Noname manuscript No.
(will be inserted by the editor)

A Constraint-Reduced MPC Algorithm for Convex Quadratic Programming, with a Modified Active Set Identification Scheme

M. Paul Laiu · André L. Tits

February 20, 2019

Abstract A constraint-reduced Mehrotra–Predictor–Corrector algorithm for convex quadratic programming is proposed. (At each iteration, such algorithms use only a subset of the inequality constraints in constructing the search direction, resulting in CPU savings.) The proposed algorithm makes use of a regularization scheme to cater to cases where the reduced constraint matrix is rank deficient. Global and local convergence properties are established under arbitrary working-set selection rules subject to satisfaction of a general condition. A modified active-set identification scheme that fulfills this condition is introduced. Numerical tests show great promise for the proposed algorithm, in particular for its active-set identification scheme. While the focus of the present paper is on dense systems, application of the main ideas to large sparse systems is briefly discussed.

M. Paul Laiu
Computational and Applied Mathematics Group, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA
E-mail: laiump@ornl.gov
This author's research was sponsored by the Office of Advanced Scientific Computing Research and performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725.

André L. Tits
Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland College Park, MD 20742 USA,
E-mail: andre@umd.edu

1 Introduction

Consider a strictly feasible convex quadratic program (CQP) in standard inequality form,[1]

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x} \quad \text{subject to } A\mathbf{x} \geq \mathbf{b}\,, \tag{P}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of optimization variables, $f\colon \mathbb{R}^n \to \mathbb{R}$ is the objective function with $\mathbf{c} \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n\times n}$ a symmetric positive semi-definite matrix, $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, with $m > 0$, define the m linear inequality constraints, and (here and elsewhere) all inequalities ($\geq$ or $\leq$) are meant component-wise. H, A, and $\mathbf{c}$ are not all zero. The dual problem associated to (P) is

$$\underset{\mathbf{x}\in\mathbb{R}^n,\,\boldsymbol{\lambda}\in\mathbb{R}^m}{\text{maximize}}\ -\frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{b}^T\boldsymbol{\lambda} \quad \text{subject to } H\mathbf{x} + \mathbf{c} - A^T\boldsymbol{\lambda} = \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}\,, \tag{D}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ is the vector of multipliers. Since the objective function f is convex and the constraints are linear, solving (P)–(D) is equivalent to solving the Karush-Kuhn-Tucker (KKT) system

$$H\mathbf{x} - A^T\boldsymbol{\lambda} + \mathbf{c} = \mathbf{0}, \quad A\mathbf{x} - \mathbf{b} - \mathbf{s} = \mathbf{0}, \quad S\boldsymbol{\lambda} = \mathbf{0}, \quad \mathbf{s}, \boldsymbol{\lambda} \geq \mathbf{0}\,, \tag{1}$$

where $\mathbf{s} \in \mathbb{R}^n$ is a vector of slack variables associated to the inequality constraints in (P), and $S = \text{diag}(\mathbf{s})$.

Primal–dual interior–point methods (PDIPM) solve (P)–(D) by iteratively applying a Newton-like iteration to the three *equations* in (1). Especially popular for its numerical behavior is S. Mehrotra's predictor–corrector (MPC) variant, which was introduced in [21] for the case of linear optimization (i.e., $H = \mathbf{0}$) with straightforward extension to CQP (e.g., [22, Section 16.6]). (Extension to linear complementarity problems was studied in [39].)

A number of authors have paid special attention to "imbalanced" problems, in which the number of active constraints at the solution is small compared to the total number of constraints (in particular, cases in which $m \gg n$). In solving such problems, while most constraints are in a sense redundant, traditional IPMs devote much effort per iteration to solving large systems of linear equations that involve all m constraints. In the 1990s, work toward reducing the computational burden by using only a small portion ("working set") of the constraints in the search direction computation focused mainly on *linear* optimization [6,15,31,38]. This was also the case for [28], which may have been the first to consider such "constraint-reduction" schemes in the context of PDIPMs (vs. purely dual interior–point methods), and for its extensions [14,34,35]. Exceptions include the works of Jung *et al.* [17,18] and of Park *et al.* [23,24]; in the former, an extension to CQP was considered, with an affine-scaling variant used as the "base" algorithm; in the latter, a constraint-reduced PDIPM for

[1] See end of Sections 2.3 and 2.5 below for a brief discussion of how linear *equality* constraints can be incorporated.

semi-definite optimization (which includes CQP as a special case) was proposed, for which polynomial complexity was proved. Another exception is the "QP-free" algorithm for inequality-constrained (smooth) nonlinear programming of Chen *et al.* [5]. There, a constraint-reduction approach is used where working sets are selected by means of the Facchinei–Fischer–Kanzow active set identification technique [8].

In the linear-optimization case, the size of the working set is usually kept above (or no lower than) the number n of variables in (P). It is indeed known that, in that case, if the set of solutions to (P) is nonempty and bounded, then a solution exists at which at least n constraints are active. Further if fewer than n constraints are included in the linear system that defines the search direction, then the default (Newton-KKT) system matrix is structurally singular.

When $H \neq \mathbf{0}$ though, the number of active constraints at solutions may be much less than n and, at strictly feasible iterates, the Newton-KKT matrix is non-singular whenever the subspace spanned by the columns of H and the working-set columns of A^T has (full) dimension n. (In particular, of course, if H is non-singular (i.e., is positive definite) the Newton-KKT matrix is non-singular even when the working set is empty—in which case the unconstrained Newton direction is obtained.) Hence, when solving a CQP, forcing the working set to have size at least n is usually wasteful.

The present paper proposes a constraint-reduced MPC algorithm for CQP. The work borrows from [18] (affine-scaling, convex quadratic programming) and is significantly inspired from [34] (MPC, linear optimization), but improves on both in a number of ways—even for the case of linear optimization (i.e., when $H = \mathbf{0}$). Specifically,

- in contrast with [18, 34] (and [5]), it does not involve a (CPU-expensive) rank estimation combined with an increase of the size of the working set when a rank condition fails; rather, it makes use of a regularization scheme adapted from [35];
- a general condition (Condition CSR) to be satisfied by the constraint-selection rule is proposed which, when satisfied, guarantees global and local quadratic convergence of the overall algorithm (under appropriate assumptions);
- a specific constraint-selection rule is introduced which, like in [5] (but unlike in [18]), does not impose any *a priori* lower bound on the size of the working set; this rule involves a modified active set identification scheme that builds on results from [8]; numerical comparison shows that the new rule outperforms previously used rules.

Other improvements over [18, 34] include (i) a modified stopping criterion and a proof of termination of the algorithm (in [18, 34], termination is only guaranteed under uniqueness of primal-dual solution and strict complementarity), (ii) a potentially larger value of the "mixing parameter" in the definition of the primal search direction (see footnote 2(iii)), and (iii) an improved update formula (compared to that used in [35]) for the regularization parameter, which fosters a smooth evolution of the regularized Hessian W from an initial (matrix) value $H + R$, where R is specified by the user, at a rate no faster than that required for local q-quadratic convergence.

In Section 2 below, we introduce the proposed algorithm (Algorithm CR-MPC) and a general condition (Condition CSR) to be satisfied by the constraint-selection rule, and we state global and local quadratic convergence results for Algorithm CR-MPC, subject to Condition CSR. We conclude the section by proposing a specific rule (Rule R), and proving that it satisfies Condition CSR. Numerical results are reported in Section 3. While the focus of the present paper is on dense systems, application of the main ideas to large sparse systems is briefly discussed in the Conclusion (Section 4), which also includes other concluding remarks. Convergence proofs are given in two appendices.

The following notation is used throughout the paper. To the number m of inequality constraints, we associate the index set $\mathbf{m} := \{1, 2, \dots, m\}$. The primal feasible and primal strictly feasible sets are

$$\mathcal{F}_P := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \geq \mathbf{b}\} \quad \text{and} \quad \mathcal{F}_P^o := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} > \mathbf{b}\},$$

and the primal and primal-dual solution sets are

$$\mathcal{F}_P^* := \{\mathbf{x} \in \mathcal{F}_P : f(\mathbf{x}) \leq f(\tilde{\mathbf{x}}), \forall \tilde{\mathbf{x}} \in \mathcal{F}_P\} \quad \text{and} \quad \mathcal{F}^* := \{(\mathbf{x}, \boldsymbol{\lambda}) : (1) \text{ holds}\} .$$

Of course, $\mathbf{x}^* \in \mathcal{F}_P^*$ if and only if, for some $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{F}^*$ Also, we term *stationary* a point $\hat{\mathbf{x}} \in \mathcal{F}_P$ for which there exists $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^m$ such that $(\hat{\mathbf{x}}, \hat{\boldsymbol{\lambda}})$ satisfies (1) except possibly for non-negativity of the components of $\hat{\boldsymbol{\lambda}}$. Next, for $\mathbf{x} \in \mathcal{F}_P$, the (primal) *active-constraint* set at $\mathbf{x}$ is

$$\mathcal{A}(\mathbf{x}) := \{i \in \mathbf{m} : \mathbf{a}_i^T \mathbf{x} = b_i\} ,$$

where $\mathbf{a}_i$ is the transpose of the i-th row of A. Given a subset $Q \subseteq \mathbf{m}$, Q^{c} indicates its complement in $\mathbf{m}$ and $|Q|$ its cardinality; for a vector $\mathbf{v} \in \mathbb{R}^m$, $\mathbf{v}_Q$ is a sub-vector consisting of those entries with index in Q, and for an $m \times n$ matrix L, L_Q is a $|Q| \times n$ sub-matrix of L consisting of those rows with index in Q. An exception to this rule, which should not create any confusion, is that for an $m \times m$ *diagonal* matrix $V = \mathrm{diag}(\mathbf{v})$, V_Q is $\mathrm{diag}(\mathbf{v}_Q)$, a $|Q| \times |Q|$ diagonal sub-matrix of V. For symmetric matrices W and H, $W \succeq H$ (resp. $W \succ H$) means that $W - H$ is positive semi-definite (resp. positive definite). Finally, given a vector $\mathbf{v}$, $[\mathbf{v}]_+$ and $[\mathbf{v}]_-$ denote the positive and negative parts of $\mathbf{v}$, i.e., vectors with components respectively given by $\max\{v_i, 0\}$ and $\min\{v_i, 0\}$, $\mathbf{1}$ is a vector of all ones, and given a Euclidean space $\mathbb{R}^p$, the ball centered at $\mathbf{v}^* \in \mathbb{R}^p$ with radius $\rho > 0$ is denoted by $B(\mathbf{v}^*, \rho) := \{\mathbf{v} \in \mathbb{R}^p : \|\mathbf{v} - \mathbf{v}^*\| \leq \rho\}$.

2 A Regularized, Constraint-Reduced MPC Iteration

2.1 A Modified MPC Algorithm

In [34], a constraint-reduced MPC algorithm was proposed for *linear* optimization problems, as a constraint-reduced extension of a globally and locally superlinearly convergent variant of Mehrotra's original algorithm [21,36]. Transposed to the CQP context, that variant proceeds as follows.

In a first step (following the basic MPC paradigm), given $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ with $\boldsymbol{\lambda} > \mathbf{0}, \mathbf{s} > \mathbf{0}$, it computes the primal–dual *affine-scaling* direction $(\Delta\mathbf{x}^{\mathrm{a}}, \Delta\boldsymbol{\lambda}^{\mathrm{a}}, \Delta\mathbf{s}^{\mathrm{a}})$ at $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$, viz., the Newton direction for the solution of the equations portion of (1). Thus, it solves

$$J(H, A, \mathbf{s}, \boldsymbol{\lambda}) \begin{bmatrix} \Delta\mathbf{x}^{\mathrm{a}} \\ \Delta\boldsymbol{\lambda}^{\mathrm{a}} \\ \Delta\mathbf{s}^{\mathrm{a}} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}) + A^T\boldsymbol{\lambda} \\ \mathbf{0} \\ -S\boldsymbol{\lambda} \end{bmatrix}, \tag{2}$$

where, given a symmetric matrix $W \succeq \mathbf{0}$, we define

$$J(W, A, \mathbf{s}, \boldsymbol{\lambda}) := \begin{bmatrix} W & -A^T & \mathbf{0} \\ A & \mathbf{0} & -I \\ \mathbf{0} & S & \Lambda \end{bmatrix},$$

with $\Lambda = \mathrm{diag}(\boldsymbol{\lambda})$. Conditions for unique solvability of system (2) are given in the following standard result (invoked in its full form later in this paper); see, e.g., [16, Lemma B.1].

Lemma 1 *Suppose $s_i, \lambda_i \geq 0$ for all i and $W \succeq \mathbf{0}$. Then $J(W, A, \mathbf{s}, \boldsymbol{\lambda})$ is invertible if and only if the following three conditions hold:*

(i) $s_i + \lambda_i > 0$ for all i;
(ii) $A_{\{i:s_i=0\}}$ has full row rank; and
(iii) $\begin{bmatrix} W & \left(A_{\{i:\lambda_i\neq 0\}}\right)^T \end{bmatrix}$ has full row rank.

In particular, with $\boldsymbol{\lambda} > \mathbf{0}$ and $\mathbf{s} > \mathbf{0}$, $J(H, A, \mathbf{s}, \boldsymbol{\lambda})$ is invertible if and only if $[H \;\; A^T]$ has full row rank. By means of two steps of block Gaussian elimination, system (2) reduces to the *normal* system

$$\begin{aligned} M\Delta\mathbf{x}^{\mathrm{a}} &= -\nabla f(\mathbf{x}), \\ \Delta\mathbf{s}^{\mathrm{a}} &= A\Delta\mathbf{x}^{\mathrm{a}}, \\ \Delta\boldsymbol{\lambda}^{\mathrm{a}} &= -\boldsymbol{\lambda} - S^{-1}\Lambda\Delta\mathbf{s}^{\mathrm{a}}, \end{aligned} \tag{3}$$

where M is given by

$$M := H + A^T S^{-1}\Lambda A = H + \sum_{i=1}^{m} \frac{\lambda_i}{s_i}\mathbf{a}_i\mathbf{a}_i^T. \tag{4}$$

Given positive definite S and Λ, M is invertible whenever $J(H, A, \mathbf{s}, \boldsymbol{\lambda})$ is.

In a second step, MPC algorithms construct a *centering/corrector* direction, which in the CQP case (e.g., [22, Section 16.6]) is the solution $(\Delta\mathbf{x}^{\mathrm{c}}, \Delta\boldsymbol{\lambda}^{\mathrm{c}}, \Delta\mathbf{s}^{\mathrm{c}})$ to (same coefficient matrix as in (2))

$$J(H, A, \mathbf{s}, \boldsymbol{\lambda}) \begin{bmatrix} \Delta\mathbf{x}^{\mathrm{c}} \\ \Delta\boldsymbol{\lambda}^{\mathrm{c}} \\ \Delta\mathbf{s}^{\mathrm{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \sigma\mu\mathbf{1} - \Delta S^{\mathrm{a}}\Delta\boldsymbol{\lambda}^{\mathrm{a}} \end{bmatrix}, \tag{5}$$

where $\mu := \mathbf{s}^T\boldsymbol{\lambda}/m$ is the "duality measure" and $\sigma := (1-\alpha^{\mathrm{a}})^3$ is the centering parameter, with

$$\alpha^{\mathrm{a}} := \operatorname{argmax}\{\alpha \in [0,\, 1] \,|\, \mathbf{s} + \alpha\Delta\mathbf{s}^{\mathrm{a}} \geq \mathbf{0},\ \boldsymbol{\lambda} + \alpha\Delta\boldsymbol{\lambda}^{\mathrm{a}} \geq \mathbf{0}\}\,.$$

While most MPC algorithms use as search direction the sum of the affine-scaling and centering/corrector directions, to force global convergence, we borrow from [34][2] and define

$$(\Delta\mathbf{x},\, \Delta\boldsymbol{\lambda},\, \Delta\mathbf{s}) = (\Delta\mathbf{x}^{\mathrm{a}},\, \Delta\boldsymbol{\lambda}^{\mathrm{a}},\, \Delta\mathbf{s}^{\mathrm{a}}) + \gamma(\Delta\mathbf{x}^{\mathrm{c}},\, \Delta\boldsymbol{\lambda}^{\mathrm{c}},\, \Delta\mathbf{s}^{\mathrm{c}}), \tag{6}$$

where the "mixing" parameter $\gamma \in [0,\, 1]$ is one when $\Delta\mathbf{x}^{\mathrm{c}} = \mathbf{0}$ and otherwise

$$\gamma := \min\left\{\gamma_1,\, \tau\frac{\|\Delta\mathbf{x}^{\mathrm{a}}\|}{\|\Delta\mathbf{x}^{\mathrm{c}}\|},\, \tau\frac{\|\Delta\mathbf{x}^{\mathrm{a}}\|}{\sigma\mu}\right\}, \tag{7}$$

where $\tau \in [0, 1)$ and

$$\gamma_1 := \operatorname{argmax}\{\tilde{\gamma} \in [0,\, 1] \mid f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}} + \tilde{\gamma}\Delta\mathbf{x}^{\mathrm{c}}) \geq \omega(f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}}))\}\,, \tag{8}$$

with $\omega \in (0,\, 1)$. The first term in (7) guarantees that the search direction is a direction of significant descent for the objective function (which in our context is central to forcing global convergence) while the other two terms ensures that the magnitude of the centering/corrector direction is not too large compared to the magnitude of the affine-scaling direction.

As for the line search, we again borrow from [34], where specific safeguards are imposed to guarantee global and local q-quadratic convergence. We set

$$\bar{\alpha}_{\mathrm{p}} := \operatorname{argmax}\{\alpha : \mathbf{s} + \alpha\Delta\mathbf{s} \geq \mathbf{0}\}, \quad \alpha_{\mathrm{p}} := \min\{1,\, \max\{\varkappa\bar{\alpha}_{\mathrm{p}},\, \bar{\alpha}_{\mathrm{p}} - \|\Delta\mathbf{x}\|\}\},$$
$$\bar{\alpha}_{\mathrm{d}} := \operatorname{argmax}\{\alpha : \boldsymbol{\lambda} + \alpha\Delta\boldsymbol{\lambda} \geq \mathbf{0}\}, \quad \alpha_{\mathrm{d}} := \min\{1,\, \max\{\varkappa\bar{\alpha}_{\mathrm{d}},\, \bar{\alpha}_{\mathrm{d}} - \|\Delta\mathbf{x}\|\}\}\,,$$

with $\varkappa \in (0, 1)$, then

$$(\mathbf{x}^+, \mathbf{s}^+) := (\mathbf{x}, \mathbf{s}) + \alpha_{\mathrm{p}}(\Delta\mathbf{x}, \Delta\mathbf{s})\,.$$

and finally

$$\lambda_i^+ := \min\{\lambda^{\max},\, \max\{\lambda_i + \alpha_{\mathrm{d}}\Delta\lambda_i,\, \min\{\underline{\lambda},\, \chi\}\}\},\, i = 1, \ldots, m\,, \tag{9}$$

where $\lambda^{\max} > 0$ and $\underline{\lambda} \in (0, \lambda^{\max})$ are algorithm parameters, and

$$\chi := \|\Delta\mathbf{x}^{\mathrm{a}}\|^{\nu} + \|[\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}^{\mathrm{a}}]_-\|^{\nu}\,,$$

with $\nu \geq 2$.

We verified via numerical tests that for the problems considered in Section 3, the modified MPC algorithm outlined in this section is at least as efficient as the MPC algorithm for CQPs given in [22, Algorithm 16.4].

[2] We however do not fully follow [34]: (i) Equation (8) generalizes (22) of [34] to CQP; (ii) In (9) we explicitly bound $\boldsymbol{\lambda}^+$ (x^+ in [34]), by $\lambda^{\max}$; in the linear case, such boundedness is guaranteed (Lemma 3.3 in [34]); as a side-effect, in (7), we could drop the penultimate term in (24) of [34] (invoked in proving convergence of the x sequence in the proof of Lemma 3.4 of [34]); (iii) We do not restrict the primal step size as done in (25) of [34] (dual step size in the context of [34]), at the expense of a slightly more involved convergence proof: see our Proposition 3 below, to be compared to [34, Lemma 3.7].

2.2 A Regularized Constraint-Reduced MPC Algorithm

In the modified MPC algorithm described in Section 2.1, the main computational cost is incurred in forming the normal matrix M (see (4)), which requires approximately $mn^2/2$ multiplications (at each iteration) if A is dense, regardless of how many of the m inequality constraints in (P) are active at the solution. This may be wasteful when few of these constraints are active at the solution, in particular (generically) when $m \gg n$ (imbalanced problems). The constraint-reduction mechanism introduced in [28] and used in [18] in the context of an affine-scaling algorithm for the solution of CQPs modifies M by limiting the sum in (4) to a wisely selected small subset of terms, indexed by an index set $Q \subseteq \mathbf{m}$ referred to as the *working set*.

Given a working set Q, the constraint-reduction technique produces an *approximate* affine-scaling direction by solving a "reduced" version of the Newton system (2), viz.

$$J(H, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q) \begin{bmatrix} \Delta\mathbf{x}^{\mathrm{a}} \\ \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}} \\ \Delta\mathbf{s}_Q^{\mathrm{a}} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}) + (A_Q)^T \boldsymbol{\lambda}_Q \\ \mathbf{0} \\ -S_Q \boldsymbol{\lambda}_Q \end{bmatrix} . \tag{10}$$

Just like the full system, when $\mathbf{s}_Q > \mathbf{0}$, the reduced system (10) is equivalent to the reduced normal system

$$\begin{aligned} \tilde{M}_{(Q)} \Delta\mathbf{x}^{\mathrm{a}} &= -\nabla f(\mathbf{x}) , \\ \Delta\mathbf{s}_Q^{\mathrm{a}} &= A_Q \Delta\mathbf{x}^{\mathrm{a}} , \\ \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}} &= -\boldsymbol{\lambda}_Q - S_Q^{-1} \Lambda_Q \Delta\mathbf{s}_Q^{\mathrm{a}} , \end{aligned} \tag{11}$$

where the "reduced" $\tilde{M}_{(Q)}$ (still of size $n \times n$) is given by

$$\tilde{M}_{(Q)} := H + (A_Q)^T S_Q^{-1} \Lambda_Q A_Q = H + \sum_{i \in Q} \frac{\lambda_i}{s_i} \mathbf{a}_i \mathbf{a}_i^T .$$

When A is dense, the cost of forming $\tilde{M}_{(Q)}$ is approximately $qn^2/2$, where $q := |Q|$, leading to significant savings when $q \ll m$.

One difficulty that may arise, when substituting A_Q for A in the Newton-KKT matrix, is that the resulting linear system might no longer be uniquely solvable. Indeed, even when $[H \;\; A^T]$ has full row rank, $[H \;\; (A_Q)^T]$ may be rank-deficient, so the third condition in Lemma 1 would not hold. A possible remedy is to regularize the linear system. In the context of linear optimization, such regularization was implemented in [9] and explored in [25] by effectively adding a fixed scalar multiple of identity matrix into the normal matrix to improve numerical stability of the Cholesky factorization. A more general regularization was proposed in [1] where diagonal matrices that are adjusted dynamically based on the pivot values in the Cholesky factorization were used for regularization. On the other hand, quadratic regularization was applied to obtain better preconditioners in [4], where a hybrid scheme of the Cholesky factorization and a preconditioned conjugate gradient method is used to solve linear systems arising in primal block-angular problems. In [4], the regularization dies out when optimality is approached.

