I_=^0(\bar{x})$ such that $\beta_\alpha = I_=(\bar{x}) \cap (\alpha \setminus \{s_1\})$. With $w_{s_1}^k = w_{s_2}^k$ from (A2) and Lemma 4.2, we get

$$w_{s_1}^k\Big(\prod_{j\in\bar{\alpha}} w_j^k\Big) = \Big(\prod_{j\in\bar{\beta}_\alpha\setminus I_<(\bar{x})} w_j^k\Big)\Big(\prod_{j\in\bar{\alpha}\cap I_<(\bar{x})} w_j^k\Big).$$

With Lemma 4.2 we get with $\beta \in I_=^0(\bar{x})$ from (A5)

$$\Big(\prod_{j\in\bar{\alpha}} w_j^k\Big) = \frac{1}{w_{s_1}^k}\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w_j^k\Big)\Big(\prod_{j\in\bar{\alpha}\cap I_<(\bar{x})} w_j^k\Big).$$

Further, $M^k_{\alpha\setminus\{\alpha_s\},\alpha\setminus\{i\}}$ has two identical rows for all $\alpha_s \notin \{s_1, s_2\}$. Thus, the sum in (12) has only two non-zero summands and these must have opposite sign, which follows from a permuatation of the rows such that the rows belonging to s_1 and s_2 are below each other. We obtain

$$\left|\sum_{s=1}^{|\alpha|}(-1)^{i+s}\, r^k_{\alpha_s}\, \det\left(M^k_{\alpha\setminus\{\alpha_s\},\alpha\setminus\{i\}}\right)\right| = \left|\left(r^k_{s_1} - r^k_{s_2}\right)\det\left(M^k_{\alpha\setminus\{s_1\},\alpha\setminus\{i\}}\right)\right|.$$

Having a closer look at the definition of r^k, and using $w^k_{s_1} = w^k_{s_2}$, we obtain

$$\left|r^k_{s_1} - r^k_{s_2}\right| = \sigma_k\mu_k\left|\frac{1}{\lambda^k_{s_1}} - \frac{1}{\lambda^k_{s_2}}\right| \le \frac{\sigma_k\mu_k}{\min\{\lambda^k_{s_1},\lambda^k_{s_2}\}} = \frac{\sigma_k\mu_k w^k_{s_1}}{\min\{\lambda^k_{s_1}w^k_{s_1},\lambda^k_{s_2}w^k_{s_2}\}}.$$

Finally, we have

$$\begin{aligned}
&\left(\prod_{j\in\bar\alpha} w^k_j\right)\left(\prod_{j\in\alpha\setminus\{i\}}\lambda^k_j\right)\left|\det((\tilde M^k_i)_{\alpha\alpha})\right| \\
&\le \frac{1}{w^k_{s_1}}\left(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w^k_j\right)\left(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w^k_j\right)\left(\prod_{j\in\alpha\setminus\{i\}}\lambda^k_j\right)\det\left(M^k_{\alpha\setminus\{s_1\},\alpha\setminus\{i\}}\right)\frac{\sigma_k\mu_k w^k_{s_1}}{\min\{\lambda^k_{s_1}w^k_{s_1},\lambda^k_{s_2}w^k_{s_2}\}} \\
&\le \left(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w^k_j\right)\left(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w^k_j\right)\left(\prod_{j\in\alpha\setminus\{i\}}\lambda^k_j\right)\det\left(M^k_{\alpha\setminus\{s_1\},\alpha\setminus\{i\}}\right)\frac{\sigma_k\mu_k}{\min\limits_{j=1,\dots,m}\lambda^k_j w^k_j}.
\end{aligned}$$

The boundedness from (A1), (A4), and $\frac{\sigma_k\mu_k}{\min\limits_{j=1,\dots,m}\lambda^k_j w^k_j} \le c\|H(z^k)\|$ by (A3), imply the existence of a constant $C_1 > 0$ such that for all summands in this case

$$\left(\prod_{j\in\bar\alpha} w^k_j\right)\left(\prod_{j\in\alpha\setminus\{i\}}\lambda^k_j\right)\left|\det((\tilde M^k_i)_{\alpha\alpha})\right| \le C_1\left(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w^k_j\right)\|H(z^k)\|.$$

(c) $\alpha \setminus I_<(\bar x) \in I^0_=(\bar x)$:

In this case we get from Lemma 4.2 with $\beta \in I^0_=(\bar x)$ from (A5)

$$\left(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w^k_j\right) = \left(\prod_{j\in\bar\alpha\setminus I_<(\bar x)} w^k_j\right).$$

Hence, we have

$$\begin{aligned}
&\left(\prod_{j\in\bar\alpha} w^k_j\right)\left(\prod_{j\in\alpha\setminus\{i\}}\lambda^k_j\right)\det((\tilde M^k_i)_{\alpha\alpha}) \\
&= \left(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w^k_j\right)\left(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w^k_j\right)\left(\prod_{j\in\alpha\setminus\{i\}}\lambda^k_j\right)\det((\tilde M^k_i)_{\alpha\alpha}).
\end{aligned}$$

Here we have two subcases: Either $\alpha \cap I_<(\bar x) \ne \emptyset$.

Then, we have one index $s \in \alpha\setminus\{i\}$ with $\lambda^k_s \to 0$ and w^k_s is bounded away from zero. Using $\lambda^k_s = \frac{\lambda^k_s w^k_s}{w^k_s} \le \frac{\|H(z^k)\|}{w^k_s}$, and the boundedness from (A1), which together with (A4) implies boundedness of $\det((\tilde M^k_i)_{\alpha\alpha})$, yields the existence of a constant $C_2 > 0$ such that

$$\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\Big|\det((\tilde M_i^k)_{\alpha\alpha})\Big|$$
$$\le \Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big) C_2\,\|H(z^k)\|.$$

Otherwise, $\alpha \subseteq I_=(\bar x)$. By definition we have for all $j \in I_=(\bar x)$

$$r_j^k = -J_x g_j(x^k)\, J_x F(x^k,\lambda^k)^{-1} F(x^k,\lambda^k) + (g_j(x^k)+w_j^k) - w_j^k + w_j^k \frac{\sigma_k\mu_k}{\lambda_j^k w_j^k} - \sigma_k\mu_k. \quad (13)$$

By (A5) we can find for every active constraint at least one positive multiplier. Hence, using (A2) and Lemma 4.2, we have an index $\tilde j$, with

$$\frac{w_j^k}{\|H(z^k)\|} = \frac{w_{\tilde j}^k \lambda_{\tilde j}^k}{\|H(z^k)\|\lambda_{\tilde j}^k} \le \frac{1}{\lambda_{\tilde j}^k},$$

which stays bounded, since $\lambda_{\tilde j}^k \not\to 0$. Then, exploiting our assumptions (A1), (A3) and (A4), $\frac{|r_j^k|}{\|H(z^k)\|}$ is bounded. Using (12) and again (A1) and (A4), we can find a constant $\tilde C_3 > 0$ such that $|\det((\tilde M_i^k)_{\alpha\alpha})| \le \tilde C_3 \|H(z^k)\|$.

Using once more (A1), we have in this subcase a constant $C_3 > 0$ such that

$$\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\Big|\det((\tilde M_i^k)_{\alpha\alpha})\Big|$$
$$\le \Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big) C_3\,\|H(z^k)\|.$$

(d) $\alpha \setminus I_<(\bar x) \in I_=^\ell(\bar x)$ with $\ell \le -1$:

With $\beta \in I_=^0(\bar x)$ from (A5), we get in this case $(\bar\beta \setminus I_<(\bar x)) \subset (\bar\alpha \setminus I_<(\bar x))$, and there is at least one more index $s \in \bar\alpha \setminus I_<(\bar x)$ which is not in $\bar\beta$. For this index we have $w_s^k \to 0$. By (A5) for each active constraint there is at least one positive multiplier, hence we can find an index $\tilde s$, with $\lambda_{\tilde s}^k \not\to 0$ and $w_s^k = w_{\tilde s}^k$ by (A2) and Lemma 4.2.

Then, we have
$$w_s^k = \frac{w_{\tilde s}^k \lambda_{\tilde s}^k}{\lambda_{\tilde s}^k} \le \frac{\|H(z^k)\|}{\lambda_{\tilde s}^k}.$$

With (A1), (A4) and $\lambda_{\tilde s}^k \not\to 0$ we get in this case a constant $C_4 > 0$ such that

$$\Big(\prod_{j\in\bar\alpha} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)|\det((\tilde M_i^k)_{\alpha\alpha})| \le \Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big) C_4\,\|H(z^k)\|.$$

Since we have bounded all summands of the finite sum in (11), we have shown (10).

