# Primal-dual relationship between Levenberg-Marquardt and central trajectories for linearly constrained convex optimization

Roger Behling[a], Clovis Gonzaga[b] and Gabriel Haeser[c**]

March 21, 2013

[a]*Department of Management Science and Engineering, Stanford University, Stanford, USA.*;
[b]*Federal University of Santa Catarina, Florianópolis-SC, Brazil.*;
[c]*Institute of Science and Technology, Federal University of São Paulo, São José dos Campos-SP, Brazil.*

## Abstract

We consider the minimization of a convex function on a compact polyhedron defined by linear equality constraints and nonnegative variables. We define the Levenberg-Marquardt (L-M) and central trajectories starting at the analytic center and using the same parameter, and show that they satisfy a primal-dual relationship, being close to each other for large values of the parameter. Based on this we develop an algorithm that starts computing primal-dual feasible points on the L-M trajectory and eventually moves to the central path. Our main theorem is particularly relevant in quadratic programming, where points on the primal-dual L-M trajectory can be calculated by means of a system of linear equations. We present some computational tests related to box constrained trust region subproblems.

Keywords: central path, Levenberg-Marquardt, primal-dual, interior points, convex quadratic programming, trust region, initial point
Classification code: 90C25; 90C51.

## 1 Introduction

In this paper we describe a path following algorithm based both on the central path and on the Levenberg-Marquardt trajectories for solving the following problem:

---

** Corresponding author. E-mail: gabriel.haeser@unifesp.br

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array} \qquad (1)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex continuously differentiable function, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. We denote the feasible set $\Omega := \{x \in \mathbb{R}^n | Ax = b, x \geq 0\}$ and the set of interior points $\Omega^0 := \{x \in \Omega | x > 0\}$.

We assume that $\Omega$ is bounded and $\Omega^0 \neq \emptyset$. Then the logarithmic barrier function $x \in \Omega^0 \mapsto p_{\log}(x) := -\sum_{i=1}^{n} \log(x_i)$ is well-defined, as well as the analytic center of $\Omega$, $\chi = \operatorname{argmin}\{p_{\log}(x) \,|\, x \in \Omega^0\}$ (see [13]).

To simplify the treatment, we assume that

$$\chi = e = (1, 1, \dots, 1) \in \mathbb{R}^n.$$

This assumption is made without loss of generality, as we now show: given the problem

$$\begin{array}{ll} \text{minimize} & f_0(w) \\ \text{subject to} & A_0 w = b \\ & w \geq 0 \end{array} \qquad (2)$$

with the same hypotheses as above but with analytic center $w_{AN} > 0$, the format (1) is obtained by defining $W_{AN} = \operatorname{diag}(w_{AN})$ and scaling the problem by setting $w = W_{AN}x$, $A = A_0 W_{AN}$ and for $x \in \mathbb{R}^n$, $f(x) = f_0(W_{AN}x)$. The scaled problem inherits the convexity of the original one because $W_{AN}$ is positive definite, and now the analytic center becomes $\chi = e$, because $w_{AN} = W_{AN}e$.

**The trajectories.** The objective function of (1) will be modified to $f(x) + \mu p(x)$ by penalty terms in two ways:

(a) with the logarithmic barrier $p_{\log}$, which determines the primal central path

$$\mu > 0 \mapsto x_{\log}(\mu) = \operatorname*{argmin}_{x \in \Omega^0} f(x) + \mu p_{\log}(x).$$

(b) with the L-M regularization

$$x \in \mathbb{R}^n \mapsto p_{LM}(x) := \|x - e\|^2 / 2,$$

which defines the primal L-M trajectory

$$\mu > 0 \mapsto x_{LM}(\mu) = \operatorname{argmin}\{f(x) + \mu p_{LM}(x) \,|\, Ax = b, x \in \mathbb{R}^n\}. \qquad (3)$$

The idea of this quadratic penalization goes back to [7, 8].

The central path is contained in $\Omega^0$ and is scale independent, while the L-M trajectory is scale dependent and is in general not contained in $\Omega$. Points on the L-M trajectory may be easy to compute, but are in general feasible only for large values of $\mu$.

The objective of this work is to show that when $\mu$ grows both trajectories converge to $\chi$, where they share the same tangent. For large values of $\mu$ they behave similarly, which will allow the construction of a procedure to initialize

an interior point path following algorithm for solving (1) from a point on the L-M trajectory.

This procedure uses the assumption that the analytic center $\chi$ is known, and will be applied to the solution of trust region subproblems, which are described in Section 2. In Section 3 we define the primal-dual versions of both trajectories and related concepts. In Section 4 we establish primal-dual relations between the trajectories, and in Section 5 we present our procedure and numerical results.

