Infinite-Dimensional Quadratic Optimization: Interior-Point Methods and Control Applications

L. Faybusovich and J. B. Moore

1Department of Mathematics, University of Notre Dame, Mail Distribution Center, Notre Dame, IN 46556, USA
2Department of Systems Engineering and Cooperative Research Centre for Robust and Adaptive Systems, Research School of Information Sciences and Engineering, Canberra, ACT 0200, Australia

Communicated by J. Stoer

Abstract. An infinite-dimensional convex optimization problem with the linear-quadratic cost function and linear-quadratic constraints is considered. We generalize the interior-point techniques of Nesterov--Nemirovsky to this infinite-dimensional situation. The complexity estimates obtained are similar to finite-dimensional ones. We apply our results to the linear-quadratic control problem with quadratic constraints. It is shown that for this problem the Newton step is basically reduced to the standard LQ problem.

Key Words. Control problems, Quadratic constraints, Path-following algorithms.

AMS Classification. 90C20, 93C05, 49N05.

1. Introduction

In their fundamental monograph [9], Nesterov and Nemirovsky have shown that a broad class of a finite-dimensional optimization problems can be solved using interior-point
technique based on the notion of the self-concordant barrier. Many control problems (like multicriteria LQ and LQG problems) could be treated with the help of this technique provided an infinite-dimensional version were available.

In this paper we make a natural step in this direction. Namely, we consider a convex quadratic programming problem with quadratic constraints in a Hilbert space. The generalization of the technique of Nesterov–Nemirovsky is somewhat straightforward for this situation. We consider both long-step and short-step versions of a path-following algorithm and obtain estimates on the number of Newton steps analogous to the best known for the finite-dimensional situation. In [11] Renegar also considers an infinite-dimensional version of the Nesterov–Nemirovsky scheme. His approach is quite different, however, and aims at the reformulation of complexity estimates in terms of certain generalizations of condition numbers. Of course, the performance of the Newton step in our situation requires solving an infinite-dimensional linear-quadratic problem (without inequality constraints). We consider Yakubovich-type problems [13] and show that the performance of the Newton step is reduced to solving the standard LQ problem of control plus a system of \( m \) linear algebraic equations where \( m \) is the number of inequality constraints.

It is quite natural to expect that the task of finding the Newton direction will at least require solving the standard LQ problem. Thus the real situation is as good as one could hope.

The case of linear inequality constraints admits a direct treatment without using the Nesterov–Nemirovsky technique and leads to sharper complexity estimates [4].

2. Duality, Central Trajectories, and Dynamical Systems

Let \((H, \langle \cdot, \cdot \rangle)\) be a Hilbert space, with \(X\) its closed vector subspace. Given self-adjoint nonnegative definite bounded operators \(Q_0, \ldots, Q_m\) on \(H\), vectors \(x_0, a_0, \ldots, a_m\) in \(H\), and real numbers \(b_0 = 0, b_1, \ldots, b_m\), consider the following optimization problem:

\[
\begin{align*}
q_0(x) &\rightarrow \min, \\
x &\in x_0 + X, \\
q_i(x) &\leq 0, \quad i = 1, 2, \ldots, m,
\end{align*}
\]

where \(q_i(x) = \frac{1}{2} \langle x, Q_i x \rangle + \langle a_i, x \rangle + b_i\).

We assume throughout this paper that the set \(P \subset H\) defined by constraints (2), (3) is bounded and has a nonempty interior. More precisely, \(\text{int}(P) = \{x \in P : q_i(x) < 0, i = 1, 2, \ldots, m\}\) is not empty.

Along with (1)–(3) consider the following (dual) problem:

\[
\mu(y, \lambda) = q_0(y) + \sum_{i=1}^{m} \lambda_i q_i(y) \rightarrow \max,
\]
\[ \nabla q_0(y) + \sum_{i=1}^m \lambda_i \nabla q_i(y) \in X^*, \quad (6) \]

\[ y \in x_0 + X, \quad \lambda_i \geq 0, \quad i = 1, 2, \ldots, m. \quad (7) \]

**Proposition 2.1.** For every \( x \) satisfying (2), (3) and every \((y, \lambda)\) satisfying (6), (7) we have

\[ q_0(x) \geq \mu(y, \lambda). \quad (8) \]

**Proof.** Indeed, \( q_0(x) - \mu(y, \lambda) \geq \mu(x, \lambda) - \mu(y, \lambda) \geq \langle \nabla \mu(y, \lambda), x - y \rangle = 0 \), where in the second inequality we used the fact that \( \mu(x, \lambda) \) is convex in \( x \) for nonnegative \( \lambda \) (see, for example, [10]). The last equality follows by (6) and the fact that \( x - y \in X \). \( \square \)

**Remark.** If \( f \) is a smooth function on \( H \), then \( \nabla f(x), x \in H \), is uniquely determined by the relation

\[ Df(x) \cdot \xi = \langle \nabla f(x), \xi \rangle, \quad \xi \in H. \quad (9) \]

Here \( Df(x) \cdot \xi \) stands for the Fréchet derivative of \( f \) at point \( x \) evaluated on the vector \( \xi \).

Consider the family of optimization problems of the form:

\[ f_\beta(x) = \beta q_0(x) - \sum_{i=1}^m \ln(-q_i(x)) \to \min, \quad (10) \]

\[ q_i(x) < 0, \quad i = 1, 2, \ldots, m, \quad (11) \]

\[ x \in x_0 + X. \quad (12) \]

Here \( \beta \geq 0 \) is a parameter. In this section we make the following assumption:

\[ D^2 f_\beta(x)(\xi, \xi) \geq a_\beta(x)(\xi, \xi), \quad \beta \geq 0, \quad (13) \]

for some \( a_\beta(x) > 0 \). Here \( x \in \text{int}(P), \xi \in X \), and \( D^2 f_\beta(x)(\xi, \xi) \) stands for the second Fréchet derivative of \( f_\beta \) at the point \( x \in \text{int}(P) \) evaluated on the pair \((\xi, \xi) \in X \times X \). Observe that it is sufficient to verify the condition (13) for the case \( \beta = 0 \). We can easily compute

\[ D^2 f_\beta(x)(\xi, \eta) = \xi, \left( \beta Q_0 - \sum_{i=1}^m \frac{Q_i}{q_i(x)} \right) \eta + \sum_{i=1}^m \frac{\langle \nabla q_i(x), \xi \rangle \langle \nabla q_i(x), \eta \rangle}{q_i(x)^2}. \quad (14) \]

We conclude from (14) that \( f_\beta \) is convex on \( \text{int}(P) \).

**Proposition 2.2.** Under the assumption (13) the problem (10)–(12) has a unique solution \( x(\beta) \in \text{int}(P) \), \( \beta \geq 0 \).
Consider the set \( P_\sigma = \{ x \in \text{int}(P) : f_\beta(x) \leq \sigma \} \). It is clear that \( P_\sigma \) is a convex and bounded subset of \( \text{int}(P) \). Suppose that \( x_i, i = 1, 2, \ldots, \) is a sequence of points in \( P_\sigma \) such that \( \lim x_i = \bar{x}, i \to +\infty \) in the strong topology (i.e., in the topology induced by the norm). It is clear that \( \bar{x} \in \text{int}(P) \). If \( \bar{x} \in \partial P = P - \text{int}(P) \), then \( f_\beta(x_i) \to \infty, i \to +\infty \). However, \( f_\beta(x_i) \leq \sigma \) for all \( i \). Hence \( \bar{x} \in \text{int}(P) \). We proved that \( P_\sigma \) is convex, bounded, and strongly closed. Hence \( P_\sigma \) is weakly compact (see, for example, \([2]\)). Since \( f_\beta \) is convex, it is semicontinuous from below on \( \text{int}(P) \). We conclude that \( f_\beta \) attains its minimum \( x \) on \( P_\sigma \). Observe that

\[
\inf \{ f_\beta(x) : x \in \text{int}(P) \} = \inf \{ f_\beta(x) : x \in P_\sigma \}.
\]

The uniqueness of the minimum easily follows from the assumption (13).

**Remark.** In the finite-dimensional situation the curve \( x(\beta), \beta \geq 0, \) is usually called the central trajectory and the point \( x(0) \) is the analytic center (see, for example, p. 59 of \([3]\)).

