Solving a linear program via a single unconstrained minimization

Adilet Otemissov * Alina Abdikarimova *

May 27, 2025

Abstract

This paper proposes a novel approach for solving linear programs. We reformulate a primal-dual linear program as an unconstrained minimization of a convex and twice continuously differentiable merit function. When the optimal set of the primal-dual pair is nonempty, its optimal set is equal to the optimal set of the proposed merit function. Minimizing this merit function poses some challenges due to its Hessian being singular at some points in the domain, including the optimal solutions. We handle singular Hessians using the Newton method with Levenberg-Marquardt regularization. We show that the Newton method with Levenberg-Marquardt regularization yields global convergence to a solution of the primal-dual linear program in at most $O(\epsilon^{-3/2})$ iterations requiring only the assumption that the optimal set of the primal-dual linear program is bounded. Testing on random synthetic problems demonstrates convergence to optimal solutions to very high accuracy, significantly faster than the derived worst-case bound.

Keywords: linear programming, penalty methods, regularized Newton method, global convergence, unconstrained minimization, zero residual problem

1 Introduction

In this paper, we intend to solve primal and dual linear programs in standard form:

$$\begin{aligned} \min \ & c^T x \\ \text{subject to} \ & Ax = b \\ & x \geq 0 \end{aligned} \qquad \text{(P)} \qquad\qquad \begin{aligned} \max \ & b^T \lambda \\ \text{subject to} \ & A^T\lambda + s = c \\ & s \geq 0, \end{aligned} \qquad \text{(D)}$$

where $x, s \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^m$ are the variables and where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ $(m < n)$ are given.

Linear programming problems are widely encountered in applications such as engineering, finance, transport and many more [21]. Linear programs are also commonly used as subproblems to obtain solutions to more complex problems [45]. In many of these cases, there is a demand to obtain accurate solutions to linear programs in a fast and reliable way. Meeting this demand becomes increasingly challenging as problem sizes in practical applications grow bigger. These challenges are motivating researchers to modify the current methods or seek alternative ways to solve linear programs (see, e.g., [9, 43]).

*Department of Mathematics, School of Sciences and Humanities, Nazarbayev University, Kabanbay Batyr 53, Astana 010000, Kazakhstan; aotemissov, alina.abdikarimova@nu.edu.kz. This research has been funded by Nazarbayev University under the Faculty Development Competitive Research Grant Program for the 2023–2025 Grant №20122022FD4138.

Historical overview. Linear Programming is one of the success stories in optimization. Its story began in 1950's with the advent of simplex method discovered by Dantzig [12, 13]. It has been developed and modified since and, nowadays, it is one of the state-of-the-art methods for solving linear programs [36, 42]. Simplex method performs very well in practice, however, it has not been proven yet that simplex method can solve every linear program in polynomial time. For each version of simplex, one can find instances where simplex method will spend exponential number of iterations before reaching an optimal solution [25, 42]. Later in 1979, Khachiyan [24] proposed a new method for solving linear programs called ellipsoid method. It was proven that ellipsoid method can solve linear programs in polynomial time, but it turned out to be slow in practice. Polynomial time convergence of ellipsoid method was a significant theoretical breakthrough as it indicated that there could be an algorithm that can provably solve linear programs in polynomial time and perform well in practice. Such algorithm came to light in 1984 in the famous paper by Karmarkar [23] who proposed a new polynomial time algorithm that is now classified as an Interior Point Method (IPM) (see also [35]). The discovery of this method revitalized research in linear programming resulting in fast development of IPMs. Nowadays, IPMs are one of the state-of-the-art methods for solving linear programs that perform very well in practice matching performance of simplex method.

1.1 Reformulations of linear programs

Barrier reformulation. A starting point in the derivation of an IPM is the barrier formulation of (P):

$$\begin{aligned} \min \;\; & c^T x - \nu \sum_{j=1}^{n} \log(x_j) \\ \text{subject to} \;\; & Ax = b, \end{aligned} \tag{1.1}$$

whose Lagrangian function is

$$L(x, \lambda) = c^T x - \nu \sum_{j=1}^{n} \log(x_j) + \lambda^T (b - Ax).$$

Taking the derivatives with respect to x_j's and λ_i's yield the first-order optimality conditions for $L(x, \lambda)$:

$$\begin{aligned} x_j s_j &= \nu \;\; \text{for } j = 1, \ldots, n, \\ Ax - b &= 0, \\ A^T \lambda + s - c &= 0. \end{aligned} \tag{1.2}$$

If we add nonnegativity constraints to (1.2) and set $\nu = 0$ we will end up with optimality conditions for (P) and (D). However, adding nonnegativity constraints to (1.2) will complicate the nonlinear system. Instead, one can drive the iterates towards satisfaction of the optimality conditions by solving the nonlinear system (1.2) using Newton method with ν approaching zero, while keeping x and s positive. The solution to (1.2) for different values of ν creates a central path $\{(x_\nu, \lambda_\nu, s_\nu) : \nu > 0\}$ which leads us to an optimal solution. Solution of (1.2) with Newton method for different values of ν generates a sequence of iterates that are not exactly on the path but are close to it. As ν approaches zero, barrier problem (1.1) better approximates original problem (P) and the iterates approach the boundary of the feasible region but never cross it because of the logarithmic term.

It should be noted that popular versions of IPMs allow iterates to be infeasible with respect to primal and dual equality constraints; they require only that the iterates x^k and s^k remain strictly positive.

Penalty reformulation. In the barrier formulation, the nonnegativity constraints are handled through logarithmic terms, that create a 'barrier' that the iterates are never allowed to cross. In a penalty formulation the iterates are allowed to cross into the infeasible region, but are incurred a large cost if they do so. The simplest and common penalty formulation for

$$\min f(x) \text{ subject to } g_i(x) = 0 \text{ for } i = 1, \ldots, m \tag{1.3}$$

is a quadratic penalty formulation

$$\min\ h_\nu(x) := f(x) + \frac{1}{2\nu} \sum_{i=1}^{m} g_i^2(x), \tag{1.4}$$

where $\nu > 0$ is the penalty parameter. Driving ν towards zero, we increase the penalty for violating the constraints. Given a sequence of penalty parameters ν_k that gradually approach zero, we can solve (1.4) for each value of ν^k and feed the computed solution as a starting point to minimize $h_{\nu^{k+1}}(x)$. Nocedal and Wright showed in [37, Theorem 17.1] that by doing so we generate a sequence of solutions x^k that approach an optimal solution x^* of the original problem as $\nu^k \to 0$. However, as ν^k becomes smaller, problem (1.4) becomes increasingly ill-conditioned [2, 37]. This issue can be alleviated with the use of an augmented Lagrangian formulation instead of (1.4), which can force satisfaction of the constraints at a non-zero value of the penalty parameter. Augmented Lagrangian formulation for linear programs was used, for example, in 'Idiot' crash [18, 17] to obtain approximate solutions to linear programs (see also [15, 19, 29]). This solution can then be used as a good starting point for simplex method to speed up convergence. One of the main drawbacks of a typical quadratic penalty method is its inability to produce an optimal solution to the original problem in a single minimization; it usually requires solving several optimization problems.

If, in addition to equality constraints, we have nonnegativity constraints $x \geq 0$, we can use the following formulation

$$\min\ h'_\nu(x) := f(x) + \frac{1}{2\nu} \sum_{i=1}^{m} g_i^2(x) + \frac{1}{2\nu} \sum_{j=1}^{n} \max\{-x_j, 0\}^2. \tag{1.5}$$

When x_j is negative, then the penalty term $\max\{-x_j, 0\}$ becomes positive, discouraging x from being optimal at a solution with negative entries. The penalty term $\max\{-x_j, 0\}^2$ is continuously differentiable, but is not twice continuously differentiable prohibiting the use of classical second-order methods.

Recall that a penalty function is called exact if there exists a nonzero value of ν such that a single minimization of the penalty function recovers an exact solution of the original problem. It is well known that, in general, if penalty terms in the penalty function are continuously differentiable as in (1.5), then the penalty function is not exact (see [14, 37]). In general, a nondifferentiability is a necessary requirement for the penalty function to be exact.

1.2 Our contributions

Are nondifferentiabilities unavoidable if we want to solve a problem with inequality constraints in a single minimization of a penalty function? In this regard, Bertsekas [3, p. 1] writes "Except for trivial cases, nondifferentiabilities are a necessary evil if the penalty method is to yield an optimal solution in a single minimization." However, this holds true if the penalty function has the commonly used form: 'objective function + penalty parameter × constraint violation' (see [3, Eq. (5)]). Our paper demonstrates that we can avoid first- and second-order nondifferentiabilities in an exact reformulation of a linear program if we step away from the convention of including

the objective function as a separate term in the reformulation. In Section 6, we show that this holds true for a more complex constrained optimization problem.

