Optimization Methods and Software
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/goms20

Infeasible constraint-reduced interior-point methods for linear optimization

Meiyun Y. He [a] & Andréé L. Tits [a]
[a] Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD, USA

Available online: 14 Sep 2011

Optimization Methods & Software
iFirst, 2011, 1–25

Infeasible constraint-reduced interior-point methods for linear optimization

Meiyun Y. He and André L. Tits*

Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD, USA

(*Received 20 November 2010; final version received 9 May 2011*)

Constraint-reduction schemes have been proposed for the solution by means of interior-point methods of linear programs with many more inequality constraints than variables in the standard dual form. Such schemes have been shown to be provably convergent and highly efficient in practice. A critical requirement of these schemes is the availability of an initial dual-feasible point.

In this paper, building on a general framework (which encompasses several previously proposed approaches) for dual-feasible constraint-reduced interior-point optimization, for which we prove convergence to a single point of the sequence of dual iterates, we propose a framework for 'infeasible' constraint-reduced interior-point optimization. Central to this framework is an exact (ℓ_1 or ℓ_∞) penalty function scheme endowed with a mechanism for iterative adjustment of the penalty parameter, which aims at yielding, after a finite number of iterations, a value that guarantees feasibility (for the original problem) of the minimizers. Finiteness of the sequence of penalty parameter adjustments is proved under mild assumptions for all algorithms that fit within the framework, including 'infeasible' extensions of a 'dual' algorithm proposed in the early 1990s (Dantzig and Ye, *A build-up interior-point method for linear programming: Affine scaling form*, Working paper, Department of Management Science, University of Iowa, 1991) and of two recently proposed 'primal–dual' algorithms (Tits, Absil, and Woessner, *Constraint reduction for linear programs with many inequality constraints*, SIAM J. Optim. 17 (2006), pp. 119–146; Winternitz, Nicholls, Tits, and O'Leary, *A constraint-reduced variant of Mehrotra's predictor–corrector algorithm*, Comput. Optim. Appl. (published on-line as of January 2011), DOI: 10.1007/s10589-010-9389-4). The last one, a constraint-reduced variant of Mehrotra's Predictor–Corrector algorithm, is then more specifically considered: further convergence results are proved, and numerical results are reported that demonstrate that the approach is of practical interest.

Keywords: constraint reduction; infeasible initial points; interior-point methods; linear programming; many constraints

AMS Subject Classification: 90C51; 90C05; 90C06; 90C34; 65K05

1. Introduction

Consider a linear program (LP) in standard primal form,

$$\min c^{\mathrm{T}}x \quad \text{s.t. } Ax = b,\ x \geq 0 \tag{P}$$

*Corresponding author. Email: andre@umd.edu

ISSN 1055-6788 print/ISSN 1029-4937 online

DOI: 10.1080/10556788.2011.589056
http://www.informaworld.com

and its associated standard dual problem

$$\max b^{\mathrm{T}} y \quad \text{s.t. } A^{\mathrm{T}} y \leq c, \tag{D}$$

where matrix $A \in \mathbb{R}^{m \times n}$ has full row rank. When $n \gg m$, i.e. (D) involves many more constraints than variables, most constraints are inactive at the solution, and hence have no bearing on the solution. Such situations are detrimental to classical interior-point methods (IPMs), whose computational cost per iteration is typically proportional to n. Starting in the early 1990s, this has prompted a number of researchers to propose, analyse and test *constraint-reduced* versions of these methods. (See, e.g. [5,9,21,23,24]. The term 'constraint-reduced' was coined in [21].)

To the best of our knowledge, all existing constraint-reduced IPMs that are supported by analysis were obtained by grafting a constraint-reduction scheme onto a *dual-feasible* method. Accordingly, they all require a dual-feasible initial point. This is an important limitation because such point is often unavailable in practice, or may be available but poorly centred, resulting in slow progress of the algorithm. Attempts at combining constraint-reduction schemes with *infeasible* IPMs were made in [14,21] with Mehrotra's Predictor–Corrector (MPC) method [13], and in [14] with an algorithm from [16], with some numerical success; but no supporting analysis was provided, and indeed, it appears unlikely that these methods do enjoy guaranteed global convergence. In this paper, we show how the need to allow for infeasible initial points can be addressed by making use of an ℓ_1 or ℓ_∞ exact penalty function, with automatic adjustment of the penalty parameter. In related work on constraint-reduced IPMs for quadratic programming, the algorithm proposed in [11] does allow for initial infeasible points, handling them by means of an ℓ_1 exact penalty function; a convergence analysis is provided, but it assumes the *a priori* knowledge of an appropriate penalty parameter value; it does not include a scheme for determining such value.

Exact ℓ_1/ℓ_∞ penalty functions have been used in connection with IPMs in nonlinear programming [1,3,22], in particular on problems with complementarity constraints [4,12,19], and in at least one instance in linear programming [2]. The dearth of instances of use of penalty functions in linear programming is probably due to the availability of powerful algorithms, both of the simplex variety and of the interior-point variety, that accommodate infeasible initial points in a natural fashion, even guaranteeing polynomial complexity in the case of interior point, e.g. [15,16]. Combining such (possibly polynomially convergent) infeasible IPMs with constraint-reduction schemes has so far proved elusive though, and the use of exact penalty functions is a natural avenue to consider.

In this paper, as a first step, we consider a general framework (rIPM) for a class of *dual-feasible* constraint-reduced IPMs: those for which the dual objective monotonically increases. This framework encompasses, in particular, the algorithms proposed in [5,21,24]. We prove convergence to a single point of the sequence of dual iterates for all methods that fit within the framework. Second, as the main contribution of the paper, we expand this framework to allow for *dual-infeasible* initial points in the case of primal–dual interior-point methods (PDIPs); we dub the resulting framework IrPDIP. The expansion features an exact (ℓ_1 or ℓ_∞) penalty function and includes an iterative penalty adjustment scheme. The scheme is taken from [22], adapted to the linear-programming context, and augmented so as to enforce boundedness of the optimization iterates; in [22] (where no assumption of linearity or even convexity is made), such boundedness was merely *assumed*. The scheme used in [2] may be an alternative possibility, though we could not ascertain that boundedness of the sequence of penalty parameters would then be guaranteed. Under minimal assumptions (strict primal–dual feasibility), it is proved that the penalty parameter value is increased at most finitely many times, thus guaranteeing that the sequence of such values remains bounded. The proof departs significantly from that in [22], where strong non-degeneracy assumptions are invoked. Finally, we propose iteration IrMPC (*i*nfeasible, constraint-*r*educed, MPC), obtained by fitting into IrPDIP the dual-feasible constraint-reduced variant rMPC* proposed and analysed in [24]. We prove convergence to an optimal solution, starting from an arbitrary, possibly infeasible, initial point, and report promising numerical results.

The remainder of the paper is organized as follows. In Section 2, rIPM is laid out and analysed. In Section 3, rIPM is extended, by incorporating an exact penalty function, to allow for infeasible initial points in the case of constraint-reduced *primal–dual* interior point, producing IrPDIP, which is then analysed. In Section 4, IrPDIP is specialized to the case of algorithm rMPC* of [24] (a constraint-reduced variant of the MPC algorithm); the resulting algorithm is then analysed. Numerical results are reported in Section 5 and conclusions are given in Section 6.

The notation used in the paper is mostly standard. Absolute value, comparison and 'max' are meant componentwise. By e, we denote the vector of all ones with size by context. We adopt the Matlab-inspired notation $[v_1;v_2;\cdots;v_p]$ to denote a (vertical) concatenation of vectors (or matrices) v_i, $1 \le i \le p$. We denote a certain subset of $\mathbf{n} := \{1, 2, \ldots, n\}$ by Q and its complement by $\bar{Q} := \mathbf{n} \setminus Q$. Given an n-vector x, x^i is its ith element, and x^Q is a subvector of x with only those elements of x that are indexed in set Q. We denote by A^Q a submatrix of A with only those columns of A that are indexed in set Q. Given a diagonal matrix $X := \mathrm{diag}(x)$, we let $X^Q := \mathrm{diag}(x^Q)$. Except when specified, the norm $\|\cdot\|$ is arbitrary. The feasible set of the dual (D) is denoted by $\mathcal{F}$, i.e.

$$\mathcal{F} := \{y \in \mathbb{R}^m : A^{\mathrm{T}} y \le c\}.$$

The active set for (D) at point y (with y not necessarily in $\mathcal{F}$) is denoted by $I(y)$, i.e.

$$I(y) := \{i : (a^i)^{\mathrm{T}} y = c^i\}.$$

2. A framework for dual-feasible constraint-reduced IPMs

Many IPMs for the solution of (P)–(D), including the current 'champion', MPC [13], make use of an *affine scaling* direction Δy_a, solution of

$$ADA^{\mathrm{T}} \Delta y_a = b \tag{1}$$

for some diagonal positive-definite matrix D, usually updated from iteration to iteration. For such methods, when $n \gg m$, the main computational cost at each iteration resides in forming the matrix

$$ADA^{\mathrm{T}} = \sum_{i=1}^{n} d^i a^i (a^i)^{\mathrm{T}}, \tag{2}$$

where d^i is the ith diagonal entry of D and a^i the ith column of A. Forming ADA^{T} takes up roughly nm^2 multiplications and as many additions. If the sum on the right-hand side of (2) is reduced by dropping all terms except those associated with a certain small working index set Q, the cost of forming it reduces from nm^2 to roughly $|Q|m^2$. Conceivably, the cardinality $|Q|$ of Q could be as small as m in nondegenerate situations, leading to a potential computational speedup factor of n/m. Ideas along these lines are explored in [5,9,21,23,24] where schemes are proposed that enjoy strong theoretical properties and work well in practice. (Interestingly, in many cases, it has been observed that using an astutely selected small working set does not significantly increase the total number of iterations required to solve the problem, and sometimes even reduces it.) Several of these methods [5,21,24] fit within the following general iteration framework.

Iteration rIPM (constraint-reduced interior-point method)

Parameters: $\theta \in (0, 1)$ and $\tau > 0$.

Data: $y \in \mathbb{R}^m$ such that $s := c - A^{\mathrm{T}}y > 0$; $Q \subseteq \mathbf{n}$ such that A^Q has full row rank; $D \in \mathbb{R}^{|Q|\times|Q|}$, diagonal and positive definite.
Step 1: Computation of the dual search direction.
(i) Let Δy_a solve

$$A^Q D(A^Q)^{\mathrm{T}} \Delta y_a = b. \tag{3}$$

(ii) Select Δy to satisfy

$$b^{\mathrm{T}} \Delta y \geq \theta b^{\mathrm{T}} \Delta y_a, \quad \|\Delta y\| \leq \tau \|\Delta y_a\|. \tag{4}$$

Step 2: Updates
(i) Update the dual variables by choosing a stepsize $t \in (0, 1]$ such that

$$s_+ := c - A^{\mathrm{T}} y_+ > 0,$$

where

$$y_+ := y + t\Delta y. \tag{5}$$

(ii) Pick $Q_+ \subseteq \mathbf{n}$ such that A^{Q_+} has full row rank.
(iii) Select $D_+ \in \mathbb{R}^{|Q_+|\times|Q_+|}$, diagonal and positive definite.

Since A^Q has full row rank, the linear system (3) has a unique solution. Hence Iteration rIPM is well defined and, since $s_+ > 0$, it can be repeated indefinitely to generate infinite sequences. We attach subscript k to denote the kth iterate. Since $s_k > 0$ for all k, it also follows from (3) that

$$b^{\mathrm{T}} \Delta y_{a,k} > 0, \tag{6}$$

and further from (4) and (5) that the sequence $\{b^{\mathrm{T}} y_k\}$ is increasing.