Our intermediate goal is to prove that \( x(\beta) \) is a smooth function of \( \beta \) and converges to the optimal solution of (1)–(3) when \( \beta \to +\infty \). Denote by \( \pi : H \to X \) the orthogonal projection onto \( X \). The point \( x(\beta) \in \text{int}(P) \) is uniquely determined by the condition

\[
\forall x \in \text{int}(P)
\]

\[
\nabla f_\beta(x(\beta)) \in X^1.
\]

In other words,

\[
\pi \left( \beta a_0 + \beta Q_0 x(\beta) - \sum_{i=1}^m \frac{\nabla q_i(x(\beta))}{q_i(x(\beta))} \right) = 0.
\]

For any \( \beta \geq 0 \) consider the map \( \psi_\beta : \text{int}(P) \to X \),

\[
\psi_\beta(x) = \pi \left( \beta Q_0 x - \sum_{i=1}^m \frac{\nabla q_i(x)}{q_i(x)} \right).
\]

We can rewrite (16) in the form

\[
\psi_\beta(x(\beta)) = -\beta \pi a_0.
\]

**Proposition 2.3.** For each \( \beta \geq 0 \) the map \( \psi_\beta \) is a smooth isomorphism of \( \text{int}(P) \) on \( X \) such that \( \psi_\beta^{-1} \) depends smoothly on \( \beta \).

**Proof.** By Proposition 2.2 for any \( c \in X \) the problem

\[
-(c,x) + \frac{\beta}{2} (Q_0 x, x) - \sum_{i=1}^m \ln(-q_i(x)) \to \min,
\]

\( x \in \text{int}(P) \)
has a unique solution \( x_c \in \text{int}(P) \) and by (18) \( \psi_\beta(x_c) = c \). Hence \( \psi_\beta \) is surjective. On the other hand, if \( \psi_\beta(x_1) = \psi_\beta(x_2), x_1, x_2 \in \text{int}(P) \), then again by (18) both \( x_1 \) and \( x_2 \) are solutions to (19). Hence by Proposition 2.2 \( x_1 = x_2 \), i.e., \( \psi_\beta \) is injective. It remains to prove that \( \psi_\beta^{-1} \) is smooth and depends smoothly on \( \beta \). Observe that, for every \( \xi \in X, x \in \text{int}(P) \),

\[
D\psi_\beta(x) \cdot \xi = \pi \left( \beta Q_0 \xi - \sum_{i=1}^m \frac{Q_i \xi}{q_i(x)} - \frac{\langle \nabla q_i(x), \xi \rangle \nabla q_i(x)}{q_i(x)^2} \right). \quad (20)
\]

If \( D\psi_\beta(x) \cdot \xi = 0 \), then by (20)

\[
0 = \langle \xi, D\psi_\beta(x)\xi \rangle = D^2\psi_\beta(x)(\xi, \xi).
\]

Hence, by (13) \( \xi = 0 \). Thus \( D\psi_\beta(x) \) induces an injective map from \( X \) into \( X \). Further, by (20)

\[
D\psi_\beta(x)|_X = \beta \pi \circ Q \circ \text{in} - \sum_{i=1}^m \frac{\pi \circ Q_i \circ \text{in}}{q_i(x)} + \pi \circ \left( \sum_{i=1}^m \frac{\nabla q_i(x) \otimes \nabla q_i(x)}{q_i(x)^2} \right) |_X, \quad (21)
\]

where \( \text{in} : X \to H \) is the canonical immersion. The condition (13) easily implies that \( D\psi_\beta(x)|_X \) is surjective. We can now apply the implicit function theorem to conclude that \( \psi_\beta^{-1} \) is smooth and depends smoothly on \( \beta \).

**Corollary 2.1.** The solution \( x(\beta) \) of the problem (10)-(12) depends smoothly on \( \beta \).

**Proof.** By (18), \( x(\beta) = \psi_\beta^{-1}(-\beta \pi a_0) \). The result follows by Proposition 2.3.

We now introduce a family of Riemannian metrics \( g_\beta \) which are determined by Hessians of functions \( f_\beta \), namely,

\[
g_\beta(x, \xi, \eta) = D^2f_\beta(x)(\xi, \eta), \quad (22)
\]

\( \xi, \eta \in H \) and \( q_i(x) < 0, i = 1, 2, \ldots, m \).

We clearly have

\[
g_\beta(x)(\xi, \eta) = \langle \gamma_\beta(x)\xi, \eta \rangle, \quad (23)
\]

\[
\gamma_\beta(x) = \beta Q_0 \sum_{i=1}^m \frac{Q_i}{q_i(x)} + \sum_{i=1}^m \frac{\nabla q_i(x) \otimes \nabla q_i(x)}{q_i(x)^2}. \quad (24)
\]

\( \beta \geq 0 \). We can think of \( \text{int}(P) \) (see (4)) as a Riemannian submanifold of the open subset in \( H \) defined by the conditions \( q_i(x) < 0, i = 1, 2, \ldots, m \). Given a smooth function \( \varphi \) defined on an open neighborhood of \( \text{int}(P) \) in \( H \), denote by \( \nabla \varphi(x) \in X \) the gradient of \( \varphi \) relative to the metric \( g_\beta \), the gradient \( \nabla \varphi(x) \in X \) is uniquely determined by the conditions

\[
D\varphi(x) \cdot \xi = \langle \nabla \varphi(x), \xi \rangle = g_\beta(x; \xi, \nabla \varphi(x)) = D^2f_\beta(x)(\xi, \nabla \varphi(x)), \quad (25)
\]

\( \xi \in X, x \in \text{int}(P) \). The existence and uniqueness of the \( \nabla \varphi(x) \in X \) satisfying (25) follows easily by (13).
**Proposition 2.4.** We have
\[ \gamma_\beta(x) \nabla_\beta \varphi(x) - \nabla \varphi(x) \in X^\perp. \]

**Proof.** For \( \xi \in X \),
\[
\langle \xi, \gamma_\beta(x) \nabla_\beta \varphi(x) - \nabla \varphi(x) \rangle = \langle g_\beta(x; \xi, \nabla_\beta \varphi(x)) - (\xi, \nabla \varphi(x)) \rangle = 0.
\]
The result follows.

**Corollary 2.2.** We have
\[ D\psi_\beta(x) \cdot \nabla_\beta \varphi(x) = \pi(\nabla \varphi(x)). \]

**Proof.** By (20), (24) we have
\[ D\psi_\beta(x) \cdot \nabla_\beta \varphi(x) = \pi(\gamma_\beta(x) \nabla_\beta \varphi(x)). \]
The result follows by (26).

The next proposition describes a differential equation which has the central trajectory as its solution.

**Proposition 2.5.** Let \( \varphi(x) = -q_0(x) \). Then
\[ \frac{dx(\beta)}{d\beta} = -\nabla_\beta \varphi(x), \quad \beta \geq 0. \]

**Proof.** Set \( \alpha(\beta) = \psi_\beta(x(\beta)) \). We have (see (17))
\[ \frac{d\alpha}{d\beta}(\beta) = D\psi_\beta(x(\beta)) \frac{dx(\beta)}{d\beta} + \pi(Q_0 x(\beta)). \]
On the other hand, by (18)
\[ \frac{d\alpha}{d\beta}(\beta) = -\pi a_0. \]
Hence,
\[ D\psi_\beta(x(\beta)) \cdot \frac{dx(\beta)}{d\beta} = -\pi(\nabla q_0(x(\beta))). \]
Comparing this with (27) and using Proposition 2.3, we arrive at (28).