We propose a convex and twice continuously differentiable merit function $f_q(x, \lambda, s)$ defined in (MF) in Section 2 that combines the ideas of barrier and penalty methods. Our formulation relies on optimality conditions (1.2), but with complementary slackness $x_j s_j$ $(1 \leq j \leq n)$ replaced by the duality gap $c^T x - b^T \lambda$. Nonnegativity constraints are imposed through an addition of max functions, just like in (1.5), but raised to a power greater than 2 to make the max penalty terms twice continuously differentiable. For example, for x_j, we have $\max\{-x_j, 0\}^q$ in the merit function, where $q > 2$.

Our formulation (MF) has the following strengths:

1. (MF) is an unconstrained optimization problem.
2. Merit function f_q in (MF) is convex and twice continuously differentiable.
3. When primal-dual pair (P) and (D) have an optimal solution, the optimal set of f_q is equal to the optimal set of (P) and (D).
4. When primal-dual pair (P) and (D) have an optimal solution, the minimum of f_q is equal to zero, in other words, (MF) is a zero residual problem.
5. Based on the computed value of $\min f_q$ we can conclude whether (P) and (D) have an optimal optimal or unbounded/infeasible.
6. (MF) is free of a penalty parameter that needs to be increased or decreased, as usually done in a typical penalty method.
7. Merit function f_q is defined over the whole Euclidean space and so it allows initialization of a starting point at any point in $\mathbb{R}^{2n+m}$ including those that are infeasible for (P) and (D).
8. For $q = 3$, the Hessian of f_q is Lipschitz continuous with Lipschitz constant 1.

Our formulation (MF) has the following weaknesses:

1. The Hessian of f_q is singular at some points in the domain and when (P) and (D) have an optimal solution, the Hessian is singular at any optimal solution of the merit function.
2. Although the computed value of $\min f_q$ can tell us whether (P) and (D) do not have an optimal solution, i.e., whether they are unbounded/infeasible, it cannot distinguish between unbounded and infeasible cases.

Addressing the first weakness. Since f_q is convex and twice continuously differentiable, we can use second-order methods to minimize f_q. However, these methods need to be applied with special care since the Hessian of f_q is singular at the optimal solution and at some other points in the domain. In particular, we use Newton method with Levenberg-Marquardt regularization [26, 30]. For any optimal solution (x^*, λ^*, s^*) of primal-dual pair (P) and (D), we show that Newton method with Levenberg-Marquardt regularization achieves $|f_q(x^k, \lambda^k, s^k) - f_q(x^*, \lambda^*, s^*)| < \epsilon$ globally it at most $O(\epsilon^{-3/2})$ iterations[1] provided that the optimal set of (P) and (D) is bounded.

Numerical experiments demonstrate that Newton method (with Levenberg-Marquardt regularization) (see Algorithm 1 and Algorithm 2) can perform much better than the theoretical worst-case bound. Algorithm 1 minimizes the merit function f_q, whereas Algorithm 2 minimizes

[1] In one iteration we solve one linear system to find a search direction.

a slightly different merit function (see (HMF)) that improves quadratic approximations near a minimizer. We test the algorithms on synthetically generated random problems of varying sizes. Both algorithms produce very promising numerical results with Algorithm 2 outperforming Algorithm 1.

Addressing the second weakness. If after computing the minimum of the merit function $f_q(x, \lambda, s)$ we have $\min f_q(x, \lambda, s) > 0$ (see Corollary 2.3) we conclude that (P) and (D) have no optimal solution. However, we cannot, with certainty, say which of the problems (P) and (D) is unbounded and which is infeasible, or whether they are both infeasible. To reliably distinguish between the two cases, we suggest using the simplified homogeneous formulation (2.1) presented in [47] (see also [49]). Based on the solution of the homogeneous formulation, we can reliably conclude whether the primal (P), the dual (D), or both are infeasible.

1.3 Outline

In Section 2, we define the merit function and describe its properties in relation to primal-dual pair (P) and (D). We derive the Hessian of the merit function and show that it has two notable properties: i) it is Lipschitz continuous, but ii) singular at some points in the domain of the merit function. Section 3 discusses regularized Newton methods focusing on Levenberg-Marquardt regularization. Using existing global convergence results (see [31]), we prove that Newton method with Levenberg-Marquardt regularization provides a globally convergent scheme for finding an optimal solution of (P) and (D) with $O(\epsilon^{-3/2})$ global convergence rate. Section 4 begins with the discussion on the possible numerical problems that may arise in minimizing the merit function and explores their possible causes. We then introduce a different formulation similar to (MF) that alleviates these problems. Later in Section 4, we test both formulations on synthetically generated random linear programs that have an optimal solution. We also test both formulations on a random unbounded linear program. Section 5 concludes our work and discusses future work.

1.4 Notation and definitions

We use

$$(-x)_+^q := \begin{pmatrix} \max\{-x_1, 0\}^q & \dots & \max\{-x_n, 0\}^q \end{pmatrix}^T \tag{1.6}$$

where

$$\max\{-x_j, 0\}^q = \begin{cases} (-x_j)^q & \text{if } x_j < 0, \\ 0 & \text{if } x_j \geq 0. \end{cases}$$

We write $\text{diag}(x)$ to denote a diagonal matrix with entries of x on the diagonal. We use 0_n to denote a zero vector of size n where the specification is needed. We use $\|\cdot\|_2$ and $\|\cdot\|_F$ for the Euclidean norm and the Frobenius norm, respectively.

Definition 1.1. For a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we say that the Hessian $\nabla^2 f(x)$ is L-Lipschitz if and only if

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2 \text{ for all } x, y \in \mathbb{R}^n. \tag{1.7}$$

Definition 1.2. Let X be a set of points in $\mathbb{R}^n$. We say that X is bounded if and only if there exists a $D > 0$ such that, for every point x in X, $\|x\|_2 \leq D$.

2 The merit function

Recall that (x, λ, s) is an optimal solution to the primal-dual pair (P) and (D) if and only if (x, λ, s) satisfy

$$\begin{aligned} c^T x - b^T \lambda &= 0 \\ Ax - b &= 0 \\ A^T \lambda + s - c &= 0 \\ x, s &\geq 0. \end{aligned} \tag{OC}$$

We propose to find a solution to (OC) by finding a minimizer to the following merit function

$$\begin{aligned} \underset{(x,\lambda,s)\in\mathbb{R}^{2n+m}}{\arg\min} \; f_q(x,\lambda,s) := \frac{1}{2}(c^T x - b^T\lambda)^2 + \frac{1}{2}\|Ax-b\|_2^2 + \frac{1}{2}\|A^T\lambda + s - c\|_2^2 \\ + \frac{1}{q(q-1)} \sum_{j=1}^{n} \left(\max\{-x_j, 0\}^q + \max\{-s_j, 0\}^q\right), \end{aligned} \tag{MF}$$

where $q > 2$ is fixed. Note that the merit function f is convex and is twice continuously differentiable for $q > 2$ ([8, p. 59])[2].

Suppose that an optimal solution to (P) and (D) exists and let (x^*, λ^*, s^*) be any minimizer of (P) and (D). It is clear that optimality conditions (OC) are satisfied for (x^*, λ^*, s^*) if and only if $f_q(x^*, \lambda^*, s^*) = 0$. Given that $f_q(x, \lambda, s)$ is nonnegative for all $(x, \lambda, s) \in \mathbb{R}^{2n+m}$, this also implies that (x^*, λ^*, s^*) is an optimal solution to (MF). In other words, the optimal set of primal-dual pair (P) and (D) is equal to the optimal set of (MF) when (P) and (D) have an optimal solution. We formally state these simple observations in the following theorem.

Theorem 2.1. *A solution (x^*, λ^*, s^*) is optimal for primal-dual pair* (P) *and* (D) *if and only if (x^*, λ^*, s^*) is an optimal solution to* (MF) *and $f_q(x^*, \lambda^*, s^*) = 0$.*

Remark 2.2. When (P) and (D) have an optimal solution, we can view (MF) as a zero residual problem.

When (P) and (D) are unbounded and/or infeasible, there is no solution that satisfies (OC). In this case, the merit function f_q cannot have a finite optimum (x^*, λ^*, s^*) at which it is equal to zero, since if there were, then this solution must satisfy (OC). We state this result as a corollary of Theorem 2.1.

Corollary 2.3. *Primal-dual pair* (P) *and* (D) *are each either infeasible or unbounded if and only if $\min_{(x,\lambda,s)\in\mathbb{R}^{2n+m}} f_q(x, \lambda, s) > 0$, where $f_q(x, \lambda, s)$ is defined in* (MF).

Proof. The result follows from the fact that $f_q(x, \lambda, s) \geq 0$ for all $(x, \lambda, s) \in \mathbb{R}^{m+2n}$ and Theorem 2.1. □

Remark 2.4. The merit function is defined over the entire Euclidean space allowing initialization of the first iterate at any point in $\mathbb{R}^{2n+m}$, which could be infeasible for (P) and (D).