An important property of Iteration rIPM, established in Proposition 2.2, is that if the dual-feasible sequence $\{y_k\}$ remains bounded, then it must converge, and if it is unbounded, then $b^{\mathrm{T}} y_k \to +\infty$. The proof makes use of the following lemma, a direct consequence of results in [18] (see also [17]).

LEMMA 2.1 *Let G be a full row rank matrix and b be in the range of G. Then,* (*i*) *there exists* $\phi > 0$ (*depending only on G and b*) *such that, given any positive-definite diagonal matrix D, the solution Δy to*

$$GDG^{\mathrm{T}} \Delta y = b,$$

satisfies

$$\|\Delta y\| \leq \phi b^{\mathrm{T}} \Delta y;$$

and (*ii*) *if a sequence $\{y_k\}$ is such that $\{b^{\mathrm{T}} y_k\}$ is bounded and, for some $\omega > 0$, satisfies*

$$\|y_{k+1} - y_k\| \leq \omega b^{\mathrm{T}} (y_{k+1} - y_k) \quad \forall k, \tag{7}$$

then $\{y_k\}$ converges.

Proof The first claim immediately follows from Theorem 5 in [18], noting (as in [17], Section 4) that, for some $\zeta > 0$, $\zeta\,\Delta y$ solves

$$\max\{b^{\mathrm{T}}u \mid u^{\mathrm{T}}GDG^{\mathrm{T}}u \leq 1\}.$$

(See also Theorem 7 in [17].) The second claim is proved using the central argument of the proof of Theorem 9 in [18]:

$$\sum_{k=0}^{N-1}\|y_{k+1}-y_k\| \leq \omega\sum_{k=0}^{N-1} b^{\mathrm{T}}(y_{k+1}-y_k) = \omega b^{\mathrm{T}}(y_N-y_0) \leq 2\omega v \quad \forall N > 0,$$

where v is a bound on $\{|b^{\mathrm{T}}y_k|\}$, implying that $\{y_k\}$ is Cauchy, and thus converges. (See also Theorem 9 in [17].) ■

PROPOSITION 2.2 *Suppose* (D) *is strictly feasible. Then, if* $\{y_k\}$ *generated by Iteration rIPM is bounded, then* $y_k \to y_*$ *for some* $y_* \in \mathcal{F}$, *and if it is not, then* $b^{\mathrm{T}}y_k \to \infty$.

Proof We first show that $\{y_k\}$ satisfies (7) for some $\omega > 0$. In view of (5), it suffices to show that, for some $\omega > 0$,[1]

$$\|\Delta y_k\| \leq \omega b^{\mathrm{T}}\Delta y_k \quad \forall k. \tag{8}$$

Now, since $\Delta y_{a,k}$ solves (3) and since A^{Q_k} has full row rank, and $Q_k \subseteq \mathbf{n}$, a finite set, it follows from Lemma 2.1 (i) that, for some $\phi > 0$,

$$\|\Delta y_{a,k}\| \leq \phi b^{\mathrm{T}}\Delta y_{a,k} \quad \forall k.$$

With this in hand, we obtain, using (4),

$$\|\Delta y_k\| \leq \tau\|\Delta y_{a,k}\| \leq \tau\phi b^{\mathrm{T}}\Delta y_{a,k} \leq \tau\frac{\phi}{\theta}b^{\mathrm{T}}\Delta y_k \quad \forall k,$$

so (8) holds with $\omega := \tau\phi/\theta$. Hence (7) holds (with the same ω).

To complete the proof, first suppose that $\{y_k\}$ is bounded. Then so is $\{b^{\mathrm{T}}y_k\}$ and, in view of Lemma 2.1(ii) and of the fact that $\{y_k\}$ is feasible, we have $y_k \to y_*$, for some $y_* \in \mathcal{F}$. On the other hand, if $\{y_k\}$ is unbounded, then $\{b^{\mathrm{T}}y_k\}$ is also unbounded (since, in view of Lemma 2.1(ii), having $\{b^{\mathrm{T}}y_k\}$ bounded together with (7) would lead to the contradiction that the unbounded sequence $\{y_k\}$ converges). Since $\{b^{\mathrm{T}}y_k\}$ is nondecreasing, the claim follows. ■

The 'build-up' algorithm in [5], and algorithms rPDAS in [21] and rMPC* in [24], all fit within the rIPM framework. In [5], D is $\mathrm{diag}(s^Q)^{-2}$, and in rPDAS and rMPC*, D is $\mathrm{diag}((x^i/s^i)_{i\in Q})$. In [5] and rPDAS, Δy is Δy_a, and in rMPC*, Δy satisfies (4) with $\tau = 1+\psi$, where $\psi > 0$ is a parameter of rMPC*. Hence, Proposition 2.2 provides a simpler proof for the convergence of dual sequence $\{y_k\}$ of [5] than that used in proving Theorem 3 of that paper; it strengthens the convergence result for rPDAS (Theorem 12 in [21]) by establishing convergence of the dual sequence to a single optimal point; and it is used in [24]. Proposition 2.2 is also used in the next section, in the analysis of the expanded framework IrPDIP (see Proposition 3.5).

3. A framework for infeasible constraint-reduced PDIPs

3.1 *Basic ideas and algorithm statement*

The *primal–dual* affine-scaling direction for dual-feasible constraint-reduced problem

$$\begin{aligned} \max \quad & b^{\mathrm{T}} y \\ \text{s.t.} \quad & (A^Q)^{\mathrm{T}} y \leq c^Q \end{aligned}$$

is the solution $(\Delta x^Q, \Delta y_a, \Delta s^Q)$ (when it exists) to the linear system

$$\begin{bmatrix} 0 & (A^Q)^{\mathrm{T}} & I \\ A^Q & 0 & 0 \\ S^Q & 0 & X^Q \end{bmatrix} \begin{bmatrix} \Delta x^Q \\ \Delta y_a \\ \Delta s^Q \end{bmatrix} = \begin{bmatrix} 0 \\ b - A^Q x^Q \\ -X^Q s^Q \end{bmatrix}, \tag{9}$$

where $S := \mathrm{diag}(s)$ and $X := \mathrm{diag}(x)$. Gaussian elimination of Δx^Q and Δs^Q yields (1) with $D := (S^Q)^{-1} X^Q$.

Previously proposed constraint-reduced IPMs [5,9,21,23,24,26] require a strictly dual-feasible initial point. In this section, we show how this limitation can be circumvented with the help of an ℓ_1 or ℓ_∞ exact penalty function. Specifically, in the ℓ_1 case, we consider relaxing (D) with

$$\begin{aligned} \max_{y,z} \quad & b^{\mathrm{T}} y - \rho e^{\mathrm{T}} z \\ \text{s.t.} \quad & A^{\mathrm{T}} y - z \leq c, \; z \geq 0, \end{aligned} \tag{D_ρ}$$

where $\rho > 0$ is a scalar penalty parameter, with associated 'primal'

$$\left.\begin{aligned} \min_{x,u} \quad & c^{\mathrm{T}} x \\ \text{s.t.} \quad & Ax = b, \; x + u = \rho e, \\ & x \geq 0, \; u \geq 0. \end{aligned}\right\} \tag{P_ρ}$$

Strictly feasible initial points for (D_ρ) are trivially available, and any of the algorithms just mentioned can be used to solve this primal–dual pair. It is well known (e.g. Theorem 40 in [6]) that there exists a threshold value ρ_* such that for any $\rho > \rho_*$, if (y_*^ρ, z_*^ρ) solves (D_ρ), then y_*^ρ solves (D) and $z_*^\rho = 0$. But such ρ_* is not known *a priori*.

We propose a scheme inspired from that used in [22] (in a nonlinear optimization context) for iteratively identifying an appropriate value for ρ. While, in contrast with the present situation, in [22], the penalty scheme is used to eliminate equality constraints, the corresponding transformation does encompass the transformation of (D) into (D_ρ): simply consider the intermediate problem

$$\begin{aligned} \max_{y,z} \quad & b^{\mathrm{T}} y \\ \text{s.t.} \quad & A^{\mathrm{T}} y - z \leq c, \quad z = 0. \end{aligned}$$

A key difference between [22] and the present context however is that, unlike that of [22] (see Lemma 4.1 and Proposition 4.2 in that paper), our scheme requires no *a priori* assumption on the boundedness of the sequences of iterates (y_k in our case, x_k in [22]). As seen from the toy example

$$\begin{aligned} \max \quad & y \\ \text{s.t.} \quad & y \leq 0,\ 2y \leq 2, \end{aligned} \tag{10}$$

when too small a value of ρ is used, such boundedness is not guaranteed. Indeed, the penalized problem associated to (10) is

$$\begin{aligned} \max_{y,z} \quad & y - \rho z^1 - \rho z^2 \\ \text{s.t.} \quad & y - z^1 \leq 0,\ 2y - z^2 \leq 2,\ z^1 \geq 0,\ z^2 \geq 0, \end{aligned}$$

or equivalently,

$$\min\{-y + \rho \max\{0, y\} + 2\rho \max\{0, y-1\}\}, \tag{11}$$

and as seen from Figure 1, when $\rho < \frac{1}{3}$, problem (11) is unbounded, even though problem (10) is bounded.

In the ℓ_1 version of our proposed scheme, the penalty parameter ρ is increased if either

$$\|z_+\| \geq \gamma_1 \frac{\|z_0\|}{\rho_0} \rho \tag{12}$$

or

$$\text{(i) } \|[\Delta y_a; \Delta z_a]\| \leq \frac{\gamma_2}{\rho}, \quad \text{and} \quad \text{(ii) } \tilde{x}^Q \geq -\gamma_3 e, \quad \text{and} \quad \text{(iii) } \tilde{u}^Q \not\geq \gamma_4 e \tag{13}$$

is satisfied, where $\gamma_i > 0$, $i = 1, 2, 3, 4$ are parameters, z_+ is the just computed next value of z, $\tilde{x}^Q$ and $\tilde{u}^Q$ (defined in (18) and (19)) are the most recently computed Karush–Kuhn–Tucker (KKT) multipliers for constraints $(A^Q)^{\mathrm{T}} y - z^Q \leq c^Q$ and $z^Q \geq 0$, respectively, and where the factor

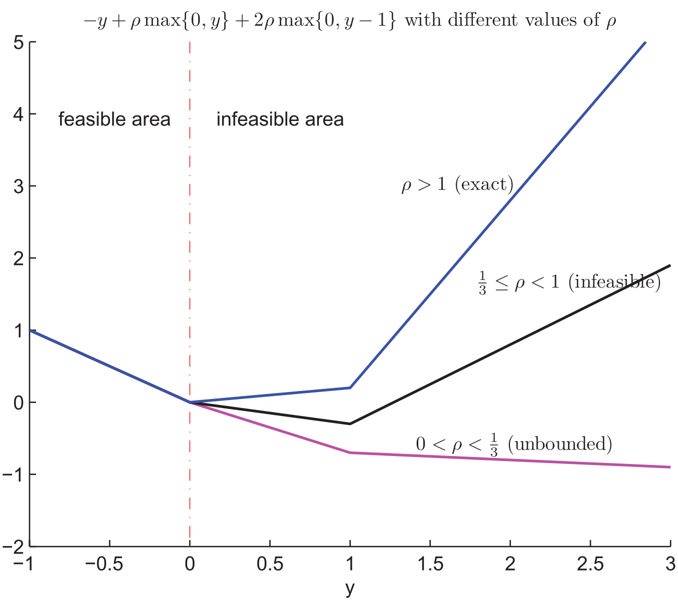

Figure 1. The objective function of problem (11) is displayed with different penalty parameter values. When $\rho < \frac{1}{3}$, problem (10) is unbounded. When $\rho \in [\frac{1}{3}, 1)$, it is bounded but the minimizer $y_*^{\rho} = 1$ is infeasible for (10). When $\rho > \rho_* = 1$, $y_*^{\rho} = 0$ solves (10) as desired.