## 2 Trust region subproblems

Our main motivation comes from the solution of trust region subproblems with the shape
$$\text{minimize}\{m(d) \mid Ad = b, \|d\| \leq \Delta\},$$

where $m(\cdot)$ is convex quadratic and $\Delta > 0$. If the Euclidean norm is used, the minimizers are points on the L-M trajectory, which may be easy to compute. The infinity norm gives larger trust regions and are very convenient whenever box constraints are present in the original problem.

The box-constrained trust region subproblem is stated in the format (1) as follows: consider the convex quadratic program

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}(\bar{z} + d)^\top \bar{Q}(\bar{z} + d) + \bar{c}^\top(\bar{z} + d) \\
\text{subject to} \quad & \bar{A}d = 0 \\
& -u \leq d \leq u,
\end{aligned}
\tag{4}
$$

where $\bar{Q} \in \mathbb{R}^{\bar{n} \times \bar{n}}$ is symmetric and positive semidefinite, $\bar{c} \in \mathbb{R}^{\bar{n}}$, $u \in \mathbb{R}^{\bar{n}}_{++}$ and $\bar{A} \in \mathbb{R}^{\bar{m} \times \bar{n}}$. The logarithmic barrier function for this problem is

$$-u < d < u \mapsto -\sum_{i=1}^{\bar{n}} \log(u_i - d_i) - \sum_{i=1}^{\bar{n}} \log(u_i + d_i),$$

and its minimizer, the analytic center of the feasible set, is $d = 0$.

The following reformulation of Problem (4) is based on the change of variables $z := u + d$ and $w := u - d$.

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}z^\top \bar{Q}z + (\bar{c} + \bar{Q}(\bar{z} - u))^\top z \\
\text{subject to} \quad & \begin{bmatrix} \bar{A} & 0 \\ I & I \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} \bar{A}u \\ 2u \end{bmatrix} \\
& z, w \geq 0.
\end{aligned}
\tag{5}
$$

Defining $U := \text{diag}(u)$ and $x := (U^{-1}z, U^{-1}w) \in \mathbb{R}^{2\bar{n}}$ we rewrite Problem (5) as

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}x^\top Qx + c^\top x \\
\text{subject to} \quad & Ax = b \\
& x \geq 0,
\end{aligned}
\tag{6}
$$

3

where $Q := \begin{bmatrix} U\bar{Q}U & 0 \\ 0 & 0 \end{bmatrix}$, $c := \begin{bmatrix} U(\bar{c} + \bar{Q}(\bar{z} - u)) \\ 0 \end{bmatrix}$, $A := \begin{bmatrix} \bar{A}U & 0 \\ U & U \end{bmatrix}$ and $b := \begin{bmatrix} \bar{A}u \\ 2u \end{bmatrix}$.

One can easily check that (6) inherits the convexity of (4) and that these problems are in an obvious correspondence with respect to optimality, due to our change of variables. Moreover, we have that $e \in \mathbb{R}^n$, with $n := 2\bar{n}$, is the analytic center of the compact polyhedron $\Omega := \{x \in \mathbb{R}^n | Ax = b \text{ and } x \geq 0\}$.

The following straightforward facts on the geometry of the trajectories are shown in Fig. 1:

(i) For $\mu > 0$, $d_{LM}(\mu) = \operatorname{argmin}\{m(d) \,|\, Ad = 0, \|d\| \leq \|d_{LM}(\mu)\|\}$,

i.e., $d_{LM}(\mu)$ is the minimizer of $m(\cdot)$ in an Euclidean trust region around 0. These trust regions are contained in the box for large values of $\mu$, and in general $d_{LM}(\mu)$ is infeasible for small $\mu$.

(ii) For $\mu > 0$, $d_{\log}(\mu) = \operatorname{argmin}\{m(d) \,|\, Ad = 0, p_{\log}(d) \leq p_{\log}(d_{\log}(\mu))\}$,

i.e., $d_{\log}(\mu)$ minimizes $m(\cdot)$ in a log-barrier trust region around 0. These trust regions are feasible and approach the feasible set as $\mu$ decreases, while $d_{\log}(\mu)$ converges to an optimal solution of (4).

Fig. 1 shows the evolution of the trajectories and trust regions for an example with a unit box and non-interior optimal solution. The figures at left and right show respectively the trust regions for large and small values of $\mu$. For large values of $\mu$ the trust regions are very similar, as well as the trajectories. As $\mu$ decreases, the trajectories branch out.

**Remark.** Although this is not a subject of this paper, such log-barrier trust regions may be used in trust region methods for convex programming, avoiding the need for solving (4) precisely.

