Mathematical Programming
https://doi.org/10.1007/s10107-019-01454-4

FULL LENGTH PAPER

Series A

On the behavior of Lagrange multipliers in convex and nonconvex infeasible interior point methods

Gabriel Haeser[1,2] · **Oliver Hinder**[2] · **Yinyu Ye**[2]

Received: 7 September 2017 / Accepted: 26 November 2019

Abstract

We analyze sequences generated by interior point methods (IPMs) in convex and nonconvex settings. We prove that moving the primal feasibility at the same rate as the barrier parameter μ ensures the Lagrange multiplier sequence remains bounded, provided the limit point of the primal sequence has a Lagrange multiplier. This result does not require constraint qualifications. We also guarantee the IPM finds a solution satisfying strict complementarity if one exists. On the other hand, if the primal feasibility is reduced too slowly, then the algorithm converges to a point of minimal complementarity; if the primal feasibility is reduced too quickly and the set of Lagrange multipliers is unbounded, then the norm of the Lagrange multiplier tends to infinity. Our theory has important implications for the design of IPMs. Specifically, we show that IPOPT, an algorithm that does not carefully control primal feasibility has practical issues with the dual multipliers values growing to unnecessarily large values. Conversely, the one-phase IPM of Hinder and Ye (A one-phase interior point method for nonconvex optimization, 2018. arXiv:1801.03072), an algorithm that controls primal feasibility as our theory suggests, has no such issue.

Keywords Interior point methods · Lagrange multipliers · Complementarity · Nonlinear optimization · Convex optimization

Gabriel Haeser was supported by the São Paulo Research Foundation (FAPESP Grants 2013/05475-7 and 2016/02092-8) and the Brazilian National Council for Scientific and Technological Development (CNPq). Oliver Hinder was supported by the PACCAR INC stanford graduate fellowship.

✉ Oliver Hinder
ohinder@stanford.edu

Gabriel Haeser
ghaeser@ime.usp.br

Yinyu Ye
yinyu-ye@stanford.edu

[1] Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil

[2] Department of Management Science and Engineering, Stanford University, Stanford, CA 94305, USA

Published online: 21 December 2019

Mathematics Subject Classification 90C25 · 90C30 · 90C46 · 90C51

1 Introduction

This paper studies sequences generated by interior point methods (IPMs) that converge to Karush-Kuhn-Tucker (KKT) points of

$$\text{minimize } f(x) \tag{1a}$$
$$\text{subject to } a(x) + s = 0 \tag{1b}$$
$$s \geq 0, \tag{1c}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the inequality constraints $a : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions.

The central path generated by sequences of log barrier problems was introduced by McLinden [31] for convex minimization subject to non-negativity constraints and generalized to linear inequalities by Sonnevend [39]. Megiddo [32] analyzed the path of primal-dual IPMs for linear programming and showed this path converges to a point satisfying strict complementarity. Güler and Ye [22] generalized this result to a large class of path-following IPMs for linear programming. Finding a strictly complementary solution is necessary to guarantee the super-linear convergence of IPMs for quadratic programs [43, Proposition 5.1]. Furthermore, finding a strictly complementary solution for problems with nonconvex constraints ensures that the critical cone is reduced to a subspace. This subspace gives an efficient way to verify the second-order conditions by computing the least eigenvalue of the Hessian of the Lagrangian restricted to this subspace [11, Theorem 4.4.2]. In the nonlinear context, a strictly complementary solution may not always exist, but if it does, we would like to obtain it.

The results mentioned above implicitly avoid the issue of unbounded dual variables by starting from a strictly feasible point. However, this is rarely done in practice, as infeasible-start algorithms are often used [29,33]. Mizuno et al. [34] studies the sequences generated by these infeasible start algorithms for linear programming without assuming the existence of an interior point. They show that moving the constraint violation at the same rate as the barrier parameter μ guarantees that the dual multipliers are bounded. The boundedness of dual multipliers is practically important because the linear system solved at each iteration of an IPM can become poorly conditioned as the dual multipliers get large, making the linear system more difficult to solve, particularly using iterative methods [21]. Some of our theoretical contributions can be viewed as extensions of this work to convex and nonconvex optimization.

One alternative and elegant solution to these issues is the homogeneous algorithm [3,4,44]. For convex problems, the homogeneous algorithm is guaranteed to produce a bounded sequence that converges to a maximally complementary solution. For linear programming, this guarantees that if the problem is feasible the algorithm will converge with bounded dual variables. However, it is unknown how to extend the homogeneous algorithm into nonconvex optimization.

While many IPMs for general nonconvex optimization problems have been developed, there is little analysis of the sequences they generate. For example, it is unclear if IPMs can generate maximal complementarity solutions in the presence of nonconvexity. Furthermore, results showing that the sequence of dual iterates are bounded rely on the set of dual multipliers being bounded (which is equivalent to the Mangasarian-Fromovitz constraint qualification [19]). This assumption may be too restrictive because many practical optimization problems may lack a strict relative interior and therefore have an unbounded set of dual multipliers. For instance, we found that this is the case for 64 out of the 95 NETLIB problems (see "Appendix A").

Primal and dual sequences generated by nonconvex optimization algorithms such as IPMs, augmented Lagrangian methods and sequential quadratic programming have been analyzed in a number of works [5,9,10,23,37]. However, we are only aware of feasible IPMs being considered. Moreover, these studies have been focused on determining primal convergence to a KKT point, despite unboundedness of the dual sequence. Instead, we focus on guaranteeing boundedness and maximal complementarity of the dual sequence.

Next, we explain the current state of knowledge of primal and dual sequences generated by IPMs for linear programming. In particular, let $f(x) := g^{\mathrm{T}}x$, with constraints $a(x) + s = 0$, $s \geq 0$, where $a(x) := Mx - p$, M is a matrix and g, p are vectors. Many IPMs for linear programming compute direction (d_x^k, d_y^k, d_s^k) at each iteration k satisfying

$$M^T d_y^k = -\eta^k (g + M^T y^k) \tag{2a}$$

$$M d_x^k + d_s^k = -\eta^k (M x^k + s^k - p) \tag{2b}$$

$$S^k d_y^k + Y^k d_s^k + S^k y^k = (1 - \eta^k)\mu^k e, \tag{2c}$$

where Y^k and S^k are the diagonal matrices defined by y^k and s^k, and e is a vector of ones. The values $\mu^{k+1} := (1 - \eta^k)\mu^k$ and $\eta^k \in (0, 1)$ are chosen, for example, using a predictor-corrector technique [33], see also [36, Algorithm 14.3]. The iterates are updated according to $(x^{k+1}, y^{k+1}, s^{k+1}) \leftarrow (x^k, y^k, s^k) + \alpha^k (d_x^k, d_y^k, d_s^k)$, where $\alpha^k \in (0, 1]$ is the step size. Methods that choose their iterates in this way reduce the primal feasibility and complementarity at approximately the same rate [30,33,44], which we formalize as follows. Suppose the IPM converges to an optimal solution as $\mu^k \to 0$. Then a subsequence of iterates satisfy $x^k \to x^*$, $s^k \to s^*$, and:

$$a(x^*) + s^* = 0 \tag{3a}$$

$$b\mu^k \leq s_i^k y_i^k \leq c\mu^k \quad \text{for all } i \tag{3b}$$

$$\ell\mu^k \leq a_i(x^k) + s_i^k \leq u\mu^k \quad \text{for all } i \tag{3c}$$

$$\|\nabla_x \mathcal{L}(x^k, y^k)\| \leq d\mu^k (\|y^k\|_1 + 1) \tag{3d}$$

$$s^k, y^k \geq 0, \tag{3e}$$

where $\mu^k > 0$ is the barrier parameter, $0 < b \leq c$, $0 < \ell \leq u$, $d \geq 0$ are real constants independent of k, $\|\cdot\|_p$ denotes the ℓ_p norm, $\|\cdot\|$ the Euclidean norm, and the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is

$$\mathcal{L}(x, y) := f(x) + y^T a(x). \tag{4}$$

Inequality (3b) ensures perturbed complementarity approximately holds. Inequality (3c) guarantees that primal feasibility is reduced at the same rate as complementarity. Inequality (3d) ensures that scaled dual feasibility is reduced fast enough. The set of inequalities (3) has a natural interpretation as a 'shifted log barrier': a sequence of approximate KKT points to the problem,

$$\text{minimize} \;\; f(x) - \mu^k \sum_{i=1}^{m} \log(\mu^k r_i - a_i(x)),$$

with the vector r satisfying $\ell \leq r_i \leq u$.

The main contribution of Mizuno, Todd, and Ye [34] was to show that IPMs for linear programming have bounded Lagrange multiplier sequences and satisfy strict complementarity when (3) holds. Hinder and Ye [26] show it is also possible to develop IPMs that satisfy (3) even if f and a are nonlinear. In particular, they give an IPM where, if the primal variables are bounded and the algorithm does not return a certificate of local primal infeasibility, a subsequence of the iterates satisfy (3). This motivates us to show given a sequence satisfying (3), even if the objective and constraints are *nonlinear* the dual multipliers are still, under general conditions, well-behaved.

1.1 Summary of contributions

Now, assuming conditions (3) and that the problem is convex (or certain sufficient conditions for local optimality hold) we show:

(a) If there exists a Lagrange multiplier at the point x^*, then the sequence of Lagrange multipliers approximations $\{y^k\}$ is bounded (see Theorems 1 and 2 for the convex and nonconvex case respectively).
(b) If $y^k \rightarrow y^*$, then among the set of Lagrange multipliers at the point x^*, the point y^* is maximally complementary (see Theorems 3 and 4).

Consider the case that (3c) does not hold, i.e., the primal feasibility is not being reduced at the same rate as complementarity. We argue that this is poor algorithm design, because if problem (1) is convex then:

(a). If we reduce the primal feasibility faster than the barrier parameter μ^k and the set of dual multipliers at the point x^* is unbounded, then $\|y^k\| \rightarrow \infty$ (see Theorem 5).
(b) If we reduce the primal feasibility slower than the barrier parameter μ^k and $y^k \rightarrow y^*$, then y^* is a minimally complementary Lagrange multiplier associated with x^* (see Theorem 6).

Our central claim is that many implemented interior point methods, especially for nonlinear optimization, such as IPOPT [42], suffer from the problems described above because they fail to control the rate at which they reduce primal feasibility (specifically IPOPT suffers from deficiency (a)). For linear programs, these methods solve systems of the form [42, equation (9)]

$$M^T d_y^k = -\left(g + M^T y^k\right) \tag{5a}$$

$$M d_x^k + d_s^k = -\left(M x^k + s^k - p\right) \tag{5b}$$

$$S^k d_y^k + Y^k d_s^k + S^k y^k = \mu^k e, \tag{5c}$$

where the notation follows (2). Equation (5b) aims to reduce the constraint violation to zero at each iteration. Contrast (5b) with Eq. (2b) that aims to reduce the constraint violation by η^k, the same amount by which complementarity is reduced. As we demonstrate in Sect. 4, a consequence of the implementation choices in IPOPT, primal feasibility is usually reduced faster than complementarity. Therefore, as our theory suggests, these IPMs have issues with the Lagrange multipliers sequence diverging.

IPOPT attempts to circumvent this issue by perturbing the original constraint $a(x) \leq 0$ to create an artificial interior as follows:

$$a(x) \leq \delta e, \tag{6}$$

for some $\delta > 0$ (see Section 3.5 of [42]). While this technically solves the issue as the theoretical assumptions of [41] are now met, it is not an elegant solution and causes undesirable behavior. For example, we show in Sect. 4 that the dual variable may still *spike* before converging. Furthermore, if δ is selected to be large, the constraints will be only loosely satisfied at the final solution.