Step 2: Let us now take an arbitrary $i \in I_<(\bar x)$. Again, we distinguish four cases, namely

$$\alpha \setminus I_<(\bar x) \in I_=^\ell(\bar x)$$

for $\ell \ge 2$, $\ell = 1$, $\ell = 0$ and $\ell \le -1$.

All cases except $\ell = 0$ can be treated identical to the cases (a), (b) and (d) of Step 1 and result in summands $o\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)$, since $\|H(z^k)\| \to 0$ by Lemma 4.1. Thus, we have

$$\begin{aligned}
&\det(\tilde M_i^k \tilde\Lambda_i^k + \tilde W_i^k)\\
&= \sum_{\substack{\alpha\subseteq\{1,\dots,m\},\\ \alpha\setminus I_<(\bar x)\in I_=^0(\bar x)}} \Big(\prod_{j\in\bar\alpha} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\det((\tilde M_i^k)_{\alpha\alpha}) + o\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)
\end{aligned}$$

As in case (c) with $\alpha \setminus I_<(\bar x) \in I_=^0(\bar x)$, we obtain for the summands

$$\begin{aligned}
&\Big(\prod_{j\in\bar\alpha} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\det((\tilde M_i^k)_{\alpha\alpha})\\
&= \Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\det((\tilde M_i^k)_{\alpha\alpha}).
\end{aligned}$$

Now, for any index set α with $(\alpha\setminus\{i\})\cap I_<(\bar x) \neq \emptyset$, we have an index $s \in \alpha\setminus\{i\}$ with $\lambda_s^k \to 0$. Hence, using (A1) and (A4), these summands are again $o\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)$.
For the remaining summands we have $(\alpha \setminus \{i\}) \in I_=^0(\bar x)$. Here, we can exploit $r_{\alpha_s}^k = o(1)$ for all $s \in \{1,\dots,|\alpha|\} \setminus \{i\}$ and $r_i^k = -w_i^k + o(1)$, which follows from (13), (A1), (A3), (A4) and $\|H(z^k)\| \to 0$ by Lemma 4.1. Inserting this in (12), we get together with (A1) and (A4)

$$\det((\tilde M_i^k)_{\alpha\alpha}) = \sum_{s=1}^{|\alpha|}(-1)^{i+s}\, r_{\alpha_s}^k \det\left(M^k_{\alpha\setminus\{\alpha_s\},\alpha\setminus\{i\}}\right) = -w_i^k \det\left(M^k_{\alpha\setminus\{i\},\alpha\setminus\{i\}}\right) + o(1),$$

resulting in

$$\begin{aligned}
&\Big(\prod_{j\in\bar\alpha} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\det((\tilde M_i^k)_{\alpha\alpha})\\
=&\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\bar\alpha\cap I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\det((\tilde M_i^k)_{\alpha\alpha})\\
=&-\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in I_<(\bar x)} w_j^k\Big)\Big(\prod_{j\in\alpha\setminus\{i\}} \lambda_j^k\Big)\det\left(M^k_{\alpha\setminus\{i\},\alpha\setminus\{i\}}\right) + o\Big(\prod_{j\in\bar\beta\setminus I_<(\bar x)} w_j^k\Big).
\end{aligned}$$

Altogether this proves (9). □

Now we can present new quantitative limiting results, that are stronger than the results presented in [1].

Theorem 4.6. *Let Assumption* 3.2 *hold. Let* $\{(x^k, \lambda^k, w^k)\}$ *be a subsequence, generated by Algorithm* 1, *that converges to* $(\bar{x}, \bar{\lambda}, \bar{w})$. *Then, we have*

$$\lim_{k\to\infty} \frac{d^k_{\lambda,i}}{\lambda^k_i} = 0 \qquad \text{for } i \in I_=(\bar{x}),$$

$$\lim_{k\to\infty} \frac{d^k_{\lambda,i}}{\lambda^k_i} = -1 \qquad \text{for } i \in I_<(\bar{x}),$$

$$\|d^k\| = \mathcal{O}(\|H(z^k)\|).$$

Proof. The proof combines Lemma 4.4 and 4.5. By (7) we have

$$\frac{d^k_{\lambda,i}}{\lambda^k_i} = \frac{\det(\tilde{M}^k_i \tilde{\Lambda}^k_i + \tilde{W}^k_i)}{\det(M^k \Lambda^k + W^k)}.$$

We can find a constant $\tilde{C} \neq 0$ with

$$\lim_{k\to\infty} \sum_{\alpha\in I^0_=(\bar{x})} \Big(\prod_{j\in I_<(\bar{x})} w^k_j\Big)\Big(\prod_{j\in\alpha}\lambda^k_j\Big) \det(M^k_{\alpha\alpha}) = \tilde{C}, \tag{14}$$

since all summands are bounded by (A1) and (A4) and have by (A5) the same sign, and at least one, and thus the entire sum, is nonzero.

For $i \in I_<(\bar{x})$ we get

$$\lim_{k\to\infty} \frac{d^k_{\lambda,i}}{\lambda^k_i}$$

$$= \lim_{k\to\infty} \frac{-\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\Big) \sum_{\alpha\in I^0_=(\bar{x})} \Big(\prod_{j\in I_<(\bar{x})} w^k_j\Big)\Big(\prod_{j\in\alpha}\lambda^k_j\Big) \det(M^k_{\alpha\alpha}) + o\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\Big)}{\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\Big) \sum_{\alpha\in I^0_=(\bar{x})} \Big(\prod_{j\in I_<(\bar{x})} w^k_j\Big)\Big(\prod_{j\in\alpha}\lambda^k_j\Big) \det(M^k_{\alpha\alpha}) + o\Big(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\Big)}$$

$$= \lim_{k\to\infty} \frac{-\sum_{\alpha\in I^0_=(\bar{x})} \Big(\prod_{j\in I_<(\bar{x})} w^k_j\Big)\Big(\prod_{j\in\alpha}\lambda^k_j\Big) \det(M^k_{\alpha\alpha}) + o(1)}{\sum_{\alpha\in I^0_=(\bar{x})} \Big(\prod_{j\in I_<(\bar{x})} w^k_j\Big)\Big(\prod_{j\in\alpha}\lambda^k_j\Big) \det(M^k_{\alpha\alpha}) + o(1)} = \frac{-\tilde{C}}{\tilde{C}} = -1.$$

Further, we have in this case $w^k_i \to \bar{w}_i \neq 0$. Hence, there is a constant $c_1 > 0$ with $w^k_i > c_1$. Since $\lim_{k\to\infty} \frac{d^k_{\lambda,i}}{\lambda^k_i} = -1$ implies boundedness of $\frac{d^k_{\lambda,i}}{\lambda^k_i}$, we also have that

$$\frac{|d^k_{\lambda,i}|}{\|H(z^k)\|} \leq \frac{|d^k_{\lambda,i}|}{\lambda^k_i w^k_i}$$

is bounded, which means $\qquad |d^k_{\lambda,i}| = \mathcal{O}(\|H(z^k)\|)$.

Using $\|H(z^k)\| \to 0$ by Lemma 4.1, we get for $i \in I_=(\bar{x})$

$$\lim_{k\to\infty}\left|\frac{d^k_{\lambda,i}}{\lambda^k_i}\right| = \lim_{k\to\infty}\frac{|\det(\tilde{M}^k_i\tilde{\Lambda}^k_i + \tilde{W}^k_i)|}{|\det(M^k\Lambda^k + W^k)|}$$

$$\leq \lim_{k\to\infty}\frac{C\left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\right)\|H(z^k)\|}{\left|\left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\right)\sum_{\alpha\in I^0_=(\bar{x})}\left(\prod_{j\in I_<(\bar{x})} w^k_j\right)\left(\prod_{j\in\alpha}\lambda^k_j\right)\det(M^k_{\alpha\alpha}) + o\left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\right)\right|}$$

$$= \lim_{k\to\infty}\frac{C\|H(z^k)\|}{\left|\sum_{\alpha\in I^0_=(\bar{x})}\left(\prod_{j\in I_<(\bar{x})} w^k_j\right)\left(\prod_{j\in\alpha}\lambda^k_j\right)\det(M^k_{\alpha\alpha}) + o(1)\right|} = \frac{0}{\tilde{C}} = 0.$$