**Corollary 2.3.** The function \( q_0(x(\beta)) \) is a monotonically decreasing function of \( \beta \), \( \beta \geq 0 \).
Finite-Dimensional Quadratic Optimization

**proof.** Indeed,
\[
\frac{d}{d\beta} q_i(x(\beta)) = \left( \nabla q_0(x(\beta)), \frac{dx(\beta)}{d\beta} \right) = g_\beta \left( x(\beta); \nabla_\beta q_0(x(\beta)), \frac{dx(\beta)}{d\beta} \right)
\]
\[
= g_\beta (x(\beta); \nabla_\beta q_0(x(\beta)), \nabla_\beta q_0(x(\beta))) \leq 0,
\]
here in the second equality we used (25) and in the third equality we used (28). \(\square\)

Consider now the dual of the problem (10)–(12):
\[
\mu_\beta(y, \lambda) = \beta \mu(y, \lambda) + \sum_{i=1}^m \ln \lambda_i + m(1 + \ln \beta) \to \max,
\]
\[
\nabla q_0(y) + \sum_{i=1}^m \lambda_i \nabla q_i(y) \in X^\perp,
\]
\[
\lambda_i > 0, \quad i = 1, 2, \ldots, m,
\]
\[
y \in x_0 + X.
\]
Here \(\mu(y, \lambda)\) was defined in (5).

**Proposition 2.6.** Let \(x(\beta)\) be a solution to (10)–(12). Set \(y(\beta) = x(\beta)\),
\[
\lambda_i(\beta) = -\frac{1}{\beta q_i(x(\beta))}, \quad i = 1, 2, \ldots, m.
\]
\[
\begin{align*}
\phi_\beta(x(\beta)) &= \mu_\beta(y(\beta), \lambda(\beta)).
\end{align*}
\]  
**proof.** A direct computation. \(\square\)

**Proposition 2.7.** For every \(x\) satisfying (11), (12) and every \((y, \lambda)\) satisfying (30), (31) have
\[
\phi_\beta(x) \geq \mu_\beta(y, \lambda).
\]
**proof.** The proof is quite similar to the one of Lemma 2.19 in [3]. \(\square\)

**Corollary 2.4.** The pair of \((x(\beta), \lambda(\beta))\) is an optimal solution to (29)–(31).
Corollary 2.5. Let $x^*$ be the optimal solution to the problem (1)–(3). Then
\[ q_0(x(\beta)) \geq q_0(x^*) \geq q_0(x(\beta)) - \frac{m}{\beta}. \]  
(33)
\[ \lim_{\beta \to +\infty} q_0(x(\beta)) = q_0(x^*), \quad \beta \to +\infty. \]  
(34)

Proof. Observe that $(x(\delta), \lambda(\delta))$ is a feasible solution to (5)–(7). By Proposition 2.1 $q_0(x^*) \geq \mu(x(\beta), \lambda(\beta)) = q_0(x(\beta)) - m/\beta$. Since by Proposition 2.5 $q_0(x(\beta))$ is a monotonically decreasing function of $\beta$, (34) follows from (33). \(\square\)

Corollary 2.6. The function $\mu(y(\beta), \lambda(\beta))$ is a monotonically increasing function of $\beta$ for $\beta \geq 0$.

Proof. The proof is quite similar to one of Theorem 2.8 in [3]. \(\square\)

Corollary 2.7. If $\beta_1 > \beta_2 > 0$, then
\[ q_0(x(\beta_2)) - q_0(x(\beta_1)) \leq \frac{m}{\beta_2} - \frac{m}{\beta_1}. \]  
(35)

Proof. By the previous corollary $\mu(x(\beta_1), \lambda(\beta_1)) \geq \mu(x(\beta_2), \lambda(\beta_2))$. We also know that
\[ q(x(\beta)) - \mu(x(\beta), \lambda(\beta)) = \frac{m}{\beta}. \]  
(36)
Hence
\[ q_0(x(\beta)) - q_0(x(\beta)) \leq [q_0(x(\beta_2)) - \mu(x(\beta_2), \lambda(\beta_2))] \]
\[ - [q_0(x(\beta)) - \mu(x(\beta_1), \lambda(\beta_1))] = \frac{m}{\beta_2} - \frac{m}{\beta_1}. \]  
\(\square\)

3. Complexity Estimates for Path-Following Algorithms

The main idea behind path-following algorithms is to produce a finite-step approximation to the central trajectory. Using the interpretation of points on the central trajectory as optimal solutions to the family of optimization problems (10)–(12), the construction of path-following algorithms can be reduced to the performance of Newton steps for functions (10) and updating the parameter $\beta$. The choice of logarithmic barriers enables us to use the powerful technique of self-concordant functions [9]. In this section we show that an infinite-dimensional generalization of the Nesterov–Nemirovsky scheme is quite straightforward. We use some simplifications of this scheme proposed by Jarre [7] and den Hertog [3]. Interior-point methods for finite-dimensional convex quadratic programming problems with quadratic constraints have also been considered in [6], [8], and [5].
Let \( q(x) = \frac{1}{2} \langle x, Qx \rangle + \langle a, x \rangle + b \) be a convex quadratic form on \( H \). Given \( x, h \in H \), consider the function
\[
\alpha(t) = q(x + th), \quad t \in \mathbb{R}.
\] (37)

We obviously have
\[
\alpha(t) = q(x) + (Dq(x) \cdot h)t + \frac{1}{2} [D^2q(x)(h, h)]t^2.
\] (38)

Suppose that \( D^2q(x)(h, h) > 0 \) and \( \alpha(0) = q(x) < 0 \). Since \( \alpha \) is convex we conclude that
\[
\alpha(t) = \ell(t - t_1)(t - t_2),
\] (39)

where \( \ell = D^2q(x)(h, h)/2 \) and \( t_1 < 0 < t_2 \). In particular,
\[
\alpha(t) < 0 \iff t_1 < t < t_2.
\] (40)

Furthermore,
\[
-\ln(-\alpha(t)) = -\ln \ell - \ln(t - t_1) - \ln(t_2 - t).
\] (41)

Using (41) we can easily compute the Taylor expansion of the function \( \beta(x) = -\ln(-q(x)) \). Indeed, by (41)
\[
\beta(x + th) = \sum_{n=0}^{\infty} \frac{t^n}{n!} D^n \beta(x)(h, \ldots, h) = -\ln(\ell \cdot t_2 \cdot (-t_1)) + \sum_{n=1}^{\infty} \frac{t^n}{n} \left( \frac{1}{t_1^n} + \frac{1}{t_2^n} \right).
\] (42)

Hence
\[
D^n \beta(x)(h, \ldots, h) = (n - 1)! \left( \frac{1}{t_1^n} + \frac{1}{t_2^n} \right), \quad n \geq 1.
\] (43)

Here \( D^n \beta(x) \) denotes the \( n \)th Fréchet derivative of \( \beta \) at the point \( x \in H \). If \( D^2q(x)(h, h) = 0 \) but \( Dq(x) \cdot h \neq 0 \), then \( \alpha(t) \) is a linear function with a single root \( t_1 \neq 0 \). The corresponding expansion for \( \beta(t) \) will have the form
\[
\beta(t) = -\ln(-q(x)) + \sum_{n=1}^{\infty} \frac{t^n}{n} \left( \frac{1}{t_1^n} \right).
\] (44)

Return now to the problem (10)--(12). Given \( x \in \text{int}(P) \), introduce a measure of proximity of \( x \) to the point \( x(\beta) \) on the central trajectory:
\[
\delta_P(x) = D^2f_\beta(x)(p_\beta(x), p_\beta(x))^{1/2},
\] (45)

where
\[
p_\beta(x) = -\nabla Pf_\beta(x)
\] (46)

(see (25)).
Proposition 3.1. Suppose that \( q_i(x) < 0 \) for \( i = 1, 2, \ldots, m \) and \( h \in H \) is such that
\[
D^2 f_i(x)(h, h) < 1,
\]
then \( q_i(x + h) < 0 \) for \( i = 1, 2, \ldots, m \).

Proof. Consider the functions \( \alpha_i(t) = q_i(x + th) \) and \( \beta_i(t) = -\ln(-\alpha_i(t)) \), \( i = 1, 2, \ldots, m \). Fix \( i \in [1, m] \). Consider first, the case \( D^2 q_i(x)(h, h) > 0 \). We have by \( (48) \)
\[
\frac{1}{|t_1^{(i)}|^2} + \frac{1}{|t_2^{(i)}|^2} < 1
\]
or \( \min(|t_1^{(i)}|^2, |t_2^{(i)}|^2) < 1 \). Hence, \( \alpha_i(t) < 0 \) for \( |t_1^{(i)}|, |t_2^{(i)}| \). This proves \( q_i(x + h) < 0 \). If \( D^2 q_i(x)(h, h) = 0 \) but \( Dq_i(x) \cdot h \neq 0 \), \( \alpha_i(t) \) is a linear (nonconstant) function with the root \( t_1^{(i)} \) such that \( |t_1^{(i)}| > 1 \). Since \( \alpha_i(0) < 0 \), we necessarily have \( \alpha_i(1) < 0 \). The case of the constant \( \alpha_i \) is trivial.