$\|z_0\|/\rho_0$ has been introduced for scaling purposes. Note that these conditions involve both the dual and primal sets of variables. As we will see though, the resulting algorithm framework IrPDIP is proved to behave adequately under rather mild restrictions on how primal variables are updated.

Condition (12) is new. It ensures boundedness of $\{z_k\}$ (which is necessary in order for $\{y_k\}$ to be bounded) whenever $\{\rho_k\}$ is bounded; with such condition, the situation just described where $\{z_k\}$ is unbounded due to $\{\rho_k\}$ being too small cannot occur. Condition (13) is adapted from [22] (see Step 1(ii) in Algorithm A of [22], as well as the discussion preceding the algorithm statement). Translated to the present context, the intuition is that ρ should be increased if a stationary point[2] for (D_ρ) is approached ($\|[\Delta y_a;\Delta z_a]\|$ small) at which not all components of the constraints $z \geq 0$ are binding (not all components of $\tilde{u}^Q$ are significantly positive), and no component of $\tilde{x}^Q$ or $\tilde{u}^Q$ takes a large negative value, suggesting that the stationary point may not be a dual maximizer. Two adaptations were in order: first, closeness to a stationary point for (D_ρ) is rather related to the size of $\rho\|[\Delta y_a;\Delta z_a]\|$; in [22], this makes no difference because the sequence of multiplier estimates ((x,u) in the present context) is bounded by construction, even when ρ grows without bound; second, the lower bound on $\tilde{u}^Q$ turns out not to be needed in the present context due to the special structure of the $z \geq 0$ constraints (compared to the general $c(x) \geq 0$ in [22]).

Iteration IrPDIP, stated next, amounts to rIPM applied to (D_ρ), rather than (D), with ρ updated as just discussed (Step 2 (iv)), as well as a specific D matrix (primal–dual affine scaling: Step 1(i)) and rather general bounds on how the primal variables x and u should be updated (Step 2 (ii)).

Iteration IrPDIP (infeasible reduced primal–dual interior point)

Parameters: $\theta \in (0,1), \tau > 0, \alpha > 0,\ \chi > 0,\ \sigma > 1,\ \gamma_i > 0$, for $i = 1,2,3,4$.
Data: $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ such that $z > \max\{0,\ A^{\mathrm{T}}y - c\}$; $s := c - A^{\mathrm{T}}y + z$; $x \in \mathbb{R}^n$, $u \in \mathbb{R}^n$, and $\rho \in \mathbb{R}$ such that $x > 0$, $u > 0$, and $\rho > 0$; $Q \subseteq \mathbf{n}$ such that A^Q has full row rank.
Step 1: Computation of the search direction.
(i) Let $(\Delta x^Q, \Delta u, \Delta y_a, \Delta z_a, \Delta s^Q)$ be the primal–dual affine-scaling direction (see (9)) for problem[3]

$$\left.\begin{array}{ll} \max\limits_{y,z} & b^{\mathrm{T}}y - \rho e^{\mathrm{T}}z \\ \text{s.t.} & (A^Q)^{\mathrm{T}}y - z^Q \leq c^Q,\ z \geq 0. \end{array}\right\} \qquad (D^Q_\rho)$$

(ii) Select $(\Delta y, \Delta z)$ to satisfy

$$b^{\mathrm{T}}\Delta y - \rho e^{\mathrm{T}}\Delta z \geq \theta(b^{\mathrm{T}}\Delta y_a - \rho e^{\mathrm{T}}\Delta z_a), \qquad \|[\Delta y;\ \Delta z]\| \leq \tau\|[\Delta y_a;\ \Delta z_a]\|. \tag{14}$$

Step 2. Updates.
(i) Update the dual variables by choosing a stepsize $t \in (0,1]$ such that

$$s_+ := c - A^{\mathrm{T}}y_+ + z_+ > 0, \quad z_+ > 0, \tag{15}$$

where

$$y_+ := y + t\Delta y, \quad z_+ := z + t\Delta z. \tag{16}$$

(ii) Select $[x_+;\ u_+] > 0$ to satisfy

$$\|[x_+;\ u_+]\| \leq \max\{\|[x;\ u]\|, \alpha\|[\tilde{x}^Q;\ \tilde{u}]\|, \chi\}, \tag{17}$$

where

$$\tilde{x}^Q := x^Q + \Delta x^Q, \tag{18}$$

$$\tilde{u} := u + \Delta u. \tag{19}$$

(iii) Pick $Q_+ \subseteq \mathbf{n}$ such that A^{Q_+} has full row rank.
(iv) Check the two cases (12) and (13). If either case is satisfied, set

$$\rho_+ := \sigma\rho;$$

otherwise set $\rho_+ := \rho$.

Note that in order to guarantee that direction $(\Delta x^Q, \Delta u, \Delta y_a, \Delta z_a, \Delta s^Q)$ (see (20)) is well defined, it is sufficient that A^Q have full row rank (see Step 2(iii) in Iteration IrPDIP). Indeed, this makes $[A^Q\ 0; -E^Q\ -I]$ full row rank, so that the solution $(\Delta y_a, \Delta z_a)$ to (21) is well defined.

3.2 *Computational issues*

The main computation in Iteration IrPDIP is the calculation of the affine-scaling direction in Step 1(i). The primal–dual affine-scaling direction $(\Delta x^Q, \Delta u, \Delta y_a, \Delta z_a, \Delta s^Q)$ for (D_ρ^Q) is obtained by solving system (derived from (9))

$$\begin{bmatrix} 0 & 0 & 0 & (A^Q)^{\mathrm{T}} & -I & 0 & I \\ A^Q & 0 & 0 & 0 & 0 & 0 & 0 \\ I & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ S^Q & 0 & 0 & 0 & 0 & 0 & X^Q \\ 0 & Z^Q & 0 & 0 & U^Q & 0 & 0 \\ 0 & 0 & Z^{\bar{Q}} & 0 & 0 & U^{\bar{Q}} & 0 \end{bmatrix} \begin{bmatrix} \Delta x^Q \\ \Delta u^Q \\ \Delta u^{\bar{Q}} \\ \Delta y_a \\ \Delta z_a^Q \\ \Delta z_a^{\bar{Q}} \\ \Delta s^Q \end{bmatrix} = \begin{bmatrix} 0 \\ b - A^Q x^Q \\ \rho e - x^Q - u^Q \\ \rho e - u^{\bar{Q}} \\ -X^Q s^Q \\ -Z^Q u^Q \\ -Z^{\bar{Q}} u^{\bar{Q}} \end{bmatrix}, \tag{20}$$

where $Z := \mathrm{diag}(z)$ and $U := \mathrm{diag}(u)$. Eliminating $(\Delta x^Q, \Delta u)$ and Δs^Q in system (20), we obtain the reduced normal system

$$\begin{bmatrix} A^Q & 0 \\ -E^Q & -I \end{bmatrix} \begin{bmatrix} X^Q & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} S^Q & 0 \\ 0 & Z \end{bmatrix}^{-1} \begin{bmatrix} A^Q & 0 \\ -E^Q & -I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \Delta y_a \\ \Delta z_a \end{bmatrix} = \begin{bmatrix} b \\ -\rho e \end{bmatrix}, \tag{21}$$

$$\Delta s^Q = -(A^Q)^{\mathrm{T}} \Delta y_a + \Delta z_a^Q, \tag{22}$$

$$\begin{bmatrix} \Delta x^Q \\ \Delta u \end{bmatrix} = -\begin{bmatrix} x^Q \\ u \end{bmatrix} - \begin{bmatrix} X^Q & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} S^Q & 0 \\ 0 & Z \end{bmatrix}^{-1} \begin{bmatrix} \Delta s^Q \\ \Delta z_a \end{bmatrix}, \tag{23}$$

where E^Q is a submatrix of the $n \times n$ identity matrix consisting of only those columns that are indexed in set Q. Further eliminating Δz_a, we can reduce (21) to

$$A^Q D^{(Q)} (A^Q)^{\mathrm{T}} \Delta y_a = b - A^Q X^Q (S^Q)^{-1} (E^Q)^{\mathrm{T}} (D_2^{(Q)})^{-1} \rho e, \tag{24}$$
$$D_2^{(Q)} \Delta z_a = -\rho e + E^Q X^Q (S^Q)^{-1} (A^Q)^{\mathrm{T}} \Delta y_a,$$

where diagonal positive definite matrices $D^{(Q)}$ and $D_2^{(Q)}$ are given by

$$D^{(Q)} := X^Q (S^Q)^{-1} - X^Q (S^Q)^{-1} (E^Q)^{\mathrm{T}} (D_2^{(Q)})^{-1} E^Q X^Q (S^Q)^{-1}, \tag{25}$$
$$D_2^{(Q)} := U Z^{-1} + E^Q X^Q (S^Q)^{-1} (E^Q)^{\mathrm{T}}.$$

(Since Q is selected in such a way that A^Q has full row rank, (24) yields a unique Δy_a.) By means of the Sherman–Morrison–Woodbury matrix identity, (25) can be simplified to

$$D^{(Q)} = (S^Q (X^Q)^{-1} + (E^Q)^{\mathrm{T}} U^{-1} Z E^Q)^{-1} = (S^Q (X^Q)^{-1} + Z^Q (U^Q)^{-1})^{-1}.$$

The dominant cost in computing $(\Delta x^Q, \Delta u, \Delta y_a, \Delta z_a, \Delta s^Q)$ is to solve (21), with cost dominated by forming the coefficient matrix $A^Q D^{(Q)} (A^Q)^{\mathrm{T}}$ of (24). When A is dense, this operation takes

$|Q|m^2$ multiplications. In the case of $n \gg m$, this can be much less than nm^2. Indeed, the same speedup factor can be obtained as in the case of the dual-feasible rIPM.

3.3 *Convergence analysis*

Iteration IrPDIP can be repeated indefinitely, generating an infinite sequence of iterates with the dual sequence $\{(y_k, z_k, s_k)\}$ feasible for problem (D_ρ). In Section 2, the sole assumption on (P)–(D) was that A has full row rank. Below, we further selectively assume (strict) feasibility of (P)–(D).

In this section, we show that under mild assumptions the penalty parameter ρ in Iteration IrPDIP will be increased no more than a finite number of times. First, as a direct application of (6) transposed to problem (D_ρ), and of (14), $(\Delta y, \Delta z)$ is an ascent direction for (D_ρ). We state this as a lemma.

LEMMA 3.1 *Step* 1(*i*) *of IrPDIP is well defined and* $b^{\mathrm{T}}\Delta y - \rho e^{\mathrm{T}}\Delta z > 0$.

In view of (12), a necessary condition for $\{\rho_k\}$ to remain bounded is that $\{z_k\}$ be bounded. The latter does hold, as we show next. A direct consequence is boundedness from above of $\{b^{\mathrm{T}}y_k\}$.