Furthermore, we have in this case

$$\left|\frac{d^k_{\lambda,i}}{\|H(z^k)\|}\right| = \frac{|d^k_{\lambda,i}|\,\lambda^k_i}{\lambda^k_i\,\|H(z^k)\|}$$

$$\leq \frac{C\,\lambda^k_i\left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\right)}{\left|\left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\right)\sum_{\alpha\in I^0_=(\bar{x})}\left(\prod_{j\in I_<(\bar{x})} w^k_j\right)\left(\prod_{j\in\alpha}\lambda^k_j\right)\det(M^k_{\alpha\alpha}) + o\left(\prod_{j\in\bar{\beta}\setminus I_<(\bar{x})} w^k_j\right)\right|}$$

$$= \frac{C\lambda^k_i}{\left|\sum_{\alpha\in I^0_=(\bar{x})}\left(\prod_{j\in I_<(\bar{x})} w^k_j\right)\left(\prod_{j\in\alpha}\lambda^k_j\right)\det(M^k_{\alpha\alpha}) + o(1)\right|},$$

which is bounded by (A1) and (14), which implies $|d^k_{\lambda,i}| = \mathcal{O}(\|H(z^k)\|)$. Both cases together show

$$\left\|d^k_\lambda\right\| = \mathcal{O}(\|H(z^k)\|).$$

From (3), (A1), (A4) and $\left\|d^k_\lambda\right\| = \mathcal{O}(\|H(z^k)\|)$, we have

$$\left\|d^k_x\right\| = \left\|J_xF(x^k,\lambda^k)^{-1}\left(-E(x^k)d^k_\lambda - F(x^k,\lambda^k)\right)\right\| = \mathcal{O}(\|H(z^k)\|).$$

Finally, we get from this, (4), and (A3)

$$\left\|d^k_w\right\| = \left\|-J_xg(x^k)d^k_x - (g(x^k)+w^k) + \sigma_k\mu_k\right\| = \mathcal{O}(\|H(z^k)\|).$$

Altogether, we have shown: $\|d^k\| = \mathcal{O}(\|H(z^k)\|)$. □

Next, we show that $\|d^k\| = \mathcal{O}(\|H(z^k)\|)$ holds on the entire sequence generated by Algorithm 1.

Lemma 4.7. *Let Assumption* 3.2 *hold and* $\{z^k\}$ *be generated by Algorithm* 1. *Then we have*

$$\|d^k\| = \mathcal{O}(\|H(z^k)\|).$$

Proof. Assume for contradiction, that there is a subsequence $\{z^k\}_{k\in K}$ violating the condition, i.e., we have a subsequence with

$$\lim_{k\in K} \frac{\|d^k\|}{\|H(z^k)\|} = \infty.$$

The subsequence $\{z^k\}_{k\in K}$ is by (A1) bounded, hence has an accumulation point. This point is by Theorem 2.1 a solution and hence it must satisfy (A5). Thus, we can apply Theorem 4.6 to get $\|d^k\| = \mathcal{O}(\|H(z^k)\|)$ on an infinite subset of K, and we have a contradiction. □

Let us comment on our results. By the convergence of $d^k_{\lambda,i}/\lambda^k_i$ to 0 or -1 depending on whether we have an active or inactive constraint, we found an expression that can be used as an indicator to identify the active set at the accumulation point. This could be exploited numerically, if we switch to a reduced system by dropping the inactive constraints, as soon as we are confident that the active set is estimated accurately enough.

Next we strengthen the assertion of [1, Lemma 5.35] in showing that the residual term $q(z^k,t)$, defined in the next Lemma, is of order $\mathcal{O}(\|H(z^k)\|^2)$ also if the step-size does not converge to 0.

Lemma 4.8. *Let Assumption* 3.2 *hold. Let* $\{z^k\}$ *be generated by Algorithm 1.*

Define
$$q(z^k,t) := \begin{pmatrix} q_1(z^k,t) \\ q_2(z^k,t) \\ q_3(z^k,t) \end{pmatrix} \in \mathbb{R}^{n+m+m} \quad \text{by}$$

$$\begin{aligned} q_1(z^k,t) &:= \left(\int_0^1 \left(J_xF(x^k + std^k_x, \lambda^k + std^k_\lambda) - J_xF(x^k,\lambda^k) \right) ds \right) d^k_x \\ &\quad + \left(\int_0^1 \left(E(x^k + std^k_x) - E(x^k) \right) ds \right) d^k_\lambda, \\ q_2(z^k,t) &:= \left(\int_0^1 \left(J_xg(x^k + std^k_x) - J_xg(x^k) \right) ds \right) d^k_x + \sigma_k\mu_k 1_m, \\ q_3(z^k,t) &:= \sigma_k\mu_k 1_m + t(d^k_w \circ d^k_\lambda). \end{aligned}$$

Then, we have
$$H(z^k + td^k) = (1-t)H(z^k) + t\, q(z^k,t), \tag{15}$$

and
$$\|q(z^k,t)\| = \mathcal{O}(\|H(z^k)\|^2). \tag{16}$$

Proof. The first part of the proof is the same as in [1, Lemma 5.35], and we restate it here for convenience. Using the mean value theorem we obtain for the first n-dimensional part of H

$$\begin{aligned} F(x^k + td^k_x, \lambda^k + td^k_\lambda) &= F(x^k,\lambda^k) + t\left(J_xF(x^k,\lambda^k)d^k_x + E(x^k)d^k_\lambda\right) \\ &\quad + t\left(\int_0^1 \left(J_xF(x^k + std^k_x, \lambda^k + std^k_\lambda) - J_xF(x^k,\lambda^k) \right) ds \right) d^k_x \\ &\quad + t\left(\int_0^1 \left(E(x^k + std^k_x) - E(x^k) \right) ds \right) d^k_\lambda \end{aligned}$$

$$
\begin{aligned}
&\overset{(3)}{=} (1-t)F(x^k,\lambda^k) \\
&\quad + t\left(\int_0^1 \big(J_xF(x^k+std_x^k,\lambda^k+std_\lambda^k) - J_xF(x^k,\lambda^k)\big)\,ds\right) d_x^k \\
&\quad + t\left(\int_0^1 \big(E(x^k+std_x^k) - E(x^k)\big)\,ds\right) d_\lambda^k \\
&= (1-t)F(x^k,\lambda^k) + tq_1(z^k,t),
\end{aligned}
$$

for the second m-dimensional part

$$
\begin{aligned}
&g(x^k+td_x^k) + w^k + td_w^k \\
&= g(x^k) + w^k + t\left(J_xg(x^k)d_x^k + d_w^k\right) + t\left(\int_0^1 \big(J_xg(x^k+std_x^k) - J_xg(x^k)\big)\,ds\right) d_x^k \\
&\overset{(4)}{=} (1-t)\left(g(x^k)+w^k\right) + t\sigma_k\mu_k 1_m + t\left(\int_0^1 \big(J_xg(x^k+std_x^k) - J_xg(x^k)\big)\,ds\right) d_x^k \\
&= (1-t)\left(g(x^k)+w^k\right) + tq_2(z^k,t),
\end{aligned}
$$

and for the third m-dimensional part

$$
\begin{aligned}
&(\lambda^k+td_\lambda^k)\circ(w^k+td_w^k) = \lambda^k\circ w^k + t(\lambda^k\circ d_w^k + w^k\circ d_\lambda^k) + t^2(d_w^k\circ d_\lambda^k) \\
&\overset{(5)}{=} (1-t)\lambda^k\circ w^k + t\sigma_k\mu_k 1_m + t^2(d_w^k\circ d_\lambda^k) = (1-t)\lambda^k\circ w^k + tq_3(z^k,t).
\end{aligned}
$$

This proves $$H(z^k+td^k) = (1-t)H(z^k) + tq(z^k,t).$$

The second part of this proof is new. By the assumed local Lipschitz continuity of all derivatives, and the boundedness assumption (A1), we get a constant $L_F > 0$ with

$$
\begin{aligned}
&\left\|\int_0^1 \big(J_xF(x^k+std_x^k,\lambda^k+std_\lambda^k) - J_xF(x^k,\lambda^k)\big)\,ds\,d_x^k\right\| \\
&\le \int_0^1 \left\|J_xF(x^k+std_x^k,\lambda^k+std_\lambda^k) - J_xF(x^k,\lambda^k)\right\|\,ds\,\|d_x^k\| \\
&\le \int_0^1 L_F\,s\,t\,\left\|((d_x^k)^\top,(d_\lambda^k)^\top)^\top\right\|\,ds\,\|d_x^k\| \\
&\le \frac{L_F}{2}t\left(\|d_x^k\|^2 + \|d_x^k\|\cdot\|d_\lambda^k\|\right) = \mathcal{O}(\|H(z^k)\|^2),
\end{aligned}
$$

where the last equation follows from $\|d^k\| = \mathcal{O}(\|H(z^k)\|)$ by Lemma 4.7. Similarly, we get

$$
\left\|\int_0^1 \big(E(x^k+std_x^k) - E(x^k)\big)\,ds\,d_\lambda^k\right\| = \mathcal{O}(\|H(z^k)\|^2),
$$

$$
\left\|\int_0^1 \big(J_xg(x^k+std_x^k) - J_xg(x^k)\big)\,ds\,d_x^k\right\| = \mathcal{O}(\|H(z^k)\|^2).
$$