We proceed as follows. Section 1.2 gives a simple example illustrating the phenomena studied. Section 2 shows that reducing the primal feasibility at the same rate as complementarity ensures the dual multiplier sequence remains bounded and satisfies maximal complementarity. Section 3 explains that reducing the constraint violation too quickly causes the dual multiplier sequence to be unbounded, while reducing it too fast causes the them to tend towards a minimal complementarity solution. Section 4 shows empirically how strategies that reduce the constraint violation too fast, such as the one employed by IPOPT, can have issues with extremely large dual multipliers. Section 5 presents our final remarks.

1.2 A simple example demonstrating phenomena

Consider the following simple linear programming problem:

$$\text{minimize } 0 \tag{7a}$$

$$\text{subject to } x \leq 1 \tag{7b}$$

$$x \geq 1. \tag{7c}$$

By adding a feasibility perturbation $\delta > 0$ and a log barrier term $\mu \geq 0$, we get

$$\text{minimize} \ -\mu \log(x-1+\delta) - \mu \log(1-x+\delta) \tag{8a}$$
$$\text{subject to } x \geq 1-\delta \tag{8b}$$
$$x \leq 1+\delta. \tag{8c}$$

The associated KKT system is

$$y_1 - y_2 = 0 \tag{9a}$$
$$x + s_1 = 1+\delta \tag{9b}$$
$$x - s_2 = 1-\delta \tag{9c}$$
$$s_1 y_1 = \mu \quad s_2 y_2 = \mu \tag{9d}$$
$$y_1, y_2 \geq 0 \quad s_1, s_2 \geq 0. \tag{9e}$$

Observe that the orignal problem (corresponding to $\delta = \mu = 0$) has a unique optimal primal solution at $x^* := 1$ and $s^* := (0, 0)$, with dual solutions $y_1^* = y_2^*$ for any $y_2^* \geq 0$. Therefore the set of dual variables is unbounded. However, for any $\delta, \mu > 0$, the solution to system (9) is

$$x = 1 \quad s_1 = \delta \quad s_2 = \delta \quad y_1 = \frac{\mu}{\delta} \quad y_2 = \frac{\mu}{\delta}.$$

From these equations, we can see that if δ and μ move at the same rate, then both strict complementarity and boundedness of the dual variables will be achieved. But if δ reduces faster than μ, i.e., $\delta/\mu \to 0^+$, then the dual variables sequence is unbounded. Alternatively, if δ moves slower than μ, i.e., $\delta/\mu \to \infty$, then strict complementarity will not hold.

Now, if $\delta > 0$ is fixed at a small value as in the IPOPT strategy (6), the dual sequence will initially grow very fast before stabilizing when the barrier parameter μ is sufficiently reduced. We confirm this hypothesis by solving the linear programming problem (7) with perturbations $\delta > 0$ using IPOPT, and we compare it with a *well-behaved IPM* [26] that moves complementarity at the same rate as primal feasibility, that is, satisfies (3a)–(3e). For this experiment, we turn off IPOPT's native perturbation strategy (6). In Fig. 1 we plot the maximum dual variables at each iteration, given by the two methods for different perturbation sizes. While perturbing the linear program prevents the dual variables of IPOPT from increasing indefinitely, the dual variables still *spike*. For example, with $\delta = 10^{-8}$ the maximum dual variable of IPOPT peaks at 10^4 on iteration 4 before sharply dropping on the next iteration to 9. Picking a smaller δ, e.g., $\delta = 10^{-5}$, ensures a smaller peak at the cost of solving the problem to a lower accuracy. On the other hand, the maximum dual variable for the well-behaved IPM remains below 1.5 irrespective of the perturbation size. Furthermore, with $\delta = 0.0$ the curve for the well-behaved IPM is essentially flat.

More thorough numerical experiments are given in Sect. 4, but first we establish our general theory.

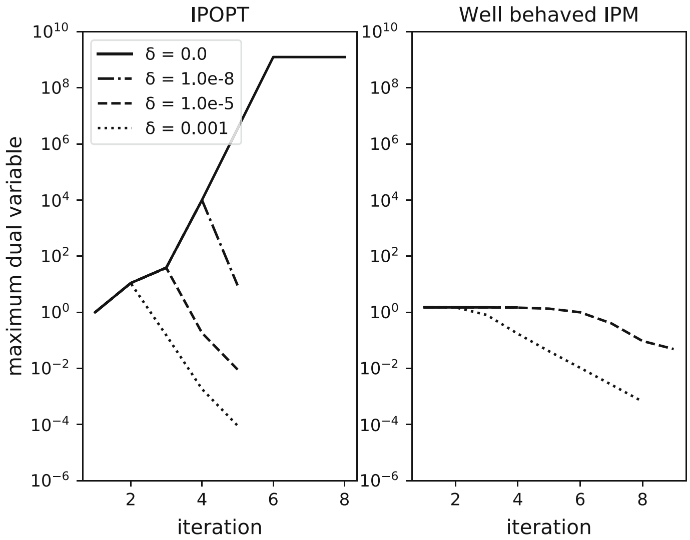

Fig. 1 Comparison of the maximum dual variable value (vertical axis) against iterations (horizontal axis) using IPOPT and a well-behaved IPM [26] as the perturbation δ is changed

2 Boundedness and maximal complementarity

In this section, we show that when feasibility is reduced at the same rate as complementarity, the dual variables are bounded and satisfy maximal complementarity. But first we establish some basic results for convex problems on the optimality of limit points of sequences satisfying (3).

Notation. When it is clear from the context, we omit a quantifier "$\forall k$" when stating properties of every sufficiently large element of a sequence indexed by $k = 1, 2, \ldots, \infty$. The Euclidean norm is denoted by $\| \cdot \|$; otherwise the ℓ_p-norm (we only use $p = 1, \infty$) is denoted by $\| \cdot \|_p$.

The following lemma gives a sufficient sequential condition for global optimality in the convex case. In our setting, the lemma is slightly more general than results found in the literature, e.g., see [27, Corollary 3.1], [9, Theorem 4.2], [24, Theorem 2.2], and [20, Theorem 3.2]. Our condition is in fact equivalent to the one from [20] but with a redundant assumption omitted.

Lemma 1 *If f and a_i for $i = 1, \ldots, m$ are convex functions, and $\{(x^k, y^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m$ are such that*

1. $x^k \to x^*$ *with* $a(x^*) \leq 0$,
2. $y^k \geq 0$,
3. $\liminf a(x^k)^{\mathrm{T}} y^k \geq 0$,
4. $\nabla_x \mathcal{L}(x^k, y^k) \to 0$,

then, x^* is a solution of (1).

Proof Given x with $a(x) \leq 0$, we have

$$f(x) \geq \mathcal{L}(x, y^k) \geq \mathcal{L}(x^k, y^k) + \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x - x^k).$$

Hence,

$$a(x^k)^{\mathrm{T}} y^k \leq f(x) - f(x^k) + \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^k - x). \tag{10}$$

Thus, for $x = x^*$, we have $\limsup a(x^k)^{\mathrm{T}} y^k \leq 0$. The assumption gives $a(x^k)^{\mathrm{T}} y^k \to 0$. Taking the limit in (10) we have $f(x) \geq f(x^*)$ and the result follows. □

The following lemma gives a sufficient condition for verifying the conditions of Lemma 1 under our slack variable formulation, which suits better our interior point framework.

Lemma 2 *If f and a_i for $i = 1, \ldots, m$ are convex functions, and $\{(x^k, y^k, s^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ are such that*

1. $x^k \to x^*$ *with* $a(x^*) \leq 0$ *and* $s^k \to -a(x^*)$,
2. $y^k \geq 0$ *and* $s^k \geq 0$,
3. $(y^k)^{\mathrm{T}} s^k \to 0$,
4. $a_i(x^k) + s_i^k \geq 0$ *for all* $i : a_i(x^*) = 0$,
5. $\nabla_x \mathcal{L}(x^k, y^k) \to 0$,

then, x^* is a solution of (1).

Proof For $i : a_i(x^*) = 0$, we have $a_i(x^k) y_i^k \geq -s_i^k y_i^k \to 0$, while if $a_i(x^*) < 0$, we have $y_i^k \to 0$. The result follows from Lemma 1. □

We note that even in the nonconvex case, the existence of sequences satisfying the conditions of Lemmas 1 and 2 are necessary at a local solution x^*, without constraint qualifications. This follows from the necessary existence of sequences $x^k \to x^*$, $y^k \geq 0$ with $\nabla_x \mathcal{L}(x^k, y^k) \to 0$, $a_i(x^k) y_i^k \to 0$ for all i, when x^* is a local solution, given in [9, Theorem 3.3], by defining $s_i^k := \max\{0, -a_i(x^k)\}$ for all i and all k. See also [23].

2.1 Boundedness of the dual sequence

The boundedness of the dual sequence is an important property because the algorithm is otherwise prone to numerical instabilities.

In Theorem 1, we consider problems involving convex functions where the algorithm is converging to a KKT point. We show that if the primal feasibility, (scaled) dual feasibility and complementarity converge at the same rate, then the dual sequence $\{y^k\}$ is bounded. We refer to [34, Theorem 4] for a more general result when the functions f and a are linear. This result is extended in Theorem 2 to situations where the optimization problem may involve nonconvex functions.

We try to present as few assumptions as possible; for example, assumptions are often placed only on constraints that are active at the limit. However, since in practice the active constraints are unknown, we advocate using IPMs that satisfy (3). All assumptions on the sequence of iterates made on theorems in this section can be subsumed by (3) ignoring constant factors.

Theorem 1 *Suppose f and a_i for $i = 1, \ldots, m$ are convex functions and $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ with $\mu^k > 0$ for all k and $\mu^k \to 0$ are such that:*

1. $x^k \to x^*$ *with* $a(x^*) \le 0$ *and* $s^k \to -a(x^*)$,
2. $y^k \ge 0$ *and* $s^k \ge 0$,
3. *for some* $c \ge 0$, $(y^k)^{\mathrm{T}} s^k \le \mu^k c$,
4. *for some* $0 < \ell \le u$, $\mu^k \ell \le a_i(x^k) + s_i^k \le \mu^k u$ *for all* $i : a_i(x^*) = 0$,
5. *for some* $d \ge 0$, $\|\nabla_x \mathcal{L}(x^k, y^k)\| \le d\mu^k(\|y^k\|_1 + 1)$.

Then, x^* is a solution of (1). If x^* is a KKT point, then

$$\limsup \|y^k\|_1 \le \frac{2u}{\ell}\|y^*\|_1 + \frac{4c}{\ell} m + \frac{2(c+d)}{\ell},$$

where y^* is any Lagrange multiplier associated with x^*, i.e., $\nabla \mathcal{L}(x^*, y^*) = 0$, $y^* \ge 0$, and $a(x^*)^T y^* = 0$.

