On the behavior of Lagrange multipliers in convex and non-convex infeasible interior point methods*

Gabriel Haeser† Oliver Hinder‡ Yinyu Ye‡

July 23, 2017

Abstract

This paper analyzes sequences generated by infeasible interior point methods. In convex and non-convex settings, we prove that moving the primal feasibility at the same rate as complementarity will ensure that the Lagrange multiplier sequence will remain bounded, provided the limit point of the primal sequence has a Lagrange multiplier, without constraint qualification assumptions. We also show that maximal complementarity holds, which guarantees the algorithm finds a strictly complementary solution, if one exists. Alternatively, in the convex case, if the primal feasibility is reduced too fast and the set of Lagrange multipliers is unbounded, then the Lagrange multiplier sequence generated will be unbounded. Conversely, if the primal feasibility is reduced too slowly, the algorithm will find a minimally complementary solution. We also demonstrate that the dual variables of the interior point solver IPOPT become unnecessarily large on Netlib problems, and we attribute this to the solver reducing the constraint violation too quickly.

Keywords: Interior point methods, Lagrange multipliers, Complementarity, Nonlinear Optimization, Convex optimization

1 Introduction

We study sequences generated by interior point methods (IPMs) to generate Karush-Kuhn-Tucker (KKT) points of smooth nonlinear optimization problems of the form:

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

The central path generated by sequences of log barrier problems was introduced by [McLinden, 1980] for convex minimization subject to non-negativity constraints and generalized to linear inequalities by [Sonnevend, 1986]. In [Megiddo, 1989], the path of primal-dual IPMs for linear programming was analyzed and it was showed that the sequence converges to a point satisfying strict complementarity. In [Güler and Ye, 1993], this result was rigorously proved and generalized to a large class of path following IPMs for linear programming. Finding a strictly complementary solution is useful because it is needed to guarantee super-linear convergence of IPMs for quadratic programs [Ye and Anstreicher, 1993, Proposition 5.1]. Furthermore, for problems with non-convex constraints, finding a strictly complementary solution gives a way to efficiently verifying second-order conditions by computing the least eigenvalue of the Hessian of the Lagrangian function restricted to a subspace. However, in the nonlinear context, a strictly complementary

*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.

†Department of Applied Mathematics, University of São Paulo, São Paulo SP, Brazil. Visiting Scholar at Department of Management Science and Engineering, Stanford University, Stanford CA 94305, USA. E-mail: ghaeser@ime.usp.br.

‡Department of Management Science and Engineering, Stanford University, Stanford CA 94305, USA. E-mail: {ohinder,yinyu-ye}@stanford.edu

solution may not always exist, so one may aim to find a maximally complementary solution, that is, one with a Lagrange multiplier with as few as possible zero entries.

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 [Lustig, 1990, Mehrotra, 1992]. The paper [Mizuno et al., 1995] studies the sequences generated by these infeasible start algorithms for linear programming without assuming the existence of an interior point. Some of our theoretical contributions can be viewed as extensions of this work to convex and non-convex optimization. In particular, they show that by keeping moving the constraint violation at the same rate as the barrier parameter μ, one can guarantee that the dual multipliers are bounded. The boundedness of dual multipliers is an important practical property, since otherwise numerical issues can arise.

One alternative and elegant solution to these issues is the homogenous algorithm [Andersen and Ye, 1998, Andersen and Ye, 1999, Ye et al., 1994]. For convex problems, the homogenous algorithm is guaranteed to produce a bounded sequence that converges to a maximally complementary solution. This guarantees, for linear programming, that if the problem is feasible the algorithm will converge with bounded dual variables. However, due to the non-existence of a central path [Shanno and Vanderbei, 2000] it is unknown how to extend the homogenous algorithm to non-convex optimization.

While many IPMs for general non-convex optimization problems have been developed, there is little analysis of the sequences they generate. For example, it is unclear if interior point methods can generate maximally complementary solutions in the presence of non-convexity. Furthermore, results that show 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 [Gauvin, 1977]). This assumption may be too restrictive since 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 non-convex optimization algorithms such as IPMs, augmented lagrangian methods and sequential quadratic programming have been analyzed, for instance, in [Qi and Wei, 2000, Andreani et al., 2011, Andreani et al., 2010, Andreani et al., 2012b, Haeser, 2016], however, we are only aware of feasible interior point methods being considered. Moreover, these studies have been focused on determining primal convergence to a KKT point, despite unboundedness of the dual sequence, while we focus on guaranteeing boundedness and maximal complementarity of the dual sequence.

1.1 Summary of contributions

Many practical IPMs for linear programming reduce the primal, dual and complementarity at approximately the same rate [Ye et al., 1994, Mehrotra, 1992, Lustig et al., 1994]. In particular, consider an interior point method for minimizing $f(x) := \mathbf{c}^{\intercal}x$, subject to $a(x) + s = 0, s \geq 0$, where $a(x) := \mathbf{A}x - \mathbf{b}$, that at iteration k, the step (d_x^k, d_s^k) to update the primal variables (x^k, s^k) and the step d_y^k to update the dual variables y^k are taken satisfying:

$$\mathbf{A}^T d_y^k = -\eta^k(\mathbf{c} + \mathbf{A}^T y^k), \tag{2a}$$

$$\mathbf{A}d_x^k + d_s^k = -\eta^k(\mathbf{A}x^k + s^k - \mathbf{b}), \tag{2b}$$

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

where Y^k and S^k are the diagonal matrices defined by y^k and s^k, respectively, $\mu^{k+1} := (1 - \eta^k)\mu^k$ and $\eta^k \in (0, 1)$ is chosen, for example, using a predictor-corrector technique [Mehrotra, 1992], see also [Nocedal and Wright, 2006, Algorithm 14.3]. If these algorithms converge to an optimal solution i.e. $\mu^k \to 0$

then a sub-sequence of iterates satisfy $x^k \to x^*$, $s^k \to s^*$ and:

