

A feasible active set method for strictly convex quadratic problems with simple bounds *

P. Hungerländer, F. Rendl
Alpen-Adria Universität Klagenfurt
Institut für Mathematik
9020 Klagenfurt, Austria
philipp.hungerlaender@uni-klu.ac.at, franz.rendl@uni-klu.ac.at

May 22, 2015

Abstract

A primal-dual active set method for quadratic problems with bound constraints is presented which extends the infeasible active set approach of Kunisch and Rendl [17]. Based on a guess of the active set, a primal-dual pair (x, α) is computed that satisfies stationarity and the complementary condition. If x is not feasible, the variables connected to the infeasibilities are added to the active set and a new primal-dual pair (x, α) is computed. This process is iterated until a primal feasible solution is generated. Then a new active set is determined based on the feasibility information of the dual variable α . Strict convexity of the quadratic problem is sufficient for the algorithm to stop after a finite number of steps with an optimal solution. Computational experience indicates that this approach also performs well in practice.

Key words. primal-dual active set methods; quadratic programming; box constraints; convex programming

1 Introduction

We consider the convex quadratic optimization problem

$$\min_x J(x), \tag{1a}$$

$$\text{subject to } x \leq b, \tag{1b}$$

where

$$J(x) = \frac{1}{2}x^\top Qx + q^\top x, \tag{2}$$

Q is a positive-definite $n \times n$ matrix, and $b, q \in \mathbb{R}^n$. This problem is well understood from a theoretical point of view with global optimality characterized by the Karush-Kuhn-Tucker conditions, see (3) below. The problem appears as a basic building block in many applications, for instance in the context of sequential quadratic optimization methods to solve more general nonlinear problems. It also appears in problems from mathematical physics (obstacle problems, elasto-plastic torsion) and computer science (compressed sensing, face recognition).

First order methods based on projected gradients are among the oldest approaches to the problem. We refer to Figueiredo et al. [8] for a recent paper on the topic. Active set and primal-dual interior point

*partial financial support from Marie-Curie ITN Project MINO, MC-ITN 316647, is gratefully acknowledged

methods are among the other major solution strategies to solve (1). D’Apuzzo and Marino [6] propose a potential-reduction interior point method for (1) with box constraints and focus on parallel implementation issues. Kim et al. [16] introduce a specially tailored interior point method with a preconditioned conjugate gradient algorithm for the search direction and apply it to large sparse l_1 -regularized least squares problems, which can be formulated in the form (1). The classical active set approach is investigated by Moré and Toraldo [19] and Dostal and Schöberl [7], who combine it with projected gradients. Bergounioux et al. [1, 2] introduce an infeasible active set approach for constrained optimal control problems, which was further investigated by Kunisch and Rendl [17] for general convex quadratic problems of type (1). Hintermüller et al. [12] show that this primal-dual active set approach can in fact be rephrased as a semi-smooth Newton method. This leads to strong local and global convergence results in the case that the Hessian Q satisfies some additional properties. In [12] it is shown that the method converges globally if Q is an M -matrix, and it converges superlinearly if the starting point is sufficiently close to the optimum. In [17] global convergence of this infeasible active set method is shown under a diagonal dominance condition on Q . Convergence in the general case can not be ensured, as shown by Ben Gharbia and Gilbert [9] for the case of P -matrices of order $n \geq 3$ and Curtis et al. [5] in case of general symmetric positive definite matrices of order $n \geq 3$.

The main goal of this paper is to investigate modifications of the Kunisch-Rendl-Algorithm (KR-Algorithm, [17]) to make it globally convergent for all problems with positive definite Hessian Q . Globalizing this method in the context of semi-smooth Newton methods has also been proposed by Byrd et al. [4] and by Curtis et al. [5]. In [4], global convergence is insured by maintaining descent directions in combination with a safeguarding procedure proposed by Judice and Pires [15]. In [5], a more general setting is considered, which includes also linear equations in the constraints. A special treatment is given to variables which change index sets very often, leading to global convergence. A common feature of these approaches and ours is that the change in the size of the active set from one iteration to the next is not bounded, which makes it attractive for large scale problems. The approaches differ however in the way, global convergence is insured.

Our new feasible active set method works as follows. Based on a guess on the active set, a primal-dual pair (x, α) is computed that satisfies stationarity and the complementary condition. If x is not feasible, the primal variables outside their bounds are added to the active set and a new primal-dual pair (x, α) is computed. This process is repeated until a primal feasible solution is generated, thereby ending one iteration. Then a new active set is defined based on the dual feasibility information of α .

The new approach inherits the preferable features of the KR-Algorithm, like simplicity (no tuning parameters), finding the exact numerical solution, insensitivity with respect to initialisation. Moreover strict complementary is not required to be satisfied. We show convergence for any positive-definite Q while the original KR-Algorithm is only guaranteed to converge under additional properties of Q such as Q being an M -matrix [12] or having some form of strong diagonal dominance [17].

The paper is organized as follows. At the end of this section we summarize notation used throughout the paper. In Section 2 we repeat the relevant details of the KR-Algorithm. An in-depth analysis of the convergence behaviour of the KR-Algorithm is contained in Section 3. The main theoretical contributions are contained in Section 4, where we extend the KR-Algorithm and prove the finite termination of the modified KR-Algorithm for any problem of type (1). In Section 6 we show the efficiency of the new algorithm by comparing it to the KR-Algorithm, standard active set, interior point and gradient projection methods on a variety of test problems.

Notation: The following notation will be used throughout. We write $Q \succ 0$ to denote that the matrix Q is positive-definite. For a subset $\mathcal{A} \subseteq \mathcal{N} := \{1, \dots, n\}$ and $x \in \mathbb{R}^n$ we write $x_{\mathcal{A}}$ for the components of x indexed by \mathcal{A} , i.e. $x_{\mathcal{A}} := (x_i)_{i \in \mathcal{A}}$. The complement of \mathcal{A} with respect to \mathcal{N} is denoted by $\overline{\mathcal{A}}$. If Q is a matrix and $\mathcal{A}, \mathcal{B} \subseteq \mathcal{N}$, then $Q_{\mathcal{A}, \mathcal{B}}$ is the submatrix of Q , given by $Q_{\mathcal{A}, \mathcal{B}} = (q_{ij})_{i \in \mathcal{A}, j \in \mathcal{B}}$. We write $Q_{\mathcal{A}}$ for $Q_{\mathcal{A}, \mathcal{A}}$. For $a, b \in \mathbb{R}^n$ we write $a \circ b$ to denote the vector of element-wise products, i.e. $a \circ b := (a_i b_i)_{i \in \mathcal{N}}$.

2 The KR-Algorithm from [17]

It is well known that x together with a vector $\alpha \in \mathbb{R}^n$ of Lagrange multipliers for the simple bound constraints furnished the unique global minimizer of (1) if and only if (x, α) satisfies the KKT system

$$Qx + q + \alpha = 0, \tag{3a}$$

$$\alpha \circ (b - x) = 0, \tag{3b}$$

$$b - x \geq 0, \tag{3c}$$

$$\alpha \geq 0. \tag{3d}$$

The crucial step in solving (1) is to identify those inequalities which are active, i.e. the set $\mathcal{A} \subseteq \mathcal{N}$, where the solution to (1) satisfies $x_{\mathcal{A}} = b_{\mathcal{A}}$. Then, with $\mathcal{I} := \mathcal{N} \setminus \mathcal{A}$, we require $\alpha_{\mathcal{I}} = 0$ for (3b) to hold.

The starting point for the present paper is the infeasible active set method from Kunisch and Rendl [17], which we call the KR-Algorithm for short. Let us briefly describe this method.

The KR-Algorithm is iterative, generating a sequence of active sets $\mathcal{A} \subseteq \mathcal{N}$, until an optimal active set is reached. Specifically, given \mathcal{A} and $\mathcal{I} := \mathcal{N} \setminus \mathcal{A}$, the $2n$ equations (3a), (3b) are solved under the condition $x_{\mathcal{A}} = b_{\mathcal{A}}, \alpha_{\mathcal{I}} = 0$. To simplify notation we introduce for a given \mathcal{A} the following set $\text{KKT}(\mathcal{A})$ of equations:

$$\text{KKT}(\mathcal{A}) \quad Qx + q + \alpha = 0, \quad x_{\mathcal{A}} = b_{\mathcal{A}}, \quad \alpha_{\mathcal{I}} = 0.$$

The solution of $\text{KKT}(\mathcal{A})$ is given by

$$x_{\mathcal{A}} = b_{\mathcal{A}}, x_{\mathcal{I}} = -Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I},\mathcal{A}}b_{\mathcal{A}}), \tag{4}$$

$$\alpha_{\mathcal{I}} = 0, \alpha_{\mathcal{A}} = -q_{\mathcal{A}} - Q_{\mathcal{A}}b_{\mathcal{A}} - Q_{\mathcal{A},\mathcal{I}}x_{\mathcal{I}}. \tag{5}$$

We also write $[x, \alpha] = \text{KKT}(\mathcal{A})$ to indicate that x and α satisfy $\text{KKT}(\mathcal{A})$. In some cases we also write $x = \text{KKT}(\mathcal{A})$ to emphasize that we only use x and therefore do not need the backsolve to get α . When we only carry out the backsolve to get α , we write $\alpha = \text{KKT}(\mathcal{A})$, and it is assumed that the corresponding x is available. The set \mathcal{A} is called **optimal** if $[x, \alpha] = \text{KKT}(\mathcal{A})$ also satisfies $x \leq b, \alpha \geq 0$ since x is then the unique solution.

The iterates of the KR-Algorithm given in Table 1 are well defined, because $\text{KKT}(\mathcal{A})$ has a unique solution for every $\mathcal{A} \subseteq \mathcal{N}$, due to $Q \succ 0$. The update rule for the next set

$$\mathcal{B} := \{i : x_i > b_i \text{ or } \alpha_i > 0\} \tag{6}$$

includes elements from the previous active set where the dual variables have the right sign and also inactive variables which are outside their bounds.

KR-Algorithm
Start: $\mathcal{A} \subseteq \mathcal{N}$
while \mathcal{A} not optimal
$[x, \alpha] = \text{KKT}(\mathcal{A}), \quad \mathcal{A} \leftarrow \{i : x_i > b_i \text{ or } \alpha_i > 0\}$

Table 1: The iterates of the KR-Algorithm.

We are going to modify and extend the convergence arguments from [17], so it will be useful to briefly recall the basic ideas from [17]. The key idea to show finite convergence consists in establishing that some merit function will decrease during successive iterates of the algorithm. This will ensure that no active set will be generated more than once.

Let us look at two consecutive iterations. Suppose that some iteration is carried out with the active set $\mathcal{A} \subseteq \mathcal{N}$. Then we get

$$[x, \alpha] = \text{KKT}(\mathcal{A}), \quad \mathcal{B} := \{i : x_i > b_i \text{ or } \alpha_i > 0\}, \quad \mathcal{J} = \mathcal{N} \setminus \mathcal{B}, \quad [y, \beta] = \text{KKT}(\mathcal{B}).$$

It is a simple exercise to verify that the active sets change in every iteration unless the current primal and dual variables are optimal and therefore the KR-Algorithm stops, see [17]. Let us take a look at a small numerical example to further clarify the workings of the KR-Algorithm:

$$Q = \begin{pmatrix} 1 & 1 & 1/2 \\ 1 & 4/3 & 1/3 \\ 1/2 & 1/3 & 3 \end{pmatrix}, \quad q = \begin{pmatrix} -10 \\ -10 \\ -10 \end{pmatrix}, \quad b = \begin{pmatrix} 8 \\ 1 \\ 2 \end{pmatrix}.$$

The given data yields the active-set-transition-graph depicted in Figure 1.

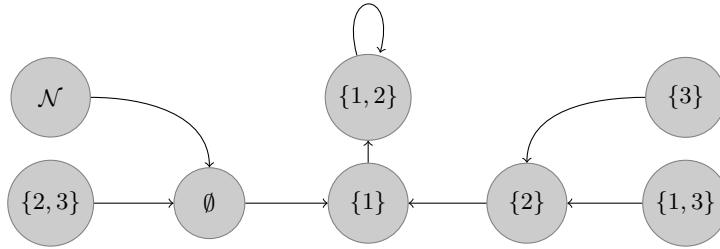


Figure 1: Active-set-transition-graph of the KR-Algorithm for a small numerical example with optimal $\mathcal{A} = \{1, 2\}$.

We close this section with an investigation of the change in the objective function of two consecutive iterations shown in [17].

Lemma 1. [17] *Let x and y be the primal solutions of two consecutive iterations. Then, we have*

$$J(y) - J(x) = -\frac{1}{2}(x - y)^\top Q(x - y) - (x - y)^\top (Qy + q).$$

When moving from set \mathcal{A} to set \mathcal{B} , the elements of \mathcal{A} will either belong to \mathcal{B} or to \mathcal{J} . We denote these sets by \mathcal{B}_1 and \mathcal{J}_1 respectively. In a similar way the elements of \mathcal{I} will also either move to \mathcal{B} or to \mathcal{J} . The respective sets are denoted by \mathcal{B}_2 and \mathcal{J}_2 . Thus,

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2, \quad \mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_2.$$

In Table 2 we summarize the relevant information about x, y, α and β of the KR-Algorithm for two consecutive sets \mathcal{A} and \mathcal{B} . A nonspecified entry indicates that the value of the associated variable cannot be known for certain. Using this information we can work out the change in the objective function.

Lemma 2. [17] *Let (x, α) , (y, β) and \mathcal{B}_2 be given as above. Then, we have*

$$J(y) - J(x) = \frac{1}{2}(x - y)^\top \begin{pmatrix} Q_{\mathcal{B}_2} & 0 \\ 0 & -Q_{\mathcal{B}_2} \end{pmatrix} (x - y).$$

If $\mathcal{B}_2 = \emptyset$, then J decreases, but there is no reason that this should hold in every iteration. Since the primal iterates are not required to be primal feasible, the objective function values of two consecutive iterates may move in an uncontrolled fashion. Thus in [17] a merit function is introduced which also includes a measure of primal infeasibility. The behavior of such merit functions is investigated in the next section.

	\mathcal{A}		\mathcal{I}	
	\mathcal{B}_1	\mathcal{J}_1	\mathcal{J}_2	\mathcal{B}_2
x	=b		$\leq b$	$> b$
y	=b	?		=b
α	$> b$	$\leq b$	=0	
β	?	=0		?

