Kybernetika

Matus Benko; Helmut Gfrerer

An SQP method for mathematical programs with complementarity constraints with strong convergence properties

Kybernetika, Vol. 52 (2016), No. 2, 169–208

Persistent URL: http://dml.cz/dmlcz/145769

KYBERNETIKA — VOLUME *52* (2016), NUMBER 2, PAGES 169–208

AN SQP METHOD FOR MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS WITH STRONG CONVERGENCE PROPERTIES

Matus Benko and Helmut Gfrerer

We propose an SQP algorithm for mathematical programs with complementarity constraints which solves at each iteration a quadratic program with linear complementarity constraints. We demonstrate how strongly M-stationary solutions of this quadratic program can be obtained by an active set method without using enumeration techniques. We show that all limit points of the sequence of iterates generated by our SQP method are at least M-stationary.

Keywords: SQP method, active set method, mathematical program with complementarity constraints, strong M-stationarity

Classification: 49M37, 90C26, 90C33, 90C55

1. INTRODUCTION

Consider the following *mathematical program with complementarity constraints* (MPCC)

$$\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f(x) & \\ \text{subject to} & h_i(x)=0 & i\in E, \\ & g_i(x)\le 0 & i\in I, \\ & 0\le G_i(x)\perp H_i(x)\ge 0 & i\in C, \end{array} \tag{1}$$

with continuously differentiable functions f, $h_i, i \in E$, $g_i, i \in I$, $G_i, H_i, i \in C$ and finite index sets E, I and C. The notation $0 \le u \perp v \ge 0$ for two vectors $u, v \in \mathbb{R}^n$ is a shortcut for $u \ge 0, v \ge 0, u^T v = 0$. MPCCs are more specialized class of *mathematical programs with equilibrium constraints* (MPECs). For more background and several applications we refer the reader to the textbooks [22, 25].

Theoretically, MPCCs can be viewed as standard nonlinear optimization problems, but due to the complementarity constraints, many of the standard constraint qualifications of nonlinear programming are violated at any feasible point. This makes it necessary, both from a theoretical and numerical point of view, to consider special tailored algorithms for solving MPCCs. Recent numerical methods follow different directions. In [7, 8] the direct application of an SQP solver was investigated whereas the

DOI: 10.14736/kyb-2016-2-0169

application of interior point methods was considered in [3, 4, 20]. Penalization techniques are suggested in [12, 24, 30]. Another class of methods deals with piece-wise decomposition [9, 11, 14, 15, 22, 23, 33, 34]. Also smoothing, lifting and relaxation methods have been suggested in order to deal with the inherent difficulties of MPCCs in [1, 2, 5, 13, 16, 17, 21, 27, 29, 31, 32].

All these approaches have in common that one can prove only convergence to weakly stationary or C-stationary points unless some strong assumptions are made. For the two methods from [16, 17] the stronger and interesting property can be shown that they converge to M-stationary points. However, the subproblems which are to be solved in these methods, do not satisfy a constraint qualification resulting in the effect that approximate solutions of the subproblems only converge to weakly stationary points. For a more detailed analysis of convergence properties of relaxation methods we refer to the recent paper [18].

In this paper, we carry over a well known SQP-method from nonlinear programming to MPCCs. The main task of our method is to solve in each iteration step a quadratic program with linear complementarity constraints. We present an active set method, very similar to the active set method for quadratic programming [6], which computes at least a *strongly M-stationary* solution of the subproblems. The concept of strong M-stationarity was introduced in the recent paper by Gfrerer [10]. This active set method is based on an active set method used to show the existence of strongly M-stationary solutions for MPCCs [10, Theorem 4.3]. Surprisingly, the inherent combinatorial structure of M-stationarity can be resolved by this active set method and one does not depend on enumeration techniques.

Then we compute the next iterate by reducing a certain merit function along some polygonal line which is given by the solution procedure for the subproblem. Then we show that every limit point of the generated sequence is at least M-stationary, provided that the solutions of the subproblems remain bounded. Numerical tests indicate that our method behaves very reliable.

A short outline of this paper is as follows. In section 2 we recall the basic stationarity concepts for MPCCs as well as the concept of strong M-stationarity. In section 3 we recall an outline of an SQP method for nonlinear programming and we provide an outline of our SQP method for MPCCs. In section 4 we describe an active set method for solving the auxiliary problem occurring in every iteration of our SQP method. This auxiliary problem is a quadratic program with linear complementarity constraints. We prove the finiteness of this algorithm and we summarize some of the properties of quantities computed during the algorithm. In section 5 we describe how the next iterate is computed by means of the solution of the auxiliary problem. Further, we consider the convergence of the overall algorithm. Section 6 is a summary of numerical results we obtained by implementing our algorithm in MATLAB and testing it on the MacMPEC collection of MPECs maintained by Leyffer [19].

In what follows we use the following notation. Given a set M we denote by $\mathcal{P}(M) := \{(M_1, M_2) \mid M_1 \cup M_2 = M,\ M_1 \cap M_2 = \emptyset\}$ the set of all partitions of M. Further, for a real number a we use the notation $(a)^+ := \max(0, a)$, $(a)^- := \min(0, a)$.

2. STATIONARY POINTS FOR MPCCS

Given a point $\bar{x}$ feasible for (1) we define the following index sets

$$\begin{aligned} I^g(\bar{x}) &:= \{i \in I \mid g_i(\bar{x}) = 0\}, \\ I^{0+}(\bar{x}) &:= \{i \in C \mid G_i(\bar{x}) = 0 < H_i(\bar{x})\}, \\ I^{+0}(\bar{x}) &:= \{i \in C \mid G_i(\bar{x}) > 0 = H_i(\bar{x})\}, \\ I^{00}(\bar{x}) &:= \{i \in C \mid G_i(\bar{x}) = 0 = H_i(\bar{x})\}. \end{aligned}$$

Further we call a triple of index sets $J = (J_g, J_G, J_H)$ with $J_g \subset I^g(\bar{x})$, $J_G \subset I^{0+}(\bar{x}) \cup I^{00}(\bar{x})$, $J_H \subset I^{+0}(\bar{x}) \cup I^{00}(\bar{x})$ an *MPEC working set* with respect to $\bar{x}$, if $J_G \cup J_H = C$, $|E| + |J_g| + |J_G| + |J_H|$ is equal to the rank of the family of vectors

$$\begin{aligned} \{\nabla h_i(\bar{x}) \mid i \in E\} \quad &\cup \quad \{\nabla g_i(\bar{x}) \mid i \in I^g(\bar{x})\} \cup \{\nabla G_i(\bar{x}) \mid i \in I^{0+}(\bar{x}) \cup I^{00}(\bar{x})\} \\ &\cup \quad \{\nabla H_i(\bar{x}) \mid i \in I^{+0}(\bar{x}) \cup I^{00}(\bar{x})\} \end{aligned}$$

and the family of vectors

$$\{\nabla h_i(\bar{x}) \mid i \in E\} \cup \{\nabla g_i(\bar{x}) \mid i \in J_g\} \cup \{\nabla G_i(\bar{x}) \mid i \in J_G\} \cup \{\nabla H_i(\bar{x}) \mid i \in J_H\}$$

is linearly independent.

In contrast to nonlinear programming there exist a lot stationarity concepts for MPCCs.

Definition 2.1. Let $\bar{x}$ be feasible for (1). Then $\bar{x}$ is called

1. *weakly stationary*, if there are multipliers $\lambda_i^g, i \in I$, $\lambda_i^h, i \in E$, $\lambda_i^G, \lambda_i^H, i \in C$ such that

$$\nabla f(\bar{x}) + \sum_{i \in E} \lambda_i^h \nabla h_i(\bar{x}) + \sum_{i \in I} \lambda_i^g \nabla g_i(\bar{x}) - \sum_{i \in C} \left(\lambda_i^G \nabla G_i(\bar{x}) + \lambda_i^H \nabla H_i(\bar{x})\right) = 0$$

   and

$$\lambda_i^g \geq 0, \lambda_i^g g_i(\bar{x}) = 0, i \in I, \qquad \lambda_i^G = 0, i \in I^{+0}(\bar{x}), \quad \lambda_i^H = 0, i \in I^{0+}(\bar{x}).$$

2. *C-stationary*, if it is weakly stationary and

$$\lambda_i^G \lambda_i^H \geq 0, i \in I^{00}(\bar{x}).$$

3. *M-stationary*, if it is C-stationary and

$$\text{either } \lambda_i^G, \lambda_i^H > 0 \text{ or } \lambda_i^G \lambda_i^H = 0, i \in I^{00}(\bar{x}).$$

4. *strongly M-stationary*, if it is M-stationary and there exists a MPEC working set J_g, J_G, J_H such that

$$\lambda_i^G, \lambda_i^H \geq 0, i \in J_G \cap J_H.$$

5. *S-stationary*, if it is M-stationary and

$$\lambda_i^G, \lambda_i^H \geq 0, i \in I^{00}(\bar{x}).$$

Strong M-stationarity was introduced in the recent paper by Gfrerer [10], whereas the other stationarity concepts are very common in the literature, see e. g. Scheel and Scholtes [28]. Obviously, the following implications hold:

$$\text{S-stationarity} \Rightarrow \text{M-stationarity} \Rightarrow \text{C-stationarity} \Rightarrow \text{weak stationarity},$$

$$\text{strong M-stationarity} \Rightarrow \text{M-stationarity}$$

Further, S-stationarity implies strong M-stationarity provided at least one MPEC working set exists.

Note that the S-stationarity conditions are nothing else than the Karush–Kuhn–Tucker conditions for the problem (1). Unfortunately, a local minimizer is S-stationary only under some comparatively strong constraint qualification, e. g. that the gradients of the active constraints are linearly independent. On the other hand, a local minimizer is strongly M-stationary under very weak constraint qualifications. Recall that the *contingent* (also *Bouligand* or *tangent*) cone to a closed set $\Omega \subset \mathbb{R}^n$ at $u \in \Omega$ is defined by

$$T_\Omega(u) := \{d \in \mathbb{R}^n \mid \exists (d_k) \to d, \exists (\tau_k) \downarrow 0 : u + \tau_k d_k \in \Omega \forall k\}.$$

Definition 2.2. Let $\bar{x}$ be feasible for (1). We say that the MPEC Guignard constraint qualification (MPEC-GCQ) holds at $\bar{x}$ if the polar cone of the contingent cone to the feasible set of (1) at $\bar{x}$ equals the polar cone of the cone

$$\begin{aligned}\{d \in \mathbb{R}^n \quad | \quad & (\nabla h_i)^T(\bar{x})d = 0,\ i \in E,\\ & (\nabla g_i)^T(\bar{x})d \leq 0,\ i \in I^g(\bar{x}),\\ & (\nabla G_i)^T(\bar{x})d = 0,\ i \in I^{0+}(\bar{x}),\\ & (\nabla H_i)^T(\bar{x})d = 0,\ i \in I^{+0}(\bar{x}),\\ & 0 \leq (\nabla G_i)^T(\bar{x})d \perp (\nabla H_i)^T(\bar{x})d \geq 0,\ i \in I^{00}(\bar{x})\}.\end{aligned}$$

Theorem 2.3. (c.f. Gfrerer [10, Theorem 4.3, Theorem 3.9]) Let $\bar{x}$ be a local minimizer of the MPCC (1) at which MPEC-GCQ holds and assume that an MPEC working set exists. Then $\bar{x}$ is strongly M-stationary.

The assumption, that one MPEC working set exists, is fulfilled, if there are index sets $(\tilde{J}_G, \tilde{J}_H) \in \mathcal{P}(I^{00}(\bar{x}))$ such that the family of gradients

$$\{\nabla h_i(\bar{x}),\, i \in E\} \cup \{\nabla G_i(\bar{x}),\, i \in I^{0+}(\bar{x}) \cup \tilde{J}_G\} \cup \{\nabla H_i(\bar{x}),\, i \in I^{+0}(\bar{x}) \cup \tilde{J}_H\}$$

is linearly independent and this seems to be a rather weak assumption.

Note that [10, Theorem 4.3] was constructively proved by some active set method and our algorithm is based on this procedure.

3. ON SQP METHODS IN MATHEMATICAL PROGRAMMING

We recall the structure of well-known SQP method for solving nonlinear optimization problems

$$\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f(x) & \\ \text{subject to} & h_i(x)=0 & i\in E, \\ & g_i(x)\le 0 & i\in I. \end{array} \tag{2}$$

If x_k denotes the kth estimate for the optimal solution and B_k a symmetric positive definite matrix the following quadratic programming subproblem is solved

$$\begin{array}{lll} \min\limits_{(s,\delta)\in\mathbb{R}^{n+1}} & \frac{1}{2}s^TB_ks+(\nabla f(x_k))^Ts+\frac{1}{2}\rho\delta^2 & \\ \text{subject to} & (1-\delta)h_i(x_k)+(\nabla h_i(x_k))^Ts=0 & i\in E, \\ & (1-\beta_i^g\delta)g_i(x_k)+(\nabla g_i(x_k))^Ts\le 0 & i\in I, \\ & -\delta\le 0, & \end{array} \tag{3}$$

where

$$\beta_i^g=\begin{cases}1 & \text{if } g_i(x_k)>0\\ 0 & \text{else.}\end{cases}$$

The additional variable δ is introduced to avoid inconsistent constraints and to have a feasible point $(s,\delta)=(0,1)$ at hand. The penalty parameter ρ has to be chosen sufficiently large such that δ is close to 0 at a solution of the problem.

Let us denote the unique solution of (3) by (s_k,δ_k). If $\delta_k>\zeta$ for some $\zeta\in(0,1)$ we increase ρ and solve problem (3) again. If this loop fails within some given upper bound for ρ the calculation finishes because it is assumed that the constraints of (2) have no feasible point. Otherwise the new iterate x_{k+1} is given by $x_{k+1}=x_k+\alpha_ks_k$ where α_k is a positive step length that is chosen by a line search procedure to give a reduction $\Phi(x_{k+1})<\Phi(x_k)$ where Φ is a suitable merit function, e. g. the ℓ_1 penalty function

$$\Phi(x)=f(x)+\sum_{i\in E}\sigma_i^h|h_i(x)|+\sum_{i\in I}\sigma_i^g(g_i(x))^+$$

with appropriately chosen penalty parameters $\sigma=(\sigma^h,\sigma^g)$.

The structure of our SQP method is very similar to the above procedure. The quadratic auxiliary problem which we have to solve at each iterate is given by

$$\begin{array}{lll} \min\limits_{(s,\delta)\in\mathbb{R}^{n+1}} & \frac{1}{2}s^TB_ks+(\nabla f(x_k))^Ts+\rho(\frac{1}{2}\delta^2+\delta) & \\ \text{subject to} & (1-\delta)h_i(x_k)+(\nabla h_i(x_k))^Ts=0 & i\in E, \\ & (1-\beta_i^g\delta)g_i(x_k)+(\nabla g_i(x_k))^Ts\le 0 & i\in I, \\ & 0\le(1-\beta_i^G\delta)G_i(x_k)+(\nabla G_i(x_k))^Ts & \\ & \ \perp\ (1-\beta_i^H\delta)H_i(x_k)+(\nabla H_i(x_k))^Ts\ge 0 & i\in C, \\ & -\delta\le 0, & \end{array} \tag{4}$$

where the vector $\beta=(\beta^g,\beta^G,\beta^H)$ is chosen such that the point $(s,\delta)=(0,1)$ is feasible. Note that we add to the quadratic penalty term $\frac{1}{2}\rho\delta^2$, as normally used in SQP methods, the term $\rho\delta$, which acts like an exact penalty term and forces δ to be zero at the solution also for moderate values of ρ. For the convergence proof we use the property of the new iterate that we have a decrease $\Phi(x_{k+1}) < \Phi(x_k)$ with respect to the merit function

$$\begin{aligned}\Phi(x) \quad := \quad & f(x) + \sum_{i\in E} \sigma_i^h |h_i(x)| + \sum_{i\in I} \sigma_i^g (g_i(x))^+ \\ & + \sum_{i\in C} \sigma_i^C (|\min\{G_i(x), H_i(x)\}| - (\max\{G_i(x), H_i(x)\})^-).\end{aligned}$$

The main difference to the SQP method for nonlinear programs is that we perform the line search not along some single direction, but along some polygonal line $s_k^0, s_k^1, \dots, s_k^{N_k}$ connecting the solutions of convex subproblems of (4).

4. SOLVING THE AUXILIARY PROBLEM

In this section, we describe an algorithm for solving quadratic problems with complementarity constraints of the type

$$\begin{array}{lll}\min\limits_{(s,\delta)\in\mathbb{R}^{n+1}} & \frac{1}{2}s^T B s + (\nabla f)^T s + \rho(\frac{1}{2}\delta^2 + \delta) & QPCC(\beta,\rho) \\ \text{subject to} & (1-\delta)h_i + (\nabla h_i)^T s = 0 & i\in E, \\ & (1-\beta_i^g\delta)g_i + (\nabla g_i)^T s \le 0 & i \in I, \\ & 0 \le (1-\beta_i^G\delta)G_i + (\nabla G_i)^T s \ \perp \ (1-\beta_i^H\delta)H_i + (\nabla H_i)^T s \ge 0 & i\in C, \\ & -\delta \le 0. & \end{array} \tag{5}$$

Here the vector $\beta = (\beta^g, \beta^G, \beta^H) \in \{0,1\}^{|I|+2|C|} =: \mathcal{B}$ is chosen at the beginning of the algorithm such that some feasible point is known in advance, e. g. $(s,\delta) = (0,1)$. During the course of the solution procedure components β_i^H, β_i^G with $G_i, H_i > 0$ can change but only in such a way that the current iterate (s,δ) remains feasible. The parameter ρ has to be chosen sufficiently large and acts like a penalty parameter forcing δ to be near zero at the solution. B is a symmetric positive definite $n\times n$ matrix, ∇f, ∇h_i, ∇g_i, ∇G_i, ∇H_i denote vectors in $\mathbb{R}^n$ and h_i, g_i, G_i, H_i are real numbers. We use this notation to point out to the reader that the problem (5) is auxiliary problem (4) which we have to solve at each iterate.