## 3   Preliminaries

In this section we define the primal-dual versions of the central path and L-M trajectories and neighborhoods of the central path. Consider the Karush-Kuhn-Tucker (KKT) conditions of Problem (1), given by

$$P_{\mathcal{K}(A)}\left(\nabla f(x) - s\right) = 0 \tag{7}$$

$$Ax = b \tag{8}$$

$$x \geq 0 \tag{9}$$

$$s \geq 0 \tag{10}$$

$$x \cdot s = 0. \tag{11}$$

where $P$ is the Euclidean projection operator, $x \cdot s := (x_1 s_1, \ldots, x_n s_n) \in \mathbb{R}^n$ is the Hadamard product and $\mathcal{K}(A) := \{x \in \mathbb{R}^n | Ax = 0\}$. Due to convexity, these conditions are necessary and sufficient for optimality of (1).

We say that a strictly positive pair $(x, s) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ is an *interior feasible primal-dual point* associated to (1), if it satisfies both the Lagrange condition
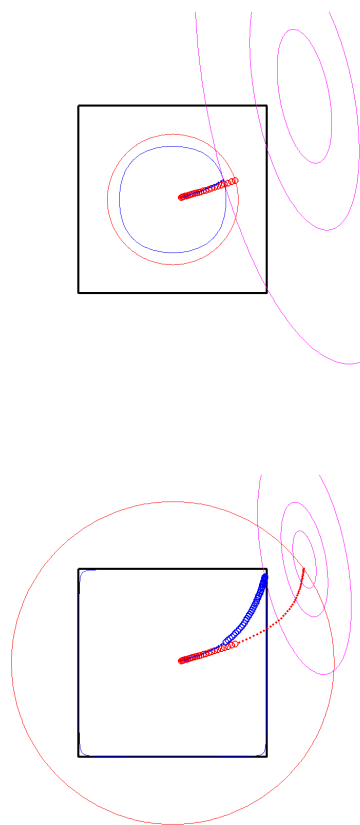
Figure 1: The trajectories.

(7) and (8). We denote by $\Gamma^\circ$ the set of all interior feasible primal-dual points, i.e.,

$$\Gamma^\circ := \{(x,s) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} | Ax = b, \ P_{\mathcal{K}(A)}(\nabla f(x) - s) = 0\}.$$

## 3.1  The primal-dual central path

The primal-dual central path associated with Problem 1 is the curve

$$\mu > 0 \mapsto (x_{CP}(\mu), s_{CP}(\mu)) \in \Gamma^\circ \text{ such that } x_{CP}(\mu) \cdot s_{CP}(\mu) = \mu e.$$

Under our hypotheses, [6] ensures that the central path is well defined, and it is consistent with the former definition of primal central path, as $x_{\log}(\mu) = x_{CP}(\mu)$ for $\mu > 0$. The primal central path $\{x_{CP}(\mu)|\mu > 0\}$ converges to an optimal solution of Problem 1 as $\mu \to 0_+$. For related results under weaker assumption, see [1]. For primal-dual convergence properties of the central path we refer the reader to [11]. In this paper we are interested in the features of the central path close to the analytic center, i.e., when $\mu$ is large.

## 3.2  Neighborhoods of the central path

Sequences generated by path following algorithms ([3, 4, 10]) remain in the proximity of the central path along the iterates, in the sense described by proximity measures. The proximity measure we are going to use in the theoretical treatment is defined by

$$(x,s) \in \Gamma^\circ, \mu > 0 \mapsto \sigma(x,s,\mu) := \left\| \frac{x \cdot s}{\mu} - e \right\|. \tag{12}$$

Clearly, $(x,s) \in \Gamma^\circ$ is the central point associated with $\mu > 0$ if and only if $\sigma(x,s,\mu) = 0$.

Low complexity path following algorithms compute sequences $(x_k, s_k, \mu_k)$ such that $\sigma(x_k, s_k, \mu_k) \leq \rho$, where $\rho$ is a fixed positive tolerance. These points belong to a neighborhood of the central path

$$\mathcal{N}_2(\rho) := \{(x,s,\mu) \in \Gamma^\circ \times R_{++} \mid \sigma(x,s,\mu) \leq \rho\}.$$

Practical algorithms (with a higher proved complexity) use a larger neighborhood, defined by (see [14], Ch. 5)

$$\mathcal{N}_{-\infty}(\rho) := \left\{(x,s,\mu) \in \Gamma^\circ \times R_{++} \mid \frac{x_i s_i}{\mu} \geq 1 - \rho, \ \forall i = 1, \ldots, n\right\}.$$

This neighborhood will be used in our implementation. For a study of other proximity measures, see [5].

The next lemma shows than one can easily get an interior feasible primal-dual point with its primal part being the analytic center.