Table 2: Analysis of two consecutive primal-dual iterates of the KR-Algorithm.

3 Convergence for different merit functions

In this section we recall the convergence result from [17] and discuss some extensions and modifications which will be the basis for a new method discussed in detail in the following section.

Let us introduce $r(x) := -\max(x - b, 0) \leq 0$ and the projection $\Pi_\Omega(x)$ of x onto the feasible region $\Omega = \{x \in \mathbb{R}^n : x \leq b\}$. $\Pi_\Omega(x)$ can also be written as

$$\Pi_\Omega(x) = x + r(x) = \begin{cases} b_i, & x_i > b_i, \\ x_i, & x_i \leq b_i. \end{cases} \quad (7)$$

In [17] it is shown that the following merit function

$$M_c(x) = J(x) + \frac{c}{2} \|r(x)\|^2 \quad (8)$$

decreases under the conditions (C1) or (C2) described below. We recall the precise statement of this result and need some more notation to do so. Let $\lambda_{\min} := \lambda_{\min}(Q) > 0$ denote the smallest eigenvalue of Q . Further, let $\nu := \max\{\|Q_{\mathcal{A}, \bar{\mathcal{A}}}\| : \mathcal{A} \subset \mathcal{N}, \mathcal{A} \neq \emptyset, \mathcal{A} \neq \mathcal{N}\}$. We also use the diagonal matrix $D := \text{diag}(q_{11}, \dots, q_{nn})$ consisting of the main diagonal elements of Q and define $d_{\min} := \min\{d_{11}, \dots, d_{nn}\}$. Finally let $r := \|Q - D\|$ denote the norm of Q with the elements from the main diagonal removed. Now we can state the following sufficient conditions (C1) and (C2) for a strict decrease of merit function (8)

$$\text{condition (C1)} \quad \text{cond}(Q) < \left(\frac{\lambda_{\min}}{\nu}\right)^2 - 1,$$

$$\text{condition (C2)} \quad \text{cond}(Q) < \left(\frac{d_{\min}}{r}\right)^2 - 1,$$

where $\text{cond}(Q) = \frac{\|Q\|}{\lambda_{\min}}$.

Theorem 3. [17] *Let (x, α) , (y, β) be two consecutive primal-dual iterates of the KR-Algorithm, and $c \geq \|Q\| + \lambda_{\min}$. If (C1) holds, then we have*

$$2(M_c(y) - M_c(x)) \leq c_1 \|y - x\| < 0,$$

with $c_1 := c \left(\frac{\nu}{\lambda_{\min}}\right)^2 - \lambda_{\min} < 0$. Similarly, (C2) implies that

$$2(M_c(y) - M_c(x)) \leq c_2 \|y - x\| < 0,$$

with $c_2 := c \left(\frac{r}{d_{\min}}\right)^2 - \lambda_{\min} < 0$. In both cases the KR-Algorithm stops after a finite number of iterations with the solution for every $b, q \in \mathbb{R}^n$.

Let us now consider the following merit function

$$L_c(x) = J(x) + \frac{c}{2}r(x)^\top Qr(x). \quad (9)$$

Our aim now is to compile conditions for a decrease of (9) with $c = 1$ and use this knowledge to develop an algorithm with guaranteed convergence behaviour. The merit function L_1 has the following attractive property.

Lemma 4. *Let $[x, \alpha] = KKT(\mathcal{A})$ be a primal-dual iterate with $\Pi_\Omega(x)$ given in (7). Then we can rewrite the merit function (9) with $c = 1$ as the objective value of the projection $\Pi_\Omega(x)$ of x on the feasible set:*

$$L_1(x) = J(\Pi_\Omega(x)).$$

Proof. First we rewrite the term

$$\frac{1}{2}r(x)^\top Qr(x) = \frac{1}{2}r(x)^\top Qr(x) - r(x)^\top \alpha = \frac{1}{2}r(x)^\top Qr(x) + r(x)^\top (Qx + q) = J(r(x)) + r(x)^\top Qx$$

by using (3a) and the fact that $r(x)_i = 0$ for $i \in \mathcal{A}$ and $\alpha_i = 0$ for $i \in \mathcal{I}$. Now we can conclude by applying (9) and (2)

$$\begin{aligned} L_1(x) &= J(x) + \frac{1}{2}r(x)^\top Qr(x) = J(x) + J(r(x)) + r(x)^\top Qx = \\ &= \frac{1}{2}x^\top Qx + q^\top x + \frac{1}{2}r(x)^\top Qr(x) + q^\top r(x) + \frac{1}{2}r(x)^\top Qx + \frac{1}{2}x^\top Qr(x) = \\ &= \frac{1}{2}(x + r(x))^\top Q(x + r(x)) + q^\top (x + r(x)) = J(x + r(x)) = J(\Pi_\Omega(x)). \end{aligned}$$

□

The change of this merit function for two consecutive iterations has the following form.

Lemma 5. *Let (x, α) , (y, β) , $r(y)$ and \mathcal{B}_2 be given as above. Then we have*

$$L_1(y) - L_1(x) = -\frac{1}{2}(x - y)^\top_{\mathcal{J}} Q_{\mathcal{J}} (x - y)_{\mathcal{J}} + \frac{1}{2}r(y)^\top_{\mathcal{J}} Q_{\mathcal{J}} r(y)_{\mathcal{J}}.$$

Proof. Using Lemma 2 and the following identities

$$\begin{aligned} r(y)_i &= 0 \text{ for } i \in \mathcal{B}_1, \mathcal{B}_2, & r(x)_i &= 0 \text{ for } i \in \mathcal{B}_1, \overline{\mathcal{B}}_2 \\ r(x)_i &= (x - y)_i \text{ for } i \in \mathcal{B}_2, & (x - y)_i &= 0 \text{ for } i \in \mathcal{B}_1 \end{aligned}$$

we get

$$\begin{aligned} L_1(y) - L_1(x) &= -\frac{1}{2}(x - y)^\top_{\overline{\mathcal{B}}_2} Q_{\overline{\mathcal{B}}_2} (x - y)_{\overline{\mathcal{B}}_2} + \frac{1}{2}(x - y)^\top_{\mathcal{B}_2} Q_{\mathcal{B}_2} (x - y)_{\mathcal{B}_2} + \\ &+ \frac{1}{2}r(y)^\top_{\overline{\mathcal{B}}_2} Q_{\overline{\mathcal{B}}_2} r(y)_{\overline{\mathcal{B}}_2} - \frac{1}{2}(x - y)^\top_{\mathcal{B}_2} Q_{\mathcal{B}_2} (x - y)_{\mathcal{B}_2} = -\frac{1}{2}(x - y)^\top_{\mathcal{J}} Q_{\mathcal{J}} (x - y)_{\mathcal{J}} + \frac{1}{2}r(y)^\top_{\mathcal{J}} Q_{\mathcal{J}} r(y)_{\mathcal{J}}. \end{aligned}$$

□

We note that the first term is negative. Moreover $r(y) = 0$ in the case that y is primal feasible. In the case where y is not primal feasible, it is possible that the merit function increases. This motivates us to modify the original KR-method to generate only primal feasible iterates.

Remark 1. *Finally let us recall a small example from [5] that causes the KR-Algorithm to cycle for most starting active sets. We are given the following problem data:*

$$Q = \begin{pmatrix} 4 & 5 & -5 \\ 5 & 9 & -5 \\ -5 & -5 & 7 \end{pmatrix}, \quad q = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The corresponding active-set-transition-graph is depicted in Figure 2.

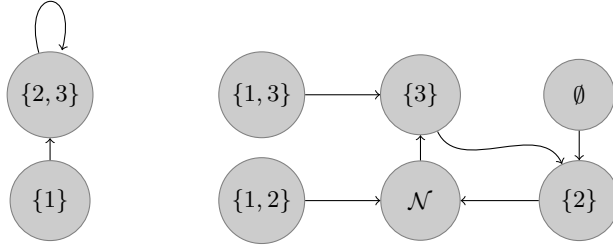


Figure 2: Active-set-transition-graph for a small numerical example with optimal $\mathcal{A} = \{2, 3\}$. The KR-Algorithm runs into cycling for 6 out of 8 active starting sets.

4 A new primal variant of the KR-method

Let us call $\mathcal{A} \subseteq \mathcal{N}$ a **primal feasible set**, if the solution x to $\text{KKT}(\mathcal{A})$ is primal feasible, $x \leq b$. Given a primal-feasible set \mathcal{A} , our goal is to find a new primal-feasible set, say \mathcal{B} . We now describe that procedure.

Suppose that $\mathcal{A} \subseteq \mathcal{N}$ is a primal feasible set. Then the KR-method would start a new iteration with the set

$$\mathcal{B}_s := \{i \in \mathcal{A} : \alpha_i \geq 0\}. \quad (10)$$

The set \mathcal{B}_s does not need to be primal feasible, because $[x, \alpha] = \text{KKT}(\mathcal{B}_s)$ may have $i \in \overline{\mathcal{B}_s}$ such that $x_i > b_i$. To turn it into a primal feasible set \mathcal{B} , we carry out the following iterations.

$$\begin{aligned} \mathcal{B} &\leftarrow \mathcal{B}_s \\ \text{while } \mathcal{B} \text{ not primal feasible} & \\ x = \text{KKT}(\mathcal{B}), \mathcal{B} &\leftarrow \mathcal{B} \cup \{i \in \overline{\mathcal{B}} : x_i \geq b_i\} \end{aligned} \quad (11)$$

This iterative scheme will clearly terminate because \mathcal{B} is augmented by adding only elements of $\overline{\mathcal{B}_s}$. The modified KR-method starts with a primal feasible set \mathcal{A} , then generates a new primal feasible set \mathcal{B} as described above. Then a new iteration is started with \mathcal{B} in place of \mathcal{A} . Let us recall the small numerical example from Section 2 to further clarify the workings of the modified KR-Algorithm. The given data yields the active-set-transition-graph given in Figure 3.

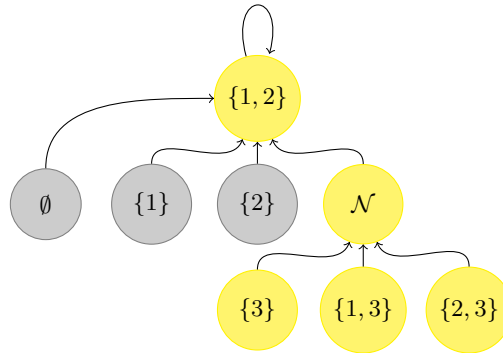


Figure 3: Active-set-transition-graph of the modified KR-Algorithm for the numerical example from Section 2 with optimal $\mathcal{A} = \{1, 2\}$. The yellow circles correspond to primal feasible sets and hence could potentially have ingoing edges.

We apply the modified KR-Algorithm on the data that causes the KR-Algorithm to cycle, and note that it avoids cycling as can be seen from the active-set-transition-graph depicted in Figure 4.

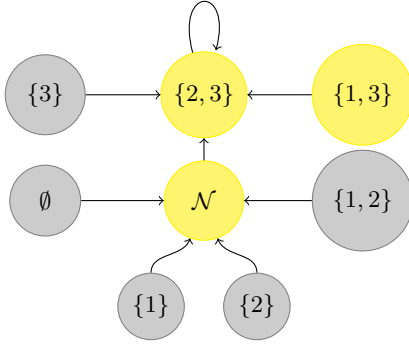


Figure 4: Active-set-transition-graph of the modified KR-Algorithm for the numerical example from Section 3 with optimal $\mathcal{A} = \{2, 3\}$ that causes the KR-Algorithm to cycle.

In the following we will prove convergence of the modified KR-Algorithm for $Q \succ 0$. Note that the idea of monotonic decrease combined with the minimization on the expanding active sets goes back at least to B.T. Polyak and his seminal paper [21] in 1969. Before going into the technical details we give an outline of the proof idea and the workings of the modified KR-Algorithm proposed above by means of the flow chart in Figure 5.

Let us start the convergence analysis with taking a closer look at the change of the objective function between two consecutive iterations given by the primal feasible sets \mathcal{A} and \mathcal{B} .

We let $[x, \alpha] = \text{KKT}(\mathcal{A})$ and $[y, \beta] = \text{KKT}(\mathcal{B})$. Using Lemma 1 and $Qy + q + \beta = 0$, we conclude

$$J(y) - J(x) = -\frac{1}{2}(x - y)^\top Q(x - y) + (x - y)^\top \beta. \quad (12)$$

The first term on the right hand side is nonpositive, so we need to investigate the second one in detail. In order to do so, we have to take a closer look at how \mathcal{B}_s and \mathcal{B} change relative to \mathcal{A} . Formally, the situation looks as indicated in Table 3.

	\mathcal{A}		\mathcal{I}	
\mathcal{B}_s	\mathcal{J}_s			
\mathcal{B}	\mathcal{B}_1	\mathcal{J}_1	\mathcal{J}_2	\mathcal{B}_2

Table 3: The change from the primal feasible set \mathcal{A} to the primal feasible set $\mathcal{B} = \mathcal{B}_s \cup \mathcal{B}_1 \cup \mathcal{B}_2$.

To explain this diagram, we note that by definition we have $\mathcal{B}_s \subseteq \mathcal{A}$ and $\mathcal{J}_s := \mathcal{N} \setminus \mathcal{B}_s$. The extension of \mathcal{B}_s to a primal feasible set \mathcal{B} is done by adding either elements from $\mathcal{A} \setminus \mathcal{B}_s$, which are contained in \mathcal{B}_1 , or from \mathcal{I} , which are contained in \mathcal{B}_2 . Thus $\mathcal{J}_1 = \mathcal{A} \setminus (\mathcal{B}_s \cup \mathcal{B}_1)$ and $\mathcal{J}_2 = \mathcal{I} \setminus \mathcal{B}_2$. What do we know about $[x, \alpha] = \text{KKT}(\mathcal{A})$ and $[y, \beta] = \text{KKT}(\mathcal{B})$? Clearly $x_{\mathcal{A}} = b_{\mathcal{A}}$, $x_{\mathcal{I}} < b_{\mathcal{I}}$ and $\alpha_{\mathcal{I}} = 0$. The definition of \mathcal{B}_s yields $\alpha_{\mathcal{B}_s} \geq 0$ and $\alpha_{\mathcal{A} \setminus \mathcal{B}_s} < 0$. The set \mathcal{B} is the union of \mathcal{B}_s , \mathcal{B}_1 and \mathcal{B}_2 , and $\mathcal{J} := \mathcal{N} \setminus \mathcal{B}$. Hence we have $y_{\mathcal{B}} = b_{\mathcal{B}}$, $y_{\mathcal{J}} < b_{\mathcal{J}}$ and $\beta_{\mathcal{J}} = 0$. This is summarized in Table 4.