Given a feasible point (s,δ) for the problem $QPCC(\beta,\rho)$ we define the active inequality constraints and the active complementarity constraints, respectively, by

$$\begin{aligned} I^g(s,\delta,\beta) &:= \{i\in I \,|\, (1-\beta_i^g\delta)g_i + (\nabla g_i)^T s = 0\} \cup \{0 \,|\, \text{if } \delta = 0\}, \\ I^{0+}(s,\delta,\beta) &:= \{i\in C \,|\, (1-\beta_i^G\delta)G_i + (\nabla G_i)^T s = 0 < (1-\beta_i^H\delta)H_i + (\nabla H_i)^T s\}, \\ I^{+0}(s,\delta,\beta) &:= \{i\in C \,|\, (1-\beta_i^G\delta)G_i + (\nabla G_i)^T s > 0 = (1-\beta_i^H\delta)H_i + (\nabla H_i)^T s\}, \\ I^{00}(s,\delta,\beta) &:= \{i\in C \,|\, (1-\beta_i^G\delta)G_i + (\nabla G_i)^T s = 0 = (1-\beta_i^H\delta)H_i + (\nabla H_i)^T s\}. \end{aligned}$$

Due to the disjunctive structure of the auxiliary problem we can subdivide it in a several QP-pieces. For every partition $(C_G, C_H) \in \mathcal{P}(C)$ we define

$$
\begin{array}{llll}
\min\limits_{(s,\delta)\in\mathbb{R}^{n+1}} & \frac{1}{2}s^TBs + (\nabla f)^Ts + \rho(\frac{1}{2}\delta^2+\delta) & QP(\beta,\rho,C_G,C_H) & \\
\text{subject to} & (1-\delta)h_i + (\nabla h_i)^Ts = 0 & i\in E, & \\
& (1-\beta_i^g\delta)g_i + (\nabla g_i)^Ts \le 0 & i\in I, & \\
& (1-\beta_i^G\delta)G_i + (\nabla G_i)^Ts = 0 & i\in C_G, & \\
& (1-\beta_i^H\delta)H_i + (\nabla H_i)^Ts \ge 0 & i\in C_G, & \\
& (1-\beta_i^G\delta)G_i + (\nabla G_i)^Ts \ge 0 & i\in C_H, & \\
& (1-\beta_i^H\delta)H_i + (\nabla H_i)^Ts = 0 & i\in C_H, & \\
& -\delta \le 0. & &
\end{array} \tag{6}
$$

The method we use is an active set method very similar to those known for quadratic programming. Given a vector $\beta = (\beta^g, \beta^G, \beta^H)$ and a pair (s,δ) feasible for $QPCC(\beta,\rho)$, a triple of index sets $J = (J_g, J_G, J_H)$ with $J_g \subset I^g(s,\delta,\beta)$, $J_G \subset I^{0+}(s,\delta,\beta) \cup I^{00}(s,\delta,\beta)$, $J_H \subset I^{+0}(s,\delta,\beta) \cup I^{00}(s,\delta,\beta)$ and $J_G \cup J_H = C$ is called a working set with respect to (β, s, δ), if the family of vectors

$$
\begin{gathered}
\{(\nabla h_i, -h_i)\,|\, i\in E\} \cup \{(\nabla g_i, -\beta_i^g g_i)\,|\, i\in J_g\setminus\{0\}\} \cup \{(0,-1)\,|\, \text{if } 0\in J_g\} \\
\cup\{(\nabla G_i, -\beta_i^G G_i)\,|\, i\in J_G\} \cup \{(\nabla H_i, -\beta_i^H H_i)\,|\, i\in J_H\}
\end{gathered}
$$

is linearly independent. Note that in case $0 \in J_g$ this linear independence requirement is fulfilled if and only if the family of vectors

$$
\{\nabla h_i\,|\, i\in E\} \cup \{\nabla g_i\,|\, i\in J_g\setminus\{0\}\} \cup \{\nabla G_i\,|\, i\in J_G\} \cup \{\nabla H_i\,|\, i\in J_H\} \tag{7}
$$

is linearly independent. We now define for every vector β, every scalar ρ and each working set J the equality constrained quadratic program

$$
\begin{array}{lll}
\min\limits_{(s,\delta)\in\mathbb{R}^{n+1}} & \frac{1}{2}s^TBs + (\nabla f)^Ts + \rho(\frac{1}{2}\delta^2+\delta) & EQP(\beta,\rho,J) \\
\text{subject to} & (1-\delta)h_i + (\nabla h_i)^Ts = 0 & i\in E, \\
& (1-\beta_i^g\delta)g_i + (\nabla g_i)^Ts = 0 & i\in J_g\setminus\{0\}, \\
& (1-\beta_i^G\delta)G_i + (\nabla G_i)^Ts = 0 & i\in J_G, \\
& (1-\beta_i^H\delta)H_i + (\nabla H_i)^Ts = 0 & i\in J_H, \\
& \delta = 0 & \text{if } 0\in J_g.
\end{array} \tag{8}
$$

At a solution (s,δ) we can define corresponding multipliers $\lambda(\beta,\rho,J) = (\lambda^h, \lambda^g, \lambda^G, \lambda^H) \in \mathbb{R}^{|E|+|I|+2|C|}$ fulfilling the first order necessary conditions

$$
Bs + \nabla f + \sum_{i\in E}\lambda_i^h \nabla h_i + \sum_{i\in J_g\setminus\{0\}}\lambda_i^g\nabla g_i - \sum_{i\in J_G}\lambda_i^G\nabla G_i - \sum_{i\in J_H}\lambda_i^H\nabla H_i = 0, \tag{9}
$$

$$
\rho(\delta+1) - \sum_{i\in E}\lambda_i^h h_i - \sum_{i\in J_g\setminus\{0\}}\lambda_i^g\beta_i^g g_i + \sum_{i\in J_G}\lambda_i^G\beta_i^G G_i + \sum_{i\in J_H}\lambda_i^H\beta_i^H H_i = 0 \text{ if } 0\notin J_g, \tag{10}
$$

$$
\lambda_i^g = 0, i\in I\setminus J_g, \quad \lambda_i^G = 0, i\in C\setminus J_G, \quad \lambda_i^H = 0, i\in C\setminus J_H. \tag{11}
$$

Due to the definition of a working set those multipliers are uniquely determined.

Fix a constants $\zeta \in (0,1)$, $\bar{\rho} > 1$ and take some $\rho > 0$. An outline of the algorithm is as follows.

Algorithm 4.1. (Solving the QPCC)

1: Initialization:
  Set the starting point $(s,\delta) := (0,1)$, set the vector β and take some working set J with respect to β, set $t := 0$ and $\beta^0 := \beta, (s^0,\delta^0) := (s,\delta), \tilde{J}^0 := J$.
  If we could not find a working set, stop the algorithm.

2: Improvement step:
  If (s,δ) is not a solution of $EQP(\beta,\rho,J)$ then
  either
    we find a new working set J with respect to β and a solution (s,δ) of $EQP(\beta,\rho,J)$ which is feasible for $QPCC(\beta,\rho)$
  or
    we must perform a restart: set $\rho = \rho\bar{\rho}$ and go to step 1.

3: Test for optimality:
  We try to find a new working set J and a descent direction (d,τ) for problem $QPCC(\beta,\rho)$ at the point (s,δ). If we have to switch to a new QP-piece, we increase the counter t of pieces by 1 and set $(s^t,\delta^t) := (s,\delta), J^t := J$, $\lambda^t := \lambda(\beta,\rho,J)$. J and eventually β are updated and we set $\beta^t := \beta, \tilde{J}^t := J$.
  This step can be terminated by one of the following possibilities:
  Either
    we proved that (s,δ) is strongly M-stationary for $QPCC(\beta,\rho)$.
    If $\delta < \zeta$ set $N := t+1$ and save $(s^N,\delta^N) := (s,\delta), J^N := J, \lambda^N := \lambda(\beta,\rho,J)$. Stop the algorithm and return.
    Else
      if the degeneracy condition (14) is fulfilled, stop the algorithm
      else set $\rho = \rho\bar{\rho}$ and go to step 1.
  Or
    we found a new working set J and a descent direction (d,τ) for problem $QPCC(\beta,\rho)$ at the point (s,δ).
    If $\tau > 0$ set $\rho = \rho\bar{\rho}$ and go to step 1.
    Otherwise set $(s,\delta) := (s,\delta) + \alpha(d,\tau)$ for some appropriate positive step size α (e. g. by formula in line 4 in Procedure 4.1 below), update J and go to step 2.
  Or
    β was changed, go to step 2.

The test for degeneracy looks as follows. Consider a solution $\eta = (\eta^h,\eta^g,\eta^G,\eta^H)$ of the system

$$\sum_{i\in E}\eta_i^h\nabla h_i + \sum_{i\in J_g\setminus\{0\}}\eta_i^g\nabla g_i - \sum_{i\in J_G}\eta_i^G\nabla G_i - \sum_{i\in J_H}\eta_i^H\nabla H_i = 0, \tag{12}$$

$$-\sum_{i\in E}\eta_i^h h_i - \sum_{i\in J_g\setminus\{0\}}\eta_i^g\beta_i^g g_i + \sum_{i\in J_G}\eta_i^G\beta_i^G G_i + \sum_{i\in J_H}\eta_i^H\beta_i^H H_i = 1, \tag{13}$$

which is unique by the definition of a working set, provided this system is solvable. We say that the degeneracy condition is fulfilled, if the system (12), (13) is solvable and the unique solution η fulfills

$$\eta_i^g \geq 0, i \in J_g \backslash \{0\}, \eta_i^G, \eta_i^H \geq 0, i \in J_G \cap J_H \text{ or } \eta_i^g \leq 0, i \in J_g \backslash \{0\}, \eta_i^G, \eta_i^H \leq 0, i \in J_G \cap J_H. \tag{14}$$

At each step the current iterate (s, δ) together with the parameters β, ρ are such that (s, δ) is feasible for $QPCC(\beta, \rho)$. Moreover, (s, δ) is also feasible for $EQP(\beta, \rho, J)$ for the actual working set J and the actual β.

In the following subsections we describe the individual steps in detail.

4.1. Initialization

At the beginning we set

$$\beta_i^g := \begin{cases} 1 & \text{if } g_i > 0 \\ 0 & \text{if } g_i \leq 0, \end{cases} \qquad (\beta_i^G, \beta_i^H) := \begin{cases} (1,1) & \text{if } G_i < 0 \text{ and } H_i < 0 \\ (1,0) & \text{if } G_i \leq H_i \text{ and } H_i \geq 0 \\ (0,1) & \text{if } H_i < G_i \text{ and } G_i \geq 0. \end{cases}$$

Note that in case $G_i = H_i \geq 0$ we do not set $(\beta_i^G, \beta_i^H) = (1,1)$, because we want to hold the set $I^{00}(s^0, \delta^0, \beta)$ of bi-active indices as small as possible. From the remaining two alternatives we fix $(\beta_i^G, \beta_i^H) = (1,0)$. Further, by definition of β it follows that $I^{0+}(s^0, \delta^0, \beta) \subset \{i \in C \,|\, G_i \leq H_i \wedge H_i \geq 0\}$ and $I^{+0}(s^0, \delta^0, \beta) \subset \{i \in C \,|\, H_i < G_i \wedge G_i \geq 0\}$ and hence we can proceed as follows to construct the initial working set J. First we set $J_g := \emptyset$, $J_G := \{i \in C \,|\, (G_i < 0 \wedge H_i < 0) \vee (G_i \leq H_i \wedge H_i \geq 0)\}$, $J_H := \{i \in C \,|\, H_i < G_i \wedge G_i \geq 0\}$, where it is to note that any other splitting of the set $I^{00}(s^0, \delta^0, \beta)$ to J_G and J_H would work as well.

Then, as long as J is not a working set we try to move indices belonging to $I^{00}(s^0, \delta^0, \beta)$ from J_G to J_H and vice-versa. E.g., if $i \in J_G \cap I^{00}(s^0, \delta^0, \beta)$ and $(\nabla G_i, -\beta_i^G G_i)$ depends linearly on the other gradients belonging to $J \cup E$, we can move i to J_H if $(\nabla H_i, -\beta_i^H H_i)$ does not depend on those gradients. Otherwise, if $(\nabla H_i, -\beta_i^H H_i)$ also depends linearly on those gradients we stop and say we cannot find a working set.

Note that β is chosen in such a way that

$$\begin{array}{rclll} \beta_i^g((1-\beta_i^g\delta^0)g_i + (\nabla g_i)^T s^0) & = & \beta_i^g(1-\beta_i^g)g_i & = 0, & \forall i \in I, \\ \beta_i^G((1-\beta_i^G\delta^0)G_i + (\nabla G_i)^T s^0) & = & \beta_i^G(1-\beta_i^G)G_i & = 0, & \forall i \in C, \\ \beta_i^H((1-\beta_i^H\delta^0)H_i + (\nabla H_i)^T s^0) & = & \beta_i^H(1-\beta_i^H)H_i & = 0, & \forall i \in C. \end{array} \tag{15}$$

4.2. Improvement step

If the current point (s, δ) is not a solution of the problem $EQP(\beta, \rho, J)$ then we can easily compute a direction (d, τ) pointing from (s, δ) to the solution of $EQP(\beta, \rho, J)$ by solving the quadratic program

$$\begin{array}{lll}\min\limits_{(d,\tau)\in\mathbb{R}^{n+1}} & \frac{1}{2}d^TBd+(Bs+\nabla f)^Td+\rho(\frac{1}{2}\tau^2+(1+\delta)\tau) & RQP(\beta,\rho,J,(s,\delta))\\ \text{subject to} & -\tau h_i+(\nabla h_i)^Td=0 & i\in E,\\ & -\beta_i^g\tau g_i+(\nabla g_i)^Td=0 & i\in J_g\setminus\{0\},\\ & -\beta_i^G\tau G_i+(\nabla G_i)^Td=0 & i\in J_G,\\ & -\beta_i^H\tau H_i+(\nabla H_i)^Td=0 & i\in J_H,\\ & \tau=0 & \text{if } 0\in J_g.\end{array} \tag{16}$$

Consider now the following procedure.

Procedure 4.1. (Improvement step)

1: `while` $((s,\delta)$ is not solution of problem $EQP(\beta,\rho,J))$

2: { compute search direction (d,τ) as solution of problem $RQP(\beta,\rho,J,(s,\delta))$

3: `if` $\tau>0$ `return` (parameter ρ is not large enough and we must perform a restart)

4: compute step length α by

$$\hat\alpha := \min\Bigg\{\min_{\substack{i\in I\setminus J_g\\ -\beta_i^g\tau g_i+(\nabla g_i)^Td>0}}\frac{-(1-\beta_i^g\delta)g_i-(\nabla g_i)^Ts}{-\beta_i^g\tau g_i+(\nabla g_i)^Td},\ \min_{\substack{i\in C\setminus J_G\\ -\beta_i^G\tau G_i+(\nabla G_i)^Td<0}}\frac{-(1-\beta_i^G\delta)G_i-(\nabla G_i)^Ts}{-\beta_i^G\tau G_i+(\nabla G_i)^Td},$$
$$\min_{\substack{i\in C\setminus J_H\\ -\beta_i^H\tau H_i+(\nabla H_i)^Td<0}}\frac{-(1-\beta_i^H\delta)H_i-(\nabla H_i)^Ts}{-\beta_i^H\tau H_i+(\nabla H_i)^Td},1\Bigg\},$$

$$\alpha := \begin{cases}\hat\alpha & \text{if } 0\in J_g \text{ or } \tau=0,\\ \min\{\hat\alpha,\frac{-\delta}{\tau}\} & \text{otherwise.}\end{cases}$$

5: $(s,\delta):=(s,\delta)+\alpha(d,\tau)$

6: `if` $\alpha<1$ set either $J_g:=J_g\cup\{i\}$ or $J_G:=J_G\cup\{i\}$ or $J_H:=J_H\cup\{i\}$, depending in which part the minimum is attained, where adding 0 to J_g has to be done with priority.

7: }

Note that during the course of the procedure J remains a working set. For instance if we add the ith inequality constraint we have $-\beta_i^g\tau g_i+(\nabla g_i)^Td\neq 0$ and hence $((\nabla g_i)^T,-\beta_i^g g_i)$ is linearly independent of

$$\{(\nabla h_i,-h_i)\,|\,i\in E\}\cup\{(\nabla g_i,-\beta_i^g g_i)\,|\,i\in J_g\setminus\{0\}\}\cup\{(0,-1)\,|\,\text{if } 0\in J_g\}$$
$$\cup\{(\nabla G_i,-\beta_i^G G_i)\,|\,i\in J_G\}\cup\{(\nabla H_i,-\beta_i^H H_i)\,|\,i\in J_H\}$$

because (d,τ) fulfills the constraints of $RQP(\beta,\rho,J,(s,\delta))$.

Lemma 4.1. Procedure 4.1 is finite.

P r o o f. Since we terminate the procedure if the condition at line 3 is fulfilled we may assume that always $\tau\le 0$. Whenever $\alpha<1$ we add some index to our working set and hence this can occur only finitely many times. Hence after finitely many steps we have $\alpha=1$ and then (s,δ) is a solution of $EQP(\beta,\rho,J)$. □

4.3. Test for optimality

This part of the algorithm is very similar to the algorithm in [10, p. 926]. At the beginning we have available a feasible point (s,δ) and an MPEC working set J for the problem $QPCC(\beta,\rho)$. It is now crucial to find a vector $b=(b^g,b^G,b^H)\in\mathbb{R}_+^{l+1}\times\mathbb{R}_-^q\times\mathbb{R}_-^q$ with $b_i^g=0, i\in J_g, b_i^G=0, i\in J_G, b_i^H=0, i\in J_H$ such that for every $(u,\mu)\in\mathbb{R}^n\times\mathbb{R}$ the family of gradients

$$\begin{gathered}\{(\nabla h_i,-h_i)\,|\,i\in E,(\nabla h_i)^Tu-h_i\mu=0\}\\ \cup\{(\nabla g_i,-\beta_i^g g_i)\,|\,i\in I^g(s,\delta,\beta)\setminus\{0\},(\nabla g_i)^Tu-\beta_i^g g_i\mu=b_i^g\}\\ \cup\{(0,-1)\,|\,\text{if }0\in I^g(s,\delta,\beta),\mu=b_0^g\}\\ \cup\{(\nabla G_i,-\beta_i^G G_i)\,|\,i\in I^{0+}(s,\delta,\beta)\cup I^{00}(s,\delta,\beta),(\nabla G_i)^Tu-\beta_i^G G_i\mu=b_i^G\}\\ \cup\{(\nabla H_i,-\beta_i^H H_i)\,|\,i\in I^{+0}(s,\delta,\beta)\cup I^{00}(s,\delta,\beta),(\nabla H_i)^Tu-\beta_i^H H_i\mu=b_i^H\}\end{gathered}\tag{17}$$

is linearly independent. It was pointed out in [10] that a random choice of b with

$$b_i^g>0, i\in I^g(s,\delta,\beta)\setminus J_g, b_i^G<0, i\in I^{00}(s,\delta,\beta)\setminus J_G, b_i^H<0, i\in I^{00}(s,\delta,\beta)\setminus J_H$$

and fixing the other components to zero will yield a suitable vector b with probability 1, see also Remark 4.3.

Procedure 4.2. (Test for optimality)

if J is not an MPEC working set for $QPCC(\beta,\rho)$, add some active constraints to get an MPEC working set

set $(u,\mu):=0$, set the vector b and compute multipliers $\lambda(J)$

`while` $((\exists i\in J_g:\lambda_i^g<0)\vee(\exists i\in J_G\cap J_H:\lambda_i^G<0\vee\lambda_i^H<0))$

{ remove an index corresponding to some negative multiplier from J as described in Procedure 4.3 (storing of the objects and redefinition of the vector β might occur)

  `if` (β was changed)

    `return`

  compute search direction (d,τ) as a solution of problem $RQP(\beta,\rho,J,(s,\delta))$

  `if` $(((\nabla g_i)^Td-\beta_i^g g_i\tau\le 0, i\in I^g(s,\delta,\beta)\setminus J_g)$
  $\wedge\ ((\nabla G_i)^Td-\beta_i^G G_i\tau\ge 0, i\in I^{00}(s,\delta,\beta)\setminus J_G)$
  $\wedge\ ((\nabla H_i)^Td-\beta_i^H H_i\tau\ge 0, i\in I^{00}(s,\delta,\beta)\setminus J_H))$

    `return` ((d,τ) is a feasible descent direction for $QPCC(\beta,\rho)$ at (s,δ))

  compute step length

$$\alpha:=\min\left\{\min_{\substack{i\in(I^g(s,\delta,\beta)\cap I)\setminus J_g\\ -\beta_i^g\tau g_i+(\nabla g_i)^Td>0}}\frac{b_i^g-(-\beta_i^g g_i\mu+(\nabla g_i)^Tu)}{-\beta_i^g\tau g_i+(\nabla g_i)^Td},\ \min_{\substack{i\in I^{00}(s,\delta,\beta)\setminus J_G\\ -\beta_i^G\tau G_i+(\nabla G_i)^Td<0}}\frac{b_i^G-(-\beta_i^G G_i\mu+(\nabla G_i)^Tu)}{-\beta_i^G\tau G_i+(\nabla G_i)^Td},\right.$$
$$\left.\min_{\substack{i\in I^{00}(s,\delta,\beta)\setminus J_H\\ -\beta_i^H\tau H_i+(\nabla H_i)^Td<0}}\frac{b_i^H-(-\beta_i^H H_i\mu+(\nabla H_i)^Tu)}{-\beta_i^H\tau H_i+(\nabla H_i)^Td}\right\}$$

  either set $J_g:=J_g\cup\{i\}$ or $J_G:=J_G\cup\{i\}$ or $J_H:=J_H\cup\{i\}$, depending

in which part the minimum is attained when computing α.
12: $(u, \mu) := (u, \mu) + \alpha(d, \tau)$, compute multipliers $\lambda(J)$
13: }
14: `return` ((s, δ) is strongly M-stationary for $QPCC(\beta, \rho)$)

We see that in case when we return a descent direction (d, τ) for the problem $QPCC(\beta, \rho)$ at (s, δ) this direction is solution of $RQP(\beta, \rho, J, (s, \delta))$. Hence, for choosing α in step 3 of Algorithm 4.1 we can use the formula from line 4 of Procedure 4.1.

Lemma 4.2. Procedure 4.2 is finite.

Proof. If either the condition at line 5 or the condition at line 8 is fulfilled, we terminate the procedure. Thus we assume that this never occurs. Then, by taking into account that the direction (d, τ) computed in line 7 always is a descent direction, we can use similar argument as in the proof of [10, Theorem 4.3] to show that the procedure is finite. □

Remark 4.3. We can easily detect whether the vector b has been chosen properly or not. We test whether the minimum when computing the step length α in line 10 of Procedure 4.2 is uniquely attained. If this is not the case, enlarge the corresponding components of b randomly such that the minimum becomes unique.