Applying regularization to address rank-deficiency of the normal matrix due to constraint reduction was first considered in [35], in the context of linear optimization. There a similar regularization as in [9, 25] is applied, while the scheme lets the regularization die out as a solution to the optimization problem is approached, to preserve fast local convergence. Adapting such approach to the present context, we replace $J(H, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$ by $J(W, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$ and $\tilde{M}_{(Q)}$ by

$$M_{(Q)} := W + (A_Q)^T S_Q^{-1} \Lambda_Q A_Q \,, \tag{12}$$

with $W := H + \varrho R$, where $\varrho \in (0, 1]$ is a regularization parameter that is updated at each iteration and $R \succeq \mathbf{0}$ a constant symmetric matrix such that $H + R \succ \mathbf{0}$. Thus the inequality $W \succeq H$ is enforced, ensuring $f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}}) < f(\mathbf{x})$ (see Proposition 2 below), which in our context is critical for global convergence. In the proposed algorithm, the modified coefficient matrix is used in the computation of both a modified affine-scaling direction and a modified centering/corrector direction, which thus solves

$$J(W, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q) \begin{bmatrix} \Delta\mathbf{x}^{\mathrm{c}} \\ \Delta\boldsymbol{\lambda}_Q^{\mathrm{c}} \\ \Delta\mathbf{s}_Q^{\mathrm{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \sigma\mu_{(Q)}\mathbf{1} - \Delta S_Q^{\mathrm{a}} \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}} \end{bmatrix} . \tag{13}$$

In the bottom block of the right-hand side of (13) (compared to (5)) we have substituted ΔS_Q^{a} and $\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}$ for ΔS^{a} and $\Delta\boldsymbol{\lambda}^{\mathrm{a}}$, and replaced μ with

$$\mu_{(Q)} := \begin{cases} \mathbf{s}_Q^T \boldsymbol{\lambda}_Q / q \,, & \text{if } q \neq 0 \\ 0 \,, & \text{otherwise} \end{cases} , \tag{14}$$

the duality measure for the "reduced" problem. The corresponding normal equation system is given by

$$\begin{aligned} M_{(Q)} \Delta\mathbf{x}^{\mathrm{c}} &= (A_Q)^T S_Q^{-1} (\sigma\mu_{(Q)}\mathbf{1} - \Delta S_Q^{\mathrm{a}} \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}), \\ \Delta\mathbf{s}_Q^{\mathrm{c}} &= A_Q \Delta\mathbf{x}^{\mathrm{c}}, \\ \Delta\boldsymbol{\lambda}_Q^{\mathrm{c}} &= S_Q^{-1} (-\Lambda_Q \Delta\mathbf{s}_Q^{\mathrm{c}} + \sigma\mu_{(Q)}\mathbf{1} - \Delta S_Q^{\mathrm{a}} \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}) \,. \end{aligned} \tag{15}$$

A partial search direction for the constraint-reduced MPC algorithm at $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ is then given by (see (6))

$$(\Delta\mathbf{x}, \Delta\boldsymbol{\lambda}_Q, \Delta\mathbf{s}_Q) = (\Delta\mathbf{x}^{\mathrm{a}}, \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}, \Delta\mathbf{s}_Q^{\mathrm{a}}) + \gamma(\Delta\mathbf{x}^{\mathrm{c}}, \Delta\boldsymbol{\lambda}_Q^{\mathrm{c}}, \Delta\mathbf{s}_Q^{\mathrm{c}}), \tag{16}$$

where γ is given by (7)–(8), with $\mu_{(Q)}$ replacing μ.[3]

Algorithm CR-MPC, including a stopping criterion, a simple update rule for ϱ, and update rules (adapted from [34]) for the components λ_i of the dual variable with $i \in Q^{\mathrm{c}}$, but with the constraint-selection rule (in Step 2)

[3] In the case that $q = 0$ (Q is empty), γ is chosen to be zero. Note that, in such case, there is no corrector direction, as the right-hand side of (13) vanishes.

left unspecified, is formally stated below.[4] Its core, Iteration CR-MPC, takes as input the current iterates $\mathbf{x}$, $\mathbf{s} > 0$, $\boldsymbol{\lambda} > 0$, $\tilde{\boldsymbol{\lambda}}$, and produces the next iterates $\mathbf{x}^+$, $\mathbf{s}^+ > 0$, $\boldsymbol{\lambda}^+ > 0$, $\tilde{\boldsymbol{\lambda}}^+$, used as input to the next iteration. Here $\tilde{\boldsymbol{\lambda}}$, with possibly $\tilde{\boldsymbol{\lambda}} \not\geq 0$, is asymptotically slightly closer to optimality than $\boldsymbol{\lambda}$, and is used in the stopping criterion. While dual feasibility of $(\mathbf{x}, \boldsymbol{\lambda})$ is not enforced along the sequence of iterates, a primal strictly feasible starting point $\mathbf{x} \in \mathcal{F}_P^o$ is required, and primal feasibility of subsequent iterates is enforced, as it allows for monotone descent of f, which in the present context is key to global convergence. (An extension of Algorithm CR-MPC that allows for infeasible starting points is discussed in Section 2.3 below.) Algorithm CR-MPC makes use (in its stopping criterion and ϱ update) of an "error" function $E\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ (also used in the constraint-selection Rule R in Section 2.6 below) given by

$$E(\mathbf{x}, \boldsymbol{\lambda}) := \|(\|\mathbf{v}(\mathbf{x}, \boldsymbol{\lambda})\|, \|\mathbf{w}(\mathbf{x}, \boldsymbol{\lambda})\|)\| \ , \tag{17}$$

where

$$\mathbf{v}(\mathbf{x}, \boldsymbol{\lambda}) := H\mathbf{x} + \mathbf{c} - A^T\boldsymbol{\lambda}, \quad w_i(\mathbf{x}, \boldsymbol{\lambda}) := \min\{|s_i|, |\lambda_i|\},\ i = 1, \ldots, m, \tag{18}$$

with $\mathbf{s} := A\mathbf{x} - \mathbf{b}$, and where the norms are arbitrary. Here E measures both dual feasibility (via $\mathbf{v}$) and complementary slackness (via $\mathbf{w}$). Note that, for $\mathbf{x} \in \mathcal{F}_P$ and $\boldsymbol{\lambda} \geq \mathbf{0}$, $E(\mathbf{x}, \boldsymbol{\lambda}) = 0$ if and only if $(\mathbf{x}, \boldsymbol{\lambda})$ solves (P)–(D).

Algorithm CR-MPC: A Constraint-Reduced variant of MPC Algorithm for CQP

Parameters: $\varepsilon \geq 0$, $\tau \in [0, 1)$, $\omega \in (0, 1)$, $\varkappa \in (0, 1)$, $\nu \geq 2$, $\lambda^{\max} > 0$, $\underline{\lambda} \in (0, \lambda^{\max})$, and $\bar{E} > 0$.[5] A symmetric $n \times n$ matrix $R \succeq \mathbf{0}$ such that $H + R \succ \mathbf{0}$.

Initialization: $\mathbf{x} \in \mathcal{F}_P^o$,[6] $\boldsymbol{\lambda} > \mathbf{0}$, $\mathbf{s} := A\mathbf{x} - \mathbf{b} > \mathbf{0}$, $\tilde{\boldsymbol{\lambda}} := \boldsymbol{\lambda}$.

Iteration CR-MPC:

Step 1. Terminate if either (i) $\nabla f(\mathbf{x}) = \mathbf{0}$, in which case $(\mathbf{x}, \mathbf{0})$ is optimal for (P)–(D), or (ii)

$$\min\left\{E(\mathbf{x}, \boldsymbol{\lambda}), E(\mathbf{x}, [\tilde{\boldsymbol{\lambda}}]_+)\right\} < \varepsilon, \tag{19}$$

in which case $(\mathbf{x}, [\tilde{\boldsymbol{\lambda}}]_+)$ is declared ε-optimal for (P)–(D) if $E(\mathbf{x}, \boldsymbol{\lambda}) \geq E(\mathbf{x}, [\tilde{\boldsymbol{\lambda}}]_+)$, and $(\mathbf{x}, \boldsymbol{\lambda})$ is otherwise.

Step 2. Select a working set Q. Set $q := |Q|$. Set $\varrho := \min\{1, \frac{E(\mathbf{x}, \boldsymbol{\lambda})}{\bar{E}}\}$. Set $W := H + \varrho R$.

Step 3. Compute approximate normal matrix $M_{(Q)} := W + \sum_{i \in Q} \frac{\lambda_i}{s_i} \mathbf{a}_i \mathbf{a}_i^T$.

Step 4. Solve

$$M_{(Q)} \Delta\mathbf{x}^{\mathrm{a}} = -\nabla f(\mathbf{x})\,, \tag{20}$$

and set

$$\Delta\mathbf{s}^{\mathrm{a}} := A\Delta\mathbf{x}^{\mathrm{a}}\,, \quad \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}} := -\boldsymbol{\lambda}_Q - S_Q^{-1} \Lambda_Q \Delta\mathbf{s}_Q^{\mathrm{a}}\,. \tag{21}$$

Step 5. Compute the affine-scaling step

$$\alpha^{\mathrm{a}} := \operatorname{argmax}\{\alpha \in [0, 1] \,|\, \mathbf{s} + \alpha\Delta\mathbf{s}^{\mathrm{a}} \geq \mathbf{0},\ \boldsymbol{\lambda}_Q + \alpha\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}} \geq \mathbf{0}\}\,. \tag{22}$$

---

[4] The "modified MPC algorithm" outlined in Section 2.1 is recovered as a special case by setting $\varrho^+ = 0$ and $Q = \{1, \ldots, m\}$ in Step 2 of Algorithm CR-MPC.

[5] For scaling reasons, it may be advisable to set the value of $\bar{E}$ to the initial value of $E(\mathbf{x}, \boldsymbol{\lambda})$ (so that, in Step 2 of the initial iteration, ϱ is set to 1, and W to $H + R$). This was done in the numerical tests reported in Section 3.

[6] Here it is implicitly assumed that $\mathcal{F}_P^o$ is nonempty. This assumption is subsumed by Assumption 1 below.

Step 6. Set $\mu_{(Q)}$ as in (14). Compute centering parameter $\sigma := (1-\alpha^{\rm a})^3$.
Step 7. Solve (13) for the corrector direction $(\Delta\mathbf{x}^{\rm c},\, \Delta\boldsymbol{\lambda}_Q^{\rm c},\, \Delta\mathbf{s}_Q^{\rm c})$, and set $\Delta\mathbf{s}^{\rm c} := A\Delta\mathbf{x}^{\rm c}$.
Step 8. If $q = 0$, set $\gamma := 0$. Otherwise, compute γ as in (7)–(8), with $\mu_{(Q)}$ replacing μ. Compute the search direction

$$(\Delta\mathbf{x},\, \Delta\boldsymbol{\lambda}_Q,\, \Delta\mathbf{s}) := (\Delta\mathbf{x}^{\rm a},\, \Delta\boldsymbol{\lambda}_Q^{\rm a},\, \Delta\mathbf{s}^{\rm a}) + \gamma(\Delta\mathbf{x}^{\rm c},\, \Delta\boldsymbol{\lambda}_Q^{\rm c},\, \Delta\mathbf{s}^{\rm c})\,. \tag{23}$$

Set $\tilde{\lambda}_i^+ = \lambda_i + \Delta\lambda_i$, $\forall i \in Q$, and $\tilde{\lambda}_i^+ = 0$, $\forall i \in Q^{\rm c}$.
Step 9. Compute the primal and dual steps $\alpha_{\rm p}$ and $\alpha_{\rm d}$ by

$$\begin{aligned} \bar{\alpha}_{\rm p} &:= \operatorname{argmax}\{\alpha : \mathbf{s} + \alpha\Delta\mathbf{s} \geq \mathbf{0}\}, & \alpha_{\rm p} &:= \min\{1,\, \max\{\varkappa\bar{\alpha}_{\rm p},\, \bar{\alpha}_{\rm p} - \|\Delta\mathbf{x}\|\}\}, \\ \bar{\alpha}_{\rm d} &:= \operatorname{argmax}\{\alpha : \boldsymbol{\lambda}_Q + \alpha\Delta\boldsymbol{\lambda}_Q \geq \mathbf{0}\}, & \alpha_{\rm d} &:= \min\{1,\, \max\{\varkappa\bar{\alpha}_{\rm d},\, \bar{\alpha}_{\rm d} - \|\Delta\mathbf{x}\|\}\}\,. \end{aligned} \tag{24}$$

Step 10. Updates:

$$(\mathbf{x}^+, \mathbf{s}^+) := (\mathbf{x}, \mathbf{s}) + (\alpha_{\rm p}\Delta\mathbf{x}, \alpha_{\rm p}\Delta\mathbf{s})\,. \tag{25}$$

Set $\chi := \|\Delta\mathbf{x}^{\rm a}\|^\nu + \|[\boldsymbol{\lambda}_Q + \Delta\boldsymbol{\lambda}_Q^{\rm a}]_-\|^\nu$. Set

$$\lambda_i^+ := \max\{\min\{\lambda_i + \alpha_{\rm d}\Delta\lambda_i,\, \lambda^{\max}\},\, \min\{\chi,\, \underline{\lambda}\}\},\ \forall i \in Q\,. \tag{26}$$

Set $\mu_{(Q)}^+ := (\mathbf{s}_Q^+)^T(\boldsymbol{\lambda}_Q^+)/q$ if $q \neq 0$, otherwise set $\mu_{(Q)}^+ := 0$. Set

$$\lambda_i^+ := \max\{\min\{\mu_{(Q)}^+/s_i^+,\, \lambda^{\max}\},\, \min\{\chi,\, \underline{\lambda}\}\},\ \forall i \in Q^{\rm c}\,. \tag{27}$$

A few more comments are in order concerning Algorithm CR-MPC. First, the stopping criterion is a variation on that of [18, 34], involving both $\boldsymbol{\lambda}$ and $[\tilde{\boldsymbol{\lambda}}]_+$ instead of only $\boldsymbol{\lambda}$; in fact the latter will fail when the parameter $\lambda^{\max}$ (see (26)–(27)) is not large enough and may fail when second order sufficient conditions are not satisfied, while we prove below (Theorem 1(iv)) that the new criterion is eventually satisfied indeed, in that the iterate $(\mathbf{x}, \boldsymbol{\lambda})$ converges to a solution (even if it is not unique), be it on a mere subsequence. Second, our update formula for the regularization parameter ϱ in Step 2 improves on that in [35] ($\varrho^+ = \min\{\chi, \chi_{\max}\}$ in the notation of this paper, where $\chi_{\max}$ is a user-defined constant) as it fosters a "smooth" evolution of W from the initial value of $H + R$, with R specified by the user, at a rate no faster than that required for local q-quadratic convergence. And third, $R \succeq 0$ should be selected to compensate for possible ill-conditioning of H—so as to mitigate possible early ill-conditioning of $M_{(Q)}$. (Note that a nonzero R may be beneficial even when H is non-singular.)

It is readily established that, starting from a strictly feasible point, regardless of the choice made for Q in Step 2, Algorithm CR-MPC either stops at Step 1 after finitely many iterations, or generates infinite sequences $\{E_k\}_{k=0}^\infty$, $\{\mathbf{x}^k\}_{k=0}^\infty$, $\{\boldsymbol{\lambda}^k\}_{k=0}^\infty$, $\{\tilde{\boldsymbol{\lambda}}^k\}_{k=0}^\infty$, $\{\mathbf{s}^k\}_{k=0}^\infty$, $\{\chi_k\}_{k=0}^\infty$, $\{Q_k\}_{k=0}^\infty$, $\{\varrho_k\}_{k=0}^\infty$, and $\{W_k\}_{k=0}^\infty$, with $\mathbf{s}^k = A\mathbf{x}^k - \mathbf{b} > \mathbf{0}$ and $\boldsymbol{\lambda}^k > \mathbf{0}$ for all k. (E_0, χ_0, ϱ_0, and W_0 correspond to the values of E_k, χ_k, ϱ_k, and W_k computed in the initial iteration, while the other initial values are provided in the "Initialization" step.) Indeed, if the algorithm does not terminate at Step 1, then $\nabla f(\mathbf{x}) \neq \mathbf{0}$, i.e., $\Delta\mathbf{x}^{\rm a} \neq \mathbf{0}$ (from (20), since $M_{(Q)}$ is invertible); it follows that $\Delta\mathbf{x} \neq \mathbf{0}$ (if $\Delta\mathbf{x}^{\rm c} \neq \mathbf{0}$, since $\tau \in [0, 1)$, (7) yields $\gamma < \|\Delta\mathbf{x}^{\rm a}\|/\|\Delta\mathbf{x}^{\rm c}\|$) and, since $\Delta\mathbf{x}^{\rm a} \neq \mathbf{0}$ implies $\chi > 0$, (24), (23), (25), (26), and (27) imply that $\mathbf{s}^+ = A\mathbf{x}^+ - \mathbf{b} > \mathbf{0}$ and $\boldsymbol{\lambda}^+ > \mathbf{0}$. From now on, we assume that infinite sequences are generated.

2.3 Extensions: Infeasible Starting Point, Equality Constraints

Because, in our constraint-reduction context, convergence is achieved by enforcing descent of the objective function at every iteration, infeasible starts cannot be accommodated as, e.g., in S. Mehrotra's original paper [21]. The penalty-function approach proposed and analyzed in [13, Chapter 3] in the context of constraint-reduced affine scaling for CQP (adapted from a scheme introduced in [27] for a nonlinear optimization context) fits right in however. (Also see [14] for the linear optimization case.) Translated to the notation of the present paper, it substitutes for (P)–(D) the primal-dual pair[7]

$$\underset{\mathbf{x}\in\mathbb{R}^n,\,\mathbf{z}\in\mathbb{R}^m}{\text{minimize}}\ f(\mathbf{x})+\varphi\mathbf{1}^T\mathbf{z}\quad\text{s.t. } A\mathbf{x}+\mathbf{z}\geq\mathbf{b}\,,\mathbf{z}\geq\mathbf{0}\,, \tag{P$_\varphi$}$$

$$\underset{\mathbf{x}\in\mathbb{R}^n,\,\boldsymbol{\lambda}\in\mathbb{R}^m}{\text{maximize}}\ -\frac{1}{2}\mathbf{x}^TH\mathbf{x}+\mathbf{b}^T\boldsymbol{\lambda}\text{ s.t. } H\mathbf{x}+\mathbf{c}-A^T\boldsymbol{\lambda}=\mathbf{0}\,,\ \mathbf{0}\leq\boldsymbol{\lambda}\leq\mathbf{1}\,, \tag{D$_\varphi$}$$

with $\varphi > 0$ a scalar penalty parameter, for which primal-strictly-feasible points $(\mathbf{x}, \mathbf{z})$ are readily available. Hence, given φ, this problem can be handled by Algorithm CR-MPC.[8] Such ℓ_1 penalty function is known to be exact, i.e., for some unknown, sufficiently large (but still moderate) value of φ, solutions $(\mathbf{x}^*, \mathbf{z}^*)$ to (P$_\varphi$) are such that $\mathbf{x}^*$ solves (P); further ([13,14]), $\mathbf{z}^* = \mathbf{0}$. In [13, 14], an adaptive scheme is proposed for increasing φ to such value. Applying this scheme on (P$_\varphi$)–(D$_\varphi$) allows Algorithm CR-MPC to handle infeasible starting points for (P)–(D). We refer the reader to [13, Chapter 3] for details.

Linear equality constraints of course can be handled by first projecting the problem on the associated affine space, and then run Algorithm CR-MPC on that affine space. A weakness of this approach though is that it does not adequately extend to the case of sparse problems (discussed in the Conclusion section (Section 4) of this paper), as projection may destroy sparsity. An alternative approach is, again, via augmentation: Given the constraints $C\mathbf{x} = \mathbf{d}$, with $\mathbf{d} \in \mathbb{R}^p$, solve the problem

$$\underset{\mathbf{x}\in\mathbb{R}^n,\,\mathbf{y}\in\mathbb{R}^p}{\text{minimize}}\ f(\mathbf{x})+\varphi\mathbf{1}^T\mathbf{y}\quad\text{s.t. } C\mathbf{x}+\mathbf{y}\geq\mathbf{d}\,,\ C\mathbf{x}-\mathbf{y}\leq\mathbf{d}\,. \tag{28}$$

(Note that, taken together, the two constraints imply $\mathbf{y} \geq \mathbf{0}$.) Again, given $\varphi > 0$ and using the same adaptive scheme from [13, 14], this problem can be tackled by Algorithm CR-MPC.

2.4 A Class of Constraint-Selection Rules

Of course, the quality of the search directions is highly dependent on the choice of the working set Q. Several constraint-selection rules have been proposed

[7] An ℓ_∞ penalty function can be substituted for this ℓ_1 penalty function with minor adjustments: see [13, 14] for details.

[8] It is readily checked that, given the simple form in which $\mathbf{z}$ enters the constraints, for dense problems, the cost of forming $M^{(Q)}$ still dominates and is still approximately $|Q|n^2/2$, with still, typically, $|Q| \ll m$.

for constraint-reduced algorithms on various classes of optimization problems, such as linear optimization [28, 34, 35], convex quadratic optimization [17, 18], semi-definite optimization [23,24], and nonlinear optimization [5]. In [28,34,35], the cardinality q of Q is constant and decided at the outset. Because in the non-degenerate case the set of active constraints at the solution of (P) with $H = \mathbf{0}$ is at least equal to the number n of primal variables, $q \geq n$ is usually enforced in that context. In [18], which like this paper deals with quadratic problems, q was allowed to vary from iteration from iteration, but $q \geq n$ was still enforced throughout (owing to the fact that, in the regular case, there are no more than n active constraints at the solution). Here we propose to again allow q to vary, but in addition to not *a priori* impose a positive lower bound on q.

The convergence results stated in Section 2.5 below are valid with any constraint-selection rule that satisfies the following condition.

Condition CSR *Let $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ be the sequence constructed by Algorithm CR-MPC with the constraint-selection rule under consideration, and let Q_k be the working set generated by the constraint-selection rule at iteration k. Then the following holds: (i) if $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ is bounded away from $\mathcal{F}^*$, then, for all limit points $\mathbf{x}'$ such that $\{\mathbf{x}^k\} \to \mathbf{x}'$ on some infinite index set K, $\mathcal{A}(\mathbf{x}') \subseteq Q_k$ for all large enough $k \in K$; and (ii) if* (P)–(D) *has a unique solution $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ and strict complementarity holds at $\mathbf{x}^*$, and if $\{\mathbf{x}^k\} \to \mathbf{x}^*$, then $\mathcal{A}(\mathbf{x}^*) \subseteq Q_k$ for all k large enough.*

Condition CSR(i) aims at preventing convergence to non-optimal primal point, and hence (given a bounded sequence of iterates) forcing convergence to solution points. Condition CSR(ii) is important for fast local convergence to set in. A specific rule that satisfies Condition CSR, Rule R, used in our numerical experiments, is presented in Section 2.6 below.