With $\sigma_k \mu_k = \mathcal{O}(\|H(z^k)\|^2)$ by (A3), $t \in (0, 1]$, and further

$$\|d_\lambda^k \circ d_w^k\| \leq \|d_\lambda^k\| \cdot \|d_w^k\| = \mathcal{O}(\|H(z^k)\|^2),$$

we obtain

$$\|q(z^k, t)\| = \mathcal{O}(\|H(z^k)\|^2).$$

□

Now, we have analyzed the behavior of the direction d^k and found some useful relations (15) and (16). It remains to have a detailed look at the stepsize t_k, restricted by the set Z_I and the Armijo linesearch with the potential function. One motivation for using the potential function is to bring some of the components of $H(z^k)$ commonly to zero at around the same rate. The following lemma follows from the definition of the potential function without further assumptions.

Lemma 4.9. *Let the sequence $\{z^k\}$ be generated by Algorithm 1. Then, we have for all $i = 1, \dots, 2m$ and all $k \in \mathbb{N}$*

$$H_{n+i}(z^k) \geq e^{-\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{\frac{\zeta}{m}}.$$

Proof. By (S.3) of Algorithm 1 and the descent property

$$\nabla \psi(z^k)^\top d^k \leq 2(\zeta - m)(\sigma_k - 1) < 0$$

from [1, Lemma 5.11] the sequence $\{\psi(z^k)\}_{k \in \mathbb{N}}$ is monotonically decreasing for $\sigma_k < 1$ and $\zeta > m$. By definition of the potential function we obtain

$$\begin{aligned}
\psi(z^0) &\geq \psi(z^k) \\
&= \zeta \ln(\|H(z^k)\|^2) - \sum_{i=1}^{2m} \ln(H_{n+i}(z^k)) \\
&\geq 2\zeta \ln(\|H(z^k)\|) - 2m \ln\left(\min_{i=1,\dots,2m} H_{n+i}(z^k)\right) \\
&= 2(\zeta - m) \ln(\|H(z^k)\|) + 2m \ln\left(\frac{\|H(z^k)\|}{\min_{i=1,\dots,2m} H_{n+i}(z^k)}\right).
\end{aligned}$$

for all $k \in \mathbb{N}$. Reordering yields

$$\ln\left(\frac{\|H(z^k)\|}{\min_{i=1,\dots,2m} H_{n+i}(z^k)}\right) \leq \frac{\psi(z^0)}{2m} - \frac{2(\zeta - m)}{2m} \ln(\|H(z^k)\|),$$

which, in turn, implies

$$\frac{\|H(z^k)\|}{\min_{i=1,\dots,2m} H_{n+i}(z^k)} \leq e^{\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{-\frac{\zeta - m}{m}}.$$

Finally, we obtain the assertion

$$\min_{i=1,\dots,2m} H_{n+i}(z^k) \geq e^{-\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{\frac{\zeta}{m}}.$$

□

Now, we analyze the conditions restricting the stepsize in (S.3) of Algorithm 1. Here, we use a fixed parameter $\eta_k = \eta$.

Lemma 4.10. *Let Assumption* 3.2 *hold and* $\{z^k\}$ *be generated by Algorithm* 1. *Let any fixed* $\eta \in (0,1)$ *be given, and assume* $\zeta \in (m, 2m)$. *Then, for all sufficiently large* $k \in \mathbb{N}$, *the stepsize* $t_k = \eta$ *satisfies* $z^k + t_k d^k \in Z_I$ *and*

$$\psi(z^k + t_k d^k) \le \psi(z^k) + \gamma t_k \nabla\psi(z^k)^\top d^k.$$

In particular, there is some lower bound $\bar{t} > 0$ *such that* $t_k \ge \bar{t}$ *for all* $k \in \mathbb{N}$, *i.e., the stepsize is bounded away from zero.*

Proof. Let us recall the definition

$$Z_I := \{z = (x, \lambda, w) \in \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^m_{++} \mid g(x) + w > 0\}.$$

By Lemma 4.8 and Lemma 4.9 we have for all $i = 1, \dots, 2m$

$$\begin{aligned} H_{n+i}(z^k + \eta d^k) &= (1-\eta)\, H_{n+i}(z^k) + \eta\, q_{n+i}(z^k, \eta) \\ &\ge (1-\eta)\, e^{-\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{\frac{\zeta}{m}} + \eta\, q_{n+i}(z^k, \eta). \end{aligned}$$

By (16) we have a constant $\tilde{C}$ such that

$$q_{n+i}(z^k, \eta) \ge -\tilde{C}\|H(z^k)\|^2$$

for all $k \in \mathbb{N}$. With $\frac{\zeta}{m} < 2$, $\eta < 1$ and $\|H(z^k)\| \to 0$ by Lemma 4.1, we obtain

$$H_{n+i}(z^k + \eta d^k) > 0 \tag{17}$$

for all $k \in \mathbb{N}$ sufficiently large and $i = 1, \dots, 2m$. In particular, we have

$$g(x^k + \eta d_x^k) + w^k + \eta d_w^k = H_{(n+1):(n+m)}(z^k + \eta d^k) > 0.$$

Further, we have to show $\lambda^k + \eta d^k > 0$ and $w^k + \eta d^k > 0$. Since for sufficiently large $k \in \mathbb{N}$ and $i \in \{1, \dots, m\}$

$$(\lambda_i^k + \eta d_{\lambda,i}^k)(w_i^k + \eta d_{w,i}^k) = H_{n+m+i}(z^k + \eta d^k) > 0,$$

by (17), both factors must have the same positive or negative signum for all sufficiently large $k \in \mathbb{N}$. By (5) we have for all $i = 1, \dots, m$

$$w_i^k d_{\lambda,i}^k + \lambda_i^k d_{w,i}^k = -\lambda_i^k w_i^k + \sigma_k \mu_k > -\lambda_i^k w_i^k.$$

Hence, we cannot have simultaneously the relations $d_{\lambda,i}^k \le -\lambda_i^k$ and $d_{w,i}^k \le -w_i^k$. Thus, $\lambda^k + \eta d^k > 0$ and $w^k + \eta d^k > 0$ must hold for all sufficiently large $k \in \mathbb{N}$. This shows $z^k + \eta d^k \in Z_I$ for all sufficiently large $k \in \mathbb{N}$.

For the Armijo-condition on the potential function we use that

$$\lim_{k \to \infty} \frac{\|q(z^k, \eta)\|}{\|H(z^k)\|^{\frac{\zeta}{m}}} = 0$$

by Lemma 4.8, $\zeta < 2m$ and Lemma 4.1.

Note that by (17) all the following logarithmic terms are well-defined for all sufficiently large $k \in \mathbb{N}$. On the one hand, we get