LEMMA 3.2 *If* (P) *is feasible, then* $\{z_k\}$ *is bounded, and* $\{b^{\mathrm{T}}y_k\}$ *is bounded from above.*

Proof We first show that $\{z_k\}$ is bounded. If ρ_k is increased finitely many times to a finite value, say ρ_*, then condition (12) must fail for k large enough, i.e. $\|z_k\| \le \gamma_1\|z_0\|\rho_*/\rho_0$ for k large enough, proving the claim. It remains to prove that $\{z_k\}$ is bounded when ρ_k is increased infinitely many times, i.e. when $\rho_k \to \infty$ as $k \to \infty$.

By assumption, (P) has a feasible point, say x^0, i.e.

$$Ax^0 = b, \quad x^0 \ge 0. \tag{26}$$

Since $\rho_k \to \infty$ as $k \to \infty$, there exists k_0 such that

$$\rho_k > \|x^0\|_\infty \quad \forall k \ge k_0. \tag{27}$$

Since $(y_k,\, z_k)$ is feasible for (D_ρ) for all k, we have

$$A^{\mathrm{T}}y_k \le z_k + c \quad \forall k, \tag{28}$$

$$z_k \ge 0 \quad \forall k. \tag{29}$$

Left-multiplying both sides of (28) by $(x^0)^{\mathrm{T}} \ge 0$ and using (26) yields

$$b^{\mathrm{T}}y_k \le (x^0)^{\mathrm{T}}z_k + c^{\mathrm{T}}x^0 \quad \forall k. \tag{30}$$

Adding $\rho_k e^{\mathrm{T}}z_k$ to both sides of (30), we get

$$(\rho_k e - x^0)^{\mathrm{T}}z_k \le \pi_k + \rho_k e^{\mathrm{T}}z_k \quad \forall k, \tag{31}$$

where we have defined

$$\pi_k := c^{\mathrm{T}}x^0 - b^{\mathrm{T}}y_k. \tag{32}$$

In view of (27) and (29), we conclude that z_k satisfies

$$0 \le z_k^i \le \frac{\pi_k + \rho_k e^{\mathrm{T}}z_k}{\rho_k - (x^0)^i} \le \frac{\pi_k + \rho_k e^{\mathrm{T}}z_k}{\rho_k - \|x^0\|_\infty} =: \nu_k \quad \forall i,\ \forall k \ge k_0,$$

so that the just introduced sequence $\{\nu_k\}$ satisfies

$$\|z_k\|_\infty \leq \nu_k \quad \forall k \geq k_0. \tag{33}$$

Hence, in order to show that $\{z_k\}$ is bounded, it suffices to prove that $\{\nu_k\}$ is bounded. We show that $\nu_{k+1} \leq \nu_k$, $\forall k \geq k_0$; since in view of (33), ν_k is nonnegative for all k, this will prove boundedness of $\{\nu_k\}$. To this end, first note that for each k, Lemma 3.1 implies that

$$b^{\mathrm{T}} y_{k+1} - \rho_k e^{\mathrm{T}} z_{k+1} = b^{\mathrm{T}} y_k - \rho_k e^{\mathrm{T}} z_k + t_k(b^{\mathrm{T}} \Delta y_k - \rho_k e^{\mathrm{T}} \Delta z_k) \geq b^{\mathrm{T}} y_k - \rho_k e^{\mathrm{T}} z_k,$$

where we have used (16). Together with (27), this implies that

$$\nu_k = \frac{\pi_k + \rho_k e^{\mathrm{T}} z_k}{\rho_k - \|x^0\|_\infty} \geq \frac{\pi_{k+1} + \rho_k e^{\mathrm{T}} z_{k+1}}{\rho_k - \|x^0\|_\infty} \quad \forall k \geq k_0. \tag{34}$$

Since $\rho_{k+1} \geq \rho_k$ and since

$$\nu_{k+1} = \frac{\pi_{k+1} + \rho_{k+1} e^{\mathrm{T}} z_{k+1}}{\rho_{k+1} - \|x^0\|_\infty}, \tag{35}$$

in order to conclude that $\nu_{k+1} \leq \nu_k$ for $k \geq k_0$, it is sufficient to verify that the function f given by

$$f(\rho) := \frac{\pi_{k+1} + \rho e^{\mathrm{T}} z_{k+1}}{\rho - \|x^0\|_\infty}$$

has a nonpositive derivative $f'(\rho)$ for all ρ satisfying (27). Now,

$$\begin{aligned}
\pi_{k+1} + \|x^0\|_\infty e^{\mathrm{T}} z_{k+1} &= c^{\mathrm{T}} x^0 - b^{\mathrm{T}} y_{k+1} + \|x^0\|_\infty e^{\mathrm{T}} z_{k+1} \\
&= (x^0)^{\mathrm{T}} c - (x^0)^{\mathrm{T}} A^{\mathrm{T}} y_{k+1} + \|x^0\|_\infty e^{\mathrm{T}} z_{k+1} \\
&\geq -(x^0)^{\mathrm{T}} z_{k+1} + \|x^0\|_\infty e^{\mathrm{T}} z_{k+1} \\
&\geq 0,
\end{aligned}$$

where the first equality comes from (32), the second one from (26), the first inequality from (28) and (26), and the second one from (29). In view of (27), it follows that

$$f'(\rho) = -\frac{\pi_{k+1} + \|x^0\|_\infty e^{\mathrm{T}} z_{k+1}}{(\rho - \|x^0\|_\infty)^2} \leq 0.$$

Hence $\{z_k\}$ is bounded, proving the first claim. It follows immediately from (30) that $\{b^{\mathrm{T}} y_k\}$ is bounded from above, proving the second claim. ■

With boundedness of $\{z_k\}$ in hand, the possibility that $\{\rho_k\}$ be unbounded will be ruled out by a contradiction argument. But first, we prove that the primal variables are bounded by a linear function of ρ_k.

LEMMA 3.3 *There exists a constant $C > 0$ such that*

$$\|[\tilde{x}_k^{Q_k}; \tilde{u}_k; x_k; u_k]\| \leq C\rho_k. \tag{36}$$

Proof In view of the triangle inequality, it suffices to show that there exist C_1 and C_2 such that

$$\|[\tilde{x}_k^{Q_k}; \tilde{u}_k]\| \le C_1\rho_k, \quad \|[x_k; u_k]\| \le C_2\rho_k. \tag{37}$$

Substituting (22) into (23), and using (18) and (19), we have

$$\begin{bmatrix} \tilde{x}_k^{Q_k} \\ \tilde{u}_k \end{bmatrix} = \begin{bmatrix} X_k^{Q_k}(S_k^{Q_k})^{-1} & 0 \\ 0 & U_k(Z_k)^{-1} \end{bmatrix} \begin{bmatrix} A^{Q_k} & 0 \\ -E^{Q_k} & -I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \Delta y_{a,k} \\ \Delta z_{a,k} \end{bmatrix}. \tag{38}$$

Solving (21) for $[\Delta y_{a,k};\Delta z_{a,k}]$ and substituting it into (38) yields

$$\begin{bmatrix} \tilde{x}_k^{Q_k} \\ \tilde{u}_k \end{bmatrix} = H_k \begin{bmatrix} b \\ -\rho_k e \end{bmatrix} \tag{39}$$

with

$$H_k := \begin{bmatrix} X_k^{Q_k}(S_k^{Q_k})^{-1} & 0 \\ 0 & U_k(Z_k)^{-1} \end{bmatrix} \begin{bmatrix} A^{Q_k} & 0 \\ -E^{Q_k} & -I \end{bmatrix}^{\mathrm{T}}$$
$$\times \left(\begin{bmatrix} A^{Q_k} & 0 \\ -E^{Q_k} & -I \end{bmatrix} \begin{bmatrix} X_k^{Q_k}(S_k^{Q_k})^{-1} & 0 \\ 0 & U_k(Z_k)^{-1} \end{bmatrix} \begin{bmatrix} A^{Q_k} & 0 \\ -E^{Q_k} & -I \end{bmatrix}^{\mathrm{T}} \right)^{-1}.$$

Because diagonal matrices $X_k^{Q_k}$, $S_k^{Q_k}$, U_k and Z_k are positive definite for all k, it follows from Theorem 1 in [20] that the sequence $\{H_k\}$ is bounded. Therefore, (39) implies that there exist $C' > 0$ and $C_1 > 0$, both independent of k, such that

$$\left\| \begin{bmatrix} \tilde{x}_k^{Q_k} \\ \tilde{u}_k \end{bmatrix} \right\| \le C' \left\| \begin{bmatrix} b \\ -\rho_k e \end{bmatrix} \right\| \le C_1\rho_k \quad \forall k, \tag{40}$$

proving the first inequality in (37). Now, without loss of generality, suppose

$$C_1 \ge \frac{\max\{\|[x_0; u_0]\|, \chi\}}{\alpha\rho_0},$$

where α is a parameter in Iteration IrPDIP, and let $C_2 \ge \alpha C_1$. That $\|[x_k; u_k]\| \le C_2\rho_k$ follows by induction. Indeed, it clearly holds at $k=0$, and if $\|[x_k; u_k]\| \le C_2\rho_k$ at some iterate k, then since $\{\rho_k\}$ is nondecreasing, it follows from (17) and (40) that

$$\|[x_{k+1}; u_{k+1}]\| \le \max\{C_2\rho_k, \alpha C_1\rho_k, \chi\} \le C_2 \max\{\rho_k, \rho_0\} \le C_2\rho_{k+1}. \tag{41}$$

■

If (P) is feasible, then Lemma 3.2 rules out the possibility that condition (12) is satisfied on an infinite sequence. Therefore, if, as we will assume by contradiction, $\rho_k \to \infty$ as $k \to \infty$, conditions (13) must be satisfied on an infinite subsequence. The next lemma exploits this. In that lemma and in Proposition 3.5, K_ρ denotes the index sequence on which ρ_k is updated, i.e.

$$K_\rho = \{k : \rho_{k+1} > \rho_k\}.$$

LEMMA 3.4 *If $\rho_k \to \infty$ as $k \to \infty$ and* (P) *is feasible, then $\{Z_k\tilde{u}_k\}$ and $\{S_k^{Q_k}\tilde{x}_k^{Q_k}\}$ are bounded on K_ρ. If in addition* (D) *is feasible, then $z_k \to 0$ as $k \to \infty$, $k \in K_\rho$, and if furthermore* (P) *is strictly feasible, then $\{y_k\}$ is bounded on K_ρ.*

Proof Since ρ_k goes to infinity on K_ρ and (P) is feasible, Lemma 3.2 implies that condition (12) is eventually violated, so conditions (13) must be satisfied for $k \in K_\rho$ large enough. In particular, there exists k_0 such that for all $k \geq k_0$, $k \in K_\rho$,

$$\|[\Delta y_{a,k};\ \Delta z_{a,k}]\| \leq \frac{\gamma_2}{\rho_k}, \tag{42}$$

and

$$\tilde{x}_k^{Q_k} \geq -\gamma_3 e. \tag{43}$$

Since (first block row of (20))

$$\Delta s_k^{Q_k} = -(A^{Q_k})^{\mathrm{T}} \Delta y_{a,k} + \Delta z_{a,k}^{Q_k},$$

it follows from (42) that there exists $\delta > 0$ such that

$$\|\Delta s_k^{Q_k}\| \leq \frac{\delta}{\rho_k}, \quad k \geq k_0,\ \ k \in K_\rho. \tag{44}$$