2.5 Convergence Properties

The following standard assumptions are used in portions of the analysis.

Assumption 1 *$\mathcal{F}_P^o$ is nonempty and $\mathcal{F}_P^*$ is nonempty and bounded.* [9]

Assumption 2 [10] *At every stationary point $\mathbf{x}$, $A_{\mathcal{A}(\mathbf{x})}$ has full row rank.*

Assumption 3 *There exists a (unique) $\mathbf{x}^*$ where the second-order sufficient condition of optimality with strict complementarity holds, with (unique) $\boldsymbol{\lambda}^*$.*

---

[9] Nonemptiness and boundedness of $\mathcal{F}_P^*$ are equivalent to dual strict feasibility (e.g., [7, Theorem 2.1]).

[10] Equivalently (under the sole assumption that $\mathcal{F}_P^*$ is nonempty) $A_{\mathcal{A}(\mathbf{x})}$ has full row rank at all $\mathbf{x} \in \mathcal{F}_P$. In fact, while we were not able to carry out the analysis without such (strong) assumption (the difficulty being to rule out convergence to non-optimal stationary points), numerical experimentation suggests that the assumption is immaterial.

Assumption 3, mostly used in the local analysis, subsumes Assumption 1.

Theorem 1, proved in Appendix A, addresses global convergence.

Theorem 1 *Suppose that the constraint-selection rule invoked in Step 2 of Algorithm CR-MPC satisfies Condition CSR. First suppose that* $\varepsilon = 0$*, that the iteration never stops, and that Assumptions 1 and 2 hold. Then (i) the infinite sequence* $\{\mathbf{x}^k\}$ *it constructs converges to the primal solution set* $\mathcal{F}_P^*$*; if in addition, Assumption 3 holds, then (ii)* $\{(\mathbf{x}^k, \tilde{\boldsymbol{\lambda}}^k)\}$ *converges to the unique primal–dual solution* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *and* $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ *converges to* $(\mathbf{x}^*, \boldsymbol{\xi}^*)$*, with* $\xi_i^* := \min\{\lambda_i, \lambda^{\max}\}$ *for all* $i \in \mathbf{m}$*, and (iii) for sufficiently large* k*, the working set* Q_k *contains* $\mathcal{A}(\mathbf{x}^*)$*.*

Finally, suppose again that Assumptions 1 and 2 hold. Then (iv) if $\varepsilon > 0$*, Algorithm CR-MPC stops (in Step 1) after finitely many iterations.*

Fast local convergence is addressed next; Theorem 2 is proved in Appendix B.

Theorem 2 *Suppose that Assumption 3 holds, that* $\varepsilon = 0$*, that the iteration never stops, and that* $\lambda_i^* < \lambda^{\max}$ *for all* $i \in \mathbf{m}$*. Then there exist* $\rho > 0$ *and* $C > 0$ *such that, if* $\|(\mathbf{x} - \mathbf{x}^*, \boldsymbol{\lambda} - \boldsymbol{\lambda}^*)\| < \rho$ *and* $Q \supseteq \mathcal{A}(\mathbf{x}^*)$*, then*

$$\|(\mathbf{x}^+ - \mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^+ - \boldsymbol{\lambda}^*)\| \leq C\|(\mathbf{x} - \mathbf{x}^*, \boldsymbol{\lambda} - \boldsymbol{\lambda}^*)\|^2. \tag{29}$$

When the constraint-selection rule satisfies Condition CSR(ii), local q-quadratic convergence is an immediate consequence of Theorems 1 and 2.

Corollary 1 *Suppose that Assumptions 1–3 hold, that* $\varepsilon = 0$*, that the iteration never stops, and that* $\lambda_i^* < \lambda^{\max}$ *for all* $i \in \mathbf{m}$*. Further suppose that the constraint-selection rule invoked in Step 2 satisfies Condition CSR. Then Algorithm CR-MPC is locally q-quadratically convergent. Specifically, there exists* $C > 0$ *such that, given any initial point* $(\mathbf{x}^0, \boldsymbol{\lambda}^0)$*, for some* $k' > 0$*,*

$$\mathcal{A}(\mathbf{x}^*) \subseteq Q_k \quad \textit{and} \quad \|(\mathbf{x}^{k+1} - \mathbf{x}^*, \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*)\| \leq C\|(\mathbf{x}^k - \mathbf{x}^*, \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)\|^2, \quad \forall k > k'.$$

The foregoing theorems and corollary (essentially) extend to the case of infeasible starting point discussed in Section 2.3. The proof follows the lines of that in [13, Theorem 3.2]. While Assumptions 1 and 3 remain unchanged, Assumption 2 must be tightened to: For every $\mathbf{x} \in \mathbb{R}^n$, $\{\mathbf{a}_i : \mathbf{a}_i^T\mathbf{x} \leq b_i\}$ is a linearly independent set.[11] (While this assumption appears to be rather restrictive—a milder condition is used in [13, Theorem 3.2] and [14], but we believe it is insufficient—footnote 10 applies here as well.)

Subject to such tightening of Assumption 2, Theorem 1 still holds. Further, Theorem 2 and Corollary 1 (local quadratic convergence) also hold, but for the *augmented* set of primal–dual variables, $(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda})$. While proving the results for $(\mathbf{x}, \boldsymbol{\lambda})$ might turn out to be possible, an immediate consequence of q-quadratic convergence for $(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda})$ is *r-quadratic* convergence for $(\mathbf{x}, \boldsymbol{\lambda})$.

[11] In fact, given any known upper bound $\overline{z}$ to $\{\mathbf{z}^k\}$, this assumption can be relaxed to merely requiring linear independence of the set $\{\mathbf{a}_i : b_i - \overline{z} \leq \mathbf{a}_i^T\mathbf{x} \leq b_i\}$, which tends to the set of active constraints when $\overline{z}$ goes to zero. This can be done, e.g., with $\overline{z} = c\|\mathbf{z}^0\|_\infty$, with any $c > 1$, if the constraint $\mathbf{z} \leq \overline{z}\mathbf{1}$ is added to the augmented problem.

Under the same assumptions, Theorems 1 and 2 and Corollary 1 still hold in the presence of equality constraints $C\mathbf{x} = \mathbf{d}$ via transforming the problem to (28), provided $\{\mathbf{a}_i : \mathbf{a}_i^T\mathbf{x} \leq b_i\} \cup \{\mathbf{c}_i : i = 1, \ldots, p\}$ is a linearly independent set for every $\mathbf{x} \in \mathbb{R}^n$, with $\mathbf{c}_i$ the ith row of C. Note that it may be beneficial to choose $\mathbf{x}^0$ to lie on the affine space defined by $C\mathbf{x} = \mathbf{d}$, in which case the components of $\mathbf{y}^0$ can be chosen quite small, and to include the constraint $\mathbf{y} \leq c\mathbf{y}^0$ for some $c > 1$ as suggested in footnote 11.

2.6 A New Constraint-Selection Rule

The proposed Rule R, stated below, first computes a threshold value based on the amount of decrease of the error $E_k := E(\mathbf{x}^k, \boldsymbol{\lambda}^k)$, and then selects the working set by including all constraints with slack values less than the computed threshold.

Rule R Proposed Constraint-Selection Rule

Parameters: $\bar{\delta} > 0$, $0 < \beta < \theta < 1$.
Input: Iteration: k, Slack variable: $\mathbf{s}^k$, Error: $E_{\min}$ (when $k > 0$), $E_k := E(\mathbf{x}^k, \boldsymbol{\lambda}^k)$, Threshold: δ_{k-1}.
Output: Working set: Q_k, Threshold: δ_k, Error: $E_{\min}$.

```
if k = 0 then
    δ_k := δ̄
    E_min := E_k
else if E_k ≤ βE_min then
    δ_k := θδ_{k−1}
    E_min := E_k
else
    δ_k := δ_{k−1}
end if
Select Q_k := {i ∈ m | s_i^k ≤ δ_k}.
```

A property of E that plays a key role in proving that Rule R satisfies Condition CSR is stated next; it does not require strict complementarity. It was established in [8], within the proof of Theorem 3.12 (equation (3.13)); also see [12, Theorem 1], [37, Theorem A.1], as well as [3, Lemma 2, with the "vector of perturbations" set to zero] for an equivalent, yet global inequality in the case of linear optimization ($H = \mathbf{0}$), under an additional dual (primal in the context of [3]) feasibility assumption ($(H\mathbf{x}) - A^T\boldsymbol{\lambda} + \mathbf{c} = \mathbf{0}$). A self-contained proof in the case of CQP is given here for the sake of completeness and ease of reference.

Lemma 2 *Suppose* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *solves (P)–(D), let* $\mathcal{I} := \{i \in \mathbf{m} : \lambda_i^* > 0\}$*, and suppose that (i)* $A_{\mathcal{A}(\mathbf{x}^*)}$ *and (ii)* $[H\ (A_{\mathcal{I}})^T]$ *have full row rank. Then there exists* $c > 0$ *and some neighborhood* V *of the origin such that*

$$E(\mathbf{x}, \boldsymbol{\lambda}) \geq c\|(\mathbf{x} - \mathbf{x}^*, \boldsymbol{\lambda} - \boldsymbol{\lambda}^*)\| \quad \text{whenever} \quad (\mathbf{x} - \mathbf{x}^*, \boldsymbol{\lambda} - \boldsymbol{\lambda}^*) \in V.$$

Proof Let $\mathbf{z}^* := (\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathbb{R}^{n+m}$, let $\mathbf{s} := A\mathbf{x} - \mathbf{b}$, $\mathbf{s}^* := A\mathbf{x}^* - \mathbf{b}$, and let $\Psi\colon \mathbb{R}^{n+m} \to \mathbb{R}$ be given by $\Psi(\boldsymbol{\zeta}) := E(\mathbf{z}^* + \boldsymbol{\zeta})$. We show that, restricted to an appropriate punctured convex neighborhood of the origin, Ψ is strictly positive and absolutely homogeneous, so that the convex hull $\hat{\Psi}$ of such restriction generates a norm on $\mathbb{R}^{n+m}$, proving the claim (with $c = 1$ for the norm generated by $\hat{\Psi}$). To proceed, let $\boldsymbol{\zeta}^x \in \mathbb{R}^n$ and $\boldsymbol{\zeta}^\lambda \in \mathbb{R}^m$ denote respectively the first n and last m components of $\boldsymbol{\zeta}$, and let $\boldsymbol{\zeta}^s := A\boldsymbol{\zeta}^x$.

First, since $\mathbf{v}(\mathbf{z}^*) = \mathbf{0}$, $\mathbf{v}(\mathbf{z}^* + \boldsymbol{\zeta}) = H\boldsymbol{\zeta}^x - A^T\boldsymbol{\zeta}^\lambda$ is linear in $\boldsymbol{\zeta}$, making its norm absolutely homogeneous in $\boldsymbol{\zeta}$; and since $s_i = s_i^* + \zeta_i^s$ and $\lambda_i = \lambda_i^* + \zeta_i^\lambda$, complementarity slackness ($s_i^*\lambda_i^* = 0$) implies that $\|\mathbf{w}(\mathbf{x}^* + \boldsymbol{\zeta}^x, \boldsymbol{\lambda}^* + \boldsymbol{\zeta}^\lambda)\|$ is absolutely homogeneous in $\boldsymbol{\zeta}$ as well, in some neighborhood V_1 of the origin. Hence Ψ is indeed absolutely homogeneous in V_1.

Next, turning to strict positiveness and proceeding by contradiction, suppose that for every $\delta > 0$ there exists $\boldsymbol{\zeta} \neq \mathbf{0}$, with $\|\boldsymbol{\zeta}\| < \delta$, such that $\Psi(\boldsymbol{\zeta}) = 0$, i.e., $\mathbf{v}(\mathbf{z}^* + \boldsymbol{\zeta}) = \mathbf{0}$ and $\mathbf{w}(\mathbf{z}^* + \boldsymbol{\zeta}) = \mathbf{0}$. In view of (i), which implies uniqueness (over all of $\mathbb{R}^m$) of the KKT multiplier vector associated to $\mathbf{x}^*$, and given that $\boldsymbol{\zeta} \neq \mathbf{0}$, we must have $\boldsymbol{\zeta}^x \neq \mathbf{0}$. In view of (ii), this implies that $H\boldsymbol{\zeta}^x$ and $\boldsymbol{\zeta}^s_{\mathcal{I}} = A_{\mathcal{I}}\boldsymbol{\zeta}^x$ cannot vanish concurrently. On the other hand, for $i \in \mathcal{I}$ and for small enough δ, $w_i(\mathbf{z}^* + \boldsymbol{\zeta}) = 0$ implies $\zeta_i^s = 0$. Hence, $H\boldsymbol{\zeta}^x$ cannot vanish, and it must hold that $(\boldsymbol{\zeta}^x)^T H\boldsymbol{\zeta}^x > 0$. Since $\mathbf{v}(\mathbf{z}^* + \boldsymbol{\zeta}) = H\boldsymbol{\zeta}^x - A^T\boldsymbol{\zeta}^\lambda$, we conclude from $\mathbf{v}(\mathbf{z}^* + \boldsymbol{\zeta}) = \mathbf{0}$ that $(\boldsymbol{\zeta}^x)^T A^T\boldsymbol{\zeta}^\lambda > 0$, i.e., $(\boldsymbol{\zeta}^s)^T\boldsymbol{\zeta}^\lambda > 0$. Now, the argument that shows that $\zeta_i^s = 0$ when $\lambda_i^* > 0$ also shows that $\zeta_i^\lambda = 0$ when $s_i^* > 0$. Hence our inequality reduces to $\sum_{\{i:s_i^*=\lambda_i^*=0\}} \zeta_i^s\zeta_i^\lambda > 0$, in contradiction with $\mathbf{w}(\mathbf{z}^* + \boldsymbol{\zeta}) = \mathbf{0}$. Taking V to be a convex neighborhood of the origin contained in $V_1 \cap \{\boldsymbol{\zeta} : \|\boldsymbol{\zeta}\| < \delta\}$ completes the proof. □

Proposition 1 *Algorithm CR-MPC with Rule R satisfies Condition CSR.*

Proof To prove that Condition CSR(i) holds, let $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ be bounded away from $\mathcal{F}^*$, let $\mathbf{x}'$ be a limit point of $\{\mathbf{x}^k\}$, and let K be an infinite subsequence such that $\{\mathbf{x}^k\} \to \mathbf{x}'$ on K. By (17)–(18), $\{E_k\}$ is bounded away from zero so that, under Rule R, there exists $\delta' > 0$ such that $\delta_k > \delta'$ for all k. Now, with $\mathbf{s}' := A\mathbf{x}' - \mathbf{b}$ and $\mathbf{s}^k := A\mathbf{x}^k - \mathbf{b}$ for all k, since $\mathbf{s}'_{\mathcal{A}(\mathbf{x}')} = \mathbf{0}$, we have that, for all $i \in \mathcal{A}(\mathbf{x}')$, $s_i^k < \delta'$ for all large enough $k \in K$. Hence, for all large enough $k \in K$,

$$s_i^k < \delta' < \delta_k\,, \quad \forall i \in \mathcal{A}(\mathbf{x}')\,.$$

Since Rule R chooses the working set $Q_k := \{i \in \mathbf{m} \,|\, s_i^k \leq \delta_k\}$ for all k, we conclude that $\mathcal{A}(\mathbf{x}') \subseteq Q_k$ for all large enough $k \in K$, which proves Claim (i).

Turning now to Condition CSR (ii), suppose that (P)–(D) has a unique solution $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, that strict complementarity holds at $\mathbf{x}^*$, and that $\{\mathbf{x}^k\} \to \mathbf{x}^*$. If δ_k is reduced no more than finitely many times, then of course it is bounded away from zero, and the proof concludes as for Condition CSR(i); thus suppose $\{\delta_k\} \to 0$. Let $K := \{k \geq 1 : \delta_k = \theta\delta_{k-1}\}$ (an infinite index set) and, for given k, let $\ell(k)$ be the cardinality of $\{k' \leq k : k' \in K\}$. Then we have $\delta_k = \bar{\delta}\theta^{\ell(k)}$ for all k and $E_k \leq \beta^{\ell(k)}E_0$ for all $k \in K$. Since $\beta < \theta$ (see Rule R), this implies that $\{\frac{E_k}{\delta_k}\}_{k\in K} \to 0$. And from the definition of E_k and uniqueness of the solution to (P)–(D), it follows that $\{\boldsymbol{\lambda}^k\} \to \boldsymbol{\lambda}^*$ as $k \to \infty$, $k \in K$. We use these two facts to prove that, for all $i \in \mathcal{A}(\mathbf{x}^*)$ and some k_0,

$$s_i^k \leq \delta_k \quad \forall k \geq k_0; \tag{30}$$

in view of Rule R, this will complete the proof of Claim (ii). From Lemma 2, there exist $C > 0$ and $\eta > 0$ such that

$$\|(\mathbf{x} - \mathbf{x}^*, \boldsymbol{\psi} - \boldsymbol{\lambda}^*)\| \leq CE(\mathbf{x}, \boldsymbol{\psi})$$

for all $(\mathbf{x}, \boldsymbol{\psi})$ satisfying $\|(\mathbf{x} - \mathbf{x}^*, \boldsymbol{\psi} - \boldsymbol{\lambda}^*)\| < \eta$. Since $\{\frac{E_k}{\delta_k}\}_{k \in K} \to 0$ and $\{\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\}_{k \in K} \to \mathbf{0}$, and since $\|\mathbf{s}^k - \mathbf{s}^*\| \leq \|A\| \|\mathbf{x}^k - \mathbf{x}^*\|$ (since $\mathbf{s}^k - \mathbf{s}^* = A(\mathbf{x}^k - \mathbf{x}^*)$), there exists k_0 such that, for all $i \in \mathcal{A}(\mathbf{x}^*)$,

$$s_i^k \leq \|A\| \|(\mathbf{x}^k - \mathbf{x}^*, \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)\| \leq \|A\| CE_k \leq \delta_k, \quad \forall k \in K,\ k \geq k_0, \tag{31}$$

establishing (30) for $k \in K$. It remains to show that (30) does hold for *all* k large enough. Let ρ and C be as in Theorem 2 and without loss of generality suppose $C\rho \leq \theta$. Since $\{(\mathbf{x}^k - \mathbf{x}^*, \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)\}_{k \in K} \to \mathbf{0}$, there exists $k \in K$ (w.l.o.g. $k \geq k_0$), such that $\|(\mathbf{x}^k - \mathbf{x}^*, \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)\| < \rho$. Theorem 2 together with (31) then imply that

$$\|(\mathbf{x}^{k+1} - \mathbf{x}^*, \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*)\| \leq C \|(\mathbf{x}^k - \mathbf{x}^*, \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)\|^2 < \theta \|(\mathbf{x}^k - \mathbf{x}^*, \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)\| \leq \theta \delta_k / \|A\|.$$

(When $A = \mathbf{0}$, Proposition 1 holds trivially.) Hence, $\|(\mathbf{x}^{k+1} - \mathbf{x}^*, \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*)\| < \rho$ and since (in view of Rule R) δ_{k+1} is equal to either δ_k or $\theta\delta_k$ and $\theta \in (0, 1)$, we get, for all $i \in \mathcal{A}(\mathbf{x}^*)$,

$$s_i^{k+1} \leq \|A\| \|(\mathbf{x}^{k+1} - \mathbf{x}^*, \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*)\| \leq \delta_{k+1},$$

so that $\mathcal{A}(\mathbf{x}^*) \subseteq Q_{k+1}$. Theorem 2 can be applied recursively, yielding $\mathcal{A}(\mathbf{x}^*) \subseteq Q_k$ for all k large enough, concluding the proof of Claim (ii). □

Note that if a constraint-selection rule satisfies Condition CSR, rules derived from it by replacing Q_k by a superset of it also satisfy Condition CSR so that our convergence results still hold. Such augmentation of Q_k is often helpful; e.g., see Section 5.3 in [34]. Note however that the following corollary to Theorems 1–2 and Proposition 1, proved in Appendix A, of course does *not* apply when Rule R is thus augmented.

Corollary 2 *Suppose that Rule R is used in Step 2 of Algorithm CR-MPC,* $\varepsilon = 0$, *and that Assumptions 1–3 hold. Let* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *be the unique primal–dual solution. Further suppose that* $\lambda_i^* < \lambda^{\max}$ *for all* $i \in \mathbf{m}$. *Then, for sufficiently large* k, *Rule R gives* $Q_k = \mathcal{A}(\mathbf{x}^*)$.[12]

[12] In particular, if $\mathbf{x}^*$ is an unconstrained minimizer, the working set Q is eventually empty, and Algorithm CR-MPC reverts to a simple regularized Newton method (and terminates in one additional iteration if $H \succ \mathbf{0}$ and $R = \mathbf{0}$).

3 Numerical Experiments

We report computational results obtained with Algorithm CR-MPC on randomly generated problems and on data-fitting problems of various sizes.[13] Comparisons are made across different constraint-selection rules, including the unreduced case ($Q = \mathbf{m}$).[14]

3.1 Other Constraint-Selection Rules

As noted, the convergence properties of Algorithm CR-MPC that are given in Section 2.5 hold with any working-set selection rule that satisfies Condition CSR. The rules used in the numerical tests are our Rule R, Rule JOT of [18], Rule FFK–CWH of [5,8], and Rule `All` ($Q = \mathbf{m}$, i.e., no reduction). The details of Rule JOT and Rule FFK–CWH are stated below.

Rule JOT

Parameters: $\kappa > 0$, $q_U \in [n, m]$ (integer).
Input: Iteration: k, Slack variable: $\mathbf{s}^k$, Duality measure: $\mu := (\boldsymbol{\lambda}^k)^T \mathbf{s}^k / m$.
Output: Working set: Q_k.
Set $q := \min\{\max\{n, \lceil \mu^\kappa m \rceil\}, q_U\}$, and let η be the q-th smallest slack value.
Select $Q_k := \{i \in \mathbf{m} \,|\, s_i^k \leq \eta\}$.

Rule FFK–CWH

Parameter: $0 < r < 1$.
Input: Iteration: k, Slack variable: $\mathbf{s}^k$, Error: $E_k := E(\mathbf{x}^k, \boldsymbol{\lambda}^k)$ (see (17)).
Output: Working set: Q_k.
Select $Q_k := \{i \in \mathbf{m} \,|\, s_i^k \leq (E_k)^r\}$.

Note that the thresholds in Rule R and Rule FFK–CWH depend on both the duality measure μ and dual feasibility (see (17)) and these two rules impose no restriction on $|Q|$. On the other hand, the threshold in Rule JOT involves only μ, while it is required that $|Q| \geq n$. In addition, it is readily verified that Rule FFK–CWH satisfies Condition CSR, and that so does Rule JOT under Assumption 2.

It is worth noting that Rule R, Rule FFK–CWH, and Rule JOT all select constraints by comparing the values of primal slack variables s_i to some threshold values (independent of i), while the associated dual variables λ_i are not taken into account individually. Of course, variations with respect to such choice are possible. In fact, it was shown in [8] (also see an implementation

[13] In addition, a preliminary version of the proposed algorithm (with a modified version of Rule JOT, see [19] for details) was successfully tested in [20] on CQPs arising from a positivity-preserving numerical scheme for solving linear kinetic transport equations.