2.1 Detecting infeasibility

Corollary 2.3 can help us identify the cases when the primal and dual problems are infeasible and/or unbounded. If, during solution of (MF), the gradient of f_q approaches zero, but the merit function itself fails to do so, this might be an indication that (P) and (D) do not have an

[2]Note that the requirement $q > 2$ is needed for $\max\{-x_j, 0\}^q + \max\{-s_j, 0\}^q$ to be twice continuously differentiable.

optimal solution. One of the pitfalls of using this criterion, however, is its inability to distinguish between unbounded and infeasible problems. To distinguish between the two cases, we suggest using the simplified homogeneous formulation of a linear program presented in [47] (see also [49]):

$$\begin{aligned} c^T x - b^T \lambda + \kappa &= 0, \\ Ax - b\tau &= 0, \\ A^T \lambda + s - c\tau &= 0, \\ x, s, \tau, \kappa &\geq 0. \end{aligned} \tag{2.1}$$

The simplified homogeneous formulation has been one of the reliable ways to detect infeasibility in IPMs [1, 46]. Its solution can reveal whether the problem is primal infeasible, dual infeasible, or both. Simplified homogeneous formulation (2.1) is satisfied for a trivial solution $(x, \lambda, s, \tau, \kappa) = (0, 0, 0, 0, 0)$, but it can be shown that (2.1) also has a strictly complementary solution $(\hat{x}, \hat{\lambda}, \hat{s}, \hat{\tau}, \hat{\kappa})$, which satisfies $\hat{\tau}\hat{\kappa} = 0$ with $\hat{\tau} + \hat{\kappa} > 0$ and

$$\hat{x}_j \hat{s}_j = 0 \text{ and } \hat{x}_j + \hat{s}_j > 0 \text{ for } j = 1, 2, \ldots, n. \tag{2.2}$$

Having computed a strictly complementary solution $(\hat{x}, \hat{\lambda}, \hat{s}, \hat{\tau}, \hat{\kappa})$, we can compute the optimal solution for (P) and (D) or determine whether they are infeasible based on the following criteria.

1. $\hat{\tau} > 0$ if and only if (P) and (D) have an optimal solution. The optimal solution is given by $(x^*, \lambda^*, s^*) = (\hat{x}/\hat{\tau}, \hat{\lambda}/\hat{\tau}, \hat{s}/\hat{\tau})$. Note that, in this case, $\hat{\kappa} = 0$ due to complementarity.

If $\hat{\kappa} > 0$ then at least one of $-c^T\hat{x}$ and $b^T\hat{\lambda}$ must be positive.

2. If $\hat{\kappa} > 0$ and if $-c^T\hat{x} > 0$ then (D) is infeasible.

3. If $\hat{\kappa} > 0$ and if $b^T\hat{\lambda} > 0$ then (P) is infeasible.

For the proofs of the above results and for more details, refer to [1, 46, 47].

We formulate a merit function for the simplified homogeneous formulation in Appendix A. The new merit function is not much more complicated than (MF), as it includes only two additional variables, and can be solved using the same approach (see Algorithm 1 and 2) as the one we use to solve (MF).

Having said the above, when testing Algorithm 1 and 2 on an unbounded problem, we could correctly identify the problem to be unbounded without relying on the simplified homogeneous formulation. See Section 4.4 for more details.

2.2 The Hessian of the merit function

Second-order methods are a standard approach for solving (MF) and, therefore, it is imperative to understand the properties of the Hessian of the merit function. In this section, we show that the Hessian of $f_q(x, \lambda, s)$ has two notable properties: i) it is Lipschitz continuous, but ii) singular at some points in the domain of f_q.

Let us first define

$$\begin{aligned} \gamma &:= c^T x - b^T \lambda \\ \rho &:= b - Ax \\ \sigma &:= c - A^T \lambda - s \end{aligned} \tag{2.3}$$

The gradient and Hessian of $f_q(x, \lambda, s)$ are given by

$$\nabla f_q(x, \lambda, s) = \gamma \begin{pmatrix} c \\ -b \\ 0 \end{pmatrix} - \begin{pmatrix} A^T\rho \\ A\sigma \\ \sigma \end{pmatrix} - \frac{1}{q-1} \begin{pmatrix} (-x)_+^{q-1} \\ 0_m \\ (-s)_+^{q-1} \end{pmatrix}$$

and

$$\nabla^2 f_q(x,\lambda,s) = H + \operatorname{diag}((-x)_+^{q-2}, 0_m, (-s)_+^{q-2}),$$

where

$$H = \begin{pmatrix} cc^T + A^TA & -cb^T & 0 \\ -bc^T & bb^T + AA^T & A \\ 0 & A^T & I \end{pmatrix}, \tag{2.4}$$

and where $(-x)_+^q$ is defined in (1.6). Note that if (x,λ,s) satisfies optimality conditions (OC) then the gradient at (x,λ,s) is identically zero. However, the converse is not always true, for example, when (P) is unbounded.

For $(x,\lambda,s) < 0$, the Hessian is positive definite as it is a sum of H (a positive semidefinite matrix) and a diagonal matrix with strictly positive diagonal entries. However, for $(x,\lambda,s) \geq 0$, the Hessian is not positive definite as it is equal to H, whose rank is at most $m+n+1$.

Lemma 2.5. *The rank of matrix H defined in* (2.4) *is at most $m+n+1$.*

Proof. We can express H as a sum of two matrices, whose ranks are at most 1 and $m+n$:

$$H = \begin{pmatrix} R^TR & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} A^TA & 0 \\ 0 & B^TB \end{pmatrix},$$

where $R = \begin{pmatrix} c^T & -b^T \end{pmatrix}$ and $B = \begin{pmatrix} A^T & I \end{pmatrix}$. The result follows from the subadditivity of rank. □

Remark 2.6. Suppose that (P) and (D) have an optimal solution (x^*,λ^*,s^*). By Theorem 2.1, (x^*,λ^*,s^*) must be one of the minimizers of $f_q(x,\lambda,s)$. Since $(x^*,\lambda^*,s^*) \geq 0$, $\nabla^2 f_q(x^*,\lambda^*,s^*) = H$. In other words, the Hessian $\nabla^2 f_q(x^*,\lambda^*,s^*)$ is singular at any optimal solution (x^*,λ^*,s^*) of (P) and (D).

The good news is that the Hessian is Lipschitz continuous for $q=3$ (see Definition 1.1).

Lemma 2.7. *For $q=3$, the Hessian $\nabla^2 f_q(x,\lambda,s)$ is 1-Lipschitz.*

Proof. For any (x,λ,s) and (x',λ',s') we have

$$\begin{aligned} \|\nabla^2 f_3(x,\lambda,s) - \nabla^2 f_3(x',\lambda',s')\|_2^2 &= \|\operatorname{diag}((-x)_+ - (-x')_+, 0_m, (-s)_+ - (-s')_+))\|_2^2 \\ &\leq \|\operatorname{diag}((-x)_+ - (-x')_+, 0_m, (-s)_+ - (-s')_+))\|_F^2 \\ &\leq \sum_{j=1}^n (x_j - x'_j)^2 + (s_j - s'_j)^2 \\ &\leq \sum_{j=1}^n (x_j - x'_j)^2 + (s_j - s'_j)^2 + \sum_{i=1}^m (\lambda_j - \lambda'_j)^2, \end{aligned} \tag{2.5}$$

where the penultimate inequality follows from $|\max\{-y,0\} - \max\{-y',0\}| \leq |y - y'|$ for any scalars y and y'. □

3 Regularized Newton method to minimize the merit function

Classical methods to solve (MF) include gradient descent, Newton, Gauss-Newton, quasi-Newton and trust-region methods [4, 10, 37]. Gradient descent method is known to perform poorly for convex functions with even mildly ill-conditioned Hessians [4], let alone singular Hessians. Functions with singular Hessians pose challenges for second-order methods, too. For example, defining Newton search direction, which is determined by solving a linear system that involves

Algorithm 1 Newton method with Levenberg–Marquardt regularization

1: Initialize $x^0, s^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Let $\nabla^2 f^k := \nabla^2 f_q(x^k, \lambda^k, s^k)$ and $\nabla f^k := \nabla f_q(x^k, \lambda^k, s^k)$, where f_q is defined in (MF).
4: Update

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} - \alpha^k (\nabla^2 f^k + \mu^k I)^{-1} \nabla f^k, \tag{3.1}$$

5: where α^k is equal to 1 or chosen using a line search method such as Armijo backtracking.
6: **end for**

the Hessian, becomes impossible. Therefore, second-order methods must be applied with special care when minimizing functions with singular Hessians.

