

Primal-Dual Relationship Between Levenberg–Marquardt and Central Trajectories for Linearly Constrained Convex Optimization

Roger Behling · Clovis Gonzaga · Gabriel Haeser

Received: 30 July 2013 / Accepted: 19 November 2013 / Published online: 28 November 2013
© Springer Science+Business Media New York 2013

Abstract We consider the minimization of a convex function on a bounded polyhedron (polytope) represented by linear equality constraints and non-negative variables. We define the Levenberg–Marquardt and central trajectories starting at the analytic center 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 Levenberg–Marquardt trajectory and eventually moves to the central path. Our main theorem is particularly relevant in quadratic programming, where points on the primal-dual Levenberg–Marquardt 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

Communicated by Alfredo N. Iusem.

R. Behling
Católica SC, Jaraguá do Sul-SC, Brazil
e-mail: rogerbehling@gmail.com

C. Gonzaga
Federal University of Santa Catarina, Florianópolis-SC, Brazil
e-mail: clovis@mtm.ufsc.br

G. Haeser (✉)
Institute of Science and Technology, Federal University of São Paulo, São José dos Campos-SP,
Brazil
e-mail: gabriel.haeser@unifesp.br

1 Introduction

In this paper, we describe a path following algorithm based both on the central path and on the Levenberg–Marquardt (LM) trajectories for solving the problem of minimizing a convex and continuously differentiable function on a bounded polyhedron with nonempty algebraic interior.

The objective of this work is to show that, near the analytic center, the trajectories share the same tangent, which will allow one the construction of a procedure to initialize an interior point path following algorithm for solving the problem from a point on the LM trajectory. This procedure assumes that the analytic center is known, which is the case in trust region subproblems.

In Sect. 2, we define both trajectories. In Sect. 3, we describe trust region subproblems, in which we apply our procedure. In Sects. 4 and 5, we establish primal-dual relations between the trajectories and, in Sects. 6 and 7, we present our procedure and numerical results. In Sect. 8, we give some conclusions.

2 The Trajectories

Consider the problem

$$\text{minimize } f(x) \quad \text{s.t. } Ax = b, x \geq 0, \quad (1)$$

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

We assume that Ω 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 Ω , $\chi = \operatorname{argmin}\{p_{\log}(x) | x \in \Omega^0\}$ (see [1]).

To simplify the treatment, we assume that

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

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

$$\text{minimize } f_0(w) \quad \text{s.t. } A_0 w = b, w \geq 0, \quad (2)$$

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 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 LM regularization

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

which defines the primal LM trajectory

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

The idea of this quadratic penalization goes back to [2, 3].

The central path is contained in Ω^0 and is scale-independent, while the LM trajectory is scale-dependent and is in general not contained in Ω . Points on the LM trajectory may be easy to compute, but are in general feasible only for large values of μ .

3 Trust Region Subproblems

Our main motivation comes from the solution of trust region subproblems with the shape

$$\operatorname{minimize} \{m(d) \mid Ad = b, \|d\| \leq \Delta\},$$

where $m(\cdot)$ is convex quadratic and $\Delta > 0$. If the Euclidean norm is used, the minimizers are points on the LM trajectory, which may be easy to compute. The infinity norm gives larger trust regions, which 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

$$\begin{aligned} &\operatorname{minimize} \quad \left[\frac{1}{2}(\bar{z} + d)^\top \bar{Q}(\bar{z} + d) + \bar{c}^\top(\bar{z} + d) \right] \\ &\text{s.t.} \quad \bar{A}d = 0, -u \leq d \leq u, \end{aligned} \tag{4}$$

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

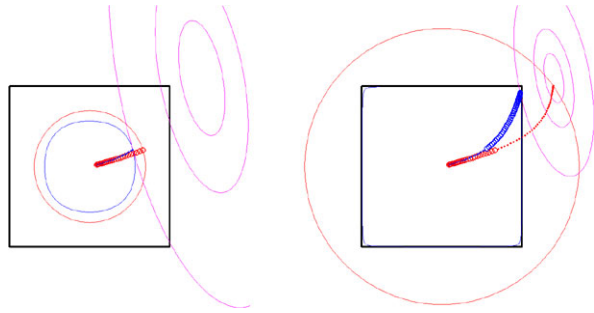
$$-u < d < u \mapsto -\sum_{i=1}^{\bar{n}} \log(u_i - d_i) - \sum_{i=1}^{\bar{n}} \log(u_i + d_i),$$

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 $w := u - d$:

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

Fig. 1 The trajectories



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

$$\begin{aligned} & \text{minimize} && \left[\frac{1}{2}x^\top Qx + c^\top x \right] \\ & \text{s.t.} && Ax = b, x \geq 0, \end{aligned} \tag{6}$$

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

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 $\Omega := \{x \in \mathbb{R}^n \mid Ax = b \text{ and } x \geq 0\}$.