**Lemma 1** *Let $s_e(\mu) := \nabla f(e) + \mu e$. Then, for all sufficiently large $\mu > 0$ we have $(e, s_e(\mu)) \in \Gamma^\circ$. Moreover, $\sigma(e, s_e(\mu), \mu) = \mu^{-1}\|\nabla f(e)\|$.*

**Proof:** The proof that $(e, s_e(\mu)) \in \Gamma^\circ$ for all large $\mu > 0$ is straightforward. Now, note that

$$\sigma(e, s_e(\mu), \mu) = \left\| \frac{e \cdot (\nabla f(e) + \mu e)}{\mu} - e \right\| = \left\| \frac{\nabla f(e)}{\mu} + e - e \right\| = \mu^{-1} \|\nabla f(e)\|.$$

$\square$

Note that for large values of $\mu$, the point $(e, s_e(\mu))$ is centralized.

### 3.3 The primal-dual L-M trajectory

Consider the primal L-M trajectory $x_{LM}$ defined in (3).

In our construction, it is easy to verify that for a given $\mu > 0$, $x_{LM}(\mu)$ can be obtained solving the problem in (3) with $p_{LM}$ replaced by the penalization $\frac{1}{2}\|\cdot\|^2$.

For convenience, we define the dual L-M trajectory by taking

$$s_{LM}(\mu) := \mu(2\chi - x_{LM}(\mu)), \text{ for each } \mu > 0.$$

This will allow us to relate the primal-dual trajectories in the next section.

## 4 Primal-dual relations between the L-M and the central path trajectories

The first lemma in this section describes properties of the trajectory associated with a general strongly convex penalty function, and will be used to establish primal relations between the central path and the L-M trajectory.

### 4.1 An auxiliary result

For each $\mu > 0$ consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) + \mu p(x) \\ \text{subject to} & Ax = b, \end{array} \tag{13}$$

where $p : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a strongly convex function.

**Lemma 2** *Let $\bar{x}$ be the minimizer of $p$ subject to $Ax = b$ and suppose that $p$ is twice continuously differentiable with a locally Lipschitz Hessian $\nabla^2 p(\cdot)$ at $\bar{x}$. Assume also that $\nabla^2 p(\bar{x}) = I$. Then, considering that $x(\mu)$ denotes the solution of the optimization problem (13), with $\mu > 0$, it follows that:*

(i) *The trajectory $\{x(\mu) \in \mathbb{R}^n | \mu > 0\}$ is well defined and $\lim_{\mu \to \infty} x(\mu) = \bar{x}$;*

(ii) *$x(\mu)$ minimizes $f$ in the convex trust region $\{x \in \mathbb{R}^n | p(x) \leq p(x(\mu)), Ax = b\}$ and $\mu > 0 \mapsto p(x(\mu))$ is a non-increasing function;*

(iii) *$\lim_{\mu \to \infty} \mu (x(\mu) - \bar{x}) = -P_{\mathcal{K}(A)}(\nabla f(\bar{x}))$.*

**Proof:** For any $\mu > 0$, the function $f + \mu p$ is strongly convex. Hence, (13) has the unique solution $x(\mu)$. In particular, $\{x(\mu) \in \mathbb{R}^n | \mu > 0\}$ is well defined. From the definition of $x(\mu)$ and $\bar{x}$ we have, for all $\mu \geq 1$, that

$$f(x(\mu)) + p(x(\mu)) - p(\bar{x}) \leq f(x(\mu)) + \mu \left( p(x(\mu)) - p(\bar{x}) \right)$$
$$\leq f(\bar{x}) + \mu \left( p(\bar{x}) - p(\bar{x}) \right) = f(\bar{x}).$$

This inequality, together with the compactness of the level sets of $f + p$, implies that $\{x(\mu) \in \mathbb{R}^n | \mu \geq 1\}$ is bounded. Consequently, $\{f(x(\mu))\}_{\mu \geq 1}$ must be bounded. Therefore, the second inequality leads to

$$\lim_{\mu \to \infty} \mu(p(x(\mu)) - p(\bar{x})) < \infty.$$

Hence,

$$\lim_{\mu \to \infty} p(x(\mu)) - p(\bar{x}) = 0.$$

Then, the strong convexity of $p$ guarantees that

$$\lim_{\mu \to \infty} x(\mu) = \bar{x},$$

proving (i).