Proof We have by convexity of $\mathcal{L}(x, y^k)$ in x that

$$f(x^*) \ge \mathcal{L}(x^*, y^k) \ge \mathcal{L}(x^k, y^k) + \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^* - x^k),$$

which gives

$$f(x^*) - f(x^k) \ge a(x^k)^{\mathrm{T}} y^k + \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^* - x^k). \tag{11}$$

Also,

$$a(x^k)^{\mathrm{T}} y^k = (a(x^k) + s^k)^{\mathrm{T}} y^k - (s^k)^{\mathrm{T}} y^k \ge \sum_{i: a_i(x^*)=0} \mu^k \ell y_i^k + \sum_{i: a_i(x^*)<0} (a_i(x^k) + s_i^k) y_i^k - \mu^k c.$$

Since $s_i^k y_i^k \ge 0$ and $a_i(x^k) y_i^k \ge \frac{a_i(x^k)}{s_i^k} \mu^k c \ge -2\mu^k c$ for $i : a_i(x^*) < 0$ and sufficiently large k, we have

$$\begin{aligned} a(x^k)^{\mathrm{T}} y^k &\ge \ell \mu^k \sum_{i: a_i(x^*)=0} y_i^k - \sum_{i: a_i(x^*)<0} 2c\mu^k - c\mu^k \\ &= \ell \mu^k \|y^k\|_1 - \sum_{i: a_i(x^*)<0} (2c + \ell y_i^k)\mu^k - c\mu^k. \end{aligned} \tag{12}$$

Also, $\nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^* - x^k) \ge -d\mu^k(\|y^k\|_1 + 1)\|x^* - x^k\|$. Hence, substituting this and (12) back in (11) we get

$$\begin{aligned} f(x^*) - f(x^k) &\ge \ell \mu^k \|y^k\|_1 - d\mu^k(\|y^k\|_1 + 1)\|x^* - x^k\| \\ &- \sum_{i: a_i(x^*)<0} (2c + \ell y_i^k)\mu^k - c\mu^k. \end{aligned}$$

We can take k large enough that $\ell\mu^k\|y^k\|_1 - d\mu^k(\|y^k\|_1+1)\|x^*-x^k\| \geq \frac{\ell}{2}\mu^k\|y^k\|_1 - d\mu^k$, so that,

$$f(x^*) - f(x^k) \geq \frac{\ell}{2}\mu^k\|y^k\|_1 - \sum_{i:a_i(x^*)<0}(2c+\ell y_i^k)\mu^k - (c+d)\mu^k. \tag{13}$$

Since $\ell > 0$ and $y_i^k \to 0$ for $i : a_i(x^*) < 0$, we have by (13) that $\mu^k\|y^k\|_1 \to 0$. This implies that $\nabla_x \mathcal{L}(x^k, y^k) \to 0$, and we can use Lemma 2 to conclude that x^* is a solution.

On the other hand, let $y^* \in \mathbb{R}^m$ be a Lagrange multiplier associated with x^*. Then, $f(x^*) = \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x^k, y^*)$, which, combining with (13) yields

$$\frac{\ell}{2}\mu^k\|y^k\|_1 - \sum_{i:a_i(x^*)<0}(2c+\ell y_i^k)\mu^k - (c+d)\mu^k \leq f(x^*) - f(x^k) \leq a(x^k)^{\mathrm{T}}y^*. \tag{14}$$

But $a(x^k)^{\mathrm{T}}y^* = (a(x^k)+s^k)^{\mathrm{T}}y^* - (s^k)^{\mathrm{T}}y^* \leq \mu^k u\|y^*\|_1$, which implies, by dividing (14) by $\frac{\ell}{2}\mu^k$, that

$$\|y^k\|_1 \leq \frac{2u}{\ell}\|y^*\|_1 + \sum_{i:a_i(x^*)<0}\left(\frac{4c}{\ell} + 2y_i^k\right) + \frac{2(c+d)}{\ell}.$$

Since $y_i^k \to 0$ for $i : a_i(x^*) < 0$, the result follows. □

Optimization problems with complementarity constraints are an important class of nonconvex optimization problems where the Mangasarian-Fromovitz constraint qualification fails. Typically, specialized IPMs for these problems are developed [12,28]. The following corollary focuses on convex programs with complementarity constraints. It shows that any general purpose IPM satisfying (3) has a bounded dual multipliers sequence under general conditions.

Corollary 1 *Let $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ with $\mu^k > 0$ and $\mu^k \to 0$ be such that assumptions 1-5 of Theorem 1 hold. Assume that problem (1) is a convex program with complementarity constraints, that is:*

$$\text{minimize } f(x) \tag{15a}$$
$$\text{subject to } g(x) \leq 0 \tag{15b}$$
$$x_i x_j \leq 0 \quad (i,j) \in \mathbf{C} \tag{15c}$$
$$x \geq 0, \tag{15d}$$

where $\mathbf{C} \subseteq \{1,\ldots,n\} \times \{1,\ldots,n\}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex for $i = 1,\ldots,p$ ($p \leq m$). Assume $x_i^* + x_j^* > 0$ for all $(i,j) \in \mathbf{C}$. Under these assumptions, x^* is a local minimizer. Furthermore, if x^* is a KKT point then $\{y^k\}$ is bounded.

Proof To prove this result, it is sufficient to show that we are implicitly generating a sequence satisfying the assumptions of Theorem 1, where (1) is replaced by the convex program

$$\text{minimize } f(x) \tag{16a}$$
$$\text{subject to } g(x) \le 0 \tag{16b}$$
$$x_i^* x_j \le 0 \quad (i, j) \in \mathbf{F} \tag{16c}$$
$$x \ge 0, \tag{16d}$$

where $\mathbf{F} = \{(l, k) \in \mathbf{C} : x_l^* > 0\} \cup \{(k, l) : (l, k) \in \mathbf{C},\ x_k^* > 0\}$. By the strict complementarity assumption, we deduce that if $(i, j) \in \mathbf{C}$ then either $(i, j) \in \mathbf{F}$ or $(j, i) \in \mathbf{F}$; i.e., there is a one-to-one correspondence between constraints in (15) and (16). Therefore if $\tilde{x}^* \in \{x : \|\tilde{x}^* - x^*\|_\infty \le \frac{1}{2} \min_{(i,j)\in F} x_i^*\}$ is feasible for (15), then $\tilde{x}^*$ is feasible for (16). We deduce that any minimizer for (16) is a local minimizer for (15). Now, $\frac{|x_i^k x_j^k - x_i^* x_j^k|}{\mu^k} = |x_i^k - x_i^*| \frac{x_j^k}{\mu^k} \le |x_i^k - x_i^*| \frac{u}{x_i^k} \to 0$ for all $(i, j) \in \mathbf{F}$. Hence, the sequence (x^k, y^k, s^k, μ^k) satisfies the assumptions of Theorem 1, where (1) is replaced by (16). □

We now present a nonconvex version of Theorem 1. For this, we assume that the limit point x^* satisfies a sufficient optimality condition based on the star-convexity concept described below. This definition is a local version of the one from [35].

Definition 1 Let a function $q : \mathbb{R}^n \to \mathbb{R}$, a point $x^* \in \mathbb{R}^n$, and a set $S \subseteq \mathbb{R}^n$ be given. We say that q is star-convex around x^* on S when

$$q(\alpha x + (1-\alpha)x^*) \le \alpha q(x) + (1-\alpha) q(x^*) \text{ for all } \alpha \in [0, 1] \text{ and } x \in S.$$

Theorem 2 *Let* $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ *with* $\mu^k > 0$ *and* $\mu^k \to 0$ *be such that*

1. $x^k \to x^*$ *with* $a(x^*) \le 0$ *and* $s^k \to -a(x^*)$,
2. $y^k \ge 0$ *and* $s^k \ge 0$,
3. *for some* $c \ge 0$, $(y^k)^{\mathrm{T}} s^k \le \mu^k c$,
4. *for some* $0 < \ell \le u$, $\mu^k \ell \le a_i(x^k) + s_i^k \le \mu^k u$ *for all* $i : a_i(x^*) = 0$,
5. *for some* $d \ge 0$, $\|\nabla_x \mathcal{L}(x^k, y^k)\| \le d\mu^k(\|y^k\|_1 + 1)$,
6. x^* *is a KKT point with Lagrange multiplier* y^*,
7. *There exist* $\theta \ge 0$ *and a neighborhood* $\mathcal{B}$ *of* x^* *such that* $\hat{\mathcal{L}}_k(x) := \mathcal{L}(x, y^k) + \theta a(x)^{\mathrm{T}} Y^k a(x)$ *for all* k *and* $\hat{\mathcal{L}}_*(x) := \mathcal{L}(x, y^*) + \theta a(x)^{\mathrm{T}} Y^* a(x)$ *are star-convex around* x^* *on* $\mathcal{B}$, *where* $Y^k = diag(y^k)$ *and* $Y^* = diag(y^*)$.

Then, $\{y^k\}$ *is bounded.*

Proof From the definition of star-convexity of $\hat{\mathcal{L}}_k$, taking limit in α, we have

$$
\begin{aligned}
f(x^*) + \theta a(x^*)^{\mathrm{T}} Y^k a(x^*) &\geq \hat{\mathcal{L}}_k(x^*) \geq \hat{\mathcal{L}}_k(x^k) + \nabla_x \hat{\mathcal{L}}_k(x^k)^{\mathrm{T}}(x^* - x^k) \\
&= \mathcal{L}(x^k, y^k) + \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^* - x^k) + \theta a(x^k)^{\mathrm{T}} Y^k a(x^k) \\
&\quad + \sum_{i=1}^m 2\theta y_i^k a_i(x^k) \nabla a_i(x^k)^{\mathrm{T}}(x^* - x^k).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f(x^*) - f(x^k) &\geq -\theta a(x^*)^{\mathrm{T}} Y^k a(x^*) + \theta a(x^k)^{\mathrm{T}} Y^k a(x^k) + a(x^k)^{\mathrm{T}} y^k \\
&\quad + \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^* - x^k) + \sum_{i=1}^m 2\theta y_i^k a_i(x^k) \nabla a_i(x^k)^{\mathrm{T}}(x^* - x^k). \qquad (17)
\end{aligned}
$$

We proceed to bound the right-hand side of (17). Note that $-\theta a(x^*)^{\mathrm{T}} Y^k a(x^*) = -\sum_{i:a_i(x^*)<0} \theta y_i^k a_i(x^*)^2 \geq -\sum_{i:a_i(x^*)<0} \frac{a_i(x^*)^2}{s_i^k} \theta c \mu^k \geq \sum_{i:a_i(x^*)<0} 2a_i(x^*) c\theta \mu^k$, while $\theta a(x^k)^{\mathrm{T}} Y^k a(x^k) \geq 0$.

As in the proof of Theorem 1, we have that (12) holds; that is,

$$
a(x^k)^{\mathrm{T}} y^k \geq \ell \mu^k \|y^k\|_1 - \sum_{i:a_i(x^*)<0} (2c + \ell y_i^k)\mu^k - c\mu^k.
$$

Clearly, $\nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^* - x^k) \geq -d\mu^k \|y^k\|_1 \|x^* - x^k\| - d\mu^k$ when $\|x^* - x^k\| \leq 1$. To bound the last term in (17), note that for $i : a_i(x^*) < 0$, $-|y_i^k a_i(x^k)| \geq \frac{a_i(x^k)}{s_i^k} c\mu^k \geq -2c\mu^k$, and for $i : a_i(x^*) = 0$, we have $-|a_i(x^k) y_i^k| \geq -u\mu^k y_i^k$ if $a_i(x^k) \geq 0$ and $-|a_i(x^k) y_i^k| \geq \ell \mu^k y_i^k - c\mu^k$ if $a_i(x^k) < 0$. Therefore,

$$
\begin{aligned}
&\sum_{i=1}^m 2\theta y_i^k a_i(x^k) \nabla a_i(x^k)^{\mathrm{T}}(x^* - x^k) \\
&\quad \geq -2\theta \sum_{i=1}^m |a_i(x^k) y_i^k| \|\nabla a_i(x^k)\| \|x^k - x^*\| \\
&\quad \geq \sum_{i:a_i(x^*)<0} -4\theta c\mu^k \|\nabla a_i(x^k)\| \\
&\quad\quad + \sum_{i:a_i(x^*)=0} 2\theta \min\{-u\mu^k y_i^k, \ell \mu^k y_i^k - c\mu^k\} \|\nabla a_i(x^k)\| \|x^k - x^*\|. \qquad (18)
\end{aligned}
$$

Note that $\min\{-u\mu^k y_i^k, \ell\mu^k y_i^k - c\mu^k\}$ is equal to $-u\mu^k y_i^k$ if $y_i^k \geq \frac{c}{\ell + u}$, while it is bounded by a constant times μ^k otherwise. Hence, substituting all bounds obtained back into (17), we get for some constant $C \geq 0$ the following:

$$f(x^*) - f(x^k) \ge -C\mu^k + \ell\mu^k\|y^k\|_1 - d\mu^k\|y^k\|_1\|x^* - x^k\| + \sum_{i:y_i^k \ge \frac{c}{\ell+u}} -u\mu^k y_i^k \|\nabla a_i(x^k)\|\|x^* - x^k\|.$$

Thus, we can take k large enough such that $f(x^*) - f(x^k) \ge -C\mu^k + \frac{\ell}{2}\mu^k\|y^k\|_1$.