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

- (i) For $\mu > 0$, $d_{LM}(\mu) = \text{argmin}\{m(d) \mid Ad = 0, \|d\| \leq \|d_{LM}(\mu)\|\}$, 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 μ , and in general $d_{LM}(\mu)$ is infeasible for small μ .
- (ii) For $\mu > 0$, $d_{\log}(\mu) = \text{argmin}\{m(d) \mid Ad = 0, p_{\log}(d) \leq p_{\log}(d_{\log}(\mu))\}$, 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 μ decreases, while $d_{\log}(\mu)$ converges to an optimal solution of (4).

Figure 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 μ . For large values of μ the trust regions are very similar, as well as the trajectories. As μ decreases, the trajectories branch out.

Remark 3.1 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.

4 Preliminaries

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

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

$$Ax = b, \quad x \geq 0, \quad s \geq 0 \tag{8}$$

$$x \cdot s = 0, \tag{9}$$

where P is the Euclidean projection operator, $x \cdot s := (x_1s_1, \dots, x_ns_n) \in \mathbb{R}^n$ 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 with (1), if it satisfies both the Lagrange condition (7) and (8). We denote by Γ° the set of all interior feasible primal-dual points, i.e.,

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

4.1 The Primal-Dual Central Path

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

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

Under our hypotheses, [4] 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 $\mu \rightarrow 0_+$. For related results under weaker assumption, see [5]. For primal-dual convergence properties of the central path we refer the reader to [6]. In this paper we are interested in the features of the central path close to the analytic center, i.e., when μ is large.

4.2 Neighborhoods of the Central Path

Sequences generated by path following algorithms [7–9] 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

$$(x, s) \in \Gamma^\circ, \mu > 0 \mapsto \sigma(x, s, \mu) := \left\| \frac{x \cdot s}{\mu} - e \right\|. \tag{10}$$

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, μ_k) such that $\sigma(x_k, s_k, \mu_k) \leq \rho$, where ρ is a fixed positive tolerance. These points belong to a neighborhood of the central path

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

Practical algorithms (with a higher proved complexity) use a larger neighborhood, defined by (see [10, Chap. 5]):

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

This neighborhood will be used in our implementation. For a study of other proximity measures, see [11].

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

Lemma 4.1 *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

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

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

4.3 The Primal-Dual LM Trajectory

Consider the primal LM 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 LM trajectory by taking

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

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

5 Primal-Dual Relations Between the LM 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 LM trajectory.

5.1 An Auxiliary Result

For each $\mu > 0$ consider the problem

$$\text{minimize } [f(x) + \mu p(x)], \quad \text{s.t. } Ax = b, \tag{11}$$

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

Lemma 5.1 *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 (11), with $\mu > 0$, it follows that:*

- (i) *The trajectory $\{x(\mu) \in \mathbb{R}^n \mid \mu > 0\}$ is well defined and $\lim_{\mu \rightarrow \infty} x(\mu) = \bar{x}$;*
- (ii) *$x(\mu)$ minimizes f in the convex trust region $\{x \in \mathbb{R}^n \mid p(x) \leq p(x(\mu)), Ax = b\}$ and $\mu > 0 \mapsto p(x(\mu))$ is a non-increasing function;*
- (iii) *$\lim_{\mu \rightarrow \infty} \mu(x(\mu) - \bar{x}) = -P_{\mathcal{K}(A)}(\nabla f(\bar{x}))$.*

Proof For any $\mu > 0$, the function $f + \mu p$ is strongly convex. Hence, (11) has the unique solution $x(\mu)$. In particular, $\{x(\mu) \in \mathbb{R}^n \mid \mu > 0\}$ is well defined. From the definition of $x(\mu)$ and \bar{x} we have, for all $\mu \geq 1$, that

$$\begin{aligned} f(x(\mu)) + p(x(\mu)) - p(\bar{x}) &\leq f(x(\mu)) + \mu(p(x(\mu)) - p(\bar{x})) \\ &\leq f(\bar{x}) + \mu(p(\bar{x}) - p(\bar{x})) = f(\bar{x}). \end{aligned}$$

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

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

Hence,

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

Then, the strong convexity of p guarantees that

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

proving (i).