Using Lemma 3.3, equations (42) and (44) and the last three block rows of (20), we obtain

$$\|Z_k \tilde{u}_k\| = \|U_k \Delta z_{a,k}\| \leq C\rho_k \cdot \frac{\gamma_2}{\rho_k} = C\gamma_2, \quad k \geq k_0,\ \ k \in K_\rho \tag{45}$$

and

$$\|S_k^{Q_k} \tilde{x}_k^{Q_k}\| = \|X_k^{Q_k} \Delta s_k^{Q_k}\| \leq C\rho_k \cdot \frac{\delta}{\rho_k} = C\delta, \quad k \geq k_0,\ k \in K_\rho, \tag{46}$$

which proves the first claim. Now, without loss of generality, assume that $\rho_{k_0} > \|x^0\|_\infty$ with x^0 a feasible point of (P), so that

$$u_k^0 := \rho_k e - x^0 > 0, \quad \forall k \geq k_0. \tag{47}$$

Then, by our assumption, in the second claim, that (P)–(D) is feasible, there exist y^0 and $s^0 \geq 0$ which, together with x^0, satisfy

$$\begin{aligned} A^{Q_k}(x^0)^{Q_k} + A^{\bar{Q}_k}(x^0)^{\bar{Q}_k} &= Ax^0 = b, \\ x^0 + u_k^0 &= \rho_k e, \\ A^{\mathrm{T}} y^0 + s^0 &= c. \end{aligned}$$

On the hand other, from the second, third and fourth block rows of (20), and definitions (18), (19) and (15), we obtain

$$\begin{aligned} A^{Q_k}\tilde{x}_k^{Q_k} &= b, \\ (\tilde{x}_k + \tilde{u}_k)^{Q_k} &= \rho_k e, \quad \tilde{u}_k^{\bar{Q}_k} = \rho_k e, \\ A^{\mathrm{T}} y_k + s_k - z_k &= c. \end{aligned} \tag{48}$$

These two groups of equations yield

$$\begin{bmatrix} A^{Q_k} & A^{\bar{Q}_k} & 0 & 0 \\ I & 0 & I & 0 \\ 0 & I & 0 & I \end{bmatrix} \begin{bmatrix} (\tilde{x}_k - x^0)^{Q_k} \\ -(x^0)^{\bar{Q}_k} \\ (\tilde{u}_k - u_k^0)^{Q_k} \\ (\tilde{u}_k - u_k^0)^{\bar{Q}_k} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} A^{Q_k} & A^{\bar{Q}_k} & 0 & 0 \\ I & 0 & I & 0 \\ 0 & I & 0 & I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} y^0 - y_k \\ z_k^{Q_k} \\ z_k^{\bar{Q}_k} \end{bmatrix} = \begin{bmatrix} (s_k - s^0)^{Q_k} \\ (s_k - s^0)^{\bar{Q}_k} \\ z_k^{Q_k} \\ z_k^{\bar{Q}_k} \end{bmatrix}.$$

It follows that

$$[(\tilde{x}_k - x^0)^{Q_k}; -(x^0)^{\bar{Q}_k}; (\tilde{u}_k - u_k^0)] \perp [(s_k - s^0)^{Q_k}; (s_k - s^0)^{\bar{Q}_k}; z_k],$$

i.e.

$$(\tilde{x}_k^{Q_k})^{\mathrm{T}}(s_k - s^0)^{Q_k} - (x^0)^{\mathrm{T}}(s_k - s^0) + (\tilde{u}_k - u_k^0)^{\mathrm{T}} z_k = 0. \tag{49}$$

Hence, for C' large enough, we obtain

$$\begin{aligned} (u_k^0)^{\mathrm{T}} z_k + (x^0)^{\mathrm{T}} s_k &= (x^0)^{\mathrm{T}} s^0 + (\tilde{x}_k^{Q_k})^{\mathrm{T}} s_k^{Q_k} - (\tilde{x}_k^{Q_k})^{\mathrm{T}} (s^0)^{Q_k} + \tilde{u}_k^{\mathrm{T}} z_k \\ &\leq (x^0)^{\mathrm{T}} s^0 + C'\delta + \gamma_3 e^{\mathrm{T}} s^0 + C'\gamma_2 \end{aligned} \tag{50}$$

where the equality comes from the expansion of (49), and the inequality from (46), (43), and (45). Since u_k^0, z_k, x^0 and s_k are nonnegative for $k \geq k_0$, we obtain

$$z_k^i \leq \frac{(x^0)^{\mathrm{T}} s^0 + C'\delta + \gamma_3 e^{\mathrm{T}} s^0 + C'\gamma_2}{(u_k^0)^i}, \quad \forall i,\ k \geq k_0,\ k \in K_\rho.$$

Since (see (47)) $(u_k^0)^i \to \infty$, $i \in \mathbf{n}$ as $k \to \infty$ on K_ρ, this proves that

$$\lim_{k \to \infty, k \in K_\rho} z_k = 0,$$

proving the second claim. Finally, if in addition (P) is strictly feasible, we can select $x^0 > 0$, and (50) yields

$$s_k^i \leq \frac{(x^0)^{\mathrm{T}} s^0 + C'\delta + \gamma_3 e^{\mathrm{T}} s^0 + C'\gamma_2}{(x^0)^i}, \quad \forall i, k \geq k_0, k \in K_\rho,$$

proving that $\{s_k\}$ is bounded on K_ρ. Boundednesses of $\{s_k\}$ and $\{z_k\}$, together with equation (48) and the full-rank property of A, imply that $\{y_k\}$ is bounded on K_ρ. ∎

We are now ready to prove that ρ_k is increased at most finitely many times. The proof uses the fact that if (D) has a strictly feasible point, then for all $y \in \mathcal{F}$, $\{a_i : i \in I(y)\}$ must be a positively linearly independent set of vectors.

Proposition 3.5 *If* (P)–(D) *is strictly feasible, then* ρ_k *is increased at most finitely many times, i.e.* K_ρ *is finite. Furthermore,* $\{y_k\}$ *and* $\{z_k\}$ *converge to some* y_* *and* z_*.

Proof If the first claim holds, then after finitely many iterations, IrPDIP reduces to rIPM applied to (D_ρ), so the second claim follows by Proposition 2.2. It remains to prove the first claim. Proceeding by contradiction, suppose K_ρ is infinite. Then there exists an infinite index set $K \subseteq K_\rho$ and some $Q \subseteq \mathbf{n}$ such that $Q_k = Q$, for all $k \in K$. In view of Lemma 3.2, since $K \subseteq K_\rho$, there must exist $k_0 > 0$ such that conditions (13) are satisfied for $k \geq k_0,\ k \in K$; in particular,

$$\tilde{x}_k^Q \geq -\gamma_3 e, \quad k \geq k_0,\ k \in K, \tag{51}$$

$$\tilde{u}_k^Q \ngeq \gamma_4 e, \quad k \geq k_0,\ k \in K. \tag{52}$$

Since $\lim_{k\to\infty} \rho_k = \infty$, it follows from (18), (19), the third block row of (20), and (52) that

$$\lambda_k := \|\tilde{x}_k^Q\|_\infty = \|\rho_k e - \tilde{u}_k^Q\|_\infty \longrightarrow \infty, \quad \text{as } k \longrightarrow \infty,\ \ k \in K. \tag{53}$$

Hence

$$\|\hat{x}_k^Q\|_\infty = 1, \quad k \geq k_0,\ \ k \in K, \tag{54}$$

where we have defined

$$\hat{x}_k^Q := \frac{\tilde{x}_k^Q}{\lambda_k}, \quad k \geq k_0,\ \ \forall k \in K. \tag{55}$$

(Without loss of generality, we have assumed that $\lambda_k \neq 0,\ \forall k \geq k_0, k \in K$.) Now, in view of Lemma 3.4, we have for certain constant $C > 0$ large enough,

$$\|S_k^Q \tilde{x}_k^Q\| \leq C, \quad \forall k \in K, \tag{56}$$

$$\|y_k\| \leq C, \quad \forall k \in K, \tag{57}$$

$$\lim_{k\to\infty} z_k = 0, \quad k \in K. \tag{58}$$

Note that by (57) and (54), $\{y_k\}$ and $\{\hat{x}_k^Q\}$ are bounded on K, so in view of (54) and (58), there exists an infinite index set $K' \subseteq K$ such that

$$\hat{x}_k^Q \longrightarrow \hat{x}_*^Q \neq 0, \quad y_k \longrightarrow y_*, \quad z_k \longrightarrow z_* = 0 \quad \text{as } k \longrightarrow \infty,\ k \in K', \tag{59}$$

for some $\hat{x}_*^Q$ and some $y_* \in \mathcal{F}$ (since $z_* = 0$). Dividing by λ_k and taking the limit on both sides of (56), we obtain

$$S_k^Q \hat{x}_k^Q \longrightarrow 0 \quad \text{as } k \longrightarrow \infty,\ k \in K',$$

which implies that

$$\hat{x}_*^i = 0, \quad \forall i \in Q \backslash I(y_*). \tag{60}$$

On the other hand, the second block equation in (20) and equation (18) give

$$A^Q \tilde{x}_k^Q = b \quad \forall k.$$

Dividing by λ_k, taking the limit of both sides, and using (60), we obtain

$$\sum_{i \in I(y_*) \cap Q} \hat{x}_*^i a^i = 0. \tag{61}$$

Now note from (51), (55) and (53) that

$$\hat{x}_*^Q = \lim_{k\to\infty, k\in K'} \frac{\tilde{x}_k^Q}{\lambda_k} \geq \lim_{k\to\infty, k\in K'} \frac{-\gamma_3 e}{\lambda_k} = 0. \tag{62}$$

Since strict feasibility of (D) implies positive linear independence of vectors $\{a^i : i \in I(y_*) \cap Q,\ y_* \in \mathcal{F}\}$, it follows from (61) and (62) that

$$\hat{x}_*^i = 0, \quad \forall i \in I(y_*) \cap Q.$$

Together with (60), this implies that

$$\hat{x}_*^Q = 0,$$

which is a contradiction to (59). ■

3.4 *An ℓ_∞ version*

Instead of the ℓ_1 exact penalty function used in (P$_\rho$)–(D$_\rho$), we can use an ℓ_∞ exact penalty function and consider the problem

$$\begin{aligned} \max \quad & b^{\mathrm{T}} y - \rho z \\ \text{s.t.} \quad & A^{\mathrm{T}} y - z e \leq c, \; z \geq 0 \end{aligned} \tag{63}$$

with its associated primal

$$\begin{aligned} \min \quad & c^{\mathrm{T}} x \\ \text{s.t.} \quad & Ax = b, \; e^{\mathrm{T}} x + u = \rho, \\ & x \geq 0, \; u \geq 0, \end{aligned}$$

where $z \in \mathbb{R}$ and $u \in \mathbb{R}$. Again, strictly feasible points for (63) are readily available. Conditions akin to (12)–(13) can again be used to iteratively obtain an appropriate value of ρ. Since both z and u are scalar variables, the scheme can be slightly simplified: Increase ρ if either

$$z_+ \geq \gamma_1 \frac{z_0}{\rho_0} \rho,$$

or

$$\text{(i) } \|[\Delta y_a; \Delta z_a]\| \leq \frac{\gamma_2}{\rho}, \quad \text{and} \quad \text{(ii) } \tilde{x}^Q \geq -\gamma_3 e, \quad \text{and} \quad \text{(iii) } \tilde{u} < \gamma_4. \tag{64}$$

An analysis very similar to that of Section 3.3 shows that the resulting ℓ_∞ variant of IrPDIP enjoys the same theoretical properties as the ℓ_1 version. Minor changes include substitution of the ℓ_∞-dual norm $\|\cdot\|_1$ for the ℓ_1-dual norm $\|\cdot\|_\infty$.