Now we provide some additional useful properties for two consecutive iterations given by the sets \mathcal{A} and \mathcal{B} .

Lemma 6. *Let \mathcal{A} be a primal feasible set. Then $\mathcal{B}_s \subseteq \mathcal{A}$ and $\mathcal{B}_s = \mathcal{A}$ if and only if \mathcal{A} is optimal. If \mathcal{A} is not optimal, then $\mathcal{A} \neq \emptyset$, $\mathcal{A} \neq \mathcal{B}$ and $x - y \neq 0$.*

Proof. $\mathcal{B}_s \subseteq \mathcal{A}$ holds due to (10). If \mathcal{A} is optimal then all elements in \mathcal{A} are dual feasible and hence $\mathcal{B}_s = \mathcal{A}$. If $\mathcal{B}_s = \mathcal{A}$ then all elements in \mathcal{A} are dual feasible. As the algorithm ensures all other KKT conditions

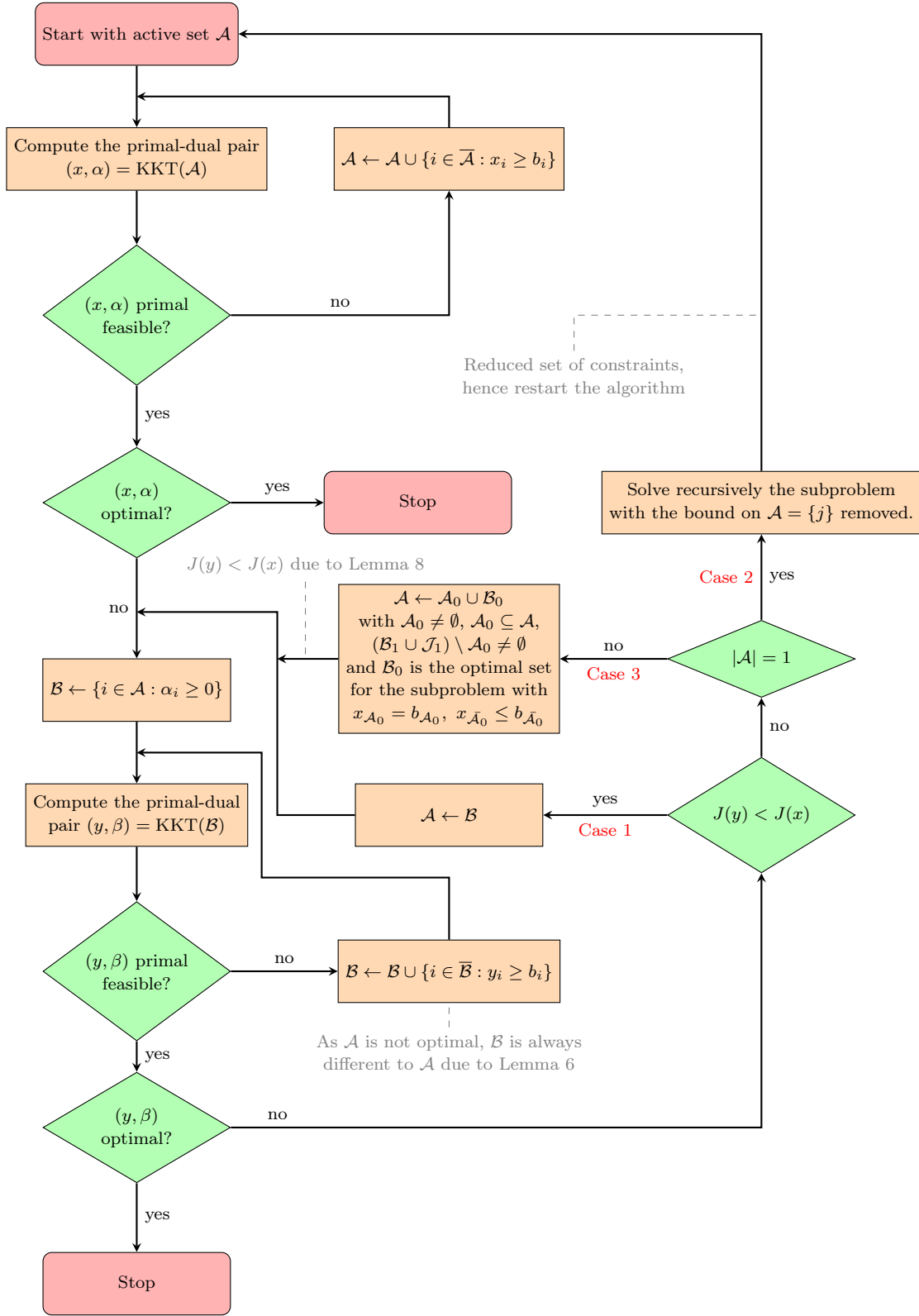


Figure 5: Outline of the workings of the modified KR-Algorithm.

	\mathcal{A}			\mathcal{I}	
	\mathcal{B}_s	\mathcal{B}_1	\mathcal{J}_1	\mathcal{J}_2	\mathcal{B}_2
x	$= b$	$= b$	$= b$	$< b$	$< b$
y	$= b$	$= b$	$< b$	$< b$	$= b$
α	≥ 0	< 0	< 0	$= 0$	$= 0$
β	$?$	$?$	$= 0$	$= 0$	$?$

Table 4: Information on x , y , α and β associated to \mathcal{A} and \mathcal{B} . The questionmarks indicate that the sign of β can not be specified for the respective sets.

including primal feasibility throughout all iterations, \mathcal{A} is optimal. If $\mathcal{A} = \emptyset$ then the unconstrained optimum is feasible and hence \mathcal{A} is optimal.

Now let us suppose that \mathcal{A} is not optimal and $\mathcal{A} = \mathcal{B}$. Then $\mathcal{J}_1 = \mathcal{B}_2 = \emptyset$ and hence $\mathcal{J}_2 = \mathcal{J} = \mathcal{I}$ and $\mathcal{B}_1 = \mathcal{A} \setminus \mathcal{B}_s \neq \emptyset$ as $\mathcal{B}_s \neq \mathcal{A}$. Due to the workings of the modified KR-method the following implications hold

$$\begin{aligned} [x, \alpha] = KKT(\mathcal{A}) &\Rightarrow \alpha_{\mathcal{B}_1} < 0, \\ [y, \beta] = KKT(\mathcal{B}_s) &\Rightarrow \beta_{\mathcal{B}_1} = 0. \end{aligned} \tag{13}$$

As $\mathcal{I} = \mathcal{J}$ we can substitute the primal variables associated with \mathcal{I} by

$$\begin{aligned} x_{\mathcal{I}} &= -Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I},\mathcal{B}_s}b_{\mathcal{B}_s} + Q_{\mathcal{I},\mathcal{B}_1}b_{\mathcal{B}_1}), \\ y_{\mathcal{I}} &= -Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I},\mathcal{B}_s}b_{\mathcal{B}_s} + Q_{\mathcal{I},\mathcal{B}_1}y_{\mathcal{B}_1}). \end{aligned} \tag{14}$$

Throughout the algorithm the equations $Qx + q + \alpha = 0$ hold. Using (13) we obtain

$$\begin{aligned} Q_{\mathcal{B}_1,\mathcal{N}} \begin{pmatrix} b_{\mathcal{B}_s} \\ b_{\mathcal{B}_1} \\ x_{\mathcal{I}} \end{pmatrix} + q_{\mathcal{B}_1} &= -\alpha_{\mathcal{B}_1} > 0, \\ Q_{\mathcal{B}_1,\mathcal{N}} \begin{pmatrix} b_{\mathcal{B}_s} \\ y_{\mathcal{B}_1} \\ y_{\mathcal{I}} \end{pmatrix} + q_{\mathcal{B}_1} &= -\beta_{\mathcal{B}_1} = 0. \end{aligned}$$

Next applying (14) to the above equations yields

$$\begin{aligned} Q_{\mathcal{B}_1,\mathcal{B}_s}b_{\mathcal{B}_s} + Q_{\mathcal{B}_1}b_{\mathcal{B}_1} - Q_{\mathcal{B}_1,\mathcal{I}}Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I},\mathcal{B}_s}b_{\mathcal{B}_s} + Q_{\mathcal{I},\mathcal{B}_1}b_{\mathcal{B}_1}) + q_{\mathcal{B}_1} &> \\ Q_{\mathcal{B}_1,\mathcal{B}_s}b_{\mathcal{B}_s} + Q_{\mathcal{B}_1}y_{\mathcal{B}_1} - Q_{\mathcal{B}_1,\mathcal{I}}Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I},\mathcal{B}_s}b_{\mathcal{B}_s} + Q_{\mathcal{I},\mathcal{B}_1}y_{\mathcal{B}_1}) + q_{\mathcal{B}_1}. & \end{aligned}$$

Simplifying the above inequality gives

$$(Q_{\mathcal{B}_1} - Q_{\mathcal{B}_1,\mathcal{I}}Q_{\mathcal{I}}^{-1}Q_{\mathcal{I},\mathcal{B}_1})(b_{\mathcal{B}_1} - y_{\mathcal{B}_1}) > 0. \tag{15}$$

The matrix $(Q_{\mathcal{B}_1} - Q_{\mathcal{B}_1,\mathcal{I}}Q_{\mathcal{I}}^{-1}Q_{\mathcal{I},\mathcal{B}_1})$ is positive definite due to the Schur-complement lemma. To assure $\mathcal{A} = \mathcal{B}$ the inequality

$$b_{\mathcal{B}_1} - y_{\mathcal{B}_1} \leq 0, \tag{16}$$

has to hold. Combining (15) and (16) yields a contradiction to $(Q_{\mathcal{B}_1} - Q_{\mathcal{B}_1,\mathcal{I}}Q_{\mathcal{I}}^{-1}Q_{\mathcal{I},\mathcal{B}_1}) \succ 0$:

$$(b_{\mathcal{B}_1} - y_{\mathcal{B}_1})^\top (Q_{\mathcal{B}_1} - Q_{\mathcal{B}_1,\mathcal{I}}Q_{\mathcal{I}}^{-1}Q_{\mathcal{I},\mathcal{B}_1})(b_{\mathcal{B}_1} - y_{\mathcal{B}_1}) \leq 0.$$

Finally using Table 4 we have $x_{\mathcal{I}} < b_{\mathcal{I}}$, $y_{\mathcal{J}} < b_{\mathcal{J}}$, $x_{\mathcal{A}} = b_{\mathcal{A}}$, $y_{\mathcal{B}} = b_{\mathcal{B}}$ and hence $x - y \neq 0$ for $\mathcal{A} \neq \mathcal{B}$. \square

As $x - y = 0$ on $\mathcal{B}_s \cup \mathcal{B}_1$ and $\beta = 0$ on $\mathcal{J}_1 \cup \mathcal{J}_2$ it follows that $(x - y)^\top \beta = \sum_{i \in \mathcal{B}_2} (x - y)_i \beta_i$. Since $(x - y)_i < 0$ on \mathcal{B}_2 , the contribution of $(x - y)^\top \beta$ is guaranteed to be nonpositive in case $\beta \geq 0$ on \mathcal{B}_2 . We summarize the discussion in the following lemma.

Lemma 7. *Let x, y, α, β and \mathcal{B}_2 be as above. Then*

$$J(y) - J(x) = -\frac{1}{2}(x - y)^\top Q(x - y) + \sum_{i \in \mathcal{B}_2} (x - y)_i \beta_i.$$

Moreover, $(x - y)_{\mathcal{B}_2} < 0$.

Remark 2. *We observe that $\beta_i < 0$ for $i \in \mathcal{B}_2$ occurs in a situation where $i \in \mathcal{I} \cap \mathcal{J}_s$. Thus x_i was inactive at the beginning of the iterations, corresponding to set \mathcal{A} . Then in the course of removing primal infeasibility, x_i went above b_i , resulting in $i \in \mathcal{B}$, more precisely, $i \in \mathcal{B}_2$. But then the solution $[y, \beta] = \text{KKT}(\mathcal{B})$ yields $\beta_i < 0$, hence the objective function would improve if x_i would be allowed to move away from the boundary again. This situation is peculiar, but there is no mathematical reason why it should not occur.*

We have just seen that we have no means to avoid $J(y) \geq J(x)$ using the current setup. To avoid cycling, we therefore need to take additional measures in case $J(y) \geq J(x)$. Let us assume \mathcal{A} is not optimal for the following case distinction. We first recall that in this case $|\mathcal{A}| \geq 1$ due to Lemma 6. We consider the following cases separately. Let x, y, α, β and \mathcal{A} and \mathcal{B} be as above.

Case 1: $J(y) < J(x)$. In this case we can set $\mathcal{A} \leftarrow \mathcal{B}$, and continue with the next iteration.

Case 2: $J(y) \geq J(x)$ and $|\mathcal{A}| = 1$. The set $\mathcal{A} = \{j\}$ is primal feasible. Since x is primal feasible, but not optimal, we must have $\alpha_j < 0$. Thus the objective function would improve by allowing x_j to move away from the boundary. Hence in an optimal solution we have $x_j < b_j$ and thus a problem with only $n - 1$ constraints which we solve by induction on the number of constraints.

Case 3: $J(y) \geq J(x)$ and $|\mathcal{A}| > 1$. In this case we will again solve a problem with less than n constraints to optimality. First suppose that $\mathcal{B}_s \neq \emptyset$. In this case $\mathcal{B}_s \neq \mathcal{A}$, due to Lemma 6. Set $\mathcal{A}_0 := \mathcal{B}_s$.

In the other case, where $\mathcal{B}_s = \emptyset$, we know that $|\mathcal{B}_1 \cup \mathcal{J}_1| > 1$, because $|\mathcal{A}| > 1$. In this case set $\mathcal{A}_0 = \{j\}$ for some $j \in \mathcal{B}_1 \cup \mathcal{J}_1$.