The Karush-Kuhn-Tucker conditions of (13), necessary and sufficient due to convexity, read as follows

$$P_{\mathcal{K}(A)} \left( \nabla f(x(\mu)) + \mu \nabla p(x(\mu)) \right) = 0; \tag{14}$$

$$A \left( x(\mu) \right) = b. \tag{15}$$

From (14) and (15) we conclude that $x(\mu)$ minimizes $f$ in the convex set $\{x \in \mathbb{R}^n | p(x) \leq p(x(\mu)), Ax = b\}$. Moreover, the Lagrange multiplier correspondent to the inequality constraint is $\mu$. Now take $0 < \mu_2 < \mu_1$. Then,

$$f \left( x(\mu_2) \right) + \mu_2 p \left( x(\mu_2) \right) \leq f \left( x(\mu_1) \right) + \mu_2 p \left( x(\mu_1) \right)$$

$$f \left( x(\mu_1) \right) + \mu_1 p \left( x(\mu_1) \right) \leq f \left( x(\mu_2) \right) + \mu_1 p \left( x(\mu_2) \right).$$

Adding these inequalities we get $p \left( x(\mu_1) \right) \leq p \left( x(\mu_2) \right)$, proving (ii).

Note that if for some $\bar{\mu} > 0$ we have $x(\bar{\mu}) = \bar{x}$, the strong convexity of $p$ and (ii) guarantee that $x(\mu) = \bar{x}$ for every $\mu > \bar{\mu}$. In this case, (14) implies (iii). Therefore, let us assume from now on that $x(\mu) \neq \bar{x}$ for all $\mu > 0$. From Taylor's formula we have that

$$p(x) - p(\bar{x}) - \nabla p(\bar{x})^\top (x - \bar{x}) - \frac{1}{2}(x - \bar{x})^\top \nabla^2 p(\bar{x})(x - \bar{x}) = o(\|x - \bar{x}\|^2), \tag{16}$$

with

$$\lim_{x \to \bar{x}} \frac{o(\|x - \bar{x}\|^2)}{\|x - \bar{x}\|^2} = 0.$$

8

Moreover, from the local Lipschitz continuity of $\nabla^2 p$ around $\bar{x}$, we know that there exist $L > 0$ and $\delta > 0$, so that for all $x \in \mathcal{B}(\bar{x}, \delta)$ it holds that

$$\|v(x)\| \leq L \|x - \bar{x}\|^2, \tag{17}$$

where

$$v(x) := \nabla p(x) - \nabla p(\bar{x}) - \underbrace{\nabla^2 p(\bar{x})}_{I}(x - \bar{x}).$$

Since $\bar{x}$ is the minimizer of $p$ subject to $Ax = b$, we conclude that $\nabla p(\bar{x})$ is orthogonal to $\mathcal{K}(A)$. Then, taking also (16) and $\nabla^2 p(\bar{x}) = I$ into account, we get

$$p(x(\mu)) - p(\bar{x}) = \|x(\mu) - \bar{x}\|^2 \left( \frac{1}{2} + \frac{o(\|x(\mu) - \bar{x}\|^2)}{\|x(\mu) - \bar{x}\|^2} \right), \tag{18}$$

for all $\mu$ sufficiently large. Due to (15) and the feasibility of $\bar{x}$, it follows that $x(\mu) - \bar{x} \in \mathcal{K}(A)$, for all $\mu > 0$. Then, using the linearity of $P_{\mathcal{K}(A)}$ and (14) we obtain

$$
\begin{aligned}
0 &= P_{\mathcal{K}(A)}\left(\nabla f(x(\mu)) + \mu \nabla p(x(\mu))\right) \\
&= P_{\mathcal{K}(A)}\left(\nabla f(x(\mu)) + \mu \nabla p(\bar{x}) + \mu(x(\mu) - \bar{x}) + \mu v(x(\mu))\right) \\
&= P_{\mathcal{K}(A)}(\nabla f(x(\mu))) + \mu P_{\mathcal{K}(A)}(\nabla p(\bar{x})) + \mu P_{\mathcal{K}(A)}(x(\mu) - \bar{x}) + P_{\mathcal{K}(A)}(\mu v(x(\mu))) \\
&= P_{\mathcal{K}(A)}(\nabla f(x(\mu))) + \mu(x(\mu) - \bar{x}) + P_{\mathcal{K}(A)}(\mu v(x(\mu))). \tag{19}
\end{aligned}
$$

In order to finish the proof we will show that $\lim_{\mu \to \infty} \mu v(x(\mu)) = 0$. Note that the definition of $x(\mu)$ implies that

$$0 \leq \mu(p(x(\mu)) - p(\bar{x})) \leq f(\bar{x}) - f(x(\mu)).$$

Using (i) we get

$$\lim_{\mu \to \infty} \mu(p(x(\mu)) - p(\bar{x})) = 0. \tag{20}$$

Now, multiplying (18) by $\mu$ and taking limits we obtain

$$\lim_{\mu \to \infty} \mu \|x(\mu) - \bar{x}\|^2 = 0. \tag{21}$$

Using (17) we conclude that $\lim_{\mu \to \infty} \mu v(x(\mu)) = 0$. This fact, combined with (19), implies (iii). $\qquad\square$

One can trivially check that $p_{\log}$ and $p_{LM}$ satisfy the hypotheses of Lemma 3.1 with $\bar{x} = \chi = e$. Note also that $P_{\mathcal{K}(A)}(e) = 0$.

