Multi-Parameter Surfaces of Analytic Centers and Long-step Surface-Following Interior Point Methods

Yu. Nesterov and A. Nemirovski

We develop a long-step polynomial time version of the Method of Analytic Centers for nonlinear convex problems. The method traces a multi-parameter surface of analytic centers rather than the usual path, which allows to handle cases with non-centered and possibly infeasible starting point.

Key words: Convex optimization, interior point methods, polynomial time methods, method of analytic centers, quadratic programming, semidefinite programming, geometric programming

1 Introduction

Consider a convex program in the following standard form:

$$\textit{minimize} \ \ c^T x \ \ \textit{s.t.} \quad x \in G; \tag{1}$$

here G is a closed and bounded convex subset of $\mathbf{R}^n$ with a nonempty interior. One of the most attractive theoretically ways to solve the problem is to trace the path of *analytic centers*, i.e., the minimizers over $x \in \operatorname{int} G$ of the penalized family of functions

$$F^t(x) = F(x) + tc^T x; \tag{2}$$

here F is a barrier (interior penalty function) for G. Under the above parameterization of the path, in order to converge to the optimal set one should trace the path as $t \to \infty$. The path of analytic centers, however, can be parameterized in another way, say, as the path of minimizers of the family

$$F_t(x) = F(x) - \vartheta \ln(t - c^T x); \tag{3}$$

($\vartheta > 0$ is fixed); here in order to get close to the optimal set one should approach the parameter t to the optimal value of the problem. As it is well-known, both the parameterizations, under appropriate choice of F, imply polynomial-time interior-point methods for (1). If G is a polytope, then it is reasonable to choose as F the standard logarithmic barrier for G; polynomiality of the associated path-following methods for Linear Programming was first established in the seminal papers of Renegar (1988), parameterization (3), and Gonzaga (1989), parameterization (2). For the nonpolyhedral case polynomial time results for both the parameterizations can be obtained if F is a *self-concordant barrier* for G (see below), as it is the case with the standard logarithmic barrier for a polytope.

Now, in order to trace the path of analytic centers F one should once get close to the path; this is the aim of a special *preliminary phase* of a path-following method, which, theoretically, is of the same complexity as following the path itself. Moreover, to initialize the preliminary phase one should know in advance a strictly feasible solution $\widehat{x} \in \operatorname{int} G$. To get such a point, it, generally speaking, again requires an additional phase of the method; at this phase we, basically, solve an auxiliary problem of the same type as (1), but with a known in advance strictly feasible solution. There are numerous strategies of combining all these phases; one of the main goals of this paper is to present a kind of a general framework, based on the notion of a "multi-parameter surface of analytic centers", for these combined strategies. The notion is introduced in Section 2, along with motivating the advantages of tracing surface as compared to tracing the usual single-parameter path. Section 2 contains also a generic predictor-corrector scheme of tracing a surface of analytic centers. When tracing a surface, one should decide, first, *where* to move - what should be the strategy of choosing the subsequent search directions - and, second, *how* to move - what should be the tactics of choosing the stepsize in the chosen direction in order to move as fast as possible. The "tactics" issues are discussed in Sections 3 and 4. In Section 3 we develop, under some reasonable assumptions on the structure of the underlying barriers, a duality-based technique which, roughly speaking, allows to adjust the stepsizes to the "local curvature" of the surface and thus results, under favourable circumstances, in "long steps". Main theoretical properties of the resulting scheme are presented in Section 4. In particular, we demonstrate that under reasonable assumptions on the underlying barriers "long steps" indeed are long – they form a significant fraction of the way to the boundary of the feasible set. Note that our "long steps" technique is closely related to the one recently developed in Nesterov (1993) for path-following methods as applied to primal-dual conic formulations of convex problems, and the results on the length of the steps are similar to those of Nesterov and Todd (1994,1995). The advantage of our approach as compared to Nesterov (1993) and Nesterov and Todd (1994,1995) is not only in the fact that now we are able to trace surfaces rather than paths, but also that now we need neither explicit conic reformulation of the initial problem, nor the self-scaled property of the associated cone; this allows to avoid necessity to increase the number of variables and enables to work with problems (e.g., the Geometric Programming ones) which cannot be covered by methods of Nesterov and Todd (1994,1995).

In Section 5 we present a strategy of tracing the surface for both the cases of feasible and infeasible start; the strategy in question fits the standard polynomial time complexity bounds and seems to be computationally reasonable.

As it was already mentioned, the "long step" technique presented in the paper requires certain assumption on the structure of the barriers in question; Section 6 presents a number of barriers satisfying this assumption and thus allows to understand what might be the applications of the developed technique.

2 Surfaces of analytic centers: preliminaries

We start with specifying the basic for what follows notions of a self-concordant function/barrier (Nesterov and Nemirovski (1994), Chapter 2; what is called below a self-concordant function, in the indicated book is a "strongly self-concordant function").

2.1 Self-concordant functions and barriers

Definition 2.1 Let Q be an open nonempty convex domain in certain $\mathbf{R}^N$. A function $\Psi : Q \to \mathbf{R}$ is called *self-concordant* (s.-c. for short), if Ψ is a C^3 smooth convex function on Q which tends to ∞ along every sequence of points from Q converging to a boundary point of Q and satisfies the differential inequality

$$\left|D^3\Psi(u)[h,h,h]\right| \leq 2\left(D^2\Psi(u)[h,h]\right)^{3/2}, \ u \in Q, h \in \mathbf{R}^N; \tag{4}$$

from now on $D^sF(u)[h_1,...,h_s]$ denotes s-th differential of a function F taken at a point u along the set of directions $h_1,...,h_s$.
If, in addition,

$$|D\Psi(u)[h]|^2 \leq \vartheta D^2\Psi(u)[h,h],\ u \in Q, h \in \mathbf{R}^N, \tag{5}$$

for some $\vartheta \geq 1$, we say that Ψ is a *ϑ-self-concordant barrier* (ϑ-s.-c.b. for short) for the closed convex domain $G = \mathrm{cl}\, Q$.

Let $\alpha \geq 1$. We say that a s.-c. function Ψ is *α-regular* on its domain Q, if Ψ is C^4 function such that

$$\left|D^4\Psi(u)[h,h,h,h]\right| \leq \alpha(\alpha+1)D^2\Psi(u)[h,h]\|h\|^2_{Q,u},\ u \in Q, h \in \mathbf{R}^N, \tag{6}$$

where

$$\|h\|_{Q,u} = \inf\{t^{-1} \mid t > 0, u \pm th \in Q\}$$

is the (semi)norm on $\mathbf{R}^N$ with the unit ball being the closure of the symmeterization $Q \cap (2u - Q)$ of Q with respect to u.

E.g., the standard logarithmic barrier $-\sum_{i=1}^m \ln(b_i - a_i^T u)$ for a nonempty polytope $G = \mathrm{cl}\{u \mid a_i^T u < b_i,\ i = 1,...,m\}$ is both m-s.-c.b. for G and 2-regular s.-c. function on int G.

The important for us properties of self-concordant functions/barriers are as follows (for proofs, see Nesterov and Nemirovski (1994), Chapter 2):

Proposition 2.1 [Combination rules]

(i) [summation] *Let Q_i, $i = 1,...,k$, be open convex domains in $\mathbf{R}^N$ with a nonempty intersection Q and let Ψ_i be s.-c. functions on Q_i. Then the function $\Psi(x) = \sum_i \Psi_i(x)$ is s.-c. on Q. If all Ψ_i are ϑ_i-s.-c.b.'s for $\mathrm{cl}\, Q_i$, then Ψ is a $(\sum_i \vartheta_i)$-s.-c.b. for $\mathrm{cl}\, Q$, and if all Ψ_i are α-regular, then so is Ψ.*

(ii) [direct summation] *Let $Q_i \subset \mathbf{R}^{N_i}$ be open convex domains, $i = 1,...,k$, and let Ψ_i be s.-c. on Q_i. Then the function*

$$\Psi(u_1...,u_k) = \sum_i \Psi_i(u_i) : Q \equiv \prod_i Q_i \to \mathbf{R}$$

is s.-c. on Q. If all Ψ_i are ϑ_i-s.-c.b.'s for $\mathrm{cl}\, Q_i$, then Ψ is a $(\sum_i \vartheta_i)$-s.-c.b. for $\mathrm{cl}\, Q$, and if all Ψ_i are α-regular, then so is Ψ.

(iii) [affine substitutions of argument] *Let Q^+ be an open convex set in $\mathbf{R}^M$ and $\mathcal{A}$ be an affine mapping from $\mathbf{R}^N$ into $\mathbf{R}^M$ with the image intersecting Q^+. Let Ψ^+ be s.-c. on Q^+; then $\Psi(\cdot) \equiv \Psi^+(\mathcal{A}(\cdot))$ is s.-c. on $Q = \mathcal{A}^{-1}(Q^+)$. If Ψ^+ is a ϑ-s.-c.b. for $\mathrm{cl}\, Q^+$, then Ψ is a ϑ-s.-c.b. for $\mathrm{cl}\, Q$, and if Ψ^+ is α-regular, then so is Ψ.*

From now on, for a positive semidefinite symmetric matrix A and a vector h of the corresponding dimension,

$$|h|_A = (h^T A h)^{1/2}.$$

Proposition 2.2 *Let Q be a nonempty open convex domain in $\mathbf{R}^N$, $G = \mathrm{cl}\, Q$ and Ψ be a s.-c. function on Q. Then*

(i) [behaviour in Dikin's ellipsoid] *For any $u \in Q$ the Dikin ellipsoid*

$$W_\Psi(u) = \{v \mid |v - u|_{\Psi''(u)} < 1\}$$

is contained in Q, and and

$$\Psi(u+h) \leq \Psi(u) + h^T\Psi'(u) + \rho(|h|_{\Psi''(u)}); \tag{7}$$

from now on,

$$\rho(r) = -\ln(1-r) - r. \tag{8}$$

(ii) [nondegeneracy] *If $\Psi''(u)$ is nondegenerate at certain $u \in Q$, then Ψ'' is nondegenerate everywhere on Q; this for sure is the case when Q does not contain lines. If Q is bounded, then Ψ attains its minimum over Q, and the minimizer is unique.*

(iii) [stability with respect to Legendre transformation] *Let $\Psi''(u)$ be nondegenerate for some $u \in Q$ (and then, by (ii), for any $u \in Q$). Consider the Legendre transformation*

$$\Psi_*(v) = \sup\{u^T v - \Psi(u) \mid u \in Q\}$$

regarded as a function on the domain Q^* comprised of those v for which the right hand side is finite. Then Q^* is an open nonempty convex domain, the mapping $x \mapsto \Psi'(x)$ is a one-to-one correspondence between Q and Q^* and Ψ_* is s.-c. and with nondegenerate Hessian on Q^*; the Legendre transformation of Ψ_* is exactly Ψ.

2.2 Surface of analytic centers

Let F be a ϑ-s.-c.b. for a closed and *bounded* convex domain $G \subset \mathbf{R}^n$ with a nonempty interior. Aside from the parameterization issues, the path of analytic centers associated with the barrier can be defined as the set of points $x \in \operatorname{int} G$ where $-\nabla F(x) = \lambda c$ for some positive λ. A natural "multi-parameter" extension of this description is as follows: let us fix k nonzero vectors $c_1, ..., c_k$ and associate with this collection the "surface" $\mathcal{S}_k = \mathcal{S}_k(c_1, ..., c_k)$ comprised of all points $x \in \operatorname{int} G$ where

$$-\nabla F(x) = \sum_{i=1}^{k} \lambda_i c_i$$

with certain positive λ_i. A convenient way to parameterize the surface is to introduce the k-dimensional "parameter"

$$t = (t_1, ..., t_k)^T$$

and associate with this parameter the barrier (cf. (3))

$$F_t(x) = \vartheta \sum_{i=1}^{k} \psi_i(t_i - c_i^T x) + F(x), \quad \psi_i(r) = -\ln r,$$

for the convex set

$$G_t = \{x \in G \mid c_i^T x \le t_i, \; i = 1, ..., k\}.$$

In what follows we are interested only in those values of t for which $\operatorname{int} G_t \neq \emptyset$; the corresponding set $T = T_k(e_1, ..., e_k)$ of values of t clearly is a nonempty open convex and monotone ($t' \ge t \in T \Rightarrow t' \in T$) subset in $\mathbf{R}^k$. Now, since F is a ϑ-s.-c.b. for G, the functions F_t, $t \in T$, are ϑ_*-s.-c.b.'s for the domains G_t with

$$\vartheta_* = (k+1)\vartheta \tag{9}$$

(Proposition 2.1; note that the function $-\vartheta \ln(s)$ for $\vartheta \ge 1$ clearly is a ϑ-s.-c.b. for $\mathbf{R}_+$). Since G_t is bounded, F_t attains its minimum over $\operatorname{int} G_t$, and the corresponding minimizer $x_k^*(t)$ - the *analytic center of* G_t - is unique (Proposition 2.2.(ii)). At this minimizer, of course,

$-\nabla F(x) = \vartheta \sum_{i=1}^k (t_i - c_i^T x)^{-1} c_i$, so that $x_k^*(t) \in \mathcal{S}_k$. Vice versa, every point of $\mathcal{S}_k$ is $x_k^*(t)$ for certain $t \in T$: immediate computation demonstrates that

$$\{\lambda_i > 0,\, i = 1, ..., k\} \& \{-\nabla F(x) = \sum_{i=1}^k \lambda_i c_i\} \Rightarrow x = x_k^*(\vartheta \lambda_1^{-1} + c_1^T x, ..., \vartheta \lambda_k^{-1} + c_k^T x).$$

Thus, we do have defined certain parameterization of $\mathcal{S}_k$.

The following property of the surfaces of analytic centers is immediate:

Lemma 2.1 *Let $\widehat{t} \in T_k(e_1, ..., e_k)$ and let $t^j \geq \widehat{t}$, $j = 1, 2, ...$ be such that $t_i^j = \widehat{t}_i$ for $i \in I \subset \{1, ..., k\}$ and $t_i^j \to \infty, j \to \infty$, for $i \notin I$. Then the points $x_k^*(t^j)$ converge as $j \to \infty$ to the point $x_l^*(\{\widehat{t}_i\}_{i \in I})$ of the surface $\mathcal{S}_l(\{c_i\}_{i \in I})$, $l = \mathrm{card}\, I$ (from now on $\mathcal{S}_0$ is the 0-dimensional surface comprised of the (unique) minimizer x_F^* of F over $\mathrm{int}\, G$). Thus, the surfaces of dimensions $< k$ obtained from $\mathcal{S}_k(c_1, ..., c_k)$ by eliminating some of the vectors c_i are contained in the closure of $\mathcal{S}_k$. In particular, the closures of all surfaces of analytic centers have a point in common, namely, x_F^*.*

2.3 Tracing surfaces of analytic centers: motivation

To solve problem (1), we may take an arbitrary set $c_2, ..., c_k$ of vectors, set $c_1 = c$ and trace the surface $\mathcal{S}_k(c_1, ..., c_k)$ along certain sequence $\{t^i\}$ of values of the parameter, i.e., to produce approximations $x^i \in \mathrm{int}\, G_{t^i}$ of the analytic centers $x^*(t^i)$. If the sequence $\{t^i\}$ is such that $t_1^i \to c^*$ as $i \to \infty$, c^* being the optimal value in (1), then x^i clearly form a sequence of feasible solutions to (1) converging to the optimal set. In what follows we demonstrate that there are basically the same possibilities to trace the surface $\mathcal{S}_k$ as in the standard case when $\mathcal{S}_k$ is a single-parameter path; with this in mind, let us explain what are the advantages of tracing a multi-parameter surface $\mathcal{S}_k$ rather than the usual path $\mathcal{S}_1(c)$.