In both cases we have the following situation: $\mathcal{A}_0 \neq \emptyset$, $\mathcal{A}_0 \subseteq \mathcal{A}$ and $(\mathcal{B}_1 \cup \mathcal{J}_1) \setminus \mathcal{A}_0 \neq \emptyset$. We now solve (recursively) the smaller subproblem with $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}$, $x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}$ which has $|\bar{\mathcal{A}}_0| < n$ bound constrained variables. Its optimal active set is denoted by \mathcal{B}_0 . We set $\mathcal{B} := \mathcal{A}_0 \cup \mathcal{B}_0$ and get $[y, \beta] = \text{KKT}(\mathcal{B})$. We note that the solution $[x, \alpha] = \text{KKT}(\mathcal{A})$ is feasible for this subproblem but not optimal, because $\alpha_i < 0$ for $i \in (\mathcal{B}_1 \cup \mathcal{J}_1) \setminus \mathcal{A}_0 \neq \emptyset$. Therefore $J(y) < J(x)$ because y is the optimal solution of the subproblem. We summarize the discussion in the following lemma.

Lemma 8. *Let \mathcal{A} be a primal feasible index set with nonoptimal solution $[x, \alpha] = \text{KKT}(\mathcal{A})$ and index partition as in Table 4 and assume $|\mathcal{A}| > 1$. Then there exists $\mathcal{A}_0 \neq \emptyset$, $\mathcal{A}_0 \subseteq \mathcal{A}$ and $(\mathcal{B}_1 \cup \mathcal{J}_1) \setminus \mathcal{A}_0 \neq \emptyset$. Let \mathcal{B}_0 be the optimal active set for the subproblem with $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}$, $x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}$ and consider $\mathcal{B} = \mathcal{A}_0 \cup \mathcal{B}_0$. Then $J(y) < J(x)$ holds for $[y, \beta] = \text{KKT}(\mathcal{B})$.*

We summarize the modified KR-method, which produces primal feasible iterates in each iteration, in Table 5. Since the objective value strictly decreases in each iteration, we have shown the following main result of this paper.

Theorem 9. *The modified KR-Algorithm as given in Table 5 terminates in a finite number of iterations with the optimal solution.*

Modified KR-Algorithm

Input: $Q \succ 0, b, d \in \mathbb{R}^n. \mathcal{A} \subseteq \mathcal{N}.$

Output: optimal active set $\mathcal{A}.$

$x = \text{KKT}(\mathcal{A}).$

while \mathcal{A} not primal feasible: $\mathcal{A} \leftarrow \mathcal{A} \cup \{i \in \bar{\mathcal{A}} : x_i \geq b_i\}; x = \text{KKT}(\mathcal{A});$ **endwhile**

$\alpha = \text{KKT}(\mathcal{A}).$

while \mathcal{A} not optimal

$\mathcal{B} \leftarrow \{i \in \mathcal{A} : \alpha_i \geq 0\}, y = \text{KKT}(\mathcal{B}).$

while \mathcal{B} not primal feasible: $\mathcal{B} \leftarrow \mathcal{B} \cup \{i \in \bar{\mathcal{B}} : y_i \geq b_i\}; y = \text{KKT}(\mathcal{B});$ **endwhile**

Case 1: $J(y) < J(x)$

$\mathcal{A} \leftarrow \mathcal{B}.$

Case 2: $J(y) \geq J(x)$ and $|\mathcal{A}| = 1$

Solve recursively the subproblem with the bound on $\mathcal{A} = \{j\}$ removed.

Case 3: $J(y) \geq J(x)$ and $|\mathcal{A}| > 1$

Select \mathcal{A}_0 such that $\mathcal{A}_0 \neq \emptyset, \mathcal{A}_0 \subseteq \mathcal{A}$ and $(\mathcal{B}_1 \cup \mathcal{J}_1) \setminus \mathcal{A}_0 \neq \emptyset.$

Let \mathcal{B}_0 be the optimal active set for the subproblem with $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}, x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}.$

$\mathcal{A} \leftarrow \mathcal{A}_0 \cup \mathcal{B}_0.$

$\alpha = \text{KKT}(\mathcal{A}).$

endwhile

Table 5: Description of the modified KR-Algorithm.

5 Practical Comments and Extensions

Remark 3. *There exist many ways to choose \mathcal{A}_0 in case 3. In our implementation we use the following rule to determine \mathcal{A}_0 .*

$$\mathcal{A}_0 = \begin{cases} \mathcal{B}_s & \text{if } \mathcal{B}_s \neq \emptyset \\ j \in \mathcal{B}_1 & \text{if } \mathcal{B}_s = \emptyset, \mathcal{B}_1 \neq \emptyset \\ j \in \mathcal{J}_1 & \text{if } \mathcal{B}_s = \emptyset, \mathcal{B}_1 = \emptyset \end{cases}$$

The intuition for this choice is as follows. In case that \mathcal{B}_s is nonempty, we set $\mathcal{A}_0 = \mathcal{B}_s$ because the elements in \mathcal{B}_s were active in the previous iteration and their dual variables had the correct sign. It seems therefore plausible to fix them again to their bounds to define the subproblem. Otherwise we know from Lemma 6 that $\mathcal{B}_1 \cup \mathcal{J}_1$ is nonempty. If possible, we select an element from \mathcal{B}_1 arbitrarily and fix it to its bound, otherwise an element from \mathcal{J}_1 is fixed.

Denoting the sign constrained variables in the subproblem by $\mathcal{S} := \mathcal{N} \setminus \mathcal{A}_0$, the subproblem is again a convex quadratic problem with bound constraints and given as follows

$$\begin{aligned} \min_{x_{\mathcal{S}}} \quad & \frac{1}{2} x_{\mathcal{S}}^{\top} Q_{\mathcal{S}, \mathcal{S}} x_{\mathcal{S}} + (q_{\mathcal{S}} + Q_{\mathcal{N}, \mathcal{A}_0} b_{\mathcal{A}_0})^{\top} x_{\mathcal{S}}, \\ \text{subject to} \quad & x_{\mathcal{S}} \leq b_{\mathcal{S}}. \end{aligned}$$

As a starting active set of the subproblem we choose all elements from $\mathcal{B}_1 \cup \mathcal{B}_2$ for which the associated dual variables are ≥ 0 .

Remark 4. *In our implementation, case 2 is carried out as follows. If \mathcal{A} is not optimal and $|\mathcal{A}| = 1$, we remove the bound of the associated primal variable and then restart the algorithm with the current active set. In this way we make use of the current information to restart.*

Remark 5. Regarding the subproblems generated in cases 2 and 3, one may be tempted to avoid finding the optimal solution and return from the subproblem, once $J(y) < J(x)$. We experimented with both strategies, but our computational experiments showed that prematurely returning from the subproblems resulted in an overall increase of computation time, especially on “difficult” instances.

Remark 6. We found examples where both “Case 2” and “Case 3” occur. Hence all cases discussed in the convergence analysis are also of practical relevance. Let us also emphasize that we cannot guarantee a polynomial running time for the modified KR-Algorithm as one might already expect for combinatorial, active set methods (see e.g. the simplex algorithm).

Remark 7. We also investigated a variant of our algorithm, where dual instead of primal feasibility is assured in every iteration but we do not describe this variant in detail because it has the following drawbacks compared to the primal version. First one has to solve $KKT(\mathcal{A})$ in each iteration, and not only compute $x_{\mathcal{I}}$. Secondly the optimal inactive set is in general approached by supersets, resulting in higher computation times per iteration. Finally we observed experimentally that assuring primal instead of dual feasibility results in fewer iterations on average.

Remark 8. The ideas used for proving Theorem 3 cannot be generalized to both upper and lower bounds, i.e. to constraints of the form $a \leq x \leq b$ because of the possible existence of indices that are active at the upper/lower bound at iteration k and are active at the lower/upper bound at iteration $k + 2$. This would spoil the analysis used in Theorem 3 (for a detailed discussion of this issue see [13]).

The convergence argument for the modified KR-method presented in the previous section can be generalized to convex QPs with box constraints $a \leq x \leq b$, for details see the Appendix.

6 Computational Experience

In this section we compare the modified KR-Algorithm to the KR-Algorithm and standard conjugate gradient, active set and interior point methods on several benchmark sets. In all the tables below, the column labels have the following meaning.

- [KR] denotes the KR-Algorithm and [modified KR] is our modified KR-Algorithm. [quadprog] denotes the MATLAB function `quadprog`, which incorporates a reflective trust region method. Finally, [CPLEX-IP] and [CPLEX-AS] stand for the interior point and active set solvers of ILOG CPLEX 12.1, see [14].
- The condition number of Q is estimated with the MATLAB command `cond`. Its average is given in the column labeled $cond(Q)$.
- The average number of outer iterations is denoted by $iter$. In case of the KR-method it agrees with the number of times the system (4) is solved to get $x_{\mathcal{I}}$.
- In each iteration a linear system of size $n_{\mathcal{I}} := n - |\mathcal{A}|$ has to be solved to get $x_{\mathcal{I}}$. The column labeled $n_{\mathcal{I}}$ contains the average system size.
- In the modified KR-method it may take several steps in (11) until a primal feasible x is reached. In the columns labeled $solve$ and $max\ solve$ we show how often the system (4) was solved on average and in the worst case. We recall that this reflects the main computational effort of our method.
- The column labeled $fail$ contains the number of trials, where the KR-method ran into a cycle and failed to solve the problem.
- Finally m_{ds} gives the maximal depth of levels of subproblems (recursive calls) in the modified KR-method.

- In some tables we also provide the average density of Q , which is denoted by *dens*.
- All experiments were conducted on an Intel Xeon 5160 processor with 3 GHz and 2 GB RAM. The corresponding Matlab code for generating all instances and solving them with the various methods discussed is available from <http://philipphungerlaender.jimdo.com/qp-code/>.

6.1 Randomly Generated Problems

First we investigate the computational overhead of globalizing the original KR-Algorithm. We compare it to the original KR-Algorithm for randomly generated problems (see [17, Section 4.1.]) of fixed dimension $n = 2000$. We use the same generator for the data and vary the matrix Q by making it increasingly ill-conditioned. The random instances are generated starting with a matrix Q_0 which is a sparse positive semidefinite and singular matrix. The matrix Q is obtained from Q_0 by adding a small multiple of the identity matrix I ,

$$Q = Q_0 + \epsilon I,$$

where $\epsilon \in \{1, 10^{-5}, 10^{-10}, 10^{-14}\}$. The Table 6 summarizes the performance of the two algorithms. Each line is averaged over 2000 trials for each value of ϵ . We used the initial active set $\mathcal{A}^0 := \mathcal{N}$ (initial active set contains all indices) to avoid solving a full system of order n .

cond(Q)	log(ϵ)	[KR]			[modified KR]				
		iter	n_I	fail	iter	solve	max solve	n_I	mds
$\approx 10^3$	0	5.84	828	0	5.07	9.48	13	854	0
$\approx 10^8$	-5	9.03	888	0	6.67	12.10	40	868	4
$\approx 10^{13}$	-10	9.05	889	0	6.68	15.06	32	686	1
$\approx 10^{17}$	-14	9.05	889	0	6.68	15.08	32	686	1

Table 6: Comparison of KR- and modified KR-Algorithm, initial active set is \mathcal{N} and $n = 2000$. Each line is the average over 2000 trials. The column max solve provides the maximum number of solves during a single trial.

The modified KR-Algorithm reduces the average number of outer iterations compared to the KR-Algorithm at the cost of solving the primal system for the inactive variables more often in order to reach primal feasibility. The average number of solves for the original KR-Algorithm is around 9, while it is around 15 for the new method. The maximum number of solves during a single run is not much higher than twice the average. We consider this an acceptable price to pay for guaranteed convergence. The maximal depth of levels of subproblems of the modified KR-Algorithm stays below 5 even for the most ill-conditioned data. It is surprising that the original KR-method does not cycle in these experiments even though in principle this is possible.

We also experimented with different starting active sets but observed that the initial starting set does not seem to have a serious influence on the overall performance of the algorithms. Therefore we suggest to use the full initial active set to avoid solving a linear system of dimension n . The effort for the backsolve in (5) to get the dual variables is negligible.

In Table 7 we compare the modified KR-Algorithm to the interior point and active set solvers of ILOG CPLEX. We use the same random data as above with $\epsilon \in \{1, 10^{-10}\}$, vary n between 2000 and 50.000 and take the average over 10 instances. The matrix Q_0 is generated as a matrix with bandwidth = 100. In this way we are able to go to larger sizes for n .

The modified KR-Algorithm is superior to both CPLEX methods in terms of computation times, where the relative gap between the methods is growing with problem size. It seems especially remarkable that the modified KR-Algorithm needs only about one fifth of the number of iterations of the interior point

n	$\text{cond}(Q)$	[CPLEX-IP]		[CPLEX-AS]		[modified KR]		
		iter	time	iter	time	iter	time	mds
2000	$\approx 10^3$	16.2	1.34	1000	0.49	5.0	0.29	0
2000	$\approx 10^{13}$	21.4	1.69	1063	0.53	6.5	0.42	0
10000	$\approx 10^3$	20.8	8.95	5024	8.03	5.2	1.77	0
10000	$\approx 10^{13}$	27.1	11.25	5273	9.04	6.6	2.44	0
50000	$\approx 10^3$	26.2	66.78	25243	99.86	5.8	8.32	0
50000	$\approx 10^{13}$	32.1	80.13	26263	108.13	6.9	10.68	0

Table 7: Comparison with standard active set and interior point methods, each line is the average over 10 trials.

solver and that the number of iterations is only slightly influenced by the size and the condition number of the instances. In summary this means that the modified KR-Algorithm solves on average fewer primal systems than the interior point solver to find the optimal solution.

The outcome of the tests is exactly what one would expect: The interior point method always solves the full system of order n and needs more iterations, so one would expect the active set methods to be a lot faster, as they solve on average fewer (and smaller) systems. Further comparisons with interior point methods therefore do not provide additional insight as long as our number of solves stays well below 30.

Next we consider randomly generated problems in the style of [5] in Tables 8 and 9. Specifically, we generated Q via Matlab's `sprandsym` routine. Both the average density and the average condition number are given in the following two tables. Then we generated the optimal solution x via the Matlab `randn` routine. With the solution in hand, we choose the objective linear term coefficients, lower and upper bounds so that at the optimal solution the number of

- the active variables and the inactive variables are roughly equal for the case with one-sided bounds and
- the active variables on the upper bound, the active variables on the lower bound and the inactive variables are all roughly equal for the case with box constraints.