Since $\nabla\hat{\mathcal{L}}_*(x^*) = 0$ and $\hat{\mathcal{L}}_*$ is star-convex, we have $\hat{\mathcal{L}}_*(x^k) \ge \hat{\mathcal{L}}_*(x^*) = f(x^*)$, giving

$$\begin{aligned} -C\mu^k + \frac{\ell}{2}\mu^k\|y^k\|_1 &\le a(x^k)^{\mathrm{T}}y^* + \theta a(x^k)^{\mathrm{T}}Y^*a(x^k) \\ &= \sum_{i:a_i(x^*)=0} (a_i(x^k) + \theta a_i(x^k)^2)y_i^*. \end{aligned} \tag{19}$$

For $i : a_i(x^*) = 0$, we have for k large enough that $a_i(x^k) + \theta a_i(x^k)^2 \le 2a_i(x^k)$ if $a_i(x^k) \ge 0$ and $a_i(x^k) + \theta a_i(x^k)^2 \le \frac{1}{2}a_i(x^k)$ if $a_i(x^k) \le 0$, where $a_i(x^k) \le u\mu^k - s_i^k \le u\mu^k$. It follows that the right-hand side of (19) is bounded by a constant times μ^k. Therefore, dividing by μ^k shows that $\{y^k\}$ is bounded. □

Remark 1 Given the bound $|a_i(x^k)y_i^k| \le \max\{u\mu^k y_i^k, c\mu^k - \ell\mu^k y_i^k\}$ obtained in (18) for $i : a_i(x^*) = 0$, assumptions 1-5 in Theorem 2 together with the assumption $\mu^k\|y^k\|_1 \to 0$ imply assumption 6 under weak constraint qualifications [5–8]. Also, assumption 6 and the star-convexity of $\hat{\mathcal{L}}_*$ in assumption 7 imply that x^* is a local solution.

Remark 2 Although we have decided by a clearer presentation, one could get the result under a weaker assumption than Assumption 7 of Theorem 2. In particular, suppose that $\hat{\mathcal{L}}_k(x) := \mathcal{L}(x, y^k) + \sum_{i=1}^m \theta_i^k a_i(x)^2$. Further assume $\theta_i^k \le C\mu^k$ for some $C \ge 0$ when $a_i(x^*) < 0$, and one of the following two conditions hold for all i such that $a_i(x^*) = 0$,

1. $\theta_i^k \le \theta(y_i^k + \|y^k\|_1\mathcal{I}[a_i \equiv -a_j \text{ for some } j \ne i])$, where $\mathcal{I}[\cdot]$ is the indicator function, or
2. $\theta_i^k \le \theta\|y^k\|_1$, under a strict complementarity assumption, namely, that $\{y_i^k\}$ is bounded away from zero,

with some $\theta > 0$. Note that by taking $\theta_i^k := \theta y_i^k$ with condition one we subsume Assumption 7 of Theorem 2. Condition one is useful when an equality constraint $a_i(x) = 0$ is represented as two inequalities $a_i(x) \le 0$ and $a_j(x) := -a_i(x) \le 0$. In that case, we may select θ_i^k and θ_j^k considerably larger, namely, proportional to the sum of all dual variables (instead of only the one correspondent to constraint i and j, respectively). The second condition says that we may consider this larger θ_i^k for all constraints (proportional to the sum of all dual variables), as long as we have strict complementarity.

The main modification to the proof of Theorem 2 would be on the bound of $|\theta_i^k a_i(x^k)|$ in (18). For the first condition, with equality constraints split as two inequalities, the bound $-|a_i(x^k)y_i^k| \ge -u\mu^k y_i^k$ holds regardless of the sign of $a_i(x^k)$. For the second condition one gets $-|\theta_i^k a_i(x^k)| \geq \theta_i^k \min\{-u\mu^k, \ell\mu^k - s_i^k\}$, and the strict complementarity assumption would give $-s_i^k \geq -\mu^k \frac{u}{y_i^k}$ with $\frac{u}{y_i^k}$ bounded. The result would now follow as in the proof of Theorem 2.

Note that functions $\hat{\mathcal{L}}_k$, in which we require star-convexity, are closely related to the sharp Lagrangian function [38], where we replace the ℓ_2-norm of $a(x)$ by a weighted ℓ_2-norm squared.

2.2 Maximal complementarity

We now focus our attention on obtaining maximal complementarity of the dual sequence under a set of algorithmic assumptions more general than the ones described in (3).

We say that a Lagrange multiplier y^* associated with x^* is maximally complementary if it has the maximum number of non-zero components among all Lagrange multipliers associated with x^*. Note that a maximally complementary multiplier always exists, because any convex combination of Lagrange multipliers is also a Lagrange multiplier. If a maximally complementary Lagrange multiplier y^* has a component $y_i^* = 0$ with $a_i(x^*) = 0$, then the ith component of all Lagrange multipliers associated with x^* are equal to zero. An interesting property of an algorithm that finds a maximally complementary Lagrange multiplier y^* is that if a strictly complementary Lagrange multiplier exists, then y^* satisfies strict complementarity.

There are benefits of algorithms with iterates that converge, in a subsequence, to a point satisfying strict complementarity. In particular, strict complementarity implies the critical cone is a subspace. One can therefore efficiently check if the second-order sufficient conditions hold by checking if the matrix $\nabla_x^2 \mathcal{L}(x^*, y^*)$ projected onto this subspace is positive definite. This allows us to confirm strict local optimality. Furthermore, when iterates converge to a point satisfying second-order sufficient conditions, strict complementarity and Mangasarian-Fromovitz, then the assumptions of Vicente and Wright [40] hold. Therefore the IPM they studied has superlinear convergence. Our work complements theirs because they could guarantee the premise of their theorems on nonconvex problems unless the optimal dual multipliers were unique, in which case standard results prove superlinear convergence [18].

In the next theorem, we show that if the constraint violation is reduced quickly enough relative to complementarity, then the dual sequence will be maximally complementary. To prove this result we assume either the problem is convex, or that the following "extended" Lagrangian function is locally star-convex

$$\tilde{\mathcal{L}}(x, y) := \mathcal{L}(x, y) + \theta \sum_{i: a_i(x^*)=0} (\nabla a_i(x^*)^{\mathrm{T}}(x - x^*))^2, \tag{20}$$

and that $\|x^k - x^*\| \leq C\sqrt{\mu^k}$ for some constant $C > 0$.

Similar results to Theorem 3 are well known when the functions are convex [22], and therefore our main contribution is when the functions f and a_i for $i = 1, \ldots, m$ are not convex.

Theorem 3 *Let $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ with $\mu^k > 0$ and $\mu^k \to 0$ be such that:*

1. *$x^k \to x^*$ with $a(x^*) \le 0$ and $s^k \to s^* := -a(x^*)$,*
2. *$y^k \ge 0$ and $s^k \ge 0$ with $y^k \to y^*$ (y^* is necessarily a Lagrange multiplier associated with x^*),*
3. *for some $0 < b \le c$, $\mu^k b \le y_i^k s_i^k$ for all $i : a_i(x^*) = 0$ and $(y^k)^{\mathrm{T}} s^k \le \mu^k c$,*
4. *for some $u \ge 0$, $|a_i(x^k) + s_i^k| \le \mu^k u$ for all $i : a_i(x^*) = 0$,*
5. *for some $d \ge 0$, $\|\nabla_x \mathcal{L}(x^k, y^k)\| \le d\mu^k(\|y^k\|_1 + 1)$,*
6. *the functions f and a_i for $i = 1, \ldots, m$ are convex functions, or*
   - *there is a neighborhood S of x^* and W of y^* such that for all $y \in W$, the function $\tilde{\mathcal{L}}(x, y)$ is star-convex around x^* on S, and*
   - *there is a constant $C \ge 0$ such that $\|x^k - x^*\| \le C\sqrt{\mu^k}$.*

Then, y^* is maximally complementary, i.e., $y_i^* > 0$ whenever there exists some Lagrange multiplier $\tilde{y}$ associated with x^* with $\tilde{y}_i > 0$.

Proof First, observe that for any Lagrange multiplier $\tilde{y}$ associated with x^* we have

$$\begin{aligned}
\sum_{i:a_i(x^*)=0} \frac{\tilde{y}_i}{y_i^k} &\le \sum_{i:a_i(x^*)=0} \frac{1}{\mu^k b} s_i^k \tilde{y}_i \\
&= \sum_{i:a_i(x^*)=0} \frac{1}{\mu^k b} \left(s_i^k (\tilde{y}_i - y_i^k) + s_i^k y_i^k \right) \\
&= \sum_{i:a_i(x^*)=0} \frac{1}{\mu^k b} \left((-a_i(x^k))(\tilde{y}_i - y_i^k) + (a_i(x^k) + s_i^k)(\tilde{y}_i - y_i^k) + s_i^k y_i^k \right) \\
&\le \sum_{i:a_i(x^*)=0} \frac{a_i(x^k)(y_i^k - \tilde{y}_i)}{\mu^k b} + \frac{u}{b}\|y^k - \tilde{y}\|_1 + \frac{c}{b}. \qquad (21)
\end{aligned}$$

If we can show that $a_i(x^k)(y_i^k - \tilde{y}_i)$ is bounded by a constant times μ^k, then the boundedness of the expression in (21) would imply that y_i^k can only converge to zero when $\tilde{y}_i = 0$ for all Lagrange multipliers, which gives the result. The remainder of the proof is dedicated to showing this and separately considers the two cases given in assumption 6.

First, we consider the case where f and a_i for $i = 1, \ldots, m$ are convex functions. Since $\nabla_x \mathcal{L}(x^*, \tilde{y}) = 0$, we have $\mathcal{L}(x^k, \tilde{y}) \ge \mathcal{L}(x^*, \tilde{y})$, and thus,

$$\begin{aligned}
(a(x^k) - a(x^*))^{\mathrm{T}}(y^k - \tilde{y}) &= \left(\mathcal{L}(x^*, \tilde{y}) - \mathcal{L}(x^*, y^k) \right) + \left(\mathcal{L}(x^k, y^k) - \mathcal{L}(x^k, \tilde{y}) \right) \\
&\le \mathcal{L}(x^k, y^k) - \mathcal{L}(x^*, y^k) \\
&\le \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^k - x^*), \qquad (22)
\end{aligned}$$

where the last inequality uses the convexity of $\mathcal{L}(x, y^k)$ with respect to x. Since

$$(a(x^k)-a(x^*))^{\mathrm{T}}(y^k-\tilde{y})=\sum_{i:a_i(x^*)=0} a_i(x^k)(y_i^k-\tilde{y}_i)+\sum_{i:a_i(x^*)<0}(a_i(x^k)-a_i(x^*))y_i^k$$

and $a_i(x^*)y_i^k \le 0$, we have

$$\sum_{i:a_i(x^*)=0} a_i(x^k)(y_i^k-\tilde{y}_i) \le \nabla_x \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^k-x^*) - \sum_{i:a_i(x^*)<0} a_i(x^k)y_i^k.$$

It remains to bound the right-hand side of the previous expression. For $i : a_i(x^*) < 0$ we have $-a_i(x^k)y_i^k \le \frac{-a_i(x^k)}{s_i^k}\mu^k c \le 2\mu^k c$. Also, $\|\nabla_x \mathcal{L}(x^k, y^k)\| \le d\mu^k(\|y^k\|_1 + 1) \le d\mu^k(\|y^*\|_1 + 2)$. This concludes the proof in the convex case.