Lemma 3.1. Given \( A_1 > 0, \ldots, A_m > 0 \), we have
\[
(A_1^y + A_2^y + \cdots + A_m^y)^\frac{1}{y} \leq (A_1^x + A_2^x + \cdots + A_m^x)^\frac{1}{x}
\]
for \( 0 < y < x \).

Proof. It is sufficient to prove that the function
\[
\mathcal{X}(x) = \frac{1}{x} \ln(A_1^x + \cdots + A_m^x)
\]
is monotonically decreasing for \( x > 0 \). However
\[
\mathcal{X}'(x) = \frac{1}{x^2(A_1^x + \cdots + A_m^x)}[(A_1^x \ln(A_1^x) + \cdots + A_m^x \ln(A_m^x)) - (A_1^y + \cdots + A_m^y) \ln(A_1^y + A_2^y + \cdots + A_m^y)]
\]
The function \( S(p_1, \ldots, p_m) = \sum_{i=1}^m p_i \ln p_i \) is convex on the positive orthant \( p_i \geq 0 \).

Hence, \( \max(S(p) : p_i \geq 0, i = 1, 2, \ldots, m, p_1 + p_2 + \cdots + p_m = c) \) is attained at one of the extreme points of the simplex \( \{p \in \mathbb{R}^m : p_i \geq 0, i = 1, 2, \ldots, m, p_1 + \cdots + p_m = c\} \).
\( p_m = c \). We conclude \( S(p_1, \ldots, p_m) \leq (p_1 + \cdots + p_m) \ln(p_1 + \cdots + p_m) \). Hence \( X'(x) \leq 0 \) for any positive \( x \).

**Proposition 3.2.** For any \( x \in \text{int}(P), h \in X, \beta \geq 0 \) we have
\[
\left( \frac{|D^j f_0(x)(h, \ldots, h)|}{(j - 1)!} \right)^{1/j} \leq \left( \frac{D^j f_0(x)(h, \ldots, h)}{(k - 1)!} \right)^{1/k}
\]
for any \( k, j \) such that \( k \) is even and \( 2 \leq k \leq j \). In particular,
\[
|D^j f_0(x)(h, h, h)| \leq 2D^2 f_0(x)(h, h)^{3/2}.
\]

**Remark.** The condition (50) means that the function \( f_0 \) is 1-self-concordant. It is crucial in the Nesterov–Nemirovsky scheme [3], [9].

**Proof.** Observe that \( f_0(x) = \beta q_0(x) + \sum_{i=1}^m \beta_i(x) \) with \( \beta_i(x) = - \ln(-q_i(x)) \). Inequality (49) easily follows by (43) and Lemma 3.1.

**Proposition 3.3.** Suppose that \( x \in \text{int}(P), i = 1/(1 + \delta_\beta(x)) \). Then \( x + \hat{t} p_\beta(x) \in \text{int}(P) \) and
\[
f_0(x + \hat{t} p_\beta(x)) = f_0(x) + \ln(1 + \delta_\beta(x)) - \delta_\beta(x).
\]

**Remark.** Recall (46), that \( p_\beta(x) = - \nabla_\beta f_\beta(x) \) is the Newton direction for the \( f_\beta \) at the point \( x; \delta_\beta(x) \) was defined in (45).

**Proof.** Set \( p_\beta(x) = p, \delta_\beta(x) = \delta \) in this proof. We have
\[
f_0(x + tp) = f_0(x) + tDf_0(x) \cdot p + \frac{t^2}{2} D^2 f_0(x)(p, p)
\]
\[
+ \sum_{j \geq 3} D^j f_0(x)(p, \ldots, p) \frac{t^j}{j!}.
\]

Let \( \alpha(t) = q_0(x + tp) = \ell(t - t_1^{(i)})(t - t_2^{(i)}) \) with \( t_i^{(i)} < 0 < t_2^{(i)} \). By (43)
\[
D^j f_0(x)(p, \ldots, p) = (j - 1)! \sum_{i=1}^m \left( \frac{1}{|t_1^{(i)}|^j} + \frac{1}{|t_2^{(i)}|^j} \right).
\]

for \( j \geq 3 \) and
\[
\delta_\beta(x)^2 = D^2 f_0(x)(p, p) = \beta D^2 q_0(x)(p, p) + \sum_{i=1}^m \left( \frac{1}{|t_1^{(i)}|^2} + \frac{1}{|t_2^{(i)}|^2} \right).
\]

(If \( \alpha(t) \) is linear, the formulas (53), (54) hold true if we set \( t_2^{(i)} = +\infty \). See (44).) Observe, further, that by (25)
\[
D f_0(x) \cdot p = - D^2 f_0(x)(p, p) = - \delta_\beta(x)^2
\]
and by (53), (54), and Lemma 3.1

$$\left| \sum_{j \geq 3} D^j f_\delta(x)(p_1 \cdots p_j) \frac{t^j}{j!} \right| \leq \sum_{j \geq 3} \frac{|t|}{j!} (j-1)! \delta_\delta(x)^j. \quad (56)$$

Combining (52), (55), (56), we obtain

$$f_\delta(x + tp) - f_\delta(x) \leq -\delta t (1 - \delta) - \ln(1 - \delta), \quad t \geq 0. \quad (57)$$

For $t = \delta$ we obtain (51). Observe now that

$$D^2 f_\delta(x)(\delta p, \delta p) \leq D^2 f_\delta(x)(\delta p, \delta p) = \frac{\delta_\delta(x)^2}{(1 + \delta_\delta(x))^2} < 1.$$

Hence, $x + \delta p \in \text{int}(P)$ by Proposition 3.1.

**Lemma 3.2.** Suppose that $F$ is a symmetric trilinear form and $G$ is a symmetric bilinear form on a Hilbert space $H$ such that

$$F(h, h, h)^2 \leq \zeta G(h, h)^3 \quad (58)$$

for some $\zeta > 0$ and every $h \in H$. Then

$$F(h_1, h_2, h_3) \leq \zeta G(h_1, h_1)G(h_2, h_2)G(h_3, h_3) \quad (59)$$

for any $h_1, h_2, h_3 \in H$.

**Proof.** Given $h_1, h_2, h_3 \in H$ consider the vector subspace $V \subset H$ spanned by vectors $h_1, h_2, h_3$. It is clear that $V$ is at most three-dimensional. Consider the restrictions of $F, G$ to $V$. It is sufficient to establish (59) for any three vectors in $V$. For the proof of this see, for example, [7].

**Proposition 3.4.** Let $x \in \text{int}(P), \delta_\delta(x) < 1$. Then $x^+ = x + p_\delta(x) \in \text{int}(P)$ and

$$\delta_\delta(x^+) \leq \frac{\delta_\delta(x)^2}{(1 - \delta_\delta(x))^2}. \quad (60)$$

**Proof.** Given $\xi \in H$, consider the function

$$\varphi_\xi(t) = Df_\delta(x + tp) \cdot \xi - (1 - t)Df_\delta(x) \cdot \xi, \quad (61)$$

where $p = p_\delta(x)$. It is clear that $\varphi_\xi(1) = Df_\delta(x + p), \varphi_\xi(0) = 0$. We have

$$\varphi_\xi'(t) = D^2 f_\delta(x + tp)(\xi, p) + Df_\delta(x) \cdot \xi, \quad (62)$$

$$\varphi_\xi''(t) = D^3 f_\delta(x + tp)(\xi, p, p). \quad (63)$$

Using (50) and Lemma 3.2 we obtain

$$|\varphi_\xi''(t)| \leq 2D^2 f_\delta(x + tp)(p, p)[D^2 f_\delta(x + tp)(\xi, \xi)]^{1/2}. \quad (64)$$
Observe now that
\[ D^2f_\delta(x)(\xi, \xi)(1 - t[D^2f_\delta(x)(p, p)]^{1/2})^2 \leq D^2f_\delta(x + tp)(\xi, \xi) \]
\[ \leq \frac{D^2f_\delta(x)(\xi, \xi)}{(1 - t[D^2f_\delta(x)(p, p)]^{1/2})^2} \]  \hspace{1cm} (65)

for 0 ≤ t < 1. Indeed, let \( \varphi(t) = D^2f_\delta(x + tp)(\xi, \xi) \). Using (50) and Lemma 3.2, we have
\[ |\varphi'(t)| \leq 2\varphi(t)^{1/2} |t|, \] \hspace{1cm} (66)
where \( \varphi(t) = D^2f_\delta(x + tp)(p, p) \). Now (again using (50) and Lemma 3.2)
\[ |\varphi'(t)| \leq 2\varphi(t)^{3/2}. \] \hspace{1cm} (67)