[14] We also ran comparison tests with the constraint-reduced algorithm of [23], for which polynomial complexity was established (as was superlinear convergence) for general semidefinite optimization problems. As was expected, that algorithm could not compete (orders of magnitude slower) with algorithms specifically targeting CQP.

in [5]) that the *strongly* active constraint set $\{i \in \mathbf{m}\colon \lambda_i^* > 0\}$ can be (asymptotically) identified at iteration k by the set $\{i \in \mathbf{m}\colon \lambda_i^k \geq \delta_k\}$ with a properly chosen threshold δ_k. Modifying the constraint selection rules considered in the present paper to include such information might improve the efficiency of the rules, especially when the constraints are poorly scaled. (Such modification does not affect the convergence properties of Algorithm CR-MPC as long as the modified rules still satisfy Condition CSR.) Numerical tests were carried out with an "augmented" Rule R that also includes $\{i \in \mathbf{m}\colon \lambda_i^k \geq \delta_k\}$ (with the same δ_k as in the original Rule R). The results suggest that, on the class of imbalanced problems considered in this section, while introducing some overhead, such augmentation (with the same δ_k) brings no benefit.

3.2 Implementation Details

All numerical tests were run with a Matlab implementation of Algorithm CR-MPC on a machine with Intel(R) Core(TM) i5-4200 CPU(3.1GHz), 4GB RAM, Windows 7 Enterprise, and Matlab 7.12.0(R2011a). In the implementation, $E(\mathbf{x}, \boldsymbol{\lambda})$ (see (17)–(18)) is normalized via division by the factor of $\max\{\|A\|_\infty, \|H\|_\infty, \|\mathbf{c}\|_\infty\}$, and 2-norms are used in (17) and Steps 9 and 10. In addition, for scaling purposes (see, for example, [18]), we used the normalized constraints $(DA)\mathbf{x} \geq D\mathbf{b}$, where $D = \mathrm{diag}\,(1/\|\mathbf{a}_i\|_2)$.

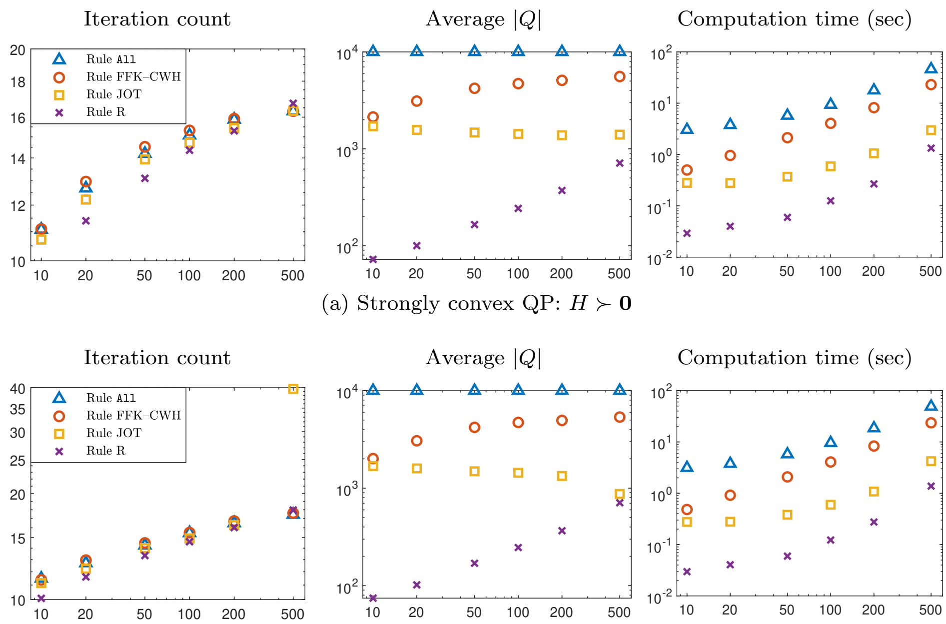
To highlight the significance of constraint-selection rules, a dense direct Cholesky solver was used to solve normal equations (20) and (15). Following [28] and [18], we set $s_i := \max\{s_i, 10^{-14}\}$ for all i when computing $M_{(Q)}$ in (12). Such safeguard prevents $M_{(Q)}$ from being too ill-conditioned and mitigates numerical difficulties in solving (20) and (15). When the Cholesky factorization of the modified $M_{(Q)}$ failed, we then doubled the regularization parameter ϱ and recomputed $M_{(Q)}$ in (12), and repeated this process until $M_{(Q)}$ was successfully factored.[15]

In the implementation, Algorithm CR-MPC is set to terminate as soon as either the stopping criterion (19) is satisfied or the iteration count reaches 200. The algorithm parameter values used in the tests were $\varepsilon = 10^{-8}$, $\tau = 0.5$, $\omega = 0.9$, $\varkappa = 0.98$, $\nu = 3$, $\lambda^{\max} = 10^{30}$, $\underline{\lambda} = 10^{-6}$, $R = I_{n\times n}$ (the $n \times n$ identity matrix), and $\bar{E} = E(\mathbf{x}^0, \boldsymbol{\lambda}^0)$, as suggested in footnote 5. The parameters in Rule R were given values $\beta = 0.4$, $\theta = 0.5$, and $\bar{\delta} =$ the $2n$-th smallest initial slack value. In Rule JOT, $\kappa = 0.25$ is used as in [18], and $q_U = m$ was selected (although $q_U = 3n$ is suggested as a "good heuristic" in [18]) to protect against a possible very large number of active constraints at the solution; the numerical results in [18] suggest that there is no significant downside in using $q_U = m$. In Rule FFK–CWH, $r = 0.5$ is used as in [5, 8].

[15] An alternative approach to take care of ill-conditioned $M_{(Q)}$ is to apply a variant of the Cholesky factorization that handles positive semi-definite matrices, such as the Cholesky-infinity factorization (i.e., `cholinc(X,'inf')` in Matlab) or the diagonal pivoting strategy discussed in [36, Chapter 11]. Either implementation does not make notable difference in the numerical results reported in this paper, since the Cholesky factorization fails in fewer than 1% of the tested problems.

3.3 Randomly Generated Problems

We first applied Algorithm CR-MPC on imbalanced ($m \gg n$) randomly generated problems; we used $m := 10\,000$ and n ranging from 10 to 500. Problems of the form (P) were generated in a similar way as those used in [18,28,34]. The entries of A and $\mathbf{c}$ were taken from a standard normal distribution $\mathcal{N}(0,1)$, those of $\mathbf{x}^0$ and $\mathbf{s}^0$ from uniform distributions $\mathcal{U}(0,1)$ and $\mathcal{U}(1,2)$, respectively, and we set $\mathbf{b} := A\mathbf{x}^0 - \mathbf{s}^0$, which guarantees that $\mathbf{x}^0$ is strictly feasible. We considered two sub-classes of problems: (i) strongly convex, with H diagonal and positive, with random diagonal entries from $\mathcal{U}(0,1)$, and (ii) linear, with $H = \mathbf{0}$. We solved 50 randomly generated problems for each sub-class of H and for each problem size, and report the results averaged over the 50 problems. There was no instance of failure on these problems. Figure 1 shows the results. (We also ran tests with H rank-deficient but nonzero, with similar results.)



(a) Strongly convex QP: $H \succ \mathbf{0}$

(b) linear optimization: $H = \mathbf{0}$

Fig. 1: Randomly generated problems with $m = 10\,000$ constraints– Numerical results on two types of randomly generated problems. In each figure, the x-axis is the number of variables (n) and the y-axis is iteration count, average size of working set, or computation time, all averaged over the 50 problem instances and plotted in logarithmic scale. The results of Rule `All`, Rule FFK–CWH, Rule JOT, and Rule R are plotted as blue triangles, red circles, yellow squares, and purple crosses, respectively.

It is clear from the plots that, in terms of computation time, Rule R outperforms other constraint-selection rules for the randomly generated problems we tested.[16] When the number of variables (n) is 1% $\sim$ 5% of the number of constraints (m) (i.e, n=100 to 500), Algorithm CR-MPC with Rule R is two to five times faster than with the second best rule, or 20 to 50 times faster than the unreduced algorithm (Rule `All`). When n is lowered to less than 1% of m (i.e., $n < 100$), the time advantage of using Rule R further doubles. Note that, as n decreases, Rule R is asymptotically more restrictive than Rule FFK–CWH and Rule JOT. We believe this may be the key reason that Rule R outperforms other rules, especially on problems with small n.

3.4 Data-Fitting Problems

We also applied Algorithm CR-MPC on CQPs arising from two instances of a data-fitting problem: trigonometric curve fitting to noisy observed data points. This problem was formulated in [28] as a linear optimization problem, and then in [18] reformulated as a CQP by imposing a regularization term. The CQP formulation of this problem, taken from [18], is as follows. Let $g : [0,1] \to \mathbb{R}$ be a given function of time, and let $\bar{\mathbf{b}} := [\bar{b}_1, \ldots, \bar{b}_{\bar{m}}]^T \in \mathbb{R}^{\bar{m}}$ be a vector that collects noisy observations of g at sample time $t = t_1, \ldots, t_{\bar{m}}$. The problem aims at finding a trigonometric expansion $u(t)$ from the noisy data $\bar{\mathbf{b}}$ that best approximates g. Here $u(t) := \sum_{j=1}^{\bar{n}} \bar{x}_j \psi_j(t)$, with the trigonometric basis

$$\psi_j(t) := \begin{cases} \cos(2(j-1)\pi t) \,, & j = 1, \ldots, \lceil \frac{\bar{n}}{2} \rceil \\ \sin(2(j - \lceil \frac{\bar{n}}{2} \rceil)\pi t) & j = \lceil \frac{\bar{n}}{2} \rceil + 1, \ldots, \bar{n} \end{cases} .$$

Equivalently, $u(t) = \bar{A}\bar{\mathbf{x}}$, where $\bar{\mathbf{x}} := [\bar{x}_1, \ldots, \bar{x}_{\bar{n}}]^T$ and $\bar{A}$ is a $\bar{m} \times \bar{n}$ matrix with entries $\bar{a}_{ij} = \psi_j(t_i)$. Based on a regularized minimax approach, the problem is then formulated as

$$\underset{\bar{\mathbf{x}} \in \mathbb{R}^{\bar{n}}}{\text{minimize}} \; \|\bar{A}\bar{\mathbf{x}} - \bar{\mathbf{b}}\|_\infty + \frac{1}{2}\bar{\alpha}\bar{\mathbf{x}}^T \bar{H} \bar{\mathbf{x}} \,,$$

where $\bar{H} \succeq \mathbf{0}$ is a symmetric $\bar{n} \times \bar{n}$ matrix, $\bar{\alpha}$ is a regularization parameter, and $\bar{\mathbf{x}}^T \bar{H} \bar{\mathbf{x}}$ is a regularization term that helps resist over-fitting. This problem can be rewritten as

$$\begin{aligned} \underset{\bar{\mathbf{x}} \in \mathbb{R}^{\bar{n}} v \in \mathbb{R}}{\text{minimize}} \; & v + \frac{1}{2}\bar{\alpha}\bar{\mathbf{x}}^T \bar{H} \bar{\mathbf{x}} \\ \text{subject to } & \bar{A}\bar{\mathbf{x}} - \bar{\mathbf{b}} \geq -v\mathbf{1} \,, \\ & -\bar{A}\bar{\mathbf{x}} + \bar{\mathbf{b}} \geq -v\mathbf{1} \,, \end{aligned}$$

which is a CQP in the form of (P) with number of variables $n = \bar{n} + 1$ and number of constraints $m = 2\bar{m}$.

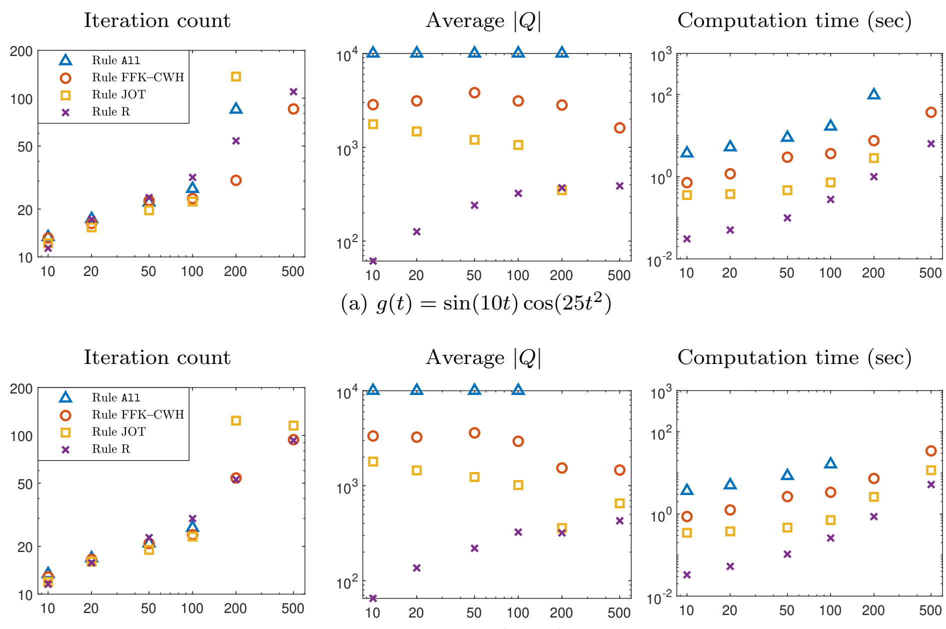
Following [18], we tested Algorithm CR-MPC on this problem with two target functions

$$g(t) = \sin(10t)\cos(25t^2) \quad \text{and} \quad g(t) = \sin(5t^3)\cos^2(10t) \,.$$

[16] Interestingly, on strongly convex problems, most rules (and especially Rule R) need a smaller number of iterations than Rule `All` (except for $n = 500$)!

In each case, as in [18], we sampled the data uniformly in time and set $\bar{b}_i := g(\frac{i-1}{\bar{m}}) + \epsilon_i$, where ϵ_i is an independent and identically distributed noise that takes values from $\mathcal{N}(0, 0.09)$,[17] for $i = 1, \ldots, \bar{m}$ and, as in [18], the regularization parameters were chosen as $\bar{\alpha} := 10^{-6}$ and $\bar{H} = \mathrm{diag}(\bar{\mathbf{h}})$, with $\bar{h}_1 := 0$ and $\bar{h}_j = \bar{h}_{j+\lceil \frac{\bar{n}}{2} \rceil - 1} := 2(j-1)\pi$, for $j = 2, \ldots, \lceil \frac{\bar{n}}{2} \rceil$, and $\bar{h}_{\bar{n}} := 2(\lfloor \frac{\bar{n}}{2} \rfloor)\pi$.

Figure 2 reports our numerical results. Since these problems involve noise, we solved the problem 50 times for each target function and report the average results. (The average is not reported—the corresponding symbol is not plotted—for problems on which one or more of the 50 runs failed, i.e., did not converge when iteration count reaches 200.) The sizes of the tested problems are $m := 10\,000$ and n ranging from 10 to 500.



(a) $g(t) = \sin(10t)\cos(25t^2)$

(b) $g(t) = \sin(5t^3)\cos^2(10t)$

Fig. 2: Data-fitting problems with $m = 10\,000$ constraints – Numerical results on two data-fitting problems. In each figure, the x-axis is the number of variables (n) and the y-axis is iteration count, average size of working set, or computation time, all averaged over the 50 problem instances and plotted in logarithmic scale. The results of Rule `All`, Rule FFK–CWH, Rule JOT, and Rule R are plotted as blue triangles, red circles, yellow squares, and purple crosses, respectively.

The results show that Rule R still outperforms other constraint-selection rules in terms of computation time, especially on problems with relatively small n. In general, Rule R is two to ten times faster than the second best

[17] We also ran the tests without noise and with noise of variance between 0 and 1, and the results were very similar to the ones reported here.

rule. We observe in Figure 2 that Rule JOT and Rule `All` failed to converge within 200 iterations in a few instances on problems with relatively large n. Numerical results suggest that these failures are due to ill-conditioning of M_Q apparently producing poor search directions. Thus, we conjecture that accurate identification of active constraints not only reduces computation time, but also alleviates the ill-conditioning issue of M_Q near optimal points.

3.5 Comparison with Broadly Used, Mature Code[18]

With a view toward calibrating the performance of Algorithm CR-MPC reported in Sections 3.3 and 3.4, we carried out a numerical comparison with two widely used solvers, SDPT3 [30, 32] and SeDuMi [26].[19] The tests were performed on the problems considered in Sections 3.3 and 3.4 with sizes $m = 10\,000$ and $n = 10, 20, 50, 100, 200, 500$. For CR-MPC, the exact same implementation, including starting points and stopping criterion, as outlined in Section 3.2 was used. As for the SDPT3 and SeDuMi solvers, we set the solver precision to 10^{-8} and let the solvers decide the starting points.

Table 1 reports the iteration counts and computation time of SDPT3, SeDuMi, and Algorithm CR-MPC with Rule `All` and Rule R, on each type of tested problems. The numbers in Table 1 are average values over 50 runs and over all six tested values of n. These results show that, for such significantly imbalanced problems, constraint reduction brings a clear edge. In particular, for such problems, Algorithm CR-MPC with Rule R shows a significantly better time-performance than two mature solvers.

	Randomly generated problems				Data fitting problems			
Algorithm	$H \succ \mathbf{0}$		$H = \mathbf{0}$		$\sin(10t)\cos(25t^2)$		$\sin(5t^3)\cos^2(10t)$	
	iteration	time	iteration	time	iteration	time	iteration	time
SDPT3	23.6	35.8	21.2	22.3	26.2	46.1	26.7	48.7
SeDuMi	22.0	4.0	16.3	4.4	26.9	5.1	26.1	5.5
Rule `All`	14.1	16.5	14.7	18.7	48.8	94.3	54.3	119.4
Rule R	13.2	0.3	14.3	0.4	38.7	1.4	43.8	1.6

Table 1: Comparison of Algorithm CR-MPC with popular codes – This table reports the iteration count and computation time (sec) for each of the compared algorithms on each type of tested problems, averaged over 50 runs. Every reported number is also averaged over various problems sizes: $m = 10\,000$ and $n = 10, 20, 50, 100, 200, 500$.

[18] It may also be worth pointing out that a short decade ago, in [33], the performance of an early version of a constraint-reduced MPC algorithm (with a more elementary constraint-selection rule than Rule JOT) was compared, on imbalanced filter-design applications (linear optimization), to the "revised primal simplex with partial pricing" algorithm discussed in [2], with encouraging results: on the tested problems, the constraint-reduced code proved competitive with the simplex code on some such problems and superior on others.

[19] While these two solvers have a broader scope (second-order cone optimization, semidefinite optimization) than Algorithm CR-MPC, they allow a close comparison with our code, as Matlab implementations are freely available within the CVX Matlab package [10, 11].

4 Conclusion

Convergence properties of the constraint-reduced algorithm proposed in this paper, which includes a number of novelties, were proved independently of the choice of the working-set selection rule, provided the rule satisfies Condition CSR. Under a specific such rule, based on a modified active-set identification scheme, the algorithm performs remarkably well in practice, both on randomly generated problems (CQPs as well as linear optimization problems) as well as data-fitting problems.

Of course, while the focus of the present paper was on dense problems, the concept of constraint reduction also applies to imbalanced large, sparse problems. Indeed, whichever technique is used for solving the Newton-KKT system, solving instead a reduced Newton-KKT system, of like sparsity but of drastically reduced size, is bound to bring in major computational savings when the total number of inequality constraints is much larger than the number of inequality constraints that are active at the solution—at least when the number of variables is reasonably small compared to the number of inequality constraints. In the case of sparse problems, the main computation cost in an IPM iteration would be that of a (sparse) Cholesky factorization or, in an iterative approach to solving the linear system, would be linked to the number of necessary iterations for reaching needed accuracy. In both cases, a major reduction in the dimension of the Newton-KKT system is bound to reduce the computation time, and like savings as in the dense case should be expected.

Appendix

The following results are used in the proofs in Appendices A and B. Here we assume that $Q \subseteq \mathbf{m}$ and W is symmetric, with $W \succeq H \succ \mathbf{0}$. First, from (10) and (13), the approximate MPC search direction $(\Delta\mathbf{x}, \Delta\boldsymbol{\lambda}_Q, \Delta\mathbf{s}_Q)$ defined in (16) solves

$$J(W, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q) \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda}_Q \\ \Delta\mathbf{s}_Q \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}) + (A_Q)^T \boldsymbol{\lambda}_Q \\ \mathbf{0} \\ -S_Q\boldsymbol{\lambda}_Q + \gamma\sigma\mu_{(Q)}\mathbf{1} - \gamma\Delta S_Q^{\mathrm{a}}\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}} \end{bmatrix}, \tag{32}$$

and equivalently, when $\mathbf{s}_Q > \mathbf{0}$, from (20), (21) and (15),

$$\begin{aligned} M_{(Q)}\Delta\mathbf{x} &= -\nabla f(\mathbf{x}) + (A_Q)^T S_Q^{-1}(\gamma\sigma\mu_{(Q)}\mathbf{1} - \gamma\Delta S_Q^{\mathrm{a}}\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}), \\ \Delta\mathbf{s}_Q &= A_Q\Delta\mathbf{x}, \\ \Delta\boldsymbol{\lambda}_Q &= -\boldsymbol{\lambda}_Q + S_Q^{-1}(-\Lambda_Q\Delta\mathbf{s}_Q + \gamma\sigma\mu_{(Q)}\mathbf{1} - \gamma\Delta S_Q^{\mathrm{a}}\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}})\,. \end{aligned} \tag{33}$$

Next, with $\tilde{\boldsymbol{\lambda}}^+$ and $\tilde{\boldsymbol{\lambda}}^{\mathrm{a},+}$ given by

$$\tilde{\lambda}_i^+ := \begin{cases} \lambda_i + \Delta\lambda_i & i \in Q, \\ 0 & i \in Q^{\mathrm{c}}, \end{cases} \quad \text{and} \quad \tilde{\lambda}_i^{\mathrm{a},+} := \begin{cases} \lambda_i + \Delta\lambda_i^{\mathrm{a}} & i \in Q, \\ 0 & i \in Q^{\mathrm{c}}, \end{cases} \tag{34}$$

from the last equation of (33) and from (21), we have

$$\tilde{\boldsymbol{\lambda}}_Q^+ = S_Q^{-1}(-\Lambda_Q\Delta\mathbf{s}_Q + \gamma\sigma\mu_{(Q)}\mathbf{1} - \gamma\Delta S_Q^{\mathrm{a}}\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}})\,, \tag{35}$$

$$\tilde{\boldsymbol{\lambda}}_Q^{\mathrm{a},+} = -S_Q^{-1}\Lambda_Q \Delta\mathbf{s}_Q^{\mathrm{a}} = -S_Q^{-1}\Lambda_Q A_Q \Delta\mathbf{x}_Q^{\mathrm{a}}\,, \tag{36}$$

and hence

$$(\Delta\mathbf{x}^{\mathrm{a}})^T (A_Q)^T \tilde{\boldsymbol{\lambda}}_Q^{\mathrm{a},+} = -(\Delta\mathbf{x}^{\mathrm{a}})^T (A_Q)^T S_Q^{-1}\Lambda_Q A_Q \Delta\mathbf{x}^{\mathrm{a}} \le 0\,, \tag{37}$$

so that, when in addition $\boldsymbol{\lambda} > \mathbf{0}$, $(\Delta\mathbf{x}^{\mathrm{a}})^T (A_Q)^T \tilde{\boldsymbol{\lambda}}^{\mathrm{a},+} = 0$ if and only if $A_Q \Delta\mathbf{x}^{\mathrm{a}} = \mathbf{0}$. Also, (20) yields

$$\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} = -(\Delta\mathbf{x}^{\mathrm{a}})^T M_{(Q)} \Delta\mathbf{x}^{\mathrm{a}} = -(\Delta\mathbf{x}^{\mathrm{a}})^T W \Delta\mathbf{x}^{\mathrm{a}} - (\Delta\mathbf{x}^{\mathrm{a}})^T (A_Q)^T S_Q^{-1}\Lambda_Q A_Q \Delta\mathbf{x}^{\mathrm{a}}\,. \tag{38}$$

Since $W \succeq H$, it follows from (37) that,

$$\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} + (\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} \le 0\,. \tag{39}$$

In addition, when $S_Q \succ \mathbf{0}$, $\Lambda_Q \succ \mathbf{0}$ and since $W \succeq \mathbf{0}$, the right-hand side of (38) is strictly negative as long as $W\Delta\mathbf{x}^{\mathrm{a}}$ and $A_Q\Delta\mathbf{x}^{\mathrm{a}}$ are not both zero. In particular, when $[W\ (A_Q)^T]$ has full row rank,

$$\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} < 0 \quad \text{if } \Delta\mathbf{x}^{\mathrm{a}} \neq \mathbf{0}\,. \tag{40}$$

Finally, we state and prove two technical lemmas.