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

$$b \leq \frac{S^k y^k}{\mu^k} \leq c, \tag{3b}$$

$$\ell \leq \frac{a(x^k) + s^k}{\mu^k} \leq u, \tag{3c}$$

$$\frac{\|\nabla f(x^k) + \sum_{i=1}^m y_i^k \nabla a_i(x^k)\|}{\mu^k(\|y^k\| + 1)} \leq d, \tag{3d}$$

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

where $\mu^k > 0$ is the barrier parameter, $0 < b \leq c, 0 < \ell \leq u$ and $d \geq 0$ are real constants independent of k. Equation (3b) ensures complementarity approximately holds, Equation (3c) guarantees that primal feasibility moves at the same rate as complementarity and Equation (3d) ensures that scaled dual feasibility is reduced fast enough.

For general non-linear optimization work by [Hinder, 2017] (in preparation) gives an IPM where conditions (3) holds for a subsequence of the iterates, if the primal variables are bounded and the algorithm does not return a certificate of local primal infeasibility. Now, assuming conditions (3); and that the problem is convex or under certain sufficient conditions for local optimality we show:

1. If there exists a Lagrange multiplier at the point x^*, then the sequence of Lagrange multipliers approximations $\{\|y^k\|\}$ is bounded (see Theorem 1 and Theorem 2 for convex and non-convex case respectively).

2. If $y^k \to y^*$, 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 a bad idea, because if the functions f and a are convex then:

1. 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 $\|y^k\| \to \infty$ (see Theorem 5).

2. If we reduce the primal feasibility slower than the barrier parameter μ^k and $y^k \to y^*$, then y^* is a minimally complementary Lagrange multiplier associated with x^* (see Theorem 6).

The central claim of this paper is that many implemented interior point methods, especially for non-linear optimization, such as IPOPT [Wächter and Biegler, 2006], suffer from the problems described above because they fail to control the rate which they reduce primal feasibility.

In our linear programming example, these methods solve systems of the form [Wächter and Biegler, 2006, Equation (9)]:

$$\mathbf{A}^T d_y^k = -\left(\mathbf{c} + \mathbf{A}^T y^k\right), \tag{4a}$$

$$\mathbf{A} d_x^k + d_s^k = -\left(\mathbf{A} x^k + s^k - \mathbf{b}\right), \tag{4b}$$

$$S^k d_y^k + Y^k d_s^k + S^k y^k = \mu^k. \tag{4c}$$

Equation (4b) aims to reduce the constraint violation to zero at each iteration. We can contrast with Equation (2b) that aims to reduce the constraint violation by η^k, the same amount that complementarity is reduced. Furthermore, in IPOPT and KNITRO [Byrd et al., 2006], μ^k is reduced only if the total KKT residual is less than μ^k [Nocedal and Wright, 2006, Algorithm 19.1]. As we demonstrate in Section 4, a consequence of these implementation choices is that primal feasibility is usually reduced faster than complementarity. Therefore, as our theory suggests, these IPMs tend to generate Lagrange multipliers sequences that diverge towards infinity.

The solver IPOPT attempts to circumnavigate these issues by perturbing the original constraint $a(x) \leq 0$ in order to artificially create an interior as follows:

$$a(x) \leq \delta, \tag{5}$$

for some $\delta > 0$ (see Section 3.5. of [Wächter and Biegler, 2006]). While this technically solves the issue as the theoretical assumptions of [Wächter and Biegler, 2005] are now met, it is not an elegant solution and causes undesirable behavior. Firstly, it reduces the accuracy that the constraints are satisfied. Furthermore, as we show in Section 4, this may cause the dual variable to *spike* before converging.

The paper proceeds as follows. Section 1.2 gives a simple example illustrating the phenomena studied in this paper. Section 2 shows that reducing the primal feasibility at the same rate as complementarity will ensure the dual multiplier sequence remains bounded and satisfy maximal complementarity. Section 3 explains that reducing the constraint violation too quickly will cause the dual multipliers sequence to diverge towards infinity, while reducing it too fast will cause the dual multipliers sequence to be minimally complementary. Section 4 shows empirically how strategies, such as the one employed by IPOPT, that reduce the constraint violation too fast, can have issues with the dual multipliers *spiking* or the norm diverging towards infinity. Section 5 presents our final remarks.

1.2 A simple example

Consider the following simple linear programming problem:

$$\begin{aligned} &\text{Minimize } 0, && (6a)\\ &\text{Subject to } x + s_1 = 1, && (6b)\\ &x - s_2 = 1, && (6c)\\ &s_1 \geq 0, s_2 \geq 0. && (6d)\end{aligned}$$

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

$$\begin{aligned} &\text{Minimize } -\mu \log(s_1 + \delta) - \mu \log(s_2 + \delta), && (7a)\\ &\text{Subject to } x + s_1 = 1, && (7b)\\ &x - s_2 = 1, && (7c)\\ &s_1, s_2 \geq -\delta. && (7d)\end{aligned}$$

The KKT system associated with this is:

$$\begin{aligned} &y_1 - y_2 = 0, && (8a)\\ &x + s_1 = 1, && (8b)\\ &x - s_2 = 1, && (8c)\\ &(s_1 + \delta)y_1 = \mu, (s_2 + \delta)y_2 = \mu, && (8d)\\ &y_1, y_2 \geq 0,\ \ s_1, s_2 \geq -\delta. && (8e)\end{aligned}$$