$$\begin{aligned}
&\limsup_{k\to\infty} \left[\psi(z^k+\eta d^k)-\psi(z^k)\right] \\
= \ &\limsup_{k\to\infty} \left[\ln\left(\frac{\|H(z^k+\eta d^k)\|}{\|H(z^k)\|}\right)^{2\zeta} - \sum_{i=1}^{2m} \ln\left(\frac{H_{n+i}(z^k+\eta d^k)}{H_{n+i}(z^k)}\right)\right] \\
\overset{(15)}{\leq} \ &\limsup_{k\to\infty} \left[\ln\left(1-\eta+\eta\frac{\|q(z^k,\eta)\|}{\|H(z^k)\|}\right)^{2\zeta} - \sum_{i=1}^{2m} \ln\left(1-\eta+\eta\frac{q_{n+i}(z^k,\eta)}{H_{n+i}(z^k)}\right)\right] \\
\leq \ &\limsup_{k\to\infty} \left[2\zeta\ln\left(1-\eta+\eta\frac{\|q(z^k,\eta)\|}{\|H(z^k)\|}\right) - 2m\ln\left(1-\eta-\eta\frac{\|q(z^k,\eta)\|}{\min\limits_{i=1,\dots,2m} H_{n+i}(z^k)}\right)\right] \\
\leq \ &\limsup_{k\to\infty} \left[2\zeta\ln\left(1-\eta+\eta\frac{\|q(z^k,\eta)\|}{\|H(z^k)\|}\right) - 2m\ln\left(1-\eta-\eta\frac{\|q(z^k,\eta)\|}{e^{-\frac{\psi(z^0)}{2m}}\|H(z^k)\|^{\frac{\zeta}{m}}}\right)\right] \\
\overset{(16)}{=} \ &2\,\zeta\ln(1-\eta)-2m\ln(1-\eta) \leq -2(\zeta-m)\eta.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\lim_{k\to\infty}\nabla\Psi(z^k)^\top d^k &= \lim_{k\to\infty}\left(\zeta\frac{2\,JH(z^k)^\top H(z^k)}{\|H(z^k)\|^2} - \sum_{i=1}^{2m}\frac{JH_{n+i,\bullet}(z^k)^\top}{H_{n+i}(z^k)}\right)^\top d^k \\
&= \lim_{k\to\infty}\left[2\zeta\frac{H(z^k)^\top(-H(z^k)+\sigma_k\mu_k a)}{\|H(z^k)\|^2} - \sum_{i=1}^{2m}\frac{-H_{n+i}(z^k)+\sigma_k\mu_k}{H_{n+i}(z^k)}\right] \\
&= \lim_{k\to\infty}\left[-2(\zeta-m)+4\zeta m\frac{\sigma_k\mu_k^2}{\|H(z^k)\|^2} - \sum_{i=1}^{2m}\frac{\sigma_k\mu_k}{H_{n+i}(z^k)}\right] \\
&\overset{(A3)}{=} -2(\zeta-m).
\end{aligned}$$

Since $\zeta > m$ and $\gamma \in (0,1)$ we have shown

$$\limsup_{k\to\infty}\psi(z^k+\eta d^k)-\psi(z^k) \leq -2(\zeta-m)\eta < -2(\zeta-m)\gamma\eta = \lim_{k\to\infty}\gamma\eta\nabla\Psi(z^k)^\top d^k,$$

hence, the stepsize $t_k = \eta$ satisfies the Armijo condition for all sufficiently large k. Furthermore, this implies that the stepsize does not converge to zero on any subsequence, and hence is bounded away from zero. □

With this Lemma, we have also the stepsize boundedness that was necessary for the proof of the finite termination criterion in [1, Theorem 5.23]. With $\lceil\cdot\rceil$ denoting the rounding to the next larger integer, we have [1, Theorem 5.23].

Corollary 4.11. *Let $\{z^k\}$ be generated by Algorithm* 1. *Let Assumption* 3.2 *hold, $\eta_k = \eta$ for all $k \in \mathbb{N}$, $\tilde\sigma := \sup_{k\in\mathbb{N}}\sigma_k < 1$, and let $\bar t > 0$ be a lower bound on the stepsize, i.e., $t_k \geq \bar t$ for all $k \in \mathbb{N}$. Then, the termination criterion $\|H(z^k)\| \leq \varepsilon$ is satisfied after at most k iterations, where*

$$k = \left\lceil \frac{\psi(z^0)}{2\gamma(\zeta-m)(1-\tilde\sigma)\bar t} - \frac{\ln\varepsilon}{4\gamma(1-\tilde\sigma)\bar t}\right\rceil$$

Finally, let us state the linear convergence properties of the potential reduction algorithm.

Theorem 4.12. *Let Assumption* 3.2 *hold and* $\{z^k\}$ *be generated by Algorithm* 1. *Let* $\eta_k = \eta \in (0,1)$ *for all* $k \in \mathbb{N}_0$ *be fixed, and assume* $\zeta \in (m, 2m)$. *Then, the sequence* $\{\|H(z^k)\|\}_{k\in\mathbb{N}}$ *converges Q-linearly to zero. Further, the sequence* $\{dist(z^k, S)\}_{k\in\mathbb{N}}$ *converges R-linearly to zero.*

Proof. The Q-linear convergence of $\{\|H(z^k)\|\}_{k\in\mathbb{N}}$ to 0 follows from (15) and (16) in Lemma 4.8 and the boundedness of the stepsize from below from Lemma 4.10.

To show the R-linear convergence of $\{\text{dist}(z^k, S)\}_{k\in\mathbb{N}}$ we need a local error bound condition. By (A5) we can use [3, Theorem 1] for every solution $\bar{z} \in S$ to get constants $\delta_{\bar{z}} > 0$ and $\ell_{\bar{z}} > 0$ such that the local error bound condition

$$\text{dist}(z, S) \le \ell_{\bar{z}} \, \|H(z)\| \qquad \text{for all } z \in \mathbb{B}(\bar{z}, \delta_{\bar{z}}) \cap (\mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^m_+)$$

holds. Herein, $\mathbb{B}(\bar{z}, \delta_{\bar{z}})$ is an open ball around $\bar{z}$ with radius $\delta_{\bar{z}}$. By the boundedness assumption (A1) all iterates z^k, and hence also all accumulation points of this sequence, are contained in a compact set Z. Hence, we have for the open covering

$$\bigcup_{\bar{z}\in S\cap Z} \mathbb{B}(\bar{z}, \delta_{\bar{z}})$$

of the compact set $S \cap Z$ a finite sub-covering, i.e., we can find a finite number of points $\bar{z}^1, \dots, \bar{z}^J \in S \cap Z$ such that

$$S \cap Z \subset \bigcup_{j=1,\dots,J} \mathbb{B}(\bar{z}^j, \delta_{\bar{z}^j}).$$

Then we obtain positive constants

$$\ell := \max_{j=1,\dots,J} \ell_{\bar{z}^j} \qquad \text{and} \qquad \delta := \min_{j=1,\dots,J} \delta_{\bar{z}^j} > 0.$$

with $\text{dist}(z, S) \le \ell \, \|H(z)\|$ for all $z \in ((S \cap Z) + \mathbb{B}(0, \delta)) \cap (\mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^m_+)$.

All iterates are contained in $((S\cap Z)+\mathbb{B}(0,\delta))\cap(\mathbb{R}^n \times\mathbb{R}^m_+\times\mathbb{R}^m_+)$, for sufficiently large $k \in \mathbb{N}$ by (A1) and Theorem 2.1. Thus, the Q-linear convergence of $\{\|H(z^k)\|\}_{k\in\mathbb{N}}$ implies the R-linear convergence of $\{\text{dist}(z^k, S)\}_{k\in\mathbb{N}}$ to 0. □

5. Superlinear convergence

Inspecting (15) and (16) again, we see that faster (quadratic) convergence would be possible, if $t_k = 1$ would be obtained for all $k \in \mathbb{N}$ sufficiently large. However, it is not clear under which conditions this might be possible. For the proof of Lemma 4.10, it was crucial, that $\eta < 1$ to obtain that $t_k = \eta$ is a feasible stepsize for sufficiently large $k \in \mathbb{N}$. What we can do, is to use a sequence $\{\eta_k\}$, which satisfies $(1 - \eta_k) = \mathcal{O}(\|H(z^k)\|^\delta)$ with $\delta \in \left(0, 2 - \frac{\zeta}{m}\right)$, for example of the form

$$\eta_k := \max\left\{\frac{1}{2}, 1 - \hat{c}\|H(z^k)\|^\delta\right\}, \tag{18}$$

with some constant $\hat{c} > 0$.

For sufficiently large $k \in \mathbb{N}$ this means that $\eta_k = 1 - \hat{c}\|H(z^k)\|^\delta$, which converges to 1 since $\|H(z^k)\| \to 0$ by Lemma 4.1. Then, we can show the following.