The way how we remove an index from J is described in the following procedure.

Procedure 4.3. (Remove index)
1: `if` $(\exists i \in J_g : \lambda_i^g < 0)$
2: $J_g := J_g \setminus \{i\}$
3: `else if` $(\exists i \in J_G \cap J_H \cap \tilde{J}_H^t : \lambda_i^G < 0)$
4: $J_G := J_G \setminus \{i\}$
5: `else if` $(\exists i \in J_G \cap J_H \cap \tilde{J}_G^t : \lambda_i^H < 0)$
6: $J_H := J_H \setminus \{i\}$
7: `else` (we switch to a new piece so we need to save the objects)
8: { save the objects corresponding to the old piece:
$(s^{t+1}, \delta^{t+1}) := (s, \delta), J^{t+1} := J, \lambda^{t+1} := \lambda$
9: remove $i \in J_G \cap J_H$ corresponding to the negative multiplier:
either set $J_G := J_G \setminus \{i\}$ or $J_H := J_H \setminus \{i\}$
10: eventually redefine some of components of the vector β as follows:

$$(\beta_i^G, \beta_i^H) := (0, 0) \quad \text{if } G_i, H_i > 0 \wedge i \in J_G \cap J_H \tag{18}$$

this can violate the condition $(J_G \cap J_H) \subset I^{00}(s, \delta, \beta)$
11: secure the condition $(J_G \cap J_H) \subset I^{00}(s, \delta, \beta)$:
$\forall i \in J_G \cap J_H \setminus I^{00}(s, \delta, \beta)$ either set $J_G := J_G \setminus \{i\}$ or $J_H := J_H \setminus \{i\}$
12: save the objects corresponding to the new piece:
$\tilde{J}^{t+1} := J, \beta^{t+1} := \beta$, $(t := t + 1)$
13:}

Note that we call the Procedure 4.3 only in case that there exists a negative multiplier either to an inequality constraint or to the ith complementarity constraint with $i \in J_G \cap J_H$. Hence, at line 9 at least one index $i \in J_G \cap J_H$ corresponding to a negative multiplier exists. Moreover, for those i exactly one of the multipliers λ_i^G, λ_i^H is negative because the case when $i \in J_G \cap J_H$ and $\lambda_i^G, \lambda_i^H < 0$ is handled by either the branch in line 3 or the branch in line 5.

Between lines 8 and 12 in Procedure 4.3 we only remove indices from J and so we always have

$$\tilde{J}^t \subset J^t \tag{19}$$

in the sense $\tilde{J}_g^t \subset J_g^t, \tilde{J}_G^t \subset J_G^t, \tilde{J}_H^t \subset J_H^t$. Moreover, we remove indices either from J_G or from J_H, but not from both and removed indices all belong to $J_G \cap J_H$. This results in $J_G^t \cap J_H^t \subset \tilde{J}_G^t \cup \tilde{J}_H^t$, $J_G^t \setminus J_H^t \subset \tilde{J}_G^t$, $J_H^t \setminus J_G^t \subset \tilde{J}_H^t$ and hence

$$J_G^t \cup J_H^t = (J_G^t \setminus J_H^t) \cup (J_H^t \setminus J_G^t) \cup (J_G^t \cap J_H^t) \subset \tilde{J}_G^t \cup \tilde{J}_H^t \cup (\tilde{J}_G^t \cup \tilde{J}_H^t) = \tilde{J}_G^t \cup \tilde{J}_H^t \subset J_G^t \cup J_H^t$$

yielding

$$J_G^t \cup J_H^t = \tilde{J}_G^t \cup \tilde{J}_H^t.$$

Remark 4.4. Note that (18) yields following trivial properties of β^t for all $t = 1, \ldots$ $\ldots, N-1$:

$$\beta^t \in \mathcal{B}; \qquad \beta^t \leq \beta^{t-1}; \qquad \beta_i^t = \beta_i^0,\ i \in I \text{ or } i \in C : G_i \leq 0 \vee H_i \leq 0. \tag{20}$$

Lemma 4.5. At every stage of the algorithm the point (s, δ) is feasible for $QPCC(\beta, \rho)$ and J is a working set with respect to (β, s, δ).

P r o o f. We prove the assertion by induction with respect to t. At the initialization step the chosen starting point (s^0, δ^0) is feasible for $QPCC(\beta^0, \rho)$ and J is a working set with respect to (β^0, s^0, δ^0). It is easy to see, that in the improvement step the step size α guarantees that (s, δ) remains feasible and we only add a constraint to J which becomes active and is linearly independent from the others contained in J. Since removing an index $i \in J_G \cap J_H$ does not affect the property of J being a working set, the same holds obviously true in Procedure 4.2 as long as we do not switch to a new piece in the Procedure 4.3. Hence the assertion holds true until line 10 of Procedure 4.3 is entered the first time.

As an induction hypothesis we now assume that at every stage of the algorithm before we reach line 10 of Procedure 4.3 the tth time the iterate (s, δ) is feasible for $QPCC(\beta, \rho)$ and J is a working set with respect to (β, s, δ). Note that the current value of β is $\beta = \beta^{t-1}$. It is clear that the assertion holds true if we do not actually change β and hence lets consider the case that

$$\exists (i \in C) \text{ with } G_i, H_i > 0, i \in J_G \cap J_H \text{ and } (\beta_i^G, \beta_i^H) \neq (0,0). \tag{21}$$

Consider i as in (21), implying $(\beta_i^{G,t}, \beta_i^{H,t}) = (0,0)$ by (18). Note that in the initialization step we set one component of β_i, without loss of generality $\beta_i^{H,0}$, to zero and so obviously $\beta_i^H = 0$. Together with $G_i > 0$ and $i \in J_G \cap J_H$ we obtain

$$\begin{array}{rcl} 0 & = & (1-\beta_i^G\delta)G_i + (\nabla G_i)^T s \le (1-\beta_i^{G,t}\delta)G_i + (\nabla G_i)^T s = G_i + (\nabla G_i)^T s \\ & \perp & H_i + (\nabla H_i)^T s = (1-\beta_i^{H,t}\delta)H_i + (\nabla H_i)^T s = (1-\beta_i^H\delta)H_i + (\nabla H_i)^T s = 0 \end{array}$$

showing feasibility and complementarity of the ith complementarity constraint. But now it could happen that $i \in I^{+0}(s,\delta,\beta^t)$ and therefore $i \notin I^{00}(s,\delta,\beta^t)$. This situation is repaired in line 11 of Procedure 4.3 so that $J_G \subset I^{0+}(s,\delta,\beta^t) \cup I^{00}(s,\delta,\beta^t)$ holds.

Now let us argue that J is a working set after line 11 of Procedure 4.3. Again, take i as in (21) and assume without loss of generality $\beta_i^H = 0$ and $\beta_i^G > 0$. Since $G_i > 0$ it follows that i is removed from J_G in line 11 of Procedure 4.3 and we see that the linear independence requirement of a working set is certainly fulfilled.

Again it is easy to see, that at every stage of the improvement step Procedure 4.1 and of Procedure 4.2 the point (s,δ) remains feasible for $QPCC(\beta,\rho)$ and J is a working set with respect to (β,s,δ), until we reach the next time line 10 in Procedure 4.3. This completes the induction step and the lemma is proved. □

The following theorem summarizes basic properties of the quantities computed in the test for optimality stage. Let us yet define the index sets C_G^t and C_H^t for $t = 1,\dots,N$ which will be important for definition of merit functions in the next section by

$$C_G^t := \tilde{J}_G^{t-1} \cap J_G^t, \quad C_H^t := C \setminus C_G^t \text{ for } t = 1,\dots,N. \tag{22}$$

Note that obviously $(C_G^t, C_H^t) \in \mathcal{P}(C)$.

Theorem 4.6.

1. Partitioning property of J and its consequences: For all $t = 1,\dots,N$ it holds that

$$\left(\tilde{J}_G^{t-1} \cap J_G^t\right) \cup \left(\tilde{J}_H^{t-1} \cap J_H^t\right) = C \tag{23}$$

and so

$$C_H^t \subset \tilde{J}_H^{t-1} \cap J_H^t, \quad (C_G^t \cap C_H^{t+1}) \cup (C_H^t \cap C_G^{t+1}) \subset J_G^t \cap J_H^t. \tag{24}$$

2. Efficiency property of β: If $G_iH_i > 0$, then for all $t = 0,\dots,N-1$ we have

$$\beta_i^{G,t} \neq 0 \Rightarrow i \notin J_H^\tau, \forall \tau \le t, \quad \beta_i^{H,t} \neq 0 \Rightarrow i \notin J_G^\tau, \forall \tau \le t. \tag{25}$$

3. Non-negativity property of multipliers λ: For all $t = 1,\dots,N$ it holds that

$$\lambda_i^{g,t} \ge 0,\ \forall i \in J_g^t, \qquad \lambda_i^{G,t} \ge 0,\ \forall i \in J_G^t \cap C_H^t, \qquad \lambda_i^{H,t} \ge 0,\ \forall i \in J_H^t \cap C_G^t. \tag{26}$$

4. Feasibility and working set properties of (s,δ) and J: For all $t = 0,\dots,N-1$ it holds that (s^t,δ^t) and (s^{t+1},δ^{t+1}) are feasible points for $QPCC(\beta^t,\rho)$ and $QP(\beta^t,\rho,C_G^{t+1},C_H^{t+1})$, that (s^{t+1},δ^{t+1}) is also a solution for $QP(\beta^t,\rho,C_G^{t+1},C_H^{t+1})$ and that $\tilde{J}^t$ is a working set and J^{t+1} is even an MPEC working set with respect to $(\beta^t,s^{t+1},\delta^{t+1})$.

P r o o f. 1. First, we show by contraposition that

$$\tilde{J}_G^{t-1} \cup J_H^t = C. \tag{27}$$

We assume on the contrary that there exists some $i \in C$ such that $i \notin \tilde{J}_G^{t-1} \wedge i \notin J_H^t$. Since $\tilde{J}^{t-1}$ is a working set with respect to $(\beta^{t-1}, s^{t-1}, \delta^{t-1})$ by Lemma 4.5, we know $\tilde{J}_G^{t-1} \cup \tilde{J}_H^{t-1} = C$ and so we have $i \in \tilde{J}_H^{t-1}$. Since we assume $i \notin J_H^t$, the index i was removed from J_H in line 6 of procedure 4.3 in some intermediate stage implying $i \in \tilde{J}_G^{t-1}$, a contradiction, and our claim is proved. In a similar way one can show the relation

$$\tilde{J}_H^{t-1} \cup J_G^t = C. \tag{28}$$

Now, for arbitrary sets A_1, A_2, A_3, A_4 it holds that

$$(A_1 \cup A_2) \cap (A_3 \cup A_4) = (A_1 \cap A_3) \cup (A_1 \cap A_4) \cup (A_2 \cap A_3) \cup (A_2 \cap A_4). \tag{29}$$

Assuming $A_1 \cap A_4 \subset A_2 \cup A_3$ we obtain

$$A_1 \cap A_4 = (A_1 \cap A_4) \cap (A_2 \cup A_3) = (A_1 \cap A_4 \cap A_2) \cup (A_1 \cap A_4 \cap A_3) \subset (A_2 \cap A_4) \cup (A_1 \cap A_3)$$

and, similarly, by assuming $A_2 \cap A_3 \subset A_1 \cup A_4$ we conclude $A_2 \cap A_3 \subset (A_2 \cap A_4) \cup (A_1 \cap A_3)$. Hence we obtain from (29)

$$(A_1 \cup A_2) \cap (A_3 \cup A_4) = (A_1 \cap A_3) \cup (A_1 \cap A_4) \cup (A_2 \cap A_3) \cup (A_2 \cap A_4) = (A_1 \cap A_3) \cup (A_2 \cap A_4). \tag{30}$$

Setting $A_1 := \tilde{J}_G^{t-1}, A_2 := J_H^t, A_3 := J_G^t, A_4 := \tilde{J}_H^{t-1}$, and taking into account that $\tilde{J}^{t-1}$ and J^t are working sets with respect to $(\beta^{t-1}, s^{t-1}, \delta^{t-1})$ and with respect to $(\beta^{t-1}, s^t, \delta^t)$ by Lemma 4.5, we have

$$J_G^t \cap J_H^t \subset C = \tilde{J}_G^{t-1} \cup \tilde{J}_H^{t-1} \text{ and } \tilde{J}_G^{t-1} \cap \tilde{J}_H^{t-1} \subset C = J_G^t \cup J_H^t$$

and so we obtain from (27), (28) and (30)

$$C = C \cap C = (\tilde{J}_G^{t-1} \cup J_H^t) \cap (J_G^t \cup \tilde{J}_H^{t-1}) = (\tilde{J}_G^{t-1} \cap J_G^t) \cup (J_H^t \cap \tilde{J}_H^{t-1}).$$

First of two formulas in (24) follows immediately from definition of C_H^t and (23). To show the second one we use the first one that gives us $C_G^t \subset \tilde{J}_G^{t-1} \cap J_G^t \subset J_G^t$ and $C_H^{t+1} \subset \tilde{J}_H^t \cap J_H^{t+1} \subset \tilde{J}_H^t \subset J_H^t$ and so $C_G^t \cap C_H^{t+1} \subset J_G^t \cap J_H^t$. Analogously, $C_H^t \cap C_G^{t+1} \subset J_G^t \cap J_H^t$.

2. Fix any $i \in C$ with $G_i, H_i > 0$ and assume first that $\beta_i^{G,t} > 0$. From Initialization step and the monotonicity of β we conclude that $\beta_i^{H,\tau} = 0, \forall \tau$. Now assume that there is some $\tau \leq t$ with $i \in J_H^\tau$ and let denote by $\bar{\tau}$ the smallest index with this property. It follows that $\bar{\tau} > 0$ and, by using 1.,

$$i \notin J_H^{\bar{\tau}-1} \Rightarrow i \notin \tilde{J}_H^{\bar{\tau}-1} \Rightarrow i \notin C_H^{\bar{\tau}} \Rightarrow i \in C_G^{\bar{\tau}} \Rightarrow i \in J_G^{\bar{\tau}}.$$

From this we conclude $i \in J_G^{\bar{\tau}} \cap J_H^{\bar{\tau}}$ therefore $\beta_i^{G,\bar{\tau}} = 0$ and consequently $\beta_i^{G,t} = 0$, a contradiction. Hence, $i \notin J_H^\tau, \forall \tau \leq t$. The case $\beta_i^{H,t} > 0$ can be treated in a similar manner.

3. This follows easily from Procedure 4.3. If we switch to a new piece then the sets $\{i \in J_g \,|\, \lambda_i^g < 0\}$, $\{i \in J_G \cap J_H \cap \tilde{J}_H^{t-1} \,|\, \lambda_i^G < 0\}$, $\{i \in J_G \cap J_H \cap \tilde{J}_G^{t-1} \,|\, \lambda_i^H < 0\}$ are empty and therefore the assertion follows by using the relations $C_H^t \subset \tilde{J}_H^{t-1} \cap J_H^t$ and $C_G^t = \tilde{J}_G^{t-1} \cap J_G^t$.

4. Feasibility of (s^t, δ^t) and (s^{t+1}, δ^{t+1}) for $QPCC(\beta^t, \rho)$ is an immediate consequence of Lemma 4.5. Since $C_G^{t+1} \subset \tilde{J}_G^t \cap J_G^{t+1}$ and $C_H^{t+1} \subset \tilde{J}_H^t \cap J_H^{t+1}$, feasibility of the pair (s^{t+1}, δ^{t+1}) for $EQP(\beta^t, \rho, J^{t+1})$ and feasibility of (s^t, δ^t) for $EQP(\beta^t, \rho, \tilde{J}^t)$, together with feasibility of both pairs for $QPCC(\beta^t, \rho)$, imply their feasibility for $QP(\beta^t, \rho, C_G^{t+1}, C_H^{t+1})$. Since

$$\begin{aligned}
&Bs^{t+1} + \nabla f + \textstyle\sum_{i \in E} \lambda_i^{h,t+1} \nabla h_i + \sum_{i \in J_g^{t+1} \setminus \{0\}} \lambda_i^{g,t+1} \nabla g_i \\
&\qquad - \textstyle\sum_{i \in J_G^{t+1}} \lambda_i^{G,t+1} \nabla G_i - \sum_{i \in J_H^{t+1}} \lambda_i^{H,t+1} \nabla H_i &&= 0 \qquad (31)\\
&\rho(\delta^{t+1} + 1) - \textstyle\sum_{i \in E} \lambda_i^{h,t+1} h_i - \sum_{i \in J_g^{t+1} \setminus \{0\}} \lambda_i^{g,t+1} \beta_i^{g,t} g_i \\
&\qquad + \textstyle\sum_{i \in J_G^{t+1}} \lambda_i^{G,t+1} \beta_i^{G,t} G_i + \sum_{i \in J_H^{t+1}} \lambda_i^{H,t+1} \beta_i^{H,t} H_i &&= 0 \quad \text{if } 0 \notin J_g^{t+1}, \qquad (32)
\end{aligned}$$

$J_G^{t+1} = C_G^{t+1} \cup (J_G^{t+1} \setminus C_G^{t+1}) = C_G^{t+1} \cup (J_G^{t+1} \cap C_H^{t+1})$, $J_H^{t+1} = C_H^{t+1} \cup (J_H^{t+1} \setminus C_H^{t+1}) = C_H^{t+1} \cup (J_H^{t+1} \cap C_G^{t+1})$, and the previous property holds, the point (s^{t+1}, δ^{t+1}) fulfills the KKT conditions for $QP(\beta^t, \rho, C_G^{t+1}, C_H^{t+1})$ with multipliers λ^{t+1} and is therefore also a solution by convexity of the problem.

The statement that J^{t+1} is an MPEC working set with respect to $(\beta^t, s^{t+1}, \delta^{t+1})$ follows from the same arguments as used in the proof of [10, Theorem 4.3]. □

4.4. Finiteness of Algorithm 4.1

For a given working set J let us denote by A_J the matrix with rows

$$(\nabla h_i)^T, i \in E; \quad (\nabla g_i)^T, i \in J_g \setminus \{0\}; \quad (\nabla G_i)^T, i \in J_G; \quad (\nabla H_i)^T, i \in J_H.$$

Further, for every working set J we denote by b_J the vector

$$(h_i, i \in E; g_i, i \in J_g \setminus \{0\}; G_i, i \in J_G; H_i, i \in J_H)^T$$

and for every $\beta \in \mathcal{B}$ we denote by b_J^β the vector

$$\big(h_i, i \in E; \beta_i^g g_i, i \in J_g \setminus \{0\}; \beta_i^G G_i, i \in J_G; \beta_i^H H_i, i \in J_H\big)^T.$$

Note that $\|b_J\|, \|b_J^\beta\| \le C_b := \big(\sum_{i \in E} h_i^2 + \sum_{i \in I} g_i^2 + \sum_{i \in C} (G_i^2 + H_i^2)\big)^{1/2}$. Let us denote by κ the number $\kappa := \max \|A_J^+\|$, where maximum is taken over all working sets J occurring in Algorithm 4.1 and A_J^+ denotes the pseudo-inverse of A_J. Finally, we define the constants $\bar{C}_d := 2\kappa C_b$ and $\bar{C}_s := \max\{4\|\nabla f\|/\lambda(B), 2(\bar{f}/\lambda(B))^{1/2}\}$, where $\bar{f} := \max_{\|s\| \le \bar{C}_d} \frac{1}{2} s^T B s + (\nabla f)^T s$ and $\lambda(B)$ denotes the smallest eigenvalue of B.

Lemma 4.7. 1. At every stage of the Algorithm 4.1 we have $\|s\| \le \bar{C}_s$ and $\delta \le 1$.

2. If

$$\rho \geq \bar{C}_\rho := \frac{2}{(1+\delta)^2}\left(\left(\frac{1}{2}\|B\|\bar{C}_d + \|B\|\bar{C}_s + \|\nabla f\|\right)\bar{C}_d + \frac{1}{2\lambda(B)}(\|B\|\bar{C}_s + \|\nabla f\|)^2\right)$$

then at every subsequent stage the solution (d,τ) of $RQP(\beta,\rho,J,(s,\delta))$ fulfills $\tau \leq 0$ and the case $\tau = 0$ is possible only if every feasible point $(\hat{d},\hat{\tau})$ for $RQP(\beta,\rho,J,(s,\delta))$ fulfills $\hat{\tau} = 0$.