Observe that the orignal problem (which corresponds 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 (8) is:

$$\begin{aligned} x = 1, s_1 = 0, s_2 = 0\\ y_1 = \frac{\mu}{\delta}, y_2 = \frac{\mu}{\delta}. \end{aligned}$$

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 taken fixed at a small tolerance as in the IPOPT strategy (5), 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 (6) with perturbations $\delta > 0$ using IPOPT and we compare it with a *well-behaved IPM* [Hinder, 2017] that moves complementarity at the same rate as primal feasibility, that is, satisfies the bounds given in Equations (3a)-(3e). For this experiment we also turn off IPOPT's native perturbation strategy (5). In Figure 1 we plot the maximum dual variables at each iteration, given by the two methods for different perturbation sizes. One can see that while perturbing the linear program prevents the dual variables iterates of IPOPT increasing indefinitely – the dual variables still *spike* and the iteration count increases. For example, with $\delta = 10^{-9}$ the maximum dual iterates of IPOPT peaks at 2×10^7 on iteration 20 before sharply dropping to 1×10^2 in the final iteration 21. This is contrasted with a well-behaved IPM [Hinder, 2017] (that reduces the complementary and primal feasibility at the same rate) where the dual variable sequence is indifferent to the perturbation size.

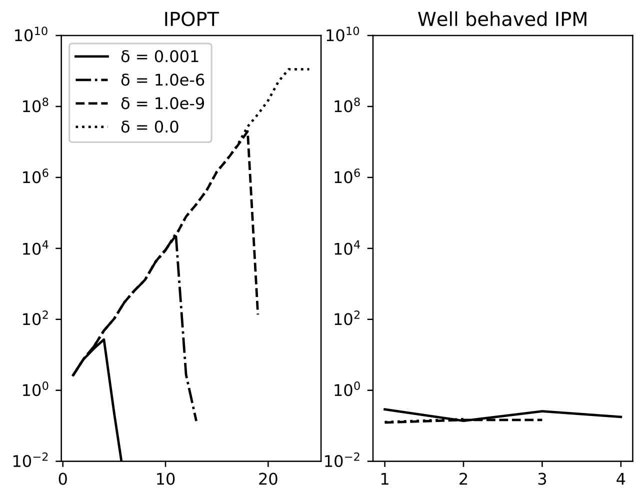

Figure 1: Comparison of the dual variable value (vertical axis) along iterations (horizontal axis) of constraint (7d) using IPOPT and a well-behaved IPM [Hinder, 2017] as the perturbation δ is changed.

A more thorough numerical experiment on the Netlib collection will be given in Section 4, but first we establish our general theory.

2 Boundedness and maximal complementarity

Let us consider the general optimization problem (1) where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the inequality constraints $a : \mathbb{R}^n \to \mathbb{R}^m$ are at least continuously differentiable functions. For $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ with $y \geq 0$, we define the Lagrangian function as:

$$\mathcal{L}(x, y) := f(x) + y^T a(x), \tag{9}$$

while $\nabla_x \mathcal{L}(x, y)$ will denote its gradient vector, where derivatives are taken with respect to x. For simplicitly, we talk about a solution x^* of (1) meaning that $(x, s) := (x^*, -a(x^*))$ is a solution of (1). The Euclidean norm will be denoted by $\|\cdot\|$, while the ℓ_1-norm will be denoted by $\|\cdot\|_1$. 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 $k = 1, 2, \ldots$.

In this section, we will show that when feasibility is reduced at the same rate of complementarity, the Lagrange multipliers sequence will be bounded and satisfies maximal complementarity. But first we establish some basic results on the optimality of solutions for convex problems.

The following lemma gives a sufficient sequential condition for global optimality in the convex case. It is a stronger version in our setting of [Jeyakumar et al., 2003, Corollary 3.1], [Andreani et al., 2010, Theorem 4.2], [Haeser and Schuverdt, 2011, Theorem 2.2] and [Giorgi et al., 2016, Theorem 3.2]. Our condition is in fact equivalent to the one from [Giorgi et al., 2016], but with a redundant assumption omitted.

Lemma 1. *If f and a 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. $\liminf a(x^k)^\intercal y^k \geq 0$,
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

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

Hence,

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

Thus, for $x = x^*$, we have $\limsup a(x^k)^\intercal y^k \leq 0$. The assumption gives $a(x^k)^\intercal y^k \to 0$. Taking 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 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. $(y^k)^\intercal s^k \to 0$,
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 non-convex case, the existence of sequences satisfying the conditions of Lemmas 1 and 2 are also 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 [Andreani et al., 2010, Theorem 3.3], by defining $s_i^k := \max\{0, -a_i(x^k)\}$ for all i and all k. See also [Haeser, 2016].

2.1 Boundedness of the dual sequence

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

In Theorem 1 we consider problem involving convex functions where the algorithm is converging to a KKT point. We show if the primal feasibility, (scaled) dual feasibility and complementarity converge at the same rate then the dual sequence $\{y^k\}$ is bounded. We refer the reader to [Mizuno et al., 1995, 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 non-convex functions.

Note that we try to present as few assumptions as possible, in the sense that often assumptions are placed only on constraints that are active at the limit. However, in practice, since the active constraints are unknown, we advocate that an interior point method with good properties should be such that the conditions given in (3) are met. All theorems in this section apply under this set of assumptions.

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

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

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

where y^* is any Lagrange multiplier associated with x^*. In particular, $\{y^k\}$ is bounded.