Numerous methods have been proposed for minimizing convex functions with singular Hessians such as cubic regularization [5, 6, 7, 20, 36, 50], trust-region-based [10, 16, 41, 44] and Levenberg-Marquardt [26, 30, 38, 40, 48] methods. Below, we briefly describe the Levenberg-Marquardt approach and, using existing results, in Theorem 3.2 we prove that Newton method with Levenberg-Marquardt regularization (Algorithm 1) converges globally to an optimal solution of the primal-dual pair (P) and (D) in at most $O(\epsilon^{-3/2})$ iterations.

For details on cubic regularization and trust-region-based approaches, we refer the reader to the respective literature.

Classical Newton method. Recall that, for a twice continuously differentiable function f at iterate x^k, a standard Newton step is defined as

$$x^{k+1} = x^k - \alpha^k \nabla^2 f(x^k)^{-1} \nabla f(x^k),$$

where α^k is the step length that is commonly defined by Armijo backtracking, Wolfe, Goldstein conditions or set to 1 for all $k \geq 1$, in which case it is pure Newton [37]. For strongly convex f, $\nabla^2 f(x^k)$ is positive definite so that $-\nabla^2 f(x^k)^{-1} \nabla f(x^k)$ is a descent direction at all iterations. For strongly convex functions, Newton method with backtracking line search is globally convergent and, with the additional assumption of a Lipschitz continuous Hessian, one can show that it enjoys local quadratic convergence to the global minimizer [4]. For general convex functions, however, such properties are not guaranteed since Newton step may not even be well defined as the inverse of $\nabla^2 f(x^k)$ may not exist. When this happens for a convex f, it implies that some eigenvalues of its Hessian $\nabla^2 f(x^k)$ are equal to zero, which means that f is locally flat in certain directions. This is the case for our merit function f in (MF). For $(x, \lambda, s) \geq 0$, the Hessian $\nabla^2 f_q(x, \lambda, s)$ has rank at most $m + n + 1$ and therefore $f_q(x, \lambda, s)$ is locally constant along a subspace of dimension at least $n - 1$.

3.1 Levenberg-Marquardt regularization

In the Levenberg-Marquardt approach (see Algorithm 1), the Hessian is modified by adding $\mu^k I$ for some positive parameter μ^k, rendering the Hessian positive definite. When μ^k is small but positive, the Newton step is well defined and the modified Hessian does not deviate much from the original Hessian. The sequence of parameters μ^k can be defined in various ways. Perhaps the simplest idea is to set μ^k to some small scalar $\mu > 0$ so that the modified Hessian is sufficiently positive definite [36, Section 3.4]. One can prove that Levenberg-Marquardt with fixed μ^k is globally convergent given that the condition numbers of the Hessians $\nabla^2 f(x^k)$ are uniformly

bounded above (see [32] and [36, Section 3.4]). In the literature, μ^k has also commonly been expressed in terms $\|\nabla f(x^k)\|$ as done in Polyak [38] who proves global convergence for convex functions and establishes $O(\epsilon^{-4})$ global convergence rate provided that the Hessian is Lipschitz continuous and that the level set of f at the initial point x^0 is compact. Ueda and Yamashita [40] extended Polyak's results to nonconvex functions and suggest $\mu^k \propto \|\nabla f(x^k)\|^\delta$. They show that with $\delta \leq 1/2$ the global convergence rate improves to $O(\epsilon^{-2})$. Recently, Mishchenko [31] improved these results showing that with the same assumptions we can achieve $O(\epsilon^{-1/2})$ rate with $\mu^k = \sqrt{(L/2)\|\nabla f(x^k)\|}$ where L is the Lipchitz constant. Below we state Mishchenko's result formally without proof.

Theorem 3.1. *[31, Theorem 1] Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let the following assumptions hold*

Assumption 1. the Hessian of f is L-Lipschitz.

Assumption 2. the objective function f has a finite optimum x^* such that $f(x^*) = \min_{x\in\mathbb{R}^n} f(x)$, and the level set $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ is bounded.

Then, Newton method with Levenberg-Marquardt regularization with $\mu^k = \sqrt{L\|\nabla f(x^k)\|/2}$ and $\alpha^k = 1$ for $k \geq 0$ achieves $f(x^k) - f(x^*) < \epsilon$ in at most $O(\epsilon^{-1/2})$ iterations.

3.2 Global convergence and its rate

Using Theorem 3.1 and the fact that the solution sets of (MF) and the primal-dual pair (P) and (D) coincide when (P) and (D) have an optimal solution, we can show that Newton method with Levenberg-Marquardt regularization (Algorithm 1) provides a globally convergent scheme for finding an optimal solution to (P) and (D):

Theorem 3.2. *Suppose that the primal-dual pair* (P) *and* (D) *have an optimal solution and that the set of optimal solutions to* (P) *and* (D) *is bounded. Then, minimizing the merit function $f_3(x, \lambda, s)$ defined in* (MF) *using Newton method with Levenberg-Marquardt regularization (Algorithm 1) with $\mu^k = \sqrt{\|\nabla f(x^k)\|/2}$ and $\alpha^k = 1$ for $k \geq 0$ achieves*

$$\|c^T x^k - b^T \lambda^k\| \leq \epsilon, \tag{3.2}$$

$$\|Ax^k - b\| \leq \epsilon, \tag{3.3}$$

$$\|A^T \lambda^k + s^k - c\| \leq \epsilon, \tag{3.4}$$

$$x_j^k \geq -\epsilon \quad \text{for } j = 1, \ldots, n, \tag{3.5}$$

$$s_j^k \geq -\epsilon \quad \text{for } j = 1, \ldots, n \tag{3.6}$$

in at most $O(\epsilon^{-3/2})$ iterations.

Proof. In Lemma 2.7 we showed that $f_q(x, \lambda, s)$ for $q = 3$ is 1-Lipschitz, so the first assumption in Theorem 3.1 is satisfied for $f_3(x, \lambda, s)$ and $L = 1$.

Now, suppose that the primal-dual pair (P) and (D) have an optimal solution and that the set of optimal solutions to (P) and (D) is a bounded set. The solution sets of (MF) and the primal-dual pair (P) and (D) coincide and hence the set of minimizers of $f_3(x, \lambda, s)$ is also a bounded set. This together with the fact that $f_3(x, \lambda, s)$ is bounded below by zero imply that $f_3(x, \lambda, s)$ has a finite optimum (x^*, λ^*, s^*). Moreover, since $f_3(x, \lambda, s)$ is convex, the level set $\{(x, \lambda, s) \in \mathbb{R}^{2n+m} : f_3(x, \lambda, s) \leq f_3(x^0, \lambda^0, s^0)\}$ is bounded for all $(x^0, \lambda^0, s^0) \in \mathbb{R}^{2n+m}$ (see [38]). Theorem 3.1 says that we have

$$f_3(x^k, \lambda^k, s^k) - f_3(x^*, \lambda^*, s^*) \leq \epsilon \tag{3.7}$$

in at most $O(\epsilon^{-1/2})$ iterations. By Theorem 2.1, we have $f_3(x^*, \lambda^*, s^*) = 0$ and hence (3.7) becomes $f_3(x^k, \lambda^k, s^k) \leq \epsilon$. Since each term of $f_3(x^k, \lambda^k, s^k)$ is nonnegative, each term of $f_3(x^k, \lambda^k, s^k)$ must be less than ϵ. In other words,

$$\begin{aligned} \|c^T x^k - b^T \lambda^k\| &\leq \sqrt{2\epsilon} \\ \|Ax^k - b\| &\leq \sqrt{2\epsilon} \\ \|A^T \lambda^k + s^k - c\| &\leq \sqrt{2\epsilon} \\ \max\{-x_j^k, 0\} &\leq \sqrt[3]{q(q-1)\epsilon} \text{ for } j = 1, \ldots, n \\ \max\{-s_j^k, 0\} &\leq \sqrt[3]{q(q-1)\epsilon} \text{ for } j = 1, \ldots, n \end{aligned} \tag{3.8}$$

is achieved in $O(\epsilon^{-1/2})$ iterations. Hence, (3.2), (3.3) and (3.4) are achieved in $O(\epsilon^{-1})$ iterations and (3.5), (3.6) are achieved in $O(\epsilon^{-3/2})$ iterations. □

The assumption in Theorem 3.1 that the optimal set of (P) and (D) has to be bounded includes the common scenario of (P) and (D) having a unique solution. Furthermore, note that the number of iterations is independent of problem dimension. If at each iteration a linear system is solved in $O(n^3)$ time then the total complexity of finding an optimal solution to (P) and (D) using Algorithm 1 is $O(n^3\epsilon^{-3/2})$.

4 Numerical considerations and results

This section is organized as follows. In Section 4.1, we consider a simple example to illustrate the challenges that formulation (MF) may pose. In Section 4.2, we propose a slightly different merit function (see (HMF)) that addresses these challenges and results in faster convergence. In Section 4.3 and Section 4.4, we test the performance of the old and new merit functions on synthetically generated random linear programs of varying dimensions. Our test set includes three linear programs with optimal solutions of dimensions $(m, n) = (100, 150)$, $(200, 300)$, $(500, 750)$ and one unbounded problem of dimension $(m, n) = (50, 150)$.