Lemma 3 *Given an infinite index set K, $\{\Delta\mathbf{x}^k\} \to \mathbf{0}$ as $k \to \infty$, $k \in K$ if and only if $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$ as $k \to \infty$, $k \in K$.*

Proof We show that $\|\Delta\mathbf{x}^k\|$ is sandwiched between constant multiples of $\|\Delta\mathbf{x}^{\mathrm{a},k}\|$. We have from the search direction given in (16) that, for all k, $\|\Delta\mathbf{x}^k - \Delta\mathbf{x}^{\mathrm{a},k}\| = \|\gamma\Delta\mathbf{x}^{\mathrm{c},k}\| \le \tau\|\Delta\mathbf{x}^{\mathrm{a},k}\|$, where $\tau \in (0,1)$ and the inequality follows from (7). Apply triangle inequality leads to $(1-\tau)\|\Delta\mathbf{x}^{\mathrm{a},k}\| \le \|\Delta\mathbf{x}^k\| \le (1+\tau)\|\Delta\mathbf{x}^{\mathrm{a},k}\|$ for all k, proving the claim. □

Lemma 4 *Suppose Assumption 1 holds. Let $Q \subset \mathbf{m}$, $\mathcal{A} \subseteq Q$, $\mathbf{x} \in \mathcal{F}_P^o$, $\mathbf{s} := A\mathbf{x} - \mathbf{b}\ (> \mathbf{0})$, and $\boldsymbol{\lambda} > \mathbf{0}$ enjoy the following property: With $\Delta\boldsymbol{\lambda}_Q$, $\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}$, $\Delta\mathbf{s}$, and $\Delta\mathbf{s}^{\mathrm{a}}$ produced by Iteration CR-MPC, $\lambda_i + \Delta\lambda_i > 0$ for all $i \in \mathcal{A}$ and $s_i + \Delta s_i > 0$ for all $i \in Q \setminus \mathcal{A}$. Then*

$$\bar{\alpha}_{\mathrm{d}} \ge \min\left\{1,\ \min_{i\in Q\setminus\mathcal{A}}\left\{\frac{s_i}{|s_i + \Delta s_i^{\mathrm{a}}|}\right\},\ \min_{i\in Q\setminus\mathcal{A}}\left\{\frac{s_i - |\Delta s_i^{\mathrm{a}}|}{|s_i + \Delta s_i|}\right\}\right\} \tag{41}$$

and

$$\bar{\alpha}_{\mathrm{p}} \ge \min\left\{1,\ \min_{i\in\mathcal{A}}\left\{\frac{\lambda_i}{|\lambda_i + \Delta\lambda_i^{\mathrm{a}}|}\right\},\ \min_{i\in\mathcal{A}}\left\{\frac{\lambda_i - |\Delta\lambda_i^{\mathrm{a}}|}{|\lambda_i + \Delta\lambda_i|}\right\}\right\}\,. \tag{42}$$

Proof If $\bar{\alpha}_{\mathrm{d}} \ge 1$, (41) holds trivially, hence suppose $\bar{\alpha}_{\mathrm{d}} < 1$. Then, from the definition of $\bar{\alpha}_{\mathrm{d}}$ in (24), we know that there exists some index $i_0 \in Q$ such that

$$\Delta\lambda_{i_0} < -\lambda_{i_0} < 0 \quad \text{and} \quad \bar{\alpha}_{\mathrm{d}} = \frac{\lambda_{i_0}}{|\Delta\lambda_{i_0}|}\,. \tag{43}$$

Since $\lambda_i + \Delta\lambda_i > 0$ for all $i \in \mathcal{A}$, we have $i_0 \in Q \setminus \mathcal{A}$. Now we consider two cases: $|\Delta\lambda_{i_0}^{\mathrm{a}}| \ge |\Delta\lambda_{i_0}|$ and $|\Delta\lambda_{i_0}^{\mathrm{a}}| < |\Delta\lambda_{i_0}|$. If $|\Delta\lambda_{i_0}^{\mathrm{a}}| \ge |\Delta\lambda_{i_0}|$, then, since the second equation in (21) is equivalently written as $\lambda_i s_i + s_i\Delta\lambda_i^{\mathrm{a}} + \lambda_i\Delta s_i^{\mathrm{a}} = 0$ for all $i \in Q$ and since $\lambda_s s_i > 0$ for all $i \in \mathbf{m}$, it follows from (43) that

$$\bar{\alpha}_{\mathrm{d}} = \frac{\lambda_{i_0}}{|\Delta\lambda_{i_0}|} \ge \frac{\lambda_{i_0}}{|\Delta\lambda_{i_0}^{\mathrm{a}}|} = \frac{s_{i_0}}{|s_{i_0} + \Delta s_{i_0}^{\mathrm{a}}|}\,,$$

proving (41). To conclude, suppose now that $|\Delta\lambda_{i_0}^{\mathrm{a}}| < |\Delta\lambda_{i_0}|$. Since (i) $s_i + \Delta s_i > 0$ for $i \in Q \setminus \mathcal{A}$; (ii) γ, σ, and $\mu_{(Q)}$ in (35) are non-negative; and (iii) $\Delta\lambda_{i_0} < 0$ (from (43)), (34)–(35) yield

$$\lambda_{i_0}(s_{i_0} + \Delta s_{i_0}) \ge s_{i_0}|\Delta\lambda_{i_0}| - \gamma|\Delta s_{i_0}^{\mathrm{a}}||\Delta\lambda_{i_0}^{\mathrm{a}}|\,.$$

Applying this inequality to (43) leads to

$$\bar{\alpha}_{\mathrm{d}} = \frac{\lambda_{i_0}}{|\Delta\lambda_{i_0}|} \ge \frac{s_{i_0}}{|s_{i_0} + \Delta s_{i_0}|} - \frac{\gamma|\Delta\lambda_{i_0}^{\mathrm{a}}||\Delta s_{i_0}^{\mathrm{a}}|}{|s_{i_0} + \Delta s_{i_0}||\Delta\lambda_{i_0}|} \ge \frac{s_{i_0} - |\Delta s_{i_0}^{\mathrm{a}}|}{|s_{i_0} + \Delta s_{i_0}|},$$

where the last inequality holds since $\gamma \le 1$ and $|\Delta\lambda_{i_0}^{\mathrm{a}}| < |\Delta\lambda_{i_0}|$. Following a very similar argument that flips the roles of $\mathbf{s}$ and $\boldsymbol{\lambda}$, one can prove that (42) also holds. □

A Proof of Theorem 1 and Corollary 2

Parts of this proof are inspired from [29], [18], [16], [34], and [35]. Throughout, we assume that the constraint-selection rule used by the algorithm is such that Condition CSR is satisfied and (except in the proof of Lemma 13) we let $\varepsilon = 0$ and assume that the iteration never stops.

A central feature of Algorithm CR-MPC, which plays a key role in the convergence proofs, is that it forces descent with respect of the primal objective function. The next proposition establishes some related facts.

Proposition 2 *Suppose $\boldsymbol{\lambda} > \mathbf{0}$ and $\mathbf{s} > \mathbf{0}$, and W satisfies $W \succ \mathbf{0}$ and $W \succeq H$. If $\Delta\mathbf{x}^{\mathrm{a}} \neq \mathbf{0}$, then the following inequalities hold:*

$$f(\mathbf{x} + \alpha\Delta\mathbf{x}^{\mathrm{a}}) < f(\mathbf{x}) \,, \quad \forall \alpha \in (0,2) \,, \tag{44}$$

$$\frac{\partial}{\partial\alpha} f(\mathbf{x} + \alpha\Delta\mathbf{x}^{\mathrm{a}}) < 0 \,, \quad \forall \alpha \in [0,1] \,, \tag{45}$$

$$f(\mathbf{x}) - f(\mathbf{x} + \alpha\Delta\mathbf{x}) \geq \frac{\omega}{2}(f(\mathbf{x}) - f(\mathbf{x} + \alpha\Delta\mathbf{x}^{\mathrm{a}})) \,, \quad \forall \alpha \in [0,1] \,, \tag{46}$$

$$f(\mathbf{x} + \alpha\Delta\mathbf{x}) < f(\mathbf{x}) \,, \quad \forall \alpha \in (0,1] \,. \tag{47}$$

Proof When $f(\mathbf{x} + \alpha\Delta\mathbf{x}^{\mathrm{a}})$ is linear in α, i.e., when $(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} = 0$, then in view of (40), (44)–(45) hold trivially. When, on the other hand, $(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} > 0$, $f(\mathbf{x} + \alpha\Delta\mathbf{x}^{\mathrm{a}})$ is quadratic and strictly convex in α and is minimized at

$$\hat{\alpha} = -\frac{\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}}}{(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}}} = 1 + \frac{(\Delta\mathbf{x}^{\mathrm{a}})^T \left(W - H + (A_Q)^T S_Q^{-1} \Lambda_Q A_Q \right) \Delta\mathbf{x}^{\mathrm{a}}}{(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}}} \geq 1,$$

where we have used (38), (37), and the fact that $W \succeq H$, and (44)–(45) again follow. Next, note that, since $\omega > 0$,

$$\psi(\theta) := \omega(f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}})) - (f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}} + \theta\Delta\mathbf{x}^{\mathrm{c}}))$$

is quadratic and convex. Now, since γ_1 satisfies the constraints in its definition (8), we see that $\psi(\gamma_1) \leq 0$, and since $\omega \leq 1$, it follows from (44) that $\psi(0) = (\omega - 1)(f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}})) \leq 0$. Since $\gamma \in [0, \gamma_1]$ (see (7)), it follows that $\psi(\gamma) \leq 0$, i.e., since from (16) $\Delta\mathbf{x} = \Delta\mathbf{x}^{\mathrm{a}} + \gamma\Delta\mathbf{x}^{\mathrm{c}}$,

$$f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}) \geq \omega(f(\mathbf{x}) - f(\mathbf{x} + \Delta\mathbf{x}^{\mathrm{a}})) \,,$$

i.e.,

$$-\nabla f(\mathbf{x})^T \Delta\mathbf{x} - \frac{1}{2}\Delta\mathbf{x}^T H \Delta\mathbf{x} \geq \omega \left(-\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} - \frac{1}{2}(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} \right) . \tag{48}$$

Now, for all $\alpha \in [0,1]$, invoking (48), (39), and the fact that $H \succeq \mathbf{0}$, we can write

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x} + \alpha\Delta\mathbf{x}) &= -\alpha\nabla f(\mathbf{x})^T \Delta\mathbf{x} - \frac{\alpha^2}{2}\Delta\mathbf{x}^T H \Delta\mathbf{x} \geq \alpha \left(-\nabla f(\mathbf{x})^T \Delta\mathbf{x} - \tfrac{1}{2}\Delta\mathbf{x}^T H \Delta\mathbf{x} \right) \\ &\geq \omega\alpha \left(-\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} - \frac{1}{2}(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} \right) \\ &= \frac{\omega\alpha}{2} \left(-\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} - \left(\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} + (\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} \right) \right) \geq \tfrac{\alpha\omega}{2} \left(-\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} \right) \\ &\geq \frac{\alpha\omega}{2} \left(-\nabla f(\mathbf{x})^T \Delta\mathbf{x}^{\mathrm{a}} - \frac{\alpha}{2}(\Delta\mathbf{x}^{\mathrm{a}})^T H \Delta\mathbf{x}^{\mathrm{a}} \right) = \tfrac{\omega}{2}(f(\mathbf{x}) - f(\mathbf{x} + \alpha\Delta\mathbf{x}^{\mathrm{a}})) \,, \end{aligned}$$

proving (46). Finally, since $\omega > 0$, (47) is a direct consequence of (46) and (44). □

In particular, in view of (23)–(24), $\{f(\mathbf{x}^k)\}$ is monotonic decreasing. Since the iterates are primal–feasible, an immediate consequence of Proposition 2, stated next, is that, under Assumption 1, the primal sequence is bounded.

Lemma 5 *Suppose Assumption 1 holds. Then* $\{\mathbf{x}^k\}$ *is bounded.*

We are now ready to prove a key result, relating two successive iterates, that plays a central role in the remainder of the proof of Theorem 1.

Proposition 3 *Suppose Assumptions 1 and 2 hold, and either* $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ *is bounded away from* $\mathcal{F}^*$*, or Assumption 3 also holds and* $\{\mathbf{x}^k\}$ *converges to the unique primal solution* $\mathbf{x}^*$*. Let* K *be an infinite index set such that*

$$\left(\inf_{k\in K}\{\chi_{k-1}\} =\right) \ \inf\left\{\|\Delta\mathbf{x}^{\mathrm{a},k-1}\|^\nu + \|[\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k}_{Q_{k-1}}]_-\|^\nu : k \in K\right\} > 0\,. \tag{49}$$

Then $\{\Delta\mathbf{x}^k\} \to \mathbf{0}$ *as* $k \to \infty$*,* $k \in K$*.*

Proof From Lemma 5, $\{\mathbf{x}^k\}$ is bounded, and hence so is $\{\mathbf{s}^k\}$; by construction, $\mathbf{s}^k$ and $\boldsymbol{\lambda}^k$ have positive components for all k, and $\{\boldsymbol{\lambda}^k\}$ ((26)–(27)) and $\{W_k\}$ are bounded. Further, for any infinite index set K' such that (49) holds, (26) and (27) imply that all components of $\{\boldsymbol{\lambda}^k\}$ are bounded away from zero on K'. Since, in addition, Q_k can take no more than finitely many different (set) values, it follows that there exist $\hat{\mathbf{x}}$, $\hat{\boldsymbol{\lambda}} > \mathbf{0}$, $\hat{W} \succeq \mathbf{0}$, an index set $\hat{Q} \subseteq \mathbf{m}$, and some infinite index set $\hat{K} \subseteq K'$ such that

$$\{\mathbf{x}^k\} \to \hat{\mathbf{x}} \text{ as } k \to \infty\,,\ k \in \hat{K}\,,$$

$$\{\mathbf{s}^k\} \to \hat{\mathbf{s}} := \{A\hat{\mathbf{x}} - \mathbf{b}\} \geq \mathbf{0} \text{ as } k \to \infty\,,\ k \in \hat{K}\,. \tag{50}$$

$$\{\boldsymbol{\lambda}^k\} \to \hat{\boldsymbol{\lambda}} > \mathbf{0} \text{ as } k \to \infty\,,\ k \in \hat{K}\,, \tag{51}$$

$$\{W_k\} \to \hat{W} \text{ as } k \to \infty\,,\ k \in \hat{K}\,,$$

$$Q_k = \hat{Q}\,,\ \forall k \in \hat{K}\,. \tag{52}$$

Next, under the stated assumptions, $J(\hat{W}, A_{\hat{Q}}, \hat{\mathbf{s}}_{\hat{Q}}, \hat{\boldsymbol{\lambda}}_{\hat{Q}})$ is non-singular. Indeed, if $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ is bounded away from $\mathcal{F}^*$, then $E(\mathbf{x}^k, \boldsymbol{\lambda}^k)$ is bounded away from zero and since $H + R \succ \mathbf{0}$, $W_k = H + \varrho_k R = H + \min\left\{1, \frac{E(\mathbf{x}^k,\boldsymbol{\lambda}^k)}{\bar{E}}\right\} R$ is bounded away from singularity and the claim follows from Assumption 2 and Lemma 1. On the other hand, if Assumption 3 also holds and $\{\mathbf{x}^k\} \to \mathbf{x}^*$, then the claim follows from Condition CSR(ii) and Lemma 1. As a consequence of this claim, and by continuity of J, it follows from Newton-KKT systems (10) and (32) that there exist $\Delta\hat{\mathbf{x}}^{\mathrm{a}}$, $\Delta\hat{\mathbf{x}}$, $\bar{\boldsymbol{\lambda}}^{\mathrm{a}}_{\hat{Q}}$, $\bar{\boldsymbol{\lambda}}_{\hat{Q}}$ such that

$$\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \Delta\hat{\mathbf{x}}^{\mathrm{a}} \text{ as } k \to \infty\,,\ k \in \hat{K}\,, \tag{53}$$

$$\{\Delta\mathbf{x}^k\} \to \Delta\hat{\mathbf{x}} \text{ as } k \to \infty\,,\ k \in \hat{K}\,,$$

$$\{\Delta\mathbf{s}^k\} \to \Delta\hat{\mathbf{s}} := A\Delta\hat{\mathbf{x}} \text{ as } k \to \infty\,,\ k \in \hat{K}\,, \tag{54}$$

$$\{\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}_{\hat{Q}}\} \to \bar{\boldsymbol{\lambda}}^{\mathrm{a}}_{\hat{Q}} \text{ as } k \to \infty\,,\ k \in \hat{K}\,, \tag{55}$$

$$\{\tilde{\boldsymbol{\lambda}}^{k+1}_{\hat{Q}}\} \to \bar{\boldsymbol{\lambda}}_{\hat{Q}} \text{ as } k \to \infty\,,\ k \in \hat{K}\,, \tag{56}$$

The remainder of the proof proceeds by contradiction. Thus suppose that, for the infinite index set K in the statement of this lemma, $\{\Delta\mathbf{x}^k\} \not\to \mathbf{0}$ as $k \to \infty$, $k \in K$, i.e., for some $K'' \subseteq K$, $\|\Delta\mathbf{x}^k\|$ is bounded away from zero on K''. Use K'' as our K' above, so that (since $\hat{K} \subseteq K'$), $\|\Delta\mathbf{x}^k\|$ is bounded away from zero on $\hat{K}$. Then, in view of Lemma 3 (w.l.o.g.),

$$\inf_{k\in\hat{K}} \|\Delta\mathbf{x}^{\mathrm{a},k}\| > 0\,. \tag{57}$$

In addition, we have $\mathcal{A}(\hat{\mathbf{x}}) \subseteq \hat{Q}$, an implication of Condition CSR(i) when $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ is bounded away from $\mathcal{F}^*$ and of Condition CSR(ii) when Assumption 3 holds and $\{\mathbf{x}^k\}$

converges to $\mathbf{x}^*$. With these facts in hand, we next show that the sequence of primal step sizes $\{\alpha_{\mathrm{p}}^k\}$ is bounded away from zero for $k \in \hat{K}$. To this end, let us define

$$\tilde{\boldsymbol{\lambda}}'^{,k+1} := -(S^k)^{-1} \Lambda^k \Delta \mathbf{s}^k \ , \quad \forall k \, , \tag{58}$$

so that, for all $i \in \mathbf{m}$ and all k, $\tilde{\lambda}_i'^{,k+1} > 0$ if and only if $\Delta s_i^k < 0$, and the primal portion of (24) can be written as

$$\bar{\alpha}_{\mathrm{p}}^k := \begin{cases} \infty & \text{if } \tilde{\boldsymbol{\lambda}}'^{,k+1} \leq \mathbf{0} \, , \\ \min_i \left\{ \frac{\lambda_i^k}{\tilde{\lambda}_i'^{,k+1}} : \tilde{\lambda}_i'^{,k+1} > 0 \right\} & \text{otherwise.} \end{cases}$$

$$\alpha_{\mathrm{p}}^k := \min \left\{ 1, \, \max\{\varkappa \bar{\alpha}_{\mathrm{p}}, \, \bar{\alpha}_{\mathrm{p}} - \|\Delta \mathbf{x}^k\|\} \right\} \, .$$

Clearly, it is now sufficient to show that, for all i, $\{\tilde{\lambda}_i'^{,k+1}\}$ is bounded above on $\hat{K}$. On the one hand, this is clearly so for $i \notin \hat{Q}$ (whence $i \notin \mathcal{A}(\hat{\mathbf{x}})$), in view of (58) and (54), since $\{\boldsymbol{\lambda}^k\}$ is bounded and $\{s_i^k\}$ is bounded away from zero on $\hat{K}$ for $i \notin \mathcal{A}(\hat{\mathbf{x}})$ (from (50)). On the other hand, in view of (52), subtracting (58) from (35) yields, for all $k \in \hat{K}$,

$$\tilde{\boldsymbol{\lambda}}_{\hat{Q}}'^{k+1} = \tilde{\boldsymbol{\lambda}}_{\hat{Q}}^{k+1} - \gamma_k \sigma_k \mu_{(\hat{Q})}^k (S_{\hat{Q}}^k)^{-1} \mathbf{1} + \gamma_k (S_{\hat{Q}}^k)^{-1} \Delta S_{\hat{Q}}^{\mathrm{a},k} \Delta \boldsymbol{\lambda}_{\hat{Q}}^{\mathrm{a},k} \, .$$

From (56), $\{\tilde{\boldsymbol{\lambda}}_{\hat{Q}}^{k+1}\}$ is bounded on $\hat{K}$, and clearly the second term in the right-hand side of the above equation is non-positive component-wise. As for the third term, the second equation in (21) gives $(S_{Q_k}^k)^{-1} \Delta S_{Q_k}^{\mathrm{a},k} = (\Lambda_{Q_k}^k)^{-1} \tilde{\Lambda}_{Q_k}^{\mathrm{a},k+1}$, so that we have

$$\gamma_k (S_{\hat{Q}}^k)^{-1} \Delta S_{\hat{Q}}^{\mathrm{a},k} \Delta \boldsymbol{\lambda}_{\hat{Q}}^{\mathrm{a},k} = \gamma_k (\Lambda_{\hat{Q}}^k)^{-1} \tilde{\Lambda}_{\hat{Q}}^{\mathrm{a},k+1} \Delta \boldsymbol{\lambda}_{\hat{Q}}^{\mathrm{a},k} \, , \quad \forall k \in \hat{K} \, ,$$

which is bounded on $\hat{K}$ since, from (51), (55), and the definition (34) of $\{\tilde{\lambda}^{\mathrm{a},+}\}$, both $\{\tilde{\Lambda}_{\hat{Q}}^{\mathrm{a},k+1}\}$ and $\{\Delta \boldsymbol{\lambda}_{\hat{Q}}^{\mathrm{a},k}\}$ are bounded, and from (51), $\{\boldsymbol{\lambda}_{\hat{Q}}^k\}$ is bounded away from zero on $\hat{K}$. Therefore, $\{\tilde{\lambda}_i'^{,k+1}\}$ is bounded above on $\hat{K}$ for $i \in \hat{Q}$ as well, proving that $\{\alpha_{\mathrm{p}}^k\}$ is bounded away from zero on $\hat{K}$, i.e., that there exists $\underline{\alpha} > 0$ such that $\alpha_{\mathrm{p}}^k > \underline{\alpha}$, for all $k \in \hat{K}$, as claimed. Without loss of generality, choose $\underline{\alpha}$ in $(0, 1)$.