By (67) (see p. 54 of [3])
\[ \psi(t) \leq \frac{\psi(0)}{(1 - t\psi(0)^{1/2})^2}, \quad 0 \leq t < 1. \]

By (66) we will have [3, p. 54]
\[ |\varphi'(t)| \leq \frac{2\varphi(t)^{1/2}}{1 - t\psi(0)^{1/2}}, \quad 0 \leq t < 1. \] \hspace{1cm} (68)

By (68) [3, p. 55],
\[ \varphi(0)(1 - t\psi(0)^{1/2}) \leq \varphi(t) \leq \frac{\varphi(0)}{(1 - t\psi(0)^{1/2})^2}, \] \hspace{1cm} (69)

which coincides with (65). Observe now that \( \varphi'_2(0) = D^2f_\delta(x)(p, \xi) + Df_\delta(x) \cdot \xi = -Df_\delta(x) \cdot \xi + Df_\delta(x) \cdot \xi = 0 \), where we used (25), (46). Hence from (64), (65) we obtain
\[ |\varphi'_2(t)| = \left| \int_0^t \varphi'_2(\tau) \, d\tau \right| \leq \int_0^t |\varphi'_2(\tau)| \, d\tau \]
\[ \leq 2\delta^2[D^2f_\delta(x)(\xi, \xi)]^{1/2} \cdot \int_0^t \frac{d\tau}{(1 - t\delta)^3} \]
\[ = \delta[D^2f_\delta(x)(\xi, \xi)]^{1/2} \cdot \left( \frac{1}{(1 - t\delta)^2} - 1 \right). \] \hspace{1cm} (70)

where \( \delta = \delta_\delta(x) \). Further, by taking into account \( \varphi'_2(0) = 0 \), we obtain
\[ |\varphi'_2(t)| \leq [D^2f_\delta(x)(\xi, \xi)]^{1/2} \int_0^t \left( \frac{1}{(1 - t\delta)^2} - 1 \right) \, d\tau \]
\[ = \frac{\delta^2[D^2f_\delta(x)(\xi, \xi)]^{1/2}}{1 - t\delta}. \] \hspace{1cm} (71)

Take \( t = 1, \xi = -p_\delta(x^+) \). We have
\[ \varphi'_2(1) = -Df_\delta(x^+) \cdot p_\delta(x^+) = D^2f_\delta(x^+)(p_\delta(x^+), p_\delta(x^+)) = \delta_\delta(x^+)^2. \]
Hence, by (71)
\[
\delta_f(x^+)^2 \leq \frac{\delta^2}{1-\delta} \left[ D^2 f_F(x)(p_F(x^*), p_F(x^+)) \right]^{1/2}
\]
\[
\leq \frac{\delta^2}{1-\delta} \frac{\delta p(x^*)}{1-\delta},
\]
which coincides with (60). In the last inequality we used (65) again. \qed

Remark. The main idea of the proof of Proposition 3.4 is due to Nesterov and Nemirovsky [9]. We used simplifications suggested by [7] and [3].

Given \( \sigma \in \mathbb{R} \), denote the set \( \{ x \in \text{int}(P) : f_F(x) \leq \sigma \} \) by \( P_{\sigma} \). We know (see the proof of Proposition 2.2) that \( P_{\sigma} \) is a convex, closed, and bounded set. Suppose that for every \( \sigma \in \mathbb{R} \) there exist positive constants \( m_{\sigma} \) and \( M_{\sigma} \) such that

\[
M_{\sigma} \| \xi \|^2 \geq D^2 f_F(x)(\xi, \xi) \geq m_{\sigma} \| \xi \|^2, \quad \xi \in X, \quad x \in P_{\sigma}.
\]  

(72)

The next proposition states the convergence of the Newton method for the function \( f_F \).

**Proposition 3.5.** Given \( \beta \geq 0, x \in \text{int}(P), \delta_f(x) < \frac{1}{3} \), consider the sequence

\[
x_0 = x, \quad x_1 = x_0 + p_F(x_0), \ldots, \quad x_i = x_{i-1} + p_F(x_{i-1}), \quad i \geq 2.
\]

Then \( x_i \in \text{int}(P) \) for all \( i \) and moreover under the assumption (72)

\[
\lim_{i \to +\infty} x_i = x(\beta),
\]

(73)

**Proof.** By (60)

\[
\delta_f(x_i) \leq \frac{\delta_f(x_0)^2}{(1-\delta_f(x_0))^2} \leq \frac{2}{3} \delta_f(x_0)^2.
\]

Using induction on \( i \), we can easily conclude from (60):

\[
\delta_f(x_i) \leq \left( \frac{2}{3} \right)^{i-1} \delta_f(x_0)^2 \leq \left( \frac{3}{4} \right)^{i-1} \cdot \frac{1}{3}.
\]

(74)

In particular,

\[
\delta_f(x_i) \to 0, \quad i \to +\infty.
\]

(75)

It is also clear from (74) that \( \delta_f(x_i) < 1 \) for all \( i \). Hence, by Proposition 3.1 \( x_i \in \text{int}(P) \) for all \( i \). Moreover, by (74) \( \delta_f(x_i) < \frac{1}{3} \) for all \( i \geq 1 \). Since the function \( \delta^2 + \delta + \ln(1-\delta) \) is positive for \( 0 < \delta < \frac{1}{2} \), we conclude by (57) that \( f_F(x_i) < f_F(x_{i-1}), i \geq 2 \). Hence, all \( x_i \) belong to the set \( P_{\sigma} \) for an appropriately chosen \( \sigma \). Since by (72)

\[
\delta_f(x_i) \geq m_{\sigma} \| p_F(x_i) \|, \quad i \geq 2,
\]
we have by (75)

\[ p_\beta(x_i) \to 0, \quad i \to +\infty. \]  

(76)

Recall that by (26)

\[ \pi \circ \gamma_\beta(x_i) p_\beta(x_i) = \pi \nabla f_\beta(x_i). \]

The condition (72) implies that \( ||\pi \circ \gamma_\beta(x)|| \leq M_a \) for all \( x \) in \( P_a \). Hence, (76) implies

\[ \pi \nabla f_\beta(x_i) \to 0, \quad i \to +\infty. \]  

(77)

Observe now that

\[ (Df_\beta(y) - Df_\beta(x)) \cdot (y - x) \geq m_a ||x - y||^2 \]

for all \( x, y \in P_a \) (see, for example, [10]). Take \( x = x(\beta), y = x_i \). We have

\[ (Df_\beta(y) - Df_\beta(x)) \cdot (y - x) = (\nabla f_\beta(x_i) - \nabla f_\beta(x(\beta)), x_i - x(\beta)) \]

\[ = (\nabla f_\beta(x_i) - \nabla f_\beta(x(\beta), \pi(x_i - x(\beta)))) \]

\[ = (\pi \nabla f_\beta(x_i) - \pi \nabla f_\beta(x(\beta)), x_i - x(\beta)) \]

\[ = (\pi \nabla f_\beta(x_i), x_i - x(\beta)) \geq m_a ||x_i - x(\beta)||^2. \]  

(78)

where in the second equality we used \( x_i - x(\beta) \in X \), but \( \pi \nabla f_\beta(x(\beta)) = 0 \) in the fourth equality. Now (77), (78), and boundedness of \( P_a \) imply \( ||x_i - x(\beta)|| \to 0, i \to +\infty. \)

\[ \hfill \square \]

**Proposition 3.6.** Given \( x \in \text{int}(P), \delta_\beta(x) < \frac{1}{3} \), we have

\[ f_\beta(x) - f_\beta(x(\beta)) \leq \frac{\delta_\beta(x)^2}{1 - [\frac{3}{4}]^2 \delta_\beta(x)^2}. \]

(79)

**Proof.** Set \( p = p_\beta(x) \) in this proof. Since \( f_\beta \) is convex, we have

\[ f_\beta(x + p) \geq f_\beta(x) + Df_\beta(x) \cdot p = f_\beta(x) - \delta_\beta(x)^2, \]

(80)

where we used (55). Let \( x_0 = x \) and \( x_1, x_2, \ldots \) be the sequence of points obtained by repeating Newton's steps. We have by (74)

\[ \delta_\beta(x_i) \leq \left( \frac{3}{4} \right)^{i-1} \delta_\beta(x_0)^2 \]

(81)

and \( x_i \to x(\beta), i \to +\infty \), by Proposition 3.5.