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

$$f(x^*) \geq \mathcal{L}(x^*, y^k) \geq \mathcal{L}(x^k, y^k) + \nabla_x \mathcal{L}(x^k, y^k)^\intercal (x^* - x^k),$$

which gives

$$f(x^*) - f(x^k) \geq a(x^k)^\intercal y^k + \nabla_x \mathcal{L}(x^k, y^k)^\intercal (x^* - x^k).$$

Also,

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

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

$$\begin{aligned} a(x^k)^\intercal y^k &\geq \ell\mu^k \sum_{i:a_i(x^*)=0} y_i^k - \sum_{i:a_i(x^*)<0} 2c\mu^k - c\mu^k = \\ &\ell\mu^k\|y^k\|_1 - \sum_{i:a_i(x^*)<0} (2c + \ell y_i^k)\mu^k - c\mu^k. \end{aligned} \tag{11}$$

Also, $\nabla_x \mathcal{L}(x^k, y^k)^\intercal (x^* - x^k) \geq -d\mu^k(\|y^k\|_1 + 1)\|x^* - x^k\|$, hence,

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

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

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

Since $\ell > 0$ and $y_i^k \to 0$ for $i : a_i(x^*) < 0$, we have $\mu^k\|y^k\|_1 \to 0$. Now 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 the previous calculations yields

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

But $a(x^k)^\intercal y^* = (a(x^k)+s^k)^\intercal y^* - (s^k)^\intercal y^* \leq \mu^k u\|y^*\|_1$ which implies

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

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

We now present a non-convex version of Theorem 1. For this, we will 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 [Nesterov and Polyak, 2006].

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

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

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 $a(x^*) \leq 0$ and $s^k \to -a(x^*)$,*
2. *$y^k \geq 0$ and $s^k \geq 0$,*
3. *for some $c \geq 0$, $(y^k)^\intercal s^k \leq \mu^k c$,*
4. *for some $0 < \ell \leq u$, $\mu^k\ell \leq a_i(x^k) + s_i^k \leq \mu^k u$ 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. *x^* is a KKT point with Lagrange multiplier y^*,*
7. *There exist $\theta \geq 0$ and a neighborhood $\mathcal{B}$ of x^* such that $\hat{\mathcal{L}}_k(x) := \mathcal{L}(x, y^k) + \theta a(x)^\intercal Y^k a(x)$ for all k and $\hat{\mathcal{L}}_*(x) := \mathcal{L}(x, y^*) + \theta a(x)^\intercal Y_* a(x)$ 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:

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

Therefore,

$$f(x^*) - f(x^k) \geq -\theta a(x^*)^{\intercal} Y^k a(x^*) + \theta a(x^k)^{\intercal} Y^k a(x^k) + a(x^k)^{\intercal} y^k$$
$$+\nabla_x \mathcal{L}(x^k, y^k)^{\intercal}(x^* - x^k) + \sum_{i=1}^{m} 2\theta y_i^k a_i(x^k) \nabla a_i(x^k)^{\intercal}(x^* - x^k).$$

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

By the calculation in the proof of Theorem 1, we have:

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

Clearly, $\nabla_x \mathcal{L}(x^k, y^k)^{\intercal}(x^* - x^k) \geq -d\mu^k \|y^k\|_1 \|x^* - x^k\| - d\mu^k$ when $\|x^* - x^k\| \leq 1$. 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,

$$\sum_{i=1}^{m} 2\theta y_i^k a_i(x^k) \nabla a_i(x^k)^{\intercal}(x^* - x^k) \geq$$
$$-2\theta \sum_{i=1}^{m} |a_i(x^k) y_i^k| \|\nabla a_i(x^k)\| \|x^k - x^*\| \geq$$
$$\sum_{i:a_i(x^*)<0} -4\theta c\mu^k \|\nabla a_i(x^k)\| +$$
$$\sum_{i:a_i(x^*)=0} 2\theta \min\{-u\mu^k y_i^k, \ell\mu^k y_i^k - c\mu^k\} \|\nabla a_i(x^k)\| \|x^k - x^*\|. \tag{12}$$

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, for some constant $C \geq 0$, we have the following:

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

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

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

$$-C\mu^k + \frac{\ell}{2}\mu^k \|y^k\|_1 \leq a(x^k)^{\intercal} y^* + \theta a(x^k)^{\intercal} Y_* a(x^k) =$$
$$\sum_{i:a_i(x^*)=0} (a_i(x^k) + \theta a_i(x^k)^2) y_i^*.$$

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

Remark 1. *Given the bound $|a_i(x^k) y_i^k| \leq \max\{u\mu^k y_i^k, c\mu^k - \ell\mu^k y_i^k\}$ obtained in* (12) *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 [Andreani et al., 2012a, Andreani et al., 2012b, Andreani et al., 2016, Andreani et al., 2017]. 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 star-convexity assumption, namely, considering the function* $\hat{\mathcal{L}}_k(x) := \mathcal{L}(x, y^k) + \sum_{i=1}^m \theta_i^k a_i(x)^2$*, where* $\theta_i^k \le C\mu^k$ *for some* $C \ge 0$ *when* $a_i(x^*) < 0$ *and, for* $i : a_i(x^*) = 0$*,*

- $\theta_i^k := \theta(y_i^k + \|y^k\|_1 \mathcal{I}[a_i \equiv -a_j \text{ for some } j \neq i])$*, where* $\mathcal{I}[\cdot]$ *is the indicator function, or*

- $\theta_i^k := \theta\|y^k\|_1$*, under a strict complementarity assumption, namely, that* $\{y_i^k\}$ *is bounded away from zero,*

with some $\theta > 0$*.*

The main modification in the proof of Theorem 2 would be on the bound of $|\theta_i^k a_i(x^k)|$ *in (12). In the first case, for 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* $a_i(x^k)$*. In the second case, one gets* $-|\theta_i^k a_i(x^k)| \ge \theta_i^k \min\{-u\mu^k, \ell\mu^k - s_i^k\}$ *and the strict complementarity assumption would give* $-s_i^k \ge -\mu^k \frac{u}{y_i^k}$*, with* $\frac{u}{y_i^k}$ *bounded. The result now would 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 [Rockafellar and Wets R., 1998], 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, since 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 i-th 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 it is not the case that some active constraint $i : a_i(x^*) = 0$ has all its Lagrange multipliers equal to zero, that is, a strictly complementary Lagrange multiplier exists, then y^* satisfies strict complementarity.