Finally, we show that $\{f(\mathbf{x}^k)\} \to -\infty$ as $k \to \infty$ on $\hat{K}$, which contradicts boundedness of $\{\mathbf{x}^k\}$ (Lemma 5). For all $k \in \hat{K}$, since $\Delta \mathbf{x}^{\mathrm{a},k} \neq \mathbf{0}$ (by (57)) and $\alpha_{\mathrm{p}}^k \in (\underline{\alpha}, 1]$, Proposition 2 implies that $\{f(\mathbf{x}^k)\}$ is monotonically decreasing and that, for all $k \in \hat{K}$,

$$f(\mathbf{x}^k + \alpha_{\mathrm{p}}^k \Delta \mathbf{x}^{\mathrm{a},k}) < f(\mathbf{x}^k + \underline{\alpha} \Delta \mathbf{x}^{\mathrm{a},k}) \, .$$

Expanding the right-hand side yields

$$f(\mathbf{x}^k + \underline{\alpha} \Delta \mathbf{x}^{\mathrm{a},k}) = f(\mathbf{x}^k) + \underline{\alpha} \nabla f(\mathbf{x}^k)^T \Delta \mathbf{x}^{\mathrm{a},k} + \frac{\underline{\alpha}^2}{2} (\Delta \mathbf{x}^{\mathrm{a},k})^T H \Delta \mathbf{x}^{\mathrm{a},k}$$
$$= f(\mathbf{x}^k) + \underline{\alpha} \left(\nabla f(\mathbf{x}^k)^T \Delta \mathbf{x}^{\mathrm{a},k} + (\Delta \mathbf{x}^{\mathrm{a},k})^T H \Delta \mathbf{x}^{\mathrm{a},k} \right) - \left(\underline{\alpha} - \frac{\underline{\alpha}^2}{2} \right) (\Delta \mathbf{x}^{\mathrm{a},k})^T H \Delta \mathbf{x}^{\mathrm{a},k} \, ,$$

where the sum of the last two terms tends to a strictly negative limit as $k \to \infty$, $k \in \hat{K}$. Indeed, in view of (39), the second term is non-positive and (i) if $(\Delta \hat{\mathbf{x}}^{\mathrm{a}})^T H \Delta \hat{\mathbf{x}}^{\mathrm{a}} > 0$, since $\underline{\alpha} > \underline{\alpha}^2/2$, from (53) and (57), the third term tends to a negative limit, and (ii) if $(\Delta \hat{\mathbf{x}}^{\mathrm{a}})^T H \Delta \hat{\mathbf{x}}^{\mathrm{a}} = 0$ then the sum of the last two terms tends to $\underline{\alpha} \nabla f(\hat{\mathbf{x}})^T \Delta \hat{\mathbf{x}}^{\mathrm{a}}$ which is also strictly negative in view of (40), since we either have $\hat{W} \succ \mathbf{0}$ (in the case that $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ bounded away from $\mathcal{F}^*$) or at least $[\hat{W} \, (A_{\hat{Q}})^T]$ full row rank (in the case that Assumption 3 holds and using the fact that $\mathcal{A}(\hat{\mathbf{x}}) \subseteq \hat{Q}$). It follows that, for some $\delta > 0$, $f(\mathbf{x}^k + \alpha_{\mathrm{p}}^k \Delta \mathbf{x}^{\mathrm{a},k}) < f(\mathbf{x}^k) - \delta$ for all $k \in \hat{K}$ large enough. Proposition 2 (eq. (46)) then gives that $f(\mathbf{x}^{k+1}) := f(\mathbf{x}^k + \alpha_{\mathrm{p}}^k \Delta \mathbf{x}^k) < f(\mathbf{x}^k) - \frac{\omega}{2} \delta$ for all $k \in \hat{K}$ large enough, where $\omega > 0$ is an algorithm parameter. Since $\{f(\mathbf{x}^k)\}$ is monotonically decreasing, the proof is now complete. □

We now conclude the proof of Theorem 1 via a string of eight lemmas, each of which builds on the previous one. First, on any subsequence, if $\{\Delta\mathbf{x}^{\mathrm{a},k}\}$ tends to zero, then $\{\mathbf{x}^k\}$ approaches stationary points. (Here both $\{\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}\}$ and $\{\tilde{\boldsymbol{\lambda}}^{k+1}\}$ are as defined in (34).)

Lemma 6 *Suppose that Assumption 1 holds and that $\{\mathbf{x}^k\}$ converges to some limit point $\hat{\mathbf{x}}$ on an infinite index set K. If $\{\Delta\mathbf{x}^{\mathrm{a},k}\}$ converges to zero on K, then (i) $\hat{\mathbf{x}}$ is stationary and*

$$\nabla f(\mathbf{x}^k) - \left(A_{\mathcal{A}(\hat{\mathbf{x}})}\right)^T \tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}_{\mathcal{A}(\hat{\mathbf{x}})} \to \mathbf{0}\,, \text{ as } k \to \infty\,,\ k \in K\,. \tag{59}$$

If, in addition, Assumption 2 holds, then (ii) $\{\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}\}$ and $\{\tilde{\boldsymbol{\lambda}}^{k+1}\}$ converge on K to $\hat{\boldsymbol{\lambda}}$, the unique multiplier associated with $\hat{\mathbf{x}}$.

Proof Suppose $\{\mathbf{x}^k\} \to \hat{\mathbf{x}}$ on K and $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$ on K. Let $\mathbf{s}^k := A\mathbf{x}^k - \mathbf{b}(> \mathbf{0})$ for all $k \in K$ and $\hat{\mathbf{s}} := A\hat{\mathbf{x}} - \mathbf{b}(\geq \mathbf{0})$, so that $\{\mathbf{s}^k\} \to \hat{\mathbf{s}}$ on K. As a first step toward proving Claim (i), we show that, for any $i \notin \mathcal{A}(\hat{\mathbf{x}})$, $\{\tilde{\lambda}_i^{\mathrm{a},k+1}\} \to 0$ on K. For $i \notin \mathcal{A}(\hat{\mathbf{x}})$, since $\hat{s}_i > 0$, $\{s_i^k\}$ is bounded away from zero on K. Since it follows from (34) and (36) that, for all k,

$$\tilde{\lambda}_i^{\mathrm{a},k+1} = 0\,,\ \forall i \notin Q_k \quad \text{and} \quad \tilde{\lambda}_i^{\mathrm{a},k+1} = -(s_i^k)^{-1}\lambda_i^k\Delta s_i^{\mathrm{a},k}\,,\ \forall i \in Q_k\,,$$

and since $\{\lambda_i^k\}$ is bounded (by construction) and $\Delta\mathbf{s}^{\mathrm{a},k} = A\Delta\mathbf{x}^{\mathrm{a},k}$ (by (21)), we have $\{\tilde{\lambda}_i^{\mathrm{a},k+1}\} \to 0$ on K. To complete the proof of Claim (i), note that the first equation of (10) (with H replaced by W) yields

$$\nabla f(\mathbf{x}^k) - (A_{Q_k})^T \tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}_{Q_k} = -W_k\Delta\mathbf{x}^{\mathrm{a},k}\,.$$

Since (i) $\{\tilde{\lambda}_i^{\mathrm{a},k+1}\} \to 0$ on K for $i \notin \mathcal{A}(\hat{\mathbf{x}})$, (ii) $\{W_k\}$ is bounded (since $H \preceq W_k \preceq H + R$), (iii) $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$ on K, and (iv) by definition (34), $\tilde{\lambda}_i^{\mathrm{a},+} = 0$ for $i \in Q^{\mathrm{c}}$, we conclude that (59) holds, hence $\{\left(A_{\mathcal{A}(\hat{\mathbf{x}})}\right)^T \tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}_{\mathcal{A}(\hat{\mathbf{x}})}\}$ converges (since $\nabla f(\mathbf{x}^k)$ does) as $k \to \infty$, $k \in K$, to a point in the range of $\left(A_{\mathcal{A}(\hat{\mathbf{x}})}\right)^T$, say $\left(A_{\mathcal{A}(\hat{\mathbf{x}})}\right)^T \hat{\boldsymbol{\lambda}}_{\mathcal{A}(\hat{\mathbf{x}})}$. We get $\nabla f(\hat{\mathbf{x}}) - \left(A_{\mathcal{A}(\hat{\mathbf{x}})}\right)^T \hat{\boldsymbol{\lambda}}_{\mathcal{A}(\hat{\mathbf{x}})} = \mathbf{0}$, proving Claim (i). Finally, Claim (ii) follows from (59), Assumption 2, and the fact that for $i \notin \mathcal{A}(\hat{\mathbf{x}})$, $\{\tilde{\lambda}_i^{\mathrm{a},k+1}\} \to 0$ as $k \to \infty$, $k \in K$, noting that the same argument applies to $\{\tilde{\boldsymbol{\lambda}}^{k+1}\}$, using a modified version of (59), with $\tilde{\boldsymbol{\lambda}}$ replacing $\tilde{\boldsymbol{\lambda}}^{\mathrm{a}}$, obtained by starting from the first equation of (32) instead of that of (10) and using the fact, proved next, that $\{\tilde{\lambda}_i^{k+1}\} \to 0$ on K for all $i \notin \mathcal{A}(\hat{\mathbf{x}})$. From its definition in (34) and the last equation in (33), we have that, for all k,

$$\begin{aligned} \tilde{\lambda}_i^{k+1} &= 0\,, \quad \forall i \notin Q_k\,, \\ \tilde{\lambda}_i^{k+1} &= (s_i^k)^{-1}(-\lambda_i^k\Delta s_i^k + \gamma_k\sigma_k\mu^{(k)}_{(Q_k)} - \gamma_k\Delta s_i^{\mathrm{a},k}\Delta\lambda_i^{\mathrm{a},k})\,, \quad \forall i \in Q_k\,. \end{aligned}$$

Since $\{\tilde{\lambda}_i^{\mathrm{a},k+1}\}$ converges (to zero) on K, $\{\Delta\lambda_i^{\mathrm{a},k}\}$ is bounded on K. Furthermore, from its definition (7)–(8) (see also (16)), $\{\gamma_k\}$ is bounded and $|\gamma_k\sigma_k\mu^{(k)}_{(Q_k)}| \leq \tau\|\Delta\mathbf{x}^{\mathrm{a},k}\|$ for all k. Since $\Delta\mathbf{s}^{\mathrm{a},k} = A\Delta\mathbf{x}^{\mathrm{a},k}$ and $\Delta\mathbf{s}^k = A\Delta\mathbf{x}^k$, in view of Lemma 3, it follows that, for $i \notin \mathcal{A}(\hat{\mathbf{x}})$, $\{\tilde{\lambda}_i^{k+1}\} \to 0$ on K. □

Lemma 6, combined with Proposition 3 via a contradiction argument, then implies that (on a subsequence), if $\{\mathbf{x}^k\}$ does not approach $\mathcal{F}_P^*$, then $\{\Delta\mathbf{x}^k\}$ approaches zero.

Lemma 7 *Suppose that Assumptions 1 and 2 hold and that $\{\mathbf{x}^k\}$ is bounded away from $\mathcal{F}_P^*$ on some infinite index set K. Then $\{\Delta\mathbf{x}^k\} \to \mathbf{0}$ as $k \to \infty$, $k \in K$.*

Proof Proceeding by contradiction, let K be an infinite index set such that $\{\mathbf{x}^k\}$ is bounded away from $\mathcal{F}_P^*$ on K and $\{\Delta\mathbf{x}^k\} \not\to \mathbf{0}$ as $k \to \infty$, $k \in K$. Then, in view of Proposition 3 and boundedness of $\{\mathbf{x}^k\}$ (Lemma 5), there exist $\hat{Q} \subseteq \mathbf{m}$, $\hat{\mathbf{x}} \notin \mathcal{F}_P^*$, and an infinite index set $\hat{K} \subseteq K$ such that $Q_k = \hat{Q}$ for all $k \in \hat{K}$ and

$$\begin{aligned} &\{\mathbf{x}^k\} \to \hat{\mathbf{x}}\,, \text{ as } k \to \infty,\ k \in \hat{K}\,, \\ &\{\Delta\mathbf{x}^{\mathrm{a},k-1}\} \to \mathbf{0}\,, \text{ as } k \to \infty,\ k \in \hat{K}\,, \\ &\{[\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k}_{\hat{Q}}]_-\} \to \mathbf{0}\,, \text{ as } k \to \infty,\ k \in \hat{K}\,. \end{aligned}$$

On the other hand, from (25), (16) and (7)–(8),

$$\|\mathbf{x}^k - \mathbf{x}^{k-1}\| = \|\alpha_{\mathrm{p}}^{k-1}\Delta\mathbf{x}^{k-1}\| \le \|\Delta\mathbf{x}^{k-1}\| \le (1+\tau)\|\Delta\mathbf{x}^{\mathrm{a},k-1}\|\,,$$

which implies that $\{\mathbf{x}^{k-1}\} \to \hat{\mathbf{x}}$ as $k \to \infty$, $k \in \hat{K}$. It then follows from Lemma 6 that $\hat{\mathbf{x}}$ is stationary and that $[\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k}_{\mathcal{A}(\hat{\mathbf{x}})}]_+$ converges to the associated multiplier vector. Hence the multipliers are non-negative, contradicting the fact that $\hat{\mathbf{x}} \notin \mathcal{F}_P^*$. □

A contradiction argument based on Lemmas 6 and 7 then shows that $\{\mathbf{x}^k\}$ approaches the set of stationary points of (P).

Lemma 8 *Suppose Assumptions 1 and 2 hold. Then the sequence $\{\mathbf{x}^k\}$ approaches the set of stationary points of* (P)*, i.e., there exists a sequence $\{\hat{\mathbf{x}}^k\}$ of stationary points such that $\|\mathbf{x}^k - \hat{\mathbf{x}}^k\|$ goes to zero as $k \to \infty$.*

Proof Proceeding by contradiction, suppose the claim does not hold, i.e., (invoking Lemma 5) suppose $\{\mathbf{x}^k\}$ converges to some non-stationary point $\hat{\mathbf{x}}$ on some infinite index set K. Then $\{\Delta\mathbf{x}^{\mathrm{a},k}\}$ does not converge to zero on K (Lemma 6(i)) and nor does $\{\Delta\mathbf{x}^k\}$ (Lemma 3). Since $\hat{\mathbf{x}}$ is non-stationary, this is in contradiction with Lemma 7. □

The next technical result, proved in [29, Lemma 3.6], invokes analogues of Lemmas 5, 7 and 8.

Lemma 9 *Suppose Assumptions 1 and 2 hold. Suppose $\{\mathbf{x}^k\}$ is bounded away from $\mathcal{F}_P^*$. Let $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ be limit points of $\{\mathbf{x}^k\}$ and let $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\lambda}}'$ be the associated KKT multipliers. Then $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}'$.*

Convergence of $\{\mathbf{x}^k\}$ to $\mathcal{F}_P^*$ ensues, proving Claim (i) of Theorem 1.

Lemma 10 *Suppose Assumptions 1 and 2 hold. Then $\{\mathbf{x}^k\}$ converges to $\mathcal{F}_P^*$.*

Proof We proceed by contradiction. Thus suppose $\{\mathbf{x}^k\}$ does not converge to $\mathcal{F}_P^*$. Then, since $\{\mathbf{x}^k\}$ is bounded (Lemma 5), it has at least one limit point $\hat{\mathbf{x}}$ that is not in $\mathcal{F}_P^*$, and since (by Proposition 2) $\{f(\mathbf{x}^k)\}$ is a bounded, monotonically decreasing sequence, $f(\hat{\mathbf{x}}) = \inf_k f(\mathbf{x}^k)$. Then, by Lemmas 7 and 3, $\{\Delta\mathbf{x}^k\}$ and $\{\Delta\mathbf{x}^{\mathrm{a},k}\}$ converge to zero as $k \to \infty$. It follows from Lemmas 6 and 9 that all limit points of $\{\mathbf{x}^k\}$ are stationary, and that both $\{\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k}\}$ and $\{\tilde{\boldsymbol{\lambda}}^k\}$ converge to $\hat{\boldsymbol{\lambda}}$, the common KKT multiplier vector associated to all limit points of $\{\mathbf{x}^k\}$. Since $\hat{\mathbf{x}} \notin \mathcal{F}_P^*$, there exists i_0 such that $\hat{\lambda}_{i_0} < 0$, so that, for some $\hat{k} > 0$,

$$\tilde{\lambda}_{i_0}^{\mathrm{a},k+1} < 0 \text{ and } \tilde{\lambda}_{i_0}^{k+1} < 0\,, \quad \forall k > \hat{k}\,, \tag{60}$$

which, in view of Step 8 of the algorithm, implies that $i_0 \in Q_k$ for all $k > \hat{k}$. Then (36) gives

$$\Delta s_{i_0}^{\mathrm{a},k} = -(\lambda_{i_0}^k)^{-1} s_{i_0}^k \tilde{\lambda}_{i_0}^{\mathrm{a},k+1}\,, \quad \forall k > \hat{k}\,,$$

where $s_{i_0}^k > 0$, $\lambda_{i_0}^k > 0$ by construction. Thus, in view of (60), $\Delta s_{i_0}^{\mathrm{a},k} > 0$ for all $k > \hat{k}$. On the other hand, the last equation of (33) gives

$$\Delta s_{i_0}^k = (\lambda_{i_0}^k)^{-1}(-s_{i_0}^k \tilde{\lambda}_{i_0}^{k+1} + \gamma_k \sigma_k \mu_{(Q_k)}^k - \gamma_k \Delta s_{i_0}^{\mathrm{a},k} \Delta\lambda_{i_0}^{\mathrm{a},k}), \quad \forall k > \hat{k}\,, \tag{61}$$

where $\gamma_k \geq 0$, $\sigma_k \geq 0$, and $\mu^k_{(Q_k)} \geq 0$ by construction. Further, for $k > \hat{k}$, $\Delta\lambda^{\mathrm{a},k}_{i_0} < 0$ since $\lambda^k_{i_0} > 0$ and $\tilde{\lambda}^{\mathrm{a},k+1}_{i_0} (= \lambda^k_{i_0} + \Delta\lambda^{\mathrm{a},k}_{i_0}) < 0$. It follows that all terms in (61) are non-negative and the first term is positive, so that $\Delta s^k_{i_0} > 0$ for all $k > k'$. Moreover, for all $k > \hat{k}$, we have $s^{k+1}_{i_0} = s^k_{i_0} + \alpha^k_{\mathrm{p}} \Delta s^k_{i_0} > s^k_{i_0} > 0$, where $\alpha^k_{\mathrm{p}} > 0$ since $\mathbf{s}^k > \mathbf{0}$. Since $\{\mathbf{s}^k\}$ is bounded (Lemma 5), we then conclude that $\{s^k_{i_0}\} \to \hat{s}_{i_0} > 0$ so that $\hat{s}_{i_0} \hat{\lambda}_{i_0} < 0$, in contradiction with the stationarity of limit points. □

Under strict complementarity, the next lemma then establishes appropriate convergence of the multipliers, setting the stage for the proof of part (ii) of Theorem 1 in the following lemma.

Lemma 11 *Suppose Assumptions 1 to 3 hold and let* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *be the unique primal-dual solution. Then, given any infinite index set* K *such that* $\{\Delta\mathbf{x}^{\mathrm{a},k}\}_{k\in K} \to \mathbf{0}$*, it holds that* $\{\boldsymbol{\lambda}^{k+1}\}_{k\in K} \to \boldsymbol{\xi}^*$*, where* $\xi^*_i := \min\{\lambda^*_i, \lambda^{\max}\}$*, for all* $i \in \mathbf{m}$.

Proof Lemma 10 guarantees that $\{\mathbf{x}^k\} \to \mathbf{x}^*$. Let K be an infinite index set such that $\{\Delta\mathbf{x}^{\mathrm{a},k}\}_{k\in K} \to \mathbf{0}$. Then, in view of Lemma 6(ii), $\{\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}\}_{k\in K} \to \boldsymbol{\lambda}^* \geq \mathbf{0}$. Accordingly, $\{\chi_k\} = \{\|\Delta\mathbf{x}^{\mathrm{a},k}\|^\nu + \|[\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k+1}_{Q_k}]_-\|^\nu\} \to 0$ as $k \to \infty$, $k \in K$. Hence, in view of (26) and (27), the proof will be complete if we show that $\{\breve{\boldsymbol{\lambda}}^{k+1}\}_{k\in K} \to \boldsymbol{\lambda}^*$, where

$$\breve{\lambda}^{k+1}_i := \lambda^k_i + \alpha^k_{\mathrm{d}} \Delta\lambda^k_i\,, \quad i \in Q_k, \quad \text{and} \quad \breve{\lambda}^{k+1}_i := \mu^{k+1}_{(Q_k)} / s^{k+1}_i\,, \quad i \in Q^{\mathrm{c}}_k\,,$$

or equivalently, $\{\tilde{\boldsymbol{\lambda}}^{k+1} - \breve{\boldsymbol{\lambda}}^{k+1}\}_{k\in K} \to \mathbf{0}$, which we do now.