4. Infeasible constraint-reduced MPC: IrMPC

As an instance of IrPDIP, we apply rMPC* of [24] to (P$_\rho$)–(D$_\rho$), and we dub the resulting full algorithm IrMPC. (Indeed the search direction in rMPC* satisfies condition (4) of rIPM and condition (17) of IrPDIP.) In view of Proposition 3.5, subject to strict feasibility of (P)–(D), after finitely many iterations, the ℓ_1 and ℓ_∞ versions of IrMPC reduce to rMPC* applied to problem (D$_\rho$) and (63), respectively, with ρ equal to a fixed value $\bar{\rho}$. Thus, we can invoke results from [24] under appropriate assumptions.

Proposition 4.1 *Suppose (P)–(D) is strictly feasible. Then $\{(y_k, z_k)\}$ generated by the ℓ_1 or ℓ_∞ version of IrMPC converges to a stationary point (y_*, z_*) of problem* (D$_\rho$) *with $\rho = \bar{\rho}$.*

Proof We prove the claim for the ℓ_1 version; the ℓ_∞ case follows similarly. It follows from Theorem 3.8 in [24] that $\{(y_k, z_k)\}$ converges to a stationary point of problem (D$_\rho$) if and only if the penalized dual objective function is bounded. To conclude the proof, we now establish that $\{b^{\mathrm{T}} y_k - \rho_k e^{\mathrm{T}} z_k\}$ is bounded indeed. Lemma 3.1 implies that $\{b^{\mathrm{T}} y_k - \rho_k e^{\mathrm{T}} z_k\}$ is increasing for k large enough that $\rho_k = \bar{\rho}$, so it is sufficient to prove that $\{b^{\mathrm{T}} y_k - \rho_k e^{\mathrm{T}} z_k\}$ is bounded from above. Since Lemma 3.2 implies that $\{b^{\mathrm{T}} y_k\}$ is bounded from above, this claim follows from boundedness of $\{z_k\}$ and $\{\rho_k\}$ (from Lemma 3.2 and Proposition 3.5, respectively). ■

Under a non-degeneracy assumption,[4] $\{z_k\}$ converges to zero, and thus $\{y_k\}$ converges to an optimal solution of (D). The proof of the following lemma is routine and hence omitted.

LEMMA 4.2 *The gradients of active constraints of problem* (D_ρ) *are linearly independent for all* (y, z) *if and only if* $\{a^i : (a^i)^T y = c^i\}$ *is a linearly independent set of vectors for all* $y \in \mathbb{R}^m$.

THEOREM 4.3 *Suppose* (P)–(D) *is strictly feasible, and for all* $y \in \mathbb{R}^m$, $\{a^i : (a^i)^T y = c^i\}$ *is a linearly independent set of vectors. Let* $\{(y_k, z_k)\}$ *be generated by the* ℓ_1 *or* ℓ_∞ *version of IrMPC. Then* $\{(y_k, z_k)\}$ *converges to* $(y_*, 0)$, *a solution of* $(D_{\bar\rho})$, *and* y_* *solves* (D). *Further, if the dual optimal set is a singleton* $\{y_*\}$, *then* $\{(\tilde{x}_k, \tilde{u}_k)\}$ *converges to the unique KKT multiplier* $(x_*, \bar\rho - x_*)$ *associated with* $(y_*, 0)$ *for* $(D_{\bar\rho})$, *and* x_* *is the unique KKT multiplier associated with* y_* *for* (D). *Moreover,* $\{(x_k, u_k), (y_k, z_k)\}$ *converges to* $\{(x_*, \bar\rho - x_*), (y_*, 0)\}$ *q-quadratically.*

Proof Lemma 4.2 implies that the gradients of active constraints of problem (D_ρ) are linearly independent for all feasible (y, z). Applying the latter portion of Theorem 3.8 in [24], we conclude that (y_k, z_k) converges to a maximizer (y_*, z_*) of problem ($D_{\bar\rho}$). Next, Proposition 3.9 of [24] implies that there exists an infinite index set K on which $[\tilde{x}_k; \tilde{u}_k]$ converges to an optimal solution $[x_*; u_*]$ of problem ($P_{\bar\rho}$) with $u_* = \bar\rho - x_*$ and on which

$$[\Delta y_{a,k}; \Delta z_{a,k}] \longrightarrow 0 \quad \text{as } k \longrightarrow \infty,\ k \in K.$$

Thus conditions (i) and (ii) of (13) or (64) are satisfied on K. On the other hand, since $\rho_k = \bar\rho$ for $k \in K$ large enough, one condition in (13) or (64) must fail. In the ℓ_1 case, where (13) applies, it follows that $\tilde{u}_k^{Q_k} \geq \gamma_4 e$ for $k \in K$ large enough. Since it follows from the fourth block row of (20) and definition (19) that $\tilde{u}_k^{\bar{Q}_k} = \rho_k e$, we conclude that

$$\tilde{u}_k \geq \min(\gamma_4, \bar\rho)e,\ k \in K \text{ large enough.}$$

It follows that

$$u_* \geq \min(\gamma_4, \bar\rho)e.$$

In the ℓ_∞ case, where (64) applies, it similarly follows that $u_* \geq \gamma_4$. Hence, in both cases, complementary slackness implies that $z_* = 0$, and as a consequence, y_* is an optimal solution of problem (D). Linear independence of $\{a^i : (a^i)^T y = c^i\}$ implies uniqueness of the KKT multiplier x_*, and the remaining claims are consequences of Theorem 4.1 in [24]. ∎

5. Numerical results

5.1 *Implementation*

IrMPC was implemented in MATLAB R2009a.[5] All tests were run on a laptop machine (Intel R/1.83 GHz, 1 GB of RAM, Windows XP Professional 2002). To mitigate random errors in measured CPU time, we report averages over 10 repeated runs.

The parameters for rMPC* (in Step 1(ii) and Step 2(i)–(iii) of IrMPC) were set to the same values as in Section 5 (Numerical Experiments) of [24]. As for the adaptive scheme (12)–(13), parameters were set to $\sigma := 10$, $\gamma_1 := 10$, $\gamma_2 := 1$, $\gamma_3 := 100$, $\gamma_4 := 100$, and the Euclidean norm was used in (12) and (13). We chose Q according to the most active rule (Rule 2.1 in [24] with $\epsilon = \infty$), which selects the constraints that have smallest slacks s. Analogously to [24], we terminated when

$$\max\left\{\frac{\|[b - Ax; \rho e - x - u]\|}{1 + \|[x; u]\|}, \frac{c^T x - b^T y + \rho e^T z}{1 + |b^T y - \rho e^T z|}\right\} < \text{tol},$$

where we used tol $= 10^{-8}$ and where, again, the Euclidean norm was used.

We applied algorithm IrMPC on two types of examples: randomly generated problems and a problem in model predictive control.

5.2 *Randomly generated problems*

We generated standard-form LPs of size $m = 100$ and $n = 20{,}000$. Entries of matrix A and vector b were normally distributed according to $\mathcal{N}(0, 1)$. We set vector $c := A^{\mathrm{T}}y + s$ with a normally distributed vector $y \sim \mathcal{N}(0, 1)$ and with a uniformly distributed vector $s \sim \mathcal{U}[0, 1]$, guaranteeing that the generated dual (D) is strictly feasible. We adopted (typically *infeasible* for (D)) initial conditions (x_0, y_0, s_0) from [13] for (P)–(D). Namely, we first computed

$$\begin{aligned}
&\tilde{y} := (AA^{\mathrm{T}})^{-1}Ac, \quad \tilde{s} := c - A^{\mathrm{T}}\tilde{y}, \quad \tilde{x} := A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}b,\\
&\delta_x := \max\{-1.5\min(\tilde{x}), 0\}, \quad \delta_s := \max\{-1.5\min(\tilde{s}), 0\},\\
&\tilde{\delta}_x := \delta_x + 0.5\,\frac{(\tilde{x} + \delta_x e)^{\mathrm{T}}(\tilde{s} + \delta_s e)}{\sum_{i=1}^{n}(\tilde{x}^i + \delta_x)}, \quad \tilde{\delta}_s := \delta_s + 0.5\frac{(\tilde{x} + \delta_x e)^{\mathrm{T}}(\tilde{s} + \delta_s e)}{\sum_{i=1}^{n}(\tilde{s}^i + \delta_s)}
\end{aligned}$$

and selected (x_0, y_0, s_0) to be

$$x_0 := \tilde{x} + \tilde{\delta}_x e, \quad y_0 := \tilde{y}, \quad s_0 := \tilde{s} + \tilde{\delta}_s e. \tag{65}$$

Hence $s_0 > c - A^{\mathrm{T}}y_0$. A *strictly feasible* initial point for the penalized problem (D_ρ) was then generated by setting z_0 to

$$z_0 := A^{\mathrm{T}}y_0 - c + s_0 > 0,$$

and to promote centrality, initial point u_0 was computed as

$$u_0^i := \frac{\mu_0}{z_0^i}, \quad i \in \mathbf{n},$$

where $\mu_0 := (x_0)^{\mathrm{T}}s_0/n$. The penalty parameter was initialized with $\rho_0 := \|x_0 + u_0\|_\infty$ for the version with the ℓ_1 exact penalty function, and with $\rho_0 := e^{\mathrm{T}}x_0 + u_0$ for the ℓ_∞ version.

We generated 10 different random problems. The average CPU time and iteration count for solving those 10 problems for various values of $|Q|/n$ (10 runs for each problem, hence 100 runs for each value of $|Q|/n$) are shown in Figures 2 and 3 for the ℓ_1 and ℓ_∞ versions, respectively. Point y_0 initialized as in (65) was infeasible for (D) for all generated problems. The fraction $|Q|/n$ of kept constraints is showed in the horizontal axis with a logarithmic scale. The rightmost point, with $|Q| = n$, corresponds to no constraint reduction. As seen from both Figures 2 and 3, the CPU time decreases as $|Q|/n$ decreases, down to as little as 1% of the constraints. As was already observed in [21,24], the number of iterations remains approximately constant for a large range of fractions $|Q|/n$. Note that with no, or moderate, constraint reduction, the ℓ_∞ version takes fewer iterations and less time to solve the problems than the ℓ_1 version, but the respective performances are similar with more aggressive constraint reduction. We have no explanation for this phenomenon.

Interestingly, even with no constraint reduction, the ℓ_∞ version of IrMPC performs better than the original MPC in our experiments (we used the version from [25] on these problems): see dashed magenta lines on Figures 2 and 3.

5.3 *Model-predictive control*

Model-predictive control (RHC)[6] is a paradigm for controlling a physical dynamical system, by which the state of the system is measured at every (discrete) time t and, during time interval

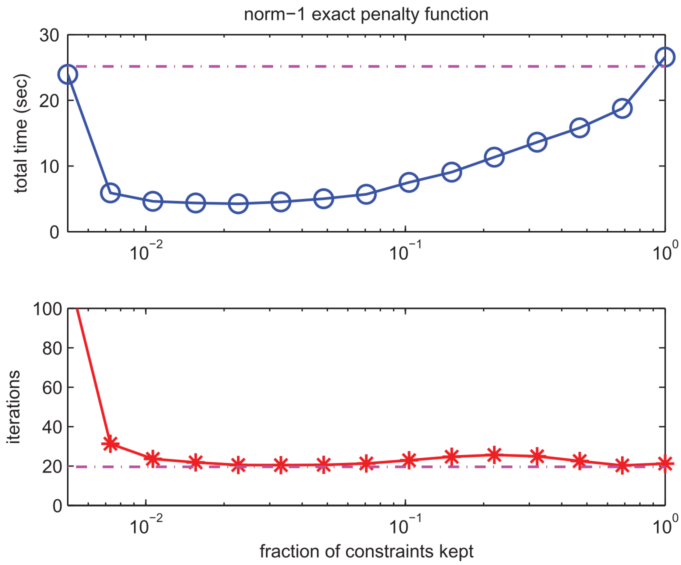

Figure 2. CPU time and iterations in solving the randomly generated problem by IrMPC with a varying fraction of kept constraints for the ℓ_1 exact penalty function; see (blue) circles and (red) stars. The time and iteration count for the original MPC are indicated by the dashed (magenta) lines.