In Tables 8 and 9 we compare again the KR-Method, the modified KR-Method and Matlabs function `quadprog`. The column labeled *error* gives the relative error of this method with respect to the optimal objective value,

$$\text{error} = \frac{\text{obtained objective value} - \text{optimal objective value}}{\text{optimal objective value}}.$$

Each line represents the average over 10 trials. These are difficult examples for the active set methods. The original KR-method fails on larger instances, once the condition number gets larger. In these cases the corresponding *iter*-column gives the average number of outer iterations for those problems that were solved successfully by the KR-Algorithm. The modified KR-method needs to go into recursions, but the depth of the recursion is small, no larger than 4. We consider the same sizes n as in [5], but allow the condition number to go up to $\approx 10^{10}$ as compared to $\approx 10^6$ in [5]. Since the code from [5] is not available to us, we can not make a direct comparison.

These experiments suggest the following first conclusion. The original KR method is extremely efficient, but it may fail completely on more difficult, badly conditioned problems, as can be seen in Tables 8 and 9. The modified KR method typically requires a slightly larger number of solves and also goes into recursion (of very small depth). The total number of solves is not much larger than in the original KR method for well-conditioned problems. It stays typically below 50 for problems with $\text{cond}(Q) \approx 10^6$ and it may rise up to 250 for problems with large condition number $\text{cond}(Q) \approx 10^{10}$. Compared to Matlab's large scale solver however, it is both much faster and also more accurate on difficult problems.

n	dens	cond(Q)	[KR]			[modified KR]				[quadprog]		
			time	iter	fail	time	iter	solve	mds	time	iter	error
5000	0.1	$\approx 10^2$	9.23	4.6	0	15.96	4.4	6.9	0	6.14	15.2	$2.9 \cdot 10^{-15}$
5000	0.1	$\approx 10^6$	15.35	13.7	0	33.77	9.6	45.6	1	40.00	55.0	$1.1 \cdot 10^{-5}$
5000	0.1	$\approx 10^{10}$	33.06	33.4	0	66.17	12.3	124.4	2	163.75	113.4	$9.1 \cdot 10^{-5}$
5000	0.01	$\approx 10^2$	3.79	4.4	0	6.51	4.1	6.3	0	0.73	13.7	$3.4 \cdot 10^{-11}$
5000	0.01	$\approx 10^6$	4.71	16.8	4	10.64	10.9	47.0	2	6.78	58.7	$4.3 \cdot 10^{-6}$
5000	0.01	$\approx 10^{10}$	-	-	10	15.05	9.7	96.0	3	40.8	112.6	$2.4 \cdot 10^{-5}$
5000	0.001	$\approx 10^2$	0.09	5.1	0	0.17	4.5	6.7	1	0.22	11.3	$1.3 \cdot 10^{-8}$
5000	0.001	$\approx 10^6$	0.10	8.6	5	0.22	6.3	13.9	1	1.87	36.3	$8.8 \cdot 10^{-7}$
5000	0.001	$\approx 10^{10}$	-	-	10	0.37	8.3	21.3	1	11.4	48.9	$5.8 \cdot 10^{-6}$
10000	0.001	$\approx 10^2$	3.09	4.9	0	4.66	4.1	6.4	0	0.60	12.6	$2.5 \cdot 10^{-9}$
10000	0.001	$\approx 10^6$	1.77	11.0	9	9.31	8.3	25.3	1	6.40	51.6	$1.2 \cdot 10^{-6}$
10000	0.001	$\approx 10^{10}$	-	-	10	16.81	10.0	52.8	2	47.32	82.2	$1.7 \cdot 10^{-5}$

Table 8: Random problems in the style of [5] with one-sided bounds. The first iter-column gives the average number of outer iterations for those problems that were solved successfully by the KR-Algorithm.

n	dens	cond(Q)	[KR]			[modified KR]				[quadprog]		
			time	iter	fail	time	iter	solve	mds	time	iter	error
5000	0.1	$\approx 10^2$	13.04	6.1	0	16.68	4.2	12.0	0	16.32	19.0	$6.1 \cdot 10^{-11}$
5000	0.1	$\approx 10^6$	43.97	26.1	0	43.21	14.0	66.5	1	115.03	40.6	$7.9 \cdot 10^{-7}$
5000	0.1	$\approx 10^{10}$	-	-	10	104.20	14.3	258.1	2	401.83	98.1	$8.0 \cdot 10^{-6}$
5000	0.01	$\approx 10^2$	4.57	6.2	0	5.83	3.6	10.6	0	2.19	18.9	$6.9 \cdot 10^{-10}$
5000	0.01	$\approx 10^6$	8.64	27.6	5	8.83	12.0	57.1	2	11.34	42.1	$9.4 \cdot 10^{-7}$
5000	0.01	$\approx 10^{10}$	-	-	10	21.11	10.5	157.9	3	31.0	58.1	$9.8 \cdot 10^{-6}$
5000	0.001	$\approx 10^2$	0.15	5.8	0	0.15	3.6	9.4	0	0.96	19.0	$9.1 \cdot 10^{-8}$
5000	0.001	$\approx 10^6$	0.25	12.8	4	0.23	6.7	19.2	1	4.37	44.2	$3.5 \cdot 10^{-6}$
5000	0.001	$\approx 10^{10}$	-	-	10	0.30	7.4	25.6	2	9.90	45.2	$6.4 \cdot 10^{-6}$
10000	0.001	$\approx 10^2$	3.28	6.0	0	3.11	3.6	9.6	0	1.71	17.2	$7.3 \cdot 10^{-9}$
10000	0.001	$\approx 10^6$	-	-	10	5.07	8.9	31.5	2	9.32	47.4	$3.3 \cdot 10^{-6}$
10000	0.001	$\approx 10^{10}$	-	-	10	10.18	10.7	62.7	4	16.49	49.2	$2.0 \cdot 10^{-5}$

Table 9: Random problems in the style of [5] with box constraints. The first iter-column gives the average number of outer iterations for those problems that were solved successfully by the KR-Algorithm.

6.2 Harmonic Equation

We now consider problems where the data are not random, but come from applications in mathematical physics. There are several problem types which lead to optimization problems of the form (1), where the Hessian Q_m represents a discretization of the Laplace operator, acting on a square grid of size $m \times m$. The linear term and the boundary conditions describe the precise nature of the problem. Moré and Toraldo [19] discuss several such applications. We now consider the obstacle problem from section 7.1 in [19] and use the same setup to make the results comparable. Instances of this type are also contained in the CUTer [10] and CUTEst [11] test library and are denoted by ‘‘OBSTCL*’’. In this collection however we have $m \leq 125$. The elastic-plastic torsion problem as well as the journal bearing problem from [19] both have the same Hessian Q_m . We do not include computational results for these classes, as they are very similar to the obstacle problem. We summarize the results of our runs on obstacle problems with one- and two-sided bounds in Tables 10 and 11 respectively.

As a second application of the harmonic operator we consider the circus tent problem taken from Matlab’s optimization demo as an example of large-scale quadratic programming with simple bounds.

m	[KR]			[modified KR]				
	iter	time	largest system	iter	solve	time	largest system	mds
16	3	0.01	76	3	3	0.01	76	0
32	4	0.02	340	4	5	0.02	684	0
64	4	0.05	1519	4	5	0.06	1519	0
128	4	0.19	6385	4	5	0.24	6385	0
256	4	1.04	26236	4	5	1.25	26236	0
512	4	7.13	106184	4	5	8.62	106184	0

Table 10: Obstacle problem from [19] with one-sided bounds. The matrix Q_m represents a discretization of the harmonic operator acting on a square grid of size $m \times m$.

m	[KR]			[modified KR]				
	iter	time	largest system	iter	solve	time	largest system	mds
16	3	0.02	143	3	4	0.02	143	0
32	4	0.03	675	3	6	0.03	675	0
64	5	0.08	3000	5	8	0.10	3000	0
128	4	0.28	12573	4	6	0.39	12573	0
256	4	2.02	51484	5	7	2.78	51484	0
512	4	16.68	208357	5	7	23.63	208357	0

Table 11: Obstacle problem from [19] with two-sided bounds. The matrix Q_m represents a discretization of the harmonic operator acting on a square grid of size $m \times m$.

The problem is to build a circus tent to cover a square lot. The tent is elastic and is to be supported by five poles. The question is to find the shape of the tent at equilibrium, that corresponds to the minimum of the energy computed from the surface position and squared norm of its gradient. The problem has only lower bounds imposed by the five poles and the ground. We also solve this problem with `quadprog` from MATLAB. In Table 12 we compare it with the modified KR-Algorithm for different grid sizes m . We note that in all cases our approach is extremely efficient. Since the matrix Q_m is very sparse for this problem, we are able to solve problems where the system size is bigger than 200,000.

m	n	optimal objective value	[quadprog]		[modified KR]				
			iter	time	iter	solve	time	largest system	mds
50	2500	0.4561	15	0.34	2	1	0.02	2480	0
100	10000	0.3420	17	1.39	2	1	0.12	9980	0
200	40000	0.2938	19	12.58	2	1	0.97	39980	0
300	90000	0.2647	21	45.24	2	1	3.29	89980	0
400	160000	0.2519	22	110.12	2	1	7.95	159980	0
500	250000	0.2425	24	219.92	2	1	16.89	249980	0

Table 12: Comparison of the modified KR-Algorithm with Matlab’s large-scale reflective trust-region algorithm on the circus tent problem.

6.3 Biharmonic Equation

As a second class of problems with a physical interpretation we consider a thin elastic square plate, clamped on its boundary with a vertical force acting on it. The plate deforms but is constrained to remain below an obstacle. Discretizing again the square into an $m \times m$ rectangular grid, we obtain a problem of the form (1), with matrix Q_m of order m^2 representing the discretized biharmonic operator. For specific details

about the force acting on the plate and the obstacle, we refer to [17]. We conducted experiments for m going up to 512 and summarize the results in Table 13

m	[KR]			[modified KR]				
	iter	time	largest system	iter	solve	time	largest system	mds
16	7	0.03	139	7	12	0.03	139	0
32	10	0.08	647	9	15	0.10	647	0
64	6	0.19	2738	6	7	0.23	2738	0
128	7	0.95	11251	7	8	1.15	11251	0
256	8	8.12	45580	8	12	12.25	45580	0
512	12	117.27	183486	12	20	196.99	183486	0

Table 13: Biharmonic Equation: Comparison with [17]. Here the matrix Q_m is the discretisation of the biharmonic operator on a square $m \times m$ grid of order m^2 .

Since the original KR-method has no difficulty with this problem, the safeguarding procedure in the modified method causes an increase in solves. We note that even though the dimension m^2 as well as the largest system being solved is quite large, we are still able to solve this problem rather quickly, because of the sparse structure of Q_m .

6.4 Sparse Quadratic 0-1 Optimization Problems by Pardalos and Rodgers

The box-constrained version of the problem, as analyzed in the appendix, has applications in combinatorial optimization. Quadratic 0-1 optimization is one of the fundamental NP-hard problems in discrete optimization. It consists of minimizing $J(x)$ on $\{0, 1\}^n$, but Q need not be semidefinite. The integrality conditions on x imply that $x_i^2 = x_i$ can be used to convexify the objective function. A possible convex relaxation is to minimize

$$\frac{1}{2}x^T(Q - \lambda_{\min}(Q)I)x + (q + \lambda_{\min}(Q)e)^T x \quad (17)$$

on the box $0 \leq x \leq 1$. Both objective functions agree on 0-1 vectors x . This opens the way to solve the original integer problem by solving the relaxation (17) using Branch-and-Bound techniques. This amounts to solving a problem of type (17) in every node of the branching tree. The nodes of the branching tree indicate which of the variables are fixed to their bounds. Since the number of nodes in such branching trees may become quite large, it is critical to solve each subproblem as quickly as possible. For more details on solving quadratic 0-1 problems see e.g. [3]. In Table 14 we provide computation times to solve (17) by the interior point solver of CPLEX and our new method averaged over 10 runs. Our method is faster by an order of magnitude, especially for larger instances. Q was generated by the Sparse Problem Generator of Pardalos and Rodgers [20]. Its diagonal coefficients are in the range $[-100, 100]$ and the off-diagonal entries are in the range $[-50, 50]$.

n	cond(Q)	dens(Q)	[CPLEX-IP]		[modified KR]	
			iter	time	iter	time
1000	$\approx 10^8$.1	8.0	3.16	6.0	0.41
2000	$\approx 2 \cdot 10^8$.1	8.0	38.36	6.0	2.19
5000	$\approx 5 \cdot 10^8$.1	8.0	533.44	7.0	31.95

Table 14: Convex quadratic minimization over the unit cube. Comparison with interior point methods.

6.5 Lack of Strict Complementarity

As a final experiment we now consider problems where strict complementarity does not hold at the optimum. We follow the setup introduced by Moré and Toraldo [18]. In this paper the dimension $n = 100$ of the problem is kept fixed but the data are generated in such a way that both the condition number of Q and the values of dual variables corresponding to active bounds are varied. Specifically, $ncond \in \{3, 12\}$ indicates condition numbers of order $\approx 10^3$ and $\approx 10^{12}$ respectively. Similarly, $ndeg \in \{3, 12\}$ results in multipliers $\alpha_i \approx 10^{-ndeg}$ for active constraints $x_i - b_i = 0$.

We note that the modified KR-method solves substantially more problems within a given number of iterations, and that the number of iterations is never more than 12. For details see Table 15.

6.6 Comparison with Gradient Projection Methods

We conclude the computational section by providing a comparison of computation times between the following codes, all implemented in Matlab, and running on the same machine.

- [KR] and [modified KR],
- [CTK], a Matlab implementation of the projected gradient method, from C. T. Kelley’s website http://www4.ncsu.edu/~ctk/matlab_darts.html,
- [MT], a basic implementation of the method proposed by Moré and Toraldo in [18]. It combines a standard active set strategy with the gradient projection method. The code from [18] is not publicly available, so we use our own Matlab implementation of the approach from [18].

We provide the computation times as well as the number of system solves for [modified KR] and [MT]. In case of [KR] it is the same as the number of iterations, and [CTK] does not solve any linear system.