Lemma 5.1. *Let Assumption* 3.2 *hold and* $\{(x^k, \lambda^k, w^k)\}$ *be generated by Algorithm* 1. *Assume* $\zeta \in (m, 2m)$ *and that the sequence* $\{\eta_k\}$ *satisfies* (18) *with* $\delta \in \left(0, 2 - \frac{\zeta}{m}\right)$. *Then, for all sufficiently large* $k \in \mathbb{N}$, *the stepsize* $t_k = \eta_k$ *satisfies* $z^k + t_k d^k \in Z_I$ *and*

$$\psi(z^k + t_k d^k) \leq \psi(z^k) + \gamma t_k \nabla\psi(z^k)^\top d^k.$$

Proof. We follow the proof of Lemma 4.10. Using Lemma 4.8, Lemma 4.9 and (18), we have for all $i = 1, \ldots, 2m$

$$\begin{aligned} H_{n+i}(z^k + \eta_k d^k) =& (1 - \eta_k)\, H_{n+i}(z^k) + \eta_k\, q_{n+i}(z^k, \eta_k) \\ \geq& \hat{c}\, e^{-\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{\delta + \frac{\zeta}{m}} - \tilde{C}\|H(z^k)\|^2 > 0, \end{aligned}$$

for all $k \in \mathbb{N}$ sufficiently large, since $\delta + \frac{\zeta}{m} < 2$ by assumption. Having $\eta_k < 1$, the proof of $z^k + \eta_k d^k \in Z_I$ for all $k \in \mathbb{N}$ sufficiently large is analogous to that part of the proof of Lemma 4.10. Further, we have

$$\begin{aligned} &\limsup_{k\to\infty} \frac{\psi(z^k + \eta_k d^k) - \psi(z^k)}{\eta_k} \\ =& \limsup_{k\to\infty} \frac{1}{\eta_k} \left(\ln\left(\frac{\|H(z^k + \eta_k d^k)\|}{\|H(z^k)\|} \right)^{2\zeta} - \sum_{i=1}^{2m} \ln\left(\frac{H_{n+i}(z^k + \eta_k d^k)}{H_{n+i}(z^k)} \right) \right) \\ \leq& \limsup_{k\to\infty} \frac{1}{\eta_k} \left(\ln\left(1 - \eta_k + \eta_k \frac{\|q(z^k, \eta_k)\|}{\|H(z^k)\|}\right)^{2\zeta} \right. \\ &\qquad \left. - \sum_{i=1}^{2m} \ln\left(1 - \eta_k + \eta_k \frac{q_{n+i}(z^k, \eta_k)}{H_{n+i}(z^k)}\right) \right) \\ \leq& \limsup_{k\to\infty} \frac{1}{\eta_k} \left(\ln\left(1 - \eta_k + \eta_k \frac{\|q(z^k, \eta_k)\|}{\|H(z^k)\|}\right)^{2\zeta} \right. \\ &\qquad \left. - \sum_{i=1}^{2m} \ln\left(1 - \eta_k - \eta_k \frac{\|q(z^k, \eta)\|}{e^{-\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{\frac{\zeta}{m}}}\right) \right) \\ =& \limsup_{k\to\infty} \frac{1}{\eta_k} \left(\ln\left(1 - \eta_k + \eta_k \frac{\|q(z^k, \eta_k)\|}{\|H(z^k)\|}\right)^{2(\zeta - m)} \right. \\ &\qquad \left. + 2m \ln\left(\frac{1 - \eta_k + \eta_k \frac{\|q(z^k, \eta_k)\|}{\|H(z^k)\|}}{1 - \eta_k - \eta_k \frac{\|q(z^k, \eta)\|}{e^{-\frac{\psi(z^0)}{2m}} \|H(z^k)\|^{\frac{\zeta}{m}}}} \right) \right). \end{aligned}$$

For the first term, we can use

$$\lim_{k\to\infty} \frac{1}{\eta_k} \ln\left(1 - \eta_k + \eta_k \frac{\|q(z^k, \eta_k)\|}{\|H(z^k)\|}\right) \leq \lim_{k\to\infty} \frac{1}{\eta_k} \eta_k \left(\frac{\|q(z^k, \eta_k)\|}{\|H(z^k)\|} - 1 \right) = -1.$$

For the second term, with $\frac{\zeta}{m} + \delta < 2$ by assumption and $\|q(z^k, \eta_k)\| = \mathcal{O}(\|H(z^k)\|^2)$ by Lemma 4.8, we can write

$$\lim_{k\to\infty} \ln\left(\frac{1-\eta_k+\eta_k\frac{\|q(z^k,\eta_k)\|}{\|H(z^k)\|}}{1-\eta_k-\eta_k\frac{\|q(z^k,\eta_k)\|}{e^{-\frac{\psi(z^0)}{2m}}\|H(z^k)\|^{\frac{\zeta}{m}}}}\right) = \lim_{k\to\infty} \ln\left(\frac{1+\frac{\eta_k\,\|q(z^k,\eta_k)\|}{(1-\eta_k)\|H(z^k)\|}}{1-\frac{\eta_k\,\|q(z^k,\eta_k)\|}{e^{-\frac{\psi(z^0)}{2m}}(1-\eta_k)\|H(z^k)\|^{\frac{\zeta}{m}}}}\right)$$

$$= \lim_{k\to\infty} \ln\left(\frac{1+\frac{\eta_k\,\|q(z^k,\eta_k)\|}{\hat{c}\|H(z^k)\|^{1+\delta}}}{1-\frac{\eta_k\,\|q(z^k,\eta_k)\|}{e^{-\frac{\psi(z^0)}{2m}}\hat{c}\,\|H(z^k)\|^{\frac{\zeta}{m}+\delta}}}\right) = 0.$$

Altogether, we obtain

$$\limsup_{k\to\infty} \frac{\psi(z^k+\eta_k d^k)-\psi(z^k)}{\eta_k} \le -2(\zeta-m). \qquad \square$$

Having this, the assertion follows analogously as in the proof of Lemma 4.10.

This allows us to prove a superlinear convergence rate.

Theorem 5.2. *Let Assumption* 3.2 *hold and* $\{(x^k, \lambda^k, w^k)\}$ *be generated by Algorithm* 1. *Assume* $\zeta \in (m, 2m)$ *and that the sequence* $\{\eta_k\}$ *satisfies* (18) *with* $\delta \in \left(0, 2-\frac{\zeta}{m}\right)$. *Then,* $\{\|H(z^k)\|\}_{k\in\mathbb{N}}$ *converges superlinearly to zero with order at least* $1+\delta$.

Proof. By Lemma 5.1 the stepsize is for sufficiently large $k \in \mathbb{N}$ given by

$$t_k = \eta_k = 1 - \hat{c}\|H(z^k)\|^\delta \quad \text{or by} \quad t_k = 1.$$

In the first case, with $\|q(z^k, t)\| = \mathcal{O}(\|H(z^k)\|^2)$ by (16) in Lemma 4.8, we obtain from (15)

$$\begin{aligned}\|H(z^k+t_kd^k)\| =& \|(1-\eta_k)\,H(z^k)+\eta_k\,q(z^k,\eta_k)\|\\ \le& \hat{c}\,\|H(z^k)\|^{\delta+1}+(1-\hat{c}\,\|H(z^k)\|^\delta)\,\|q(z^k,\eta_k)\| = \mathcal{O}(\|H(z^k)\|^{\delta+1}).\end{aligned}$$

Whenever we have $t_k = 1$, we obtain

$$\|H(z^k+t_kd^k)\| = \|q(z^k,1)\| = \mathcal{O}(\|H(z^k)\|^2).$$

This proves superlinear convergence and the order of convergence is at least $1+\delta$ with $\delta \in \left(0, 2-\frac{\zeta}{m}\right)$. $\square$

Let us stress, that for superlinear convergence the convergence order of $\|q(z^k, t_k)\|$ is crucial. If this is only linear, which would be in general the case, if we have only linear constraints, resulting in $q_2(z^k, t_k) = \sigma_k\mu_k$, and the sequence $\{\sigma_k\}$ does not converge to zero (in contrast to (A3)), we would get from (15) only linear convergence, even if the stepsize is always $t_k = 1$. A constant value of $\{\sigma_k\}$ was for example used in the implementation in [1].

6. Inexact potential reduction algorithm

In [4] an inexact version of the potential reduction algorithm for GNEPs was discussed. This allows in (S.2) any direction $d \in \mathbb{R}^{n+2m}$ satisfying

$$\|JH(z^k)d + H(z^k) - \sigma_k \mu_k \, a\| \leq \delta_k \|H(z^k)\|,$$

with a sequence $\{\delta_k\}$ converging to zero. We will now discuss the convergence analysis for this inexact version, without repeating all the details.

Defining the residual vector $\rho^k := \begin{pmatrix} \rho_1^k \\ \rho_2^k \\ \rho_3^k \end{pmatrix} \in \mathbb{R}^{n+m+m}$ by

$$\rho^k := JH(z^k)d^k + H(z^k) - \sigma_k \mu_k \, a,$$

we have $\|\rho^k\| \leq \delta_k \|H(z^k)\|$. With

$$\Delta r^k := J_x g(x^k) J_x F(x^k, \lambda^k)^{-1} \rho_1^k - \rho_2^k + (\Lambda^k)^{-1} \rho_3^k,$$

we obtain, analogous to (6), the linear equation system

$$\left(M^k \Lambda^k + W^k\right) \frac{d_\lambda^k}{\lambda^k} = r^k + \Delta r^k.$$

Inspecting our convergence analysis, we see that the combination of the convergence proofs for the inexact version of the potential reduction algorithm from [4, Theorem 4.6] and the linesearch with non-fixed $\{\eta_k\}$ from Theorem 2.1 yield the assertion of Theorem 2.1 also for the inexact version. Then, Lemma 4.1 holds with the same proof. Lemma 4.2 only holds, if we make an additional assumption.