For every $Q \subseteq \mathbf{m}$, define the index set $K(Q) := \{k \in K : Q_k = Q\}$, and let $\mathcal{Q} := \{Q \subseteq \mathbf{m} : |K(Q)| = \infty\}$. We first show that for all $Q \in \mathcal{Q}$, $\{\tilde{\boldsymbol{\lambda}}^{k+1}_Q - \breve{\boldsymbol{\lambda}}^{k+1}_Q\}_{k\in K(Q)} \to \mathbf{0}$. For $Q \in \mathcal{Q}$, the definition (34) of $\tilde{\boldsymbol{\lambda}}^+$ yields

$$\|\tilde{\boldsymbol{\lambda}}^{k+1}_Q - \breve{\boldsymbol{\lambda}}^{k+1}_Q\| = (1 - \alpha^k_{\mathrm{d}}) \|\Delta\boldsymbol{\lambda}^k_Q\|, \quad k \in K(Q)\,.$$

Since boundedness of $\{\boldsymbol{\lambda}^k_Q\}$ (by construction) and of $\{\tilde{\boldsymbol{\lambda}}^{k+1}_Q\}_{k\in K}$ $(=\{\boldsymbol{\lambda}^k_Q + \Delta\boldsymbol{\lambda}^k_Q)\}_{k\in K})$ (by Lemma 6(ii)) implies boundedness of $\{\Delta\boldsymbol{\lambda}^k_Q\}_{k\in K}$, we only need $\{\alpha^k_{\mathrm{d}}\}_{k\in K(Q)} \to 1$ in order to guarantee that $\|\tilde{\boldsymbol{\lambda}}^{k+1}_Q - \breve{\boldsymbol{\lambda}}^{k+1}_Q\| \to 0$ on $K(Q)$. Now, $\{\Delta\mathbf{x}^{\mathrm{a},k}\}_{k\in K} \to \mathbf{0}$ implies that $\{\Delta\mathbf{s}^{\mathrm{a},k}\}_{k\in K} \to \mathbf{0}$, and from Lemma 3 that $\{\Delta\mathbf{x}^k\}_{k\in K} \to \mathbf{0}$, implying that $\{\Delta\mathbf{s}^k\}_{k\in K} \to \mathbf{0}$; and $\{\mathbf{x}^k\} \to \mathbf{x}^*$ yields $\{\mathbf{s}^k\} \to \mathbf{s}^* := A\mathbf{x}^* - \mathbf{b}$, so $s^k_i + \Delta s^k_i > 0$ for all $i \in \mathcal{A}(\mathbf{x}^*)^{\mathrm{c}}$, $k \in K$ large enough. Moreover, Assumption 3 gives $\lambda^*_i > 0$ for all $i \in \mathcal{A}(\mathbf{x}^*)$ so that, for sufficiently large $k \in K$, $\tilde{\lambda}^{k+1}_i > 0$ for all $i \in \mathcal{A}(\mathbf{x}^*)$, and Condition CSR(ii) implies that $\mathcal{A}(\mathbf{x}^*) \subseteq Q$, so Lemma 4 applies, with $\mathcal{A} := \mathcal{A}(\mathbf{x}^*)$. It follows that $\{\bar{\alpha}^k_{\mathrm{d}}\}_{k\in K} \to 1$, since all terms on the right-hand side of (41) converge to one on K. Thus, from the definition of α^k_{d} in (24) and the fact that $\{\Delta\mathbf{x}^k\}_{k\in K} \to \mathbf{0}$, we have $\{\alpha^k_{\mathrm{d}}\}_{k\in K} \to 1$ indeed, establishing that $\{\tilde{\boldsymbol{\lambda}}^{k+1}_Q - \breve{\boldsymbol{\lambda}}^{k+1}_Q\}_{k\in K(Q)} \to \mathbf{0}$.

It remains to show that, for all $Q \in \mathcal{Q}$, $\{\tilde{\boldsymbol{\lambda}}^{k+1}_{Q^{\mathrm{c}}} - \breve{\boldsymbol{\lambda}}^{k+1}_{Q^{\mathrm{c}}}\}_{k\in K(Q)} \to \mathbf{0}$. To show this, we first note that, since $\{\chi_k\}_{k\in K} \to 0$, it follows from (26) and (27) and the fact established above that $\{\breve{\boldsymbol{\lambda}}^{k+1}_Q\}_{K(Q)} \to \boldsymbol{\lambda}^*_Q$ that, for all $Q \in \mathcal{Q}$,

$$\{\boldsymbol{\lambda}^{k+1}_Q\} \to \boldsymbol{\xi}^*_Q, \quad k \to \infty,\ k \in K(Q). \tag{62}$$

Next, from (26), (27), and the definition (34) of $\tilde{\boldsymbol{\lambda}}^+$, we have, for $Q \in \mathcal{Q}$ and sufficiently large $k \in K(Q)$,

$$|\tilde{\lambda}^{k+1}_i - \breve{\lambda}^{k+1}_i| = \breve{\lambda}^{k+1}_i = \frac{\mu^{k+1}_{(Q)}}{s^{k+1}_i}, \quad i \in Q^{\mathrm{c}}\,. \tag{63}$$

Clearly, since $\mathcal{A}(\mathbf{x}^*) \subseteq Q$, we have $s_i^* > 0$ for $i \in Q^{\mathrm{c}}$. Hence, since $\{\Delta\mathbf{s}^k\}_{k\in K} \to \mathbf{0}$, $\{s_i^{k+1}\}$ is bounded away from zero on K for $i \in Q^{\mathrm{c}}$. When Q is empty, the right-hand side of (63) is set to zero (see definition (14) of $\mu_{(Q)}$). When Q is not empty, since $\xi_i^* = 0$ whenever $\lambda_i^* = 0$, (62) and complementary slackness gives

$$\left\{\mu_{(Q)}^{k+1}\right\} = \left\{\frac{(\mathbf{s}_Q^{k+1})^T\boldsymbol{\lambda}_Q^{k+1}}{|Q|}\right\} \to \left\{\frac{(\mathbf{s}_Q^*)^T\boldsymbol{\xi}_Q^*}{|Q|}\right\} = 0, \quad k \in K(Q)\,,$$

and it follows from (63) that $\{\tilde{\boldsymbol{\lambda}}_{Q^{\mathrm{c}}}^{k+1} - \breve{\boldsymbol{\lambda}}_{Q^{\mathrm{c}}}^{k+1}\}_{k\in K(Q)} \to \mathbf{0}$, completing the proof. □

Claim (ii) of Theorem 1 can now be proved.

Lemma 12 *Suppose Assumptions 1 to 3 hold and let* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *be the unique primal-dual solution. Then* $\{\tilde{\boldsymbol{\lambda}}^k\} \to \boldsymbol{\lambda}^*$ *and* $\{\boldsymbol{\lambda}^k\} \to \boldsymbol{\xi}^*$*, with* $\xi_i^* := \min\{\lambda_i^*, \lambda^{\max}\}$ *for all* $i \in \mathbf{m}$.

Proof Again, Lemma 10 guarantees that $\{\mathbf{x}^k\} \to \mathbf{x}^*$ and $\{\mathbf{s}^k\} \to \mathbf{s}^* := A\mathbf{x}^* - \mathbf{b}$. Note that if $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$, the claims are immediate consequences of Lemmas 6 and 11. We now prove by contradiction that $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$. Thus, suppose that for some infinite index set K, $\inf_{k\in K}\|\Delta\mathbf{x}^{\mathrm{a},k}\| > 0$. Then, Lemma 3 gives $\inf_{k\in K}\|\Delta\mathbf{x}^k\| > 0$. It follows from Proposition 3 that, on some infinite index set $K' \subseteq K$, $\{\Delta\mathbf{x}^{\mathrm{a},k-1}\} \to \mathbf{0}$ and $\{[\tilde{\boldsymbol{\lambda}}^{\mathrm{a},k}]_-\} \to \mathbf{0}$. Since Q_k is selected from a finite set and $\{W_k\}$ is bounded, we can assume without loss of generality that $Q_k = Q$ on K' for some $Q \subseteq \mathbf{m}$, and that $\{W_k\} \to W^* \succeq H$ on K'. Further, from Lemma 11, $\{\boldsymbol{\lambda}^k\}_{k\in K'} \to \boldsymbol{\xi}^*$. Therefore, $\{J(W_k, A_{Q_k}, \mathbf{s}_{Q_k}, \boldsymbol{\lambda}_{Q_k})\}_{k\in K'} \to J(W^*, A_Q, \mathbf{s}_Q^*, \boldsymbol{\xi}_Q^*)$, and in view of Assumptions 2 and 3 and Lemma 1, $J(W^*, A_Q, \mathbf{s}_Q^*, \boldsymbol{\xi}_Q^*)$ is non-singular (since $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is optimal). It follows from (10), with W substituted for H, that $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$ on K', a contradiction, proving that $\{\Delta\mathbf{x}^{\mathrm{a},k}\} \to \mathbf{0}$. □

Claim (iv) of Theorem 1 follows as well.

Lemma 13 *Suppose Assumptions 1 and 2 hold and* $\varepsilon > 0$*. Then Algorithm CR-MPC terminates (in Step 1) after finitely many iterations.*

Proof If $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ has a limit point in $\mathcal{F}^*$, then $\inf_k\{E_k\} = 0$, proving the claim. Thus, suppose that $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ is bounded away from $\mathcal{F}^*$. In view of Lemmas 5 and 10, $\{\mathbf{x}^k\}$ has a limit point $\mathbf{x}^* \in \mathcal{F}_P^*$. Assumption 2 then implies that there exists a unique KKT multiplier vector $\boldsymbol{\lambda}^* \geq \mathbf{0}$ associated to $\mathbf{x}^*$. If $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{F}^*$ is a limit point of $\{(\mathbf{x}^k, \tilde{\boldsymbol{\lambda}}^k)\}$, which also implies that $\inf_k\{E(\mathbf{x}^k, \tilde{\boldsymbol{\lambda}}^k)\} = 0$, then in view of the stopping criterion, the claim again follows. Thus, further suppose that there is an infinite index set K such that $\{\mathbf{x}^k\}_{k\in K} \to \mathbf{x}^*$, but $\inf_{k\in K}\|\tilde{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^*\| > 0$. It then follows from Lemma 6(ii) that $\{\Delta\mathbf{x}^{\mathrm{a},k-1}\}_{k\in K} \not\to \mathbf{0}$, and from Lemma 3 that $\{\Delta\mathbf{x}^{k-1}\}_{k\in K} \not\to \mathbf{0}$. Proposition 3 and Lemma 3 then imply that $\{\Delta\mathbf{x}^{\mathrm{a},k-2}\}_{k\in K'} \to \mathbf{0}$ and $\{\Delta\mathbf{x}^{k-2}\}_{k\in K'} \to \mathbf{0}$ for some infinite index set $K' \subseteq K$. Next, from Lemmas 5 and 10, we have $\{\mathbf{x}^{k-2}\}_{k\in K''} \to \mathbf{x}^{**} \in \mathcal{F}_P^*$ for some infinite index set $K'' \subseteq K'$, and in view of Lemma 6(ii) $\{\tilde{\boldsymbol{\lambda}}^{k-1}\}_{k\in K''} \to \boldsymbol{\lambda}^{**}$, where $\boldsymbol{\lambda}^{**}$ is the KKT multiplier associated to $\mathbf{x}^{**}$. Since $\alpha_{\mathrm{p}}^k \in [0,1]$ for all k, we also have $\{\mathbf{x}^{k-1}\}_{k\in K''} = \{\mathbf{x}^{k-2} + \alpha_{\mathrm{p}}^{k-2}\Delta\mathbf{x}^{k-2}\}_{k\in K''} \to \mathbf{x}^{**}$, i.e., $\{(\mathbf{x}^{k-1}, \tilde{\boldsymbol{\lambda}}^{k-1})\}_{k\in K''} \to (\mathbf{x}^{**}, \boldsymbol{\lambda}^{**}) \in \mathcal{F}^*$, completing the proof. □

Proof of Theorem 1. Claim (i) was proved in Lemma 10 and Claim (ii) in Lemma 12, Claim (iii) is a direct consequence of Condition CSR(ii), and Claim (iv) was proved in Lemma 13.

Proof of Corollary 2. From Theorem 1, $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\} \to (\mathbf{x}^*, \boldsymbol{\lambda}^*)$, i.e., $\{E_k\} \to 0$. It follows that (i) in view of Proposition 1 and Condition CSR(ii) $Q_k \supseteq \mathcal{A}(x^*)$ for all k large enough, and (ii) in view of Rule R, $\{\delta_k\} \to 0$, so that Q_k eventually excludes all indexes that are not in $\mathcal{A}(\mathbf{x}^*)$. □

B Proof of Theorem 2

Parts of this proof are adapted from [16,29,34]. Throughout, we assume that Assumption 3 holds (so that Assumption 1 also holds), that $\varepsilon = 0$ and that the iteration never stops, and that $\lambda_i^* < \lambda^{\max}$ for all i.

Newton's method plays the central role in the local analysis. The following lemma is standard or readily proved; see, e.g., [29, Proposition 3.10].

Lemma 14 *Let $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ be twice continuously differentiable and let $\mathbf{t}^* \in \mathbb{R}^n$ such that $\Phi(\mathbf{t}^*) = \mathbf{0}$. Suppose there exists $\rho > 0$ such that $\frac{\partial \Phi}{\partial \mathbf{t}}(\mathbf{t})$ is non-singular for all $\mathbf{t} \in B(\mathbf{t}^*, \rho)$. Define $\Delta^{\mathrm{N}}\mathbf{t}$ to be the Newton increment at $\mathbf{t}$, i.e., $\Delta^{\mathrm{N}}\mathbf{t} = -\left(\frac{\partial \Phi}{\partial \mathbf{t}}(\mathbf{t})\right)^{-1}\Phi(\mathbf{t})$. Then, given any $c > 0$, there exists $c^* > 0$ such that, for all $\mathbf{t} \in B(\mathbf{t}^*, \rho)$, if $\mathbf{t}^+ \in \mathbb{R}^n$ satisfies*

$$\min\{|t_i^+ - t_i^*|, |t_i^+ - (t_i + (\Delta^{\mathrm{N}} t)_i)|\} \leq c \max\{\|\Delta^{\mathrm{N}}\mathbf{t}\|^2, \|\mathbf{t} - \mathbf{t}^*\|^2\}, \quad i = 1, \ldots, n\,, \tag{64}$$

then

$$\|\mathbf{t}^+ - \mathbf{t}^*\| \leq c^* \|\mathbf{t} - \mathbf{t}^*\|^2\,.$$

For convenience, define $\mathbf{z} := (\mathbf{x}, \boldsymbol{\lambda})$ (as well as $\mathbf{z}^* := (\mathbf{x}^*, \boldsymbol{\lambda}^*)$, etc.). For $\mathbf{z} \in \mathcal{F}^o := \{\mathbf{z} : \mathbf{x} \in \mathcal{F}_P^o, \boldsymbol{\lambda} > \mathbf{0}\}$, define

$$\varrho(\mathbf{z}) := \min\left\{1, \frac{E(\mathbf{x}, \boldsymbol{\lambda})}{\bar{E}}\right\} \quad \text{and} \quad W(\mathbf{z}) := H + \varrho(\mathbf{z})R\,.$$

The gist of the remainder of this appendix is to apply Lemma 14 to

$$\Phi_Q(\mathbf{z}) := \begin{bmatrix} H\mathbf{x} - (A_Q)^T\boldsymbol{\lambda}_Q + \mathbf{c} \\ \Lambda_Q(A_Q\mathbf{x} - \mathbf{b}_Q) \end{bmatrix}, \quad Q \subseteq \mathbf{m}.$$

(Note that $\Phi_Q(\mathbf{z}^*) = \mathbf{0}$.) Let $\mathbf{z}_Q := (\mathbf{x}, \boldsymbol{\lambda}_Q)$, then the step taken on the Q components along the search direction generated by the Algorithm CR-MPC is analogously given by $\breve{\mathbf{z}}_Q^+ := (\mathbf{x}^+, \breve{\boldsymbol{\lambda}}_Q^+)$ with $\breve{\boldsymbol{\lambda}}_Q^+ := \boldsymbol{\lambda}_Q + \alpha_{\mathrm{d}}\Delta\boldsymbol{\lambda}_Q$. The first major step of the proof is achieved by Proposition 4 below, where the focus is on $\breve{\mathbf{z}}_Q^+$ rather than on $\mathbf{z}^+$. Thus we compare $\breve{\mathbf{z}}_Q^+$, with $Q \in \mathcal{Q}^*$ to the Q components of the (unregularized) Newton step, i.e., $\mathbf{z}_Q + (\Delta^{\mathrm{N}}\mathbf{z})_Q$. Define

$$\mathscr{A} := \begin{bmatrix} \alpha_{\mathrm{p}} I_n & \mathbf{0} \\ \mathbf{0} & \alpha_{\mathrm{d}} I_{|Q|} \end{bmatrix}, \quad \text{and} \quad \alpha := \min\{\alpha_{\mathrm{p}}, \alpha_{\mathrm{d}}\}\,.$$

The difference between the CR-MPC iteration and the Newton iteration can be written as

$$\begin{aligned} &\|\breve{\mathbf{z}}_Q^+ - (\mathbf{z}_Q + (\Delta^{\mathrm{N}}\mathbf{z})_Q)\| \\ &\leq \|\breve{\mathbf{z}}_Q^+ - (\mathbf{z}_Q + \Delta\mathbf{z}_Q)\| + \|\Delta\mathbf{z}_Q - \Delta\mathbf{z}_Q^{\mathrm{a}}\| + \|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^0\| + \|\Delta\mathbf{z}_Q^0 - (\Delta^{\mathrm{N}}\mathbf{z})_Q\| \\ &= \|(I - \mathscr{A})\Delta\mathbf{z}_Q\| + \gamma\|\Delta\mathbf{z}_Q^{\mathrm{c}}\| + \|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^0\| + \|\Delta\mathbf{z}_Q^0 - (\Delta^{\mathrm{N}}\mathbf{z})_Q\| \\ &\leq (1 - \alpha)\|\Delta\mathbf{z}_Q\| + \|\Delta\mathbf{z}_Q^{\mathrm{c}}\| + \|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^0\| + \|\Delta\mathbf{z}_Q^0 - (\Delta^{\mathrm{N}}\mathbf{z})_Q\|\,, \end{aligned} \tag{65}$$

where $\Delta\mathbf{z}_Q := (\Delta\mathbf{x}, \Delta\boldsymbol{\lambda}_Q)$, $\Delta\mathbf{z}_Q^{\mathrm{a}} := (\Delta\mathbf{x}^{\mathrm{a}}, \Delta\boldsymbol{\lambda}_Q^{\mathrm{a}})$, $\Delta\mathbf{z}_Q^{\mathrm{c}} := (\Delta\mathbf{x}^{\mathrm{c}}, \Delta\boldsymbol{\lambda}_Q^{\mathrm{c}})$, and $\Delta\mathbf{z}_Q^0$ is the (constraint-reduced) affine-scaling direction for the original (unregularized) system (so $\Delta^{\mathrm{N}}\mathbf{z} = \Delta\mathbf{z}_{\mathbf{m}}^0$).

Let

$$J_a(W, A, \mathbf{s}, \boldsymbol{\lambda}) := \begin{bmatrix} W & -A^T \\ \Lambda A & S \end{bmatrix}\,.$$

The following readily proved lemma will be of help. (For details, see Lemmas B.15 and B.16 in [16]; also Lemmas 13 and 1 in [28])

Lemma 15 *Let $\mathbf{s}, \boldsymbol{\lambda} \in \mathbb{R}^m$ and $Q \subseteq \mathbf{m}$ be arbitrary and let W be symmetric, with $W \succeq H$. Then (i) $J_a(W, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$ is non-singular if and only if $J(W, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$ is, and (ii) if $\mathcal{A}(\mathbf{x}^*) \subseteq Q$, then $J(W, A_Q, \mathbf{s}_Q^*, \boldsymbol{\lambda}_Q^*)$ is non-singular (and so is $J_a(W, A_Q, \mathbf{s}_Q^*, \boldsymbol{\lambda}_Q^*)$).*

With $\mathbf{s} := A\mathbf{x} - \mathbf{b}$, $J_a(H, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$, the system matrix for the (constraint-reduced) original (unregularized) "augmented" system, is the Jacobian of $\Phi_Q(\mathbf{z})$, i.e.,

$$J_a(H, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)\Delta\mathbf{z}_Q^0 = -\Phi_Q(\mathbf{z}) ,$$

and its regularized version $J_a(W(\mathbf{z}), A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$ satisfies (among other systems solved by Algorithm CR-MPC)

$$J_a(W(\mathbf{z}), A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)\Delta\mathbf{z}_Q^{\mathrm{a}} = -\Phi_Q(\mathbf{z}) .$$

Next, we verify that $J_a(W(\mathbf{z}), A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)$ is non-singular near $\mathbf{z}^*$ (so that $\Delta\mathbf{z}_Q^0$ and $\Delta\mathbf{z}_Q^{\mathrm{a}}$ in (65) are well defined) and establish other useful local properties. For convenience, we define

$$\mathcal{Q}^* := \{Q \subseteq \mathbf{m} : \mathcal{A}(\mathbf{x}^*) \subseteq Q\} .$$

and

$$\tilde{\mathbf{s}}^+ := \mathbf{s} + \Delta\mathbf{s}, \qquad \tilde{\mathbf{s}}^{\mathrm{a},+} := \mathbf{s} + \Delta\mathbf{s}^{\mathrm{a}}.$$

Lemma 16 *Let $\epsilon^* := \min\{1, \min_{i\in\mathbf{m}}(\lambda_i^* + s_i^*)\}$. There exist $\rho^* > 0$ and $r > 0$, such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$ and all $Q \in \mathcal{Q}^*$, the following hold:*

(i) $\|J_a(W(\mathbf{z}), A_Q, \boldsymbol{\lambda}_Q, \mathbf{s}_Q)^{-1}\| \leq r$,
(ii) $\max\{\|\Delta\mathbf{z}_Q^{\mathrm{a}}\|, \|\Delta\mathbf{z}_Q\|, \|\Delta\mathbf{s}_Q^{\mathrm{a}}\|, \|\Delta\mathbf{s}_Q\|\} < \epsilon^*/4$,
(iii) $\min\{\lambda_i, \tilde{\lambda}_i^{\mathrm{a},+}, \tilde{\lambda}_i^+\} > \epsilon^*/2$, $\forall i \in \mathcal{A}(\mathbf{x}^*)$,
$\max\{\lambda_i, \tilde{\lambda}_i^{\mathrm{a},+}, \tilde{\lambda}_i^+\} < \epsilon^*/2$, $\forall i \in \mathbf{m} \setminus \mathcal{A}(\mathbf{x}^*)$,
$\max\{s_i, \tilde{s}_i^{\mathrm{a},+}, \tilde{s}_i^+\} < \epsilon^*/2$, $\forall i \in \mathcal{A}(\mathbf{x}^*)$,
$\min\{s_i, \tilde{s}_i^{\mathrm{a},+}, \tilde{s}_i^+\} > \epsilon^*/2$, $\forall i \in \mathbf{m} \setminus \mathcal{A}(\mathbf{x}^*)$.
(iv) $\tilde{\lambda}_i^+ < \lambda^{\max}$, $\forall i \in \mathbf{m}$.

Proof Claim (i) follows from Lemma 15, continuity of $J_a(W(\mathbf{z}), A_Q, \boldsymbol{\lambda}_Q, \mathbf{s}_Q)$ (and the fact that $W(\mathbf{z}^*) = H$). Claims (ii) and (iv) follow from Claim (i), Lemma 15, continuity of the right-hand sides of (10) and (32), which are zero at the solution, definition (34) of $\tilde{\boldsymbol{\lambda}}^+$, and our assumption that $\lambda_i^* < \lambda^{\max}$ for all $i \in \mathbf{m}$. Claim (iii) is true due to strict complementary slackness, the definition of ϵ^*, and Claim (ii). □

In preparation for Proposition 4, Lemmas 17–20 provide bounds on the four terms in the last line of (65). The ρ^* used in these lemmas comes from Lemma 16. The proofs of Lemmas 17, 18, and 20 are omitted, as they are very similar to those of Lemmas A.9 and A.10 in the supplementary materials of [34] (where an MPC algorithm for linear optimization problems is considered) and of Lemma B.19 in [16] (also Lemma 16 in [28]).