Hence, item (ii) of Lemma 3.1 implies that $x_{LM}(\mu)$ and $x_{CP}(\mu)$ minimize $f$ subject to $Ax = b$ in the Euclidean ball centered in $\chi = e$ with radius $\|x_{LM}(\mu)\|$ and in the level set $\{x \in \mathbb{R}^n | p_{\log}(x) \leq p_{\log}(x_{CP}(\mu))\}$, respectively.

## 4.2   Our theorem

**Theorem 1** *It holds that:*

(i) $P_{\mathcal{K}(A)}\left(\nabla f(x_{LM}(\mu)) - s_{LM}(\mu)\right) = 0$ *for all $\mu > 0$;*

(ii) $\sigma(x_{LM}(\mu), s_{LM}(\mu), \mu) \le \|x_{LM}(\mu) - \chi\|^2$, *where $\sigma$ is the proximity measure given in* (12);

(iii) *the L-M trajectory and the central path associated with* (1) *are tangent at $\chi = e$;*

(iv) *for all $\mu > 0$ sufficiently large, $(x_{LM}(\mu), s_{LM}(\mu)) \in \mathbb{R}^n \times \mathbb{R}^n$ is an interior feasible primal-dual point associated with* (1).

**Proof:** From the linearity of $P_{\mathcal{K}(A)}$ and the definition of $x_{LM}(\mu)$ and $s_{LM}(\mu)$, we conclude that

$$
\begin{aligned}
P_{\mathcal{K}(A)}\left(\nabla f(x_{LM}(\mu)) - s_{LM}(\mu)\right) &= P_{\mathcal{K}(A)}\left(\nabla f(x_{LM}(\mu)) + \mu x_{LM}(\mu) - 2\mu\chi\right) \\
&= P_{\mathcal{K}(A)}\left(\nabla f(x_{LM}(\mu)) + \mu x_{LM}(\mu)\right) - 2\mu P_{\mathcal{K}(A)}(\chi) \\
&= 0 - 2\mu P_{\mathcal{K}(A)}(e) = 0.
\end{aligned}
$$

This proves (i). Let us now prove (ii).

Considering the definition of the proximity measure $\sigma$ given in (12) and remembering that $\chi = e$, we conclude that

$$
\begin{aligned}
\sigma(x_{LM}(\mu), s_{LM}(\mu), \mu) &:= \left\| \frac{x_{LM}(\mu) \cdot s_{LM}(\mu)}{\mu} - e \right\| \\
&= \| x_{LM}(\mu) \cdot (2\chi - x_{LM}(\mu)) - e \| \\
&= \| (\chi + (x_{LM}(\mu) - \chi)) \cdot (\chi - \\
& \quad (x_{LM}(\mu) - \chi)) - e \| \\
&= \| (x_{LM}(\mu) - \chi) \cdot (x_{LM}(\mu) - \chi) \| \\
&\le \| x_{LM}(\mu) - \chi \|^2.
\end{aligned}
$$

Taking into account that $p_{\log}$ and $p_{LM}$ satisfy the hypotheses of Lemma 3.1, item (iii) follows directly from item (iii) in Lemma 2.

Due to item (i), and the fact that $Ax_{LM}(\mu) = b$, we only need to check the positivity of $x_{LM}(\mu)$ and $s_{LM}(\mu)$ in order to prove (iv). But this follows for sufficiently large $\mu > 0$, since $\chi = e > 0$ and $\lim_{\mu \to \infty} x_{LM}(\mu) = \chi$.  □

Our theorem says basically that the primal-dual L-M trajectory $\{(x_{LM}(\mu), s_{LM}(\mu)) | \mu > 0\}$ is given by interior primal-dual feasible points associated with Problem (1) (for all $\mu$ sufficiently large), that lie close to the primal-dual central path in terms of the proximity measure $\sigma$. Furthermore, the primal L-M trajectory is tangent to the central path at the analytic center $\chi = e$. We established these results for a general convex continuously differentiable function $f$.

# 5    Interior point scheme

We now present a scheme for solving Problem (1) composed of two phases: phase 1 follows the primal-dual L-M trajectory until a point near the central path is found; an interior point path following method is then started at this point. Phase 1 will stop when a point in a neighborhood $\mathcal{N}(\rho)$ is found, where $\rho$ is a proximity constant, usually $\rho < 1$ and $\mathcal{N} = \mathcal{N}_2$ or $\mathcal{N} = \mathcal{N}_{-\infty}$. In the algorithm below we start following the L-M trajectory at a given value $\mu_0$ and increase it.