(A7) For all indices $i, j \in \{1, \ldots, m\}$ with $g_i \equiv g_j$ and $w_i^k = w_j^k$, we have equal residuals: $(\rho_2^k)_i = (\rho_2^k)_j$.

This is an assumption on the used solver for the inexact equation system in (S.2). The solver must treat equal components, corresponding to repeated constraints, equally. This will not be satisfied for all solvers, but guarantees that Lemma 4.2 still holds.

We can also get Corollary 4.3 by assuming, additional to (A7), that the solver guarantees for all indices $i, j \in \{1, \ldots, m\}$ with $g_i \equiv g_j$ and $\lambda_i^k = \lambda_j^k$ that

$$(\rho_3^k)_i = (\rho_3^k)_j.$$

Since our analysis is not based on Corollary 4.3, we only need the additional assumption (A7) for the further results.

Lemma 4.4 is not affected by using the inexact method, and the crucial part of our analysis is Lemma 4.5, which exploits the structure of r^k. Here, we will need an additional assumption on the sequence $\{\delta_k\}$ of the inexactness parameter.

(A8) There is a constant $\tilde{c} > 0$ such that $\dfrac{\delta_k}{\min\limits_{j=1,\ldots,2m} H_{n+j}(z^k)} \leq \tilde{c}$.

This can always be satisfied, since under (A6) the exact solution of the equation system exists, meaning that δ_k can be chosen arbitrarily small. Now, we look at the proof of Lemma 4.5:

Step 1(a) holds unchanged. In Step 1(b), we now have

$$\begin{aligned}
&\left|r^k_{s_1} + \Delta r^k_{s_1} - r^k_{s_2} - \Delta r^k_{s_2}\right| = \left|\frac{\sigma_k\mu_k + (\rho^k_3)_{s_1}}{\lambda^k_{s_1}} - \frac{\sigma_k\mu_k + (\rho^k_3)_{s_2}}{\lambda^k_{s_2}}\right| \\
&\le \frac{\sigma_k\mu_k + |(\rho^k_3)_{s_1}| + |(\rho^k_3)_{s_2}|}{\min\{\lambda^k_{s_1}, \lambda^k_{s_2}\}} = \frac{(\sigma_k\mu_k + |(\rho^k_3)_{s_1}| + |(\rho^k_3)_{s_2}|)w^k_{s_1}}{\min\{\lambda^k_{s_1} w^k_{s_1}, \lambda^k_{s_2} w^k_{s_2}\}} \\
&\le \frac{(\sigma_k\mu_k + 2\delta_k\|H(z^k)\|)w^k_{s_1}}{\min\{\lambda^k_{s_1} w^k_{s_1}, \lambda^k_{s_2} w^k_{s_2}\}}.
\end{aligned}$$

Using (A3) and (A8), this term is bounded by $w^k_{s_1}(c + 2\tilde{c})\|H(z^k)\|$, and we can conclude as before in Step 1(b). In all the remaining subcases, we exploit that under (A8) and (A1) we have

$$\|(\Lambda^k)^{-1}\rho^k_3\| = \|(\Lambda^k)^{-1}(W^k)^{-1}W^k\rho^k_3\| \le \sqrt{m}\,\frac{\delta_k\|H(z^k)\| \max\limits_{j=1,\dots,m} w^k_j}{\min\limits_{j=1,\dots,m} H_{n+m+j}(z^k)} = O(\|H(z^k)\|).$$

Hence, using (A1) and (A4)

$$\|\Delta r^k\| = \|J_x g(x^k) J_x F(x^k, \lambda^k)^{-1}\rho^k_1 - \rho^k_2 + (\Lambda^k)^{-1}\rho^k_3\| = O(\|H(z^k)\|).$$

Exploiting this allows the same conclusions, and hence, under the additional assumption (A8), Lemma 4.5 also holds for the inexact version. With these two Lemmata we can also prove Theorem 4.6 and Lemma 4.7. The assertion of Lemma 4.8 holds by adding ρ^k in the definition of the term $q(z^k, t)$, and exploiting that by $\|\rho^k\| \le \delta_k\|H(z^k)\|$ and (A8), we have $\|\rho^k\| = O(\|H(z^k)\|^2)$. Lemma 4.9 is independent of the inexactness, and Lemma 4.10 holds, since we only exploit the formula and convergence rate from Lemma 4.8. Corollary 4.11 and our main convergence result Theorem 4.12 then also hold. Hence, we obtain also the linear convergence rate for the inexact potential reduction algorithm. Furthermore, the superlinear convergence results of Section 5 can also be shown for the inexact version under the additional assumptions (A7) and (A8), by using the adapted results of Section 4.

7. A numerical example

In this section we want to show for one example that the linear or superlinear convergence rates can be observed numerically.

Example 7.1. Consider a 3 player LQGNEP with shared constraints defined by

$$\begin{aligned}
&\text{Player 1:} \quad \min_{x^1} \frac{1}{2}(x^1-1)^2 - \frac{1}{2}x^1x^2 \quad \text{s.t.} \quad x^1 \ge 0,\ x^1+x^2+x^3 \le 1,\\
&\text{Player 2:} \quad \min_{x^2} \frac{1}{2}(x^2-1)^2 + \frac{1}{2}x^1x^2 \quad \text{s.t.} \quad x^2 \ge 0,\ x^1+x^2+x^3 \le 1,\\
&\text{Player 3:} \quad \min_{x^3} \frac{1}{2}(x^3-1)^2 \quad \text{s.t.} \quad x^3 \ge 0,\ x^1+x^2+x^3 \le 1.
\end{aligned}$$

This example has infinitely many solutions, and for all of them the shared constraint is active. There is only one solution with equal multipliers for the repeated constraint. Using the coercivity and the relaxed EMFCQ mentioned in Remark 3.3, we can show that (A1) holds. Since we have an LQGNEP, we can check (A4), and that the P_0 property holds for the matrix M, implying (A6) and the second part of (A5), see Remark 3.3. Further, choosing equal starting values for the repeated constraints in λ^0 and w^0, we get (A2), and we will have positive multipliers of the active constraint at the solution (we can compute the unique solution with equal multipliers for repeated constraints analytically, or we can observe this fact numerically). The nonsingularity condition in (A5) can be checked, having the active set. As starting values we used $x^0 = (0,0,0)^\top$ and all components of λ^0 and w^0 are set to 10.

For a preliminary implementation we use the parameters $\gamma = 0.01$ and $\zeta = \lceil \frac{3m}{2} \rceil$. Further, we set $\sigma_0 = 0.1$ and

$$\sigma_{k+1} = \min \left\{ \sigma_k, m \cdot \min_{j=1,\dots,2m} H_{n+j}(z^k) \right\},$$

which satisfies (A3). As stopping criterion we use $\|H(z^k)\| \leq 10^{-9}$.

In our comparison of two algorithms, the first one has fixed values $\eta_k = 0.5$, and the second one uses $\{\eta_k\}$ as in (18) with $\hat{c} = 1$ and $\delta = 0.4 \in (0, 2 - \lceil \frac{3m}{2} \rceil / m)$. Both variants converge to the unique generalized Nash equilibrium with equal non-zero multipliers $\bar{x} = \frac{1}{13}(6,2,5)^\top$.

On the left hand side of Table 7.1 on the next page we can observe linear convergence of $\{\|H(z^k)\|\}_{k\in\mathbb{N}}$ to zero for the algorithm with fixed linesearch parameter η_k. Starting from iteration 6, only the stepsize $t_k = \eta_k = 0.5$ is accepted, which is expected from Lemma 4.10, and the convergence rate is linear, c.f., Theorem 4.12.

On the right hand side, we observe that the first 6 iterations are the same as for the method with constant η_k. But then the convergence rate improves, since the accepted stepsize is equal to $\eta_k > 0.5$, as expected by Lemma 5.1. We observe superlinear convergence as expected from Theorem 5.2.