Proof. 1. We first show the following claim: Whenever $(\hat{s},\hat{\delta})$ is a solution of $EQP(\beta,\rho,J)$ with $\hat{\delta} \leq 1$ then $\|\hat{s}\| \leq \bar{C}_s$. This is done by contraposition. Assume on the contrary that $(\hat{s},\hat{\delta})$ is a solution of some problem $EQP(\beta,\rho,J)$ with $\hat{\delta} \leq 1$ and $\|\hat{s}\| > \bar{C}_s$. Then $\hat{s}$ fulfills $A_J\hat{s} = -b_J + \hat{\delta}b_J^\beta$ and hence there exists a point $(\tilde{s},\hat{\delta})$ feasible for $EQP(\beta,\rho,J)$ with $\|\tilde{s}\| \leq \kappa\| - b_J + \hat{\delta}b_J^\beta\| \leq 2\kappa C_b = \bar{C}_d$. By the definitions of $\bar{C}_s$ and $\bar{f}$ together with $\|\hat{s}\| > \bar{C}_s$ we obtain

$$\frac{1}{2}\hat{s}^T B\hat{s} + (\nabla f)^T\hat{s} \geq \frac{1}{2}\lambda(B)\|\hat{s}\|^2 - \|\nabla f\|\|\hat{s}\| \geq \frac{1}{4}\lambda(B)\|\hat{s}\|^2 \geq \bar{f} \geq \frac{1}{2}\tilde{s}^T B\tilde{s} + (\nabla f)^T\tilde{s}$$

and by adding $\rho(1/2\hat{\delta}^2 + \hat{\delta})$ to both sides we conclude that $(\tilde{s},\hat{\delta})$ is also a solution of $EQP(\beta,\rho,J)$, which is not possible due to the strict convexity of the objective. Hence our claim is proved. Now the first assertion follows by induction. At the initialization step we have $\|s\| = 0 \leq \bar{C}_s$ and $\delta = 1$. Now assume as the induction hypothesis that we are at a point (s,δ) with $\|s\| \leq \bar{C}_s$ and $\delta \leq 1$. We shall show that the next iterate (s^+,δ^+) also fulfills $\|s^+\| \leq \bar{C}_s$ and $\delta^+ \leq 1$. We move away from (s,δ) in some direction (d,τ) pointing to the solution $(\hat{s},\hat{\delta})$ of some problem $EQP(\beta,\rho,J)$, but only if $\tau \leq 0$ because if $\tau > 0$ then we make a restart. Hence $\hat{\delta} \leq \delta \leq 1$ and from the claim just proved we conclude $\|\hat{s}\| \leq \bar{C}_s$. Now the induction argument follows from the observation that the next iterate (s^+,δ^+) is on the line segment connecting (s,δ) and $(\hat{s},\hat{\delta})$.

2. We first show the following claim by contraposition: If $\rho \geq \bar{C}_\rho$ then at every stage of Algorithm 4.1 the solution (d,τ) of the problem $RQP(\beta,\rho,J,(s,\delta))$ satisfies $\tau < 0$ whenever there is a point $(\hat{d},\hat{\tau})$ feasible for $RQP(\beta,\rho,J,(s,\delta))$ with $\hat{\tau} \neq 0$. Assume on the contrary that a solution $(\bar{d},\bar{\tau})$ of some problem $RQP(\beta,\rho,J,(s,\delta))$ fulfills $\bar{\tau} \geq 0$. Note that by the first part of the lemma we have $\|s\| \leq \bar{C}_s$ and $\delta \leq 1$. The point $(\hat{d},\hat{\tau})$ fulfills $A_J\hat{d} = b_J^\beta\hat{\tau}$, showing $A_J(-\frac{\hat{d}}{\hat{\tau}}(\delta+1)) = -(\delta+1)b_J^\beta$. Hence there exists a point $(\tilde{d},-(\delta+1))$ feasible for $RQP(\beta,\rho,J,(s,\delta))$ satisfying $\|\tilde{d}\| \leq \kappa\|(\delta+1)b_J^\beta\| \leq 2\kappa C_b = \bar{C}_d$. Denoting $q(d) := \frac{1}{2}d^T Bd + (Bs+\nabla f)^T d$ then

$$\min_{d\in\mathbb{R}^n} q(d) = -\frac{1}{2}(Bs+\nabla f)^T B^{-1}(Bs+\nabla f) \geq -\frac{1}{2\lambda(B)}\left(\|B\|\bar{C}_s + \|\nabla f\|\right)^2.$$

On the other hand, we have $q(\tilde{d}) \leq (\frac{1}{2}\|B\|\bar{C}_d + \|B\|\bar{C}_s + \|\nabla f\|)\bar{C}_d$ and therefore

$$\rho \geq \bar{C}_\rho \geq \frac{2}{(1+\delta)^2}((q(\tilde{d}) - \min_{d\in\mathbb{R}^n} q(d))).$$

Hence $q(\tilde{d}) - \rho\frac{(1+\delta)^2}{2} \leq \min_{d\in\mathbb{R}^n} q(d) \leq q(\bar{d}) + \rho\left(\frac{\bar{\tau}^2}{2} + (1+\delta)\bar{\tau}\right)$, where we have taken also in account $\bar{\tau} \geq 0$. On the other hand, since $(\bar{d}, \bar{\tau})$ is the unique global solution of $RQP(\beta, \rho, J, (s, \delta))$ and $(\bar{d}, \bar{\tau}) \neq (\tilde{d}, -(\delta+1))$ we obtain the contradiction

$$q(\bar{d}) + \rho\left(\frac{\bar{\tau}^2}{2} + (1+\delta)\bar{\tau}\right) < q(\tilde{d}) - \rho\frac{(1+\delta)^2}{2}.$$

Hence our claim is proved and the assertion of the lemma follows easily from this claim. □

Theorem 4.8. Algorithm 4.1 is finite.

Proof. By contraposition, let us assume that algorithm 4.1 is not finite. We already showed that the procedures 4.1 and 4.2 are finite.

First we show that ρ must tend to infinity. If ρ remains finite it is changed only finitely many times and hence it is constant from a certain stage. But for constant ρ, β can obviously change only finitely many times. Hence β also becomes constant from a certain stage. But then we can enter the Test for optimality step only once for each working set J, because objective function of $QPCC(\beta, \rho)$ is strictly decreasing. And since there are only finitely many working sets J we must stop the algorithm in contradiction to our assumption. Hence ρ must tend to infinity.

Because of Lemma 4.7 we know that for $\rho \geq \bar{C}_\rho$ the computed search directions (d, τ) fulfill $\tau \leq 0$. Hence the case that we increase ρ because of $\tau > 0$ can occur only finitely many often and we increase ρ infinitely many times because we found a sequence (s_j, δ_j) of strongly M-stationary solutions (s, δ) for $QPCC(\beta_j, \rho_j)$ with $\delta_j \geq \zeta$ and the non-degeneracy condition (14) is not fulfilled. Without lost of generality we can assume that $\rho_j \geq \bar{C}_\rho$. By passing to a subsequence we can assume that the MPEC working set J_j corresponding to the strongly M-stationary solution (s_j, δ_j) is the same for every j, $J_j = J$. At the strongly M-stationary solution (s_j, δ_j) we know that $(0, 0)$ is the solution of $RQP(\beta_j, \rho_j, J, (s_j, \delta_j))$ and by Lemma 4.7 we know that all feasible points (d, τ) of $RQP(\beta_j, \rho_j, J, (s_j, \delta_j))$ fulfill $\tau = 0$. But from this we can conclude from the fundamental theorem of linear algebra that (12), (13) has a solution η_j which must be unique. Now, let us consider the multipliers λ_j fulfilling the first order optimality conditions (9), (10). Since $\rho_j(\delta_j + 1) \to \infty$ and $\beta_j \in \mathcal{B}$ we conclude $\|\lambda_j\| \to \infty$ and by passing to a subsequence once more we can assume $\lambda_j/\|\lambda_j\| \to \lambda$. Then we conclude from (9) that

$$\sum_{i\in E} \lambda_i^h \nabla h_i + \sum_{i\in J_g\setminus\{0\}} \lambda_i^g \nabla g_i - \sum_{i\in J_G} \lambda_i^G \nabla G_i - \sum_{i\in J_H} \lambda_i^H \nabla H_i = 0$$

and, since λ_j was the multiplier to a strongly M-stationary solution, we have $\lambda_i^g \geq 0, i \in J_g\setminus\{0\}, \lambda_i^G, \lambda_i^H \geq 0, i \in J_G \cap J_H$. Now consider $\sigma_j := -\sum_{i\in E} \lambda_i^h h_i - \sum_{i\in J_g\setminus\{0\}} \lambda_i^g \beta_{i,j}^g g_i + \sum_{i\in J_G} \lambda_i^G \beta_{i,j}^G G_i + \sum_{i\in J_H} \lambda_i^H \beta_{i,j}^H H_i$. Since J is a working set with respect to (β_j, s_j, δ_j) it follows that $\sigma_j \neq 0$ and hence $\lambda/\sigma_j = \eta_j$. Hence we see that η_j fulfills the degeneracy condition (14), a contradiction. □

The following theorem summarizes basic properties of the quantities computed in the Algorithm 4.1.

Theorem 4.9. If the Algorithm 4.1 terminates with a strongly M-stationary solution (s, δ) with $\delta < \zeta$, then the following properties hold:

1. Monotonicity property of δ: For all $t = 1, \ldots, N$ it holds that

$$1 \geq \delta^{t-1} \geq \delta^t \geq 0. \tag{33}$$

2. Boundedness of the counter of the pieces t: There exists a constant C_t, dependent only on number of constraints such that

$$N \leq C_t. \tag{34}$$

3. Signs of the final multipliers λ:

$$\lambda_i^{g,N} \geq 0, \ \forall i \in J_g^N, \qquad \lambda_i^{G,N}, \lambda_i^{H,N} \geq 0, \ \forall i \in J_G^N \cap J_H^N. \tag{35}$$

Proof. 1. This follows easily from the fact that it is prevented for the algorithm to reach the line where setting $(s, \delta) := (s, \delta) + \alpha(d, \tau)$ occurs with $\tau > 0$.

2. Since whenever the parameter ρ is increased the algorithm goes to the step 1 and thus the counter t of the pieces is reset to 0, it follows that after the last time the algorithm enters step 1 we keep ρ constant. With a fixed vector β, it is obvious that the algorithm never returns to the same piece implying that the maximum of switches to a new piece is $2^{|C|}$. Since we only redefine β when there is i with $G_i, H_i > 0$ and we set $(\beta_i^G, \beta_i^H) := (0, 0)$, it follows that β changes at most $|\{i \in C \,|\, G_i, H_i > 0\}|$ times. Thus, the total number of switches to a new piece is certainly bounded by $|C|2^{|C|}$.

Property 3. follows from the fact that we assume that we stop at a strongly M-stationary solution. □

5. AN SQP ALGORITHM FOR MPCC

An outline of the algorithm is as follows.

Algorithm 5.1. (Solving the MPCC)

1: Initialization:

Select a starting point $x_0 \in \mathbb{R}^n$ together with a positive definite $n \times n$ matrix B_0 and a parameter $\rho_0 > 0$ and choose constants $\zeta \in (0, 1)$ and $\bar{\rho} > 1$.

Select positive penalty parameters $\sigma_{-1} = (\sigma_{-1}^h, \sigma_{-1}^g, \sigma_{-1}^C)$.

Set the iteration counter $k := 0$.

2: Solve the Auxiliary problem:

Run Algorithm 4.1 with data $\zeta, \bar{\rho}, \rho := \rho_k, B := B_k, \nabla f := \nabla f(x_k), h_i := h_i(x_k), \nabla h_i := \nabla h_i(x_k), i \in E$, etc.

If the Algorithm 4.1 stops because the degeneracy condition is fulfilled or because no working set was found, stop the Algorithm 5.1 with an error message.

If the final iterate s^N is zero, stop the Algorithm 5.1 and return x_k as a solution.

3: Next iterate:

Compute new penalty parameters σ_k.

Set $x_{k+1} := x_k + s_k$ where s_k is a point on the polygonal line connecting the points $s^0, s^1, \dots, s^N$ such that an appropriate merit function depending on σ_k is decreased.
Set $\rho_{k+1} := \rho$, the final value of ρ in Algorithm 4.1.
Update B_k to get positive definite matrix B_{k+1}.
Set $k := k+1$ and go to step 2.

Remark 5.1. We terminate the Algorithm 5.1 only in the following two cases. In the first case no sufficient reduction of the violation of the constraints can be achieved or no working set can be found. The second case will be satisfied only by chance when the current iterate is a strongly M-stationary solution. Normally, this algorithm produces an infinite sequence of iterates and we must include a stopping criterion for convergence. Such a criterion could be that the violation of the constraints at some iterate is sufficiently small,

$$\max\left\{\max_{i\in I}(g_i(x_k))^+, \max_{i\in E}|h_i(x_k)|, \max_{i\in C}|\min\{G_i(x_k), H_i(x_k)\}|\right\} \le \epsilon_C,$$

and the expected decrease in our merit function is sufficiently small,

$$(s_k^{N_k})^T B_k s_k^{N_k} \le \epsilon_1,$$

$$\sum_{i\in E}\lambda_{i,k}^{h,N_k}|h_i(x_k)| + \sum_{i\in I}\lambda_{i,k}^{g,N_k}|g_i(x_k)| + \sum_{i\in C}\left(\lambda_{i,k}^{G,N_k}|G_i(x_k)| + \lambda_{i,k}^{H,N_k}|H_i(x_k)|\right) \le \epsilon_2,$$

see Proposition 5.3 below.

5.1. The next iterate

Let the outcome of Algorithm 4.1 at the kth iterate be denoted by

$$(s_k^t, \delta_k^t), \lambda_k^t, J_k^t, \tilde{J}_k^t, \beta_k^t \text{ for } t = 0, \dots, N_k$$

and recall that the sets $C_{G,k}^t$ and $C_{H,k}^t$ are given by

$$C_{G,k}^t := \tilde{J}_{G,k}^{t-1} \cap J_{G,k}^t, \quad C_{H,k}^t := C \setminus C_{G,k}^t \text{ for } t = 1, \dots, N_k.$$

The new penalty parameters are computed by

$$\sigma_{i,k}^h = \begin{cases}\xi_2\tilde{\lambda}_{i,k}^h & \text{if } \sigma_{i,k-1}^h < \xi_1\tilde{\lambda}_{i,k}^h,\\ \sigma_{i,k-1}^h & \text{else,}\end{cases} \qquad \sigma_{i,k}^g = \begin{cases}\xi_2\tilde{\lambda}_{i,k}^g & \text{if } \sigma_{i,k-1}^g < \xi_1\tilde{\lambda}_{i,k}^g,\\ \sigma_{i,k-1}^g & \text{else,}\end{cases} \tag{36}$$

$$\sigma_{i,k}^C = \begin{cases}\xi_2\max\{\tilde{\lambda}_{i,k}^G, \tilde{\lambda}_{i,k}^H\} & \text{if } \sigma_{i,k-1}^C < \xi_1\max\{\tilde{\lambda}_{i,k}^G, \tilde{\lambda}_{i,k}^H\},\\ \sigma_{i,k-1}^C & \text{else,}\end{cases}$$

where

$$\tilde{\lambda}_{i,k}^h = \max|\lambda_{i,k}^{h,t}|, \quad \tilde{\lambda}_{i,k}^g = \max|\lambda_{i,k}^{g,t}|, \quad \tilde{\lambda}_{i,k}^G = \max|\lambda_{i,k}^{G,t}|, \quad \tilde{\lambda}_{i,k}^H = \max|\lambda_{i,k}^{H,t}| \tag{37}$$

with maximum being taken over $t \in \{1, \dots, N_k\}$ and $1 < \xi_1 < \xi_2$. Note that this choice of σ_k ensures

$$\sigma_k^h \ge \tilde{\lambda}_k^h, \ \sigma_k^g \ge \tilde{\lambda}_k^g, \ \sigma_k^C \ge \max\{\tilde{\lambda}_k^G, \tilde{\lambda}_k^H\}. \tag{38}$$

5.1.1. The merit function

We are looking for the next iterate at the polygonal line connecting the points $s_k^0, s_k^1, \dots$ $\dots, s_k^{N_k}$. For each line segment $[s_k^{t-1}, s_k^t], t = 1, \dots, N_k$ we consider the functions

$$\begin{aligned}
\phi_k^t(\alpha) \quad &:= \quad f(x_k+s) + \sum_{i\in E} \sigma_{i,k}^h |h_i(x_k+s)| + \sum_{i\in I} \sigma_{i,k}^g (g_i(x_k+s))^+ \\
&\quad + \sum_{i\in C_{G,k}^t} \sigma_{i,k}^C (|G_i(x_k+s)| - (H_i(x_k+s))^-) \\
&\quad + \sum_{i\in C_{H,k}^t} \sigma_{i,k}^C (|H_i(x_k+s)| - (G_i(x_k+s))^-), \\
\hat{\phi}_k^t(\alpha) \quad &:= \quad f + \nabla f^T s + \frac{1}{2} s^T B_k s + \sum_{i\in E} \sigma_{i,k}^h |h_i + \nabla h_i^T s| + \sum_{i\in I} \sigma_{i,k}^g (g_i + \nabla g_i^T s)^+ \\
&\quad + \sum_{i\in C_{G,k}^t} \sigma_{i,k}^C (|G_i + \nabla G_i^T s| - (H_i + \nabla H_i^T s)^-) \\
&\quad + \sum_{i\in C_{H,k}^t} \sigma_{i,k}^C (|H_i + \nabla H_i^T s| - (G_i + \nabla G_i^T s)^-),
\end{aligned}$$

where $s = (1-\alpha)s_k^{t-1} + \alpha s_k^t$ and $f = f(x_k)$, $\nabla f = \nabla f(x_k)$, $h_i = h_i(x_k), \nabla h_i = \nabla h_i(x_k), i \in E$, etc. Obviously the functions $\hat{\phi}_k^t$ are first order approximations of ϕ_k^t, that is $|\phi_k^t(\alpha) - \hat{\phi}_k^t(\alpha)| = o(\|(1-\alpha)s_k^{t-1} + \alpha s_k^t\|)$.

Lemma 5.2. For every $t \in \{1, \dots, N_k\}$ the function $\hat{\phi}_k^t$ is convex.

P r o o f. Obviously $\hat{\phi}_k^t$ is convex because it is sum of convex functions. □

We state now the main result of this subsection. For the sake of simplicity we omit the iteration index k in this part.

Proposition 5.3. For every $t \in \{1, \dots, N\}$

$$\begin{aligned}
\hat{\phi}^t(1) - \hat{\phi}^1(0) \quad &\leq \quad -\sum_{\tau=1}^t \frac{1}{2}(s^\tau - s^{\tau-1})^T B (s^\tau - s^{\tau-1}) && (39) \\
&\quad +(\delta^t - 1)\left(\sum_{i\in E} (\sigma_i^h - \tilde{\lambda}_i^h)|h_i| + \sum_{i\in I} (\sigma_i^g - \tilde{\lambda}_i^g) \beta_i^{g,0} (g_i)^+ \right. \\
&\qquad \left. - \sum_{i\in C} (\sigma_i^C - \max\{\tilde{\lambda}_i^G, \tilde{\lambda}_i^H\}) \left(\beta_i^{G,0}(G_i)^- + \beta_i^{H,0}(H_i)^- \right) \right) \\
&\leq \quad -\sum_{\tau=1}^t \frac{1}{2}(s^\tau - s^{\tau-1})^T B (s^\tau - s^{\tau-1}) \;\leq\; 0. && (40)
\end{aligned}$$

In order to prove this proposition we need the following lemma.