1. Difficulty of initialization. As it was already mentioned, in the usual path-following method we trace the path $\mathcal{S}_1(c)$; to start the process, we should, anyhow, come close to the path. Now assume that we are given an initial strictly feasible solution $\widehat{x}$. In the standard path-following scheme, to get close to $\mathcal{S}_1(c)$ we trace the auxiliary path $\mathcal{S}_1(d)$, $d = -\nabla F(\widehat{x})$ which clearly passes through $\widehat{x}$. According to Lemma 2.1, both the paths $\mathcal{S}_1(c)$ and $\mathcal{S}_1(d)$ approach, as the parameter tends to ∞, the minimizer x_F^* of F over G; therefore, tracing the auxiliary path as $t \to \infty$, we in the mean time come close to the path $\mathcal{S}_1(c)$ and then may switch to tracing this latter path. On the other hand, given $\widehat{x}$ and arbitrary $\widehat{t}_1 > c^T \widehat{x}$, we can easily find vector c_2 and real $\widehat{t}_2$ such that the 2-dimensional surface of analytic centers $\mathcal{S}_2(c_1 \equiv c, c_2)$ would pass, as $t = (\widehat{t}_1, \widehat{t}_2)$, through $\widehat{x}$; it suffices to set

$$c_2 = \vartheta^{-1} d - (\widehat{t}_1 - c^T \widehat{x})^{-1} c, \quad \widehat{t}_2 = c_2^T \widehat{x} + 1.$$

Now we have a two-dimensional surface of analytic centers which "links" $\widehat{x}$ with the optimal set, and we may use various policies of tracing the surface, starting at $\widehat{x}$, in order to approach the optimal set. Note that our "main path" $\mathcal{S}_1(c)$, due to Lemma 2.1, lies in the closure of $\mathcal{S}_2$, while the "auxiliary path" $\mathcal{S}_1(d)$, as it is immediately seen, simply belongs to the surface. Thus, the standard path-following scheme - first trace $\mathcal{S}_1(d)$ and then $\mathcal{S}_1(c)$ - is nothing but a specific way to trace the two-dimensional surface of analytic centers $\mathcal{S}_2(c, c_2)$. After this is realized, it becomes clear that there is no necessity to restrict ourselves with the above specific route; why not to move in a more "direct" manner, thus avoiding the preliminary phase where we do not take care of the objective at all?

2. Infeasible start. Now assume that we do not know in advance an initial strictly feasible solution to the problem. What should we do? Note that normally the situation, under appropriate renaming of the data, is as follows. We need to solve the problem

$$(P'): \quad minimize \;\; c^T x \;\; s.t. \;\; x \in G, \, x_n = 0,$$

where G is a solid in $\mathbf{R}^n \cap \{x_n \geq 0\}$ with a known in advance interior point $\widehat{x}$. In other words, normally we can represent the actual feasible set as a kind of a "facet" in a higher-dimensional convex solid with known in advance interior point. To support this claim, consider a standard form convex program

$$\text{(CP)} \quad f_0(u) \to \min \mid \; f_i(u) \leq 0, \, i = 1, ..., m \quad [u \in \mathbf{R}^q],$$

(f_i, $0 \leq i \leq m$, are convex lower semicontinuous functions) and assume that we know in advance

- a point u_0 such that all f_i are finite in a neighbourhood of u_0,
- an upper bound $R > |u_0|$ on the Euclidean norm of the optimal solution to (CP),
- an upper bound $V > f_0(u_0)$ on the optimal value of the problem.

Then we can equivalently rewrite (CP) in the form of (P') with the design vector $x = (u, v, w) \in \mathbf{R}^{q+2}$, the objective $c^T x \equiv v$ and

$$G = \{(u, v, w) \mid \; f_0(u) \leq v \leq V; \, u^T u \leq R^2; f_i(u) \leq w, \, i = 1, ..., m; \, 0 \leq w \leq W\},$$

where W is an arbitrary constant which is greater than $\widehat{f}(u_0) \equiv \max\{0, f_1(u_0), ..., f_m(u_0)\}$. Note that G indeed is a solid in $\mathbf{R}^{q+2}$ and that there is no difficulty to point out an interior point x_0 in G: one can set $x_0 = (u_0, v_0, w_0)$ with arbitrarily chosen $v_0 \in (f_0(u_0), V)$ and $w_0 \in (\widehat{f}(u_0), W)$.

Now assume that we are given a ϑ-s.-c.b. F for G.Note that in the case of problem (P') coming from (CP) such a barrier is readily given by ϑ_i-s.-c.b.'s $F_i(v, u)$ for the epigraphs $\{(u, v) \mid v \geq f_i(u)\}$ of the functions f_i, $i = 0, ..., m$:

$$F(u, v, w) = F_0(u, v) + \sum_{i=1}^{m} F_i(u, w) - \ln(R^2 - u^T u) - \ln(V - v) - \ln(W - w) - \ln w \qquad \left[\vartheta = 4 + \sum_{i=0}^{m} \vartheta_i\right].$$

In this situation the standard "big M" approach to (P') is to apply an interior point method to the problem

$$(P): \quad minimize \;\; (c + Mf)^T x \;\; s.t. \;\; x \in G \qquad \left[f^T x \equiv x_n\right],$$

where M is a "large enough" constant. Here we meet with unpleasant question how big should be the "big M". Now note that the path $\mathcal{S}_1(c + Mf)$ which is traced in the "big M" scheme clearly belongs to the two-dimensional surface $\mathcal{S}_2(c, f)$, which is independent of the particular value of M we choose. Thus, the "big M" approach is nothing but a specific way of tracing certain 2-dimensional surface of analytic centers. After this is realized, we may ask ourselves why should we trace the surface in this particular manner rather than to use more flexible strategies.

Note that for the Linear Programming case (G is a polytope, F is the standard logarithmic barrier for G) the surface $\mathcal{S}_2(c, f)$ was introduced and studied in details, although from a slightly different viewpoint, in Mizuno et al (1993).

Now, to trace $\mathcal{S}_2(c, f)$, we should first get close to the surface. Here again the traditional way would be to note that all surfaces $\mathcal{S}_k(c_1, ..., c_k)$ in G come close to each other, so that tracing the path $\mathcal{S}_1(-\nabla F(\widehat{x}))$ (which passes through the given point $\widehat{x}$) and pushing the parameter to ∞, we come close to x^*_F and, consequently, to $\mathcal{S}_2(c,f)$ and then can switch to tracing the surface $\mathcal{S}_2(c,f)$. But this is nothing but a particular way to trace the 3-parameter surface $\mathcal{S}_3(c,f,d)$ in G given by

$$d = -\vartheta^{-1}\nabla F(\widehat{x}) - (\widehat{t}_1 - c^T\widehat{x})^{-1}c - (\widehat{t}_2 - \widehat{x}_n)^{-1}f$$

($\widehat{t}_1 > c^T\widehat{x}, \widehat{t}_2 > \widehat{x}_n$ are arbitrary). The surface $\mathcal{S}_3$ clearly passes through $\widehat{x}$ and links $\widehat{x}$ with the optimal set of the initial problem. After this is realized, why should we restrict ourselves with certain particular route?

2.4 The "surface-following" scheme

We believe that the aforementioned discussion demonstrates that it makes sense to trace not only *paths* of analytic centers, but also *multi-parameter surfaces* of these centers, at least 2- and 3-parameter ones. The point is, of course, how to trace such a surface; this is the issue we address in this section.

2.5 Surface of analytic centers: general definition

To make the presentation more compact it is convenient to get rid of particular structure of the surfaces introduced so far and speak about general situation as follows. Assume that G^+ is a closed convex domain with a nonempty interior in certain $\mathbf{R}^N$ and Ψ is a ϑ_*-s.-c.b. for G^+ with nondegenerate Ψ''. Let, further, π and σ be $N \times n$ and $N \times k$ matrices, respectively, and let $\epsilon \in \mathbf{R}^N$. Consider the affine mapping

$$(t,x) \mapsto \mathcal{A}(t,x) = \sigma t + \pi x + \epsilon : \mathbf{R}^k \times \mathbf{R}^n \to \mathbf{R}^N,$$

and assume that

(A) the image of the mapping $\mathcal{A}$ intersects the interior of G^+.

(A) implies that the set

$$Q = \{(t,x) \mid \mathcal{A}(t,x) \in \mathrm{int}\, G^+\}$$

is a nonempty open convex subset of $\mathbf{R}^k \times \mathbf{R}^n$. Let T be the projection of Q onto the "parameter space" $\mathbf{R}^k$; for $t \in T$ let

$$Q_t = \{x \in \mathbf{R}^n \mid (t,x) \in Q\},\ t \in T; \quad G_t = \mathrm{cl}\, Q_t,$$

so that Q_t is a nonempty open convex set in $\mathbf{R}^n$.

Our second, and for the time being the last, assumption is

(B) for some (and, consequently, for all) $t \in T$ the set Q_t is bounded.

Since Ψ is a ϑ_*-s.-c.b. for G^+, (A) implies that the function

$$F(t,x) = \Psi(\mathcal{A}(t,x))$$

is a ϑ_*-s.-c.b. for $\mathrm{cl}\, Q$, and for $t \in T$ the function

$$F_t(x) \equiv F(t,x)$$

is a ϑ_*-s.-c.b. for G_t (Proposition 2.1). Since G_t is bounded, the Hessian of $F_t(\cdot)$ is nondegenerate at any point from $Q_t = \mathrm{int}\ G_t$, and $F_t(\cdot)$ attains its minimum over Q_t at exactly one point $x^*(t)$ (Proposition 2.2.(ii)). From now on we call the set

$$\mathcal{S} = \{(t,x) \in Q \mid x = x^*(t)\}$$

the *surface of analytic centers* associated with the data G^+, Ψ, $\mathcal{A}(\cdot)$. Note that the surface of analytic centers $\mathcal{S}_k(c_1, ..., c_k)$ is obtained from the general definition by setting

$$G^+ = \mathbf{R}^k_+ \times G,\ \Psi(u_1, ..., u_k, x) = -\vartheta \sum_{i=1}^{k} \ln u_i + F(x),\ \mathcal{A}(t,x) = \begin{pmatrix} t_1 - c_1^T x \\ \cdots \\ t_k - c_k^T x \\ x \end{pmatrix},\ \vartheta_* = (k+1)\vartheta, \tag{10}$$

ϑ being the parameter of the s.-c.b. F.

2.6 Tracing a surface of analytic centers: basic scheme

Our general scheme of tracing the surface $\mathcal{S}$ associated with the data $G^+, \Psi, \mathcal{A}$ is as follows. First, we fix the "tolerances"

$$\kappa \in (0, 0.125], \quad \overline{\kappa} > 2\rho(\kappa) - \kappa^2.$$

We say that a pair (t,x) is *κ-close* to $\mathcal{S}$ if it satisfies the predicate

$\mathcal{P}_\kappa(t,x):$ $\qquad (t,x) \in Q$ *and* $\lambda(t,x) \le \kappa,$

where

$$\lambda(t,x) = \left((\nabla_x F(t,x))^T [\nabla_x^2 F(t,x)]^{-1} \nabla_x F(t,x)\right)^{1/2}$$

is the *Newton decrement* of the s.-c.b. $F_t(\cdot)$ at x; the quantity is well-defined, since, as it was already mentioned, boundedness of G_t implies nondegeneracy of $\nabla^2 F_t(x)$ at any $x \in Q_t$.

We say that a pair (t,x) is *$\overline{\kappa}$-good* with respect to $\mathcal{S}$, if it satisfies the predicate

$\mathcal{R}_{\overline{\kappa}}:$ $\qquad \{(t,x) \in Q\} \,\&\, \{V(t,x) \equiv F(t,x) - \min_{u \in Q_t} F(t,u) \le \overline{\kappa}\}.$

When tracing $\mathcal{S}$, at each step we are given a κ-close to $\mathcal{S}$ pair (t,x) and transform it into a new pair (t^+, x^+) with the same property according to the following

Basic Updating Scheme:

1. Choose a *search direction* $\delta t \in \mathbf{R}^k$, "lift" it to the direction

$$(\delta t, \delta x) \in \Pi(t,x) \equiv \{(dt,dx) \in \mathbf{R}^k \times \mathbf{R}^n \mid \frac{\partial}{\partial t} \nabla_x F(t,x) dt + \nabla_x^2 F(t,x) dx = 0\} =$$

$$= \{(dx,dt) \mid dx = -[\nabla_x^2 F(t,x)]^{-1} \frac{\partial^2}{\partial t \partial x} F(t,x) dt\} =$$

$$= \{(dt,dx) \mid \pi^T \Psi''(\sigma t + \pi x + \epsilon)[\sigma dt + \pi dx] = 0\} \tag{11}$$

and define the *primal search line*

$$R = \{X(p) = (t + p\delta t, x - d_x(t,x) + p\delta x) \mid p \in \mathbf{R}\},\ d_x(t,x) = [\nabla_x^2 F(t,x)]^{-1} \nabla_x F(t,x). \tag{12}$$

2. [predictor step] Choose a stepsize $r > 0$ along the primal search line and form the forecast $(t^+, \widetilde{x}) \equiv X(r)$ which should be $\overline{\kappa}$-good with respect to $\mathcal{S}$ (this is a restriction on the stepsize; a stepsize satisfying this restriction will be called *proper*, and it will be proved that proper stepsizes do exist).

3. [corrector step] Apply to the function $F(t^+, \cdot)$ the damped Newton minimization

$$y^{s+1} = y^s - \frac{1}{1 + \lambda(t^+, y^s)} [\nabla_x^2 F(t^+, y^s)]^{-1} \nabla_x F(t^+, y^s), \quad y^0 = \widetilde{x}. \tag{13}$$

(13) is terminated when it turns out that $\lambda(t^+, y^s) \le \kappa$; the corresponding y^s is taken as x^+, which ensures $\mathcal{P}_\kappa$. The updating $(t, x) \mapsto (t^+, x^+)$ is completed.