Hence,

\[ f_\beta(x) - f_\beta(x(\beta)) = \sum_{i=0}^{\infty} [f_\beta(x_i) - f_\beta(x_{i+1})] \]

\[ \leq \sum_{i=0}^{\infty} \delta_\beta(x_i)^2 \leq \sum_{i=0}^{\infty} \left( \frac{3}{4} \right)^{2i+1} \delta_\beta(x_0)^{2i+1} \]

\[ \leq \frac{\delta_\beta(x_0)^2}{1 - \left( \frac{3}{4} \delta_\beta(x_0)^2 \right)}, \]

where we used (80) in the first inequality and (81) in the second inequality. \( \hfill \square \)
Proposition 3.7. If \( x \in \text{int}(P), \delta_p(x) < \frac{1}{3} \), then
\[
|q_0(x) - q_0(x(\beta))| \leq \frac{\delta_p(x)}{1 - \frac{9}{4}\delta_p(x)} \cdot \frac{(1 + \delta_p(x)^2)}{1 - \delta_p(x)} \cdot \frac{\sqrt{m}}{\beta}.
\] (82)

Proof. The proof is quite similar to one of Lemma 2.23 in [3]. \( \square \)

We are now in the position to describe a version of the large-step path-following algorithm for solving (1)–(3). Suppose we are given \( \beta_0 > 0 \) and \( x_0 \in \text{int}(P) \) such that \( \delta_p(x_0) \leq \frac{1}{3} \). Given \( \theta > 0 \), we perform what are called outer iterations \( \beta_1 = (1 + \theta)\beta_0, \ldots, \beta_i = (1 + \theta)^i \beta_0 \). Suppose we are able to construct a sequence of points \( x_1, x_2, \ldots \) in \( \text{int}(P) \) such that \( \delta_p(x_i) \leq \frac{1}{3} \).

Theorem 3.1. Suppose \( \varepsilon > 0 \) is given and
\[
i \geq \frac{\ln(4m/\varepsilon\beta_0)}{\ln(1 + \theta)}.
\] (83)

Then \( q_0(x_i) - q_0(x^*) \leq \varepsilon \), where \( x^* \) is an optimal solution to the problem (1)–(3).

Proof. Observe that (83) is equivalent to
\[
\beta_i \geq \frac{4m}{\varepsilon}.
\] (84)

By (33)
\[
q_0(x(\beta_i)) - q_0(x^*) \leq \frac{m}{\beta_i},
\]

By (82) with \( \delta_p(x) \leq \frac{1}{3} \) we obtain
\[
q_0(x_i) - q_0(x(\beta_i)) \leq \frac{\frac{1}{3} \cdot \frac{1}{1 - \frac{9}{4} \cdot \frac{1}{3}} \cdot \frac{1 + (\frac{1}{3})^2 \sqrt{m}}{1 - \frac{1}{3}}}{\beta_i} \leq \frac{3\sqrt{m}}{\beta_i}.
\]

Thus
\[
q_0(x_i) - q_0(x^*) = q_0(x_i) - q_0(x(\beta_i)) + q_0(x(\beta_i)) - q_0(x^*) \leq \frac{3\sqrt{m} + m}{\beta_i} \leq \varepsilon,
\]

where in the last inequality we used (84). \( \square \)

It remains to describe a procedure for updating \( x_i \). Suppose we are given \( x_i \in \text{int}(P) \) such that \( \delta_p(x_i) \leq \frac{1}{3} \). Set \( \beta_{i+1} = (1 + \theta)\beta_i \). We have to find \( x_{i+1} \in \text{int}(P) \) such that \( \delta_p(x_{i+1}) \leq \frac{1}{3} \). To do that we perform several Newton steps for the function \( f_{\beta_{i+1}} \), using \( x_i \) as a starting point. Each such Newton step is called an inner iteration.
Theorem 3.2. Each outer iteration requires at most

$$\frac{\Delta^2}{3} + 22\theta \left( \frac{\sqrt{\gamma}}{2} + \frac{\delta}{1 + \delta} \right)$$

inner iterations.

Proof. We start with a point \( x_i \in \text{int}(P) \) such that \( \delta_{\beta_i}(x_i) \leq \frac{1}{2} \). By Proposition 3.3 the point

\[
\bar{x} = x_i + \bar{t} p_{\beta_i}(x_i), \quad \bar{t} = \frac{1}{1 + \delta_{\beta_i}(x_i)},
\]

belongs to \( \text{int}(P) \) and moreover

\[
f_{\beta_i}(x_i) - f_{\beta_{i+1}}(\bar{x}) \geq \delta_{\beta_{i+1}}(x_i) - \ln(1 + \delta_{\beta_{i+1}}(x_i)) \geq \frac{1}{2} - \ln(1 + \frac{1}{2}) > \frac{1}{22}. \quad (86)
\]

We continue to perform iterations described above until \( \delta_{\beta_{i+1}}(x_{i+1}) \leq \frac{1}{2} \). By (86) the value of the function \( f_{\beta_{i+1}} \) is decreased by at least \( \frac{1}{22} \) after such an iteration. If \( N \) is the number of such iterations, we obviously have

\[
\frac{N}{22} \leq f_{\beta_{i+1}}(x_i) - f_{\beta_{i+1}}(\bar{x}_{i+1}). \quad (87)
\]

Let

\[
\eta(x, \beta) = f_{\beta}(x) - f_{\beta}(\bar{x}).
\]

We wish to estimate \( \eta(x_i, \beta_{i+1}) \). By the mean value theorem

\[
\eta(x_i, \beta_{i+1}) = \eta(x_i, \beta) + \frac{\partial \eta}{\partial \beta}(x_i, \beta) \delta_{i+1}, \quad (88)
\]

for some \( \hat{\beta} \in (\beta_i, \beta_{i+1}) \). We have

\[
\frac{\partial \eta}{\partial \beta}(x, \beta) = \frac{\partial q_0(x)}{\partial \beta} + \left( \nabla f_{\beta}(x(\beta)) \cdot \frac{dx(\beta)}{d\beta} \right).
\]

Moreover, \( \nabla f_{\beta}(x(\beta)) \in X^1, dx(\beta)/d\beta \in X \).

Hence,

\[
\frac{\partial \eta}{\partial \beta}(x, \beta) = q_0(x) - q_0(x(\beta)). \quad (89)
\]

(88), (89)

\[
\eta(x_i, \beta_{i+1}) \leq \eta(x_i, \beta_i) + (\beta_{i+1} - \beta_i)|q_0(x_i) - q_0(x(\beta))| \\
\leq \eta(x_i, \beta_i) + \delta_{i+1}|q_0(x(\beta)) - q_0(x_i)| \\
+ q_0(x(\beta)) - q_0(x(\beta_{i+1})). \quad (90)
\]
In the last inequality we used the fact that \( q_0(x(\beta)) \) is a monotonically decreasing function of \( \beta \).