On the other hand, let us assume the remaining conditions in assumption 6. We note first that we can take the Lagrange multiplier $\tilde{y}$ sufficiently close to y^* without loss of generality because for any Lagrange multiplier $\hat{y}$ associated with x^* we can take $\tilde{y}$ of the form $\tilde{y} := \eta\hat{y} + (1-\eta)y^*$, $\eta \in (0, 1)$, with the property that if $\hat{y}_i > 0$ then $\tilde{y}_i > 0$. Now, similarly to (22), from the star-convexity of $\tilde{\mathcal{L}}(x, \tilde{y})$ and $\tilde{\mathcal{L}}(x, y^k)$ we have

$$\begin{aligned}(a(x^k)-a(x^*))^{\mathrm{T}}(y^k-\tilde{y}) &= \left(\tilde{\mathcal{L}}(x^*, \tilde{y}) - \tilde{\mathcal{L}}(x^*, y^k)\right) + \left(\tilde{\mathcal{L}}(x^k, y^k) - \tilde{\mathcal{L}}(x^k, \tilde{y})\right) \\ &\le \tilde{\mathcal{L}}(x^k, y^k) - \tilde{\mathcal{L}}(x^*, y^k) \\ &\le \nabla_x \tilde{\mathcal{L}}(x^k, y^k)^{\mathrm{T}}(x^k - x^*). \end{aligned} \tag{23}$$

Hence,

$$\sum_{i:a_i(x^*)=0} a_i(x^k)(y_i^k-\tilde{y}_i) \le \nabla_x \tilde{\mathcal{L}}(x^k, y^k)^{\mathrm{T}}(x^k-x^*) - \sum_{i:a_i(x^*)<0} a_i(x^k)y_i^k.$$

It remains to bound the right-hand side of the previous expression by a constant times μ^k. Note that $-a_i(x^k)y_i^k \le 2\mu^k c$ for $i : a_i(x^*) < 0$ and

$$\|\nabla_x \tilde{\mathcal{L}}(x^k, y^k)\| \le d\mu^k(\|y^k\|_1 + 1) + 2\theta \sum_{i:a_i(x^*)=0} \|\nabla a_i(x^*)\|^2 \|x^k - x^*\|.$$

The result now follows from the bound $\|x^k - x^*\| \le C\sqrt{\mu^k}$. □

The next lemma shows that one can guarantee the upper bound on $\{\|x^k - x^*\|\}$ given in assumption 6 of Theorem 3 by assuming the standard second-order sufficient condition.

Lemma 3 (Hager and Mico-Umutesi [25]) *Let f and a be twice differentiable at a local minimizer x^* with a Lagrange multiplier $y^* \in \mathbb{R}^m$ satisfying the sufficient second-order optimality condition:*

$$d^{\mathrm{T}}\nabla^2_{xx}\mathcal{L}(x^*, y^*)d \geq \lambda\|d\|^2, \text{ for all } d \text{ such that}$$
$$\nabla f(x^*)^{\mathrm{T}}d \leq 0, \nabla a_i(x^*)^{\mathrm{T}}d \leq 0, i : a_i(x^*) = 0, \tag{24}$$

for some $\lambda > 0$. *Then, there is a neighborhood* $\mathcal{B}$ *of* $(x^*, y^*, -a(x^*))$ *such that if* $(x, v, s) \in \mathcal{B}$ *with* $v \geq 0$, $v_i = 0$ *for* $i : a_i(x^*) < 0$ *and* $s \geq 0$, *we have*

$$\|x - x^*\| \leq C\sqrt{\max\{\|\nabla_x\mathcal{L}(x, v)\|, \|[a(x) + s]_{i:a_i(x^*)=0}\|, v^{\mathrm{T}}s\}}$$

for some $C \geq 0$.

Proof The result follows from [25, Theorem 4.2] because the sufficient optimality condition is equivalently stated at constraints $a(x) \leq 0$ or at the slack variable formulation $a(x) + s = 0$, $s \geq 0$. Inactive constraints are removed from the problem and equivalence of norms is employed. □

Another useful result is the following.

Lemma 4 (Debreu [16]) *Let* $H \in \mathbb{R}^{n\times n}$ *be a symmetric matrix and* $A \in \mathbb{R}^{m\times n}$. *If* $d^T Hd > 0$ *for all* $d \in \mathbb{R}^n$ *such that* $Ad = 0$, *then there exists* $\theta \geq 0$ *such that* $H + \theta A^T A \succ 0$.

Now we can replace our nonconvex assumptions in Theorem 3 by the second-order sufficiency condition as follows.

Theorem 4 *Let* $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ *with* $\mu^k > 0$ *and* $\mu^k \to 0$ *be such that:*

1. $x^k \to x^*$ *with* $a(x^*) \leq 0$ *and* $s^k \to s^* := -a(x^*)$,
2. $y^k \geq 0$ *and* $s^k \geq 0$ *with* $y^k \to y^*$ *(*y^* *is necessarily a Lagrange multiplier associated with* x^**),*
3. *for some* $0 < b \leq c$, $\mu^k b \leq y_i^k s_i^k$ *for all* $i : a_i(x^*) = 0$ *and* $(y^k)^{\mathrm{T}}s^k \leq \mu^k c$,
4. *for some* $u \geq 0$, $|a_i(x^k) + s_i^k| \leq \mu^k u$ *for all* $i : a_i(x^*) = 0$,
5. *for some* $d \geq 0$, $\|\nabla_x\mathcal{L}(x^k, y^k)\| \leq d\mu^k(\|y^k\|_1 + 1)$,
6. f *and* a *are twice continuously differentiable and* (x^*, y^*) *satisfies the sufficient second-order optimality condition (24).*

Then, y^* *is maximally complementary, i.e.,* $y_i^* > 0$ *whenever there exists some Lagrange multiplier* $\tilde{y}$ *associated with* x^* *with* $\tilde{y}_i > 0$.

Proof Since the sufficient second-order optimality condition holds at (x^*, y^*) by Lemma 4, there exists $\theta \geq 0$ such that

$$\nabla^2_{x,x}\tilde{\mathcal{L}}(x^*, y^*) = \nabla^2_{x,x}\mathcal{L}(x^*, y^*) + \theta \sum_{i:a_i(x^*)=0} \nabla a_i(x^*)^{\mathrm{T}}\nabla a_i(x^*) \succ 0.$$

It follows that there exists some neighborhood $\mathcal{B}$ of (x^*, y^*) such that $\tilde{\mathcal{L}}(x, y)$ is convex on x for all $(x, y) \in \mathcal{B}$.

Also, for $v_i^k := y_i^k$ if $a_i(x^*) = 0$ and $v_i^k := 0$ otherwise, we have

$$\|\nabla_x \mathcal{L}(x^k, v^k)\| \leq \|\nabla_x \mathcal{L}(x^k, y^k)\| + \left\| \sum_{i : a_i(x^*) < 0} y_i^k \nabla a_i(x^k) \right\|,$$

which is bounded by a non-negative constant times μ^k. By Lemma 3 we have $\|x^k - x^*\| \leq C\sqrt{\mu^k}$ for some constant $C \geq 0$. Hence, the result follows by Theorem 3. □

From the proof of Theorem 4 we can see the term $\theta \sum_{i:a_i(x^*)=0} (\nabla a_i(x^*)^{\mathrm{T}}(x - x^*))^2$ in (20) is important because it guarantees $\tilde{\mathcal{L}}(x, y^*)$ is convex if the second-order sufficient conditions hold. Conversely, even if the second-order sufficient conditions hold, the Lagrangian $\mathcal{L}(x, y^*)$ may not be convex in a neighborhood of this point. For example, consider the problem $\min -x^2$ s.t. $x \geq 0, x \leq 0$ at the point $x = 0$; the second-order sufficient conditions are satisfied, but the Lagrangian is not convex in x. However, as we show in Theorem 4, the second-order sufficient conditions imply the nonconvex case of assumption 6 of Theorem 3.

Now that Theorem 3 and 4 are proved, we discuss possible extensions. When there are additional constraints $\tilde{a}_i(x) \leq 0, i = 1, \ldots, \tilde{m}$ that are known to have a strict interior (for instance, if they represent simple bounds on the variables), a common implementation choice is to maintain feasibility for these constraints at each iteration, instead of considering the slow reduction of feasibility suggested by (3c). Note that assumption 4 of Theorem 3 is weaker than (3c) and includes the possibility of keeping $\tilde{a}_i(x^k) + s_i^k = 0, i = 1, \ldots, \tilde{m}$, at each iteration. With respect to the results of Theorems 1 and 2, one may weaken their assumption 4 in order to consider the case $\tilde{a}_i(x^k) + s_i^k = 0, i = 1, \ldots, \tilde{m}$, by strengthening the corresponding assumption 5 by replacing the term $\|y^k\|_1$ on the bound of $\|\nabla_x \mathcal{L}(x^k, y^k)\|$ (which includes all dual multipliers) by the possibly smaller sum of the multipliers associated only with the original constraints $a_i(x) \leq 0$.

3 When things may fail

We now limit our results to the convex case, where we explore the possibility of (3c) not being satisfied (i.e., the constraint violation is not reduced at the same rate as complementarity).

In the following theorem, we show that controlling the constraint violation rate is essential for the boundedness of the dual sequence. In fact, we show that if the constraint violation reduces faster than the barrier parameter μ^k, the dual sequence is unbounded whenever the constraints are convex and the set of Lagrange multipliers is unbounded. We note that a similar result was already known when the functions f and a are linear [34, Theorem 4].

Theorem 5 *Assume that a is convex and the feasible region has empty interior. Let $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ with $\mu^k > 0$ for all k and $\mu^k \to 0$ be such that:*

1. *$x^k \to x^*$ with $a(x^*) \leq 0$ and $s^k \to -a(x^*)$,*
2. *$y^k \geq 0$ and $s^k \geq 0$,*
3. *for some $b > 0$, $\mu^k b \leq y_i^k s_i^k$ for all $i : a_i(x^*) = 0$,*
4. *$\frac{a_i(x^k)+s_i^k}{\mu^k} \to 0$ for all $i : a_i(x^*) = 0$.*

Then $\{y^k\}$ is unbounded.

Proof Note that there is no $d \in \mathbb{R}^n$, $d \neq 0$ with $\nabla a_i(x^*)^{\mathrm{T}} d < 0$ for all $i : a_i(x^*) = 0$, otherwise, $x^* + td$ would be interior for $t > 0$ sufficiently small. By Farkas's Lemma, there is some $\hat{y} \in \mathbb{R}^m$ with $\hat{y} \geq 0$, $\hat{y} \neq 0$, $a(x^*)^{\mathrm{T}}\hat{y} = 0$ and $\sum_{i=1}^m \hat{y}_i \nabla a_i(x^*) = 0$. For all i, we have $a_i(x^k) \geq a_i(x^*) + \nabla a_i(x^*)^{\mathrm{T}}(x^k - x^*)$ and hence $a(x^k)^{\mathrm{T}}\hat{y} \geq a(x^*)^{\mathrm{T}}\hat{y} + \sum_{i=1}^m \hat{y}_i \nabla a_i(x^*)^{\mathrm{T}}(x^k - x^*) = 0$. Thus,

$$\hat{y}^{\mathrm{T}}(a(x^k) + s^k) = \hat{y}^{\mathrm{T}} a(x^k) + \hat{y}^{\mathrm{T}} s^k \geq \hat{y}^{\mathrm{T}} s^k.$$

Take i such that $\hat{y}_i > 0$ and we have

$$0 < \mu^k b \hat{y}_i \leq y_i^k s_i^k \hat{y}_i \leq y_i^k \hat{y}^{\mathrm{T}} s^k \leq y_i^k \hat{y}^{\mathrm{T}}(a(x^k) + s^k).$$

Then, $\hat{y}^{\mathrm{T}}(a(x^k) + s^k) > 0$ and $y_i^k \geq b\hat{y}_i \frac{\mu^k}{\hat{y}^{\mathrm{T}}(a(x^k)+s^k)} \to +\infty$. □

The next theorem shows that the dual sequence can have a poor quality in terms of maximal complementarity if constraint violation is not reduced fast enough. We prove that in this instance the dual sequence limits to a point with minimal complementarity.