The KKT conditions of (11), necessary and sufficient due to convexity, read as follows:

$$P_{\mathcal{K}(A)}(\nabla f(x(\mu)) + \mu \nabla p(x(\mu))) = 0; \tag{12}$$

$$A(x(\mu)) = b. \tag{13}$$

From (12) and (13) we conclude that $x(\mu)$ minimizes f in the convex set $\{x \in \mathbb{R}^n \mid p(x) \leq p(x(\mu)), Ax = b\}$. Moreover, the Lagrange multiplier correspondent to the inequality constraint is μ . Now take $0 < \mu_2 < \mu_1$. Then,

$$\begin{aligned} f(x(\mu_2)) + \mu_2 p(x(\mu_2)) &\leq f(x(\mu_1)) + \mu_2 p(x(\mu_1)), \\ f(x(\mu_1)) + \mu_1 p(x(\mu_1)) &\leq f(x(\mu_2)) + \mu_1 p(x(\mu_2)). \end{aligned}$$

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

Note that, if for some $\bar{\mu} > 0$ we have $x(\bar{\mu}) = \bar{x}$, then the strong convexity of p and (ii) guarantee that $x(\mu) = \bar{x}$ for every $\mu > \bar{\mu}$. In this case, (12) 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

$$p(x) - p(\bar{x}) - \nabla p(\bar{x})^\top (x - \bar{x}) - \frac{1}{2}(x - \bar{x})^\top \nabla^2 p(\bar{x})(x - \bar{x}) = o(\|x - \bar{x}\|^2), \tag{14}$$

with

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

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

$$\|v(x)\| \leq L\|x - \bar{x}\|^2, \tag{15}$$

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 (14) and $\nabla^2 p(\bar{x}) = I$ into account, we get

$$p(x(\mu)) - p(\bar{x}) = \|x(\mu) - \bar{x}\|^2 \left(\frac{1}{2} + \frac{o(\|x(\mu) - \bar{x}\|^2)}{\|x(\mu) - \bar{x}\|^2} \right) \tag{16}$$

for all μ sufficiently large. Due to (13) 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 (12) we obtain

$$\begin{aligned} 0 &= P_{\mathcal{K}(A)}(\nabla f(x(\mu)) + \mu \nabla p(x(\mu))) \\ &= P_{\mathcal{K}(A)}(\nabla f(x(\mu)) + \mu \nabla p(\bar{x}) + \mu(x(\mu) - \bar{x}) + \mu v(x(\mu))) \\ &= P_{\mathcal{K}(A)}(\nabla f(x(\mu))) + \mu P_{\mathcal{K}(A)}(\nabla p(\bar{x})) + \mu P_{\mathcal{K}(A)}(x(\mu) - \bar{x}) \\ &\quad + P_{\mathcal{K}(A)}(\mu v(x(\mu))) \\ &= P_{\mathcal{K}(A)}(\nabla f(x(\mu))) + \mu(x(\mu) - \bar{x}) + P_{\mathcal{K}(A)}(\mu v(x(\mu))). \end{aligned} \tag{17}$$

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

$$\lim_{\mu \rightarrow \infty} \mu(p(x(\mu)) - p(\bar{x})) = 0. \tag{18}$$

Now, multiplying (16) by μ and taking limits, we obtain

$$\lim_{\mu \rightarrow \infty} \mu \|x(\mu) - \bar{x}\|^2 = 0. \tag{19}$$

Using (15) we conclude that $\lim_{\mu \rightarrow \infty} \mu v(x(\mu)) = 0$. This fact, combined with (17), implies (iii). □

One can trivially check that p_{\log} and p_{LM} satisfy the hypotheses of Lemma 5.1 with $\bar{x} = \chi = e$. Note also that $P_{\mathcal{K}(A)}(e) = 0$.

Hence, item (ii) of Lemma 5.1 implies that $x_{LM}(\mu)$ and $x_{CP}(\mu)$ minimize f subject to $Ax = b$ in the Euclidean ball centered at $\chi = e$ with radius $\|x_{LM}(\mu)\|$ and in the level set $\{x \in \mathbb{R}^n \mid p_{\log}(x) \leq p_{\log}(x_{CP}(\mu))\}$, respectively.

5.2 Our Theorem

Theorem 5.1 *It holds that:*

- (i) $P_{\mathcal{K}(A)}(\nabla f(x_{LM}(\mu)) - s_{LM}(\mu)) = 0$ for all $\mu > 0$;
- (ii) $\sigma(x_{LM}(\mu), s_{LM}(\mu), \mu) \leq \|x_{LM}(\mu) - \chi\|^2$, where σ is the proximity measure given in (10);
- (iii) the LM 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

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

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

Considering the definition of the proximity measure σ given in (10) and remembering that $\chi = e$, we conclude that

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

Taking into account that p_{\log} and p_{LM} satisfy the hypotheses of Lemma 5.1, item (iii) follows directly from item (iii) in Lemma 5.1.