4.1 A simple example

Let us consider a simple two-dimensional function that resembles the form of merit function f_q:

$$(x_1 + x_2)^2 + \max\{-x_1, 0\}^3 + \max\{-x_2, 0\}^3. \tag{4.1}$$

The contour levels of this function are shown in the left plot in Figure 1. First, note straight contour lines in the first quadrant, indicating that the Hessian is singular in that region. Moreover, we see banana-shaped contours near the optimal solution x^*; these are known to cause difficulties (recall Rosenbrock function [39]) even for Newton's method as its quadratic model may not approximate the function well near the solution.

The right plot in Figure 1 draws $\max\{-x, 0\}^3$ against its quadratic approximation near a point slightly less than zero. We see that the quadratic model approximates $\max\{-x, 0\}^3$ well in the neighbourhood of zero for $x < 0$, but does not yield a good approximation for $x > 0$.

4.2 Algorithmic enhancements

We can improve the contour lines and quadratic approximations near solutions by considering a slightly different merit function:

$$\underset{(x,\lambda,s)\in\mathbb{R}^{2n+m}}{\arg\min} \; h_{q,\nu}(x, \lambda, s) := f_q(x, \lambda, s) + \frac{\nu}{q(q-1)} \sum_{j=1}^{n} \left(\max\{x_j, 0\}^q + \max\{s_j, 0\}^q \right), \tag{HMF}$$

where $\nu > 0$ and where $f_q(x, \lambda, s)$ is the original merit function defined in (MF).

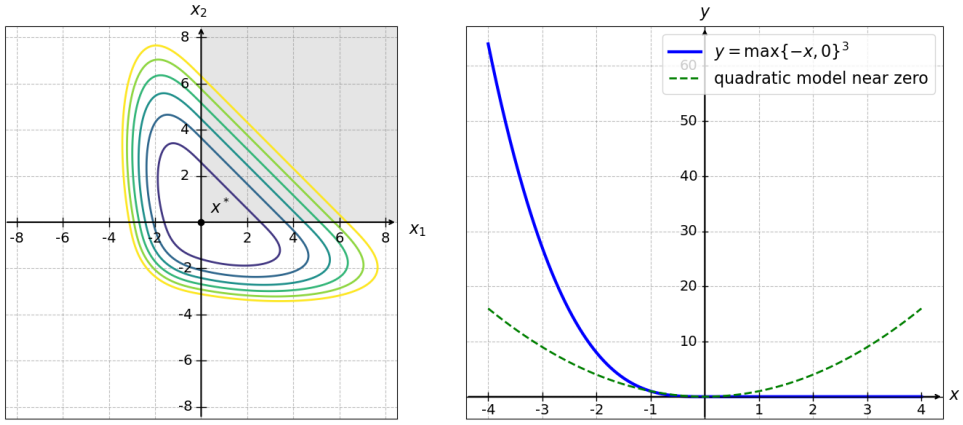

Figure 1: Left: A contour plot of (4.1). Right: Illustration of $\max\{-x, 0\}^3$ and its quadratic model near a point slightly less than zero.

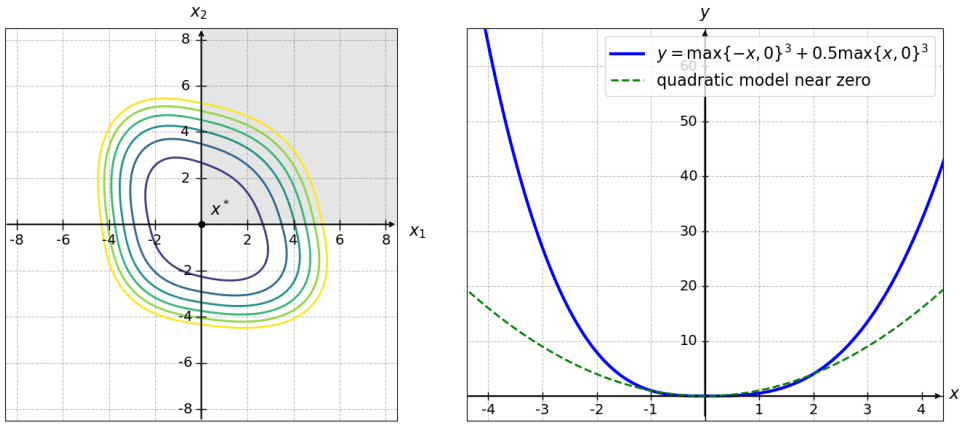

Figure 2: Left: A contour plot of (4.2). Right: Illustration of $\max\{-x, 0\}^3 + 0.5\max\{x, 0\}^3$ and its quadratic model near zero.

A simple example revisited. Let us revisit the function in (4.1) but with additional max functions to resemble the form of (HMF). We would like to see the changes in contour lines and the quadratic approximation of max functions. Consider

$$(x_1 + x_2)^2 + (\max\{-x_1, 0\}^3 + \max\{-x_2, 0\}^3) + 0.5(\max\{x_1, 0\}^3 + \max\{x_2, 0\}^3). \tag{4.2}$$

The contour lines shown in Figure 2 are now close to elliptical – a significant improvement over banana-shaped contours and more tractable for Newton's method. The right plot in Figure 2 also shows a better quadratic approximation of the max functions of (4.2) near zero.

Improved Algorithm. Instead of solving (MF), we propose to solve (HMF) following the scheme of Algorithm 1, but with gradually decreasing ν at each subsequent iteration (see Algorithm 2). We borrow the idea from homotopy methods [33, 34, 37, 46]: as v approaches 0, the new merit function $h_{q,\nu}$ approximates f_q better and, hopefully, the iterates are gradually driven towards the optimal set of f_q. For initially large values of ν, contour lines of (HMF) are near elliptical, but they gradually become banana-shaped as ν gets smaller. By that stage, we hope that the iterates are in the vicinity of the optimal set and fast local convergence takes over (see Figure 3).

Algorithm 2 Newton method with Levenberg–Marquardt regularization for the modified merit function

1: Initialize $x^0, s^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$, $\nu^0 > 0$, $\theta \in (0,1)$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Let $\nabla^2 h^k := \nabla^2 h_{q,\nu^k}(x^k, \lambda^k, s^k)$ and $\nabla h^k := \nabla h_{q,\nu^k}(x^k, \lambda^k, s^k)$, where $h_{q,\nu}$ is defined in (HMF).
4: Update

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ s^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \\ s^k \end{pmatrix} - \alpha^k (\nabla^2 h^k + \mu^k I)^{-1} \nabla h^k, \tag{4.3}$$

5: where α^k is equal to 1 or chosen using a line search method such as Armijo backtracking.
6: Update $\nu^{k+1} = \theta \nu^k$.
7: **end for**

4.3 Linear programs with an optimal solution

We test the performance of Algorithm 1 and Algorithm 2 on synthetically generated random linear programs of dimensions $(m, n) = (100, 150)$, $(200, 300)$, $(500, 750)$. All three problems have optimal solutions. We test two variants of Algorithm 1 and one variant of Algorithm 2 with parameters outlined in the following table.

	value of q	regularization parameter	line search	value of θ
Algorithm 1a	3.0	$\mu^k = \sqrt{(1/2)\|\nabla f(x^k)\|}$	$\alpha^k = 1$ for all $k \geq 0$	N/A
Algorithm 1b	2.1	$\mu^k = 10^{-9}$ for all $k \geq 0$	Armijo backtracking	N/A
Algorithm 2	2.1	$\mu^k = 10^{-9}$ for all $k \geq 0$	Armijo backtracking	0.8

Table 1: The set of parameters used in the experiments for Algorithm 1 and Algorithm 2.

The results of the experiments are provided in Figure 3. The leftmost plots show the results of Algorithm 1a for the adaptive choice of μ^k. We observe slow sublinear convergence aligning with the theoretical worst-case sublinear rate derived in Theorem 3.2. Algorithm 1b (the middle plots) and Algorithm 2 (the rightmost plots) perform better. Interestingly, Algorithm 1b exhibits slow initial convergence followed by fast convergence resembling the behaviour of the Newton method for well-conditioned problems. Algorithm 2 dramatically reduces the phase of slow convergence, resulting in faster convergence to the optimal solution. In Table 2, we provide the relative errors of x^k with respect to the primal optimal solution x^* of the last ten iterations of Algorithm 1b and Algorithm 2. Implementations of Algorithm 1 and 2 on a single random problem are available on GitHub[3].

Remark 4.1. Results for Algorithm 2 may suggest that the performance of the algorithm is independent of problem dimension. Our extensive testing showed that this, in general, is not true. Choosing smaller value of θ for smaller dimensional problems will result in faster convergence, while the same value of θ for larger dimensional problems may decrease ν too fast, resulting in stalling of the iterates similar to the initial slow convergence phase of Algorithm 1b.