Comment. The origin of equations (11) and (12) is clear: (11) is the equation in variations corresponding to the equation $\nabla_x F(s, y) = 0$ of the surface $\mathcal{S}$, so that $\Pi(t, x)$ is the "approximate tangent plane" to the surface at the point (t, x). The primal search line R is given by the linearization

$$\nabla_x F(t, x) + \frac{\partial}{\partial t} \nabla_x F(t, x) dt + \nabla_x^2 F(t, x) dx = 0$$

of the equation of the surface at the point (t, x): it is comprised of the points $(t + dt, x + dx)$ with dx and dt linked by the above equation and dt is proportional to δt.

We do not discuss here how to choose the direction δt; it depends on an "upper-level" strategy of tracing the surface, the issue to be discussed in Section 5.

Since $(t^+, \widetilde{x})$ is a $\overline{\kappa}$-good pair, the number of iterations (13) at a corrector step - the *Newton complexity* of the step - can be bounded as follows:

Proposition 2.3 [Nesterov and Nemirovski (1994), Theorem 2.2.3 and Proposition 2.2.2] *Process (13) is well-defined (i.e., it keeps the iterates y^s in* int G_{t^+}*) and results in y^s such that $\mathcal{P}_\kappa(t^+, y^s)$ is satisfied in no more than*

$$N = O(1) \{\overline{\kappa} + \ln \ln(1/\kappa)\} \tag{14}$$

Newton iterations (13); here and further $O(1)$ denote appropriately chosen absolute constants.

The point is, of course, how to choose the "large" stepsize r for which the forecast $X(r) = (t, x) + r(\delta t, \delta x)$ satisfies the predicate $\mathcal{R}_{\overline{\kappa}}$. To this end it is natural to use line search in r. A straightforward line search is impossible, since $V(\tau, y)$ involves the implicitly defined quantity

$$f^*(\tau) = \min_{u \in Q_\tau} F(\tau, u).$$

What we intend to do is to derive "computationally cheap" *lower bound* for the latter quantity. This is the issue we are coming to.

3 Dual bounds

3.1 Basic assumption

From now on we make the following assumption on the barrier Ψ under consideration:

(C) we know the Legendre transformation

$$\Psi_*(s) = \sup_u \{s^T u - \Psi(u)\} \tag{15}$$

of the barrier Ψ.

"We know Ψ_*" means that, given s, we are able to check whether s belongs to the domain $\mathrm{Dom}\,\Psi_*$ of the Legendre transformation, and if it is the case, are able to compute $\Psi_*(s)$.

Note that by assumption Ψ'' is nondegenerate, so that the domain of Ψ_* is an open convex set and Ψ_* is s.-c. on its domain (Proposition 2.2, (ii) and (iii)).

3.2 Dual bounds

Let us start with the following simple observation

Lemma 3.1 *Let $s \in \mathrm{Dom}\,\Psi_*$ satisfy the linear homogeneous equation*

$$\pi^T s = 0. \tag{16}$$

Then for any $\tau \in T$ we have

$$f^*(\tau) \geq [\sigma\tau + \epsilon]^T s - \Psi_*(s). \tag{17}$$

Proof. Since Ψ is the Legendre transformation of Ψ_* (Proposition 2.2.(iii)), we have for any $y \in Q_\tau$

$$F(\tau, y) = \Psi(\sigma\tau + \pi y + \epsilon) \geq [\sigma\tau + \pi y + \epsilon]^T s - \Psi_*(s) = [\sigma\tau + \epsilon]^T s - \Psi_*(s)$$

(we have used that $\pi^T s = 0$). ■

According to Lemma, each *dual feasible vector* s (a vector s from $\mathrm{Dom}\,\Psi_*$ satisfying (16)) results in an affine lower bound for the function $f^*(\cdot)$, and these are the bounds we intend to use in order to ensure $\mathcal{R}$. Note that dual feasible vectors belong to the subspace $\mathcal{D} \subset \mathbf{R}^N$ of all solutions to linear equation (16); the vectors from this subspace will be called *dual feasible directions*. The set $\mathcal{D}^*$ of dual feasible vectors clearly is an open convex subset of $\mathcal{D}$.

We are about to present a systematic way to generate dual feasible directions and dual feasible vectors.

3.3 Dual search line

Lemma 3.2 *Given a primal search line*

$$R = \{X(p) = (t + p\delta t, x - d_x(t,x) + p\delta x) \mid p \in \mathbf{R}\} \tag{18}$$

($(t,x) \in Q$, $\delta t \in \mathbf{R}^k$, δx and $d_x(t,x)$ are given by (11) and (12)), set

$$u = \sigma t + \pi x + \epsilon, \quad s(t,x) = \Psi'(u), \quad d_s(t,x) = \Psi''(u)\pi d_x(t,x), \quad \delta s = \Psi''(u)[\sigma\delta t + \pi\delta x] \tag{19}$$

and define the dual search line as

$$R^* = \{S(p) = s(t,x) - d_s(t,x) + p\delta s \mid p \in \mathbf{R}\}. \tag{20}$$

Then all vectors from R^* are dual feasible directions:

$$\pi^T S(p) = 0, \ p \in \mathbf{R}. \tag{21}$$

Moreover,

$$\lambda(t,x) < 1 \Rightarrow S(0) \in \mathrm{Dom}\,\Psi_*, \tag{22}$$

so that under the premise of (22) all points $S(p)$ corresponding to small enough $|p|$ are dual feasible vectors.

Proof. To simplify notation, let us omit explicit indicating the argument values in the below computations; the values of all quantities related to Ψ are taken at the point u, and the values of the quantities related to Ψ_* are taken at the point $s = s(t,x) = \Psi'(u)$.

By virtue of (12), (19) and (20) we have

$$\pi^T S(p) = \pi^T \Psi' - \pi^T \Psi'' \pi [\nabla_x^2 F(t,x)]^{-1} \nabla_x F(\tau, x) + p\pi^T \Psi'' [\sigma \delta t + \pi \delta x] =$$

[since $\nabla_x^2 F(t,x) = \pi^T \Psi'' \pi$, $\nabla_x F(t,x) = \pi^T \Psi'$ and due to (11)]

$$= p\pi^T \Psi'' [\sigma \delta t + \pi \delta x] = 0,$$

as required in (21).

To prove (22), note that

$$\lambda^2(t,x) = |\nabla_x F(t,x)|^2_{[\nabla_x^2 F(t,x)]^{-1}} = \left| [\nabla_x^2 F(t,x)]^{-1} \nabla_x F(t,x) \right|^2_{\nabla_x^2 F(t,x)} =$$

$$= |d_x(t,x)|^2_{\pi^T \Psi'' \pi} = |\pi d_x(t,x)|^2_{\Psi''} =$$

[see (19) and take into account that $[\Psi'']^{-1} = \Psi''_*$]

$$= \left| [\Psi'']^{-1} d_s(t,x) \right|^2_{\Psi''} = |d_s(t,x)|^2_{[\Psi'']^{-1}} = |d_s(t,x)|^2_{\Psi''_*}.$$

Thus, we come to

$$|d_s(t,x)|_{\Psi''_*} = |\pi d_x(t,x)|_{\Psi''} = \lambda(t,x). \tag{23}$$

It remains to note that, as we know, Ψ_* is s.-c. on its domain and that $s = \Psi' \in \operatorname{Dom} \Psi_*$, so that (23) combined with Proposition 2.2.(i) (applied to Ψ_*) implies that if $\lambda(t,x) < 1$, then $S(0) = s - d_s(t,x) \in \operatorname{Dom} \Psi_*$. ∎

Now we are ready to present the "computationally cheap" sufficient condition for a forecast $X(r)$ to satisfy the predicate $\mathcal{R}$:

Basic Test: *given r, compute $X(r)$ and $S(r)$ (see (18) - (20)) and verify whether $X(r) \in Q$ and $S(r) \in \operatorname{Dom} \Psi_*$ (then $S(r)$ is a dual feasible vector, see Lemma 3.2). If one of these inclusions is not valid, reject r, otherwise check the inequality*

$$v(r) \equiv F(t + r\delta t, x - d_x(t,x) + r\delta x) + \Psi_*(S(r)) - [\sigma[t + r\delta t] + \epsilon]^T S(r) \le \overline{\kappa}. \tag{24}$$

If it is satisfied, accept r, otherwise reject it.

An immediate consequence of lemmas 3.1 and 3.2 is as follows.

Proposition 3.1 *Let $(t,x) \in Q$, $(\delta t, \delta x) \in \Pi(t,x)$, and let $r \ge 0$ be such that r passes the Basic Test. Then the forecast $X(r) = (t + r\delta t, x - d_x(t,x) + r\delta x)$ satisfies the predicate $\mathcal{R}_{\overline{\kappa}}$.*

We should, of course, prove that the test is "reasonable", namely, that it accepts at least "small" steps in the parameters leading to the standard overall complexity of the algorithm. This is the issue we are coming to.

4 Main results on tracing a surface

4.1 Acceptable stepsizes

Our main observation is that a stepsize r such that the displacement $r\delta t$ is not too large in the Euclidean metric defined by the matrix $\nabla_t^2 F(t,x)$ for sure passes the Basic Test. We start from the following simple

Proposition 4.1 *Given $u \in \mathrm{Dom}\,\Psi$ and $du \in \mathbf{R}^N$, let us set $s = \Psi'(u)$, so that $s \in \mathrm{Dom}\,\Psi_*$, and $ds = \Psi''(u)du$. Let also*

$$\begin{array}{rcl}\rho_u^*[du] & = & \Psi(u+du) + \Psi_*(s+ds) \\ & & -\Big[\Psi(u) + \Psi_*(s) + (du)^T\Psi'(u) + (ds)^T\Psi_*'(s) + \frac{1}{2}(du)^T\Psi''(u)du + \frac{1}{2}(ds)^T\Psi_*''(s)ds\Big]\end{array}$$

be the remainder in the second-order Taylor expansion of the function $\Psi + \Psi_*$ at the point (u,s) along the displacement (du, ds) (if this displacement moves the point outside the domain of the function, then $\rho^* = \infty$). Then

(i) $|du|^2_{\Psi''(u)} = |ds|^2_{\Psi_*''(s)} = (du)^T ds$,

(ii) *if $\chi \equiv |du|_{\Psi''(u)} < 1$, then $\rho_u^*[du]$ is well-defined and, moreover,*

$$\rho_u^*[du] \le 2\rho(\chi) - \chi^2, \tag{25}$$

($\rho(\cdot)$ is given by (8)) and

(iii) *The third derivative of the function $\Psi + \Psi_*$ taken at the point (u,s) along the direction (du, ds) is zero, so that $\rho_u^*(du)$ is in fact the remainder in the third order Taylor expansion of $\Psi + \Psi_*$ at the point (u,s) along the direction (du, ds).*

Proof. (i) is an immediate consequence of the relations $ds = \Psi''(u)du$ and $\Psi_*''(s) = [\Psi''(u)]^{-1}$. (i) combined with the upper bound on the remainder in the first-order Taylor expansion of a s.-c. function (Proposition 2.2.(i)) results in (ii). To verify (iii), let us differentiate the identity (y varies, h is fixed)

$$h^T\Psi_*''(\Psi'(y))h = h^T[\Psi''(y)]^{-1}h$$

at the point $y = u$ in the direction du, which results in

$$D^3\Psi_*(s)[h, h, \Psi''(u)du] = -D^3\Psi(u)[[\Psi''(u)]^{-1}h, [\Psi''(u)]^{-1}h, du];$$

substituting $h = \Psi''(u)du \equiv ds$, we come to $D^3\Psi_*(s)[ds, ds, ds] = -D^3\Psi(u)[du, du, du]$, as required in (iii). ∎

Theorem 4.1 *Let (t,x) satisfy $\mathcal{P}_\kappa$, $(\delta t, \delta x) \in \Pi(t,x)$, let*

$$R^* = \{S(p) = s(t,x) - d_s(t,x) + p\delta s \mid p \in \mathbf{R}\}$$

be the dual search line associated with the primal search line

$$R = \{X(p) = (t, x - d_x(t,x)) + p(\delta t, \delta x) \mid p \in \mathbf{R}\}$$

(see Lemma 3.2) and let

$$u = \sigma t + \pi x + \epsilon,\ du(r) = r\sigma\delta t + \pi[-d_x(t,x) + r\delta x].$$

Then

(i) *The vector* δx *is the minimizer of the quadratic form* $|\sigma\delta t + \pi h|^2_{\Psi''(u)}$ *over* $h \in \mathbf{R}^n$, *and, in particular,*

$$\zeta \equiv |\sigma\delta t + \pi\delta x|_{\Psi''(u)} \leq |\sigma\delta t|_{\Psi''(u)}. \tag{26}$$

(ii) *One has*

$$\chi(r) \equiv |du(r)|_{\Psi''(u)} = \sqrt{r^2\zeta^2 + \lambda^2(t,x)} \tag{27}$$

and

$$v(r) \equiv F(X(r)) + \Psi_*(S(r)) - [\sigma(t + r\delta t) + \epsilon]^T S(r) = \rho_u^*[du(r)]; \tag{28}$$

in particular,

$$\chi(r) < 1 \Rightarrow v(r) \leq 2\rho(\chi(r)) - \chi^2(r), \tag{29}$$

so that if

$$\omega_* = \max\left[\omega \mid 2\rho(\sqrt{\kappa^2 + \omega^2}) - \kappa^2 - \omega^2 \leq \overline{\kappa}\right] \tag{30}$$

then all stepsizes r *with*

$$|r| \leq \frac{\omega_*}{|\sigma\delta t|_{\Psi''(u)}} \tag{31}$$

for sure pass the Basic Test.