Theorem 6 *Let f and a_i for $i = 1, \ldots, m$ be convex functions and $\{(x^k, y^k, s^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$ with $\mu^k > 0$ and $\mu^k \to 0$ be such that:*

1. *$x^k \to x^*$ with $a(x^*) \leq 0$ and $s^k \to s^* := -a(x^*)$,*
2. *$y^k \geq 0$ and $s^k \geq 0$ with $y^k \to y^*$ (y^* is necessarily a Lagrange multiplier associated with x^*),*
3. *for some $c \geq 0$, $(y^k)^{\mathrm{T}} s^k \leq \mu^k c$,*
4. *$0 \leq a_i(x^k) + s_i^k$ for all $i : a_i(x^*) = 0$,*
5. *for some $d \geq 0$, $\|\nabla_x \mathcal{L}(x^k, y^k)\| \leq d\mu^k(\|y^k\|_1 + 1)$.*

Let $\tilde{y} \in \mathbb{R}^m$ be some Lagrange multiplier associated with x^* such that for all $i : a_i(x^*) = 0$,

- *$\frac{a_i(x^k)+s_i^k}{\mu^k} \to +\infty$ when $\tilde{y}_i = 0$, and*
- *$a_i(x^k) + s_i^k \leq u\mu^k$ or $y_i^k \geq \tilde{y}_i$ when $\tilde{y}_i > 0$,*

for some $u \geq 0$. Then, $y_i^* = 0$ whenever $\tilde{y}_i = 0$. In particular, if $\tilde{y}$ is minimally complementary, that is, it has a minimal number of non-zero elements, then y^* is also minimally complementary.

Proof Let $\tilde{y}$ be a Lagrange multiplier associated with x^*. We have

$$\sum_{i:a_i(x^*)=0} \frac{1}{\mu^k}(a_i(x^k) + s_i^k)(y_i^k - \tilde{y}_i) = \frac{1}{\mu^k} \sum_{i:a_i(x^*)=0} s_i^k y_i^k + a_i(x^k)(y_i^k - \tilde{y}_i) - s_i^k \tilde{y}_i.$$

Since $s_i^k \tilde{y}_i \geq 0$, $(s^k)^{\mathrm{T}} y^k \leq \mu^k c$ and, from the proof of Theorem 3, $a_i(x^k)(y_i^k - \tilde{y}_i) \leq C\mu^k$ for some $C \geq 0$, we have

$$\sum_{i:a_i(x^*)=0} \frac{a_i(x^k) + s_i^k}{\mu^k}(y_i^k - \tilde{y}_i) \leq c + C,$$

and the result follows. □

If assumption 4 in Theorem 6 is replaced by a similar one with a strict inequality, and assumption 5 is replaced by $\nabla \mathcal{L}(x^k, y^k)^{\mathrm{T}}(x^k - x^*) \leq d\mu^k$ for some $d \geq 0$, then we can drop the assumption that $\{y^k\}$ is convergent. It will then follow that $\{y^k\}$ is bounded, and any limit point y^* will have the property stated in the theorem.

In the next section we investigate the numerical behavior of the dual sequences generated by IPOPT on the NETLIB collection.

4 Numerical experiments

In this section, we contrast a well-behaved IPM, the one-phase IPM [26] that satisfies (3), with IPOPT, an IPM that *tends* to moves the primal feasibility faster than (3) would suggest. Empirically, we demonstrate on both linear and nonlinear programs that IPOPT has issues with the dual multiplier norms exploding, but the one-phase IPM does not. This demonstrates that our theory has practical implications for the design of IPMs.

Many IPM codes, such as IPOPT, keep $\frac{s_i y_i}{\mu}$ bounded below and require an inequality similar to

$$\|\nabla_x \mathcal{L}(x, y)\| + \|a(x) + s\| + \max_i s_i y_i \leq \mu(1 + \|y\|)$$

to hold before μ is decreased [36, Algorithm 19.1]. Hence assumptions 3–5 of Theorem 4 hold, and it follows that the IPM iterates are likely to converge to a maximal complementarity solution.

Our tests do not include IPMs that risk not tending to a minimal complementarity solution, i.e., reduce the constraint violation slower than perturbed complementarity. However, such IPMs certainly could be artificially created. This phenomenon might also occur naturally, for example, in dual regularized IPMs [1] or ℓ_2-penalty IPMs [15] *if* the algorithm is not well-designed.

The code for replicating our results can be found at https://github.com/ohinder/Lagrange-multipliers-behavior.jl. We test on the NETLIB test set of real linear programs in Sect. 4.1 and then on three toy nonconvex programs in Sect. 4.2.

4.1 Linear programs

The focus of this section is showing that on the NETLIB test set – of real linear programming problems – IPMs such as IPOPT, that aggressively reduce the primal

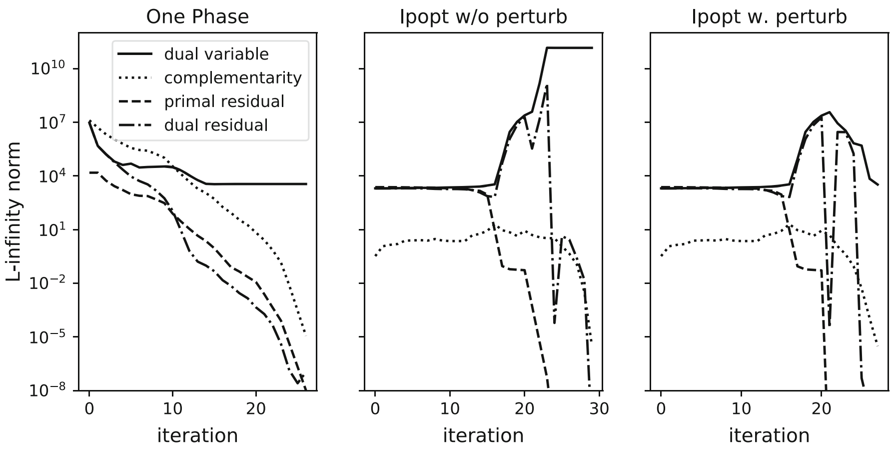

Fig. 2 Comparison of the iterates of different IPMs on the NETLIB problem ADLITTLE

feasibility, will have unnecessarily large dual iterates. As we discussed in the introduction, the convergence analysis of IPOPT and many other nonlinear optimization solvers [14,41] assumes that the set of dual multipliers at the convergence point is bounded to guarantee that the dual multipliers do not diverge. One natural question is whether these assumptions are valid on a test set like NETLIB. As documented in Table 4 in the "Appendix", we find that 64 of the 95 linear programs we tested lack a strict relative interior, and therefore Mangasarian-Fromovitz constraint qualification fails to hold. See the "Appendix" for more details on the experiments.

The next natural question is to check if the violation of these assumptions translates into undesirable behavior on these test problems. Consider Fig. 2 where we plot the performance of IPOPT on the problem ADLITTLE from the NETLIB collection. As our theory predicts when the primal feasibility is reduced faster than complementarity, the dual variables increase substantially. When IPOPT's default perturbation strategy is used, while the final dual variable value is only 3×10^3, the maximum dual variable value still spikes to 4×10^7 on iteration 22. This contrasts with the one-phase IPM [26] that smoothly reduces the constraint violation, dual feasibility, and complementarity; consequently, the maximum dual variable follows a smooth trajectory.

Next, we show that this phenomenon occurs across the whole NETLIB test set. We run these IPMs on the NETLIB problems with less than 10, 000 non-zero entries and record the maximum dual variable value (across all the IPMs iterates). All solvers successfully terminate, within the maximum number of iterations of 300, on 56 of the 68 problems. See the "Appendix" for further details. Figure 3 plots an empirical cumulative distribution over the maximum dual variable for each solver. In particular, for each solver, it plots the function $g : [0, 1] \to \mathbb{R}$ where $g(\theta)$ is the maximum dual variable value of the problem, for which, exactly a θ proportion of the problems have a smaller or equal maximum dual variable value. The plot illustrates that the maximum dual variable of IPOPT in the last few iterations (either with or without the default perturbation) is unnecessarily large for most problems that lack a strict relative interior.

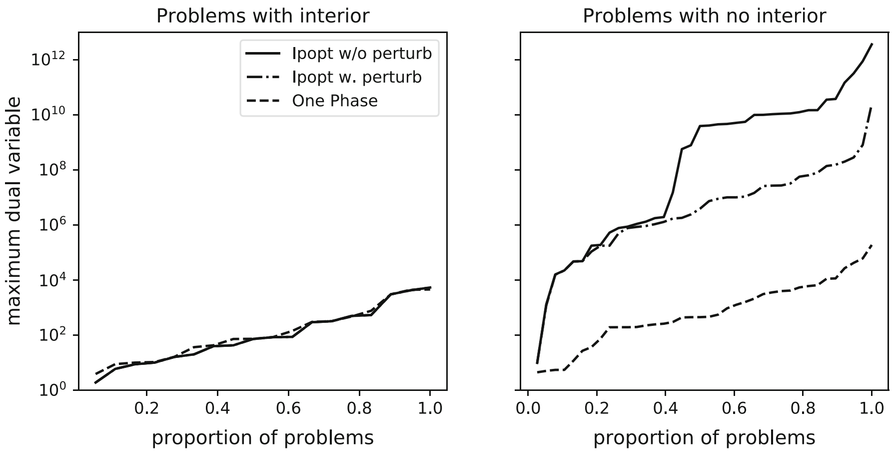

Fig. 3 Comparison of the maximum dual variable value over the last 20% of iterations for different IPMs on the NETLIB collection

4.2 Nonconvex programs

This section focuses on nonconvex programs. We test IPOPT and the one-phase IPM on three toy examples. The results for these examples are given in Table 1, and we believe validate the utility of our theory. Additional figures displaying the algorithm trajectories are given in "Appendix A.3". The first two problems were chosen to satisfy the assumptions of our theory. The final problem gives an example, derived from issues encountered in drinking water network optimization, where dual multipliers exploding is a practical issue.

Intersection of two circles. This problem is written as

$$\text{minimize} \ -(x_1-1)^2 + x_2^2 \tag{25a}$$

$$\text{subject to} \ x_1^2 + x_2^2 \le 1 \tag{25b}$$

$$(x_1-2)^2 + x_2^2 \le 1. \tag{25c}$$

The constraints require the solution to lie in the intersection of two circles, and the objective is a nonconvex quadratic. At the optimal solution (and only feasible solution) given by $x_1 = 1$, $x_2 = 0$, the Mangasarian-Fromovitz constraint qualification (MFCQ) does not hold. However, the point is a KKT point. Furthermore, the set of dual multipliers corresponding to this KKT point, contains both the point $(1, 1)$ which satisfies strict complementarity and the point $(0, 0)$ which does not satisfy strict complementarity.