We first note that our safeguard mechanism to insure global convergence increases the computation times only moderately. While the method [MT] is competitive on the obstacle problem with only one-sided bounds, it is clearly outperformed in all other cases. The code [CTK] is more efficient than [MT] for the obstacle problem with two-sided bounds, but it is much slower than the two versions of [KR]. We also note that both [CTK] and [MT] are quite sensitive to possible ill-conditioning of the data, while both versions of [KR] behave quite robustly. The last two lines provide timings on random instances like in Table 6, again averaged over 2000 runs and problem dimension $n = 2000$.

7 Conclusion

We presented an extension of the infeasible active set method described in [17]. We provided an in-depth analysis of the convergence behaviour of the KR-Algorithm and proved global convergence of the modified KR-method for strictly convex quadratic programming problems without a condition on complementarity. Moreover we demonstrated the efficiency of the modified KR-Algorithm by comparing it to the KR-Algorithm and the CPLEX interior point and active set solvers. Beside its simplicity (no tuning parameters), the modified KR-method offers the favorable features of standard active set methods like the ability to find the exact numerical solution of the problem, and the fact that at each iteration level the size of the linear system which must be solved is determined by the currently inactive set, which can be significantly smaller than the total set of variables. Additionally the modified KR-Algorithm requires even significantly less iterations than interior point methods to find the optimal solution, independent of initialisation, singularity of the system matrix and the number of variables. Thus the presented algorithm inherits and even improves the preferable features of existing methods for problem type (1) and is competitive and often superior to most of them.

Acknowledgement: We thank two anonymous referees for their constructive comments and suggestions for improvement leading to the present version of the paper.

iter	[KR]				[modified KR]			
ndeg	3	12	3	12	3	12	3	12
ncond	3	3	12	12	3	3	12	12
1					83	94	95	100
2					1625	1585	535	534
3	133		2		6302	6320	2161	2156
4	3194		61		1958	1946	2228	2226
5	5646	27	760	19	32	55	1206	1219
6	1027	578	2372	292			731	727
7		3349	3648	1398			985	816
8		4351	2181	3280			881	994
9		1628	818	3116			351	848
10		67	142	1433			43	335
11			15	415			1	43
12			1	42			1	2
total time	14.59	23.22	21.28	25.50	12.14	15.58	26.16	29.79

Table 15: Problem size is fixed to $n = 100$. Each column represents 10,000 random instances with the specified combination of condition number and degree of degeneracy given by (ncond,ndeg). The numbers in a line corresponding to $\text{iter} = k$ indicate, how many of the problems were solved with k iterations. The algorithm was stopped, once the violation of primal and dual feasibility was below $\text{tol} = 10^{-12}$ elementwise.

Obstacle problem with one-sided bounds								
	[KR]		[modified KR]			[CTK]	[MT]	
m	iter	time	iter	solve	time	time	solve	time
64	4	0.05	4	5	0.06	0.32	14	0.04
128	4	0.19	4	5	0.24	3.73	25	0.24
Obstacle problem with two-sided bounds								
	[KR]		[modified KR]			[CTK]	[MT]	
m	iter	time	iter	solve	time	time	solve	time
64	5	0.08	5	8	0.10	0.44	423	1.11
128	4	0.28	4	6	0.39	5.52	1433	19.08
Biharmonic equation								
	[KR]		[modified KR]			[CTK]	[MT]	
m	iter	time	iter	solve	time	time	solve	time
32	10	0.08	9	15	0.10	4.96	532	0.53
64	6	0.19	6	7	0.23	26.11	1739	8.24
Random data from Table 6								
	[KR]		[modified KR]			[CTK]	[MT]	
$\text{cond}(Q)$	iter	time	iter	solve	time	time	solve	time
$\approx 10^3$	5.84	0.13	5.07	9.48	0.17	-	302.09	2.89
$\approx 10^8$	9.03	0.17	6.67	12.10	0.22	-	354.69	3.21

Table 16: Computational comparison of [KR], [modified KR] and two gradient projection methods from the literature.

References

- [1] M. Bergounioux, M. Haddou, M. Hintermüller, and K. Kunisch. A comparison of interior point methods and a Moreau-Yosida based active set strategy for constrained optimal control problems. *SIAM Journal on Optimization*, 11(2):495–521, 2000.
- [2] M. Bergounioux, K. Ito, and K. Kunisch. Primal-Dual Strategy for Constrained Optimal Control Problems. *SIAM Journal on Control and Optimization*, 37:1176–1194, 1999.
- [3] A. Billionnet and S. Elloumi. Using a mixed integer quadratic programming solver for the unconstrained quadratic 0-1 problem. *Mathematical Programming*, 109(1):55–68, 2007.
- [4] R. Byrd, G. Chin, J. Nocedal, and F. Oztoprak. A family of second-order methods for convex ℓ_1 regularized optimization. Technical report, Optimization Online, version July 2012, 2012.
- [5] F. E. Curtis, Z. Han, and D. P. Robinson. A globally convergent primal-dual active-set framework for large-scale convex quadratic optimization. *Computational Optimization and Applications*, 60(2):311–341, 2015.
- [6] M. D’Apuzzo and M. Marino. Parallel computational issues of an interior point method for solving large bound-constrained quadratic programming problems. *Parallel Computing*, 29:467–483, 2003.
- [7] Z. Dostál and J. Schöberl. Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination. *Computational Optimization and Applications*, 30:23–43, 2005.
- [8] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in Signal Processing*, 1(4):586–597, 2007.
- [9] I. B. Gharbia and C. J. Gilbert. Nonconvergence of the plain newton-min algorithm for linear complementarity problems with a P-matrix. *Mathematical Programming*, 134(2):349–364, 2012.
- [10] N. I. M. Gould, D. Orban, and P. L. Toint. CUTer and sifdec: A constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, Dec. 2003.
- [11] N. I. M. Gould, D. Orban, and P. L. Toint. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3):545–557, 2015.
- [12] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semi-smooth newton method. *SIAM Journal on Optimization*, 13(3):865–888, 2003.
- [13] P. Hungerländer. *Algorithms for convex quadratic programming*. Master Thesis, Austria, 2009.
- [14] *IBM ILOG CPLEX V12.1 User’s Manual for CPLEX*. Available at ftp://public.dhe.ibm.com/software/websphere/ilog/docs/optimization/cplex/ps/_usrmanplex.pdf, 2009.
- [15] J. J. Júdice and F. M. Pires. A block principal pivoting algorithm for large-scale strictly monotone linear complementarity problems. *Computers and Operations Research*, 21(5):587–596, 1994.
- [16] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale ℓ_1 -regularized least squares. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):606–617, 2007.
- [17] K. Kunisch and F. Rendl. An infeasible active set method for convex problems with simple bounds. *SIAM Journal on Optimization*, 14(1):35–52, 2003.
- [18] J. Moré and G. Toraldo. Algorithms for bound constrained quadratic problems. *Numerische Mathematik*, 55:377–400, 1989.
- [19] J. Moré and G. Toraldo. On the solution of large quadratic programming problems with bound constraints. *SIAM Journal on Optimization*, 1:93–113, 1991.
- [20] P. M. Pardalos and G. P. Rodgers. Computational aspects of a branch and bound algorithm for quadratic zero-one programming. *Computing*, 45(2):131–144, 1990.
- [21] B. Polyak. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, 9(4):94–112, 1969.

A APPENDIX

A.1 Convergence Argument for Box Constraint Convex QPs

The convergence argument for the modified KR-method presented in Section 4 can be generalized to convex QPs with box constraints $a \leq x \leq b$ as follows.

We introduce an additional active set \mathcal{C} with $x_{\mathcal{C}} = a_{\mathcal{C}}$ and dual variables γ . The KKT system for box-constrained convex QPs is given by

$$\begin{aligned} Qx + q + \alpha + \gamma &= 0, \\ \alpha \circ (b - x) &= 0, \\ \gamma \circ (a - x) &= 0, \\ b - x &\geq 0, \\ x - a &\geq 0, \\ \alpha &\geq 0, \\ \gamma &\leq 0. \end{aligned}$$

To simplify notation let us denote $\text{KKT}(\mathcal{A}, \mathcal{C})$ as the following set of equations:

$$\text{KKT}(\mathcal{A}, \mathcal{C}) \quad Qx + q + \alpha + \gamma = 0, \quad x_{\mathcal{A}} = b_{\mathcal{A}}, \quad x_{\mathcal{C}} = a_{\mathcal{C}}, \quad \alpha_{\mathcal{I}} = \gamma_{\mathcal{I}} = 0.$$

Let us call the pair $(\mathcal{A}, \mathcal{C})$ with $\mathcal{A} \cup \mathcal{C} \subseteq \mathcal{N}$ a primal feasible pair, if the solution x to $\text{KKT}(\mathcal{A}, \mathcal{B})$ is primal feasible, $a \leq x \leq b$.

Suppose that $(\mathcal{A}, \mathcal{B})$ is a primal feasible pair. Then we start a new iteration with the sets

$$\mathcal{B}_s := \{i \in \mathcal{A} : \alpha_i \geq 0\}, \quad \mathcal{D}_s := \{i \in \mathcal{C} : \gamma_i \leq 0\}. \quad (18)$$

The pair $(\mathcal{B}_s, \mathcal{D}_s)$ does not need to be primal feasible, because $[x, \alpha, \gamma] = \text{KKT}(\mathcal{B}_s, \mathcal{D}_s)$ may have $i \in \overline{\mathcal{B}_s}$ such that $x_i > b_i$ or $i \in \overline{\mathcal{D}_s}$ such that $x_i < a_i$. To turn it into a primal feasible pair $(\mathcal{B}, \mathcal{D})$, we carry out the following iterations.

$$\mathcal{B} \leftarrow \mathcal{B}_s, \quad \mathcal{D} \leftarrow \mathcal{D}_s$$

while $(\mathcal{B}, \mathcal{D})$ not primal feasible

$$[x] = \text{KKT}(\mathcal{B}, \mathcal{D}), \quad \mathcal{B} \leftarrow \mathcal{B} \cup \{i \in \overline{\mathcal{B}} : x_i \geq b_i\}, \quad \mathcal{D} \leftarrow \mathcal{D} \cup \{i \in \overline{\mathcal{D}} : x_i \leq a_i\}$$

This iterative scheme will clearly terminate because \mathcal{B} and \mathcal{D} are augmented by adding only elements of \mathcal{J} . The modified KR-method for box constraints starts with a primal feasible pair $(\mathcal{A}, \mathcal{C})$, then generates a new primal feasible pair $(\mathcal{B}, \mathcal{D})$ as described above. Then a new iteration is started with $(\mathcal{B}, \mathcal{D})$ in place of $(\mathcal{A}, \mathcal{C})$.

We are now taking a closer look at the change of the objective function between two consecutive iterations given by the primal feasible pairs $(\mathcal{A}, \mathcal{C})$ and $(\mathcal{B}, \mathcal{D})$.

We let $[x, \alpha, \gamma] = \text{KKT}(\mathcal{A}, \mathcal{C})$ and $[y, \beta, \delta] = \text{KKT}(\mathcal{B}, \mathcal{D})$ be given. Using Lemma 1 and $Qy + q + \beta + \delta = 0$, we conclude

$$J(y) - J(x) = -\frac{1}{2}(x - y)^\top Q(x - y) + (x - y)^\top (\beta + \delta). \quad (19)$$

The first term on the right hand side is nonpositive, so we need to investigate the second one in detail. In order to do so, we have to take a closer look at how $(\mathcal{B}_s, \mathcal{D}_s)$ and $(\mathcal{B}, \mathcal{D})$ change relative to $(\mathcal{A}, \mathcal{C})$. Formally, the situation looks as indicated in Table 17.

To explain this diagram, we note that by definition we have $\mathcal{B}_s \subseteq \mathcal{A}$, $\mathcal{D}_s \subseteq \mathcal{C}$ and $\mathcal{J}_s := \mathcal{N} \setminus (\mathcal{B}_s \cup \mathcal{D}_s)$. The extension of \mathcal{B}_s to \mathcal{B} is done by adding elements from $\mathcal{A} \setminus \mathcal{B}_s$, which are contained in \mathcal{B}_1 , elements from $\mathcal{C} \setminus \mathcal{D}_s$, which are contained in \mathcal{B}_2 and elements from \mathcal{I} , which are contained in \mathcal{B}_3 . Along the same

\mathcal{A}				\mathcal{C}				\mathcal{I}			
\mathcal{B}_s	\mathcal{J}_s			\mathcal{D}_s	\mathcal{J}_s						
\mathcal{B}_s	\mathcal{B}_1	\mathcal{D}_2	\mathcal{J}_1	\mathcal{D}_s	\mathcal{D}_1	\mathcal{B}_2	\mathcal{J}_2	\mathcal{J}_3	\mathcal{B}_3	\mathcal{D}_3	

Table 17: The change from the primal feasible set \mathcal{A} to the primal feasible set $\mathcal{B} = \mathcal{B}_s \cup \mathcal{B}_1 \cup \mathcal{B}_2$.

lines \mathcal{D}_s is extended to \mathcal{D} and $(\mathcal{B}, \mathcal{D})$ then form a primal feasible pair. Finally we have $\mathcal{J}_1 = \mathcal{A} \setminus (\mathcal{B}_s \cup \mathcal{B}_1 \cup \mathcal{D}_2)$, $\mathcal{J}_2 = \mathcal{C} \setminus (\mathcal{D}_s \cup \mathcal{B}_2 \cup \mathcal{D}_1)$ and $\mathcal{J}_3 = \mathcal{I} \setminus (\mathcal{B}_3 \cup \mathcal{D}_3)$. What do we know about $[x, \alpha, \gamma] = \text{KKT}(\mathcal{A}, \mathcal{C})$ and $[y, \beta, \delta] = \text{KKT}(\mathcal{B}, \mathcal{D})$? Clearly $x_{\mathcal{A}} = b_{\mathcal{A}}$, $x_{\mathcal{C}} = a_{\mathcal{C}}$, $a_{\mathcal{I}} < x_{\mathcal{I}} < b_{\mathcal{I}}$ and $\alpha_{\mathcal{I}} = \alpha_{\mathcal{C}} = \gamma_{\mathcal{I}} = \gamma_{\mathcal{A}} = 0$. The definitions of \mathcal{B}_s and \mathcal{D}_s yield $\alpha_{\mathcal{B}_s} \geq 0$, $\gamma_{\mathcal{D}_s} \leq 0$ and $\alpha_{\mathcal{A} \setminus \mathcal{B}_s} < 0$, $\gamma_{\mathcal{A} \setminus \mathcal{D}_s} > 0$. Further we have $y_{\mathcal{B}} = b_{\mathcal{B}}$, $y_{\mathcal{D}} = a_{\mathcal{D}}$, $a_{\mathcal{J}} < y_{\mathcal{J}} < b_{\mathcal{J}}$ and $\beta_{\mathcal{J}} = \beta_{\mathcal{D}} = \delta_{\mathcal{J}} = \delta_{\mathcal{B}} = 0$. This is summarized in Table 18.