---

**Phase 1: Primal-dual L-M path following**
Set $k := 0$ and choose $\mu_0 > 0$, $\rho > 0$ and $\beta > 1$;
REPEAT
   Compute the solution $x_{LM}(\mu)$ of (3) with $\mu := \mu_k$ and define
   $s_{LM}(\mu_k) := \mu_k(2\underbrace{\chi}_{e} - x_{LM}(\mu_k))$;

   IF $(x_{LM}(\mu_k), s_{LM}(\mu_k)) \notin \mathcal{N}(\rho)$, set $\mu_{k+1} := \beta\mu_k$, $k = k+1$.
   ELSE set $(\hat{x}, \hat{s}) = (x_{LM}(\mu_k), s_{LM}(\mu_k))$ and STOP.
   **Phase 2: Path following scheme**
   Choose and implement a primal-dual interior point method taking $(\hat{x}, \hat{s})$ as the initial point.

---

Due to Theorem 1 this scheme is well defined and may be applied to any problem in the format (1).

## 5.1    Our theorem applied to convex quadratic programming

Theorem 1 is specially interesting when $f$ is quadratic. In this case, one can calculate points on the L-M trajectory by solving linear systems of equations. Indeed, if $f : \mathbb{R}^n \to \mathbb{R}$ is a convex quadratic function given by

$$f(x) := x^\top Q x + c^\top x,$$

where $c \in \mathbb{R}^n$ and $Q \in \mathbb{R}^{n \times n}$ is positive semidefinite, then the Karush-Kuhn-Tucker conditions of Problem (3), that define the L-M trajectory, read as follows

$$\begin{bmatrix} Q + \mu I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}. \tag{22}$$

For each $\mu > 0$ this system is always solvable and its unique solution is $x_{LM}(\mu)$. Uniqueness in $\lambda$ is ensured if, and only if, $A$ is full rank.

So, in the quadratic case, Phase 1 turns out to be easily implementable, since the primal-dual L-M trajectory is defined by means of the linear system (22).

# 6    Numerical experiments

We generate instances of Problem (4) for different values of $\bar{n}$ and $\bar{m}$, and solve them by an infeasible path following interior point algorithm using three different primal-dual starting points: the point given by our scheme, the Mehrotra starting point [9], and the analytic center $\chi = e$ with a dual variable $s_e(\mu)$ given by Lemma 1 for a large value of $\mu$. Since the analytic center is known, we shall see that the third choice is always better than the second.

Each test problem with given $\bar{n}$ and $\bar{m}$ is constructed as follows:
$\bar{A}$ is an $m \times n$ matrix with random entries in $[-1, 1]$.
$\bar{Q}$ is obtained by the following procedure: let $D$ be a diagonal matrix with random entries in $[1, 1000]$, and $M$ a unitary random matrix (constructed using the Matlab routine qr). Now $\bar{Q} = MDM^\top$ is a randomly generated hessian matrix with eigenvalues diag($D$).
The objective function is given by $(x - x^*)^\top \bar{Q}(x - x^*)/2$, where the global minimizer $x^*$ is obtained by projecting a random vector into the null space of $\bar{A}$ and then adjusting it so that $\|x^*\|_\infty = 1.2$. This choice is done to stress the nonlinearity in the examples: if $x^*$ is in the unit box, the minimization becomes too easy; if its norm is large, the L-M trajectory becomes nearly straight, again too easy.

We do not elaborate on the search along the L-M trajectory, studied in non linear programming texts like [12]. For our series of examples we used the simple strategy of setting $\mu_0 = \hat{\mu}/2$, where $\hat{\mu}$ is the parameter value obtained by the previous example ($\mu_0 = 1$ for the first example in the series). This is a usual scheme in trust region algorithms for non-linear programming.

We used Matlab R2012a to implement the infeasible path-following method for convex quadratic programming presented in [2]. The tolerance for optimality is fixed at $10^{-8}$ and the barrier parameter is reduced by a factor of 10 at each iteration. As a measure of performance we used the number of iterations necessary for convergence.

Table 1 lists the results for 56 problems: for each value of $\bar{n}$ in the first column and four values of $\bar{m}$, we run the three strategies. The column LM displays the number of iterations Phase 1/Phase 2 of our scheme; columns M and AC correspond to the number of iterations of the interior point method using respectively the Mehrotra starting point and the analytic center.

From the table we see that for this choice of quadratic objectives, our scheme spares about 3 iterations from the interior point algorithm. This may be very significant in trust region algorithms, in which there is no need for a high precision in the solution of Problem (4).