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 $\lim_{\mu \rightarrow \infty} x_{LM}(\mu) = \chi$. □

Our theorem says basically that the primal-dual LM trajectory $\{(x_{LM}(\mu), s_{LM}(\mu)) \mid \mu > 0\}$ is given by interior primal-dual feasible points associated with Problem (1) (for all μ sufficiently large) that lie close to the primal-dual central path in terms of the proximity measure σ . Furthermore, the primal LM 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 .

6 Interior Point Scheme

We now present a scheme for solving Problem (1) composed of two phases: Phase 1 follows the primal-dual LM 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 ρ 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 LM trajectory at a given value μ_0 and increase it.

Phase 1: Primal-Dual LM 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 5.1 this scheme is well defined and may be applied to any problem in the format (1).

6.1 Our Theorem Applied to Convex Quadratic Programming

Theorem 5.1 is specially interesting when f is quadratic. In this case, one can calculate points on the LM trajectory by solving linear systems of equations. Indeed, if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex quadratic function given by

$$f(x) := \frac{1}{2}x^\top Qx + c^\top x$$

where $c \in \mathbb{R}^n$ and $Q \in \mathbb{R}^{n \times n}$ is positive semidefinite, then the KKT conditions of Problem (3) that define the LM trajectory read as follows:

$$\begin{bmatrix} Q + \mu I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}. \tag{20}$$

For each $\mu > 0$ this system is always solvable and its unique solution is $x_{LM}(\mu)$. Uniqueness in λ 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 LM trajectory is defined by means of the linear system (20).

7 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 [12], and the analytic center $\chi = e$ with a dual variable $s_e(\mu)$ given by Lemma 4.1 for a large value of μ . 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 $\bar{Q} = MDM^T$ is a randomly generated Hessian matrix with eigenvalues $\text{diag}(D)$.

The objective function is given by $(x - x^*)^T \bar{Q}(x - x^*)/2$, 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 LM trajectory becomes nearly straight, again too easy.

We do not elaborate on the search along the LM trajectory, studied in non-linear programming texts like [13]. 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 [14]. 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).

Table 1 Number of iterations for the interior point method using the LM starting point, the Mehrotra starting point and the analytic center

\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

8 Conclusions

The main goal of this paper is to show how the primal-dual LM and central trajectories are related for general linearly constrained convex programming problems, and then to develop an initialization technique for path following 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 Sect. 2.

Initial numerical experiments indicate that our initialization technique spares about 3 iterations from the interior point algorithm. More robust numerical experiments should be carried out on a larger collection of relevant problems where the analytic center is known. In particular, the algorithm should be tested on trust region subproblems from general non-linear solvers.

Acknowledgements This work is supported by CNPq and FAPESP (Grant 2010/19720-5) from Brazil. The authors would like to thank Professor Yinyu Ye, Luiz Rafael dos Santos and the referees for their contributions to this work.

References

1. Sonnevend, A.: An Analytical Centre Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming. Lecture Notes Control Inform. Sci., vol. 84, pp. 866–876. Springer, New York (1985)
2. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. **2**, 164–168 (1944)
3. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. **11**, 431–441 (1963)

4. Graña Drummond, L.M., Svaiter, B.F.: On well definedness of the central path. *J. Optim. Theory Appl.* **102**, 223–237 (1999)
5. Gilbert, J.C., Gonzaga, C.C., Karas, E.W.: Examples of ill-behaved central paths in convex optimization. *Math. Program.* **103**, 63–94 (2005)
6. Monteiro, R.D.C., Zhou, F.: On the existence and convergence of the central path for convex programming and some duality results. *Comput. Optim. Appl.* **10**, 51–77 (1998)
7. Gonzaga, C.C.: Path following methods for linear programming. *SIAM Rev.* **34**(2), 167–227 (1992)
8. Gonzaga, C.C., Bonnans, F.: Fast convergence of the simplified largest step path following algorithm. *Math. Program.* **76**, 95–116 (1997)
9. Monteiro, R.D.C., Adler, I.: Interior path-following primal-dual algorithms, part 2: convex quadratic programming. *Math. Program.* **44**, 43–66 (1989)
10. Wright, S.J.: *Primal-Dual Interior Point Methods*. SIAM, Philadelphia (1997)
11. Gonzalez-Lima, M.D., Roos, C.: On central-path proximity measures in interior-point methods. *J. Optim. Theory Appl.* **127**, 303–328 (2005)
12. Mehrotra, S.: On the implementation of a primal-dual interior point method. *SIAM J. Optim.* **2–4**, 575–601 (1992)
13. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer Series in Operations Research. Springer, Berlin (1999)
14. Gondzio, J.: Interior point methods 25 years later. *Eur. J. Oper. Res.* **218**, 587–601 (2012)