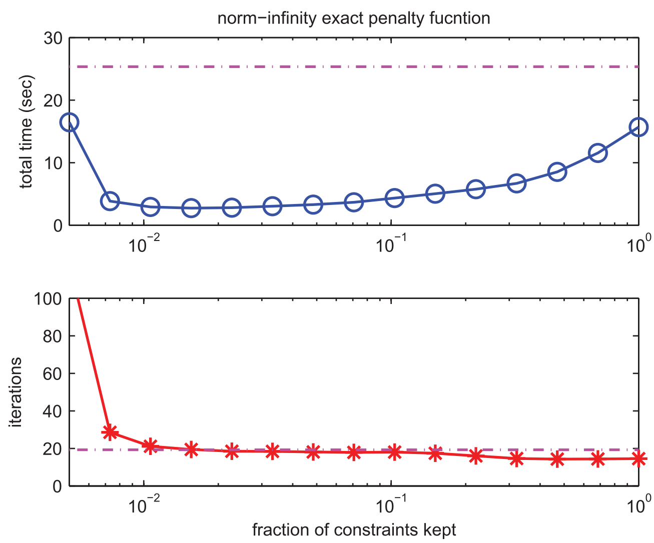

Figure 3. Same as Figure 2, but with the ℓ_∞ version of IrMPC.

$(t-1, t)$, an optimization problem such as the following is solved:

$$\min_{\boldsymbol{w},\boldsymbol{\theta}} \quad \sum_{k=0}^{M-1} \|Rw_k\|_\infty + \sum_{k=1}^{N} \|P\theta_k\|_\infty \tag{66}$$

$$\text{s.t.} \quad \theta_{k+1} = A_s\theta_k + B_s w_k \quad \text{for } k = 0, \ldots, N-1, \tag{67}$$

$$\theta_0 = \theta(t-1), \tag{68}$$

$$\theta_{\min} \leq \theta_k \leq \theta_{\max} \quad \text{for } k = 1, \ldots, N, \tag{69}$$

$$w_{\min} \leq w_k \leq w_{\max} \quad \text{for } k = 0, \ldots, M-1, \tag{70}$$

$$\delta w_{\min} \leq w_k - w_{k-1} \leq \delta w_{\max} \quad \text{for } k = 0, \ldots, M-1, \tag{71}$$

$$w_k = 0 \quad \text{for } k = M, \ldots, N-1, \tag{72}$$

with $R \in \mathbb{R}^{r \times r}$, $P \in \mathbb{R}^{p \times p}$, $A_s \in \mathbb{R}^{p \times p}$ and $B_s \in \mathbb{R}^{p \times r}$. Vectors $\theta_k \in \mathbb{R}^p$ and $w_k \in \mathbb{R}^r$, respectively denote the estimated state and the control input k time steps ahead of the current time, and positive integers M and N are the control and prediction horizons, respectively; (67) is a model of the physical system being controlled; $\theta(t-1)$ is the state of the physical system measured (sensed) at time $t-1$; parameters $\theta_{\min}$, $\theta_{\max}$, $w_{\min}$, $w_{\max}$, $\delta w_{\min}$ and $\delta w_{\max}$ are prescribed bounds; and constraints (71) restrict the rate of change of w. The optimization variables are the control sequence and state sequence, respectively denoted by

$$\boldsymbol{w} = [w_0; \cdots; w_{M-1}] \in \mathbb{R}^{Mr}, \quad \boldsymbol{\theta} = [\theta_1; \cdots; \theta_N] \in \mathbb{R}^{Np}.$$

After problem (66)–(72) is solved, yielding the optimal control sequence $[w_0^*; \cdots; w_{M-1}^*]$, only the first step $w_0^* =: w(t-1)$ of the sequence is applied as control input to the physical system (at time t). The main computational task is to solve (66)–(72). (See, e.g. [8] for background on model-predictive control.)

Problem (66)–(72) can be converted to a standard-form dual LP. First, introduce additional nonnegative optimization variables $[\epsilon_{w_0}, \ldots, \epsilon_{w_{M-1}}, \epsilon_{\theta_1}, \ldots, \epsilon_{\theta_N}]^{\mathrm{T}} \in \mathbb{R}^{M+N}$ required to satisfy

$$Rw_k - \epsilon_{w_k} e \leq 0, \quad -Rw_k - \epsilon_{w_k} e \leq 0, \quad k = 0, \ldots, M-1, \tag{73}$$

$$P\theta_k - \epsilon_{\theta_k} e \leq 0, \quad -P\theta_k - \epsilon_{\theta_k} e \leq 0, \quad k = 1, \ldots, N. \tag{74}$$

Minimizing the objective function of (66) is then equivalent to minimizing $\epsilon_{w_0} + \cdots + \epsilon_{w_{M-1}} + \epsilon_{\theta_1} + \cdots + \epsilon_{\theta_N}$ with additional constraints (73)–(74). Second, states θ_k can be expressed explicitly in terms of w_k by eliminating constraints (67)–(68),

$$\theta_k = A_s^k \theta_0 + \sum_{i=0}^{k-1} A_s^i B_s w_{k-1-i}, \quad k = 1, \ldots, N,$$

or equivalently in matrix form,

$$\boldsymbol{\theta} = \Gamma \boldsymbol{w} + \Omega \theta_0, \tag{75}$$

where

$$\Gamma := \begin{bmatrix} B_s & 0 & \cdots & 0 & 0 \\ A_s B_s & B_s & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_s^{M-1} B_s & A_s^{M-2} B_s & \cdots & A_s B_s & B_s \\ A_s^M B_s & A_s^{M-1} B_s & \cdots & A_s^2 B_s & A_s B_s \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_s^{N-1} B_s & A_s^{N-2} B_s & \cdots & \cdots & A_s^{N-M} B_s \end{bmatrix}, \quad \Omega := \begin{bmatrix} A_s \\ A_s^2 \\ \cdots \\ A_s^N \end{bmatrix}.$$

Hence, problem (66)–(72) can be rewritten into the standard-form dual LP

$$
\begin{aligned}
\min_{\boldsymbol{w},\boldsymbol{\epsilon_w},\boldsymbol{\epsilon_\theta}} \quad & \epsilon_{w_0} + \cdots + \epsilon_{w_{M-1}} + \epsilon_{\theta_1} + \cdots + \epsilon_{\theta_N} && (76)\\
\text{s.t.} \quad & w_{\min} e \le \boldsymbol{w} \le w_{\max} e, && (77)\\
& \theta_{\min} e - \Omega\theta_0 \le \Gamma \boldsymbol{w} \le \theta_{\max} e - \Omega\theta_0, && (78)\\
& \delta w_{\min} \le w_k - w_{k-1} \le \delta w_{\max} \quad \text{for } k = 0, \ldots, M-1, && (79)\\
& \mathbf{R}\boldsymbol{w} - \boldsymbol{\epsilon_w} \le 0, && (80)\\
& \mathbf{R}\boldsymbol{w} + \boldsymbol{\epsilon_w} \ge 0, && (81)\\
& \mathbf{P}\Gamma\boldsymbol{w} - \boldsymbol{\epsilon_\theta} \le -\mathbf{P}\Omega\theta_0, && (82)\\
& \mathbf{P}\Gamma\boldsymbol{w} + \boldsymbol{\epsilon_\theta} \ge -\mathbf{P}\Omega\theta_0, && (83)
\end{aligned}
$$

where

$$\boldsymbol{\epsilon_w} := [\epsilon_{w_0} e; \cdots ; \epsilon_{w_{M-1}} e] \in \mathbb{R}^{Mr}, \quad \boldsymbol{\epsilon_\theta} := [\epsilon_{\theta_1} e; \cdots ; \epsilon_{\theta_N} e] \in \mathbb{R}^{Np}$$

and

$$\mathbf{R} := \mathrm{diag}\{R, R, \ldots, R\} \in \mathbb{R}^{Mr \times Mr}, \quad \mathbf{P} := \mathrm{diag}\{P, P, \ldots, P\} \in \mathbb{R}^{Np \times Np}.$$

When all states and control inputs are constrained by bounds, problem (76)–(83) has $Mr + M + N$ variables and $6Mr + 4Np$ constraints. Because usually $p > r$ and $N > M$, the number of constraints is much larger than that of variables. Hence, constraint reduction is likely to be beneficial.

For this class of problems, IrMPC has two advantages over dual-feasible constraint reduction. First, while warm starts are readily available from the solution of the problem at the previous time step, they may be infeasible (due to the receding horizon as well as to modelling errors and perturbations affecting the dynamical system). Second, even when the warm start is strictly feasible, it is usually close to the boundary of the feasible set. In this situation, feasible IPMs (e.g. [24]) need to take many small steps to get away from that boundary, making it slow to converge. Problem (D_ρ) allows initial points to move outside the feasible set, avoiding both problems.

The warm starts were set as follows: Given a partitioned vector $v = [v^1; \cdots ; v^n]$, define $\bar{v} := [v^2; \cdots ; v^n; v^n]$ where the first block-component has been clipped and the last one repeated. Now, suppose that for time interval $(t-1, t)$, the solution for problem (76)–(83) is

$$[\boldsymbol{w}; \boldsymbol{\epsilon_w}; \boldsymbol{\epsilon_\theta}] := [w_*^0; \cdots ; w_*^{M-1}; \epsilon_{w_*^0}; \cdots ; \epsilon_{w_*^{M-1}}; \epsilon_{\theta_*^1}; \cdots ; \epsilon_{\theta_*^N}];$$

then the initial point we used for the problem (76)–(83) solved during interval $(t, t+1)$ is

$$[\bar{\boldsymbol{w}}; \bar{\boldsymbol{\epsilon}}_{\boldsymbol{w}} + 0.01; \bar{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}} + 0.01]$$

with initial penalty parameter ρ equal to $2\|x_*^t\|_\infty$, where x_*^t is the solution for the dual of problem (76)–(83) for interval $(t-1, t)$. As for the next state $\theta(t)$, we generated it using the dynamics

$$\theta(t) = A_s\theta(t-1) + B_s w(t-1),$$

i.e. we assumed for simplicity that the model used in the optimization is exact, and that there are no perturbations.

The data we tested are from a rotorcraft hover control problem. We ran the controlled system starting at $t = 1$ $(t - 1 = 0)$ with $\theta(0) = [0; 0; 0; 0; 0; 0; 0; -40\,\text{ft}]$ (40 feet initial downward deviation from desired altitude), and with the model (i.e. A_s and B_s) and parameters (such as matrices R and P, integers M and N, and constraint bounds) as in [10], where quadratic programming-based RHC is considered instead. (This model is originally from [7].) The LP to be solved during each time interval has 160 variables and 1180 inequality constraints. We used the ℓ_1 version of IrMPC. Results with the ℓ_∞ version were similar and hence are not reported.