By (82)
\[
|q_0(x_i) - q_0(x(\beta_i))| \leq \frac{\frac{1}{2} + \frac{1}{2}}{1 - \frac{1}{2}} \frac{\sqrt{m}}{\beta} \leq \frac{5}{2} \frac{\sqrt{m}}{\beta}.
\]
(91)

Further, by (79)
\[
\eta(x_i, \beta_i) \leq \frac{\frac{1}{2}^2}{1 - \frac{1}{2}^2} < \frac{1}{3}.
\]
(92)

Finally, by (35)
\[
q_0(x(\beta)) - q_0(x(\beta_{i+1})) \leq \frac{m}{\beta_i} - \frac{m}{\beta_{i+1}} = \frac{m}{\beta_{i+1}}.
\]
(93)

Substituting (91)–(93) into (90), we obtain
\[
\eta(x_i, \beta_{i+1}) \leq \frac{1}{3} + \theta \left( \frac{5\sqrt{m}}{2} + \frac{m}{1 + \theta} \right).
\]
(94)

Combining (94), (87) we arrive at (85). \( \square \)

From Theorems 3.1 and 3.2 we immediately obtain:

**Theorem 3.3.** An upper bound for the total number of Newton iterations is given by
\[
\ln(4m/\varepsilon \beta_0) \quad \text{ln}(1 + \theta) \left( \frac{1}{2} + 22\theta \left( \frac{5\sqrt{m}}{2} + \frac{m}{1 + \theta} \right) \right).
\]

We now briefly consider the situation where the number of inner iterations (per one outer iteration) does not exceed 1.

**Proposition 3.8.** Given \( x \in \text{int}(P), \beta > 0, \theta > 0 \). If \( \beta^+ = (1 + \theta)\beta \), then
\[
\delta_{\beta^+}(x) \leq (1 + \theta)\delta_{\beta}(x) + \theta \sqrt{m}.
\]
(95)

**Proof.** Given \( \xi \in X \),
\[
D^2f_{\beta^+}(x)(\xi, \xi) = \beta^+ D^2q_0(x)(\xi, \xi) + \sum_{i=1}^{m} \frac{|Dq_i(x) \cdot \xi|^2}{q_i(x)^2} \geq D^2f_{\beta}(x)(\xi, \xi).
\]
(96)

Introduce a vector \( \bar{p} \in X \) by the condition
\[
D^2f_{\beta}(x)(\xi, \bar{p}) = -Df_{\beta^+}(x) \cdot \xi, \quad \xi \in X.
\]
(97)

We have
\[
Df_{\beta^+}(x) \cdot \xi = (1 + \theta)Df_{\beta}(x) \cdot \xi + \theta \sum_{i=1}^{m} \frac{(\nabla q_i(x), \xi)}{q_i(x)}.
\]
Since, by (97)
\[ D^2_{f_{\theta}}(x)(\xi, \tilde{\rho} = (1 + \theta)D^2_{f_{\theta}}(x)(\xi, p_{\theta}(x)) - \theta \sum_{i=1}^{m} \frac{\langle \nabla q_i(x), \xi \rangle}{q_i(x)} \]  
(98)

Taking \( \xi = \tilde{\rho} \), we obtain from (98)
\[ D^2_{f_{\theta}}(x)(\tilde{\rho}, \tilde{\rho}) \leq (1 + \theta) [D^2_{f_{\theta}}(x)(\tilde{\rho}, \tilde{\rho})]^{1/2} [D^2_{f_{\theta}}(x)(p_{\theta}(x), p_{\theta}(x))]^{1/2} \]
\[ + \sqrt{m} \left( \sum_{i=1}^{m} \frac{\langle \nabla q_i(x), \tilde{\rho} \rangle^2}{q_i(x)} \right) \]
\[ \leq [D^2_{f_{\theta}}(x)(\tilde{\rho}, \tilde{\rho})]^{1/2} (1 + \theta) + \sqrt{m}. \]  
(99)

We used the Cauchy–Schwarz inequality twice. It remains to prove that
\[ D^2_{f_{\theta}}(x)(p_{\theta}(x), p_{\theta}(x)) \leq D^2_{f_{\theta}}(x)(\tilde{\rho}, \tilde{\rho}). \]  
(100)

used, by (97)
\[ -D^2_{f_{\theta}}(x)(\xi) = D^2_{f_{\theta}}(x)(\xi, \tilde{\rho}) = D^2_{f_{\theta}}(x)(\xi, p_{\theta}(x)). \]

particular, for \( \xi = p_{\theta}(x) \),
\[ D^2_{f_{\theta}}(x)(p_{\theta}(x), p_{\theta}(x)) = D^2_{f_{\theta}}(x)(\tilde{\rho}, p_{\theta}(x)) \]
\[ \leq [D^2_{f_{\theta}}(x)(p_{\theta}(x), p_{\theta}(x))]^{1/2} [D^2_{f_{\theta}}(x)(\tilde{\rho}, \tilde{\rho})]^{1/2} \]
\[ \leq [D^2_{f_{\theta}}(x)(p_{\theta}(x), p_{\theta}(x))]^{1/2} [D^2_{f_{\theta}}(x)(\tilde{\rho}, \tilde{\rho})]^{1/2}, \]
here is the last inequality we used (96). The result follows by (99), (100). \( \square \)

**Theorem 3.4.** Let \( x \in \text{int}(P) \), \( \delta_{\beta}(x) \leq \frac{1}{2} \), and \( \beta^+ = (1 + \theta)\beta \), where \( \theta = v/\sqrt{m} \), for sufficiently small \( v > 0 \). Then \( x^+ = x + p_{\theta}(x) \) is such that \( \delta_{\beta^+}(x^+) \leq \frac{1}{2} \).

*Proof.* By Propositions 3.4 and 3.8
\[ \delta_{\beta^+}(x^+) \leq \frac{\delta_{\beta^+}(x)^2}{[1 - \delta_{\beta^+}(x)]^2} \leq \frac{\delta_{\beta^+}(x)^2}{\frac{3}{4}} \leq \frac{3}{4} [(1 + \theta)\delta_{\beta}(x) + \theta \sqrt{m}]^2 \]
\[ = \frac{\sqrt{3}}{4} \left( 1 + \frac{v}{\sqrt{m}} \right)^2 \delta_{\beta}(x)^2 \leq \Delta(v). \]  
(101)

Observe that \( \Delta(v) \to \frac{\sqrt{3}}{4} \delta_{\beta}(x)^2 \leq \frac{1}{4} \) for \( v \to 0 \). Hence \( \delta_{\beta^+}(x^+) \leq \frac{1}{2} \) for sufficiently small \( v \). \( \square \)

**Example**

Consider the linear-quadratic time-dependent control problem with quadratic inequality constraints. Namely, let
\[ q_i(y, u) = \frac{1}{2} \int_0^T \left[ y^T(t)Q_i(t)y(t) + u(t)^T R_i(t)u(t) \right] dt + b_i, \quad i = 1, 2, \ldots, m. \]  
(102)
where \((u, y) \in L^2_\nu[0, T] \times L^2_u[0, T]\), \(y\) is an absolutely continuous function with values in \(\mathbb{R}^n\) and such that \(\dot{y} \in L^2_\nu[0, T]\). We use the standard notation \(L^2_\nu[0, T]\) for the Hilbert space of measurable square integrable functions on \([0, T]\) with values in \(\mathbb{R}^n\). Further, \(Q_i(t)\) (resp. \(R_i(t)\)) is a symmetric positive definite \(n \times n\) (resp. \(p \times p\)) matrix which depends continuously on \(t \in [0, T]\); \(A(t)\) (resp. \(B(t)\)) is an \(n \times n\) (resp. \(n \times p\)) matrix which depends continuously on \(t \in [0, T]\); \(b_i, i = 1, 2, \ldots, m\), are real numbers and \(b_0 = 0\); \(T\) is a fixed positive number.