4.4 An unbounded linear program

We run Algorithm 1a, Algorithm 1b and Algorithm 2 on a randomly generated unbounded linear program with the same parameters provided in Table 1. Our synthetic unbounded problem has

[3] `https://github.com/alinaabdikarimova/lp-via-unconstrained-minimization`

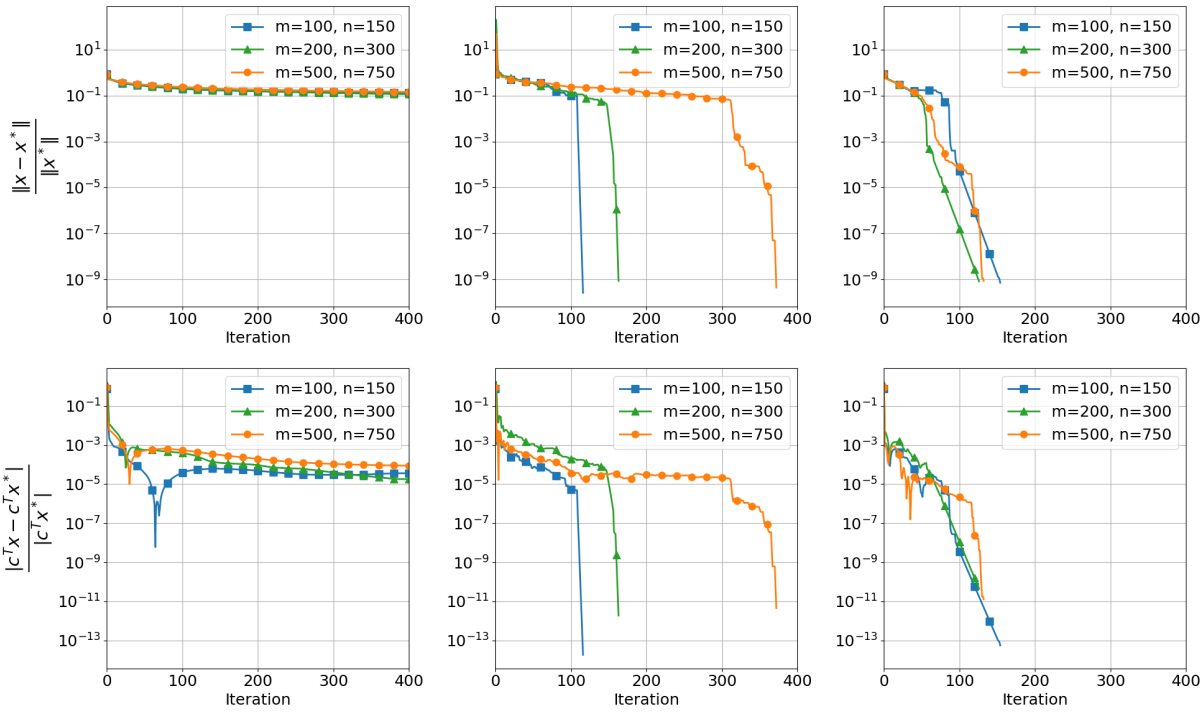

Figure 3: This figure shows results for Algorithm 1a (leftmost plots), Algorithm 1b (middle plots) and Algorithm 2 (rightmost plots) when tested on three synthetically generated random linear programs of dimensions $(m, n) = (100, 150)$, $(200, 300)$, $(500, 750)$. The plots in the first row show the relative error of x^k with respect to the primal optimal solution x^*. The plots in the second row show the relative error of $c^T x^k$ with respect to the primal objective function value at the primal optimal solution.

Algorithm 1b	Algorithm 2
9.7×10^{-2}	4.5×10^{-9}
9.7×10^{-2}	3.7×10^{-9}
4.6×10^{-3}	3.0×10^{-9}
4.1×10^{-4}	2.5×10^{-9}
3.8×10^{-5}	2.0×10^{-9}
3.5×10^{-6}	1.6×10^{-9}
3.2×10^{-7}	1.3×10^{-9}
3.0×10^{-8}	1.2×10^{-9}
2.8×10^{-9}	1.0×10^{-9}
2.6×10^{-10}	7.2×10^{-10}

$m = 100, n = 150$

Algorithm 1b	Algorithm 2
8.3×10^{-4}	5.1×10^{-9}
4.5×10^{-4}	4.2×10^{-9}
2.4×10^{-4}	3.4×10^{-9}
1.6×10^{-5}	2.8×10^{-9}
1.4×10^{-5}	2.3×10^{-9}
1.3×10^{-5}	1.9×10^{-9}
1.2×10^{-6}	1.5×10^{-9}
1.1×10^{-7}	1.2×10^{-9}
9.6×10^{-9}	1.0×10^{-9}
8.8×10^{-10}	8.3×10^{-10}

$m = 200, n = 300$

Algorithm 1b	Algorithm 2
4.8×10^{-6}	7.0×10^{-7}
4.7×10^{-6}	7.0×10^{-7}
4.7×10^{-6}	7.0×10^{-7}
4.7×10^{-7}	1.8×10^{-7}
5.0×10^{-8}	1.7×10^{-7}
5.0×10^{-8}	2.6×10^{-8}
5.0×10^{-8}	4.3×10^{-9}
5.0×10^{-8}	1.1×10^{-9}
4.7×10^{-9}	1.1×10^{-9}
4.5×10^{-10}	8.6×10^{-10}

$m = 500, n = 750$

Table 2: Relative errors of x^k with respect to the primal optimal solution x^* of the last ten iterations of Algorithm 1b and Algorithm 2.

dimensions $(m, n) = (50, 150)$. We first solved this problem with Python's HiGHS solver [22] to verify unboundedness. The results of the experiment are provided in Figure 4.

From all three plots in Figure 4 we observe that $\|\nabla f_q\|$ approaches zero, so the algorithms converge to an optimal solution. Since f_q remains bounded away from zero, from Corollary 2.3 it follows that the problem is either infeasible or unbounded. Since $\|Ax - b\|$ approaches zero, indicating feasibility of the primal problem, we conclude that the problem is unbounded.

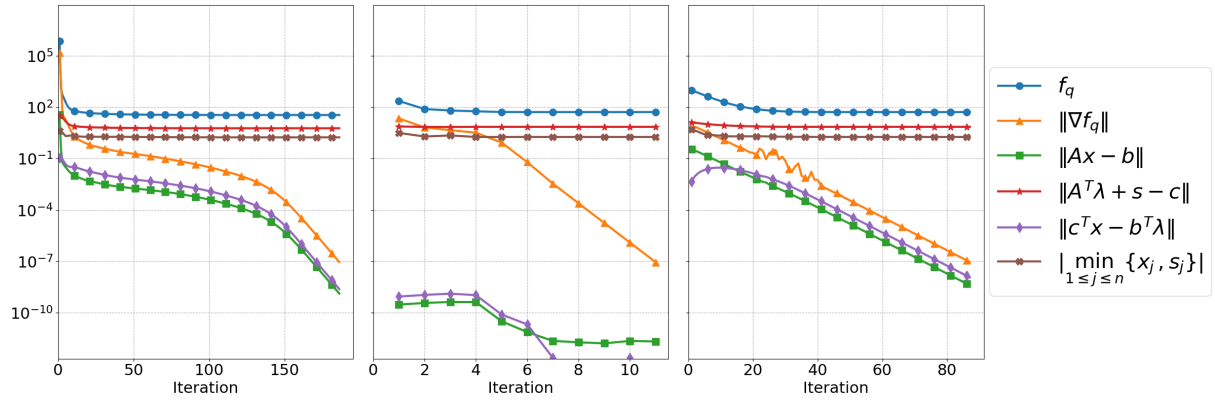

Figure 4: This figure shows results for Algorithm 1a (leftmost plot), Algorithm 1b (middle plot) and Algorithm 2 (rightmost plot) when tested on a synthetically generated random unbounded linear program of dimension $(m, n) = (50, 150)$.

5 Summary

In this paper, we have demonstrated that solving primal-dual linear program (P) and (D) is equivalent to finding an unconstrained minimizer of a convex and twice continuously differentiable merit function $f_q(x, \lambda, x)$ defined in (MF). The merit function $f_q(x, \lambda, x)$ is a sum of i) the squared duality gap $c^T x - b^T \lambda$ to encourage optimality, ii) the squared equality constraint terms $Ax - b$ and $A^T\lambda + s - c$ to encourage primal and dual feasibility, and iii) the nonnegativity penalty terms $\max\{-x_j, 0\}^q$ and $\max\{-s_j, 0\}^q$ to encourage nonnegative solutions, where $q > 2$ to ensure second-order differentiability. From the optimality conditions (OC) for linear programs, it follows that (x^*, λ^*, x^*) is a solution to the primal-dual linear program if and only if (x^*, λ^*, x^*) is a minimizer of the merit function $f_q(x, \lambda, x)$ and $f_q(x^*, \lambda^*, x^*) = 0$. In other words, when the primal-dual pair have an optimal solution, its unconstrained reformulation (MF) can be viewed as a zero residual problem. Moreover, the merit function is defined over the entire $\mathbb{R}^{2n+m}$ allowing initialization of the first iterate at a point that is infeasible for (P) and (D).