Proof. From now on the quantities related to F, Ψ and Ψ_*, if no argument values are explicitly indicated, are taken at the points (t, x), $u = \sigma t + \pi x + \epsilon$, $s = s(t, x) = \Psi'(u)$, respectively.
1^0. The minimizer of the quadratic form $|\sigma\delta t + \pi h|^2_{\Psi''} = [\sigma\delta t + \pi h]^T\Psi''[\sigma\delta t + \pi h]$ of $h \in \mathbf{R}^n$ is given by

$$\pi^T\Psi''[\sigma\delta t + \pi h] = 0;$$

this is exactly the equation defining δx, see (11), which proves the first statement of the theorem.
2^0. By definition of $d_x(t, x)$ (see (12)), $d_s(t, x)$ (see (19)) and the correspondence between δx and δs given by (19) we have

$$d_s(t, x) = \Psi''\pi d_x(t, x), \;\; \delta s = \Psi''[\sigma\delta t + \pi\delta x] \tag{32}$$

whence

$$ds(r) \equiv -d_s(t, x) + r\delta s = \Psi'' du(r). \tag{33}$$

By construction $s = \Psi'(u)$. From (23) we know that

$$|d_s(t, x)|_{\Psi''_*} = |\pi d_x(t, x)|_{\Psi''} = \lambda(t, x);$$

besides this, $[\pi d_x(t, x)]^T\Psi''[\sigma\delta t + \pi\delta x] = 0$ in view of (11), so that

$$\chi^2(r) \equiv |du(r)|^2_{\Psi''} = |-\pi d_x(t, x) + r[\sigma\delta t + \pi\delta x]\,|^2_{\Psi''} = \lambda^2(t, x) + r^2\zeta^2, \tag{34}$$

as claimed in (27).
3^0. We have (see the definitions of $X(r)$, $S(r)$, $du(r)$, $ds(r)$)

$$F(X(r)) + \Psi_*(S(r)) = \Psi(u + du(r)) + \Psi_*(s + ds(r)) =$$

[definition of $\rho_u^*[\cdot]$ combined with (33)]

$$= \Psi(u) + \Psi_*(s) + (du(r))^T\Psi' + (ds(r))^T\Psi'_* + \frac{1}{2}|du(r)|^2_{\Psi''} + \frac{1}{2}|ds(r)|^2_{\Psi''_*} + \rho_u^*[du(r)] =$$

[Proposition 4.1.(i) as applied to $du = du(r)$ combined with (33); relations $\Psi' = s, \Psi'_* = u$]

$$= u^T s + (du(r))^T s + (ds(r))^T u + |du(r)|^2_{\Psi''} + \rho^*_u[du(r)] =$$

$$= (u + du(r))^T (s + ds(r)) - (du(r))^T ds(r) + |du(r)|^2_{\Psi''} + \rho^*_u[du(r)] =$$

[Proposition 4.1.(i)]

$$= (u + du(r))^T (s + ds(r)) + \rho^*_u[du(r)]. \tag{35}$$

By construction,

$$(u + du(r))^T (s + ds(r)) = [\sigma(t + r\delta t) + \epsilon]^T S(r) + [x - d_x(t,x) + r\delta x]^T \pi^T S(r) =$$

[Lemma 3.2]

$$= [\sigma(t + r\delta t) + \epsilon]^T S(r),$$

so that (35) implies (28). Relation (29) follows from (25) and (28), and the concluding statement of the Theorem is a corollary of (26) and (29). ■

4.1.1 How long are "long steps"

Theorem 4.1 says that if the tolerances κ and $\overline{\kappa}$ are chosen reasonably (say, $\kappa = 0.125$ and $\overline{\kappa} = 2$) and (t,x) is κ-close to $\mathcal{S}$, then any step

$$(t,x) \mapsto (t^+, \widetilde{x}) = (t + \delta t, x - d_x(t,x) + \delta x)$$

"along the surface" (i.e., with $(\delta t, \delta x) \in \Pi(t,x)$) of "$O(1)$-local length", namely, with

$$\zeta \equiv |\sigma\delta t + \pi\delta x|_{\Psi''(\sigma t + \pi x + \epsilon)} \leq 0.89$$

results in a $\overline{\kappa}$-good forecast $(t^+, \widetilde{x})$ (indeed, for $\kappa = 0.125$ and $\zeta = 0.89$ the right hand side in (29) is < 2). The natural question is whether we could ensure a larger step, still resulting in a $\overline{\kappa}$-good forecast. For the sake of simplicity, let us answer this question for the case when the point (t,x) belongs to $\mathcal{S}$ (i.e., $\lambda(t,x) = 0$) rather than is κ-close to the surface (modifications required in the case of small positive $\lambda(t,x)$ are quite straightforward).

When answering the question, we can normalize the direction $(\delta t, \delta x) \in \Pi(t,x)$ to have unit local length:

$$|du|_{\Psi''} = 1, \quad du = \sigma\delta t + \pi\delta x \tag{36}$$

(here and in what follows all quantities related to Ψ, Ψ_* are evaluated at the points $u = \sigma t + \pi x + \epsilon$ and $s = \Psi'(u)$, respectively). Recall that we have associated with the data $(t, x, \delta t)$ the primal and the dual search lines R and R^*; it is convenient to aggregate these lines in a single "primal-dual" line

$$R_{pd} = (u,s) + \mathbf{R}(du, ds) \subset \mathbf{R}^{2N}_{pd} = \mathbf{R}^N_p \times \mathbf{R}^N_d,$$

where $ds = \Psi'' du$; the projection of R_{pd} onto the space $\mathbf{R}^N_d$ of dual variables is R^*, while the projection of R_{pd} onto the space $\mathbf{R}^N_p$ of the primal variables is the image of the primal search line R under the embedding $(\tau, y) \mapsto \sigma\tau + \pi y + \epsilon$. It is convenient to equip the primal-dual space $\mathbf{R}^{2N}_{pd}$ with the Euclidean norm

$$|(v_p, v_d)|_{pd} = \sqrt{|v_p|^2_{\Psi''} + |v_d|^2_{\Psi''_*}};$$

this is nothing but the local norm $|\cdot|_{\Xi''(z_0)}$ given by the Hessian of the s.-c. function

$$\Xi(v_p, v_d) = \Psi(v_p) + \Psi_*(v_d)$$

at the point $z_0 = (u, s)$.

Let us define T as the distance from z_0 to the boundary of the domain $\mathrm{Dom}\,\Xi = (\mathrm{Dom}\,\Psi) \times (\mathrm{Dom}\,\Psi_*)$ along the line R_{pd}:

$$T = \min\{|z - z_0|_{pd} \mid z \in R_{pd} \setminus \mathrm{Dom}\,\Xi\};$$

note that $T \geq 1$ due to Proposition 2.2.(i). Now, when choosing a stepsize r, forming the corresponding forecast and subjecting it to the Basic Test, we in fact generate and process the point $z_r = (u + rdu, s + rds)$ on the primal-dual search line R_{pd}, i.e., perform certain step along R_{pd}. The $|\cdot|_{pd}$-length of this step $|z_r - z_0|_{pd}$ is simply $|r|\sqrt{2}$ (indeed, $ds = \Psi'' du$, so that $|ds|_{\Psi_*''} = |du|_{\Psi''}$ by Proposition 4.1, while du is normalized by (36)). It follows that T is a natural upper bound on the acceptable step $|z_r - z_0|_{pd}$ – when $r = 2^{-1/2}T$ and du is "badly oriented", the stepsize r results in $z_r \notin \mathrm{Dom}\,\Xi$ and is therefore rejected by the Basic Test. With these remarks, the above question "how long are long steps" can be posed as follows:

Which fraction of T indeed is accepted by the Basic Test, i.e., which fraction of the way to the boundary of the primal-dual feasible set $\mathrm{Dom}\,\Xi$ along the direction (du, ds) can we cover in one step of the Basic Updating scheme ?

According to the above discussion, we for sure can move towards the boundary by the fixed distance $0.89\sqrt{2}$; this is a "short step" allowed in any path-following interior point method, and to get a result of this type, no structural assumption **(C)** on the s.-c.b. in question and no dual bounding are needed. If T is large, then a short step covers small part of the way to the boundary, and a short-step method becomes slow.

In fact our approach in many cases enables much larger steps:

Proposition 4.2 *Let both the barrier Ψ and its Legendre transformation Ψ_* be α-regular (Definition 2.1), let $(t, x) \in \mathcal{S}$ and let δt be such that (36) takes place. Then all stepsizes satisfying*

$$|r| \leq r^*(\alpha)\sqrt{T}, \tag{37}$$

same as all stepsizes satisfying

$$|r| \leq r^*(\alpha) T \vartheta_*^{-1/4} \tag{38}$$

are accepted by the Basic Test. Here $r^*(\alpha) > 0$ depends on α only and ϑ_* is the parameter of the s.-c.b. Ψ.

Proof. Let $R_{pd} = z_0 + \mathbf{R}dz$, $dz = (du, ds)$, be the primal-dual search line associated with $t, x, \delta t$, and let $\Delta = \{r \mid z_0 + 2^{-1/2} r dz \in \mathrm{cl}\,\mathrm{Dom}\,\Xi\}$, so that Δ is a closed convex set on the axis containing the segment $[-T, T]$. By Proposition 2.1, the function

$$\phi(r) = \Xi(z_0 + 2^{-1/2} r dz)$$

is self-concordant and α-regular on $\mathrm{int}\,\Delta$. Since we clearly have $\|1\|_{\mathrm{int}\,\Delta, r} \leq (T - |r|)^{-1}$ for $|r| < T$ (for notation, see Definition 2.1), inequality (6) implies that the function $\psi(r) \equiv \phi''(r)$ satisfies

$$|\psi''(r)| \leq \frac{\alpha(\alpha+1)}{(T - |r|)^2}\psi(r), \quad |r| < T. \tag{39}$$

Besides this, we have

$$\psi(0) = 1, \quad \psi'(0) = 0, \tag{40}$$

the first relation coming from $\psi(0) = |2^{-1/2}dz|^2_{pd} = 1$ (note that $|dz|_{pd} = \sqrt{2}$ due to the discussion preceding the Proposition), and the second relation being given by Proposition 4.1.

We claim that

$$0 \le \phi''(r) \equiv \psi(r) \le \frac{T^\alpha}{(T-|r|)^\alpha}, \ \ |r| < T. \tag{41}$$

The left inequality follows from convexity of ϕ. By symmetry reasons, it suffices to establish the right inequality for $0 \le r < T$. To this end note that the function

$$\omega(r) = \frac{T^\alpha}{(T-r)^\alpha}$$

clearly satisfies the relations

$$\omega''(r) = \frac{\alpha(\alpha+1)}{(T-r)^2}\omega(r), \ 0 \le r < T; \ \omega(0) = 1; \ \omega'(0) > 0. \tag{42}$$

From (42), (39), (40) we see that the function $\xi(r) = \omega(r) - \psi(r)$ satisfies the relations

$$\xi''(r) \ge \frac{\alpha(\alpha+1)}{(T-r)^2}\xi(r), \ 0 \le r < T; \quad \xi(0) = 0; \quad \xi'(0) > 0. \tag{43}$$

To establish (41) for $0 \le r < T$ is the same as to prove that $\xi \ge 0$ on $[0,T)$, which is immediate: since $0 = \xi(0) < \xi'(0)$, ξ is positive on $(0,\widehat{r})$ with some positive $\widehat{r}$. Consequently, if $\xi < 0$ somewhere on $[0,T)$, then ξ possesses a zero on $[\widehat{r},T)$. Let r^* be the smallest of zeros of ξ on $[\widehat{r},T)$; then ξ is nonnegative on $[0,r^*]$, whence, due to (43), ξ is convex on $[0,r^*]$. This observation combined with $\xi(0) = 0$, $\xi'(0) > 0$ contradicts the relation $\xi(r^*) = 0$.

Combining (41) and (39), we come to

$$|\phi^{(4)}(r)| \le \frac{\alpha(\alpha+1)T^\alpha}{(T-|r|)^{\alpha+2}}, \quad |r| < T. \tag{44}$$

According to Theorem 4.1.(ii) and Proposition 4.1.(ii), the quantity $v(2^{-1/2}r)$, $v(\cdot)$ being defined by (24), is the remainder in the third-order Taylor expansion of $\phi(\cdot)$ at the point $r = 0$. From (44) we therefore conclude that

$$v(r) \le \frac{\alpha(\alpha+1)T^\alpha r^4}{6(T-|r|\sqrt{2})^{\alpha+2}}, \quad \sqrt{2}|r| < T. \tag{45}$$

Since the Basic Test accepts all stepsizes with $v(r) \le \overline{\kappa}$, we see from (45) that it accepts all stepsizes satisfying (37), with properly chosen $r^*(\alpha)$ (recall that $T \ge 1$).

In view of already proved part of the statement, in order to demonstrate acceptability of the stepsizes (38) with properly chosen $r^*(\alpha)$, it suffices to verify that

$$T \le \gamma(\alpha)\sqrt{\vartheta_*}. \tag{46}$$

This inequality is an immediate consequence of the following

Lemma 4.1 *Let Ψ be α-regular ϑ_*-s.c.b. for convex domain G^+, let $u \in \mathrm{int}\, G^+$, let du be such that $|du|_{\Psi''(u)} = 1$, and let T be such that $u \pm T du \in G^+$. Then*

$$T \leq 2^{2+\alpha/2}\sqrt{\vartheta_*}. \tag{47}$$

Proof. Without loss of generality we may assume that $(du)^T\Psi'(u) \geq 0$ (otherwise we could replace du with $-du$). Let

$$\phi(r) = \Psi(u + rdu), \;\; r \in \Delta = \{r \mid \; u + rdu \in G^+\} \supset [-T, T];$$

by Proposition 2.1, ϕ is α-regular ϑ_*-s.-c.b. for Δ. We claim that if $r \in \mathrm{int}\,\Delta$ and $d > 0$ is such that $r \pm 2d \in \mathrm{int}\,\Delta$, then

$$0 \leq \phi''(r-d) \leq 2^{\alpha+1}\phi''(r). \tag{48}$$

Indeed, by Proposition 2.1 the function $\chi(s) = \phi(r-s) + \phi(r+s)$ is α-regular on the set $\Delta_r = \{s \mid \; r \pm s \in \mathrm{int}\,\Delta\} \supset [-2d, 2d]$ and is such that $\chi'(0) = \chi'''(0) = 0$. From these properties, same as above (cf. (41)), one can derive the inequality

$$\chi''(s) \leq \frac{(2d)^\alpha}{(2d-|s|)^\alpha}\chi''(0), \;\; |s| < 2d.$$

Substituting $s = d$, we get

$$\phi''(r-d) + \phi''(r+d) = \chi''(d) \leq 2^\alpha\chi''(0) = 2^{\alpha+1}\phi''(r),$$

which, in view of the convexity of ϕ, implies (48).

Now let $0 \leq r < T/3$. Applying (48) to $d = r$, we get

$$\phi''(r) \geq 2^{-\alpha-1}\phi''(0) = 2^{-\alpha-1}, \;\; 0 < r < T/3 \tag{49}$$

(the equality is given by $|du|_{\Psi''(u)} = 1$), whence, in view of $\phi'(0) = (du)^T\Psi'(u) \geq 0$,

$$\phi'(r) \geq 2^{-\alpha-1}r, \;\; 0 < r < T/3.$$

Now note that ϕ is ϑ_*-self-concordant barrier for Δ, whence, by Nesterov and Nemirovski (1994), Proposition 2.3.2, $\phi'(r)(s-r) \leq \vartheta_*$ for all $s \in \Delta$. Applying this inequality to $s = T \in \Delta$ and $r \in (0, T/3)$, we get $2^{-\alpha-1}r(T-r) \leq \vartheta_*$, $0 < r < T/3$, and (47) follows. ∎

Results similar to those stated by Proposition 4.2 were recently obtained in Nesterov and Todd (1994,1995) for the predictor-corrector interior point methods associated with the *self-scaled* cones. Note that the property of Ψ and Ψ_* to be α-regular seems to be less restrictive than the one imposed in Nesterov and Todd (1994,1995); we shall see in Section 6 that the property of 6-regularity is shared by the functions Ψ and Ψ_* responsible for Linear, Quadratically Constrained Quadratic and Semidefinite Programming (these applications are covered by the results of Nesterov and Todd (1994,1995) as well), same as by those responsible for Geometric Programming (where the results of just mentioned papers are inapplicable).