Figure 4 shows the CPU times used by the optimization runs during each of the 1000 time intervals in a 10 s (real-time) simulation with sample time T_s=0.01 s. (The first interval starts at 0 s ($t-1$=0) and the last one at 9.99 s ($t-1$=999).) In order to keep the figure readable, only every

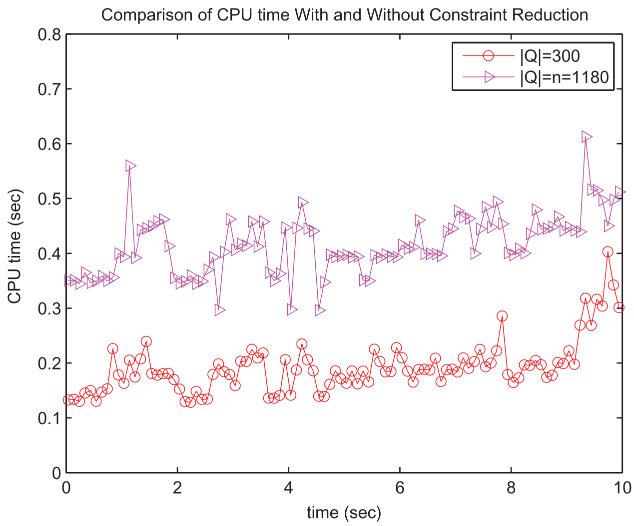

Figure 4. CPU time in seconds by IrMPC to solve the 10 s RHC process with $|Q| \approx 300$ and $|Q| = n = 1180$ (corresponding to the case with unreduced constraints). For the former, in 58 cases, $|Q|$ was increased slightly above 300 due to A^Q losing rank, as per Step 2 (iii) of Iteration IrPDIP.

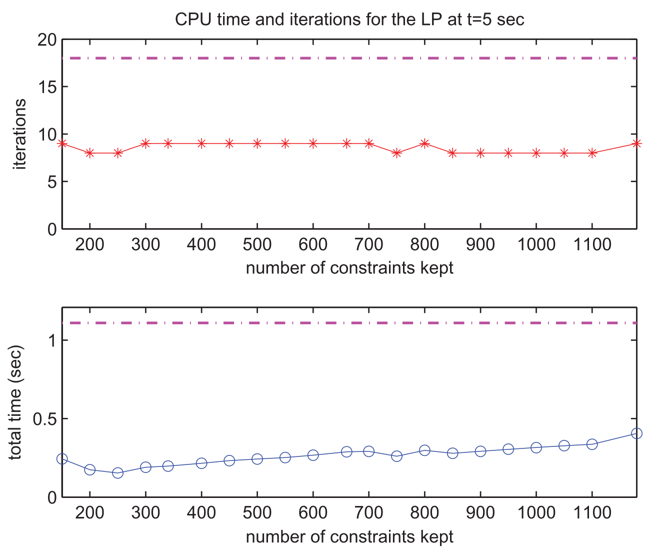

Figure 5. CPU time and iteration of solving the problem at 5 s by IrMPC with varying number $|Q|$ of kept constraints; see (blue) circles and (red) stars. MPC takes much longer to solve this problem; see the dashed (magenta) line.

Table 1. Number of problems with certain properties.

NFIPs	FIPs	SIPPs	IrMPC < rMPC*
516	484	5	461 out of 484

10th time step is shown. Note that solving every LP with constraint reduction ((red) circles) takes close to or less than half of the time it takes without constraint reduction ((magenta) triangles). Because not all constraints of (77)–(83) are dense, constraint reduction did not afford a full fourfold ($\frac{1180}{300}$) speedup.

Figure 5 shows the effect of constraint reduction on the single LP at time 5.00 s ($t-1=499$), which is a typical case. The CPU time needed to completely solve this problem decreases as the number $|Q|$ of constraints kept decreases, from 1180 constraints down to as little as 200, i.e. down to approximately 17% of all constraints. For this LP, MPC takes much more time and iterations than IrMPC.

Table 1 shows that 516 of the 1000 LPs begin with warm starts that are infeasible points (NFIPs), the remaining 484 with strictly feasible initial points (FIPs). Because we used a warm start for the initial penalty parameter, only *five* problems started with too small initial penalty parameters (SIPPs), and we observed an increase of the penalty parameter for those five problems only. For those 484 problems with strictly feasible initial points, rMPC* in [24] can be used instead of IrMPC, and we compared the respective times. For 95% (461 out of 484) of the instances, IrMPC took less time than rMPC*, presumably due to the better ability of IrMPC to handle initial points close to the constraint boundaries.

6. Conclusions

We have outlined a general framework (rIPM) for a class of constraint-reduced, dual-feasible IPMs that encompasses several previously proposed algorithms, and proved that for all methods in that class, the dual sequence converges to a single point. In order to accommodate important classes of problems for which an initial dual-feasible point is not readily available, we then proposed an ℓ_1/ℓ_∞ penalty-based extension (IrPDIP) of this framework for *infeasible* constraint-reduced primal–dual interior point. We showed that the penalty adjustment scheme in IrPDIP has the property that, under the sole assumption that the primal–dual pair is strictly feasible, the penalty parameter remains bounded.

An infeasible constraint-reduced variant of MPC (specifically, an infeasible variant of rMPC* from [24]), dubbed IrMPC, was then considered, as an instance of IrPDIP. IrMPC was analyzed, and tested on randomly generated problems and on a sequence of problems arising in an instance of model-predictive control. The results show promise for handling both infeasible initial points and nearly infeasible warm starts. Indeed, on the model-predictive control problem, IrMPC performed significantly better than (a version of) the original MPC, even when constraint reduction was turned off.

Acknowledgements

The authors wish to thank Professor Dianne O'Leary for her suggestions and insight. This work was supported by DoE grants DEFG0204ER25655, DESC0002218 and DESC0001862. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the US Department of Energy.

Notes

1. Inequality (8) is an angle condition: existence of $\omega>0$ means that the angle between b and Δy is bounded away from 90 degrees. This condition, which is weaker than (4), is sufficient for Proposition 2.2 to hold.
2. Following [24], we term 'stationary point' for (D_ρ) a point (y,z) that is feasible for (D_ρ) and, for some (x,u) such that $Ax=b$ and $x+u=\rho e$, satisfies $x(c-Ay+z)=0$ and $uz=0$. (If (y,z) is stationary for (D_ρ) and the associated (x,u) satisfies $x\geq 0$ and $u\geq 0$, (y,z) is optimal for (D_ρ).)
3. Constraints $z\geq 0$ are not 'constraint-reduced' in (D_ρ^Q). The reason is that they are known to be active at the solution, and that furthermore their contribution to normal matrix (2) is computed at no cost.
4. The question of whether Theorem 3.8 and Proposition 3.9 in [24] hold without assuming linear independence of gradients of active constraints is open. If the answer is positive, then global convergence of (y_k, z_k) as established in our Theorem 1 will hold under the sole assumption that (P)–(D) is strictly feasible (and A has full row rank).
5. The code is available from the authors.
6. Model-predictive control (MPC) is also known as receding-horizon control (RHC). In this paper, we refer to it by the acronym RHC, and reserve 'MPC' for Mehrotra's Predictor-Corrector optimization algorithm.

References

[1] P. Armand, *A quasi-Newton penalty barrier method for convex minimization problems*, Math. Program. 26 (2003), pp. 5–34.
[2] H.Y. Benson and D.F. Shanno, *An exact primal–dual penalty method approach to warmstarting interior-point methods for linear programming*, Comput. Optim. Appl. 38 (2007), pp. 371–399.
[3] H.Y. Benson, A. Sen, and D.F. Shanno, *Interior-point methods for nonconvex nonlinear programming: Convergence analysis and computational performance*, http://rutcor.rutgers.edu/~shanno/converge5.pdf (2009).
[4] H.Y. Benson, A. Sen, D.F. Shanno, and R.J. Vanderbei, *Interior-point algorithms, penalty methods and equilibrium problems*, Comput. Optim. Appl. 34 (2006), pp. 155–182.
[5] G. Dantzig and Y. Ye, *A build-up interior-point method for linear programming: Affine scaling form*. Working paper, Department of Management Science, University of Iowa, 1991.
[6] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Society for Industrial and Applied Mathmatics, Philadelphia, PA, 1990.
[7] J.W. Fletcher, *Identification of UH-60 stability derivative models in hover from flight test data*, J. Amer. Helicopter Soc. 40 (1995), pp. 32–46.
[8] G.C. Goodwin, M.M. Seron, and J.A.D. Dona, *Constrained Control and Estimation: an Optimisation Approach*, Springer, London, 2005.
[9] D. den Hertog, C. Roos, and T. Terlaky, *Adding and deleting constraints in logarithmic barrier method for linear programming problems*. Shell report, AMER 92-001, 1992.
[10] M.Y. He, M. Kiemb, A.L. Tits, A. Greenfield, and V. Sahasrabudhe, *Constraint-reduced interior-point optimization for model predictive rotorcraft control*, American Control Conference, American Automatic Control Council, Baltimore, MD, 2010, pp. 2088–2094.
[11] J.H. Jung, D.P. O'Leary, and A.L. Tits, *Adaptive constraint reduction for convex quadratic programming*, Comput. Optim. Appl. (published on-line as of March 2010), DOI: 10.1007/s10589–010–9324–8.
[12] S. Leyffer, G.L. Calva, and J. Nocedal, *Interior methods for mathematical programs with complementarity constraints*, SIAM J. Optim. 17 (2006), pp. 52–77.
[13] S. Mehrotra, *On the implementation of a primal–dual interior point method*, SIAM J. Optim. 2 (1992), pp. 575–601.
[14] S.O. Nicholls, *Column generation in infeasible predictor–corrector methods for solving linear programs*. Ph.D. Thesis, University of Maryland, 2009.
[15] F.A. Potra, *A quadratically convergent predictor–corrector method for solving linear programs from infeasible starting points*, Math. Program. Ser. A and B 67 (1994), pp. 383–406.
[16] F.A. Potra, *An infeasible-interior-point predictor–corrector algorithm for linear programming*, SIAM J. Optim. 6 (1996), pp. 19–32.
[17] R. Saigal, *On the primal–dual affine scaling method*, Technical report, Department of Industrial and Operational Engineering, The University of Michigan, 1994.
[18] R. Saigal, *A simple proof of a primal affine scaling method*, Ann. Oper. Res. 62 (1996), pp. 303–324.
[19] A. Sen and D.F. Shanno, *Convergence analysis of an interior-point method for mathematical programs with equilibrium constraints*. http://rutcor.rutgers.edu/~shanno/IPMPEC2.pdf(2006).
[20] G.W. Stewart, *On scaled projections and pseudo-inverses*, Linear Algebra Appl. 112 (1989), pp. 189–194.
[21] A.L. Tits, P.A. Absil, and W.P. Woessner, *Constraint reduction for linear programs with many inequality constraints*, SIAM J. Optim. 17 (2006), pp. 119–146.
[22] A.L. Tits, A. Wächter, S. Bakhtiari, T.J. Urban, and C.T. Lawrence, *A primal–dual interior-point method for nonlinear programming with strong global and local convergence properties*, SIAM J. Optim. 14 (2003), pp. 173–199.
[23] K. Tone, *An active-set strategy in an interior point method for linear programming*, Math. Program. 59 (1993), pp. 345–360.

[24] L.B. Winternitz, S.O. Nicholls, A.L. Tits, and D.P. O'Leary, *A constraint-reduced variant of Mehrotra's predictor–corrector algorithm*, Comput. Optim. Appl. (published on-line as of January 2011), DOI: 10.1007/s10589–010–9389–4.
[25] S.J. Wright, *Primal–Dual Interior-Point Methods*, SIAM, Philadelphia, PA, 1997.
[26] Y. Ye, *A 'build-down' scheme for linear programming*, Math. Program. 46 (1990), pp. 61–72.