We saw that minimizing $f_q(x, \lambda, x)$ is not trivial due to its Hessian being singular at the optimal solutions and at some other points in the domain (for example, for $(x, \lambda, s) > 0$). To handle singular Hessians, we regularized Newton with Levenberg-Marquardt approach. We proved that this variant of the regularized Newton method with a particular choice of the regularization parameter μ^k (see Theorem 3.2) converges globally to an optimal solution of the primal-dual pair at $O(\epsilon^{-3/2})$ rate requiring only the assumption that the optimal set of the primal-dual linear program is bounded. Note that this assumption includes the common scenario of the linear program having a unique solution.

Before conducting the numerical experiments, we illustrated the challenges that the merit function can pose during minimization and introduced a slightly modified merit function $h_{q,\nu}$ defined in (HMF) with an improved landscape and better conditioned Hessians to help Newton to converge faster. We conducted numerical experiments on synthetically generated random linear programs of varying dimensions. We minimized the merit function f_q with two variants of the Newton method with Levenberg-Marquardt regularization corresponding to two different ways of defining regularization parameter μ^k (i.e., adaptive and constant μ^k) and minimized the new merit function $h_{q,\nu}$ with the same method with constant μ^k. On problems with optimal solutions, we observed convergence to very high accuracy for both merit functions, except for the variant with adaptive μ^k. Between the two successful variants, the new merit function $h_{q,\nu}$ converged to optimal solutions faster avoiding the initial slow convergence phase observed for the merit function f_q.

6 Future work

Our work can be improved in multiple ways.

- **Modifications of the merit function.** We saw in Section 4.3 that a modification (HMF) of the old merit function in (MF) achieves faster convergence. A modification of (MF) can take different forms. For example, we can add $0.5\nu\|\lambda\|^2$ to $h_{q,\nu}$ in (HMF). This addition will make the Hessian of $h_{q,\nu}$ positive definite for all $\nu > 0$ and there is no need to regularize Newton except at the last iterations, when ν gets extremely small. Moreover, one may need an adaptive strategy for decreasing ν rather than at a fixed rate. In our experiments, we observed that for smaller dimensional problems, we can decrease ν more aggressively, achieving even faster convergence.

- **Improved convergence rate.** Our numerical experiments demonstrate that we can do much better in practice than the worst-case bound derived in Theorem 3.2, especially at the final iterations when the iterates are close to the optimal set. Regularized Newton method is well-known to converge globally to the optimal set of a convex function [38] with global convergence rates derived, for example, in [31, 38, 40]. Local convergence rates have also been studied. It was proven that (see, e.g., [11, 27]), under the so-called local error bound assumption, regularized Newton method inherits classical Newton's local quadratic convergence property. It is very possible that, with very mild assumptions, the merit function in (MF) or its modification satisfies the local error bound.

- **Cheap iterations.** Each iteration in Algorithm 1 and 2 require a solution of a linear system to find the Newton's search direction. One may consider computing the search directions approximately to cheapen each iteration's cost (see [11, 27, 28]). Dan et al. [11] showed that an inexact regularized Newton method is globally convergent under mild assumptions and, under the additional assumption of the local error bound, enjoys local superlinear convergence to the optimal set.

Of course, there are many more avenues that we can explore to enhance and extend this work. One of the important possible extensions of this work concerns unconstrained formulations of more complex constrained optimization problems. The fact that complementary slackness in linear programs can be replaced by the duality gap, that is linear in the variables of the problem, allowed us to formulate a convex merit function. If we consider KKT conditions for more complex problems and use a similar approach to put them in one unconstrained formulation, what properties would the resulting merit function have? How easy would it be to find its solutions?

Consider, for example, a nonnegative quadratic programming problem:

$$\min \ \frac{1}{2}x^TQx + c^Tx \ \text{ subject to } \ x \geq 0. \tag{6.1}$$

Its KKT conditions are given by

$$\begin{aligned} Qx + c - \lambda &= 0_n \\ x &\geq 0 \\ \lambda &\geq 0 \\ x_j\lambda_j &= 0 \text{ for } j = 1, \ldots, n. \end{aligned} \tag{6.2}$$

We can keep the complementary slackness conditions $x_j\lambda_j = 0$ as they are or replace them with $x^T\lambda = 0$, since x and λ are nonnegative. With the latter choice, a possible merit function for

(6.2) is

$$\min_{x,\lambda} m_q(x,\lambda) := \|Qx + c - \lambda\|_2^2 + \sum_{j=1}^{n}(\max\{-x_j,0\}^q + \max\{-\lambda_j,0\}^q) + (x^T\lambda)^2, \tag{6.3}$$

where $q > 2$ and where only $(x^T\lambda)^2$ is nonconvex. Note that (x^*, λ^*) is a solution to (6.2) if and only if (x^*, λ^*) is a solution to (6.3) and $m_q(x^*, \lambda^*) = 0$ (an analogue of Theorem 2.1 for linear programs). Since KKT conditions (6.2) provide sufficient and necessary conditions for optimality, constrained problem (6.1) and unconstrained problem (6.3) are equivalent.

A Simplified homogeneous model's merit function

Here, we provide the merit function and derive the gradient and Hessian for the simplified homogeneous model (2.1). Define

$$\begin{aligned} \gamma &:= c^T x - b^T y + \kappa \\ \rho &:= b\tau - Ax \\ \sigma &:= c\tau - A^T y - s \end{aligned} \tag{A.1}$$

The merit function for the simplified homogeneous model is given by

$$\begin{aligned} \arg\min_{x,\lambda,s} f_q(x,y,s,\tau,\kappa) :=& \frac{1}{2}(c^T x - b^T\lambda + \kappa)^2 + \frac{1}{2}||Ax - b\tau||_2^2 + \frac{1}{2}||A^T\lambda + s - c\tau||_2^2 \\ &+ \frac{1}{q(q-1)}\Bigg(\max\{-\tau,0\}^q + \\ &+ \max\{-\kappa,0\}^q + \sum_{j=1}^{n}\max\{-x_j,0\}^q + \max\{-s_j,0\}^q\Bigg), \end{aligned} \tag{MF1}$$

The gradient and Hessian of $f_q(x,y,s,\tau,\kappa)$ are given by

$$\nabla f_q(x,y,s,\tau,\kappa) = \gamma\begin{pmatrix} c \\ -b \\ 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} A^T\rho \\ A\sigma \\ \sigma \\ w \\ 0 \end{pmatrix} - \frac{1}{q-1}\begin{pmatrix} (-x)_+^{q-1} \\ 0_m \\ (-s)_+^{q-1} \\ (-\tau)_+^{q-1} \\ (-\kappa)_+^{q-1} \end{pmatrix}, \tag{A.2}$$

where $w = (Ax)^T b + (A^T y + s)^T c - (\|b\|^2 + \|c\|^2)\tau$, and

$$\nabla^2 f(x,y,s,\tau,\kappa) = H + \mathrm{diag}((-x)_+^{q-2}, 0_m, (-s)_+^{q-2}, (-\tau)_+^{q-2}, (-\kappa)_+^{q-2}), \tag{A.3}$$

where

$$H = \begin{pmatrix} cc^T + A^TA & -cb^T & 0 & -A^Tb & c \\ -bc^T & bb^T + AA^T & A & -Ac & -b \\ 0 & A^T & I & -c & 0 \\ -b^TA & -c^TA^T & -c^T & \|b\|^2 + \|c\|^2 & 0 \\ c^T & -b^T & 0 & 0 & 1 \end{pmatrix}. \tag{A.4}$$

References

[1] E. D. Andersen and K. D. Andersen. "The Mosek Interior Point Optimizer for Linear Programming: An Implementation of the Homogeneous Algorithm". In: *High Performance Optimization*. Ed. by H. Frenk, K. Roos, T. Terlaky, and S. Zhang. Boston, MA: Springer US, 2000, pp. 197–232.

[2] A. Ben-Tal and A. Nemirovski. *Optimization III: Convex Optimization*. Lecture notes, Georgia Institute of Technology. 2023.

[3] D. Bertsekas. "Necessary and sufficient conditions for a penalty method to be exact". English (US). In: *Mathematical Programming* 9.1 (Dec. 1975), pp. 87–99.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[5] C. Cartis, N. I. M. Gould, and P. L. Toint. "On the Complexity of Steepest Descent, Newton's and Regularized Newton's Methods for Nonconvex Unconstrained Optimization Problems". In: *SIAM Journal on Optimization* 20.6 (2010), pp. 2833–2852.