	\mathcal{A}				\mathcal{C}				\mathcal{I}			
	\mathcal{B}_s	\mathcal{B}_1	\mathcal{D}_2	\mathcal{J}_1	\mathcal{D}_s	\mathcal{D}_1	\mathcal{B}_2	\mathcal{J}_2	\mathcal{J}_3	\mathcal{B}_3	\mathcal{D}_3	
x	$= b$	$= b$	$= b$	$= b$	$= a$	$= a$	$= a$	$= a$	$a < x < b$	$a < x < b$	$a < x < b$	
y	$= b$	$= b$	$= a$	$a < y < b$	$= a$	$= a$	$= b$	$a < y < b$	$a < y < b$	$= b$	$= a$	
α	≥ 0	< 0	< 0	< 0	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$	
β	?	?	$= 0$	$= 0$	$= 0$	$= 0$?	$= 0$	$= 0$?	$= 0$	
γ	$= 0$	$= 0$	$= 0$	$= 0$	≤ 0	> 0	> 0	> 0	$= 0$	$= 0$	$= 0$	
δ	$= 0$	$= 0$?	$= 0$?	?	$= 0$	$= 0$	$= 0$	$= 0$?	

Table 18: Information on x , y , α , β , γ and δ associated to \mathcal{A} , \mathcal{B} , \mathcal{C} and \mathcal{D} . The questionmarks indicate that the signs of β and δ can not be specified for the respective sets.

Now we can provide some additional useful properties for two consecutive iterations given by the feasible pairs $(\mathcal{A}, \mathcal{C})$ and $(\mathcal{B}, \mathcal{D})$.

Lemma 10. *Let $(\mathcal{A}, \mathcal{C})$ be a primal feasible pair. Then $\mathcal{B}_s \subseteq \mathcal{A}$, $\mathcal{D}_s \subseteq \mathcal{C}$ and $(\mathcal{B}_s = \mathcal{A}) \wedge (\mathcal{D}_s = \mathcal{C})$ if and only if $(\mathcal{A}, \mathcal{C})$ is optimal. If $(\mathcal{A}, \mathcal{C})$ is not optimal, then $(\mathcal{A} \neq \emptyset) \vee (\mathcal{C} \neq \emptyset)$, $(\mathcal{A} \neq \mathcal{B}) \vee (\mathcal{C} \neq \mathcal{D})$ and $x - y \neq 0$.*

Proof. $\mathcal{B}_s \subseteq \mathcal{A}$ and $\mathcal{D}_s \subseteq \mathcal{C}$ hold due to (18). If $(\mathcal{A}, \mathcal{C})$ is optimal then all elements in \mathcal{A} and \mathcal{C} are dual feasible and hence $(\mathcal{B}_s = \mathcal{A}) \wedge (\mathcal{D}_s = \mathcal{C})$. If $(\mathcal{B}_s = \mathcal{A}) \wedge (\mathcal{D}_s = \mathcal{C})$ then all elements in \mathcal{A} and \mathcal{C} are dual feasible. As the algorithm ensures all other KKT conditions including primal feasibility throughout all iterations, $(\mathcal{A}, \mathcal{C})$ is optimal. If $\mathcal{A} = \mathcal{C} = \emptyset$ then the unconstrained optimum is feasible and hence $(\mathcal{A}, \mathcal{C})$ is optimal.

Now let's suppose $(\mathcal{A}, \mathcal{C})$ not optimal and $(\mathcal{A} = \mathcal{B}) \wedge (\mathcal{C} = \mathcal{D})$. Then $\mathcal{J}_1 = \mathcal{J}_2 = \mathcal{B}_2 = \mathcal{D}_2 = \mathcal{B}_3 = \mathcal{D}_3 = \emptyset$ and hence $\mathcal{J}_3 = \mathcal{J} = \mathcal{I}$, $\mathcal{B}_1 = \mathcal{A} \setminus \mathcal{B}_s \neq \emptyset$ and $\mathcal{D}_1 = \mathcal{C} \setminus \mathcal{D}_s \neq \emptyset$. Due to the workings of the modified KR-method the following implications hold

$$\begin{aligned}
[x, \alpha, \gamma] = \text{KKT}(\mathcal{A}, \mathcal{C}) &\Rightarrow \alpha_{\mathcal{B}_1} < 0 \wedge \gamma_{\mathcal{D}_1} > 0, \\
[y, \beta, \delta] = \text{KKT}(\mathcal{B}_s, \mathcal{D}_s) &\Rightarrow \beta_{\mathcal{B}_1} = 0 \wedge \gamma_{\mathcal{D}_1} = 0.
\end{aligned} \tag{20}$$

As $\mathcal{I} = \mathcal{J}$ we can substitute the primal variables associated with \mathcal{I} by

$$\begin{aligned}
x_{\mathcal{I}} &= -Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I}, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{I}, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{I}, \mathcal{B}_1} b_{\mathcal{B}_1} + Q_{\mathcal{I}, \mathcal{D}_1} a_{\mathcal{D}_1}), \\
y_{\mathcal{I}} &= -Q_{\mathcal{I}}^{-1}(q_{\mathcal{I}} + Q_{\mathcal{I}, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{I}, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{I}, \mathcal{B}_1} y_{\mathcal{B}_1} + Q_{\mathcal{I}, \mathcal{D}_1} y_{\mathcal{D}_1}).
\end{aligned} \tag{21}$$

Throughout the algorithm the equations $Qx + q + \alpha + \gamma = 0$ hold. Using (20) we obtain

$$\begin{aligned} Q_{\mathcal{B}_1, N} \begin{pmatrix} b_{\mathcal{B}_s} \\ b_{\mathcal{B}_1} \\ a_{\mathcal{D}_s} \\ a_{\mathcal{D}_1} \\ x_{\mathcal{I}} \end{pmatrix} + q_{\mathcal{B}_1} &= -\alpha_{\mathcal{B}_1} > 0, & Q_{\mathcal{D}_1, N} \begin{pmatrix} b_{\mathcal{B}_s} \\ b_{\mathcal{B}_1} \\ a_{\mathcal{D}_s} \\ a_{\mathcal{D}_1} \\ x_{\mathcal{I}} \end{pmatrix} + q_{\mathcal{D}_1} &= -\delta_{\mathcal{D}_1} < 0, \\ Q_{\mathcal{B}_1, N} \begin{pmatrix} b_{\mathcal{B}_s} \\ y_{\mathcal{B}_1} \\ a_{\mathcal{D}_s} \\ y_{\mathcal{D}_1} \\ y_{\mathcal{I}} \end{pmatrix} + q_{\mathcal{B}_1} &= -\beta_{\mathcal{B}_1} = 0, & Q_{\mathcal{D}_1, N} \begin{pmatrix} b_{\mathcal{B}_s} \\ y_{\mathcal{B}_1} \\ a_{\mathcal{D}_s} \\ y_{\mathcal{D}_1} \\ y_{\mathcal{I}} \end{pmatrix} + q_{\mathcal{D}_1} &= -\delta_{\mathcal{D}_1} = 0. \end{aligned}$$

Next applying (21) to the above equations yields

$$\begin{aligned} &Q_{\mathcal{B}_1, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{B}_1} b_{\mathcal{B}_1} + Q_{\mathcal{B}_1, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{B}_1, \mathcal{D}_1} a_{\mathcal{D}_1} - Q_{\mathcal{B}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} (q_{\mathcal{I}} + Q_{\mathcal{I}, \mathcal{B}_s} b_{\mathcal{B}_s} + \\ &Q_{\mathcal{I}, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{I}, \mathcal{B}_1} b_{\mathcal{B}_1} + Q_{\mathcal{I}, \mathcal{D}_1} a_{\mathcal{D}_1}) + q_{\mathcal{B}_1} > Q_{\mathcal{B}_1, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{B}_1} y_{\mathcal{B}_1} + Q_{\mathcal{B}_1, \mathcal{D}_s} a_{\mathcal{D}_s} + \\ &Q_{\mathcal{B}_1, \mathcal{D}_1} y_{\mathcal{D}_1} - Q_{\mathcal{B}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} (q_{\mathcal{I}} + Q_{\mathcal{I}, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{I}, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{I}, \mathcal{B}_1} y_{\mathcal{B}_1} + Q_{\mathcal{I}, \mathcal{D}_1} y_{\mathcal{D}_1}) + q_{\mathcal{B}_1}, \end{aligned}$$

and

$$\begin{aligned} &Q_{\mathcal{D}_1, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{D}_1, \mathcal{B}_1} b_{\mathcal{B}_1} + Q_{\mathcal{D}_1, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{D}_1} a_{\mathcal{D}_1} - Q_{\mathcal{D}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} (q_{\mathcal{I}} + Q_{\mathcal{I}, \mathcal{B}_s} b_{\mathcal{B}_s} + \\ &Q_{\mathcal{I}, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{I}, \mathcal{B}_1} b_{\mathcal{B}_1} + Q_{\mathcal{I}, \mathcal{D}_1} a_{\mathcal{D}_1}) + q_{\mathcal{D}_1} < Q_{\mathcal{D}_1, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{D}_1, \mathcal{B}_1} y_{\mathcal{B}_1} + Q_{\mathcal{D}_1, \mathcal{D}_s} a_{\mathcal{D}_s} + \\ &Q_{\mathcal{D}_1} y_{\mathcal{D}_1} - Q_{\mathcal{D}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} (q_{\mathcal{I}} + Q_{\mathcal{I}, \mathcal{B}_s} b_{\mathcal{B}_s} + Q_{\mathcal{I}, \mathcal{D}_s} a_{\mathcal{D}_s} + Q_{\mathcal{I}, \mathcal{B}_1} y_{\mathcal{B}_1} + Q_{\mathcal{I}, \mathcal{D}_1} y_{\mathcal{D}_1}) + q_{\mathcal{D}_1}. \end{aligned}$$

Simplifying the above inequalities gives

$$\begin{aligned} &(Q_{\mathcal{B}_1} - Q_{\mathcal{B}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{B}_1}) b_{\mathcal{B}_1} + (Q_{\mathcal{B}_1, \mathcal{D}_1} - Q_{\mathcal{B}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{D}_1}) a_{\mathcal{D}_1} > \\ &(Q_{\mathcal{B}_1} - Q_{\mathcal{B}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{B}_1}) y_{\mathcal{B}_1} + (Q_{\mathcal{B}_1, \mathcal{D}_1} - Q_{\mathcal{B}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{D}_1}) y_{\mathcal{D}_1}. \end{aligned}$$

and

$$\begin{aligned} &(Q_{\mathcal{D}_1} - Q_{\mathcal{D}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{D}_1}) a_{\mathcal{D}_1} + (Q_{\mathcal{D}_1, \mathcal{B}_1} - Q_{\mathcal{D}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{B}_1}) b_{\mathcal{B}_1} < \\ &(Q_{\mathcal{D}_1} - Q_{\mathcal{D}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{D}_1}) y_{\mathcal{D}_1} + (Q_{\mathcal{D}_1, \mathcal{B}_1} - Q_{\mathcal{D}_1, \mathcal{I}} Q_{\mathcal{I}}^{-1} Q_{\mathcal{I}, \mathcal{B}_1}) y_{\mathcal{B}_1}. \end{aligned}$$

Now adding up the two inequalities yields

$$\left[\begin{pmatrix} Q_{\mathcal{B}_1} & -Q_{\mathcal{B}_1, \mathcal{D}_1} \\ -Q_{\mathcal{D}_1, \mathcal{B}_1} & Q_{\mathcal{D}_1} \end{pmatrix} - \begin{pmatrix} Q_{\mathcal{B}_1, \mathcal{I}} \\ -Q_{\mathcal{D}_1, \mathcal{I}} \end{pmatrix} Q_{\mathcal{I}}^{-1} \begin{pmatrix} Q_{\mathcal{I}, \mathcal{B}_1} & -Q_{\mathcal{I}, \mathcal{D}_1} \end{pmatrix} \right] \begin{pmatrix} b_{\mathcal{B}_1} - y_{\mathcal{B}_1} \\ y_{\mathcal{D}_1} - a_{\mathcal{D}_1} \end{pmatrix} > 0. \quad (22)$$

The matrix

$$\begin{pmatrix} Q_{\mathcal{B}_1} & -Q_{\mathcal{B}_1, \mathcal{D}_1} \\ -Q_{\mathcal{D}_1, \mathcal{B}_1} & Q_{\mathcal{D}_1} \end{pmatrix} - \begin{pmatrix} Q_{\mathcal{B}_1, \mathcal{I}} \\ -Q_{\mathcal{D}_1, \mathcal{I}} \end{pmatrix} Q_{\mathcal{I}}^{-1} \begin{pmatrix} Q_{\mathcal{I}, \mathcal{B}_1} & -Q_{\mathcal{I}, \mathcal{D}_1} \end{pmatrix}, \quad (23)$$

is positive definite. To see this we use

$$\begin{pmatrix} Q_{\mathcal{B}_1} & -Q_{\mathcal{B}_1, \mathcal{D}_1} \\ -Q_{\mathcal{D}_1, \mathcal{B}_1} & Q_{\mathcal{D}_1} \end{pmatrix} \succ 0, \quad \text{and} \quad M_1 := \begin{pmatrix} Q_{\mathcal{B}_1} & -Q_{\mathcal{B}_1, \mathcal{D}_1} & Q_{\mathcal{B}_1, \mathcal{I}} \\ -Q_{\mathcal{D}_1, \mathcal{B}_1} & Q_{\mathcal{D}_1} & -Q_{\mathcal{D}_1, \mathcal{I}} \\ Q_{\mathcal{I}, \mathcal{B}_1} & -Q_{\mathcal{I}, \mathcal{D}_1} & Q_{\mathcal{I}} \end{pmatrix} \succ 0,$$

in the Schur-complement lemma. The two above matrices are positive definite because

$$M_2 := \begin{pmatrix} Q_{\mathcal{B}_1} & Q_{\mathcal{B}_1, \mathcal{D}_1} & Q_{\mathcal{B}_1, \mathcal{I}} \\ Q_{\mathcal{D}_1, \mathcal{B}_1} & Q_{\mathcal{D}_1} & Q_{\mathcal{D}_1, \mathcal{I}} \\ Q_{\mathcal{I}, \mathcal{B}_1} & Q_{\mathcal{I}, \mathcal{D}_1} & Q_{\mathcal{I}} \end{pmatrix} \succ 0, \quad \text{and} \quad \begin{pmatrix} x_{\mathcal{B}_1} \\ x_{\mathcal{D}_1} \\ x_{\mathcal{I}} \end{pmatrix}^\top M_2 \begin{pmatrix} x_{\mathcal{B}_1} \\ x_{\mathcal{D}_1} \\ x_{\mathcal{I}} \end{pmatrix} = \begin{pmatrix} x_{\mathcal{B}_1} \\ -x_{\mathcal{D}_1} \\ x_{\mathcal{I}} \end{pmatrix}^\top M_1 \begin{pmatrix} x_{\mathcal{B}_1} \\ -x_{\mathcal{D}_1} \\ x_{\mathcal{I}} \end{pmatrix}.$$