Lemma 5.4. For every $t \in \{1, \ldots, N-1\}$

$$\hat{\phi}^{t+1}(0) - \hat{\phi}^t(1) = -\sum_{i \in \tilde{C}_H^t} \sigma_i^C (H_i + (\nabla H_i)^T s^t)^+ - \sum_{i \in \tilde{C}_G^t} \sigma_i^C (G_i + (\nabla G_i)^T s^t)^+ \le 0, \quad (41)$$

where

$$\tilde{C}_H^t := C_H^t \cap C_G^{t+1} \quad \text{and} \quad \tilde{C}_G^t = C_G^t \cap C_H^{t+1}. \quad (42)$$

Proof. By the definition of $\hat{\phi}$ we have

$$\begin{aligned} \hat{\phi}^{t+1}(0) - \hat{\phi}^t(1) &= \sum_{i \in \tilde{C}_H^t} \sigma_i^C \left((G_i + (\nabla G_i)^T s^t)^+ - (H_i + (\nabla H_i)^T s^t)^+ \right) \\ &\quad + \sum_{i \in \tilde{C}_G^t} \sigma_i^C \left((H_i + (\nabla H_i)^T s^t)^+ - (G_i + (\nabla G_i)^T s^t)^+ \right) \end{aligned}$$

and thus the result follows if we show

$$\begin{aligned} (G_i + (\nabla G_i)^T s^t)^+ &= 0 \text{ for } i \in \tilde{C}_H^t, \quad (43) \\ (H_i + (\nabla H_i)^T s^t)^+ &= 0 \text{ for } i \in \tilde{C}_G^t. \quad (44) \end{aligned}$$

Fix any $i \in \tilde{C}_H^t$. From the second part of (24) we obtain $i \in J_G^t \cap J_H^t$ and consequently $G_i + (\nabla G_i)^T s^t = \beta_i^{G,t-1} \delta^t G_i$. Hence (43) follows if we show

$$\beta_i^{G,t-1} \delta^t G_i \le 0 \text{ for } i \in \tilde{C}_H^t. \quad (45)$$

This holds obviously true if $G_i \le 0$ or if $\beta_i^{G,t-1} = 0$. If $G_i > 0$ and $H_i \le 0$ then $\beta_i^{G,0} = 0$ and consequently $\beta_i^{G,t-1} = 0$ showing the validity of (45). In the remaining case $G_i, H_i > 0, \beta_i^{G,t-1} > 0$ we conclude from the first part of (25) that $i \in J_G^\tau, \forall \tau \le t-1$ implying $i \in C_G^t$, contradicting $i \in C_H^t$. Thus (45) and consequently (43) hold. Equation (44) can be shown by using similar arguments with G and H interchanged. This completes the proof. ☐

Proof. [Proof of Proposition 5.3] Consider any $t \in \{1, \ldots, N\}$. Then we have

$$\begin{aligned} \hat{\phi}^t(1) - \hat{\phi}^t(0) &= 1/2(s^t)^T B s^t - 1/2(s^{t-1})^T B s^{t-1} + (\nabla f)^T (s^t - s^{t-1}) \quad (46) \\ &\quad + \sum_{i \in E} \left(\sigma_i^h (|h_i + (\nabla h_i)^T s^t| - |h_i + (\nabla h_i)^T s^{t-1}|) \right) \\ &\quad + \sum_{i \in I} \left(\sigma_i^g ((g_i + (\nabla g_i)^T s^t)^+ - (g_i + (\nabla g_i)^T s^{t-1})^+) \right) \\ &\quad + \sum_{i \in C_G^t} \sigma_i^C \left((|G_i + (\nabla G_i)^T s^t| - |G_i + (\nabla G_i)^T s^{t-1}|) \right. \\ &\qquad \left. + (-(H_i + (\nabla H_i)^T s^t)^- + (H_i + (\nabla H_i)^T s^{t-1})^-) \right) \\ &\quad + \sum_{i \in C_H^t} \sigma_i^C \left((|H_i + (\nabla H_i)^T s^t| - |H_i + (\nabla H_i)^T s^{t-1}|) \right. \\ &\qquad \left. + (-(G_i + (\nabla G_i)^T s^t)^- + (G_i + (\nabla G_i)^T s^{t-1})^-) \right). \end{aligned}$$

Using that s^t is the solution of $EQP(\beta^{t-1}, \rho, J^t)$ and multiplying the first order optimality condition (9) by $(s^t - s^{t-1})^T$ yields

$$(s^t - s^{t-1})^T \left(Bs^t + \nabla f + \sum_{i \in E} \lambda_i^{h,t} \nabla h_i + \sum_{i \in I} \lambda_i^{g,t} \nabla g_i - \sum_{i \in C} (\lambda_i^{G,t} \nabla G_i + \lambda_i^{H,t} \nabla H_i) \right) = 0. \tag{47}$$

Subtracting the expression on the left hand side from the right hand side of (46) and taking into account the identity

$$1/2(s^t)^T Bs^t - 1/2(s^{t-1})^T Bs^{t-1} - (s^t - s^{t-1})^T Bs^t = -1/2(s^t - s^{t-1})^T B(s^t - s^{t-1})$$

we obtain

$$\begin{aligned} \hat{\phi}^t(1) - \hat{\phi}^t(0) = & -\frac{1}{2}(s^t - s^{t-1})^T B(s^t - s^{t-1}) \\ & + \sum_{i \in E} \left(\sigma_i^h (|h_i + (\nabla h_i)^T s^t| - |h_i + (\nabla h_i)^T s^{t-1}|) - \lambda_i^{h,t} (\nabla h_i)^T (s^t - s^{t-1}) \right) \\ & + \sum_{i \in I} \left(\sigma_i^g ((g_i + (\nabla g_i)^T s^t)^+ - (g_i + (\nabla g_i)^T s^{t-1})^+) - \lambda_i^{g,t} (\nabla g_i)^T (s^t - s^{t-1}) \right) \\ & + \sum_{i \in C_G^t} \mathcal{D}_i(t,t) + \sum_{i \in C_H^t} \mathcal{E}_i(t,t), \end{aligned} \tag{48}$$

where for all $1 \le t_1 \le t_2 \le N$ and every $i \in C$ we denote

$$\begin{aligned} \mathcal{D}_i(t_1, t_2) := & \sigma_i^C \left(|G_i + (\nabla G_i)^T s^{t_2}| - |G_i + (\nabla G_i)^T s^{t_1 - 1}| \right) + \sum_{t=t_1}^{t_2} \lambda_i^{G,t} (\nabla G_i)^T (s^t - s^{t-1}) \\ & + \sigma_i^C \left(-(H_i + (\nabla H_i)^T s^{t_2})^- + (H_i + (\nabla H_i)^T s^{t_1 - 1})^- \right) + \sum_{t=t_1}^{t_2} \lambda_i^{H,t} (\nabla H_i)^T (s^t - s^{t-1}), \\ \mathcal{E}_i(t_1, t_2) := & \sigma_i^C \left(|H_i + (\nabla H_i)^T s^{t_2}| - |H_i + (\nabla H_i)^T s^{t_1 - 1}| \right) + \sum_{t=t_1}^{t_2} \lambda_i^{H,t} (\nabla H_i)^T (s^t - s^{t-1}) \\ & + \sigma_i^C \left(-(G_i + (\nabla G_i)^T s^{t_2})^- + (G_i + (\nabla G_i)^T s^{t_1 - 1})^- \right) + \sum_{t=t_1}^{t_2} \lambda_i^{G,t} (\nabla G_i)^T (s^t - s^{t-1}). \end{aligned}$$

Summing up (48) from 1 to t and taking into account Lemma 5.4 we obtain

$$\begin{aligned}
\hat\phi^t(1) - \hat\phi^1(0) =& \sum_{\tau=1}^{t} \left(\hat\phi^\tau(1) - \hat\phi^\tau(0)\right) + \sum_{\tau=1}^{t-1} \left(\hat\phi^{\tau+1}(0) - \hat\phi^\tau(1)\right) \\
=& -\sum_{\tau=1}^{t} \frac{1}{2}(s^\tau - s^{\tau-1})^T B (s^\tau - s^{\tau-1}) \\
&+ \sum_{i\in E} \left(\sigma_i^h(|h_i+(\nabla h_i)^T s^t| - |h_i+(\nabla h_i)^T s^0|) - \sum_{\tau=1}^{t} \lambda_i^{h,\tau}(\nabla h_i)^T(s^\tau - s^{\tau-1})\right) \\
&+ \sum_{i\in I} \left(\sigma_i^g((g_i+(\nabla g_i)^T s^t)^+ - (g_i+(\nabla g_i)^T s^0)^+) - \sum_{\tau=1}^{t} \lambda_i^{g,\tau}(\nabla g_i)^T(s^\tau - s^{\tau-1})\right) \\
&+ \sum_{\tau=1}^{t} \left(\sum_{i\in C_G^\tau} \mathcal{D}_i(\tau,\tau) + \sum_{i\in C_H^\tau} \mathcal{E}_i(\tau,\tau) \right) \\
&- \sum_{\tau=1}^{t-1} \left(\sum_{i\in \tilde C_H^\tau} \sigma_i^C (H_i + (\nabla H_i)^T s^\tau)^+ + \sum_{i\in \tilde C_G^\tau} \sigma_i^C (G_i + (\nabla G_i)^T s^\tau)^+ \right).
\end{aligned} \tag{49}$$

We claim that for every $i \in C$ we have

$$\begin{aligned}
&\sum_{\tau\in\{1,\ldots,t\}: i\in C_G^\tau} \mathcal{D}_i(\tau,\tau) - \sum_{\tau\in\{1,\ldots,t-1\}: i\in \tilde C_G^\tau} \sigma_i^C (G_i + (\nabla G_i)^T s^\tau)^+ \\
&+ \sum_{\tau\in\{1,\ldots,t\}: i\in C_H^\tau} \mathcal{E}_i(\tau,\tau) - \sum_{\tau\in\{1,\ldots,t-1\}: i\in \tilde C_H^\tau} \sigma_i^C (H_i + (\nabla H_i)^T s^\tau)^+ \\
&\qquad \le -(\sigma_i^C - \max\{\tilde\lambda_i^G, \tilde\lambda_i^H\})(\delta^t - 1)\left(\beta_i^{G,0}(G_i)^- + \beta_i^{H,0}(H_i)^-\right),
\end{aligned} \tag{50}$$

with $\tilde\lambda$ given by (37). In order to prove this claim, fix $i \in C$ and consider $\underline{\tau} \le \bar\tau$ such that $i \in C_G^\tau, \forall \tau : \underline{\tau} \le \tau \le \bar\tau$ and either $\underline{\tau} = 1$ or $i \in C_H^{\underline{\tau}-1}$. We will first show that

$$\mathcal{D}_i(\underline{\tau},\bar\tau) - \sigma_i^C (G_i + (\nabla G_i)^T s^{\bar\tau})^+ \le \hat{\mathcal{D}}_i(\underline{\tau},\bar\tau) \tag{51}$$

and

$$\mathcal{D}_i(\underline{\tau},\bar\tau) \le \tilde{\mathcal{D}}_i(\underline{\tau},\bar\tau), \tag{52}$$

where for all $1 \le t_1 \le t_2 \le N$ and every $i \in C$ we define

$$\begin{aligned}
&\hat{\mathcal{D}}_i(t_1,t_2) \\
&\quad := (\sigma_i^C - \tilde\lambda_i^G)(\beta_i^{G,t_2}\delta^{t_2} - \beta_i^{G,t_1-1}\delta^{t_1-1})|G_i| - (\sigma_i^C - \tilde\lambda_i^H)(\beta_i^{H,t_2}\delta^{t_2} - \beta_i^{H,t_1-1}\delta^{t_1-1})(H_i)^-, \\
&\tilde{\mathcal{D}}_i(t_1,t_2) \\
&\quad := (\sigma_i^C - \tilde\lambda_i^G)(\beta_i^{G,t_2-1}\delta^{t_2} - \beta_i^{G,t_1-1}\delta^{t_1-1})|G_i| - (\sigma_i^C - \tilde\lambda_i^H)(\beta_i^{H,t_2}\delta^{t_2} - \beta_i^{H,t_1-1}\delta^{t_1-1})(H_i)^-.
\end{aligned}$$

Indeed, since the point $(s^{\bar\tau}, \delta^{\bar\tau})$ is feasible for the problem $QPCC(\beta^{\bar\tau}, \rho)$ and $(s^{\underline{\tau}-1}, \delta^{\underline{\tau}-1})$ is feasible for $QPCC(\beta^{\underline{\tau}-1}, \rho)$ by Theorem 4.6(4.), we obtain

$$-(H_i + (\nabla H_i)^T s^{\bar\tau})^- \le -\beta_i^{H,\bar\tau}\delta^{\bar\tau}(H_i)^-, \; -(H_i + (\nabla H_i)^T s^{\underline{\tau}-1})^- \le -\beta_i^{H,\underline{\tau}-1}\delta^{\underline{\tau}-1}(H_i)^- \tag{53}$$

and we claim that in fact one has

$$(H_i + (\nabla H_i)^T s^{\underline{\tau}-1})^- = \beta_i^{H,\underline{\tau}-1}\delta^{\underline{\tau}-1}(H_i)^-. \tag{54}$$

This follows easily from (15) when $\underline{\tau} = 1$ and in case when $i \in C_H^{\underline{\tau}-1}$ we have $i \in J_H^{\underline{\tau}-1}$ by virtue of (24) and the claimed equality (54) follows since either $\beta_i^{H,\underline{\tau}-1} = \beta_i^{H,\underline{\tau}-2}$ or $\beta_i^{H,\underline{\tau}-1} = 0$.

Now, using that (s^τ, δ^τ) is a solution for the problem $QP(\beta^{\tau-1}, \rho, C_G^\tau, C_H^\tau)$ and $(s^{\tau-1}, \delta^{\tau-1})$ is feasible for the problem $QPCC(\beta^{\tau-1}, \rho)$ for $\tau = \underline{\tau}, \ldots, \bar\tau$ by Theorem 4.6(4.), together with $\lambda_i^{H,\tau} \ge 0$ if $i \in J_H^\tau$ by (26) and $\lambda_i^{H,\tau} = 0$ if $i \notin J_H^\tau$, we obtain

$$\lambda_i^{H,\tau}\left((1-\beta_i^{H,\tau-1}\delta^\tau)H_i + (\nabla H_i)^T s^\tau\right) = 0, \; \lambda_i^{H,\tau}\left((1-\beta_i^{H,\tau-1}\delta^{\tau-1})H_i + (\nabla H_i)^T s^{\tau-1}\right) \ge 0.$$

Rearranging and summing up from $\underline{\tau}$ to $\bar\tau$ yields

$$\begin{aligned}
\sum_{\tau=\underline{\tau}}^{\bar\tau} \lambda_i^{H,\tau}(\nabla H_i)^T(s^\tau - s^{\tau-1}) &\le \sum_{\tau=\underline{\tau}}^{\bar\tau} \lambda_i^{H,\tau}(\beta_i^{H,\tau-1}\delta^\tau - \beta_i^{H,\tau-1}\delta^{\tau-1})H_i \\
&\le \sum_{\tau=\underline{\tau}}^{\bar\tau} \lambda_i^{H,\tau}(\beta_i^{H,\tau-1}\delta^\tau - \beta_i^{H,\tau-1}\delta^{\tau-1})(H_i)^- \\
&\le \tilde\lambda_i^H (H_i)^- \sum_{\tau=\underline{\tau}}^{\bar\tau}(\beta_i^{H,\tau}\delta^\tau - \beta_i^{H,\tau-1}\delta^{\tau-1}) \\
&= \tilde\lambda_i^H(\beta_i^{H,\bar\tau}\delta^{\bar\tau} - \beta_i^{H,\underline{\tau}-1}\delta^{\underline{\tau}-1})(H_i)^-,
\end{aligned}$$

where we have used $\delta^{\tau-1} - \delta^\tau \ge 0$ and $\beta^{H,\tau-1} \ge \beta^{H,\tau}$, and together with (53) and (54) we obtain

$$\begin{aligned}
\sigma_i^C(-(H_i + (\nabla H_i)^T s^{\bar\tau})^- + (H_i + (\nabla H_i)^T s^{\underline{\tau}-1})^-) + \sum_{\tau=\underline{\tau}}^{\bar\tau} \lambda_i^{H,\tau}(\nabla H_i)^T(s^\tau - s^{\tau-1}) \\
\le -(\sigma_i^C - \tilde\lambda_i^H)(\beta_i^{H,\bar\tau}\delta^{\bar\tau} - \beta_i^{H,\underline{\tau}-1}\delta^{\underline{\tau}-1})(H_i)^-.
\end{aligned} \tag{55}$$

By feasibility of the point $(s^{\bar\tau}, \delta^{\bar\tau})$ for the problem $QPCC(\beta^{\bar\tau}, \rho)$ we have

$$-(G_i + (\nabla G_i)^T s^{\bar\tau})^- \le -\beta_i^{G,\bar\tau}\delta^{\bar\tau}(G_i)^- \le \beta_i^{G,\bar\tau}\delta^{\bar\tau}|G_i|. \tag{56}$$

Further, feasibility of the points $(s^{\tau-1}, \delta^{\tau-1})$ and (s^τ, δ^τ) for the problem $QP(\beta^{\tau-1}, \rho, C_G^\tau, C_H^\tau)$ by Theorem 4.6(4.), together with $i \in C_G^\tau$ yields

$$G_i + (\nabla G_i)^T s^{\tau-1} = \beta_i^{G,\tau-1}\delta^{\tau-1}G_i, \quad G_i + (\nabla G_i)^T s^\tau = \beta_i^{G,\tau-1}\delta^\tau G_i \tag{57}$$

for all τ satisfying $\underline{\tau} \le \tau \le \bar{\tau}$. By multiplying with $\lambda_i^{G,\tau}$ and summing up we obtain

$$\begin{aligned}
&\sum_{\tau=\underline{\tau}}^{\bar{\tau}} \lambda_i^{G,\tau}(\nabla G_i)^T(s^\tau - s^{\tau-1}) \\
&= \sum_{\tau=\underline{\tau}}^{\bar{\tau}} \lambda_i^{G,\tau}(\beta_i^{G,\tau-1}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1})G_i \\
&\le -\sum_{\tau=\underline{\tau}}^{\bar{\tau}} |\lambda_i^{G,\tau}|(\beta_i^{G,\tau-1}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1})|G_i| \\
&\le -\sum_{\tau=\underline{\tau}}^{\bar{\tau}} \tilde{\lambda}_i^G(\beta_i^{G,\tau-1}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1})|G_i| \\
&\le \begin{cases} -\sum_{\tau=\underline{\tau}}^{\bar{\tau}} \tilde{\lambda}_i^G(\beta_i^{G,\tau}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1})|G_i| = -\tilde{\lambda}_i^G|G_i|(\beta_i^{G,\bar{\tau}}\delta^{\bar{\tau}} - \beta_i^{G,\underline{\tau}-1}\delta^{\underline{\tau}-1}) \\ -\tilde{\lambda}_i^G|G_i|\Big(\sum_{\tau=\underline{\tau}}^{\bar{\tau}-1}(\beta_i^{G,\tau}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1}) + (\beta_i^{G,\bar{\tau}-1}\delta^{\bar{\tau}} - \beta_i^{G,\bar{\tau}-1}\delta^{\bar{\tau}-1})\Big) \\ \qquad\qquad = -\tilde{\lambda}_i^G|G_i|(\beta_i^{G,\bar{\tau}-1}\delta^{\bar{\tau}} - \beta_i^{G,\underline{\tau}-1}\delta^{\underline{\tau}-1}). \end{cases}
\end{aligned}$$

Together with the bound (56) and the bounds (57) with $\tau = \underline{\tau}$ and $\tau = \bar{\tau}$, respectively, we deduce that

$$\begin{aligned}
\sigma_i^C(-(G_i + (\nabla G_i)^T s^{\bar{\tau}})^- - |G_i + (\nabla G_i)^T s^{\underline{\tau}-1}|) + \sum_{\tau=\underline{\tau}}^{\bar{\tau}} \lambda_i^{G,\tau}(\nabla G_i)^T(s^\tau - s^{\tau-1}) \\
\le (\sigma_i^C - \tilde{\lambda}_i^G)(\beta_i^{G,\bar{\tau}}\delta^{\bar{\tau}} - \beta_i^{G,\underline{\tau}-1}\delta^{\underline{\tau}-1})|G_i|.
\end{aligned}$$

and

$$\begin{aligned}
\sigma_i^C(|G_i + (\nabla G_i)^T s^{\bar{\tau}}| - |G_i + (\nabla G_i)^T s^{\underline{\tau}-1}|) + \sum_{\tau=\underline{\tau}}^{\bar{\tau}} \lambda_i^{G,\tau}(\nabla G_i)^T(s^\tau - s^{\tau-1}) \\
\le (\sigma_i^C - \tilde{\lambda}_i^G)(\beta_i^{G,\bar{\tau}-1}\delta^{\bar{\tau}} - \beta_i^{G,\underline{\tau}-1}\delta^{\underline{\tau}-1})|G_i|.
\end{aligned}$$

By combining these inequalities with (55) and taking into account the identity $|G_i + (\nabla G_i)^T s^{\bar{\tau}}| = (G_i + (\nabla G_i)^T s^{\bar{\tau}})^+ - (G_i + (\nabla G_i)^T s^{\bar{\tau}})^-$, the claimed estimates (51) and (52) follow.