4.2 Centering property

To proceed, we need certain centering property of the surface of analytic centers given by the following

Proposition 4.3 *Let (t,x) satisfy $\mathcal{P}_\kappa$ with $\kappa \le 0.125$, and let $t' \in \mathrm{cl}\,T$. Then*

$$[\Psi'(\sigma t + \pi x + \epsilon)]^T \sigma(t'-t) \le (1+6\kappa)\vartheta_* + 4\kappa. \tag{50}$$

Proof. By evident reasons, it suffices to consider the case $t' \in T$. Let $x' \in Q_{t'}$. From general properties of s.-c.b.'s (Nesterov and Nemirovski (1994), Proposition 2.3.2) it follows that

$$(\nabla F(t,x))^T (t'-t, x'-x) \le \vartheta_*,$$

so that (in what follows the derivatives of Ψ are taken at $\sigma t + \pi x + \epsilon$)

$$(\Psi')^T [\sigma(t'-t) + \pi(x'-x)] \le \vartheta_*.$$

Therefore

$$(\Psi')^T \sigma(t'-t) \le \vartheta_* + (\Psi')^T \pi(x-x') = \vartheta_* + (\nabla_x F(t,x))^T (x-x'). \tag{51}$$

On the other hand, let x^* be the minimizer of $F(t,\cdot)$, and let

$$|h| = (h^T \nabla_x^2 F(t,x^*) h)^{1/2}, \ \ |h|_* = (h^T [\nabla_x^2 F(t,x^*)]^{-1} h)^{1/2}.$$

From the relation $\lambda(t,x) \le \kappa \le 0.125$ it follows (see Nesterov and Nemirovski (1994), Theorem 2.2.2 and Proposition 2.3.2) that

$$|x - x^*| \le 1, \ \ |x' - x^*| \le 1 + 3\vartheta_*, \ \ |\nabla_x F(t,x)|_* \le 2\kappa.$$

Therefore the concluding expression in (51) does not exceed $|x-x'||\nabla_x F(t,x)|_* \le 2\kappa(3\vartheta_*+2) + \vartheta_*$, and we come to (50). ■

What we are interested in is the following consequence of the latter theorem:

Corollary 4.1 *Let $F(t,x)$ be the barrier associated with a surface of the type $\mathcal{S}_k(c_1,...,c_k)$:*

$$F(t,x) = -\vartheta \sum_{i=1}^k \ln(t_i - c_i^T x) + F(x),$$

where F is a ϑ-s.-c.b. for a closed and bounded convex domain $G \subset \mathbf{R}^n$, and let (t,x) satisfy $\mathcal{P}_\kappa$ with some $\kappa \le 0.125$. If $t' \le t$ belongs to $\mathrm{cl}\,T$, then

$$\sum_{i=1}^k \frac{t_i - t_i'}{\Delta_i} \le 2(k+1), \ \ \Delta_i = \Delta_i(t,x) = t_i - c_i^T x. \tag{52}$$

Geometrically: The part of $\mathrm{cl}\,T$ "to the left of t", i.e., comprised of $t' \le t$, belongs to the simplex

$$\Delta(t,x) = \{t' \le t \mid \sum_{i=1}^k \frac{t_i - t_i'}{\Delta_i} \le 2(k+1)\}$$

and contains the box

$$C(t,x) = \{t' \le t \mid \frac{t_i - t_i'}{\Delta_i} \le 1, \ i = 1,...,k\}.$$

In particular, if

$$t_i^*(t) = \inf\{c_i^T y \mid y \in G, \ c_j^T y \le t_j, \ j = 1,...,k\},$$

then, for every $i \le k$,

$$t_i - t_i^*(t) \le 2(k+1)\Delta_i(t,x). \tag{53}$$

Proof. In the case in question the left hand side in (50) is equal to $\vartheta \sum_{i=1}^k (t_i - t_i')\Delta_i^{-1}$, while the right hand side is $(1+6\kappa)\vartheta_* + 4\kappa \le 1.75(k+1)\vartheta + 0.5 \le 2\vartheta(k+1)$ (recall that $\vartheta \ge 1$, see Definition 2.1). Thus, (50) implies (52). The inclusion $C(t,x) \subset \mathrm{cl}\,T$ is evident, since for $t' \in \mathrm{int}\ C(t,x)$ we simply have $c_i^T x < t_i'$, $i = 1,...,k$, so that $(t',x) \in Q = \{(\tau,y) \mid y \in \mathrm{int}\ G, \ \tau_i > c_i^T y, \ i = 1,...,k\}$. Relation (53) is an immediate consequence of the preceding statements of the Corollary. ■

5 Solving convex programs via tracing surfaces

To the moment we know what are our local abilities to trace a surface of analytic centers, but did not discuss the "strategy" - where to move in order to solve the problem the surface is associated with. This question does not occur in the usual path-following approach, since there is a unique reasonable strategy: to vary the parameter in the only direction of interest at the highest possible rate compatible with the restriction on the Newton complexity of the corrector steps. In the multi-parameter case the strategy of tracing the surface requires special investigation; this is the issue we are coming to.

We intend to apply our surface-following scheme to convex programs

$$\begin{array}{ll}(P) & \textit{minimize } c^T x \textit{ s.t. } x \in G,\\ (P') & \textit{minimize } c^T x \textit{ s.t. } x \in G,\ f^T x \le 0;\end{array}$$

in both of the problems, G is a closed and bounded convex domain in $\mathbf{R}^n$ represented by a ϑ-s.-c.b. F, and we are given a starting point $\widehat{x} \in \operatorname{int} G$. In the second problem it is assumed that the quantity

$$f^* = \min_{x\in G} f^T x \tag{54}$$

is nonnegative (the case of feasible (P') clearly corresponds to the case of $f^* = 0$).

To make presentation more compact, we shall focus on (evidently more general) problem (P'); to get constructions and results for (P), it suffices to set in what follows $f = 0$.

In Section 2.3 problem (P') was associated with the barrier

$$F(t,x) = -\vartheta \ln(t_1 - c^T x) - \vartheta \ln(t_2 - f^T x) - \vartheta \ln(t_3 - d^T x) + F(x) \tag{55}$$

and the 3-parameter surface $\mathcal{S}_3(c,f,d)$; here d is readily given by the requirement that the pair $(\widehat{t}, \widehat{x})$, with certain explicit $\widehat{t}$, belongs to the surface. In what follows we deal with the setup

$$\begin{array}{rcl}\widehat{t}_1 &=& c^T\widehat{x} + \left[c^T[\nabla^2 F(\widehat{x})]^{-1}c\right]^{1/2},\\ \widehat{t}_2 &=& f^T\widehat{x} + \left[f^T[\nabla^2 F(\widehat{x})]^{-1}f\right]^{1/2},\\ d &=& -\vartheta^{-1}\nabla F(\widehat{x}) - [\widehat{t}_1 - c^T\widehat{x}]^{-1}c - [\widehat{t}_2 - f^T\widehat{x}]^{-1}f,\\ \widehat{t}_3 &=& d^T\widehat{x} + 1;\end{array} \tag{56}$$

Note that $\widehat{t}_j$, $j = 1,2$, are nothing but the maxima of the linear forms $c^T x$, respectively, $f^T x$, over the closed Dikin ellipsoid $W_F(\widehat{x})$, see Proposition 2.2; according to this Proposition, the ellipsoid is contained in G, so that

$$\widehat{t}_1 \le \max_{x\in G} c^T x; \quad \widehat{t}_2 \le \max_{x\in G} f^T x. \tag{57}$$

Below c^* denotes the optimal value in the problem in question. When measuring accuracy of an approximate solution $x \in G$, we normalize the residuals $c^T x - c^*$ and $f^T x$ by the variations of the corresponding linear forms on the domain of the problem, the variation of a linear form $e^T x$ on a bounded set U being defined as

$$\mathrm{V}_U(e) = \max_{x\in U} e^T x - \min_{x\in U} e^T x.$$

When solving (P') via tracing the surface $\mathcal{S}_3(c,f,d)$, our goal is to enforce the "objective parameter" t_1 and the "constraint parameter" t_2 to converge to the optimal value c^* and to 0, respectively. As for the "centering parameter" t_3, all we need is to control it in a way which allows us to achieve the indicated goals, and a reasonable policy is to push the parameter to ∞, since with a "small" value τ of the centering parameter the artificial constraint $d^T x \le \tau$ may vary the optimal value in the problem.

5.1 Assumption on the structure of F

In order to use the long-step tactics presented in Section 3, from now on we make the following assumption on the structure of the barrier F for the domain G:

$\mathcal{Q}$: *we are given a closed convex domain* $H \in \mathbf{R}^M$, *a* ϑ*-s.-c.b.* Φ *for* H *such that* Φ'' *is nondegenerate and the Legendre transformation* Φ_* *is known, and an affine mapping* $B(x) = \pi x + \epsilon : \mathbf{R}^n \to \mathbf{R}^M$ *with the image of the mapping intersecting* int H, *such that*

$$G = B^{-1}(H),\ F(x) = \Phi(B(x)).$$

Note that under this assumption any barrier of the type

$$F(t,x) = \vartheta \sum_{i=1}^{k} \psi_i(t_i - c_i^T x) + F(x),\ \psi_i(s) = -\ln s, \tag{58}$$

and, in particular, the barrier underlying the surface $\mathcal{S}_3(c,f,d)$, satisfies assumptions (A) - (C) from sections 2.5 and 3.1. The corresponding data are

$$G^+ = \mathbf{R}_+^k \times H,\ \Psi(u_1,...,u_k,u) = \vartheta \sum_{i=1}^{k} \psi_i(u_i) + \Phi(u),\ \vartheta_* = (k+1)\vartheta,$$

$$\mathcal{A}(t,x) = (t_1 - c_1^T x, ..., t_k - c_k^T x, B(x));$$

the Legendre transformation of Ψ is

$$\Psi_*(s_1,...,s_k,s) = \vartheta \sum_{i=1}^{k} \psi_i(-s_i) + \Phi_*(s) + k\vartheta(\ln\vartheta - 1),$$

$$\mathrm{Dom}\,\Psi_* = \{(s_1,...,s_k) < 0\} \times \mathrm{Dom}\,\Phi_*.$$

5.2 Preliminary remarks

In what follows we speak about tracing the surface $\mathcal{S}_3(c,f,d)$. Sometimes we write c_i instead of the i-th vector identifying the surface (so that $c_1 = c$, $c_2 = f$, $c_3 = d$).

Let us fix the tolerances κ, $\overline{\kappa}$ such that

$$\kappa \in (0, 0.125];\ \overline{\kappa} > 2\rho(\kappa) - \kappa^2.$$

When tracing the surface $\mathcal{S}_3$, we form a sequence of pairs (t^i, x^i) satisfying the predicate $\mathcal{P}_\kappa$ associated with the surface. To update the pair (t^{i-1}, x^{i-1}) into the new pair (t^i, x^i), we use the Basic Updating Scheme equipped with the Basic Test for choosing a proper stepsize r_i in the current direction $(\delta t^i, \delta x^i)$, so that the forecast we use is

$$(t^i, \widetilde{x}^i) = (t^{i-1}, x^{i-1} - d_x(t^{i-1}, x^{i-1})) + r_i(\delta t^i, \delta x^i).$$

The above remarks specify the method up to the following two "degrees of freedom":

(I) *strategy of choosing the direction* δt^i*;*
(II) *tactics of choosing a proper stepsize* r_i*.*

(I) is the main subject of this section. As about (II), we are not going to be too specific. The only assumption on r_i is that it is *at least the "short step"* r_i^* *given by Theorem 4.1*, i.e. (see (31)),

$$r_i \geq r_i^* \equiv \frac{\omega_*}{\sqrt{\vartheta \sum_{j=1}^{3} (\delta t_j^i)^2 / \Delta_j^2(t^{i-1}, x^{i-1})}}, \quad \Delta_j(t,x) = t_j - c_j^T x \tag{59}$$

(from now on, the superscript in notation like δt_j^i denotes the number of the step, and the subscript is the coordinate index). In view of the origin of ω_* (see (30)) and Theorem 4.1 (applied to barrier (55)), the default value r_i^* of the stepsize for sure passes the Basic Test (so that to use the default stepsize no Basic Test, and, consequently, no assumptions on the structure of F are needed). One is welcome to combine the Basic Test with any kind of line search to get a larger (proper) stepsize. Note that in what follows we sometimes impose certain "safety" upper bounds on r_i; each time it can be immediately verified that these bounds are valid for $r_i = r_i^*$, so that the safety bounds are consistent with the aforementioned lower bound on r_i.

With the above remarks, we may completely focus on the "strategic" issue (I).

As it was already explained, when tracing surface $\mathcal{S}_3(c,f,d)$, we should get rid of the centering parameter; the simplest way is to push it to ∞. How to ensure this, this is the issue we start with.

Consider a surface $\mathcal{S}_k(c_1,...,c_k)$ associated with barrier (58), and let (t,x) satisfy the predicate $\mathcal{P}_\kappa$ associated with the surface. Let

$$q = \frac{2k}{4k+3}. \tag{60}$$

We say that a direction $\delta t = (\delta t_1, ..., \delta t_k)$ in the space of parameters is *k-safe*, if

$$\delta t_k \geq 0, \frac{\delta t_k}{\Delta_k(t,x)} \geq 2(2k+1)\frac{-\delta t_i}{\Delta_i(t,x)}, \; i = 1,...,k-1, \tag{61}$$

where, same as in Corollary 4.1,

$$\Delta_i(t,x) = t_i - c_i^T x, \; i = 1,...,k,$$

and we say that a stepsize $r \geq 0$ in the direction δt is *k-safe*, if

$$t + q^{-1} r \delta t \in T \equiv \{t \in \mathbf{R}^k \mid \exists y \in \operatorname{int} G : t_i > c_i^T y, \, i = 1,...,k\}. \tag{62}$$

Lemma 5.1 *Let* (t,x) *satisfy* $\mathcal{P}_\kappa$ *with respect to* $\mathcal{S}_k(c_1,...,c_k)$*,* δt *be a safe direction,* r *be a safe stepsize, and let* $t^+ = t + r\delta t$*. Then*

$$t_k^+ - t_k^*(t^+) \geq \min\{1 + \frac{1}{2k+2}; 1 + \frac{r\delta t_k}{4(k+1)\Delta_k(t,x)}\}(t_k - t_k^*(t)), \tag{63}$$

where, same as in Corollary 4.1, $t_i^*(t) = \min\{c_i^T y \mid y \in G, \, c_j^T y \leq t_j, \, j = 1,...,k\}$*.*

Proof is given in Appendix.