8. Conclusion and outlook

There are two main contributions of this paper. First, the proof of Q-linear convergence of $\{\|H(z^k)\|\}_{k\in\mathbb{N}}$ and R-linear convergence of $\{\text{dist}(z^k, S)\}_{k\in\mathbb{N}}$ for the exact version of the potential reduction algorithm for GNEPs with a fixed linesearch parameter η under suitable assumptions. Second, the superlinear convergence, if we choose this parameter as a sequence $\{\eta_k\}$ converging at a suitable rate to 1. Both convergence rates can also be shown for the inexact version of the potential reduction algorithm if the inexactness parameter vanishes fast enough and a technical requirement for the inexact solver of the linear equation system is satisfied.

In this paper we focused on theoretical results. Numerical results of the potential reduction algorithm for GNEPs with fixed parameters η_k and σ_k can be found in [1] and [3]. The numerical realization of an algorithm with balanced choices of the parameter $\zeta \in (m, 2m)$, the sequences $\{\sigma_k\}_{k\in\mathbb{N}}$ satisfying (A3), and $\{\eta_k\}_{k\in\mathbb{N}}$ fulfilling (18), is left for future research.

	fixed parameter η_k				adapted parameter η_k			
iter	σ_k	η_k	t_k	$\|H\|$	σ_k	η_k	t_k	$\|H\|$
0	1.00e-01	0.500	1.000	2.46e+02	1.00e-01	0.500	1.000	2.46e+02
1	1.00e-01	0.500	1.000	6.90e+01	1.00e-01	0.500	1.000	6.90e+01
2	1.00e-01	0.500	1.000	1.66e+01	1.00e-01	0.500	1.000	1.66e+01
3	1.00e-01	0.500	1.000	4.01e+00	1.00e-01	0.500	1.000	4.01e+00
4	1.00e-01	0.500	1.000	1.11e+00	1.00e-01	0.500	1.000	1.11e+00
5	1.00e-01	0.500	1.000	2.60e-01	1.00e-01	0.500	1.000	2.60e-01
6	1.00e-01	0.500	0.500	1.48e-01	1.00e-01	0.500	0.500	1.48e-01
7	6.11e-02	0.500	0.500	8.32e-02	5.84e-02	0.535	0.535	7.90e-02
8	3.44e-02	0.500	0.500	4.53e-02	2.56e-02	0.638	0.638	3.36e-02
9	1.84e-02	0.500	0.500	2.40e-02	7.54e-03	0.743	0.743	1.02e-02
10	9.52e-03	0.500	0.500	1.24e-02	1.30e-03	0.840	0.840	1.91e-03
11	4.85e-03	0.500	0.500	6.36e-03	1.10e-04	0.918	0.918	1.74e-04
12	2.45e-03	0.500	0.500	3.22e-03	3.46e-06	0.969	0.969	5.65e-06
13	1.23e-03	0.500	0.500	1.62e-03	2.75e-08	0.992	0.992	4.53e-08
14	6.16e-04	0.500	0.500	8.14e-04	3.18e-11	0.999	0.999	5.23e-11
15	3.08e-04	0.500	0.500	4.08e-04				
16	1.54e-04	0.500	0.500	2.04e-04				
17	7.72e-05	0.500	0.500	1.02e-04				
18	3.86e-05	0.500	0.500	5.10e-05				
19	1.93e-05	0.500	0.500	2.55e-05				
20	9.65e-06	0.500	0.500	1.28e-05				
21	4.83e-06	0.500	0.500	6.38e-06				
22	2.41e-06	0.500	0.500	3.19e-06				
23	1.21e-06	0.500	0.500	1.60e-06				
24	6.03e-07	0.500	0.500	7.98e-07				
25	3.02e-07	0.500	0.500	3.99e-07				
26	1.51e-07	0.500	0.500	1.99e-07				
27	7.54e-08	0.500	0.500	9.97e-08				
28	3.77e-08	0.500	0.500	4.98e-08				
29	1.89e-08	0.500	0.500	2.49e-08				
30	9.43e-09	0.500	0.500	1.25e-08				
31	4.71e-09	0.500	0.500	6.23e-09				
32	2.36e-09	0.500	0.500	3.12e-09				
33	1.18e-09	0.500	0.500	1.56e-09				
34	5.89e-10	0.500	0.500	7.79e-10				

Table 7.1: Comparison of local convergence rates for potential reduction algorithms

Further, it is an interesting question, how we can exploit the convergence of the fractions in Theorem 4.6 as indicators to identify the active set of a solution. It might be a good idea to remove the non-active inequalities at some point in order to reduce the dimension of the problem. Furthermore, if one would have a different algorithm that can exploit the knowledge of the active set, this could be applied to get faster local convergence, and one might be able to design a new hybrid method.

Acknowledgments. The author wants to thank two anonymous referees for their very helpful reviews and the detailed comments that helped to improve the paper considerably.

References

[1] A. Dreves: *Globally Convergent Algorithms for the Solution of Generalized Nash Equilibrium Problems*, Dissertation, University of Würzburg (2012), available online at http://opus.bibliothek.uni-wuerzburg.de/volltexte/2012/6982.

[2] A. Dreves: *Improved error bound and a hybrid method for generalized Nash equilibrium problems*, Comput. Optimization Appl. 65/2 (2016) 431–448.

[3] A. Dreves, F. Facchinei, A. Fischer, M. Herrich: *A new error bound result for generalized Nash equilibrium problems and its algorithmic application*, Comput. Optimization Appl. 59 (2014) 63–84.

[4] A. Dreves, F. Facchinei, C. Kanzow, S. Sagratella: *On the solution of the KKT conditions of generalized Nash equilibrium problems*, SIAM J. Optimization 21 (2011) 1082–1108.

[5] A. Dreves, A. von Heusinger, C. Kanzow, M. Fukushima: *A globalized Newton method for the computation of normalized Nash equilibria*, J. Global Optimization 56 (2013) 327–340.

[6] A. Dreves, S. Sagratella: *Nonsingularity and stationarity results for quasi-variational inequalities*, J. Optimization Theory Appl. 185/3 (2020) 711–743.

[7] A. Dreves, N. Sudermann-Merx: *Solving linear generalized Nash equilibrium problems numerically*, Optimization Methods Software 31/5 (2016) 1036–1063.

[8] F. Facchinei, C. Kanzow: *Penalty methods for the solution of generalized Nash equilibrium problems*, SIAM J. Optimization 20 (2010) 2228–2253.

[9] F. Facchinei, C. Kanzow: *Generalized Nash equilibrium problems*, Ann. Oper. Res. 1755 (2010) 177–211.

[10] F. Facchinei, J.-S. Pang: *Nash equilibria: The variational approach*, in: *Convex Optimization in Signal Processing and Communications*, Y. Eldar et al. (eds.), Cambridge University Press, Cambridge (2009) 443–493.

[11] F. Facchinei, A. Fischer, M. Herrich: *A family of Newton methods for nonsmooth constrained systems with nonisolated solutions*, Math. Methods Oper. Res. 77 (2013) 433–443.

[12] F. Facchinei, A. Fischer, M. Herrich: *An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions*, Math. Programming 146 (2014) 1–36.

[13] F. Facchinei, A. Fischer, V. Piccialli: *Generalized Nash equilibrium problems and Newton methods*, Math. Programming 117 (2009) 163–194.

[14] A. Fischer, M. Herrich, A. F.Izmailov, M. V. Solodov: *A globally convergent LP-Newton method*, SIAM J. Optimization 26/4 (2016) 2012–2033.

[15] A. Fischer, M. Herrich, K. Schönefeld: *Generalized Nash equilibrium problems – recent advances and challenges*, Pesquisa Operacional 34 (2014) 521–588.

[16] M. Fukushima: *Restricted generalized Nash equilibria and controlled penalty algorithm*, Comput. Manage. Sci. 8 (2011) 201–218.

[17] A. F. Izmailov, M. V. Solodov: *On error bounds and Newton-type methods for generalized Nash equilibrium problems*, Comput. Optimization Appl. 59 (2014) 201–218.

[18] C. Kanzow, V. Karl, D. Steck, G. Wachsmuth: *The multiplier-penalty method for generalized Nash equilibrium problems in Banach spaces*, SIAM J. Optimization 29/1 (2019) 767–793.

[19] R. D. C. Monteiro, J.-S. Pang: *A potential reduction Newton method for constrained equations*, SIAM J. Optimization 9 (1999) 729–754.

[20] J.-S. Pang, M. Fukushima: *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Comput. Manage. Sci. 2 (2005) 21–56.

[21] T. Wang, R. D. C. Monteiro, J.-S. Pang: *An interior point potential reduction method for constrained equations*, Math. Programming 74 (1996) 159–195.