There are benefits of algorithms with iterates that limit 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 Magngasarian-Fromovitz, then the assumptions of [Vicente and Wright, 2002] hold and therefore the IPM they studied has super-linear convergence. This work supplements [Vicente and Wright, 2002] since they did not provide non-trivial conditions, aside from the convex case, when their algorithm would converge to a point satisfying strict complementarity.

In the next theorem we show that if the constraint violation is reduced quickly enough relative to complementarity, then the dual sequence will satisfy maximal complementarity. For this, we assume the problem is convex, or that the following "extended" Lagrangian function is locally star-convex:

$$\tilde{\mathcal{L}}(x, y) := \mathcal{L}(x, y) + \theta \sum_{i : a_i(x^*) = 0} (\nabla a_i(x^*)^{\intercal}(x - x^*))^2. \tag{13}$$

In the last case, we also assume that $\{x^k\}$ converges to x^* at least as fast as $\{\sqrt{\mu^k}\}$ converges to zero. Note that when the functions are non-convex but the sufficient second-order conditions hold at the point x^* (see equation (17)), the Lagrangian $\mathcal{L}(x, y^*)$ may not be convex in a neighborhood of this point[1]. However, we as we show in Theorem 4 the second order sufficient conditions imply Assumption 6 of Theorem 3. Note that similar results are well-known when the functions are convex [Güler and Ye, 1993], and therefore our main contribution is when the functions f and a are not convex.

[1]i.e. consider the problem $\min -x^2$ s.t. $x \ge 0, x \le 0$ at the point $x = 0$, the second-order sufficient conditions are satisfied but the Lagrangian $\nabla_x \mathcal{L}(x, y) = -x + y_1 - y_2$ is not convex in x.

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 $a(x^*) \leq 0$ 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 $(y^k)^\intercal s^k \leq \mu^k c$,*
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. *The functions f and a 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 $C \geq 0$ such that $\|x^k - x^*\| \leq C\sqrt{\mu^k}$.*

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:

$$\begin{aligned}
\sum_{i:a_i(x^*)=0} \frac{\tilde{y}_i}{y_i^k} &\leq \sum_{i:a_i(x^*)=0} \frac{1}{\mu^k b} s_i^k \tilde{y}_i \\
&= \sum_{i:a_i(x^*)=0} \frac{1}{\mu^k b} \left(s_i^k(\tilde{y}_i - y_i^k) + s_i^k y_i^k \right) \\
&= \sum_{i:a_i(x^*)=0} \frac{1}{\mu^k b} \left((-a_i(x^k))(\tilde{y}_i - y_i^k) + (a_i(x^k) + s_i^k)(\tilde{y}_i - y_i^k) + s_i^k y_i^k \right) \\
&\leq \sum_{i:a_i(x^*)=0} \frac{a_i(x^k)(y_i^k - \tilde{y}_i)}{\mu^k b} + \frac{u}{b}\|y^k - \tilde{y}\|_1 + \frac{c}{b}. \qquad (14)
\end{aligned}$$

Observe that 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 (14) 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 that f and a are convex functions, then, since $\nabla_x \mathcal{L}(x^*, \tilde{y}) = 0$, we have $\mathcal{L}(x^k, \tilde{y}) \geq \mathcal{L}(x^*, \tilde{y})$, thus,

$$\begin{aligned}
(a(x^k) - a(x^*))^\intercal (y^k - \tilde{y}) &= \left(\mathcal{L}(x^*, \tilde{y}) - \mathcal{L}(x^*, y^k) \right) + \left(\mathcal{L}(x^k, y^k) - \mathcal{L}(x^k, \tilde{y}) \right) \\
&\leq \mathcal{L}(x^k, y^k) - \mathcal{L}(x^*, y^k) \\
&\leq \nabla_x \mathcal{L}(x^k, y^k)^\intercal (x^k - x^*). \qquad (15)
\end{aligned}$$

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

Since

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

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

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

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

Now, on the other hand, let us assume the remaing conditions in Assumption 6. We note first that we can take the Lagrange multiplier $\tilde{y}$ sufficiently close to y^* without loss of generality, since 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 (15), from the star-convexity of $\tilde{\mathcal{L}}(x, \tilde{y})$ and $\tilde{\mathcal{L}}(x, y^k)$, we have

$$
\begin{aligned}
(a(x^k) - a(x^*))^{\mathsf{T}}(y^k - \tilde{y}) &= \left(\tilde{\mathcal{L}}(x^*, \tilde{y}) - \tilde{\mathcal{L}}(x^*, y^k)\right) + \left(\tilde{\mathcal{L}}(x^k, y^k) - \tilde{\mathcal{L}}(x^k, \tilde{y})\right) \\
&\leq \tilde{\mathcal{L}}(x^k, y^k) - \tilde{\mathcal{L}}(x^*, y^k) \\
&\leq \nabla_x \tilde{\mathcal{L}}(x^k, y^k)^{\mathsf{T}}(x^k - x^*). \qquad (16)
\end{aligned}
$$

Hence,

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

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

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

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

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, 2014])**.** *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:*

$$
\begin{aligned}
&d^{\mathsf{T}}\nabla^2_{xx}\mathcal{L}(x^*, y^*)d \geq \lambda\|d\|^2, \text{ for all } d \text{ such that} \\
&\quad \nabla f(x^*)^{\mathsf{T}}d \leq 0, \nabla a_i(x^*)^{\mathsf{T}}d \leq 0, i : a_i(x^*) = 0,
\end{aligned} \qquad (17)
$$

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

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

for some $C \geq 0$.