As we show next, this problem satisfies the assumptions of Theorem 2 and Theorem 4 at the point $x_1 = 1$, $x_2 = 0$. A picture of this problem is given in Fig. 4. Next, we verify that the assumptions of Theorem 2 are met. Recall $x^* = (1, 0)$. Observe, the Lagrangian, its gradient and its Hessian are

Table 1 A selection of nonlinear programming problems for testing dual multiplier behavior

Solvers	Iterations	Max dual	Strict complementarity
Intersection of two circles			
One phase	6	1.2	1.2
Ipopt w/o perturb	26	6.6×10^9	4.0×10^8
Ipopt w. perturb	19	7.3×10^6	9.1
Linear program with complementarity constraints			
One phase	8	1.2×10^1	8.5×10^{-1}
Ipopt w/o perturb	28	5.8×10^9	2.0
Ipopt w. perturb	25	3.6×10^5	2.0
Drinking water network optimization			
One phase	9	1.7	2.9×10^{-1}
Ipopt w/o perturb	6*	1.2×10^4	1.1×10^{-7}
Ipopt w. perturb	44*	1.9×10^3	1.2×10^{-6}

Suppose the algorithm is generating a sequence of primal iterates x^k, slack iterates s^k, and dual iterates y^k. 'Max dual' refers to the value $\|y^k\|_\infty$ over the last 20% of iterations. 'Strict complementarity' refers to the minimum value of $\min_i y_i^k + s_i^k$ over the last 20% of iterations

A * indicates on these problems the 'dual multiplier calculator in IPOPT failed and therefore the algorithm terminated unsuccessfully. See Sect. A.3 for plots of IPM trajectories for these problems

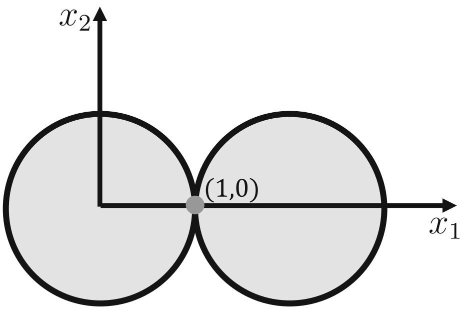

Fig. 4 Picture of the circle intersection problem given in (25)

$$
\begin{aligned}
\mathcal{L}(x, y) &= -(x_1 - 1)^2 + y_1(x_1^2 + x_2^2 - 1) + y_2((x_1 - 2)^2 + x_2^2 - 1) \\
\frac{\partial \mathcal{L}(x, y)}{\partial x_1} &= -2(x_1 - 1) + 2y_1 x_1 + 2y_2(x_1 - 2) \\
\frac{\partial \mathcal{L}(x, y)}{\partial x_2} &= 2(y_1 + y_2)x_2 \\
\nabla_{xx}^2 \mathcal{L}(x, y) &= 2\begin{pmatrix} y_1 + y_2 - 1 & 0 \\ 0 & y_1 + y_2 + 1 \end{pmatrix}.
\end{aligned}
$$

From this we observe assumption 6 holds with $y^* = (1, 1)$ and $\nabla_x \mathcal{L}(x^*, y^*) = 0$. Furthermore, from $\nabla_{xx}^2 \mathcal{L}(x, y)$ we deduce $\mathcal{L}(x, y)$ is convex in x if $y_1 + y_2 \geq 1$. This verifies assumption 7 of Theorem 2 with $\theta = 0$. The remaining assumptions of Theorem 2 are naturally satisfied by the one-phase IPM.

Next, we verify the assumptions of Theorem 4. Since y^k is bounded, there exists a convergent subsequence with limit y^*, where y^* satisfies $\nabla_x \mathcal{L}(x^*, y^*) = 0$. Furthermore, $d^T \nabla^2_{xx} \mathcal{L}(x^*, y^*) d \geq \lambda \|d\|_2^2$ on the null space of the Jacobian of the constraints ($d_1 = 0$).

Linear program with complementarity constraints. This problem is written as

$$\text{minimize } 3x_1 - 2x_2 \tag{26a}$$
$$\text{subject to } x_1 + 3x_2 \leq 2 \tag{26b}$$
$$x_1 x_2 \leq 0 \tag{26c}$$
$$x_1, x_2 \geq 0. \tag{26d}$$

At the unique local optima given by $x_1 = 2$ and $x_2 = 0$, MFCQ does not hold. However, the point is a KKT point. It is straightforward to see that this problem satisfies the assumptions of Corollary 1 and Theorem 4. The fact that Corollary 1 holds is immediate because linear functions are convex. Theorem 4 requires verifying the second-order sufficient conditions hold. They do because the null space of the Jacobian of the constraints evaluated at the solution $x_1 = 2$, $x_2 = 0$ only contains zero.

Therefore, our theory proves that for the one-phase IPM the dual multipliers remain bounded and strict complementarity holds for both the ‘intersection of two circles’ and ‘linear program with complementarity constraints’ problems. Table 1 demonstrates this behavior is seen in practice. Table 1 also shows IPOPT has issues with the dual multiplier values exploding on these problems. Both solvers maintain strict complementarity for these problems.

Drinking water network optimization. The final example is a toy drinking water network optimization problem (see [13] for the formulation of real drinking water network optimization problems as nonlinear programs). The aim is to choose the minimum inlet pressure to ensure that minimum node pressures and demand for water are met. We stumbled across this example when experimenting with our one-phase IPM [26] on real drinking water networks. A diagram representing the water network is given in Fig. 5.

$$\text{minimize } h_1 \tag{27a}$$
$$\text{subject to } x_{1,2} + x_{1,3} = 2 \tag{27b}$$
$$x_{1,2} + x_{2,3} = 1 \tag{27c}$$
$$x_{1,3} = 1 \tag{27d}$$
$$x_{1,2}^{1.8} = h_1 - h_2 \tag{27e}$$
$$x_{1,3}^{1.8} = h_1 - h_3 \tag{27f}$$

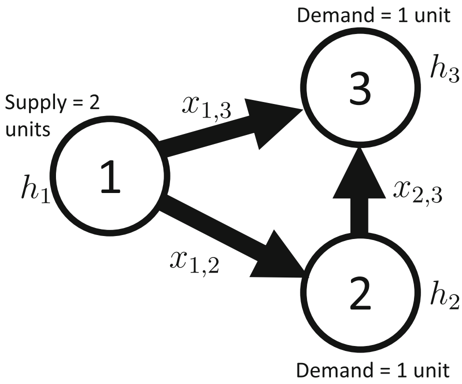

Fig. 5 Picture of the drinking water network optimization problem given in (27). Flows across the edges are given by the x variables and pressures at nodes by the h variables

$$x_{2,3}^{1.8} = h_2 - h_3 \tag{27g}$$

$$h_1, h_2, h_3, x_{1,2}, x_{1,3}, x_{2,3} \geq 0 \tag{27h}$$

Equation (27b) states that 2 units of water are available at node 1. Equation (27c) and (27d) states that 1 unit of water is demanded at both node 2 and node 3. Finally, (27e), (27f), (27g) represent the pressure loss in pipes due to friction. Our objective minimizing the inlet pressure is equivalent to minimizing the shafting speed of a variable speed pump at node 1.

The optimal solution (and unique local minimizer) occurs at $x_{1,2} = 1, x_{1,3} = 1, x_{2,3} = 0, h_1 = 0.29, h_2 = 0, h_3 = 0$. At this point, MFCQ fails but nonetheless the solution is a KKT point. We included this problem because it corresponds to a 'physically meaningful' problem where MFCQ fails to hold at the optimal solution.[1] Table 1 shows that the one-phase IPM keeps the dual variables bounded but fails to maintain strict complementarity for this problem. On the other hand, IPOPT seems to have issues with both the dual multipliers exploding and strict complementarity failing.

5 Final remarks

We demonstrated that carefully controlling both primal feasibility and the barrier parameter are important when designing IPMs to ensure the dual multipliers are well-behaved. In the linear programming community, there was awareness of this issue [34], and thus, many implemented IPMs move primal feasibility and complementarity at the same rate [2,33]. However, in the general nonlinear programming community, there is a lack of awareness of this issue. Consequently, there are few papers (e.g., Hinder and Ye [26]) that consider the relative rate of reduction of primal feasibility and complementarity.

[1] A popular misconception is that when the constraints of a optimization problem are defined by 'physics', MFCQ always holds. This is a nice counter-example.

Acknowledgements We would like to thank the anonymous referees for their helpful feedback, and Michael Saunders for carefully proof reading the paper.

A Experimental details

The code for the experiments can be found at https://github.com/ohinder/Lagrange-multipliers-behavior.

A.1 Solvers

One-phase solver. For the well-behaved interior point solver, given a problem of the form

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } c(x) = 0 \\ &\qquad x_L \leq x \leq x_U, \end{aligned}$$

we can re-write the constraints as

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } c(x) \leq 0 \\ &\qquad -c(x) \leq 0 \\ &\qquad x_L \leq x \leq x_U. \end{aligned}$$

This gives a problem of the form

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } a(x) + s = 0 \\ &\qquad s \geq 0, \end{aligned}$$

which we can pass to the one-phase solver.

The terms in Fig. 2 and Table 1 are given as follows:

- The infinity norm of the primal residual is given by $\|a(x) + s\|_\infty$.
- The infinity norm of the dual residual is measured by $\|\nabla \mathcal{L}(x, y)\|_\infty$.
- The infinity norm of complementarity is given by $\max_i s_i y_i$.
- We measure strict complementarity by $\min_i s_i + y_i$.

The optimality termination criterion of the one-phase IPM is

$$\max \left\{ \frac{100}{\max\{\|y\|_\infty, 100\}} \max\{\|\nabla_x \mathcal{L}(x, y)\|_\infty, \|Sy\|_\infty\}, \|a(x) + s\|_\infty \right\} \leq 10^{-6}.$$

For more etails on the one-phase IPM see the paper [26] and code (https://github.com/

Table 2 Solver options

Option	Value
IPOPT	
Max iter	300
tol	1.0×10^{-6}
acceptable_tol	1.0×10^{-6}
acceptable_iter	99999
acceptable_compl_inf_tol	1.0×10^{-6}
acceptable_constr_viol_tol	1.0×10^{-6}
acceptable_constr_viol_tol	1.0×10^{-6}
bound_relax_factor	0.0*
One phase	
max_it	300
tol	10^{-6}

The * indicates this option was only changed for 'IPOPT w/o perturb'. For 'IPOPT w. perturb' this was kept at its default value of 10^{-8}

Table 3 NETLIB problems where a solver failed

IPOPT w/o perturb	PEROLD, FFFFF800, SCAGR25, SHELL, SHARE1B, AGG3, VTP-BASE (7 total)
IPOPT w. perturb	PEROLD, FFFFF800, SCAGR25, SHELL, SHARE1B, VTP-BASE (6 total)
One phase	PEROLD, PILOT4, AGG2, PILOT-WE, GROW15, GROW22 (6 total)

ohinder/OnePhase.jl). The linear solver used was the default Julia Cholesky factorization (SuiteSparse).

IPOPT. We use IPOPT 3.12.4 with the linear solver MUMPS. Given any generic nonlinear problem, IPOPT rewrites it in the form (by adding slacks to inequalities, see [42])

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } c(x) = 0 \\ &\qquad\qquad x_L \leq x \leq x_U. \end{aligned}$$

For practical reasons related to the interface we use [17], we do this reformulation ourselves. We then measure

- Primal feasibility by $\|c(x)\|_\infty$.
- Dual feasibility by $\|\nabla f(x) + \nabla c(x)^T \lambda - z_L + z_U\|_\infty$, where z_L and z_U are the dual multipliers corresponding to the constraint $x \geq l$ and $x \leq u$ respectively (same notation as in [42]).
- Complementarity is given by $\max\{\max_i((z_L)_i(x_i - l_i)), \max_i((z_U)_i(x_i - u_i))\}$.