It is easy to see that the set $\{1,\dots,t\} \cap \{\tau \mid i \in C_G^\tau\}$ can be subdivided into disjoint sets $T_{j,i} = \{\underline{\tau}_{j,i},\dots,\bar{\tau}_{j,i}\}, j = 1,\dots,N_i$ where for the starting index $\underline{\tau}_{j,i}$ of each of these sets holds either $\underline{\tau}_{j,i} = 1$ or $i \in C_H^{\underline{\tau}_{j,i}-1}$ and for the final index $\bar{\tau}_{j,i}$ of each of these sets holds either $\bar{\tau}_{j,i} = t$ or $i \in C_H^{\bar{\tau}_{j,i}+1}$. Obviously we have $i \in C_G^\tau \, \forall \tau \in T_{j,i}$ and from the definitions it follows that $\mathcal{D}_i(\underline{\tau}_{j,i}, \bar{\tau}_{j,i}) = \sum_{\tau=\underline{\tau}_{j,i}}^{\bar{\tau}_{j,i}} \mathcal{D}_i(\tau,\tau)$, $\hat{\mathcal{D}}_i(\underline{\tau}_{j,i}, \bar{\tau}_{j,i}) = \sum_{\tau=\underline{\tau}_{j,i}}^{\bar{\tau}_{j,i}} \hat{\mathcal{D}}_i(\tau,\tau)$ and $\tilde{\mathcal{D}}_i(\underline{\tau}_{j,i}, \bar{\tau}_{j,i}) = \sum_{\tau=\underline{\tau}_{j,i}}^{\bar{\tau}_{j,i}-1} \hat{\mathcal{D}}_i(\tau,\tau) + \tilde{\mathcal{D}}_i(\bar{\tau}_{j,i}, \bar{\tau}_{j,i})$. Further we have

$$\{\tau \in \{1,\dots,t-1\} : i \in \tilde{C}_G^\tau\} = \begin{cases} \{\bar{\tau}_{1,i},\dots,\bar{\tau}_{N_i-1,i}\} & \text{if } \bar{\tau}_{N_i,i} = t, \\ \{\bar{\tau}_{1,i},\dots,\bar{\tau}_{N_i,i}\} & \text{if } \bar{\tau}_{N_i,i} < t. \end{cases}$$

Hence, if $\bar{\tau}_{N_i,i} = t$ we obtain

$$\begin{aligned}
&\sum_{\tau\in\{1,\dots,t\}:i\in C_G^\tau} \mathcal{D}_i(\tau,\tau) - \sum_{\tau\in\{1,\dots,t-1\}:i\in\tilde{C}_G^\tau} \sigma_i^C(G_i + (\nabla G_i)^T s^\tau)^+ \qquad (58)\\
&= \sum_{j=1}^{N_i-1}\left(\sum_{\tau=\underline{\tau}_{j,i}}^{\bar{\tau}_{j,i}} \mathcal{D}_i(\tau,\tau) - \sigma_i^C(G_i + (\nabla G_i)^T s^{\bar{\tau}_{j,i}})^+\right) + \sum_{\tau=\underline{\tau}_{N_i,i}}^{\bar{\tau}_{N_i,i}} \mathcal{D}_i(\tau,\tau)\\
&= \sum_{j=1}^{N_i-1}\left(\mathcal{D}_i(\underline{\tau}_{j,i},\bar{\tau}_{j,i}) - \sigma_i^C(G_i + (\nabla G_i)^T s^{\bar{\tau}_{j,i}})^+\right) + \mathcal{D}_i(\underline{\tau}_{N_i,i},\bar{\tau}_{N_i,i})\\
&\le \sum_{j=1}^{N_i-1}\hat{\mathcal{D}}_i(\underline{\tau}_{j,i},\bar{\tau}_{j,i}) + \tilde{\mathcal{D}}_i(\underline{\tau}_{N_i,i},\bar{\tau}_{N_i,i}) = \sum_{\tau\in\{1,\dots,t-1\}:i\in C_G^\tau}\hat{\mathcal{D}}_i(\tau,\tau) + \tilde{\mathcal{D}}_i(t,t)
\end{aligned}$$

and in case when $\bar{\tau}_{N_i,i} < t$ we have

$$\begin{aligned}
&\sum_{\tau\in\{1,\dots,t\}:i\in C_G^\tau} \mathcal{D}_i(\tau,\tau) - \sum_{\tau\in\{1,\dots,t-1\}:i\in\tilde{C}_G^\tau} \sigma_i^C(G_i + (\nabla G_i)^T s^\tau)^+ \qquad (59)\\
&= \sum_{j=1}^{N_i}\left(\sum_{\tau=\underline{\tau}_{j,i}}^{\bar{\tau}_{j,i}} \mathcal{D}_i(\tau,\tau) - \sigma_i^C(G_i + (\nabla G_i)^T s^{\bar{\tau}_{j,i}})^+\right)\\
&\le \sum_{j=1}^{N_i}\hat{\mathcal{D}}_i(\underline{\tau}_{j,i},\bar{\tau}_{j,i}) = \sum_{\tau\in\{1,\dots,t\}:i\in C_G^\tau}\hat{\mathcal{D}}_i(\tau,\tau).
\end{aligned}$$

Similar arguments show that

$$\begin{aligned}
&\sum_{\tau\in\{1,\dots,t\}:i\in C_H^\tau} \mathcal{E}_i(\tau,\tau) - \sum_{\tau\in\{1,\dots,t-1\}:i\in\tilde{C}_H^\tau} \sigma_i^C(H_i + (\nabla H_i)^T s^\tau)^+ \qquad (60)\\
&\le \begin{cases}\sum_{\tau\in\{1,\dots,t\}:i\in C_H^\tau}\hat{\mathcal{E}}_i(\tau,\tau) & \text{if } \bar{\tau}_{N_i,i} = t,\\ \sum_{\tau\in\{1,\dots,t-1\}:i\in C_H^\tau}\hat{\mathcal{E}}_i(\tau,\tau) + \tilde{\mathcal{E}}_i(t,t) & \text{if } \bar{\tau}_{N_i,i} < t,\end{cases}
\end{aligned}$$

where $\hat{\mathcal{E}}_i$ and $\tilde{\mathcal{E}}_i$ are defined analogously to $\hat{\mathcal{D}}_i$ and $\tilde{\mathcal{D}}_i$, in particular

$$\begin{aligned}
\hat{\mathcal{E}}_i(\tau,\tau) &:= (\sigma_i^C - \tilde{\lambda}_i^H)(\beta_i^{H,\tau}\delta^\tau - \beta_i^{H,\tau-1}\delta^{\tau-1})|H_i| - (\sigma_i^C - \tilde{\lambda}_i^G)(\beta_i^{G,\tau}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1})(G_i)^-,\\
\tilde{\mathcal{E}}_i(\tau,\tau) &:= (\sigma_i^C - \tilde{\lambda}_i^H)(\beta_i^{H,\tau-1}\delta^\tau - \beta_i^{H,\tau-1}\delta^{\tau-1})|H_i| - (\sigma_i^C - \tilde{\lambda}_i^G)(\beta_i^{G,\tau}\delta^\tau - \beta_i^{G,\tau-1}\delta^{\tau-1})(G_i)^-.
\end{aligned}$$

Since for every $\tau \in \{1,\dots,t\}$ the index sets C_G^τ and C_H^τ form a partition of C, the index i exactly belongs to one of the two sets C_G^τ, C_H^τ. Since we also have $\hat{\mathcal{D}}_i(t,t) \le \tilde{\mathcal{D}}_i(t,t)$

and $\hat{\mathcal{E}}_i(t,t) \le \tilde{\mathcal{E}}_i(t,t)$, we obtain from (58),(59) and (60), the estimate

$$\begin{aligned}
&\sum_{\tau\in\{1,\dots,t\}:i\in C_G^\tau} \mathcal{D}_i(\tau,\tau) - \sum_{\tau\in\{1,\dots,t-1\}:i\in\tilde{C}_G^\tau} \sigma_i^C(G_i+(\nabla G_i)^T s^\tau)^+\\
&+\sum_{\tau\in\{1,\dots,t\}:i\in C_H^\tau} \mathcal{E}_i(\tau,\tau) - \sum_{\tau\in\{1,\dots,t-1\}:i\in\tilde{C}_H^\tau} \sigma_i^C(H_i+(\nabla H_i)^T s^\tau)^+\\
&\qquad\le \sum_{\tau=1}^{t-1} \max\{\hat{\mathcal{D}}_i(\tau,\tau),\hat{\mathcal{E}}_i(\tau,\tau)\} + \max\{\tilde{\mathcal{D}}_i(t,t),\tilde{\mathcal{E}}_i(t,t)\}. \qquad (61)
\end{aligned}$$

Taking into account

$$\begin{aligned}
&\max\{\hat{\mathcal{D}}_i(\tau,\tau),\hat{\mathcal{E}}_i(\tau,\tau)\}\\
&\quad\le\ -(\sigma_i^C-\max\{\tilde\lambda_i^G,\tilde\lambda_i^H\})\Big((\beta_i^{G,\tau}\delta^\tau-\beta_i^{G,\tau-1}\delta^{\tau-1})(G_i)^-+(\beta_i^{H,\tau}\delta^\tau-\beta_i^{H,\tau-1}\delta^{\tau-1})(H_i)^-\Big)
\end{aligned}$$

and

$$\max\{\tilde{\mathcal{D}}_i(t,t),\tilde{\mathcal{E}}_i(t,t)\}\le-(\sigma_i^C-\max\{\tilde\lambda_i^G,\tilde\lambda_i^H\})(\delta^t-\delta^{t-1})\Big(\beta_i^{G,t-1}(G_i)^-+\beta_i^{H,t-1}(H_i)^-\Big)$$

it follows that

$$\begin{aligned}
&\sum_{\tau=1}^{t-1} \max\{\hat{\mathcal{D}}_i(\tau,\tau),\hat{\mathcal{E}}_i(\tau,\tau)\} + \max\{\tilde{\mathcal{D}}_i(t,t),\tilde{\mathcal{E}}_i(t,t)\}\\
&\quad\le\ -(\sigma_i^C-\max\{\tilde\lambda_i^G,\tilde\lambda_i^H\})\Big((\beta_i^{G,t-1}\delta^{t-1}-\beta_i^{G,0}\delta^0)(G_i)^-+(\beta_i^{H,t-1}\delta^{t-1}-\beta_i^{H,0}\delta^0)(H_i)^-\\
&\qquad\qquad+(\delta^t-\delta^{t-1})\big(\beta_i^{G,t-1}(G_i)^-+\beta_i^{H,t-1}(H_i)^-\big)\Big)\\
&\quad=\ -(\sigma_i^C-\max\{\tilde\lambda_i^G,\tilde\lambda_i^H\})\Big((\beta_i^{G,t-1}\delta^t-\beta_i^{G,0}\delta^0)(G_i)^-+(\beta_i^{H,t-1}\delta^t-\beta_i^{H,0}\delta^0)(H_i)^-\Big)\\
&\quad\le\ -(\sigma_i^C-\max\{\tilde\lambda_i^G,\tilde\lambda_i^H\})\Big(\beta_i^{G,0}(\delta^t-1)(G_i)^-+\beta_i^{H,0}(\delta^t-1)(H_i)^-\Big), \qquad (62)
\end{aligned}$$

where we have used $\beta_i^{G,t-1}\le\beta_i^{G,0}$, $\beta_i^{H,t-1}\le\beta_i^{H,0}$ and $\delta^0=1$. Thus the claimed inequality (50) follows from (61) and (62).

Next we prove for every $i\in I$ the estimate

$$\sigma_i^g\big((g_i+(\nabla g_i)^T s^t)^+-(g_i+(\nabla g_i)^T s^0)^+\big)-\sum_{\tau=1}^{t}\lambda_i^{g,\tau}(\nabla g_i)^T(s^\tau-s^{\tau-1})\le(\sigma_i^g-\tilde\lambda_i^g)(\delta^t-1)\beta_i^{g,0}(g_i)^+. \qquad (63)$$

For every $\tau=1,\dots,t$ we have $\lambda_i^{g,\tau}\ge 0$, $\beta_i^{g,\tau-1}=\beta_i^{g,0}$ and

$$\lambda_i^{g,\tau}\big((1-\beta_i^{g,\tau-1}\delta^\tau)g_i+(\nabla g_i)^T s^\tau\big)=0,\quad \lambda_i^{g,\tau}\big((1-\beta_i^{g,\tau-1}\delta^{\tau-1})g_i+(\nabla g_i)^T s^{\tau-1}\big))\le 0,$$

implying, together with $\delta^\tau\le\delta^{\tau-1}$,

$$\lambda_i^{g,\tau}(\nabla g_i)^T(s^\tau-s^{\tau-1})\ge\lambda_i^{g,\tau}\beta_i^{g,0}(\delta^\tau-\delta^{\tau-1})g_i\ge\lambda_i^{g,\tau}\beta_i^{g,0}(\delta^\tau-\delta^{\tau-1})(g_i)^+\ge\tilde\lambda_i^g\beta_i^{g,0}(\delta^\tau-\delta^{\tau-1})(g_i)^+.$$

Summing up this inequalities yields

$$\sum_{\tau=1}^{t} \lambda_i^{g,\tau} (\nabla g_i)^T (s^\tau - s^{\tau-1}) \geq \tilde{\lambda}_i^g \beta_i^{g,0} (\delta^t - \delta^0)(g_i)^+ = \tilde{\lambda}_i^g \beta_i^{g,0} (\delta^t - 1)(g_i)^+.$$

Since by feasibility of (s^t, δ^t) for $QPCC(\beta^{\tau-1}, \rho)$ we have $(g_i + (\nabla g_i)^T s^t)^+ \leq \beta_i^{g,0} \delta^t (g_i)^+$ and due to our choice of $\beta_i^{g,0}$ we have $(g_i + (\nabla g_i)^T s^0)^+ = (g_i)^+ = \beta_i^{g,0}(g_i)^+$, the inequality (63) follows.

Similar arguments show, that for every $i \in E$ we have

$$\sigma_i^h (|h_i + (\nabla h_i)^T s^t| - |h_i + (\nabla h_i)^T s^0|) - \sum_{\tau=1}^{t} \lambda_i^{h,\tau} (\nabla h_i)^T (s^\tau - s^{\tau-1}) \leq (\sigma_i^h - \tilde{\lambda}_i^h)(\delta^t - 1)|h_i|. \tag{64}$$

Then the relation (39) follows from (49), (50), (63) and (64) and the second estimate (40) is an easy consequence of (33) and (38). □

5.1.2. Searching for the next iterate

We choose the next iterate as a point from the polygonal line connecting the points $s_k^0, \ldots, s_k^{N_k}$. First we parametrize this line by its length as a curve $\hat{s}_k : [0,1] \to \mathbb{R}^n$ in the following way. We define $t_k(1) := N_k$, for every $\gamma \in [0,1)$ we denote by $t_k(\gamma)$ the smallest number t such that $S_k^t / S_k^{N_k} > \gamma$ and we set $\alpha_k(1) := 1$,

$$\alpha_k(\gamma) := \frac{\gamma - S_k^{t_k(\gamma)-1} / S_k^{N_k}}{S_k^{t_k(\gamma)} / S_k^{N_k} - S_k^{t_k(\gamma)-1} / S_k^{N_k}} = \frac{\gamma S_k^{N_k} - S_k^{t_k(\gamma)-1}}{\| s_k^{t_k(\gamma)} - s_k^{t_k(\gamma)-1} \|}, \gamma \in [0,1)$$

where $S_k^0 := 0, S_k^t := \sum_{\tau=1}^{t} \| s_k^\tau - s_k^{\tau-1} \|$ for $t = 1, \ldots, N_k$. Then we define

$$\hat{s}_k(\gamma) = s_k^{t_k(\gamma)-1} + \alpha_k(\gamma)(s_k^{t_k(\gamma)} - s_k^{t_k(\gamma)-1}).$$

Now consider some sequence of positive numbers $\gamma_1^k = 1, \gamma_2^k, \gamma_3^k, \ldots$ with $1 > \bar{\gamma} \geq \gamma_{j+1}^k / \gamma_j^k \geq \underline{\gamma} > 0$ for all $j \in \mathbb{N}$ and let us shortly denote $t_k^j := t_k(\gamma_j^k), \alpha_k^j := \alpha_k(\gamma_j^k)$. Consider the smallest j, denoted by $j(k)$ such that for some given constant $\xi \in (0,1)$ one has

$$\phi_k^{t_k^j}(\alpha_k^j) - \phi_k^1(0) \leq \xi \left((1 - \alpha_k^j)(\hat{\phi}_k^{t_k^j - 1}(1) - \hat{\phi}_k^1(0)) + \alpha_k^j (\hat{\phi}_k^{t_k^j}(1) - \hat{\phi}_k^1(0)) \right), \tag{65}$$

where in case $t_k^j = 1$ we define $\hat{\phi}_k^0(1) := \hat{\phi}_k^1(0)$. Then the new iterate is given by

$$x_{k+1} := x_k + \hat{s}_k(\gamma_{j(k)}^k).$$

Lemma 5.5. The new iterate x_{k+1} is well defined.

Proof. In order to show that the new iterate is well defined, we have to prove the existence of some j such that (65) is fulfilled. Let τ_k be the smallest natural number such that $s_k^{\tau_k} \neq 0$. Note that this implies $s_k^t = 0$ for $t = 0, \dots \tau_k - 1$ and consequently $S_k^{\tau_k - 1} = 0$. There is some $\delta_k > 0$ such that $|\phi_k^{\tau_k}(\alpha) - \hat{\phi}_k^{\tau_k}(\alpha)| \leq \alpha(1-\xi)(\hat{\phi}_k^1(0) - \hat{\phi}_k^{\tau_k}(1))$, whenever $0 \leq \alpha \leq \delta_k$. Since $\lim_{j\to\infty} \gamma_j^k = 0$, we can choose j as large that $t_k^j = \tau_k$ and $\alpha_k^j \leq \delta_k$ yielding

$$\phi_k^{\tau_k}(\alpha_k^j) - \hat{\phi}_k^{\tau_k}(\alpha_k^j) \leq (1-\xi)\alpha_k^j(\hat{\phi}_k^1(0) - \hat{\phi}_k^{\tau_k}(1)). \tag{66}$$

Then by convexity of $\hat{\phi}_k^{\tau_k}$ from Lemma 5.2, taking into account $\hat{\phi}_k^{\tau_k}(0) \leq \hat{\phi}_k^{\tau_k - 1}(1) \leq \hat{\phi}_k^1(0)$ by Lemma 5.4 and Proposition 5.3 and $\phi_k^1(0) = \hat{\phi}_k^1(0)$ we obtain

$$\begin{aligned}
\phi_k^{\tau_k}(\alpha_k^j) - \phi_k^1(0) &\leq \hat{\phi}_k^{\tau_k}(\alpha_k^j) + (1-\xi)\alpha_k^j(\hat{\phi}_k^1(0) - \hat{\phi}_k^{\tau_k}(1)) - \hat{\phi}_k^1(0) \\
&\leq (1-\alpha_k^j)\hat{\phi}_k^{\tau_k}(0) - (1-\alpha_k^j)\hat{\phi}_k^1(0) + \xi\alpha_k^j(\hat{\phi}_k^{\tau_k}(1) - \hat{\phi}_k^1(0)) \\
&\leq (1-\alpha_k^j)\left((\xi\hat{\phi}_k^{\tau_k-1}(1) + (1-\xi)\hat{\phi}_k^1(0)) - \hat{\phi}_k^1(0)\right) + \xi\alpha_k^j(\hat{\phi}_k^{\tau_k}(1) - \hat{\phi}_k^1(0)) \\
&= \xi\left((1-\alpha_k^j)(\hat{\phi}_k^{\tau_k-1}(1) - \hat{\phi}_k^1(0)) + \alpha_k^j(\hat{\phi}_k^{\tau_k}(1) - \hat{\phi}_k^1(0))\right).
\end{aligned}$$

Thus (65) is fulfilled for this j and the lemma is proved. □

5.2. Convergence of the algorithm

We consider the behavior of the Algorithm 5.1 when it does not prematurely stop and it generates an infinite sequence of iterates

$$x_k, B_k, (s_k^t, \delta_k^t), \lambda_k^t, J_k^t, \tilde{J}_k^t, \beta_k^t, t = 0, \dots, N_k.$$

We discuss the convergence behavior under the following assumption.