Consider the problem

\[
q_0(y, u) \rightarrow \min
\]

\[
q_i(y, u) \leq 0, \quad i = 1, 2, \ldots, m
\]

\((y, u) \in x_0 + X,\)

where

\[
X = \{(y, u) \in L^2_\nu[0, T] \times L^2_u[0, T] : y \text{ is absolutely continuous on } [0, T],
\]

\[
\dot{y} \in L^2_\nu[0, T], y(0) = 0 \text{ and } \dot{y}(t) = A(t)y(t) + B(t)u(t), t \in [0, T]\}.
\]

Theorem 3.3 gives a rather sharp estimate for the number of Newton steps for the path-following algorithm. From the computational viewpoint it is crucial to see that the performance of a Newton step is a reasonable problem. Set \(x = (y, u)\). We frequently omit the time variable \(t\) in what follows. We have

\[
f_\beta(x) = \beta q_0(y, u) - \sum_{i=1}^{m} \ln(-q_i(y, u)),
\]

\[
\nabla f_\beta(y, u) = \begin{bmatrix}
\beta Q_0 \cdot y - \sum_{i=1}^{m} \frac{Q_i \cdot y}{q_i(y, u)} \\
\beta R_0 \cdot y - \sum_{i=1}^{m} \frac{R_i \cdot u}{q_i(y, u)}
\end{bmatrix},
\]

\[
\gamma_\beta(y, u) \begin{bmatrix}
\xi \\
\eta
\end{bmatrix} = \begin{bmatrix}
\beta Q_0 \xi - \sum_{i=1}^{m} \left( \frac{Q_i \cdot \xi}{q_i(y, u)} + \frac{d_i Q_i \cdot x}{q_i^2(y, u)} \right) \\
\beta R_0 \eta - \sum_{i=1}^{m} \left( \frac{R_i \cdot \eta}{q_i(y, u)} + \frac{d_i R_i \cdot y}{q_i^2(y, u)} \right)
\end{bmatrix},
\]

\[
d_i = \int_{0}^{T} (y^T Q_i \xi + u^T R_i \eta) \, dt, \quad i = 1, 2, \ldots, m,
\]

\[
p_\beta(x) = \begin{bmatrix}
\xi \\
\eta
\end{bmatrix}.
\]
\[ X^\perp = \left\{ \left[ \begin{array}{c} \dot{p} + A^T p \\ B^T p \end{array} \right] : p \text{ is absolutely continuous on } [0, T] \text{ and } \dot{p} \in L^2_0[0, T], p(T) = 0 \right\}. \]

Equation (26) will take form

\[ \dot{p} + A^T p - \ddot{\gamma} = Q_p(x)\dot{\xi} + \sum_{i=1}^{m} \frac{(Q_i y) d_i}{q_i^2(x)}, \quad \tag{105} \]

\[ B^T p - \ddot{u} = R_p(x)\eta + \sum_{i=1}^{m} \frac{(R_i u) d_i}{q_i^2(x)}, \quad \tag{106} \]

where

\[ Q_p(x) = \beta Q_0 - \sum_{i=1}^{m} \frac{Q_i}{q_i(x)}, \]

\[ R_p(x) = \beta R_0 - \sum_{i=1}^{m} \frac{R_i}{q_i(x)}. \]

We further have

\[ \dot{\xi} = A\xi + B\eta, \quad \xi(0) = 0, \]

\[ p(T) = 0, \quad \nabla f_p(x) = \left[ \begin{array}{c} \ddot{\gamma} \\ \ddot{u} \end{array} \right]. \tag{107} \]

We have to solve (105)–(107) with respect to \( \xi, \eta \). The problem (105)–(107) is quite similar to one arising in connection with the standard linear-quadratic control problem (see, for example, [12] and [1]). The only difference is that the constants \( d_i \) are unknown. We proceed, nevertheless, as in the standard linear-quadratic case. We are looking for the solution (105)–(107) in the form

\[ p(t) = K(t)\dot{\xi}(t) + \rho(t). \tag{108} \]

By (106)

\[ \eta(t) = R_p(x)^{-1} B^T p - \ddot{u} - \sum_{i=1}^{m} \frac{d_i R_p(x)^{-1} R_i u}{q_i^2(x)}. \tag{109} \]

Substituting (108), (109) in (100), (106), respectively, we obtain

\[ K + KA + A^T K + KL_p(x)K - Q_p(x) = 0. \tag{110} \]

\[ K(T) = 0 \text{ where } L_p(x) = B(t)R_p(x)^{-1} B^T(t), \]

\[ \dot{p} = -A^T p + Q_p(x)\dot{\xi} + \alpha(t), \quad p(T) = 0. \tag{111} \]
\[ \dot{\xi} = A\xi + L_\beta(x) \cdot \rho + \beta(t), \quad \xi(0) = 0, \] (112)

\[ \alpha(t) = \tilde{y} + \sum_{i=1}^{m} \frac{(Q_i y) d_i}{q_i^2(x)}, \] (113)

\[ \beta(t) = -B \left( R_\beta(x)^{-1} \bar{u} + \sum_{i=1}^{m} \frac{(R_\beta(x)^{-1} R_i u) d_i}{q_i^2(x)} \right), \] (114)

\[ \eta = R_\beta(x)^{-1} \left( B^T p - \bar{u} - \sum_{i=1}^{m} \frac{(R_i u) d_i}{q_i^2(x)} \right), \] (115)

\[ \dot{\rho} = -(A^T + K L_\beta) \rho + \alpha(t) + \beta(t), \quad \rho(T) = 0. \] (116)

From (110) we find \( K(t), 0 \leq t \leq T. \) Let \( \varphi(t, \tau) \) be the fundamental solution to
\[ \dot{\rho} = -(A^T + K L_\beta) \rho. \] (117)

Then \( \psi(t, \tau) = [\varphi(t, \tau)^{-1}]^T \) is the fundamental solution to
\[ \dot{\xi} = (A + L_\beta K)\xi. \]

By (116)
\[ \rho(t) = \int_0^t \psi(t, \tau) (\alpha(\tau) + \beta(\tau)) \, d\tau = v(t) + \sum_{i=1}^{m} d_i v_i(t). \] (118)

(See (113) for the known functions \( v, v_i. \))

By (112)
\[ \xi(t) = \int_0^t \psi(t, \tau) (L_\beta \rho(\tau) + \beta(\tau)) \, d\tau \]
\[ = \mu(t) + \sum_{i=1}^{m} d_i \mu_i(t) \] (119)

for known functions \( \mu, \mu_i \) (see (114)).

By (108), (118), (119)
\[ \rho(t) = \chi(t) + \sum_{i=1}^{m} d_i \chi_i(t) \] (120)

for the known functions \( \chi, \chi_i \) and by (115)
\[ \eta(t) = \pi(t) + \sum_{i=1}^{m} d_i \pi_i(t) \] (121)

for the known functions \( \pi, \pi_i. \)
Substituting (119), (121) into (104), we obtain

\[ d_i = \sum_{j=1}^{m} c_{ij} d_j + e_i, \quad i = 1, 2, \ldots, m, \tag{122} \]

where

\[ c_{ij} = \int_0^T (y^T Q_i \mu_j + u^T R_i \pi_j) \, dt, \]

\[ e_i = \int_0^T (y^T Q_i \mu + u^T R_i \pi_j) \, dt. \]

Finally, we find \( d_i \) solving the system of linear algebraic equations (122). We see that the performance of the Newton step involves three essential steps:

(a) The numerical integration of the Riccati equation (110)
(b) Finding the fundamental matrix for the linear system for the linear time-dependent system (117)
(c) Solving a system of \( m \) by \( m \) linear algebraic equations (122).

The first two steps are the same as for solving the standard time-dependent LQ problem. We see that the performance of the Newton step in our situation is essentially equivalent to solving the standard LQ problem plus the system of linear algebraic equations. The standard numerical treatment of the problem (103) involves introducing the Lagrange multipliers and considering the dual problem. From the computational viewpoint, finding the value of the cost function of the dual problem at a point is more or less equivalent to performing the Newton step of the primal problem.

**Concluding Remarks**

In the present paper we have considered both long-step and short-step versions of a th-following algorithm for the infinite-dimensional quadratic programming problem with quadratic constraints.

We found an estimate for the number of Newton steps which coincides with the best own [3] in the finite-dimensional case. We then considered an example of the linear-quadratic time-dependent control problem with quadratic constraints. We have shown that in this situation the performance of the Newton step is essentially equivalent to solving the standard LQ problem without inequality constraints. Thus the interior-point methodology has been extended to an infinite-dimensional situation useful for control applications.

**References**


Accepted 24 October 1995