Proof. The result follows from [Hager and Mico-Umutesi, 2014, Theorem 4.2], by noting that 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. □

Now we can replace our non-convex 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 $(y^k)^{\mathsf{T}}s^k \leq \mu^k c$,*
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* (17).

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^*) with parameter $\lambda > 0$, then there exists some neighborhood $\mathcal{B}$ of (x^*, y^*) such that for any $(x, y) \in \mathcal{B}$ we have:

$$d^T \nabla^2_{x,x} \mathcal{L}(x, y) d \geq \frac{\lambda}{2}\|d\|^2, \text{ for all } d \text{ with}$$
$$\nabla f(x^*)^{\mathtt{T}} d \leq 0,$$
$$\nabla a_i(x^*)^{\mathtt{T}} d \leq 0, \text{ for all } i : a_i(x^*) = 0.$$

Therefore, in particular, there is a sufficiently large $\theta \geq 0$ such that the function

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

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

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

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. □

Now that Theorem 3 and 4 are proved we discuss possible extensions. We note that when there are additional constraint $\tilde{a}(x)_i \leq 0, i = 1, \ldots, \tilde{m}$, which are known to have a sufficient interior (for instance, if they represent simple bounds on the variables), a common implementation choice is to maintain feasibility of 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 strenghtening 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, that is, 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 [Mizuno et al., 1995, 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* $b > 0$, $\mu^k b \le y_i^k s_i^k$ *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 $d \in \mathbb{R}^n$, $d \ne 0$ with $\nabla a_i(x^*)^{\mathsf{T}} d < 0$ for all $i : a_i(x^*) = 0$, otherwise, $x^* + td$ would be interior for $t > 0$ sufficiently small. By Farkas' Lemma, there is some $\hat{y} \in \mathbb{R}^m$ with $\hat{y} \ge 0, \hat{y} \ne 0$, $a(x^*)^{\mathsf{T}}\hat{y} = 0$ and $\sum_{i=1}^m \hat{y}_i \nabla a_i(x^*) = 0$. For all i, we have $a_i(x^k) \ge a_i(x^*) + \nabla a_i(x^*)^{\mathsf{T}}(x^k - x^*)$, hence, $a(x^k)^{\mathsf{T}}\hat{y} \ge a(x^*)^{\mathsf{T}}\hat{y} + \sum_{i=1}^m \hat{y}_i \nabla a_i(x^*)^{\mathsf{T}}(x^k - x^*) = 0$. Thus,

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

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

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

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

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 is minimally complementary.

Theorem 6. *Let f and a 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* $a(x^*) \le 0$ *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* $c \ge 0$, $(y^k)^{\mathsf{T}} s^k \le \mu^k c$,

4. $0 \le a_i(x^k) + s_i^k$ *for all* $i : a_i(x^*) = 0$,

5. *for some* $d \ge 0$, $\|\nabla_x \mathcal{L}(x^k, y^k)\| \le d\mu^k(\|y^k\|_1 + 1)$.

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

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

for some $u \ge 0$. Then, $y_i^* = 0$ whenever $\tilde{y}_i = 0$. In particular, if $\tilde{y}$ is minimially 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 $s_i^k \tilde{y}_i \ge 0$, $(s^k)^{\mathsf{T}} y^k \le \mu^k c$ and, from the proof of Theorem 3, $a_i(x^k)(y_i^k - \tilde{y}_i) \le C\mu^k$ for some $C \ge 0$, we have

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

and the result follows. □

Note that if Assumption 4. in Theorem 6 is replaced by a similar one with a strict inequality, and Assumption 5. is replaced by $\nabla \mathcal{L}(x^k, y^k)^{\mathsf{T}}(x^k - x^*) \le d\mu^k$ for some $d \ge 0$, 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 will investigate the numerical behavior of the dual sequences generated by IPOPT in the Netlib collection.

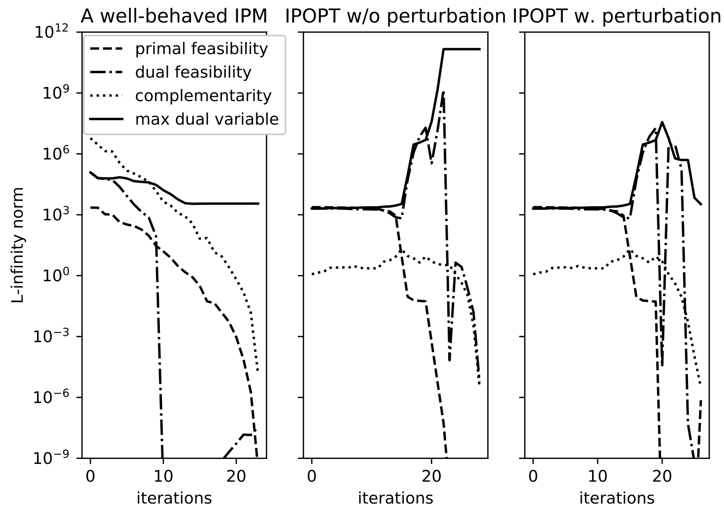

Figure 2: Comparison of the iterates of different IPMs on the Netlib problem ADLITTLE.