5.3 How to trace $\mathcal{S}_3(c,f,d)$

5.3.1 The algorithm

Our strategy for solving (P') is as follows. Given (t,x) κ-close to $\mathcal{S} \equiv \mathcal{S}_3(c,f,d)$, we say that (t,x) is *good*, if

$$t_2 \le 16\Delta_2(t,x) \tag{64}$$

and is *bad* otherwise; here and in what follows, as always,

$$\Delta_1(t,x) = t_1 - c^T x,\ \Delta_2(t,x) = t_2 - f^T x,\ \Delta_3(t,x) = t_3 - d^T x.$$

At step i (where $z^{i-1} \equiv (t^{i-1}, x^{i-1})$ is updated into $z^i \equiv (t^i, x^i)$), we act as follows.

1) If the current iterate (t^{i-1}, x^{i-1}) is bad, we apply

Infeasibility Test: *If both the inequalities*

$$\Delta_1(t^{i-1}, x^{i-1}) > 9\vartheta_* \left(c^T[\nabla_x^2 F(t^{i-1}, x^{i-1})]^{-1} c\right)^{1/2} \tag{65}$$

and

$$\Delta_3(t^{i-1}, x^{i-1}) > 9\vartheta_* \left(d^T[\nabla_x^2 F(t^{i-1}, x^{i-1})]^{-1} d\right)^{1/2} \tag{66}$$

are valid, claim that (P') is infeasible and terminate.

2) If the method is not terminated by 1), we set

$$\delta t^i = \begin{cases} (-\Delta_1(z^{i-1}), & -\Delta_2(z^{i-1}), & 16\Delta_3(z^{i-1})), & z^{i-1} \text{ is good} \\ (\ \ \Delta_1(z^{i-1}), & 0\ \ \ , & 16\Delta_3(z^{i-1})), & \text{otherwise} \end{cases}. \tag{67}$$

After the direction δt^i is determined, we use the Basic Updating Scheme to update (t^{i-1}, x^{i-1}) into (t^i, x^i); in this updating, we subject the stepsize r_i to the "safety restriction"

$$r_i |\delta t_j^i| \le \frac{1}{8}\Delta_j(t^{i-1}, x^{i-1}),\ j = 1, 2 \tag{68}$$

(this does not forbid "long steps": the short step r_i^* in our case is by factor $O(\sqrt{\vartheta})$ less than the upper bound in (68)).

The recurrence is started at $(t^0, x^0) = (\widehat{t}, \widehat{x}) \in \mathcal{S}_3$, see (56).

Remark 5.1 In the case of problem (P) – according to our convention, it means that $f = 0$ – all (t^{i-1}, x^{i-1}) are good, since $t_2^{i-1} = \Delta_2(t^{i-1}, x^{i-1})$, so that (67) always results in $\delta t_1^i < 0$: we decrease the parameter of interest, as it should be in the case of feasible start.

5.3.2 Complexity

To describe the rate of convergence of the resulting method, let us denote by N^* the index of the iteration where the method terminates (if it happens; otherwise $N^* = \infty$), and let $N(\varepsilon)$, $0 < \varepsilon \le 1$, be the index i of the first iteration starting with which all iterates are ε-solutions to (P'):

$$c^T x^j - c^* \le \varepsilon \mathrm{V}_G(c), \quad f^T x^j \le \varepsilon \mathrm{V}_G(f),\ \forall j,\ N^* \ge j > i$$

(note that in the case of $N^* < \infty$ the latter relations are for sure satisfied when $i = N^*$, so that $N(\varepsilon) \le N^*$). The efficiency estimate for the presented method is given by the following.

Theorem 5.1 *The method never claims a feasible problem (P') to be infeasible, and if (P') is infeasible, this is detected in no more than*

$$N^* = O(1)\omega_*^{-1}\sqrt{\vartheta}\ln\left(\frac{2\vartheta[G:\widehat{x}](\mathrm{V}_G(f)+f^*)}{f^*}\right) \tag{69}$$

iterations; here ω_* is given by (30), f^* is given by (54) and

$$[G:\widehat{x}] = \max\{s \mid \exists y \notin G : \widehat{x} + s(\widehat{x} - y) \in G\}$$

is the asymmetry coefficient of G with respect to $\widehat{x}$.

If (P) is feasible, then

$$N(\varepsilon) \le O(1)\omega_*{}^{-1}\sqrt{\vartheta}\ln\left(\frac{2\vartheta[G:\widehat{x}]}{\varepsilon}\right) \quad \forall\varepsilon \in (0,1). \tag{70}$$

The Newton complexity of any corrector step of the method does not exceed

$$O(1)\{\overline{\kappa} + \ln\ln(1/\kappa)\}.$$

Proof. 1^0. The upper bound on the Newton complexity of corrector steps is given by Propositions 2.3 and 3.1.

2^0. Note that (59), (67) and (68) result in

$$\frac{1}{8} \ge r_i \ge O(1)\omega_*\vartheta^{-1/2}. \tag{71}$$

Let, same as before,

$$G_t = \{x \in G \mid c^T x \le t_1, f^T x \le t_2, d^T f \le t_3\}, \quad T = \{t \in \mathbf{R}^3 \mid \mathrm{int}\ G_t \ne \emptyset\}$$

and

$$\begin{array}{rcl} t_1^*(t) &=& \min\{c^T x \mid x \in G, f^T x \le t_2, d^T x \le t_3\} \\ t_2^*(t) &=& \min\{f^T x \mid x \in G, c^T x \le t_1, d^T x \le t_3\} \\ t_3^*(t) &=& \min\{d^T x \mid x \in G, c^T x \le t_1, f^T x \le t_2\} \end{array}, \quad t \in T.$$

3^0. Let us prove that if the method terminates at certain step i, then (P') is infeasible. Let x be an arbitrary point of the domain $G_{t^{i-1}}$, let x^* be the minimizer of $F(t^{i-1},\cdot)$, and let $H_* = \nabla_x^2 F(t^{i-1}, x^*)$, $H = \nabla_x^2 F(t^{i-1}, x^{i-1})$. As we know, $F(t^{i-1},\cdot)$ is ϑ_*-s.-c.b. for $G_{t^{i-1}}$ and $\lambda(t^{i-1}, x^{i-1}) \le \kappa \le 0.125$; from these observations by Nesterov and Nemirovski (1994), Theorem 2.2.2.(iii) and Proposition 2.3.2, it follows that

$$|x^* - x|_{H_*} \le 3\vartheta_* + 1, \quad |x^{i-1} - x^*|_H \le 2\kappa, \quad H_* \le (1-2\kappa)^{-2} H,$$

whence $|x - x^{i-1}|_H \le 9\vartheta_*$, $x \in G_{t^{i-1}}$. Therefore for an arbitrary vector e we have

$$\max_{x \in G_{t^{i-1}}} e^T x \le e^T x^{i-1} + 9\vartheta_* |e|_{H^{-1}}.$$

Applying this relation to $e = c$ and $e = d$ and taking into account (65) and (66), we see that

$$t_1^{i-1} > \max_{x \in G_{t^{i-1}}} c^T x, \quad t_3^{i-1} > \max_{x \in G_{t^{i-1}}} d^T x.$$

Thus, the constraints $c^T x \le t_1^{i-1}$ and $d^T x \le t_3^{i-1}$ are redundant in the description of $G_{t^{i-1}}$, whence

$$t_2^*(t^{i-1}) = \min\{f^T x \mid x \in G\}.$$

By Corollary 4.1 applied with $k = 3$, we have (see (53))

$$t_2^{i-1} - t_2^*(t^{i-1}) \le 8\Delta_2(t^{i-1}, x^{i-1});$$

since the Infeasibility Test was applied at the step i, the pair (t^{i-1}, x^{i-1}) is bad, so that $t_2^{i-1} > 16\Delta_2(t^{i-1}, x^{i-1})$, and we come to

$$\min_{x\in G} f^T x = t_2^*(t^{i-1}) \ge t_2^{i-1} - 8\Delta_2(t^{i-1}, x^{i-1}) > 8\Delta_2(t^{i-1}, x^{i-1}) > 0,$$

and (P') is infeasible, as claimed.

The algorithm in question terminates when the Infeasibility Test detects that (P') is infeasible. In what follows it is however more convenient to think that we ignore the "reports on infeasibility", if any, and continue the process as if there were no Infeasibility Test at all.

4^0. Note that all directions δt^i satisfy (61) with $k = 3$. Besides this, (68) ensures that the stepsizes r_i are 3-safe. Indeed, in our case $q^{-1} = 5/2$, so that by (68) we have for $j = 1, 2$:

$$t_j^{i-1} + q^{-1} r_i \delta t_j^i \ge t_j^{i-1} - \Delta_j(t^{i-1}, x^{i-1}) = \begin{cases} c^T x^{i-1}, & j = 1 \\ f^T x^{i-1}, & j = 2 \end{cases};$$

since $\delta t_3^i > 0$, we also have $t^{i-1} + q^{-1} r_i \delta t_3^i > d^T x^{i-1}$, whence $t^{i-1} + q^{-1} r_i \delta t^i \in T$.

5^0. Applying Lemma 5.1 and taking into account (71), we observe that

$$t_3^i - t_3^*(t^i) \ge (1 + r_i)(t_3^{i-1} - t_3^*(t^{i-1})) \ge (1 + O(1)\omega_* \vartheta^{-1/2})(t_3^{i-1} - t_3^*(t^{i-1})). \tag{72}$$

Let us derive from this observation that there exists the first moment i, let it be called i^*, when $t_3^i \ge \max_{x\in G} d^T x$, and that

$$i^* \le O(1)\omega_*^{-1} \vartheta^{1/2} \ln(2\vartheta[G\!:\!\widehat{x}]). \tag{73}$$

To derive (73) from (72), it clearly suffices to verify that

$$\mathrm{V}_G(d) \le 14\vartheta[G\!:\!\widehat{x}](\widehat{t}_3 - t_3^*(\widehat{t})). \tag{74}$$

To get (74), note that by (57) the domain

$$G_{\rm ini} = \{x \in G \mid c^T x \le \widehat{t}_1, \; f^T x \le \widehat{t}_2\}$$

contains the closed Dikin ellipsoid

$$W = \{y \mid |y - \widehat{x}|_{F''(\widehat{x})} \le 1\}.$$

Let $\widehat{G} = G \cap (2\widehat{x} - G)$ be the symmeterization of G with respect to $\widehat{x}$; by definition of the quantity $[G\!:\!\widehat{x}]$, we have

$$G \subset \widehat{x} + [G\!:\!\widehat{x}](\widehat{G} - \widehat{x}). \tag{75}$$

On the other hand, the function

$$\widehat{F}(x) = F(x) + F(2\widehat{x} - x)$$

is $\widehat{\vartheta}$-s.-c.b. for $\widehat{G}$, $\widehat{\vartheta} = 2\vartheta$ (Proposition 2.1.(i)), and clearly $\nabla\widehat{F}(\widehat{x}) = 0$. From the latter inequality it follows (Nesterov and Nemirovski (1994), Theorem 2.2.2.(iii)) that $\widehat{G}$ is contained in $|\cdot|_{\widehat{F}''(\widehat{x})}$-ball of the radius $1+3\widehat{\vartheta} \le 7\vartheta$ centered at $\widehat{x}$, whence

$$\widehat{G} - \widehat{x} \subset 7\vartheta(W - \widehat{x});$$

combining this relation with (75), we get

$$G \subset \widehat{x} + 7\vartheta[G:\widehat{x}](W - \widehat{x}).$$

It follows that for an arbitrary vector e one has

$$\mathrm{V}_G(e) \le 7\vartheta[G:\widehat{x}]\mathrm{V}_W(e) = 14\vartheta[G:\widehat{x}](e^T\widehat{x} - \min_{x\in W} e^T x). \tag{76}$$

Taking into account that

$$\widehat{t}_3 - t_3^*(\widehat{t}) = \widehat{t}_3 - \min\{d^T x \mid x \in G,\ c^T x \le \widehat{t}_1,\ f^T x \le \widehat{t}_2,\ d^T x \le \widehat{t}_3\} =$$

$$= \widehat{t}_3 - \min\{d^T x \mid x \in G, c^T x \le \widehat{t}_1,\ f^T x \le \widehat{t}_2\} = \widehat{t}_3 - \min_{x\in G_{\mathrm{ini}}} d^T x \ge$$

[since $\widehat{t}_3 \ge d^T\widehat{x}$ and $W \subset G_{\mathrm{ini}}$]

$$\ge d^T\widehat{x} - \min_{x\in W} d^T x$$

and applying (76) to $e = d$, we come to (74).
6^0. Let

$$\Omega_i = \frac{t_1^i - t_1^*(t^i)}{(t_2^i)^{32}}.$$

Our key argument is as follows: for properly chosen $O(1)$ and all i one has

$$\Omega_i/\Omega_{i-1} \ge 1 + O(1)\omega_*\vartheta^{-1/2}. \tag{77}$$

To establish the inequality, let us fix i and consider separately the cases of good and bad (t^{i-1}, x^{i-1}).

6^0.1. Assume that (t^{i-1}, x^{i-1}) is bad. According to (67), in the case in question

$$t_1^i = t_1^{i-1} + r_i\Delta_1(t^{i-1}, x^{i-1}), \quad t_2^i = t_2^{i-1}, \quad t_3^i \ge t_3^{i-1}.$$

Since $t_1^*(t)$ clearly depends only on t_2, t_3 and is nonincreasing in (t_2, t_3), we have $t_1^*(t^i) \le t_1^*(t^{i-1})$, whence, in view of $t_2^i = t_2^{i-1}$,

$$\frac{\Omega_i}{\Omega_{i-1}} = \frac{t_1^i - t_1^*(t^i)}{t_1^{i-1} - t_1^*(t^{i-1})} \ge \frac{t_1^{i-1} + r_i\Delta_1(t^{i-1}, x^{i-1}) - t_1^*(t^{i-1})}{t_1^{i-1} - t_1^*(t^{i-1})} = 1 + r_i\frac{\Delta_1(t^{i-1}, x^{i-1})}{t_1^{i-1} - t_1^*(t^{i-1})}.$$

According to Corollary 4.1 applied with $k = 3$ (see (53)), the concluding quantity is $\ge 1 + \frac{1}{8}r_i$. Taking into account (71), we come to (77).

6^0.2. Now assume that (t^{i-1}, x^{i-1}) is good. For the sake of brevity, let us write (t, x) instead of (t^{i-1}, x^{i-1}), r instead of r_i, and let $t^+ = t^i$. By definition of $t_1^*(\cdot)$, there exists $u \in G$ such that

$$c^T u = t_1^*(t); \quad f^T u \le t_2; \quad d^T u \le t_3; \tag{78}$$

by definition of $\Delta_j(\cdot,\cdot)$ and since $x \in G_t$, we have

$$c^T x = t_1 - \Delta_1(t,x); \quad f^T x = t_2 - \Delta_2(t,x); \quad d^T x \le t_3, \tag{79}$$

and since (t,x) is good, we have

$$t_2 \le 16\Delta_2(t,x). \tag{80}$$

Let

$$v = (1-r)u + rx,$$

and let us verify that $v \in G_{t^+}$. Indeed, by (78) - (79) and due to $t_3^+ \ge t_3$ we have $d^T v \le t_3^+$. Further, by (78), (79) and (67)

$$f^T v = (1-r)f^T u + r f^T x \le (1-r)t_2 + r(t_2 - \Delta_2(t,x)) = t_2 - r\Delta_2(t,x) = t_2^+.$$

Last, by (78) and (79)

$$c^T v = (1-r)c^T u + r c^T x \le c^T x = t_1 - \Delta_1(t,x) \le t_1^+$$

(concluding inequality is given by (68)).