Lemma 17 *There exists a constant $c_1 > 0$ such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$, and for all $Q \in \mathcal{Q}^*$,*

$$\|\Delta\mathbf{z}_Q^{\mathrm{c}}\| \leq c_1\|\Delta\mathbf{z}_Q^{\mathrm{a}}\|^2 .$$

Note that an upper bound on the magnitude of the MPC search direction $\Delta\mathbf{z}_Q$ can be obtained by using Lemma 17 and Lemma 16(ii), viz.

$$\|\Delta\mathbf{z}_Q\| \leq \|\Delta\mathbf{z}_Q^{\mathrm{a}}\| + \|\Delta\mathbf{z}_Q^{\mathrm{c}}\| \leq \|\Delta\mathbf{z}_Q^{\mathrm{a}}\| + c_1\|\Delta\mathbf{z}_Q^{\mathrm{a}}\|^2 \leq \left(1 + c_1\frac{\epsilon^*}{4}\right)\|\Delta\mathbf{z}_Q^{\mathrm{a}}\| . \qquad (66)$$

This bound is used in the proofs of Lemma 18 and Proposition 4.

Lemma 18 *There exists a constant $c_2 > 0$ such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$, and for all $Q \in \mathcal{Q}^*$,*

$$|1 - \alpha| \leq c_2\|\Delta\mathbf{z}_Q^{\mathrm{a}}\| .$$

Lemma 19 *There exists a constant $c_3 > 0$ such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$ and all $Q \in \mathcal{Q}^*$,*

$$\|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^0\| \leq c_3\|\mathbf{z} - \mathbf{z}^*\|^2.$$

Proof We have

$$\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^{0} = -(J_a(W(\mathbf{z}), A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)^{-1} - J_a(H, A_Q, \mathbf{s}_Q, \boldsymbol{\lambda}_Q)^{-1})\Phi_Q(\mathbf{z})$$

so that there exist $c_{31} > 0$ such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$ and all $Q \in \mathcal{Q}^*$,

$$\|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^{0}\| \le c_{31}\|W(\mathbf{z}) - H\|\|\mathbf{z} - \mathbf{z}^*\| \,,$$

where the second inequality follows from Lemma 16(i). Since $W(\mathbf{z}) - H = \varrho(\mathbf{z})R$, $|\varrho(\mathbf{z})| \le c_{32}|E(\mathbf{z})|$, and $|E(\mathbf{z})| \le c_{33}\|\mathbf{z} - \mathbf{z}^*\|$, for some $c_{32} > 0$ and $c_{33} > 0$, the proof is complete. □

Lemma 20 *There exists a constant $c_4 > 0$ such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$, and for all $Q \in \mathcal{Q}^*$,*

$$\|\Delta\mathbf{z}_Q^{0} - (\Delta^{\mathrm{N}}\mathbf{z})_Q\| \le c_4\|\mathbf{z} - \mathbf{z}^*\|\|(\Delta^{\mathrm{N}}\mathbf{z})_Q\| \,.$$

With Lemmas 17–20 in hand, we return to inequality (65).

Proposition 4 *There exists a constant $c_5 > 0$ such that, for all $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$, and for all $Q \in \mathcal{Q}^*$,*

$$\|\breve{\mathbf{z}}_Q^{+} - (\mathbf{z}_Q + (\Delta^{\mathrm{N}}\mathbf{z})_Q)\| \le c_5 \max\{\|\Delta^{\mathrm{N}}\mathbf{z}\|^2, \|\mathbf{z} - \mathbf{z}^*\|^2\} \,. \tag{67}$$

Proof Let $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$ and $Q \in \mathcal{Q}^*$. It follows from (65), Lemmas 17–20, and (66) that

$$\begin{aligned}\|\breve{\mathbf{z}}_Q^{+} - (\mathbf{z}_Q + (\Delta^{\mathrm{N}}\mathbf{z})_Q)\| &\le (1-\alpha)\|\Delta\mathbf{z}_Q\| + \|\Delta\mathbf{z}_Q^{\mathrm{c}}\| + \|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^{0}\| + \|\Delta\mathbf{z}_Q^{0} - (\Delta^{\mathrm{N}}\mathbf{z})_Q\| \\ &\le c_2\|\Delta\mathbf{z}_Q^{\mathrm{a}}\|\|\Delta\mathbf{z}_Q\| + c_1\|\Delta\mathbf{z}_Q^{\mathrm{a}}\|^2 + c_3\|\mathbf{z} - \mathbf{z}^*\|^2 + c_4\|\mathbf{z} - \mathbf{z}^*\|\|(\Delta^{\mathrm{N}}\mathbf{z})_Q\| \\ &\le \left(c_2\left(1 + c_1\frac{\epsilon^*}{4}\right) + c_1\right)\|\Delta\mathbf{z}_Q^{\mathrm{a}}\|^2 + c_3\|\mathbf{z} - \mathbf{z}^*\|^2 + c_4\|\mathbf{z} - \mathbf{z}^*\|\|(\Delta^{\mathrm{N}}\mathbf{z})_Q\| \,.\end{aligned}$$

Also, by Lemmas 19 and 20, we have

$$\begin{aligned}\|\Delta\mathbf{z}_Q^{\mathrm{a}}\| &\le \|\Delta\mathbf{z}_Q^{\mathrm{a}} - \Delta\mathbf{z}_Q^{0}\| + \|\Delta\mathbf{z}_Q^{0} - (\Delta^{\mathrm{N}}\mathbf{z})_Q\| + \|(\Delta^{\mathrm{N}}\mathbf{z})_Q\| \\ &\le c_3\|\mathbf{z} - \mathbf{z}^*\|^2 + c_4\|\mathbf{z} - \mathbf{z}^*\|\|(\Delta^{\mathrm{N}}\mathbf{z})_Q\| + \|(\Delta^{\mathrm{N}}\mathbf{z})_Q\| \,.\end{aligned} \tag{68}$$

The claim follows (in view of boundedness of $\mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$). □

With Proposition 4 established, we proceed to the second major step of the proof of Theorem 2: to show that (67) still holds when $\mathbf{z}^+$ is substituted for $\breve{\mathbf{z}}_Q^{+}$.

Proof of Theorem 2. Again, let ρ^* be as given in Lemma 16. Let $\mathbf{z} \in \mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$ and $Q \in \mathcal{Q}^*$. Let $\rho := \rho^*$, $\mathbf{t} := \mathbf{z}$, and $\mathbf{t}^* := \mathbf{z}^*$. Then the desired q-quadratic convergence is a direct consequence of Lemma 14, provided that the condition (64) is satisfied. Hence, we now show that there exists some constant $c > 0$ such that, for each $i \in \mathbf{m}$,

$$\min\{|z_i^+ - z_i^*|, |z_i^+ - (z_i + (\Delta^{\mathrm{N}}z)_i)|\} \le c \max\{\|\Delta^{\mathrm{N}}\mathbf{z}\|^2, \|\mathbf{z} - \mathbf{z}^*\|^2\} \,. \tag{69}$$

As per Proposition 4, (69) holds for $i \in Q$ with z_i^+ replaced with $\breve{z}_i^+$. In particular, (69) holds for the $\mathbf{x}^+$ components of $\mathbf{z}^+$. It remains to show that (69) holds for the $\boldsymbol{\lambda}^+$ components of $\mathbf{z}^+$. Firstly, for all $i \in \mathcal{A}(\mathbf{x}^*)$, we show that $\lambda_i^+ = \breve{\lambda}_i^+$, thus (69) holds for all λ_i^+ such that $i \in \mathcal{A}(\mathbf{x}^*)$ by Proposition 4. From the fact that $\boldsymbol{\lambda} > \mathbf{0}$ ($\mathbf{z} \in \mathcal{F}^o$) and Lemma 16(ii), and since $\nu \ge 2$, it follows that

$$\chi := \|\Delta\mathbf{x}^{\mathrm{a}}\|^\nu + \|[\tilde{\boldsymbol{\lambda}}_Q^{\mathrm{a},+}]_-\|^\nu \le \|\Delta\mathbf{x}^{\mathrm{a}}\|^\nu + \|\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}\|^\nu \le 2\left(\frac{\epsilon^*}{4}\right)^\nu \le \frac{\epsilon^*}{2} \,, \tag{70}$$

so that $\min\{\chi, \underline{\lambda}\} \leq \epsilon^*/2$. Also, from Lemma 16(iii) and the fact that $\breve{\boldsymbol{\lambda}}_Q^+$ is a convex combination of $\boldsymbol{\lambda}_Q$ and $\tilde{\boldsymbol{\lambda}}_Q^+$, we have, for all $i \in \mathcal{A}(\mathbf{x}^*)$,

$$\frac{\epsilon^*}{2} < \min\{\lambda_i, \tilde{\lambda}_i^+\} \leq \breve{\lambda}_i^+ \,. \tag{71}$$

Hence, from (70), (71), Lemma 16(iv), and (26), we conclude that $\lambda_i^+ = \breve{\lambda}_i^+$ for all $i \in \mathcal{A}(\mathbf{x}^*)$. Secondly, we prove that there exists $d_1 > 0$ such that

$$\|\boldsymbol{\lambda}^+_{Q\setminus\mathcal{A}(\mathbf{x}^*)}\| = \|\boldsymbol{\lambda}^+_{Q\setminus\mathcal{A}(\mathbf{x}^*)} - \boldsymbol{\lambda}^*_{Q\setminus\mathcal{A}(\mathbf{x}^*)}\| \leq d_1 \max\{\|\Delta^{\mathrm{N}}\mathbf{z}\|^2, \|\mathbf{z}-\mathbf{z}^*\|^2\} \quad \forall i \in Q \setminus \mathcal{A}(\mathbf{x}^*) \,, \tag{72}$$

thus establishing (69) for λ_i^+ with $i \in Q \setminus \mathcal{A}(\mathbf{x}^*)$. For $i \in Q \setminus \mathcal{A}(\mathbf{x}^*)$, we know from (26) that, either $\lambda_i^+ = \min\{\lambda^{\max}, \breve{\lambda}_i^+\}$, or $\lambda_i^+ = \min\{\underline{\lambda}, \|\Delta\mathbf{x}^{\mathrm{a}}\|^\nu + \|[\tilde{\boldsymbol{\lambda}}_Q^{\mathrm{a},+}]_-\|^\nu\}$. In the former case, we have

$$\begin{aligned}|\lambda_i^+| \leq |\breve{\lambda}_i^+| = |\breve{\lambda}_i^+ - \lambda_i^*| &\leq |\breve{\lambda}_i^+ - (\lambda_i + (\Delta^{\mathrm{N}}\lambda)_i)| + |(\lambda_i + (\Delta^{\mathrm{N}}\lambda)_i) - \lambda_i^*| \\ &\leq d_2 \max\{\|\Delta^{\mathrm{N}}\mathbf{z}\|^2, \|\mathbf{z}-\mathbf{z}^*\|^2\} + d_3\|\mathbf{z}-\mathbf{z}^*\|^2 \,,\end{aligned}$$

for some $d_2 > 0$, $d_3 > 0$. Here the last inequality follows from Proposition 4 and the quadratic rate of the Newton step given in Lemma 14. In the latter case, since $\boldsymbol{\lambda} > \mathbf{0}$, we obtain

$$|\lambda_i^+| \leq \|\Delta\mathbf{x}^{\mathrm{a}}\|^\nu + \|[\tilde{\boldsymbol{\lambda}}_Q^{\mathrm{a},+}]_-\|^\nu \leq \|\Delta\mathbf{x}^{\mathrm{a}}\|^\nu + \|\Delta\boldsymbol{\lambda}_Q^{\mathrm{a}}\|^\nu = \|\Delta\mathbf{z}_Q^{\mathrm{a}}\|^\nu \leq d_4 \max\{\|\Delta^{\mathrm{N}}\mathbf{z}\|^2, \|\mathbf{z}-\mathbf{z}^*\|^2\}, \tag{73}$$

for some $d_4 > 0$. Here the equality is from the definition of $\Delta\mathbf{z}^{\mathrm{a}}$ and the last inequality follows from $\nu \geq 2$, (68), and boundedness of $\mathcal{F}^o \cap B(\mathbf{z}^*, \rho^*)$. Hence, we have established (72). Thirdly and finally, consider the case that $i \in Q^{\mathrm{c}}$. Since $\mathcal{A}(\mathbf{x}^*) \subseteq Q$, $\boldsymbol{\lambda}^*_{Q^{\mathrm{c}}} = \mathbf{0}$ and it follows from (27) that, either $\lambda_i^+ = \min\{\lambda^{\max}, \mu^+_{(Q)}/s_i^+\}$, or $\lambda_i^+ = \min\{\underline{\lambda}, \|\Delta\mathbf{x}^{\mathrm{a}}\|^\nu + \|[\tilde{\boldsymbol{\lambda}}_Q^{\mathrm{a}}]_-\|^\nu\}$. In the latter case, the bound in (73) follows. In the former case, we have

$$|\lambda_i^+ - \lambda_i^*| = |\lambda_i^+| \leq \mu^+_{(Q)}/s_i^+ \,.$$

By definition, $s_i^+ := s_i + \alpha_{\mathrm{p}}\Delta s_i$ is a convex combination of s_i and $\tilde{s}_i^+$. Thus, Lemma 16(iii) gives that $s_i^+ \geq \min\{s_i, \tilde{s}_i^+\} > \epsilon^*/2$. Then using the definition of $\mu^+_{(Q)}$ (see Step 10 of Algorithm CR-MPC) leads to

$$|\lambda_i^+ - \lambda_i^*| \leq \begin{cases} \frac{2}{\epsilon^*|Q|}\left((\mathbf{s}^+_{\mathcal{A}(\mathbf{x}^*)})^T(\boldsymbol{\lambda}^+_{\mathcal{A}(\mathbf{x}^*)}) + (\mathbf{s}^+_{Q\setminus\mathcal{A}(\mathbf{x}^*)})^T(\boldsymbol{\lambda}^+_{Q\setminus\mathcal{A}(\mathbf{x}^*)})\right), & \text{if } |Q| \neq 0 \\ 0\,, & \text{otherwise} \end{cases}.$$

Since $\mathbf{z} \in B(\mathbf{z}^*, \rho^*)$, $\boldsymbol{\lambda}^+_{\mathcal{A}(\mathbf{x}^*)}$ and $\mathbf{s}^+_{Q\setminus\mathcal{A}(\mathbf{x}^*)}$ are bounded by Lemma 16(ii). Also, by definition, $\mathbf{s}^*_{\mathcal{A}(\mathbf{x}^*)} = \mathbf{0}$. Thus there exist $d_5 > 0$ and $d_6 > 0$ such that

$$|\lambda_i^+ - \lambda_i^*| \leq d_5\|\mathbf{s}^+_{\mathcal{A}(\mathbf{x}^*)} - \mathbf{s}^*_{\mathcal{A}(\mathbf{x}^*)}\| + d_6\|\boldsymbol{\lambda}^+_{Q\setminus\mathcal{A}(\mathbf{x}^*)}\| \,.$$

Having already established that the second term is bounded by the right-hand side of (72), and we are left to prove that the first term also is. By definition,

$$\|\mathbf{s}^+_{\mathcal{A}(\mathbf{x}^*)} - \mathbf{s}^*_{\mathcal{A}(\mathbf{x}^*)}\| = \|A_{\mathcal{A}(\mathbf{x}^*)}\mathbf{x}^+ - A_{\mathcal{A}(\mathbf{x}^*)}\mathbf{x}^*\| \leq \|A\mathbf{x}^+ - A\mathbf{x}^*\| \leq \|A\|\|\breve{\mathbf{z}}_Q^+ - \mathbf{z}_Q^*\| \,.$$

Applying Proposition 4 and Lemma 14, we get

$$\begin{aligned}\|\mathbf{s}^+_{\mathcal{A}(\mathbf{x}^*)} - \mathbf{s}^*_{\mathcal{A}(\mathbf{x}^*)}\| &\leq \|A\|\|\breve{\mathbf{z}}_Q^+ - (\mathbf{z}_Q + (\Delta^{\mathrm{N}}\mathbf{z})_Q)\| + \|A\|\|(\mathbf{z}_Q + (\Delta^{\mathrm{N}}\mathbf{z})_Q) - \mathbf{z}_Q^*\| \\ &\leq d_7 \max\{\|\Delta^{\mathrm{N}}\mathbf{z}\|^2, \|\mathbf{z}-\mathbf{z}^*\|^2\} + d_8\|\mathbf{z}-\mathbf{z}^*\|^2 \,,\end{aligned}$$

for some $d_7 > 0$, $d_8 > 0$. Hence, we established (69) for all $i \in \mathbf{m}$, thus proving the q-quadratic convergence rate. □

References

1. Altman, A., Gondzio, J.: Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. Optim. Methods Softw. **11**(1-4), 275–302 (1999)
2. Bertsimas, D., Tsitsiklis, J.: Introduction to Linear Optimization. Athena (1997)
3. Cartis, C., Yan, Y.: Active-set prediction for interior point methods using controlled perturbations. Comput. Optim. Appl. **63**(3), 639–684 (2016)
4. Castro, J., Cuesta, J.: Quadratic regularizations in an interior-point method for primal block-angular problems. Math. Prog. **130**(2), 415–445 (2011)
5. Chen, L., Wang, Y., He, G.: A feasible active set QP-free method for nonlinear programming. SIAM J. Optimiz. **17**(2), 401–429 (2006)
6. Dantzig, G.B., Ye, Y.: A build-up interior-point method for linear programming: Affine scaling form. Tech. rep., University of Iowa, Iowa City, IA 52242, USA (July 1991)
7. Drummond, L., Svaiter, B.: On well definedness of the central path. J. Optim. Theory and Appl. **102**(2), 223–237 (1999)
8. Facchinei, F., Fischer, A., Kanzow, C.: On the accurate identification of active constraints. SIAM J Optimiz. **9**(1), 14–32 (1998)
9. Gill, P.E., Murray, W., Ponceleón, D.B., Saunders, M.A.: Solving reduced KKT systems in barrier methods for linear programming. In: G.A. Watson, D. Griffiths (eds.) Numerical Analysis 1993, pp. 89–104. Pitman Research Notes in Mathematics 303, Longmans Press (1994)
10. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: V. Blondel, S. Boyd, H. Kimura (eds.) Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited (2008). `http://stanford.edu/~boyd/graph_dcp.html`
11. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, Version 2.1. `http://cvxr.com/cvx` (2014)
12. Hager, W.W., Seetharama Gowda, M.: Stability in the presence of degeneracy and error estimation. Math. Prog. **85**(1), 181–192 (1999)
13. He, M.: Infeasible constraint reduction for linear and convex quadratic optimization. Ph.D. thesis, University of Maryland (2011). URL: `http://hdl.handle.net/1903/12772`
14. He, M.Y., Tits, A.L.: Infeasible constraint-reduced interior-point methods for linear optimization. Optim. Methods Softw. **27**(4-5), 801–825 (2012)
15. Hertog, D., Roos, C., Terlaky, T.: Adding and deleting constraints in the logarithmic barrier method for LP. In: D.Z. Du, J. Sun (eds.) Advances in Optimization and Approximation, pp. 166–185. Kluwer Academic Publishers, Dordrecht, The Netherlands (1994)
16. Jung, J.H.: Adaptive constraint reduction for convex quadratic programming and training support vector machines. Ph.D. thesis, University of Maryland (2008). URL: `http://hdl.handle.net/1903/8020`
17. Jung, J.H., O'Leary, D.P., Tits, A.L.: Adaptive constraint reduction for training support vector machines. Electron. T. Numer. Ana. **31**, 156–177 (2008)
18. Jung, J.H., O'Leary, D.P., Tits, A.L.: Adaptive constraint reduction for convex quadratic programming. Comput. Optim. Appl. **51**(1), 125 – 157 (2012)
19. Laiu, M.P.: Positive filtered P_N method for linear transport equations and the associated optimization algorithm. Ph.D. thesis, University of Maryland (2016). URL: `http://hdl.handle.net/1903/18732`
20. Laiu, M.P., Hauck, C.D., McClarren, R.G., O'Leary, D.P., Tits, A.L.: Positive filtered P_N moment closures for linear kinetic equations. SIAM J. Numer. Anal. **54**(6), 3214–3238 (2016)
21. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**(4), 575–601 (1992)
22. Nocedal, J., Wright, S.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering. Springer New York (2006)
23. Park, S.: A constraint-reduced algorithm for semidefinite optimization problems with superlinear convergence. J. Optimiz. Theory App. **170**(2), 512–527 (2016)

24. Park, S., O'Leary, D.P.: A polynomial time constraint-reduced algorithm for semidefinite optimization problems. J. Optimiz. Theory App. **166**(2), 558–571 (2015)
25. Saunders, M.A., Tomlin, J.A.: Solving regularized linear programs using barrier methods and KKT systems. Tech. rep., SOL 96-4. Department of Operations Research, Stanford University (1996)
26. Sturm, J.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optim. Methods Softw. **11–12**, 625–653 (1999). Version 1.05 available from `http://fewcal.kub.nl/sturm`
27. Tits, A., Wächter, A., Bakhtiari, S., Urban, T., Lawrence, C.: A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. SIAM J. Optimiz. **14**(1), 173–199 (2003)
28. Tits, A.L., Absil, P.A., Woessner, W.P.: Constraint reduction for linear programs with many inequality constraints. SIAM J. Optimiz. **17**(1), 119 – 146 (2006)
29. Tits, A.L., Zhou, J.L.: A simple, quadratically convergent algorithm for linear and convex quadratic programming. In: W. Hager, D. Hearn, P. Pardalos (eds.) Large Scale Optimization: State of the Art, pp. 411–427. Kluwer Academic Publishers (1994)
30. Toh, K.C., Todd, M.J., Tütüncü, R.H.: SDPT3 – A Matlab software package for semidefinite programming, Version 1.3. Optim. Methods Softw. **11**(1-4), 545–581 (1999)
31. Tone, K.: An active-set strategy in an interior point method for linear programming. Math. Prog. **59**(1), 345–360 (1993)
32. Tütüncü, R.H., Toh, K.C., Todd, M.J.: Solving semidefinite-quadratic-linear programs using SDPT3. Mathe. Prog. **95**(2), 189–217 (2003)
33. Winternitz, L.: Primal-dual interior-point algorithms for linear programming problems with many inequality constraints. Ph.D. thesis, University of Maryland (2010). URL: `http://hdl.handle.net/1903/10400`
34. Winternitz, L.B., Nicholls, S.O., Tits, A.L., O'Leary, D.P.: A constraint-reduced variant of Mehrotra's predictor-corrector algorithm. Comput. Optim. Appl. **51**(1), 1001 – 1036 (2012)
35. Winternitz, L.B., Tits, A.L., Absil, P.A.: Addressing rank degeneracy in constraint-reduced interior-point methods for linear optimization. J. Optimiz. Theory App. **160**(1), 127–157 (2014)
36. Wright, S.J.: Primal-Dual Interior-Point Methods. SIAM (1997)
37. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optimiz. **13**(2), 470–497 (2002)
38. Ye, Y.: A "build-down" scheme for linear programming. Math. Prog. **46**(1), 61–72 (1990)
39. Zhang, Y., Zhang, D.: On polynomiality of the Mehrotra-type predictor–corrector interior-point algorithms. Math. Prog. **68**(1), 303–318 (1995)