[6] C. Cartis, N. I. M. Gould, and P. L. Toint. "Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results". In: *Mathematical Programming* 127.2 (2011), pp. 245–295.

[7] C. Cartis, N. I. M. Gould, and P. L. Toint. "Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity". In: *Mathematical Programming* 130.2 (2011), pp. 295–319.

[8] E. K. P. Chong and S. H. Zak. *An Introduction to Optimization*. 2nd ed. New York: Wiley-Interscience, 2001.

[9] A. Chowdhury, G. Dexter, P. London, H. Avron, and P. Drineas. "Faster Randomized Interior Point Methods for Tall/Wide Linear Programs". In: *Journal of Machine Learning Research* 23 (2022), 336:1–336:48.

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust Region Methods*. MPS-SIAM Series on Optimization. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2000.

[11] H. Dan, N. Yamashita, and M. Fukushima. "Convergence Properties of the Inexact Levenberg-Marquardt Method under Local Error Bound Conditions". In: *Optimization Methods and Software* 17.4 (2002), pp. 605–626.

[12] G. B. Dantzig. "A Proof of the Equivalence of the Programming Problem and the Game Problem". In: *Activity Analysis of Production and Allocation* (1951). Ed. by T. C. Koopmans, pp. 330–335.

[13] G. B. Dantzig. "Application of the Simplex Method to a Transportation Problem". In: *Activity Analysis of Production and Allocation* (1951). Ed. by T. C. Koopmans, pp. 359–373.

[14] M. Dolgopolik. "A Unifying Theory of Exactness of Linear Penalty Functions". In: *Optimization* 65.6 (2016), pp. 1167–1202.

[15] Y. G. Evtushenko, A. I. Golikov, and N. Mollaverdy. "Augmented Lagrangian method for large-scale linear programming problems". In: *Optimization Methods and Software* 20.4–5 (2005), pp. 515–524.

[16] J. Fan and Y. Yuan. "A regularized Newton method for monotone nonlinear equations and its application". In: *Optimization Methods and Software* 29.1 (2014), pp. 102–119.

[17] J. Forrest, S. Vigerske, T. Ralphs, J. Forrest, L. Hafer, jpfasano, H. G. Santos, Jan-Willem, M. Saltzman, a-andre, B. Kristjansson, h-i-gassmann, A. King, Arevall, B. Mart, P. Bonami, R. Luies, S. Brito, and to-st. *coin-or/Clp: Release releases/1.17.10*. Version releases/1.17.10. Aug. 2024.

[18] I. L. Galabova and J. A. J. Hall. "The 'Idiot' crash quadratic penalty algorithm for linear programming and its application to linearizations of quadratic assignment problems". In: *Optimization Methods and Software* 35.3 (2020), pp. 488–501.

[19] O. Güler. "Augmented Lagrangian algorithms for linear programming". In: *Journal of Optimization Theory and Applications* 75.3 (1992), pp. 445–470.

[20] F. Hanzely, N. Doikov, P. Richtárik, and Y. Nesterov. "Stochastic subspace cubic Newton method". In: *Proceedings of the 37th International Conference on Machine Learning*. ICML'20. JMLR.org, 2020.

[21] F. S. Hillier and G. J. Lieberman. *Introduction to Operations Research*. 10th ed. McGraw-Hill Education, 2014.

[22] Q. Huangfu, I. Galabova, M. Feldmeier, and J. A. J. Hall. "HiGHS – High performance software for linear optimization". In: *INFORMS Journal on Computing* 35.1 (2023), pp. 1–20.

[23] N. Karmarkar. "A New Polynomial-Time Algorithm for Linear Programming". In: *Combinatorica* 4.4 (1984), pp. 373–395.

[24] L. G. Khachiyan. "A Polynomial Algorithm in Linear Programming". In: *Soviet Mathematics Doklady* 20 (1979). Originally published in Russian in Doklady Akademii Nauk SSSR, 244 (1979), 1093–1096, pp. 191–194.

[25] V. Klee and G. J. Minty. "How good is the simplex algorithm?" In: *Inequalities III* (1972). Ed. by O. Shisha, pp. 159–175.

[26] K. Levenberg. "A Method for the Solution of Certain Nonlinear Problems in Least Squares". In: *Quarterly of Applied Mathematics* 2 (1944), pp. 164–168.

[27] D.-H. Li, M. Fukushima, L. Qi, and N. Yamashita. "Regularized Newton Methods for Convex Minimization Problems with Singular Solutions". In: *Computational Optimization and Applications* 28.2 (2004), pp. 131–147.

[28] Y.-J. Li and D.-H. Li. "Truncated regularized Newton method for convex minimizations". In: *Computational Optimization and Applications* 43.1 (2009), pp. 119–131.

[29] O. L. Mangasarian. "A Newton Method for Linear Programming". In: *Journal of Optimization Theory and Applications* 121.1 (2004), pp. 1–18.

[30] D. Marquardt. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". In: *Journal of the Society for Industrial and Applied Mathematics* 11 (1963), pp. 431–441.

[31] K. Mishchenko. "Regularized Newton Method with Global $\boldsymbol{\mathcal{O}(1/k^2)}$ Convergence". In: *SIAM Journal on Optimization* 33.3 (2023), pp. 1440–1462.

[32] J. J. Moré, D. C. Sorensen, and G. H. Golub. "Newton's Method". In: *Studies in Numerical Analysis*. Vol. 24. MAA Studies in Mathematics. Washington, DC: Mathematical Association of America, 1984, pp. 29–82.

[33] J. L. Nazareth. "Homotopy techniques in linear programming". In: *Algorithmica* 1.1 (1986), pp. 529–535.

[34] J. L. Nazareth. "The homotopy principle and algorithms for linear programming". In: *SIAM Journal on Optimization* 1.3 (1991), pp. 316–332.

[35] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1994.

[36] Y. Nesterov and B. T. Polyak. "Cubic regularization of Newton method and its global performance". In: *Mathematical Programming* 108.1 (2006), pp. 177–205.

[37] J. Nocedal and S. J. Wright. *Numerical Optimization*. 2nd ed. Springer Series in Operations Research and Financial Engineering. eBook ISBN: 978-0-387-40065-5; Softcover ISBN: 978-1-4939-3711-0. Springer New York, NY, 2006, pp. XXII, 664.

[38] R. A. Polyak. "Regularized Newton method for unconstrained convex optimization". In: *Mathematical Programming* 120.1 (2009), pp. 125–145.

[39] H. H. Rosenbrock. "An Automatic Method for Finding the Greatest or Least Value of a Function". In: *The Computer Journal* 3.3 (Jan. 1960), pp. 175–184.

[40] K. Ueda and N. Yamashita. "Convergence Properties of the Regularized Newton Method for the Unconstrained Nonconvex Optimization". In: *Applied Mathematics and Optimization* 62.1 (2010), pp. 27–46.

[41] K. Ueda and N. Yamashita. "A regularized Newton method without line search for unconstrained optimization". In: *Computational Optimization and Applications* 59.1 (2014), pp. 321–351.

[42] R. J. Vanderbei. *Linear Programming: Foundations and Extensions*. 5th ed. International Series in Operations Research & Management Science. Springer, 2020.

[43] K. Vu, P.-L. Poirion, and L. Liberti. "Random projections for linear programming". In: *Mathematics of Operations Research* 43.4 (2018), pp. 1051–1071.

[44] T. Wang and Q. Huang. "A new Newton method for convex optimization problems with singular Hessian matrices". In: *AIMS Mathematics* 8.9 (2023), pp. 21161–21175.

[45] L. A. Wolsey. *Integer Programming*. 2nd ed. John Wiley & Sons, 2020.

[46] S. J. Wright. *Primal-dual interior-point methods*. USA: Society for Industrial and Applied Mathematics, 1997.

[47] X. Xu, P.-F. Hung, and Y. Ye. "A simplified homogeneous and self-dual linear programming algorithm and its implementation". In: *Annals of Operations Research* 62.1 (1996), pp. 151–171.

[48] N. Yamashita and M. Fukushima. "On the Rate of Convergence of the Levenberg-Marquardt Method". In: *Topics in Numerical Analysis*. Ed. by G. Alefeld and X. Chen. Vienna: Springer Vienna, 2001, pp. 239–249.

[49] Y. Ye, M. J. Todd, and S. Mizuno. "An $O(\sqrt{n}L)$-Iteration Homogeneous and Self-Dual Linear Programming Algorithm". In: *Mathematics of Operations Research* 19.1 (1994), pp. 53–67. (Visited on 04/22/2025).

[50] M.-C. Yue, Z. Zhou, and A. Man-Cho So. "On the Quadratic Convergence of the Cubic Regularization Method under a Local Error Bound Condition". In: *SIAM Journal on Optimization* 29.1 (2019), pp. 904–932.