Thus, $v \in G_{t^+}$; consequently,

$$t_1^*(t^+) \le c^T v = (1-r)c^T u + r c^T x = (1-r)t_1^*(t) + r(t_1 - \Delta_1(t,x)). \tag{81}$$

With this inequality, we have

$$t_1^+ - t_1^*(t^+) = t_1 - r\Delta_1(t,x) - t_1^*(t^+) \ge t_1 - r\Delta_1(t,x) - (1-r)t_1^*(t) - r(t_1 - \Delta_1(t,x)) =$$

$$= (1-r)(t_1 - t_1^*(t)),$$

whence

$$\frac{\Omega_i}{\Omega_{i-1}} \ge (1-r)\left(\frac{t_2}{t_2^+}\right)^{32} = (1-r)\left(\frac{t_2}{t_2 - r\Delta_2(t,x)}\right)^{32} \ge$$

[see (80)]

$$\ge (1-r)\left(\frac{1}{1-\frac{r}{16}}\right)^{32} \ge (1-r)\left(1+\frac{r}{16}\right)^{32} \ge (1-r)(1+2r) \ge 1 + \frac{3}{4}r$$

(we have taken into account that $r \le 1/8$, see (71)); this inequality combined with (71) results in (77).

7^0. Let (P') be infeasible, and let us prove (69). Indeed, we have $t_2^i \ge f^* > 0$, whence $\Omega_i \le \Delta_1(t^i,x^i)(f^*)^{-32}$. We conclude from (77) that

$$\Delta_1(t^i,x^i) \ge (1 + O(1)\omega_*\vartheta^{-1/2})^i \Delta_1(\widehat{t},\widehat{x}) \left(\frac{f^*}{\Delta_2(\widehat{t},\widehat{x})}\right)^{32}.$$

Taking into account that $\Delta_2(\widehat{t},\widehat{x}) \le \mathrm{V}_G(f)$ due to $\widehat{t}_2 \le \max_{x\in G} f^T x$ (see (57)) and that

$$\Delta_1(\widehat{t},\widehat{x}) \ge (14\vartheta[G:\widehat{x}])^{-1}\mathrm{V}_G(c)$$

(apply (76) to $e = c$ and note that $c^T\widehat{x} - \min_{x\in W} c^T x = \widehat{t}_1 - c^T\widehat{x}$ by (56)), we come to

$$\Delta_1(t^i, x^i) \geq (1 + O(1)\omega_*\vartheta^{-1/2})^i \frac{\mathrm{V}_G(c)}{14\vartheta[G\!:\!\widehat{x}]} \left(\frac{f^*}{\mathrm{V}_G(f)}\right)^{32}.$$

It follows that

$$i \geq i_c = O(1)\omega_*^{-1}\sqrt{\vartheta}\ln\left(\frac{2\vartheta[G\!:\!\widehat{x}]\mathrm{V}_G(f)}{f^*}\right) \Rightarrow \Delta_1(t^i, x^i) \geq 5\vartheta_*\mathrm{V}_G(c). \tag{82}$$

Further, from (53), (72) and (74) we have

$$\Delta_3(t^i, x^i) \geq \frac{1}{8}(t_3^i - t_3^*(t_i)) \geq \frac{1}{8}(1 + O(1)\omega_*\vartheta^{-1/2})^i(\widehat{t}_3 - t_3^*(\widehat{t})) \geq$$

$$\geq (1 + O(1)\omega_*\vartheta^{-1/2})^i(112\vartheta[G\!:\!\widehat{x}])^{-1}\mathrm{V}_G(d),$$

whence

$$i \geq i_d = O(1)\omega_*^{-1}\sqrt{\vartheta}\ln(2\vartheta[G\!:\!\widehat{x}]) \Rightarrow \Delta_3(t^i, x^i) \geq 5\vartheta_*\mathrm{V}_G(d). \tag{83}$$

Last, $t_2^i \leq t_2^{i-1}$ for all i, and if (t^{i-1}, x^{i-1}) is good, then

$$t_2^i = (1 - r_i\Delta_2(t^{i-1}, x^{i-1})/t_2^{i-1})t_2^{i-1} \geq (1 - r_i/16)t_2^{i-1} \geq (1 - O(1)\omega_*\vartheta^{-1/2})t_2^{i-1}$$

(we have used (71)). We clearly have $t_2^i \geq f^*$, while $\widehat{t}_2 \leq f^* + \mathrm{V}_G(f)$ by (57). Combining these observations, we see that the total number i_f of those i with good (t^{i-1}, x^{i-1}) can be bounded as follows:

$$i_f \leq O(1)\omega_*^{-1}\sqrt{\vartheta}\ln\left(\frac{2\vartheta(\mathrm{V}_G(f) + f^*)}{f^*}\right). \tag{84}$$

From (82), (83) and (84) we conclude that there exists

$$i^+ \leq O(1)\omega_*^{-1}\sqrt{\vartheta}\ln\left(\frac{2\vartheta[G\!:\!\widehat{x}](\mathrm{V}_G(f) + f^*)}{f^*}\right)$$

such that (t^{i^+-1}, x^{i^+-1}) is bad and

$$\Delta_1(t^{i^+-1}, x^{i^+-1}) \geq 5\vartheta_*\mathrm{V}_G(c), \quad \Delta_3(t^{i^+-1}, x^{i^+-1}) \geq 5\vartheta_*\mathrm{V}_G(d). \tag{85}$$

Now let $H = \nabla_x^2 F(t^{i^+-1}, x^{i^+-1})$. The Dikin ellipsoid $\{x \mid |x - x^{i^+-1}|_H \leq 1\}$ is contained in G, so that the variation $2\sqrt{e^T H^{-1} e}$ of a linear form $e^T x$ on the ellipsoid is $\leq \mathrm{V}_G(e)$. Thus, (85) implies (65) and (66); since (t^{i^+-1}, x^{i^+-1}) is bad, we see that the Infeasibility Test detects infeasibility at the iteration i^+, and (69) follows.
8^0. It remains to consider the case when (P') is feasible, as it is assumed from now on. We need the following observation:

Lemma 5.2 *Let i be such that $t_3^{i-1} \geq \max_{x\in G} d^T x$. If (t^{i-1}, x^{i-1}) is bad, then both t_1^{i-1} and t_1^i are $\leq c^*$.*

Proof. For the sake of brevity, let us write (t, x) instead of (t^{i-1}, x^{i-1}) and t^+ instead of t^i. Since $t_3 \geq \max_{x\in G} d^T x$, we have

$$t_2^*(\tau, \tau', t_3) = \phi(\tau) \equiv \min\{f^T x \mid x \in G, c^T x \leq \tau\}, \ \ (\tau, \tau', t_3) \in T.$$

Since G is compact convex set, $\phi(\tau)$ is continuous nonincreasing convex function on the ray $[\min_{x\in G} c^T x, \infty)$; this function is positive to the left of c^* and is zero to the right of c^* (recall that since (P') is feasible, we have $\min_{x\in G} f^T x = 0$ and $c^* = \min_{x\in G: f^T x\le 0} c^T x$).

By Corollary 4.1 and since (t,x) is bad, we have

$$16\Delta_2(t,x) < t_2 = t_2^*(t) + [t_2 - t_2^*(t)] \le t_2^*(t) + 8\Delta_2(t,x) = \phi(t_1) + 8\Delta_2(t,x),$$

whence

$$\phi(t_1) \ge \frac{1}{2} t_2 > 8\Delta_2(t,x) > 0. \tag{86}$$

Consequently, $t_1 < c^*$. Besides this,

$$c^T x = t_1 - \Delta_1(t,x), \quad f^T x = t_2 - \Delta_2(t,x),$$

whence

$$\phi(t_1 - \Delta_1(t,x)) \le t_2 - \Delta_2(t,x).$$

Since ϕ is convex, we have

$$\phi(t_1 + \Delta_1(t,x)) \ge 2\phi(t_1) - \phi(t_1 - \Delta_1(t,x)) \ge 2\phi(t_1) + \Delta_2(t,x) - t_2 \ge$$

[see (86)]

$$\ge \Delta_2(t,x) > 0,$$

whence $t_1 + \Delta_1(t,x) < c^*$. It remains to note that in view of (67), (71)

$$t_1^+ = t_1 + r_i\Delta_1(t,x) < t_1 + \Delta_1(t,x). \qquad \blacksquare$$

9^0. According to Lemma 5.2, the behaviour of the sequence $\{t_1^i\}$ starting with the moment i^* is as follows: if (t^i, x^i) is bad, then both t_1^i and t_1^{i+1} are less than c^*, and if (t^i, x^i) is good, then, by (67), $t_1^{i+1} < t_1^i$. Consequently, the sequence

$$\{\varepsilon_i = \max[t_1^i - c^*, 0]\}_{i\ge i^*}$$

is nonincreasing and

$$\phi_i \equiv t_1^i - \min_{x\in G} c^T x \le \phi^* \equiv \max[c^*, t_1^{i^*}] - \min_{x\in G} c^T x, \;\; i \ge i^*. \tag{87}$$

Since one clearly has $t_1^i - t_1^*(t^i) \le \phi_i$, we come to

$$t_1^i - t_1^*(t^i) \le \phi^*, \;\; i \ge i^*. \tag{88}$$

By definition of Ω_i, for $i \ge i^*$ we have

$$f^T x^i \le t_2^i = \left[\frac{\Omega_0}{\Omega_i} \frac{t_1^i - t_1^*(t^i)}{\widehat{t}_1 - t_1^*(\widehat{t})}\right]^{1/32} \widehat{t}_2 \le$$

[see (77), (57) and (88)]

$$\le (1 - O(1)\omega_*\vartheta^{-1/2})^i \left[\frac{\phi^*}{\widehat{t}_1 - t_1^*(\widehat{t})}\right]^{1/32} \max_{x\in G} f^T x.$$

Applying (76) to $e = c$ and taking into account that by definition of $\widehat{t}_1$ one has

$$\widehat{t}_1 - t_1^*(\widehat{t}) \ge \widehat{t}_1 - c^T\widehat{x} = c^T\widehat{x} - \min_{x\in W} c^T x,$$

we get

$$i \ge i^* \ \Rightarrow\ f^T x^i \le (1 - O(1)\omega_*\vartheta^{-1/2})^i (\phi^{**})^{1/32} \max_{x\in G} f^T x, \quad \phi^{**} = 14\vartheta[G\!:\!\widehat{x}]\frac{\phi^*}{\mathrm{V}_G(c)}. \tag{89}$$

10^0. Let us prove that

$$\phi^* \le 20\vartheta[G\!:\!\widehat{x}]\mathrm{V}_G(c). \tag{90}$$

Let $d_i = t_3^i - t_3^*(t^i)$. By (72) we have $d_i \ge (1 + r_i)d_{i-1}$, while (see (87), (67))

$$\frac{\phi_i}{\phi_{i-1}} \le 1 + r_i\frac{\Delta_1(t^{i-1}, x^{i-1})}{\phi_{i-1}} \le 1 + r_i$$

(we have taken into account that $\phi_{i-1} > \Delta_1(t^{i-1}, x^{i-1}) = t_1^{i-1} - c^T x^{i-1}$). Consequently,

$$\phi_i \le \phi_0\frac{d_i}{d_0} \le \mathrm{V}_G(c)\frac{d_i}{d_0}$$

(the second inequality is given by (57)). In the case of $i^* = 0$ the resulting relation clearly implies $\phi_{i^*} \le \mathrm{V}_G(c)$. If $i^* > 0$, then the above inequalities lead to

$$\phi_{i^*} \le (1 + r_{i^*})\phi_{i^*-1} \le (1 + r_{i^*})\mathrm{V}_G(c)\frac{d_{i^*-1}}{d_0} \le$$

[see (71) and note that $d_{i^*-1} \le \mathrm{V}_G(d)$ by definition of i^*, while $d_0 \ge (14\vartheta[G:\widehat{x}])^{-1}\mathrm{V}_G(d)$ by (74)]

$$\le 20\vartheta[G\!:\!\widehat{x}]\mathrm{V}_G(c).$$

To get (90), it remains to note that, by definition,

$$\phi^* = \max[c^*, t_1^{i^*}] - \min_{x\in G} c^T x = \max[c^* - \min_{x\in G} c^T x, \phi_{i^*}] \le \max[\mathrm{V}_G(c), \phi_{i^*}].$$

Combining (90) and (89). we come to

$$i \ge i^* \Rightarrow f^T x^i \le O(1)\vartheta[G\!:\!\widehat{x}](1 - O(1)\omega_*\vartheta^{-1/2})^i \max_{x\in G} f^T x. \tag{91}$$

11^0. Now let i^{**} be the first $i \ge i^*$ such that (t^i, x^i) is bad (if no such i exists, we set $i^{**} = +\infty$). As it was explained in the beginning of 9^0, we have

$$c^T x^i - c^* \le \varepsilon_i \equiv \begin{cases} 0, & i \ge i^{**} \\ \max\left[t_1^i - c^*, 0\right], & i^* \le i < i^{**} \end{cases} \tag{92}$$

According to Corollary 4.1,

$$t_1^{i-1} - t_1^*(t^{i-1}) \le 8\Delta_1(t^{i-1}, x^{i-1});$$

when $i > i^*$, we have

$$t_1^*(t^{i-1}) = \min\{c^T x \mid x \in G, f^T x \le t_2^{i-1}\} \le c^*,$$

so that for $i^* < i < i^{**}$

$$t_1^i - c^* \le t_1^{i-1} - r_i \Delta_1(t^{i-1}, x^{i-1}) - c^* \le t_1^{i-1} - c^* - \frac{r_i}{8}(t_1^{i-1} - t_1^*(t^{i-1})) \le \left(1 - \frac{r_i}{8}\right)(t_1^{i-1} - c^*),$$

whence, due to (71),

$$\varepsilon_i \le (1 - O(1)\omega_* \vartheta^{-1/2})\varepsilon_{i-1}, \ \ i > i^*.$$

Combining this observation with (92), we come to

$$c^T x^i - c^* \le (1 - O(1)\omega_* \vartheta^{-1/2})^{i-i^*} \varepsilon_{i^*}, \ i > i^*,$$

whence, in view of $\varepsilon_{i^*} \le \phi^*$ and (90)

$$c^T x^i - c^* \le O(1)\vartheta[G\!:\!\widehat{x}](1 - O(1)\omega_* \vartheta^{-1/2})^{i-i^*} \mathrm{V}_G(c), \ i \ge i^*. \tag{93}$$

Combining (93), (91) and taking into account (73), we come to (70). ■

6 Application examples

Our "long step" technique for tracing a surface of analytic centers heavily exploits assumption $\mathcal{Q}$ (Section 5.1) on structure of the s.-c.b. F for the domain G of problems (P), (P'); let us call barriers satisfying this assumption *good*. The goal of this section is to demonstrate that the s.-c.b.'s responsible for many important applications indeed are good.