4 Numerical Experiments

The focus of this section is showing that on the Netlib test set – of real linear programming problems – interior point algorithms such as IPOPT, that aggressively reduce the primal feasibility, will have unnecessarily large dual iterates. As we discussed in the introduction, the convergence analysis of IPOPT and many other non-linear optimization solvers [Wächter and Biegler, 2005, Byrd et al., 2000] assume 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 on a test set such as Netlib, these assumptions are valid. As documented in Table 1 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 Appendix A for more details on how the experiments were performed.

The next natural question is to check if the violation of these assumptions translates into undesirable behavior on these instances. On this point, first consider Figure 2 where we plot the performance of IPOPT on the problem ADLITTLE from the Netlib collection. One can see, as our theory predicts, that when the primal feasibility is reduced faster than complementarity, the dual variables increase substantially. And 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 20. This contrasts with the *well-behaved* interior point solver [Hinder, 2017] that smoothly reduces the constraint violation, dual feasibility and complementarity; consequently, the maximum dual variable remains stable throughout the algorithm trajectory.

Next, we show that this phenomena occurs across the whole Netlib test set. We run these three IPMs on the Netlib problems with less than $10,000$ non-zeros and record the maximum dual variable value (across all the IPMs iterates). Of the 68 problems, 31 were solved successfully by IPOPT and all 31 of these were solved successfully by the well-behaved solver[2]. In Figure 3 we plot 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 along iterations. The plot illustrates that the maximum dual variable of IPOPT (either with or without the default perturbation) is unnecessarily large for about 40% of the problems.

[2] The well-behaved algorithm solved 57 of the 68 problems. We remark that both these solvers are general non-linear optimization solvers and are not specialized to solving linear programs.

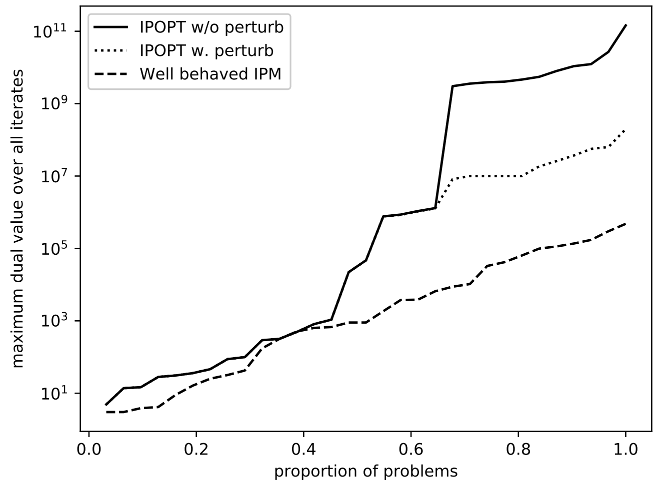

Figure 3: Comparison of the maximum dual variable value for different IPMs on the Netlib collection.

5 Final remarks

This paper shows how the behavior of the Lagrange multipliers are affected by the rate that primal feasibility and the barrier parameter are moved in IPMs. We give a theoretical explanation for this in convex and non-convex optimization and we verify that this is an issue in practice for linear programs. Preliminary results on the CUTEst collection show that this issue is also present in nonlinear problems. In the linear programming community there was awareness of this issue [Mizuno et al., 1995], and consequently, many implemented IPMs move primal feasibility and complementary at the same rate [Mehrotra, 1992, Andersen and Andersen, 2000]. However, in the general non-linear programming community, there is a lack of awareness of this issue. Consequently, with the exception of the working paper of Hinder and Ye [Hinder, 2017], general purpose infeasible start non-linear programming IPMs [Vanderbei and Shanno, 1999, Byrd et al., 2006, Wächter and Biegler, 2006] do not consider the relative rate of reduction of primal feasibility and complementarity.

References

[Andersen and Andersen, 2000] Andersen, E. D. and Andersen, K. D. (2000). The mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm. *High performance optimization*, 33:197–232.

[Andersen and Ye, 1998] Andersen, E. D. and Ye, Y. (1998). A computational study of the homogeneous algorithm for large-scale convex optimization. *Computational Optimization and Applications*, 10(3):243–269.

[Andersen and Ye, 1999] Andersen, E. D. and Ye, Y. (1999). On a homogeneous algorithm for the monotone complementarity problem. *Mathematical Programming*, 84(2):375–399.

[Andreani et al., 2011] Andreani, R., Haeser, G., and Martínez, J. (2011). On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641.

[Andreani et al., 2012a] Andreani, R., Haeser, G., Schuverdt, M. L., and Silva, P. J. S. (2012a). A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135:255–273.

[Andreani et al., 2012b] Andreani, R., Haeser, G., Schuverdt, M. L., and Silva, P. J. S. (2012b). Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22:1109–1135.

[Andreani et al., 2016] Andreani, R., Martínez, J. M., Ramos, A., and Silva, P. J. S. (2016). A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110.

[Andreani et al., 2017] Andreani, R., Martínez, J. M., Ramos, A., and Silva, P. J. S. (2017). Strict constraint qualifications and sequential optimality conditions for constrained optimization. *to appear in Mathematics of Operations Research*.

[Andreani et al., 2010] Andreani, R., Martínez, J. M., and Svaiter, B. F. (2010). A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554.

[Byrd et al., 2000] Byrd, R. H., Gilbert, J. C., and Nocedal, J. (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149–185.

[Byrd et al., 2006] Byrd, R. H., Nocedal, J., and Waltz, R. A. (2006). Knitro: An integrated package for nonlinear optimization. In *Large-scale nonlinear optimization*, pages 35–59. Springer.

[Gauvin, 1977] Gauvin, J. (1977). A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, (12):136–138.

[Giorgi et al., 2016] Giorgi, G., Jiménez, B., and Novo, V. (2016). Approximate Karush—Kuhn—Tucker condition in multiobjective optimization. *Journal of Optimization Theory and Applications*, 171(1):70–89.