Assumption 1.

1. Algorithm 4.1 can be run for every k, i. e. we can find a working set J at the initialization step.
2. There exist constants C_x, C_s, C_λ such that

   $$\|x_k\| \leq C_x, \quad S^{N_k} \leq C_s, \quad \tilde{\lambda}_k^h, \tilde{\lambda}_k^g, \tilde{\lambda}_k^G, \tilde{\lambda}_k^H \leq C_\lambda$$

   for all k, where $\tilde{\lambda}_k^h := \max_{i\in E}\{\tilde{\lambda}_{i,k}^h\}$, $\tilde{\lambda}_k^g := \max_{i\in I}\{\tilde{\lambda}_{i,k}^g\}$, $\tilde{\lambda}_k^G := \max_{i\in C}\{\tilde{\lambda}_{i,k}^G\}$, $\tilde{\lambda}_k^H := \max_{i\in C}\{\tilde{\lambda}_{i,k}^H\}$.
3. There exist constants $\bar{C}_B, \underline{C}_B$ such that $\underline{C}_B \leq \lambda(B_k), \|B_k\| \leq \bar{C}_B$ for all k, where $\lambda(B_k)$ denotes the smallest eigenvalue of B_k.

For our convergence analysis we need one more merit function

$$\begin{aligned}
\Phi_k(x) &:= f(x) + \sum_{i\in E}\sigma_{i,k}^h|h_i(x)| + \sum_{i\in I}\sigma_{i,k}^g(g_i(x))^+ \\
&\quad + \sum_{i\in C}\sigma_{i,k}^C(|\min\{G_i(x), H_i(x)\}| - (\max\{G_i(x), H_i(x)\})^-).
\end{aligned}$$

Lemma 5.6. For each k, for every $t = 1, \dots, N_k$ and for any $\alpha \in [0, 1]$ it holds

$$\Phi_k(x_k + s) \le \phi_k^t(\alpha) \quad \text{and} \quad \Phi_k(x_k) = \phi_k^1(0), \tag{67}$$

where $s = (1 - \alpha)s_k^{t-1} + \alpha s_k^t$.

Proof. Let $A, B \in \mathbb{R}$. We start with the proof of the inequality

$$|\min\{A, B\}| - (\max\{A, B\})^- \le |\max\{A, B\}| - (\min\{A, B\})^-. \tag{68}$$

If $\min\{A, B\} < 0$ we have $|\min\{A, B\}| = -(\min\{A, B\})^-$ and since it always holds that $-(\max\{A, B\})^- \le |\max\{A, B\}|$ we obtain (68).

In the case $\min\{A, B\} \ge 0$ we have $0 \le \min\{A, B\} \le \max\{A, B\}$ and hence

$$|\min\{A, B\}| - (\max\{A, B\})^- = \min\{A, B\} \le \max\{A, B\} = |\max\{A, B\}| - (\min\{A, B\})^-,$$

showing the validity of (68).

Thus, for every $i \in C_{G,k}^t$ we have

$$\begin{aligned} &|G_i(x_k + s)| - (H_i(x_k + s))^- \\ &\quad \ge \min\{|\min\{G_i(x_k + s), H_i(x_k + s)\}| - (\max\{G_i(x_k + s), H_i(x_k + s)\})^-, \\ &\qquad\quad |\max\{G_i(x_k + s), H_i(x_k + s)\}| - (\min\{G_i(x_k + s), H_i(x_k + s)\})^-\} \\ &\quad = |\min\{G_i(x_k + s), H_i(x_k + s)\}| - (\max\{G_i(x_k + s), H_i(x_k + s)\})^- \end{aligned}$$

because of (68). Analogously we obtain

$$|H_i(x_k+s)| - (G_i(x_k+s))^- \ge |\min\{G_i(x_k+s), H_i(x_k+s)\}| - (\max\{G_i(x_k+s), H_i(x_k+s)\})^-$$

for every $i \in C_{H,k}^t$ and the claimed inequality $\Phi_k(x_k + s) \le \phi_k^t(\alpha)$ follows.

To show $\Phi_k(x_k) = \phi_k^1(0)$ consider $i \in C_{G,k}^1$. If $G_i(x_k) \le H_i(x_k)$ then we obviously have

$$|G_i(x_k)| - (H_i(x_k))^- = |\min\{G_i(x_k), H_i(x_k)\}| - (\max\{G_i(x_k), H_i(x_k)\})^-. \tag{69}$$

On the other hand, if $G_i(x_k) > H_i(x_k)$ we have $0 \ge G_i(x_k) > H_i(x_k)$ because of $i \in C_{G,k}^1 \subset \tilde{J}_{G,k}^0$ and (69) again follows. Analogously, one shows

$$|H_i(x_k)| - (G_i(x_k))^- = |\min\{G_i(x_k), H_i(x_k)\}| - (\max\{G_i(x_k), H_i(x_k)\})^-$$

for $i \in C_{H,k}^1$ and the second claim follows. $\square$

An easy consequence of the way how we define the penalty parameters in (36) is the following lemma.

Lemma 5.7. Under Assumption 1 there exists some $\bar{k}$ such that for all $k \ge \bar{k}$ the penalty parameters remain constant, $\bar{\sigma} := \sigma_k$ and consequently $\Phi_k(x) = \Phi_{\bar{k}}(x)$.

Remark 5.8. Note that we do not use Φ_k for calculating the new iterate because the first order approximation is in general not convex on the line segments connecting s_k^{t-1} and s_k^t due to the involved min operation.

Lemma 5.9. Assume that Assumption 1 is fulfilled. Then

$$\lim_{k\to\infty} \phi_k^{t_k^{j(k)}}(\alpha_k^{j(k)}) - \phi_k^1(0) = 0. \tag{70}$$

P r o o f. Take an existed $\bar{k}$ from Lemma 5.7. Then we have for $k \geq \bar{k}$

$$\begin{aligned} \Phi_{k+1}(x_{k+1}) &= \Phi_{\bar{k}}(x_{k+1}) = \Phi_{\bar{k}}(x_k + \hat{s}_k(\gamma_{j(k)}^k)) = \Phi_k(x_k + \hat{s}_k(\gamma_{j(k)}^k)) \leq \phi_k^{t_k^{j(k)}}(\alpha_k^{j(k)}) \\ &< \phi_k^1(0) = \Phi_k(x_k) \end{aligned}$$

and therefore $\Phi_{k+1}(x_{k+1}) - \Phi_k(x_k) \leq \phi_k^{t_k^{j(k)}}(\alpha_k^{j(k)}) - \phi_k^1(0) < 0$. Hence the sequence $\Phi_k(x_k)$ is monotonically decreasing and therefore convergent, because it is bounded below by Assumption 1. Hence

$$-\infty < \lim_{k\to\infty} \Phi_k(x_k) - \Phi_{\bar{k}}(x_{\bar{k}}) = \sum_{k=\bar{k}}^{\infty} (\Phi_{k+1}(x_{k+1}) - \Phi_k(x_k)) \leq \sum_{k=\bar{k}}^{\infty} (\phi_k^{t_k^{j(k)}}(\alpha_k^{j(k)}) - \phi_k^1(0))$$

and the assertion follows. □

Proposition 5.10. Assume that Assumption 1 is fulfilled. Then

$$\lim_{k\to\infty} \hat{\phi}_k^{N_k}(1) - \hat{\phi}_k^1(0) = 0 \tag{71}$$

and consequently

$$\lim_{k\to\infty} \|s_k^{N_k}\| = 0. \tag{72}$$

P r o o f. We prove (71) by contraposition. Assuming on the contrary that (71) does not hold, by taking into account $\hat{\phi}_k^{N_k}(1) - \hat{\phi}_k^1(0) \leq 0$ by Proposition 5.3, there exists a subsequence $K = \{k_1, k_2, \ldots\}$ such that $\hat{\phi}_k^{N_k}(1) - \hat{\phi}_k^1(0) \leq \bar{r} < 0$. By passing to a subsequence we can assume that for all $k \in K$ we have $k \geq \bar{k}$ with $\bar{k}$ given by Lemma 5.7 and $N_k = \bar{N}$, where we have taken into account (34). By passing to a subsequence once more we can also assume that

$$\lim_{k \overset{K}{\to} \infty} S_k^t = \bar{S}^t, \ \lim_{k \overset{K}{\to} \infty} r_k^t = \bar{r}^t,$$

where $r_k^t := \hat{\phi}_k^t(1) - \hat{\phi}_k^1(0)$. Note that $\bar{r}^{\bar{N}} \leq \bar{r} < 0$.

Let us first consider the case $\bar{S}^{\bar{N}} = 0$. There exists $\delta > 0$ such that $|\phi_k^{\bar{N}}(\alpha) - \hat{\phi}_k^{\bar{N}}(\alpha)| \leq (\xi-1)\bar{r}^{\bar{N}}\|(1-\alpha)s_k^{\bar{N}-1} + \alpha s_k^{\bar{N}}\| \,\forall k \in K$, whenever $\|(1-\alpha)s_k^{\bar{N}-1} + \alpha s_k^{\bar{N}}\| \leq \delta$. Since $\bar{S}^{\bar{N}} = 0$ implies $s_k^{\bar{N}} \to 0$ we can assume that $\|s_k^{\bar{N}}\| \leq \min\{\delta, 1/2\} \,\forall k \in K$. Then

$$\begin{aligned} \phi_k^{\bar{N}}(1) - \phi_k^1(0) &\leq \hat{\phi}_k^{\bar{N}}(1) - \hat{\phi}_k^1(0) + (\xi - 1)\bar{r}^{\bar{N}}\|s_k^{\bar{N}}\| \qquad (73) \\ &\leq \hat{\phi}_k^{\bar{N}}(1) - \hat{\phi}_k^1(0) + (\xi - 1)(\hat{\phi}_k^{\bar{N}}(1) - \hat{\phi}_k^1(0)) \\ &= \xi(\hat{\phi}_k^{\bar{N}}(1) - \hat{\phi}_k^1(0)) \to \xi\bar{r}^{\bar{N}} < 0 \qquad (74) \end{aligned}$$

and this implies that for the next iterate we have $j = 1$ and hence $t_k^j = \bar{N}$ and $\alpha_k^j = 1$, contradicting (70).

Now consider the case $\bar{S}^N \neq 0$ and let us define the number $\bar{\tau} := \max\{t \,|\, \bar{S}^t = 0\} + 1$. Note that Proposition 5.3 yields

$$r_k^t \leq -\frac{\lambda(B_k)}{2} \sum_{\tau=1}^{t} \|s_k^\tau - s_k^{\tau-1}\|^2 \leq -\frac{\lambda(B_k)}{2} \sqrt{t} \left(\sum_{\tau=1}^{t} \|s_k^\tau - s_k^{\tau-1}\| \right)^2 = -\frac{\lambda(B_k)}{2} \sqrt{t} (S_k^t)^2 \tag{75}$$

and therefore $\tilde{r} := \max_{t \geq \bar{\tau}} \bar{r}^t < 0$. By passing to a subsequence we can assume that for every $t \geq \bar{\tau}$ and every $k \in K$ we have $r_k^t \leq \frac{\bar{r}^t}{2}$.

Now assume that for infinitely many $k \in K$ we have $t_k^{j(k)} > \bar{\tau}$. Then we conclude

$$\begin{aligned} \phi_k^{t_k^{j(k)}}(\alpha_k^{j(k)}) - \phi_k^1(0) &\leq \xi \left((1 - \alpha_k^{j(k)})(\hat{\phi}_k^{t_k^{j(k)}-1}(1) - \hat{\phi}_k^1(0)) + \alpha_k^{j(k)}(\hat{\phi}_k^{t_k^{j(k)}}(1) - \hat{\phi}_k^1(0)) \right) \\ &\leq \frac{\xi}{2} \left((1 - \alpha_k^{j(k)})\tilde{r} + \alpha_k^{j(k)}\tilde{r} \right) \leq \frac{\xi \tilde{r}}{2} < 0 \end{aligned}$$

contradicting (70). Hence for all but finitely many $k \in K$, without loss of generality for all $k \in K$, we have $t_k^{j(k)} \leq \bar{\tau}$.

There exists $\delta > 0$ such that $|\phi_k^{\bar{\tau}}(\alpha) - \hat{\phi}_k^{\bar{\tau}}(\alpha)| \leq \frac{|\bar{r}^{\bar{\tau}}|(1-\xi)\underline{\gamma}\|(1-\alpha)s_k^{\bar{\tau}-1} + \alpha s_k^{\bar{\tau}}\|}{4S^{\bar{\tau}}}$ $\forall k \in K$, whenever $\|(1-\alpha)s_k^{\bar{\tau}-1} + \alpha s_k^{\bar{\tau}}\| \leq \delta$. By eventually choosing δ smaller we can assume $\delta \leq \bar{S}^{\bar{\tau}}$. Note that we have $\|s_k^t\| \leq S_k^t$ for all t and thus $\|s_k^{\bar{\tau}-1}\| \to 0$. This implies $\bar{S}^{\bar{\tau}} = \lim_{k \xrightarrow{K} \infty} S_k^{\bar{\tau}} - S_k^{\bar{\tau}-1} = \lim_{k \xrightarrow{K} \infty} \|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\| = \lim_{k \xrightarrow{K} \infty} \|s_k^{\bar{\tau}}\|$. Hence, by passing to a subsequence if necessary we can assume that for all $k \in K$ we have $\|s_k^{\bar{\tau}-1}\| \leq \delta/4$,

$$\underline{\gamma} \frac{3}{4} \delta \frac{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|}{\|s_k^{\bar{\tau}}\|} > 4(1 - \underline{\gamma}) S_k^{\bar{\tau}-1}, \tag{76}$$

$$\|s_k^{\bar{\tau}}\| \leq \frac{9}{8} \bar{S}^{\bar{\tau}} \tag{77}$$

and

$$\frac{\underline{\gamma} S_k^{\bar{\tau}} - S_k^{\bar{\tau}-1}}{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|} \geq \frac{\underline{\gamma}}{2}. \tag{78}$$

Now let for each k the index $\tilde{j}(k)$ denote the smallest j such that $t_k^j = \bar{\tau}$ and $\alpha_k^j \|s_k^{\bar{\tau}}\| \leq \frac{3}{4}\delta$. Hence we either have $t_k^{\tilde{j}(k)-1} > \bar{\tau}$ or $\alpha_k^{\tilde{j}(k)-1}\|s_k^{\bar{\tau}}\| > \frac{3}{4}\delta$. In the first case we have $\frac{S_k^{\bar{\tau}}}{S_k^N} \leq \gamma_{\tilde{j}(k)-1}^k \leq \frac{\gamma_{\tilde{j}(k)}^k}{\underline{\gamma}}$ and therefore $\alpha_k^{\tilde{j}(k)} \geq \frac{\underline{\gamma} S_k^{\bar{\tau}} - S_k^{\bar{\tau}-1}}{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|} \geq \frac{\underline{\gamma}}{2} \geq \frac{\underline{\gamma}\delta}{2\bar{S}^{\bar{\tau}}}$ by (78). In the second case we have

$$\frac{3}{4}\delta < \alpha_k^{\tilde{j}(k)-1}\|s_k^{\bar{\tau}}\| \leq \frac{\frac{\gamma_{\tilde{j}(k)}^k}{\underline{\gamma}} S_k^{\bar{N}} - S_k^{\bar{\tau}-1}}{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|} \|s_k^{\bar{\tau}}\|.$$

Rearranging yields $\gamma^k_{\tilde{j}(k)} S_k^{\bar{N}} \geq \underline{\gamma}\left(\frac{3}{4}\delta \frac{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|}{\|s_k^{\bar{\tau}}\|} + S_k^{\bar{\tau}-1}\right)$ and therefore, by using (76) and (77),

$$\alpha_k^{\tilde{j}(k)} \geq \frac{\underline{\gamma}\left(\frac{3}{4}\delta \frac{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|}{\|s_k^{\bar{\tau}}\|} + S_k^{\bar{\tau}-1}\right) - S_k^{\bar{\tau}-1}}{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|} > \frac{\underline{\gamma}\frac{9}{16}\delta \frac{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|}{\|s_k^{\bar{\tau}}\|}}{\|s_k^{\bar{\tau}} - s_k^{\bar{\tau}-1}\|} = \frac{9\underline{\gamma}\delta}{16\|s_k^{\bar{\tau}}\|} \geq \frac{\underline{\gamma}\delta}{2\bar{S}^{\bar{\tau}}}.$$

We now prove that $\tilde{j}(k)$ fulfills (65). In fact, since $t_k^{\tilde{j}(k)} = \bar{\tau}$, $\alpha_k^{\tilde{j}(k)} \geq \frac{\underline{\gamma}\delta}{2\bar{S}^{\bar{\tau}}}$ and

$$\|(1 - \alpha_k^{\tilde{j}(k)}) s^{\bar{\tau}-1} + \alpha_k^{\tilde{j}(k)} s^{\bar{\tau}}\| \leq \|s^{\bar{\tau}-1}\| + \alpha_k^{\tilde{j}(k)} \|s^{\bar{\tau}}\| \leq \delta/4 + 3\delta/4 = \delta$$

we conclude

$$\begin{aligned} \phi_k^{\bar{\tau}}(\alpha_k^{\tilde{j}(k)}) - \hat{\phi}_k^{\bar{\tau}}(\alpha_k^{\tilde{j}(k)}) &\leq \frac{|\bar{r}^{\bar{\tau}}|(1-\xi)\underline{\gamma}\|(1 - \alpha_k^{\tilde{j}(k)}) s_k^{\bar{\tau}-1} + \alpha_k^{\tilde{j}(k)} s_k^{\bar{\tau}}\|}{4S^{\bar{\tau}}} \\ &\leq (1-\xi)\frac{\underline{\gamma}\delta}{2S^{\bar{\tau}}}(\hat{\phi}_k^1(0) - \hat{\phi}_k^{\bar{\tau}}(1)) \leq (1-\xi)\alpha_k^{\tilde{j}(k)}(\hat{\phi}_k^1(0) - \hat{\phi}_k^{\bar{\tau}}(1)). \end{aligned}$$

Now we can proceed as in the proof of Lemma 5.5 to show that $\tilde{j}(k)$ fulfills (65).

However, this yields $\tilde{j}(k) \geq j(k)$ by definition of $j(k)$ and hence $\bar{\tau} \geq t_k^{j(k)} \geq t_k^{\tilde{j}(k)} = \bar{\tau}$ showing $t_k^{j(k)} = \bar{\tau}$. But then we also have $\alpha_k^{j(k)} \geq \alpha_k^{\tilde{j}(k)} \geq \frac{\underline{\gamma}\delta}{2\bar{S}^{\bar{\tau}}}$ and from (65) we obtain

$$\begin{aligned} \phi_k^{t_k^{j(k)}}(\alpha_k^{j(k)}) - \phi_k^1(0) &\leq \xi\left((1 - \alpha_k^{j(k)})(\hat{\phi}_k^{t_k^{j(k)}-1}(1) - \hat{\phi}_k^1(0)) + \alpha_k^{j(k)}(\hat{\phi}_k^{t_k^{j(k)}}(1) - \hat{\phi}_k^1(0))\right) \\ &\leq \xi\alpha_k^{j(k)}(\hat{\phi}_k^{t_k^{j(k)}}(1) - \hat{\phi}_k^1(0)) \leq \xi\frac{\underline{\gamma}\delta}{2\bar{S}^{\bar{\tau}}}\frac{\tilde{r}}{2} < 0 \end{aligned}$$

contradicting (70) and so (71) is proved. Condition (72) now follows from (71) because we conclude from (75) that

$$\hat{\phi}_k^{N_k}(1) - \hat{\phi}_k^1(0) \leq -\frac{\lambda(B_k)}{2}\sqrt{N_k}(S_k^{N_k})^2 \leq -\frac{\lambda(B_k)}{2}\sqrt{N_k}(\|s_k^{N_k}\|)^2.$$

□

Now we are ready to state the main result of this section.

Theorem 5.11. Let Assumption 1 be fulfilled. Then every limit point is M-stationary for the problem (1).