6.1 Combination rules

Let us start from the following general remark. The desired structure is "stable with respect to intersections". Namely, assume that G is represented as an intersection $\cap_{i=1}^m G_i$ of closed convex domains; we shall say that G_i represents (or simply is) *i-th constraint of the problem*. The aforementioned stability means that if every G_i admits a good ϑ_i-s.-c.b. $F_i(x) = \Phi_i(\pi_i x + \epsilon_i)$ (so that Φ_i is a ϑ_i-s.-c.b. for certain closed convex domain $H_i \subset \mathbf{R}^{q_i}$ and we know the Legendre transformation $\Phi_{i,*}$ of Φ_i), then a good $(\sum_i \vartheta_i)$-s.c.b. for G is

$$F(x) = \Phi(\pi x + \epsilon),$$

where

$$\Phi(u_1, ..., u_m) = \sum_{i=1}^m \Phi_i(u_i) : \mathrm{int}\,(H_1 \times ... \times H_m) \to \mathbf{R}, \ \ \pi x + \epsilon = (\pi_1 x + \epsilon_1, ..., \pi_m x + \epsilon_m);$$

note that

$$\Phi_*(s_1, ..., s_m) = \sum_{i=1}^m \Phi_{i,*}(s_i)$$

(see Proposition 2.1).

Thus, our assumption is "separable with respect to the constraints" involved into the description of G.

The structure in question is also stable with respect to affine substitutions of argument: if G is the inverse image of certain closed convex domain G^+ under an affine mapping $\mathcal{A}$ (the image of the mapping intersects int G^+) and we know a good ϑ-s.-c.b. F^+ for G^+, then we can equip G with the ϑ-s.-c. barrier $F(x) = F^+(\mathcal{A}(x))$, and this barrier clearly is good.

6.2 "Building blocks"

The indicated combination rules can be applied to a number of "building blocks", i.e., good barriers for certain standard convex domains. These blocks are as follows:

1. *Nonnegative half-axis* $\mathbf{R}_+$: The standard 1-s.-c.b. $\Phi(x) = -\ln x$ for $\mathbf{R}_+$ is good: its Legendre transformation is $\Phi_*(s) = -\ln(-s) - 1$, $s < 0$; both Φ and Φ_* are 2-regular.

This elementary observation, in view of the combination rules, allows to handle arbitrary linear inequality constraints and, in particular, covers all needs of Linear Programming.

2. *Convex domain $G \subset \mathbf{R}^n$ which is a connectedness component of the Lebesgue set* $\mathrm{cl}\{x \mid f(x) < 0\}$ *of a quadratic function* f: Such a domain can be represented as the inverse image of the second-order cone

$$H = \{u \in \mathbf{R}^q \mid u_q \geq (\sum_{i=1}^{q-1} u_i^2)^{1/2}\}$$

under an easily computable affine mapping $x \mapsto u = \pi x + \epsilon$ with the image of the mapping intersecting the interior of H; setting

$$\Phi(u) = -\ln(u_q^2 - \sum_{i=1}^{q-1} u_i^2),$$

we obtain a 2-s.-c.b. for H (Nesterov and Nemirovski (1994), Chapter 5) with the explicit Legendre transformation

$$\Phi_*(s) = -\ln(s_q^2 - \sum_{i=1}^{q-1} s_i^2) - 2 + 2\ln 2,\ s \in -H,$$

and consequently can equip G with the good 2-s.-c.b. $F(x) = \Phi(\pi x + \epsilon)$; both Φ and Φ_* are 6-regular.

This observation covers convex quadratically constrained problems and even more general family of convex programs (note that f should not necessarily be convex, e.g., we may handle the hyperbolic domain of the type $\sum_{i=1}^{n-1} x_i^2 + 1 \leq x_n^2$, $x_n > 0$).

3. *Geometrical Programming in the exponential form*: Assume that among the constraints defining the feasible domain of a convex program there is a constraint of the form

$$\sum_{i=1}^{q} \exp\{a_i^T x + b_i\} \leq p^T x + r.$$

Adding q extra variables $y_1, ..., y_q$, one may pass from the initial problem to an equivalent one where the indicated constraint is represented by the system of convex inequalities

$$\exp\{a_i^T x\} \leq y_i,\ i = 1, ..., q;\ \sum_{i=1}^{q} \exp\{b_i\} y_i \leq p^T x + r.$$

We already know how to handle the concluding linear constraint, and all we need is to understand how to deal with the exponential inequality

$$\exp\{a_i^T x\} \leq y_i.$$

In order to penalize the latter constraint by a good barrier, it suffices to point out a good barrier for the epigraph

$$G = \{(t, x) \in \mathbf{R}^2 \mid t \geq \exp\{x\}\}$$

of the exponent and to use the combination rule related to affine substitutions of argument. A good 2-s.-c.b. for G can be written down explicitly (Nesterov and Nemirovski (1994), Chapter 5):

$$\Phi(t,x) = -\ln(\ln t - x) - \ln t,\ \ \Phi_*(\tau,\xi) = (\xi+1)\ln\left(\frac{\xi+1}{-\tau}\right) - \xi - \ln\xi - 2,$$

$$\mathrm{Dom}\,\Phi_* = \{(\tau,\xi)\in\mathbf{R}^2 \mid \tau<0, \xi>0\}.$$

Both Φ and Φ_* turn out to be 6-regular.

4. *Linear Matrix Inequality constraint:* A constraint of this type arises in numerous applications and defines the domain of the form

$$G = \{x \mid \mathcal{A}(x) \text{ is positive semidefinite}\},$$

where $\mathcal{A}(x)$ is an affine in x matrix-valued function taking values in the space of symmetric matrices of a given row size m. A good m-s.-c.b. for G is given by

$$F(x) = \Phi(\mathcal{A}(x)),\ \ \Phi(y) = -\ln\mathrm{Det}\ y;$$

Φ is the standard m-s.-c.b. for the cone $\mathbf{S}^m_+$ of symmetric $m\times m$ positive semidefinite matrices (see Nesterov and Nemirovski (1994), Chapter 5) with the Legendre transformation

$$\Phi_*(s) = -\ln\mathrm{Det}\,(-s) - m,\ \ \mathrm{Dom}\,\Phi_* = -\,\mathrm{int}\ \mathbf{S}^m_+;$$

both Φ and Φ_* are 2-regular. This example covers all needs of Semidefinite Programming.

We see that our assumption on the structure of F is compatible with a wide spectrum of important Convex Programming problems.

7 Appendix. Proof of Lemma 5.1

Proof of Lemma 5.1. The function $t^*_k(\cdot)$ clearly depends on the first $k-1$ components of argument only; let us denote the vector comprised of these first $k-1$ components by τ, so that $t^*_k = t^*(\tau)$. Since G is bounded, $t^*(\tau)$ is a convex continuous function on the closure T' of the projection of T on the plane of the first $k-1$ parameters; this projection is monotone ($\tau' \geq \tau \in T' \Rightarrow \tau' \in T'$), and $t^*(\tau)$ is monotonically nonincreasing on T'.

Let

$$\tau = (t_1, ..., t_{k-1}),\quad \delta\tau = ((\delta t_1)_-, (\delta t_2)_-, ..., (\delta t_{k-1})_-),$$

where $a_- = \min\{a,0\}$, and let $\Delta\tau = (\delta t_1 - \delta\tau_1, ..., \delta t_{k-1} - \delta\tau_{k-1})$; then $\Delta\tau \geq 0$. It is possible that $\delta\tau = 0$; then the statement in question is evident, since here

$$t^*_k(t + r\delta t) = t^*(\tau + r\Delta\tau) \leq t^*(\tau) = t^*_k(t)$$

(we have used the monotonicity of $t^*(\cdot)$), so that

$$t_k + r\delta t_k - t^*_k(t+r\delta t) \geq t_k + r\delta t_k - t^*_k(t) = (t_k - t^*_k(t))\left(1 + \frac{r\delta t_k}{t_k - t^*_k(t)}\right) \geq$$

$$\geq \left(1 + r\frac{\delta t_k}{2(k+1)\Delta_k(t,x)}\right)(t_k - t^*_k(t))$$

(the concluding inequality follows from (53)), and we come to (63).

Now consider the case when $\delta\tau \neq 0$. Let z be the largest real $p \geq 0$ such that

$$|p\delta\tau_i| \leq \Delta_i(t,x),\ i = 1, ..., k-1.$$

As we know from Corollary 4.1, one has $(\tau + z\delta\tau, t_k - \Delta_k(t,x)) \in \mathrm{cl}\, T$, and consequently

$$t^*(\tau + z\delta\tau) \leq t_k - \Delta_k(t,x). \tag{94}$$

Let, further,

$$\widehat{r} = \min\{r, z\},\ \theta = \widehat{r}/z;$$

since t^* is convex, we have

$$t^*(\tau + \widehat{r}\delta\tau) \leq t^*(\tau) + \theta(t^*(\tau + z\delta\tau) - t^*(\tau)) \leq t^*(\tau) + \theta[t_k - \Delta_k(t,x) - (t_k - 2(k+1)\Delta_k(t,x))]$$

(the latter inequality follows from (94) and (53)). Thus, we come to

$$t^*(\tau + \widehat{r}\delta\tau) \leq t^*(\tau) + (2k+1)\theta\Delta_k(t,x),$$

and since $t^*(\cdot)$ is monotonically nonincreasing, we have also

$$t^*_k(t + \widehat{r}\delta t) = t^*(\tau + \widehat{r}\delta\tau + \widehat{r}\Delta\tau) \leq t^*(\tau) + (2k+1)\theta\Delta_k(t,x),$$

or

$$t^*_k(t + \widehat{r}\delta t) \leq t^*_k(t) + (2k+1)\theta\Delta_k(t,x). \tag{95}$$

On the other hand, by definition of z one has $z|\delta t_i| = \Delta_i(t,x)$ for certain $i < k$ with $\delta t_i < 0$; since, by assumption, δt is k-safe,

$$\frac{\delta t_k}{\Delta_k(t,x)} \geq (4k+2)\frac{|\delta t_i|}{\Delta_i(t,x)},$$

and we come to

$$z\delta t_k \geq (4k+2)z|\delta t_i|\Delta_k(t,x)\Delta_i^{-1}(t,x) = (4k+2)\Delta_k(t,x), \tag{96}$$

whence

$$\Delta_k(t,x) \leq (4k+2)^{-1}z\delta t_k.$$

Combining this inequality with (95), we come to

$$(t_k + \widehat{r}\delta t_k) - t^*_k(t + \widehat{r}\delta t) \geq \widehat{r}\delta t_k + (t_k - t^*_k(t)) - \theta z(4k+2)^{-1}(2k+1)\delta t_k =$$

[since $\theta z = \widehat{r}$]

$$= \frac{\widehat{r}}{2}\delta t_k + (t_k - t^*_k(t)),$$

whence, again in view of $t_k - t^*_k(t) \leq 2(k+1)\Delta_k(t,x)$,

$$(t_k + \widehat{r}\delta t_k) - t^*_k(t + \widehat{r}\delta t) \geq (1 + \frac{\widehat{r}\delta t_k}{4(k+1)\Delta_k(t,x)})(t_k - t^*_k(t)). \tag{97}$$

It is possible that $\widehat{r} = r$; in this case (63) immediately follows from (97). It remains to consider the case when $\widehat{r} < r$. By definition of $\widehat{r}$, it means that $\widehat{r} = z < r$, and in view of (96) relation (97) implies that

$$\Delta \equiv (t_k + \widehat{r}\delta t_k) - t_k^*(t + \widehat{r}\delta t) \geq (1 + \frac{4k+2}{4k+4})(t_k - t_k^*(t)) = \frac{4k+3}{2k+2}(t_k - t_k^*(t)). \qquad (98)$$

Now consider the function

$$g(l) = (t_k + lr\delta t_k) - t_k^*(t + lr\delta t);$$

since r is a safe stepsize, it is a well-defined concave nonnegative function on the segment $[0, 1/q]$; the value of this function at the point $\widehat{l} = \widehat{r}/r \leq 1$ is, as we know from (98), $\geq (4k+3)(2k+2)^{-1}g(0)$, and from concavity it follows that

$$g(1) \geq \frac{q^{-1} - 1}{q^{-1} - \widehat{l}} g(\widehat{l}) \geq (1 - q) g(\widehat{l}),$$

whence

$$t_k + r\delta t_k - t_k^*(t + r\delta t) = g(1) \geq (1 - q)\frac{4k+3}{2k+2} g(0) = (1 - q)\frac{4k+3}{2k+2}(t_k - t_k^*(t)),$$

as required in (63). ∎

Acknowledgement. Research of the Research of the second author was partly supported by the Israeli-American Binational Science Foundation (BSF) and The Israeli Science Foundation grant # 306/94-3.

References

Gonzaga, C. (1989). An algorithm for solving Linear Programming problems in $O(n^3L)$ operations. N. Megiddo, Ed. *Progress in Mathematical Programming: Interior Point and related methods*, Springer-Verlag, New York, 1-29

Mizuno, S., M.J. Todd, Y. and Ye (1993). A surface of analytic centers and infeasible-interior-point algorithms in linear programming. *Math. Oper. Res.* **20** 135-162.

Nesterov, Yu.,A. Nemirovski (1994). *Interior point polynomial methods in Convex Programming: theory and applications*, *SIAM Series in Applied Mathematics*, SIAM, Philadelphia.

Nesterov, Yu. (1993). *Long-step strategies in interior point potential reduction methods*, Report, Departement d'Economie Commericale et Industrielle, Universite de Geneve, Geneve.

Nesterov, Yu., M. Todd (1994). *Self-scaled cones and interior-point methods in nonlinear programming*, CORE Discussion Paper # 9462, Louvain-la-Neuve, Belgium.

Nesterov, Yu., M. Todd (1995). *Primal-dual interior point methods for self-scaled cones*, Technical Report # 1125, School of OR and IE, Cornell University, Ithaca.

Renegar, J. (1988). A polynomial time algorithm, based on Newton's method, for linear programming. *Math. Progr.* **40** 55-93.

Yu. Nesterov: CORE, Voie de Roman Pays 34, 1348 Louvain-la-Neuve, Belgium; e-mail nesterov@core.ucl.ac.be

A. Nemirovski: Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel; e-mail nemirovs@ie.technion.ac.il