[Güler and Ye, 1993] Güler, O. and Ye, Y. (1993). Convergence behavior of interior-point algorithms. *Mathematical Programming*, 60(1):215–228.

[Haeser, 2016] Haeser, G. (2016). A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Optimization online*.

[Haeser and Schuverdt, 2011] Haeser, G. and Schuverdt, M. L. (2011). On approximate KKT condition and its extension to continuous variational inequalities. *Journal of Optimization Theory and Applications*, 149(3):528–539.

[Hager and Mico-Umutesi, 2014] Hager, W. W. and Mico-Umutesi, D. (2014). Error estimation in nonlinear optimization. *Journal of Global Optimization*, (59):327–341.

[Hinder, 2017] Hinder, O.; Ye, Y. (2017). A one phase IPM for non-convex optimization. *In preparation*.

[Jeyakumar et al., 2003] Jeyakumar, V., Lee, G. M., and Dinh, N. (2003). New sequential lagrange multiplier conditions characterizing optimality without constraint qualification for convex programs. *SIAM Journal on Optimization*, 14(2):534–547.

[Lustig, 1990] Lustig, I. J. (1990). Feasibility issues in a primal-dual interior-point method for linear programming. *Mathematical Programming*, 49(1-3):145–162.

[Lustig et al., 1994] Lustig, I. J., Marsten, R. E., and Shanno, D. F. (1994). Interior point methods for linear programming: Computational state of the art. *ORSA Journal on Computing*, 6(1):1–14.

[McLinden, 1980] McLinden, L. (1980). An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88(1):101–161.

[Megiddo, 1989] Megiddo, N. (1989). Pathways to the optimal set in linear programming. In *Progress in mathematical programming*, pages 131–158. Springer.

[Mehrotra, 1992] Mehrotra, S. (1992). On the implementation of a primal-dual interior point method. *SIAM Journal on optimization*, 2(4):575–601.

[Mizuno et al., 1995] Mizuno, S., Todd, M. J., and Ye, Y. (1995). A surface of analytic centers and primal-dual infeasible-interior-point algorithms for linear programming. *Mathematics of Operations Research*, 20(1):135–162.

[Nesterov and Polyak, 2006] Nesterov, Y. and Polyak, B. T. (2006). Cubic regularization of newton method and its global performance. *Mathematical Programming*, 108(1):177–205.

[Nocedal and Wright, 2006] Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

[Qi and Wei, 2000] Qi, L. and Wei, Z. (2000). On the constant positive linear dependence conditions and its application to SQP methods. *SIAM Journal on Optimization*, 10:963–981.

[Rockafellar and Wets R., 1998] Rockafellar, R. and Wets R., J.-B. (1998). *Variational Analysis*. Springer, Berlin.

[Shanno and Vanderbei, 2000] Shanno, D. F. and Vanderbei, R. J. (2000). Interior-point methods for nonconvex nonlinear programming: orderings and higher-order methods. *Mathematical Programming*, 87(2):303–316.

[Sonnevend, 1986] Sonnevend, G. (1986). An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In *System modelling and optimization*, pages 866–875. Springer.

[Vanderbei and Shanno, 1999] Vanderbei, R. J. and Shanno, D. F. (1999). An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13(1):231–252.

[Vicente and Wright, 2002] Vicente, L. N. and Wright, S. J. (2002). Local convergence of a primal-dual method for degenerate nonlinear programming. *Computational Optimization and Applications*, 22(3):311–328.

[Wächter and Biegler, 2005] Wächter, A. and Biegler, L. T. (2005). Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM Journal on Optimization*, 16(1):1–31.

[Wächter and Biegler, 2006] Wächter, A. and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57.

[Ye and Anstreicher, 1993] Ye, Y. and Anstreicher, K. (1993). On quadratic and $O(\sqrt{n}L)$ convergence of a predictor–corrector algorithm for LCP. *Mathematical Programming*, 62(1-3):537–551.

[Ye et al., 1994] Ye, Y., Todd, M. J., and Mizuno, S. (1994). An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19(1):53–67.

A Experimental details

The linear programs in the Netlib collection come in the form $\min c^T x$ s.t. $Ax = b$, $u \geq x \geq l$. We use IPOPT 3.12.4 with the linear solver MUMPS. For IPOPT we measure the primal feasibility by $\max\{\|Ax - b\|_\infty, l - x, x - u\}$, the dual feasibility by: $\|c + A^T\lambda - z_L + z_U\|_\infty$ using the same notation from [Wächter and Biegler, 2006] and the complementary by $\|Z_L(x - l); Z_U(x - u)\|_\infty$. Similarly, for the well-behaved interior point solver, we re-write the constraints as $Ax + s_1 = b$, $Ax - s_2 = b$, $x + s_3 = l$, $x - s_4 = u$. We measure the primal feasibility as $\max\{\|Ax + s_1 - b\|_\infty, \|Ax - s_2 - b\|_\infty\}$, where the dual variables y satisfy $y \geq 0$ and $s \geq 0$. The complementarity is given by $\|(S_1 y_1; S_2 y_2; S_3 y_3; S_4 y_4)\|_\infty$, where S_1, S_2, S_3 and S_4 are diagonal with entries from the slack variables s_1, s_2, s_3 and s_4 respectively. The dual feasibility is measured by $\|c + A^T(y_1 - y_2) + (y_3 - y_4)\|_\infty$.

The termination criterion of IPOPT is set to a tolerance of 10^{-6} and we use the default options, unless otherwise stated. The termination of the well-behaved IPM is set to:

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

Table 1 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.

Table 1: 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	true
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
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		