# 7    Concluding remarks

The main goal of this paper is to show how the primal-dual L-M and central trajectories are related for general linearly constrained convex programming problems, and then to develop an initialization technique for path following

| $\bar{n}$ | $\bar{m} = \bar{n}/1.5$ | | | $\bar{m} = \bar{n}/2$ | | | $\bar{m} = \bar{n}/4$ | | | $\bar{m} = 0$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LM | M | AC | LM | M | AC | LM | M | AC | LM | M | AC |
| 50 | 7/8 | 8 | 8 | 5/8 | 12 | 11 | 4/7 | 11 | 10 | 0/9 | 11 | 11 |
| 100 | 2/6 | 10 | 9 | 1/8 | 11 | 11 | 1/8 | 11 | 11 | 3/9 | 12 | 12 |
| 150 | 1/6 | 10 | 10 | 1/6 | 10 | 9 | 3/8 | 11 | 10 | 1/8 | 12 | 11 |
| 200 | 3/7 | 10 | 10 | 1/6 | 9 | 9 | 1/8 | 11 | 11 | 2/9 | 12 | 11 |
| 250 | 1/7 | 10 | 10 | 4/6 | 9 | 9 | 1/7 | 11 | 11 | 1/9 | 11 | 11 |
| 300 | 1/5 | 9 | 9 | 2/6 | 9 | 9 | 3/8 | 11 | 11 | 1/7 | 11 | 11 |
| 400 | 2/5 | 9 | 9 | 1/6 | 8 | 7 | 3/8 | 11 | 10 | 4/9 | 11 | 11 |
| 500 | 1/5 | 10 | 9 | 1/5 | 8 | 8 | 2/8 | 11 | 11 | 1/9 | 12 | 11 |
| 600 | 3/5 | 9 | 9 | 4/5 | 8 | 8 | 2/8 | 10 | 10 | 2/8 | 11 | 11 |
| 700 | 3/6 | 9 | 9 | 1/5 | 8 | 8 | 2/8 | 11 | 11 | 2/8 | 11 | 11 |
| 800 | 1/6 | 9 | 9 | 2/5 | 8 | 8 | 2/9 | 11 | 11 | 1/7 | 11 | 10 |
| 900 | 3/5 | 9 | 8 | 2/5 | 8 | 8 | 2/8 | 11 | 10 | 3/8 | 11 | 11 |
| 1000 | 1/5 | 9 | 9 | 2/5 | 8 | 8 | 2/8 | 11 | 11 | 2/8 | 11 | 11 |

Table 1: Number of iterations for the interior point method using the L-M starting point, the Mehrotra starting point and the analytic center.

methods based on this relationship.

We believe that the main application of this is in trust region methods for non-linear programming based on trust regions defined by level sets of the barrier penalty function, which are contained in the unit box, as we remarked in Section 2.

## Acknowledgements.

## References

[1] J.C. Gilbert, C.C. Gonzaga, E. Karas, *Examples of ill-behaved central paths in convex optimization*, Math Program., Ser. A, (103) 2005, 63–94.

[2] J. Gondzio, *Interior point methods 25 years later*, European Journal of Operational Research, (218) 2012, 587-601.

[3] C.C. Gonzaga, *Path following methods for linear programming*, SIAM Review, (34)(2) 1992, 167–227.

[4] C.C. Gonzaga, F. Bonnans, *Fast convergence of the simplified largest step path following algorithm*, Math Program., **76**, 95–116 (1997).

[5] M.D. Gonzalez-Lima, C. Roos, On Central-Path Proximity Measures in Interior-Point Methods, Journal of Optimization Theory and Applications, (127) 2005, 303-328.

[6] L.M. Graña Drummond, B.F. Svaiter, *On Well Definedness of the Central Path*, Journal of Optimization Theory and Applications, (102) 1999, 223–237.

[7] K. Levenberg, *A method for the solution of certain non-linear problems in least squares*, Q. Appl. Math. (2) 1944, 164–168.

[8] D.W. Marquardt, *An algorithm for least-squares estimation of nonlinear parameters*, J. Soc. Ind. Appl. Math. (11) 1963, 431–441.

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

[10] R.D.C. Monteiro, I. Adler, *Interior Path-Following Primal-Dual Algorithms, Part 2: Convex Quadratic Programming*, Math Program., (44) 1989, 43–66.

[11] R.D.C. Monteiro, F. Zhou, *On the Existence and Convergence of the Central Path for Convex Programming and Some Duality Results*, Computational Optimization and Applications, (10) 1998, 51–77.

[12] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer Verlag, 1999.

[13] G. Sonnevend, *An Analytical Centre Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming*, in Lecture Notes Control Inform. Sci. (84), Springer-Verlag, New York, NY, 1985, 866–876.

[14] S.J. Wright, *Primal-Dual Interior Point Methods*, Philadelphia, SIAM, 1997.