Table 4 Problems in NETLIB collection with a strict relative interior

Problem name	Strict interior	Problem name	Strict interior
25FV47	True	PILOT-JA	False
80BAU3B	False	PILOT-WE	False
ADLITTLE	False	PILOT	False
AFIRO	True	PILOT4	False
AGG	False	PILOTNOV	False
AGG2	False	QAP12	Tue
AGG3	False	QAP8	True
BANDM	False	RECIPELP	False
BEACONFD	False	SC105	False
BLEND	True	SC205	False
BNL1	False	SC50A	False
BNL2	False	SC50B	False
BOEING1	False	SCAGR25	True
BOEING2	False	SCAGR7	True
BORE3D	False	SCFXM1	False
BRANDY	False	SCFXM2	False
CAPRI	False	SCFXM3	False
CYCLE	False	SCORPION	False
CZPROB	False	SCRS8	False
D2Q06C	False	SCSD1	True
D6CUBE	True	SCSD6	True
DEGEN2	False	SCSD8	True
DEGEN3	False	SCTAP1	True
DFL001	False	SCTAP2	True
E226	False	SCTAP3	True
ETAMACRO	False	SEBA	False
FFFFF800	False	SHARE1B	True
FINNIS	False	SHARE2B	True
FIT1D	True	SHELL	False
FIT1P	True	SHIP04L	False
FIT2P	True	SHIP04S	False
FORPLAN	False	SHIP08L	False
GANGES	False	SHIP08S	False
GFRD-PNC	False	SHIP12L	False
GREENBEA	False	SHIP12S	False
GREENBEB	False	SIERRA	False
GROW15	True	STAIR	False
GROW22	True	STANDATA	False
GROW7	True	STANDGUB	False
ISRAEL	True	STANDMPS	False

Table 4 continued

Problem name	Strict interior	Problem name	Strict interior
KB2	True	STOCFOR1	True
LOTFI	True	STOCFOR2	True
MAROS	False	TRUSS	True
MODSZK1	False	VTP-BASE	False
NESM	False	WOOD1P	False
PEROLD	False	WOODW	False
QAP15	True		

– We measure strict complementarity by $\min\{\min_i((z_L)_i(x_i - l_i)), \min_i((z_U)_i(x_i - l_i))\}$.

The details of this computation can be found in the file 'src/shared.jl' in the function 'add_solver_results!'.

The options chosen for the solvers are given in Table 2. We turn off the acceptable termination criterion for IPOPT to try to make the termination criterion of the algorithms as similar as possible.

A.2 NETLIB LP test details

The linear programs in the NETLIB linear programming collection come in the form $\min c^T x$ s.t. $Ax = b, l \leq x \leq u$. Table 3 shows which solver failed on which problem.

Table 4 shows when there is a feasible solution according to Gurobi when the bound constraints are tightened by δ i.e. find a solution to the system $Ax = b$ and $u - \delta \geq x \geq l + \delta$. We tried $\delta = 10^{-4}, 10^{-6}, 10^{-8}$ and obtained the same results with Gurobi's feasibility tolerance set to 10^{-9}. We found 29 problems with a feasible solution and 64 without a feasible solution in the NETLIB collection. We used Gurobi version 7.02.

A.3 Additional figures for nonconvex problems

This section gives plots of solver trajectories for the nonconvex problems of Sect. 4.2.

See Figs. 6, 7 and 8.

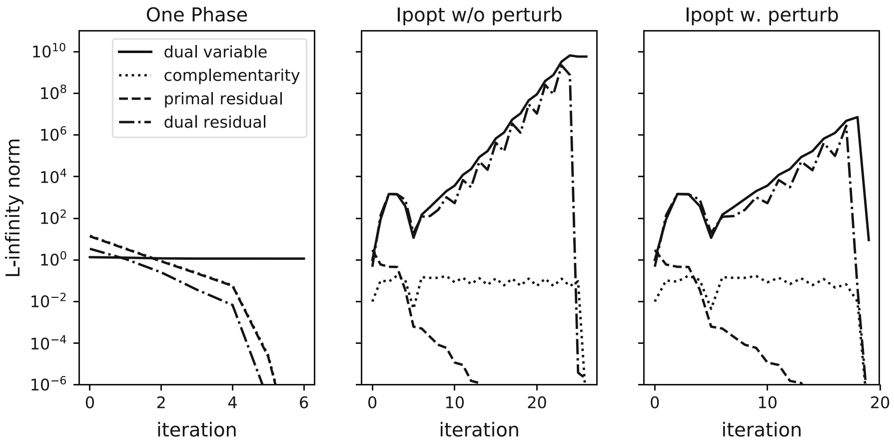

Fig. 6 Comparison on the problem of finding the intersection of two circles

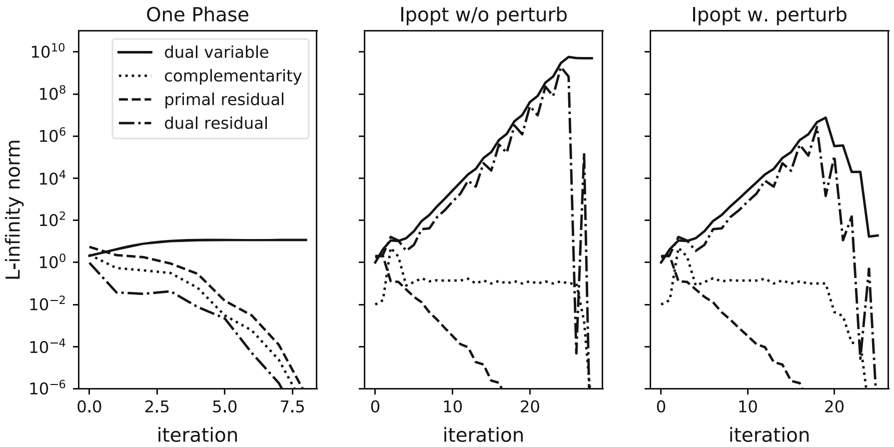

Fig. 7 Comparison on a linear program with complementarity constraints

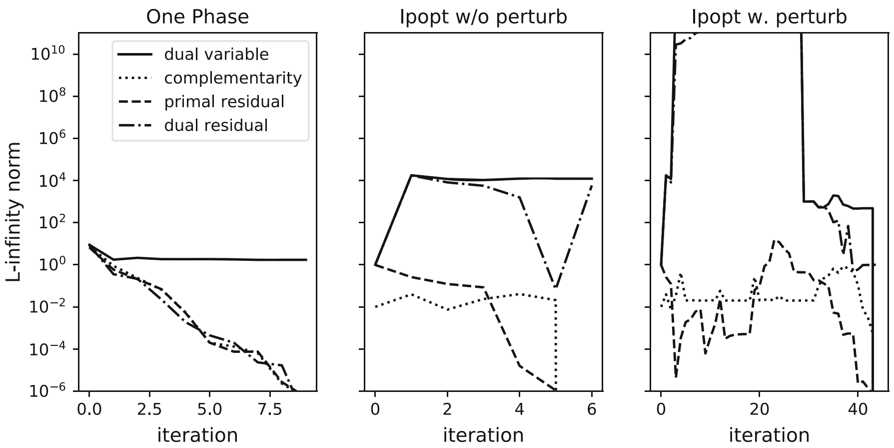

Fig. 8 Comparison on a toy drinking water network optimization problem

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. Andersen, E.D., Andersen, K.D.: The mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm. High Perform. Optim. **33**, 197–232 (2000)
3. Andersen, E.D., Ye, Y.: A computational study of the homogeneous algorithm for large-scale convex optimization. Comput. Optim. Appl. **10**(3), 243–269 (1998)
4. Andersen, E.D., Ye, Y.: On a homogeneous algorithm for the monotone complementarity problem. Math. Program. **84**(2), 375–399 (1999)
5. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**, 1109–1135 (2012a)
6. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**, 255–273 (2012b)
7. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**(1), 96–110 (2016)
8. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math. Oper. Res. (2018). https://doi.org/10.1287/moor.2017.0879
9. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010)
10. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)
11. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Practical methods of optimization: theory and algorithms. Wiley, Hoboken, New Jersey (2006)
12. Benson, H.Y., Shanno, D.F., Vanderbei, R.J.: Interior-point methods for nonconvex nonlinear programming: Complementarity constraints. Operations Res. Financ. Eng. pp. 1–20 (2002)
13. Burgschweiger, J., Gnädig, B., Steinbach, M.C.: Optimization models for operative planning in drinking water networks. Optim. Eng. **10**(1), 43–73 (2009)
14. Byrd, R.H., Gilbert, J.C., Nocedal, J.: A trust region method based on interior point techniques for nonlinear programming. Math. Program. **89**(1), 149–185 (2000)
15. Chen, L., Goldfarb, D.: Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties. Math. Program. **108**(1), 1–36 (2006)
16. Debreu, G.: Definite and semidefinite quadratic forms. Econometrica (pre-1986) **20**(2), 295 (1952)
17. Dunning, I., Huchette, J., Lubin, M.: Jump: a modeling language for mathematical optimization. SIAM Rev. **59**(2), 295–320 (2017)
18. El-Bakry, A.S., Tapia, R.A., Tsuchiya, T., Zhang, Y.: On the formulation and theory of the newton interior-point method for nonlinear programming. J. Optim. Theory Appl. **89**(3), 507–541 (1996)
19. Gauvin, J.: A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. Math. Program. **12**, 136–138 (1977)
20. Giorgi, G., Jiménez, B., Novo, V.: Approximate Karush-Kuhn-Tucker condition in multiobjective optimization. J. Optim. Theory Appl. **171**(1), 70–89 (2016)
21. Gondzio, J.: Matrix-free interior point method. Comput. Optim. Appl. **51**(2), 457–480 (2012)
22. Güler, O., Ye, Y.: Convergence behavior of interior-point algorithms. Math. Program. **60**(1), 215–228 (1993)
23. Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. Comput. Optim. Appl. **70**(2), 615–639 (2018)
24. Haeser, G., Schuverdt, M.L.: On approximate KKT condition and its extension to continuous variational inequalities. J. Optim. Theory Appl. **149**(3), 528–539 (2011)
25. Hager, W.W., Mico-Umutesi, D.: Error estimation in nonlinear optimization. J. Global Optim. **59**, 327–341 (2014)
26. Hinder, O., Ye, Y.: A one-phase interior point method for nonconvex optimization (2018) . arXiv preprint arXiv:1801.03072
27. Jeyakumar, V., Lee, G.M., Dinh, N.: New sequential lagrange multiplier conditions characterizing optimality without constraint qualification for convex programs. SIAM J. Optim. **14**(2), 534–547 (2003). https://doi.org/10.1137/S1052623402417699

28. Leyffer, S., López-Calva, G., Nocedal, J.: Interior methods for mathematical programs with complementarity constraints. SIAM J. Optim. **17**(1), 52–77 (2006)
29. Lustig, I.J.: Feasibility issues in a primal-dual interior-point method for linear programming. Math. Program. **49**(1–3), 145–162 (1990)
30. Lustig, I.J., Marsten, R.E., Shanno, D.F.: Interior point methods for linear programming: computational state of the art. ORSA J. Comput. **6**(1), 1–14 (1994)
31. McLinden, L.: An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem. Pac. J. Math. **88**(1), 101–161 (1980)
32. Megiddo, N.: Pathways to the optimal set in linear programming. In: Progress in Mathematical Programming. Springer, pp. 131–158 (1989)
33. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**(4), 575–601 (1992)
34. Mizuno, S., Todd, M.J., Ye, Y.: A surface of analytic centers and primal-dual infeasible-interior-point algorithms for linear programming. Math. Oper. Res. **20**(1), 135–162 (1995)
35. Nesterov, Y., Polyak, B.T.: Cubic regularization of newton method and its global performance. Math. Program. **108**(1), 177–205 (2006)
36. Nocedal, J., Wright, S.: Numer. Optim. Springer, Berlin (2006)
37. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
38. Rockafellar, R.T., Wets R, J.-B.: Variational Analysis. Springer, Berlin (1998)
39. Sonnevend, G.: An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In: System Modelling And Optimization. Springer, pp. 866–875 (1986)
40. Vicente, L.N., Wright, S.J.: Local convergence of a primal-dual method for degenerate nonlinear programming. Comput. Optim. Appl. **22**(3), 311–328 (2002)
41. Wächter, A., Biegler, L.T.: Line search filter methods for nonlinear programming: motivation and global convergence. SIAM J. Optim. **16**(1), 1–31 (2005)
42. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. **106**(1), 25–57 (2006)
43. Ye, Y., Anstreicher, K.: On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP. Math. Program. **62**(1–3), 537–551 (1993)
44. Ye, Y., Todd, M.J., Mizuno, S.: An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. Math. Oper. Res. **19**(1), 53–67 (1994)