Proof. Let $\bar{x}$ denote a limit point of the sequence x_k and let K denote a subsequence such that $\lim_{k \overset{K}{\to} \infty} x_k = \bar{x}$. Further let $\bar{\lambda}$ be a limit point of the bounded sequence $\lambda_k^{N_k}$ and assume without loss of generality that $\lim_{k \overset{K}{\to} \infty} \lambda_k^{N_k} = \bar{\lambda}$. First we show feasibility of $\bar{x}$ for the problem (1) and complementarity of $\bar{\lambda}$.

Consider $i \in I$. For all k it holds that

$$0 \geq \left((1 - \beta_{i,k}^{g,N_k-1}\delta_k^{N_k})g_i(x_k) + (\nabla g_i(x_k))^T s_k^{N_k}\right) \perp \lambda_{i,k}^{g,N_k} \geq 0.$$

Since $0 \le \delta_k^{N_k} \le \zeta$, $0 \le \beta_{i,k}^{g,N_k-1} \le 1$ we have $1 \ge (1-\beta_{i,k}^{g,N_k-1}\delta_k^{N_k}) \ge 1-\zeta$ and together with $s_k^{N_k} \to 0$ by Proposition 5.10 we conclude

$$0 \ge \limsup_{k \xrightarrow{K} \infty} \left(g_i(x_k) + \frac{(\nabla g_i(x_k))^T s_k^{N_k}}{(1-\beta_{i,k}^{g,N_k-1}\delta_k^{N_k})} \right) = g_i(\bar{x}),$$

$\bar{\lambda}_i^g \ge 0$ and

$$0 = \lim_{k \xrightarrow{K} \infty} \lambda_{i,k}^{g,N_k} \left(g_i(x_k) + \frac{(\nabla g_i(x_k))^T s_k^{N_k}}{(1-\beta_{i,k}^{g,N_k-1}\delta_k^{N_k})} \right) = \bar{\lambda}_i^g g_i(\bar{x}).$$

Hence $0 \le \bar{\lambda}_i^g \perp g_i(\bar{x}) \le 0$. Similar arguments show that for every $i \in E$ we have

$$0 = \lim_{k \xrightarrow{K} \infty} \left(h_i(x_k) + \frac{(\nabla h_i(x_k))^T s_k^{N_k}}{(1-\delta_k^{N_k})} \right) = h_i(\bar{x}).$$

Finally consider $i \in C$. Then for infinitely many $k \in K$ we either have

$$(1-\beta_{i,k}^{G,N_k-1}\delta_k^{N_k})G_i + (\nabla G_i)^T s_k^{N_k} \ge 0, \quad (1-\beta_{i,k}^{H,N_k-1}\delta_k^{N_k})H_i + (\nabla H_i)^T s_k^{N_k} = 0$$

or

$$(1-\beta_{i,k}^{G,N_k-1}\delta_k^{N_k})G_i + (\nabla G_i)^T s_k^{N_k} = 0, \quad (1-\beta_{i,k}^{H,N_k-1}\delta_k^{N_k})H_i + (\nabla H_i)^T s_k^{N_k} \ge 0.$$

We consider only the first case because the second one can be treated analogously. Again we have $1 \ge (1-\beta_{i,k}^{G,N_k-1}\delta_k^{N_k}), (1-\beta_{i,k}^{H,N_k-1}\delta_k^{N_k}) \ge 1-\zeta$ and hence

$$0 \le \liminf_{k \xrightarrow{K} \infty} \left(G_i(x_k) + \frac{(\nabla G_i(x_k))^T s_k^{N_k}}{(1-\beta_{i,k}^{G,N_k-1}\delta_k^{N_k})} \right) = G_i(\bar{x}),$$

$$0 = \lim_{k \xrightarrow{K} \infty} \lambda_{i,k}^{G,N_k} \left(G_i(x_k) + \frac{(\nabla G_i(x_k))^T s_k^{N_k}}{(1-\beta_{i,k}^{G,N_k-1}\delta_k^{N_k})} \right) = \bar{\lambda}_i^G G_i(\bar{x})$$

and

$$0 = \lim_{k \xrightarrow{K} \infty} \left(H_i(x_k) + \frac{(\nabla H_i(x_k))^T s_k^{N_k}}{(1-\beta_{i,k}^{H,N_k-1}\delta_k^{N_k})} \right) = H_i(\bar{x}).$$

Hence $0 \le G_i(\bar{x}) \perp H_i(\bar{x}) \ge 0$ and $\bar{\lambda}_i^G G_i(\bar{x}) = \bar{\lambda}_i^H H_i(\bar{x}) = 0$. By first order optimality condition we have

$$B_k s_k^{N_k} + \nabla f(x_k) + \sum_{i \in E} \lambda_{i,k}^{h,N_k} \nabla h_i(x_k) + \sum_{i \in I} \lambda_{i,k}^{g,N_k} \nabla g_i(x_k) - \sum_{i \in C} (\lambda_{i,k}^{G,N_k} \nabla G_i(x_k) + \lambda_{i,k}^{H,N_k} \nabla H_i(x_k)) = 0$$

for each k and by passing to a limit and by taking into account that $B_k s^{N_k} \to 0$ by Proposition 5.10 we obtain

$$\nabla f(\bar{x}) + \sum_{i \in E} \bar{\lambda}_i^h \nabla h_i(\bar{x}) + \sum_{i \in I} \bar{\lambda}_i^g \nabla g_i(\bar{x}) - \sum_{i \in C} (\bar{\lambda}_i^G \nabla G_i(\bar{x}) + \bar{\lambda}_i^H \nabla H_i(\bar{x})) = 0.$$

Now consider $i \in C$. Then for infinitely many $k \in K$ we have one of the following three possibilities:

(i) $i \in (J^{N_k}_{G,k} \setminus J^{N_k}_{H,k})$: this implies $\bar{\lambda}^H_i = 0$.

(ii) $i \in (J^{N_k}_{H,k} \setminus J^{N_k}_{G,k})$: this implies $\bar{\lambda}^G_i = 0$.

(iii) $i \in (J^{N_k}_{G,k} \cap J^{N_k}_{H,k})$: this implies $\bar{\lambda}^G_i, \bar{\lambda}^H_i \geq 0$.

This finishes the proof. □

6. NUMERICAL RESULTS

Algorithm 4.1 was implemented in MATLAB where we did not take care of the numerical performance. We did not use any update methods for factorizing the occurring matrices when one index left or entered the working set and computed the factorization from scratch. As in SQP-methods for nonlinear programming, the matrices B_k were computed by using the BFGS formula for updating the Hessian of the Lagrangian with the modification due to Powell [26] ensuring positive definiteness of B_k.

To perform numerical tests we used a subset of test problems taken from MacMPEC, which is a collection of MPECs in AMPL and is maintained by Leyffer [19]. In MacMPEC there are currently 193 test problems, 54 of them were discarded because of their size, since our AMPL license can only handle problems with up to 300 variables. Further, 11 problems could not be treated because they do not match our format (1). Our algorithm found for 101 of the remaining 128 test problems (79 %) the solution provided by MacMPEC when using the starting point from MacMPEC. For one problem (design-cent-3) we found the optimal solution (which is the same as the one of the better scaled problem design-cent-31), whereas in MacMPEC this problem is marked as infeasible. For 2 problems (2 %) we found a solution which we could prove to be globally optimal, but differs from the one from MacMPEC. For 15 problems (12 %) our algorithm converged to a different solution when using the starting point from MacMPEC, however when changing the starting point our method found the MacMPEC solution. Only for 9 problems (7 %) we could not find any solution. For one problem (taxmcp) our algorithm always stopped with the degeneracy condition fulfilled. For the remaining 8 problems we could not find a working set at some iteration. For example our algorithm did not work for the problem scholtes5 where the feasible region is given by

$$0 \leq x_1 \perp x_3 \geq 0, \quad 0 \leq x_2 \perp x_3 \geq 0.$$

The constraints of the auxiliary problem (4) which we have to solve at some iterate x^k with $x^k_1 > 0, x^k_2 > 0, x^k_3 = 0$ are given by

$$\begin{aligned} 0 \leq (1 - 0 \cdot \delta)x^k_1 + s_1 \quad &\perp \quad (1 - 1 \cdot \delta)x^k_3 + s_3 \geq 0 \\ 0 \leq (1 - 0 \cdot \delta)x^k_2 + s_2 \quad &\perp \quad (1 - 1 \cdot \delta)x^k_3 + s_3 \geq 0 \end{aligned}$$

and it is easy to see that no working set exists for the auxiliary problem at $(s, \delta) = (0, 1)$.

These numerical results indicate that our algorithm behaves very reliable as long as the assumption that we can find a working set is fulfilled.

To better demonstrate the performance of our algorithm we conclude this section by a table with more detailed information about solving 31 chosen problems. We use the following notation.

Problem	name of the test problem
(n,q)	number of variables, number of all constraints
k^*	total number of outer iterations of the SQP method
$(N_0,\dots,N_{k^*-1})$	total numbers of inner iterations corresponding to each outer iteration
$\sum_{k=0}^{k^*-1} j(k)$	overall sum of steps made during line search
$\sharp f_{eval}$	total number of function evaluations, $\sharp f_{eval} = (q+1)\left(k^* + \sum_{k=0}^{k^*-1} j(k)\right)$
$\sharp\nabla f_{eval}$	total number of gradient evaluations, $\sharp\nabla f_{eval} = (q+1)(k^*+1)$

Problem	(n,q)	k^*	$(N_0,\dots,N_{k^*-1})$	$\sum_{k=0}^{k^*-1} j(k)$	$\sharp f_{eval}$	$\sharp\nabla f_{eval}$
bar-truss-3	$(35,47)$	9	$(1,3,1,\dots,1)$	11	960	480
bard1	$(5,9)$	4	$(1,1,1,1)$	4	80	50
bard2	$(12,29)$	6	$(3,1,\dots,1)$	6	360	210
bilevel2	$(16,33)$	5	$(8,3,1,1,1)$	5	340	204
design-cent-21	$(13,19)$	8	$(1,\dots,1)$	11	380	180
design-cent-4	$(22,33)$	7	$(1,\dots,1)$	7	476	272
ex9.1.7	$(17,26)$	2	$(4,1)$	2	108	81
ex9.2.6	$(16,22)$	6	$(3,2,1,1,1,1)$	6	276	161
flp4-1	$(80,90)$	2	$(1,1)$	2	364	273
flp4-2	$(110,170)$	2	$(1,1)$	2	684	513
flp4-3	$(140,240)$	2	$(1,1)$	2	964	723
flp4-4	$(200,350)$	2	$(1,1)$	2	1404	1053
gnash16	$(13,22)$	9	$(5,1,1,1,2,3,1,1,1)$	9	414	230
incid-set1-8	$(118,225)$	5	$(34,4,4,1,1)$	5	2260	1356
incid-set1c-8	$(117,229)$	5	$(39,6,4,1,1)$	5	2300	1380
incid-set2c-8	$(117,229)$	26	$(21,1,5,2,1,\dots,1)$	51	17710	6210
liswet1-050	$(152,203)$	23	$(28,1,\dots,1)$	23	9384	4896
pack-comp2c-8	$(107,203)$	9	$(35,9,8,4,3,\dots,3)$	11	4080	2040
pack-rig1-8	$(87,169)$	14	$(16,4,1,\dots,1)$	22	6120	2550
pack-rig1c-8	$(87,176)$	12	$(15,4,1,\dots,1)$	12	4248	2301
pack-rig2-8	$(85,165)$	13	$(15,5,4,2,1,\dots,1)$	18	5146	2324
pack-rig2c-8	$(85,172)$	9	$(12,2,3,1,1,1,1,2,1)$	9	3114	1730
pack-rig2p-8	$(103,201)$	18	$(17,3,3,2,2,1,2,1,2,2,1,\dots,1)$	35	10706	3838
pack-rig3-8	$(85,165)$	17	$(17,5,5,5,2,2,1,\dots,1)$	30	7802	2988
pack-rig3c-8	$(85,172)$	10	$(16,4,3,1,\dots,1)$	10	3460	1903
portfl-i-1	$(87,99)$	12	$(7,1,\dots,1)$	12	2400	1300
qpec-100-3	$(110,204)$	28	$(307,4,4,2,1,4,3,8,5,2,$ $1,2,1,\dots,1)$	28	11480	5945
qpec-200-1	$(210,404)$	14	$(200,35,17,2,1,\dots,1)$	14	11340	6075
qpec-200-4	$(240,408)$	15	$(278,15,6,1,2,1,\dots,1)$	15	12270	6544
qpec1	$(30,40)$	4	$(11,1,1,1)$	4	328	205
sl1	$(8,14)$	4	$(2,1,1,1)$	4	120	75

We see that only in the first iterations we need a higher number of inner iterations. However, the number of inner iterations needed to solve the auxiliary problem was always less than two times the number of constraints. Usually we need only one inner iteration when we are close to a solution, but it may happen, as in problem pack-comp2c-8, that the initial working set chosen by our algorithm is different from that at the solution of the auxiliary problem.

Locally we observe very fast convergence of the algorithm. For the problems where $\sum_{k=0}^{k^*-1} j(k) \neq k^*$, we include the following table with all the numbers $(j(0),\dots,j(k^*-1))$ of steps made during line search in each outer iteration.

Problem	$(j(0),\dots,j(k^*-1))$
bar-truss-3	$(3,1,\dots,1)$
design-cent-21	$(1,2,1,1,1,1,3,1)$
incid-set2c-8	$(1,1,3,1,1,1,1,2,1,1,5,5,4,2,2,2,3,2,1,3,1,1,2,3,1,1)$
pack-comp2c-8	$(1,2,2,1,\dots,1)$
pack-rig1-8	$(1,1,2,2,1,3,1,4,2,1,\dots,1)$
pack-rig2-8	$(1,3,1,2,3,1,\dots,1)$
pack-rig2p-8	$(1,1,7,4,3,3,1,1,1,1,1,4,2,1,\dots,1)$
pack-rig3-8	$(1,3,5,3,4,2,2,1,\dots,1)$

ACKNOWLEDGEMENT

This work was supported by the Austrian Science Fund (FWF) under grant P 26132-N25.

(Received April 30, 2015)

REFERENCES

[1] M. Anitescu: Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints. SIAM J. Optim. *16* (2005), 120–145. DOI:10.1137/040606855

[2] M. Anitescu: On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints. SIAM J. Optim. *15* (2005), 1203–1236. DOI:10.1137/s1052623402401221

[3] L. T. Biegler and A. U. Raghunathan: An interior point method for mathematical programs with complementarity constraints (MPCCs). SIAM J. Optim. *15* (2005), 720–750. DOI:10.1137/s1052623403429081

[4] V. DeMiguel, M. P. Friedlander, F. J. Nogales, and S. Scholtes: A two-sided relaxation scheme for mathematical programs with equilibrium constraints. SIAM J. Optim. *16* (2005), 587–609. DOI:10.1137/04060754x

[5] F. Facchinei, H. Jiang, and L. Qi: A smoothing method for mathematical programs with equilibrium constraints. Math. Programming *85* (1999), 107–134. DOI:10.1007/s101070050048

[6] R. Fletcher: Practical Methods of Optimization 2: Constrained Optimization. John Wiley and Sons, Chichester 1981.

[7] R. Fletcher and S. Leyffer: Solving mathematical programs with complementary constraints as nonlinear programs. Optim. Methods Softw. *19* (2004), 15–40.) DOI:10.1080/1055678041001654241

[8] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes: Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. *17* (2006), 259–286. DOI:10.1137/s1052623402407382

[9] M. Fukushima and P. Tseng: An implementable active-set algorithm for computing a b-stationary point of a mathematical program with linear complementarity constraints. SIAM J. Optim. *12* (2002), 724–739. DOI:10.1137/s1052623499363232

[10] H. Gfrerer: Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints. SIAM J. Optim. *24* (2014), 898–931. DOI:10.1137/130914449

[11] G. Giallombardo and D. Ralph: Multiplier convergence in trust region methods with application to convergence of decomposition methods for MPECs. Math. Programming *112* (2008), 335–369. DOI:10.1007/s10107-006-0020-5

[12] X. M. Hu and D. Ralph: Convergence of a penalty method for mathematical programming with complementarity constraints. J. Optim. Theory Appl. *123* (2004), 365–390. DOI:10.1007/s10957-004-5154-0

[13] A. F. Izmailov, A. L. Pogosyan, and M. V. Solodov: Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints. Computational Optim. Appl. *51* (2012), 199–221. DOI:10.1007/s10589-010-9341-7

[14] H. Jiang and D. Ralph: QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints. Comput. Optim. Appl. *13* (1999), 25–59.

[15] H. Jiang and D. Ralph: Extension of quasi-newton methods to mathematical programs with complementarity constraints. Comput. Optim. Appl. *25* (2002), 123–150.

[16] A. Kadrani, J. P. Dussault, and A. Benchakroun: A new regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. *20* (2009), 78–103. DOI:10.1137/070705490

[17] C. Kanzow and A. Schwartz: A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. SIAM J. Optim. *23* (2013), 770–798.

[18] C. Kanzow and A. Schwartz: The price of inexactness: convergence properties of relaxation methods for mathematical programs with equilibrium constraints revisited. Math. Oper. Res. *40* (2015), 2, 253–275. DOI:10.1287/moor.2014.0667

[19] S. Leyffer: MacMPEC: AMPL collection of MPECs, 2000. https://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC

[20] S. Leyffer, G. López-Calva, and J. Nocedal: Interior methods for mathematical programs with complementarity constraints. SIAM J. Optim. *17* (2006), 52–77. DOI:10.1137/040621065

[21] G. H. Lin and M. Fukushima: A modified relaxation scheme for mathematical programs with complementarity constraints. Ann. Oper. Res. *133* (2005), 63–84. DOI:10.1007/s10479-004-5024-z

[22] Z. Q. Luo, J. S. Pang, and D. Ralph: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge, New York, Melbourne 1996. DOI:10.1017/cbo9780511983658

[23] Z. Q. Luo, J. S. Pang, and D. Ralph: Piecewise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints. In: Multilevel Optimization: Algorithms, Complexity, and Applications (A. Migdalas, P. Pardalos, and P. Värbrand, eds.), Kluwer Academic Publishers, Dordrecht 1998, pp. 209–229. DOI:10.1007/978-1-4613-0307-7_9

[24] Z. Q. Luo, J. S. Pang, D. Ralph, and S. Q. Wu: Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints. Math. Programming *75* (1996), 19–76. DOI:10.1007/bf02592205

[25] J. V. Outrata, M. Kočvara, and J. Zowe: Nonsmooth Approach to Optimization Problems with Equilibrium Constraints, Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht 1998. DOI:10.1007/978-1-4757-2825-5

[26] M. J. D. Powell: A fast algorithm for nonlinearly constrained optimization calculations. In: Numerical Analysis Dundee 1977 (G. A. Watson, ed.), Lecture Notes in Mathematics *630*, Springer, Berlin, 1978, pp. 144–157. DOI:10.1007/bfb0067703

[27] D. Ralph and S. J. Wright: Some properties of regularization and penalization schemes for MPECs. Optim. Methods Software *19* (2004), 527–556. DOI:10.1080/10556780410001709439

[28] H. Scheel and S. Scholtes: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Math. Oper. Res. *25* (2000), 1–22. DOI:10.1287/moor.25.1.1.15213

[29] S. Scholtes: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. *11* (2001), 918–936. DOI:10.1137/s1052623499361233

[30] S. Scholtes and M. Stöhr: Exact penalization of mathematical programs with equilibrium constraints. SIAM J. Control Optim. *37* (1999), 617–652. DOI:10.1137/s0363012996306121

[31] S. Steffensen and M. Ulbrich: A new regularization scheme for mathematical programs with equilibrium constraints. SIAM J. Optim. *20* (2010), 2504–2539. DOI:10.1137/090748883

[32] O. Stein: Lifting mathematical programs with complementarity constraints. Math. Programming *131* (2012), 71–94. DOI:10.1007/s10107-010-0345-y

[33] M. Stöhr: Nonsmooth Trust Region Methods and their Applications to Mathematical Programs with Equilibrium Constraints. Shaker-Verlag, Aachen 1999.

[34] J. Zhang and G. Liu: A new extreme point algorithm and its application in psqp algorithms for solving mathematical programs with linear complementarity constraints. J. Glob. Optim. *19* (2001), 335–361.

Matus Benko, Institute of Computational Mathematics, Johannes Kepler University Linz, A-4040 Linz. Austria.
e-mail: benko@numa.uni-linz.ac.at

Helmut Gfrerer, Institute of Computational Mathematics, Johannes Kepler University Linz, A-400 Linz. Austria.
e-mail: helmut.gfrerer@jku.at