To assure $(\mathcal{A}, \mathcal{C}) = (\mathcal{B}, \mathcal{D})$ the inequality

$$\begin{pmatrix} b_{\mathcal{B}_1} - y_{\mathcal{B}_1} \\ y_{\mathcal{D}_1} - a_{\mathcal{D}_1} \end{pmatrix} \leq 0, \quad (24)$$

has to hold. Combining (22) and (24) yields a contradiction to the positive definiteness of the matrix in (23):

$$\begin{pmatrix} b_{\mathcal{B}_1} - y_{\mathcal{B}_1} \\ y_{\mathcal{D}_1} - a_{\mathcal{D}_1} \end{pmatrix}^\top \left[\begin{pmatrix} Q_{\mathcal{B}_1} & -Q_{\mathcal{B}_1, \mathcal{D}_1} \\ -Q_{\mathcal{D}_1, \mathcal{B}_1} & Q_{\mathcal{D}_1} \end{pmatrix} - \begin{pmatrix} Q_{\mathcal{B}_1, \mathcal{I}} \\ -Q_{\mathcal{D}_1, \mathcal{I}} \end{pmatrix} Q_{\mathcal{I}}^{-1} (Q_{\mathcal{I}, \mathcal{B}_1} \quad -Q_{\mathcal{I}, \mathcal{D}_1}) \right] \begin{pmatrix} b_{\mathcal{B}_1} - y_{\mathcal{B}_1} \\ y_{\mathcal{D}_1} - a_{\mathcal{D}_1} \end{pmatrix} \leq 0. \quad (25)$$

Finally using Table 18 we have $a < x_{\mathcal{I}} < b_{\mathcal{I}}$, $a < y_{\mathcal{J}} < b_{\mathcal{J}}$, $x_{\mathcal{A}} = b_{\mathcal{A}}$, $x_{\mathcal{C}} = a_{\mathcal{C}}$, $y_{\mathcal{B}} = b_{\mathcal{B}}$, $y_{\mathcal{D}} = a_{\mathcal{D}}$ and hence $x - y \neq 0$ for $(\mathcal{A}, \mathcal{C}) \neq (\mathcal{B}, \mathcal{D})$. \square

As $x - y = 0$ on $\mathcal{B}_s \cup \mathcal{B}_1 \cup \mathcal{D}_s \cup \mathcal{D}_1$, $\beta = 0$ on $\mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3 \cup \mathcal{D}_s \cup \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3$ and $\delta = 0$ on $\mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3 \cup \mathcal{B}_s \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$ it follows that $(x - y)^\top \beta = \sum_{i \in \mathcal{B}_2 \cup \mathcal{B}_3} (x - y)_i \beta_i$ and $(x - y)^\top \delta = \sum_{i \in \mathcal{D}_2 \cup \mathcal{D}_3} (x - y)_i \delta_i$. Since $(x - y)_i < 0$ on $\mathcal{B}_2 \cup \mathcal{B}_3$ and $(x - y)_i > 0$ on $\mathcal{D}_2 \cup \mathcal{D}_3$, the contribution of $(x - y)^\top (\beta + \delta)$ is guaranteed to be nonpositive in case $\beta \geq 0$ on $\mathcal{B}_2 \cup \mathcal{B}_3$ and $\delta \leq 0$ on $\mathcal{D}_2 \cup \mathcal{D}_3$. We summarize the discussion in the following lemma.

Lemma 11. *Let $x, y, \alpha, \beta, \gamma, \delta$ and $\mathcal{B}_2, \mathcal{B}_3, \mathcal{D}_2, \mathcal{D}_3$ be as above. Then*

$$J(y) - J(x) = -\frac{1}{2}(x - y)^\top Q(x - y) - \sum_{i \in \mathcal{B}_2 \cup \mathcal{B}_3} (x - y)_i \beta_i + \sum_{i \in \mathcal{D}_2 \cup \mathcal{D}_3} (x - y)_i \delta_i.$$

Moreover, $(x - y)_{\mathcal{B}_2 \cup \mathcal{B}_3} < 0$ and $(x - y)_{\mathcal{D}_2 \cup \mathcal{D}_3} > 0$.

In general we have no means to avoid $J(y) \geq J(x)$ using the current algorithmic setup. To avoid cycling, we therefore need to take additional measures in case $J(y) \geq J(x)$. Let us assume $(\mathcal{A}, \mathcal{C})$ not optimal for the following case distinction. We first recall that in this case $|\mathcal{A}| + |\mathcal{C}| \geq 1$ due to Lemma 10. We consider the following cases separately. Let $x, y, \alpha, \beta, \gamma, \delta$ and $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and \mathcal{D} be as above.

Case 1: $J(y) < J(x)$.

In this case we can set $\mathcal{A} \leftarrow \mathcal{B}$, $\mathcal{C} \leftarrow \mathcal{D}$, and continue with the next iteration.

Case 2: $J(y) \geq J(x)$ and $|\mathcal{A}| + |\mathcal{C}| = 1$.

The pair $(\mathcal{A}, \mathcal{C})$ with $\mathcal{A} \cup \mathcal{C} = \{j\}$ is primal feasible. Since x is primal feasible, but not optimal, we must have $\alpha_j < 0$ if $j \in \mathcal{A}$ and $\gamma_j > 0$ if $j \in \mathcal{C}$. Thus the objective function would improve by allowing x_j to move away from the respective boundary. Hence in an optimal solution it is essential that $x_j < b_j$ if $j \in \mathcal{A}$ and $x_j > a_j$ if $j \in \mathcal{C}$. Thus we have a problem with only $2n - 1$ constraints which we solve by induction on the number of constraints.

Case 3: $J(y) \geq J(x)$ and $|\mathcal{A}| + |\mathcal{C}| > 1$.

In this case we will again solve a problem with less than $2n$ constraints to optimality. Its optimal solution will yield a feasible pair $(\mathcal{B}, \mathcal{D})$ and a primal feasible y with $J(y) < J(x)$. Our strategy is to identify two sets $\mathcal{A}_0 \subseteq \mathcal{A}$ and $\mathcal{C}_0 \subseteq \mathcal{C}$ with $\mathcal{A}_0 \cup \mathcal{C}_0 \neq \emptyset$ such that x is feasible but not optimal for the subproblem with $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}$, $x_{\mathcal{C}_0} = a_{\mathcal{C}_0}$, $x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}$, $x_{\bar{\mathcal{C}}_0} \geq a_{\bar{\mathcal{C}}_0}$. We are then (recursively) solving the smaller subproblem with $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}$, $x_{\mathcal{C}_0} = a_{\mathcal{C}_0}$, $x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}$, $x_{\bar{\mathcal{C}}_0} \geq a_{\bar{\mathcal{C}}_0}$ to optimality yielding a new primal feasible pair $(\mathcal{B}, \mathcal{D})$ with improved objective value.

We choose \mathcal{A}_0 as a subset of \mathcal{A} and \mathcal{C}_0 as a subset of \mathcal{C} with $(\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{J}_1 \cup \mathcal{J}_2) \setminus (\mathcal{A}_0 \cup \mathcal{C}_0) \neq \emptyset$ and $\mathcal{A}_0 \cup \mathcal{C}_0 \neq \emptyset$. Note in particular that $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{J}_1 \cup \mathcal{J}_2$ is nonempty due to Lemma 10. Hence the solution $[x, \alpha, \gamma] = \text{KKT}(\mathcal{A}, \mathcal{C})$ is feasible for $x_{\mathcal{A}_0} = b_{\mathcal{A}_0} \wedge x_{\mathcal{C}_0} = a_{\mathcal{C}_0}$, but not optimal, because $\alpha_i < 0$ for $i \in (\mathcal{B}_1 \cup \mathcal{D}_2 \cup \mathcal{J}_1) \setminus \mathcal{A}_0$, $\gamma_i > 0$ for $i \in (\mathcal{D}_1 \cup \mathcal{B}_2 \cup \mathcal{J}_2) \setminus \mathcal{C}_0$ and $(\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{J}_1 \cup \mathcal{J}_2) \setminus (\mathcal{A}_0 \cup \mathcal{C}_0) \neq \emptyset$.

Now let $(\mathcal{B}_0, \mathcal{D}_0)$ be the optimal set of the subproblem, given by $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}$, $x_{\mathcal{C}_0} = a_{\mathcal{C}_0}$, $x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}$, $x_{\bar{\mathcal{C}}_0} \geq a_{\bar{\mathcal{C}}_0}$, which can be determined by induction. We set $\mathcal{B} := \mathcal{A}_0 \cup \mathcal{B}_0$, $\mathcal{D} := \mathcal{C}_0 \cup \mathcal{D}_0$ and get $[y, \beta, \delta] = \text{KKT}(\mathcal{B}, \mathcal{D})$. By construction, $(\mathcal{B}, \mathcal{D})$ is feasible, and $J(y) < J(x)$ because y is optimal for the subproblem.

We summarize the modified KR-method for box constraints, which produces primal feasible iterates in each iteration, in Table 19. Since the objective value strictly decreases in each iteration, we have shown the following result

Theorem 12. *The modified KR-Algorithm for box constraints as given in Table 19 terminates in a finite number of iterations with the optimal solution.*

Modified KR-Algorithm for box constraints

Input: $Q \succ 0$, a , b , $d \in \mathbb{R}^n$. $\mathcal{A} \cup \mathcal{C} \subseteq \mathcal{N}$, $\mathcal{A} \cap \mathcal{C} = \emptyset$.

Output: optimal active pair $(\mathcal{A}, \mathcal{C})$.

$x = \text{KKT}(\mathcal{A}, \mathcal{C})$.

while $(\mathcal{A}, \mathcal{C})$ not primal feasible:

$\mathcal{A} \leftarrow \mathcal{A} \cup \{i \in \bar{\mathcal{A}} : x_i \geq b_i\}$.

$\mathcal{C} \leftarrow \mathcal{C} \cup \{i \in \bar{\mathcal{C}} : x_i \leq a_i\}$.

$x = \text{KKT}(\mathcal{A}, \mathcal{C})$.

endwhile

$[\alpha, \gamma] = \text{KKT}(\mathcal{A}, \mathcal{C})$

while $(\mathcal{A}, \mathcal{C})$ not optimal

$\mathcal{B}_s \leftarrow \{i \in \mathcal{A} : \alpha_i \geq 0\}; \mathcal{B} \leftarrow \mathcal{B}_s$.

$\mathcal{D}_s \leftarrow \{i \in \mathcal{C} : \gamma_i \leq 0\}; \mathcal{D} \leftarrow \mathcal{D}_s$.

$y = \text{KKT}(\mathcal{B}, \mathcal{D})$.

while $(\mathcal{B}, \mathcal{D})$ not primal feasible

$\mathcal{B} \leftarrow \mathcal{B} \cup \{i \in \bar{\mathcal{B}} : y_i \geq b_i\}$.

$\mathcal{D} \leftarrow \mathcal{D} \cup \{i \in \bar{\mathcal{D}} : y_i \leq a_i\}$.

$y = \text{KKT}(\mathcal{B}, \mathcal{D})$.

endwhile

Case 1: $J(y) < J(x)$

$\mathcal{A} \leftarrow \mathcal{B}$, $\mathcal{C} \leftarrow \mathcal{D}$.

Case 2: $J(y) \geq J(x)$ and $|\mathcal{A}| + |\mathcal{C}| = 1$.

Let $(\mathcal{A}_{opt}, \mathcal{C}_{opt})$ be the optimal pair for the subproblem with the bound on $\mathcal{A} \cup \mathcal{C} = \{j\}$ removed.

$\mathcal{A} \leftarrow \mathcal{A}_{opt}$, $\mathcal{C} \leftarrow \mathcal{C}_{opt}$ is optimal, stop.

Case 3: $J(y) \geq J(x)$ and $|\mathcal{A}| + |\mathcal{C}| > 1$

$\mathcal{A}_0 \subseteq \mathcal{A}$, $\mathcal{C}_0 \subseteq \mathcal{C}$, $\mathcal{A}_0 \cup \mathcal{C}_0 \neq \emptyset$ with $(\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{I}_1 \cup \mathcal{I}_2) \setminus (\mathcal{A}_0 \cup \mathcal{C}_0) \neq \emptyset$.

Let $(\mathcal{B}_0, \mathcal{D}_0)$ be the optimal pair for the subproblem with $x_{\mathcal{A}_0} = b_{\mathcal{A}_0}$, $x_{\mathcal{C}_0} = a_{\mathcal{C}_0}$, $x_{\bar{\mathcal{A}}_0} \leq b_{\bar{\mathcal{A}}_0}$, $x_{\bar{\mathcal{C}}_0} \geq a_{\bar{\mathcal{C}}_0}$.

$\mathcal{A} \leftarrow \mathcal{A}_0 \cup \mathcal{B}_0$, $\mathcal{C} \leftarrow \mathcal{C}_0 \cup \mathcal{D}_0$.

$[\alpha, \gamma] = \text{KKT}(\mathcal{A}, \mathcal{C})$

endwhile

Table 19: Description of the modified KR-Algorithm for box constraints.