POLYNOMIAL TIME CUTTING PLANE ALGORITHMS ASSOCIATED WITH SYMMETRIC CONES

Research Thesis

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Operations Research and System Analysis

Dmitry Gabelev

SUBMITTED TO THE SENATE OF THE TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY

TAMUZ 5762 HAIFA JANUARY 2003

The research thesis was done under the supervision of Prof. A. S. Nemirovski in the department of Industrial Engineering.

I am deeply grateful to Prof. A. S. Nemirovski for his devoted guidance and encouragement, for teaching me his clear way of thinking, and for making of me a researcher.

Contents

List of Tables

List of Figures

Abstract

In this Thesis, we develop new polynomial-time methods for black-box-represented Convex Programming problems of the form $\min_{x \in X} f(x)$, where X is a solid (convex set with a nonempty interior) in $\mathbf{R}^n$ and f is a convex continuous function on X. A "black-box" representation of such a problem is given by a pair of "black box" subroutines ("oracles"); specifically, X is given by a *Separation oracle* which, given on input a point $x \in \mathbf{R}^n$, reports whether $x \in \operatorname{int} X$, and if its is not the case, returns a linear form which separates x from $\operatorname{int} X$, while f is represented by a *First Order oracle* which, given on input $x \in \operatorname{int} X$, returns $f(x)$ and $f'(x) \in \partial f(x)$. Besides, we assume that we are given in advance reals $R \geq r > 0$ such that X is contained in the centered at the origin Euclidean ball of radius R and contains a ball of radius r. A method $\mathcal{B}$ for solving problems of this type is called *polynomial*, if, for every $\epsilon \in (0,1)$ and every problem $\min_{x \in X} f(x)$ from the outlined family, $\mathcal{B}$ generates an ϵ-solution to the problem (i.e., a point $x_\epsilon \in X$ such that $f(x_\epsilon) - \min_X f \leq \epsilon(\max_X f - \min_X f)$) in a polynomial in n, $\ln(1 + R/r)$ and $\ln(1 + 1/\epsilon)$ number of calls to the Separation and the First Order oracles, with every call accompanied by a polynomial in the same parameters number of arithmetic operations.

The discovery of polynomial time methods for black-box-represented convex problems (1976) had fundamental theoretical consequences: these methods underly general results on polynomial time solvability of generic convex optimization programs, e.g., the famous result (Khachiyan, 1978) on polynomial time solvability of Linear Programming. Aside of their theoretical importance, methods of this group yield fast and reliable tools for solving general-type convex optimization programs of low dimension (up to 30-50 variables) and can be used to solve auxiliary problems when implementing numerous algorithms (e.g., restricted memory bundle methods) for large-scale convex optimization.

For the time being, just 3 polynomial time methods for black-box-represented Convex Programming problems are known, specifically, the *Ellipsoid method*

(Nemirovski & Yudin, 1976, Shor 1977), the *Circumscribed Simplex* method (Bulatov & Shepot'ko 1982, Yamnitski & Levin 1982), and *Inscribed Ellipsoid method* (Khachiyan et al., 1987). All these methods belong to the family of *cutting plane* methods (a natural multidimensional extension of the usual Bisection); however, the crucial property of polynomiality of a method relies upon "ad hoc" geometrical considerations, specific for every one of the three methods in question. The major goal of Thesis is to develop a new insight on the intrinsic nature of *stationary* polynomial time cutting plane methods. Our main result states, essentially, that *every self-scaled cone* (i.e., a closed pointed and convex cone in a Euclidean space which is self-dual and such that the group of linear authomorphisms of the cone acts transitively on its interior) *produces, via certain standard construction, a stationary polynomial time cutting plane method.* Our construction, as applied to a particular self-scaled cone – a *Lorentz cone* $\{x \in \mathbf{R}^{n+1} : x_{n+1} \geq \sqrt{x_1^2 + ... + x_n^2}\}$, yields the Ellipsoid method. Similarly, to get the Circumscribed Simplex method, one should apply the same construction to another self-scaled cone – *nonnegative orthant* $\mathbf{R}_+^{n+1} = \{x \in \mathbf{R}^{n+1} : x \geq 0\}$[1)]. As applied to other self-scaled cones (e.g., the cones of positive semidefinite real symmetric (or Hermitian, or Hermitian quaternion) matrices, or to arbitrary direct products of the above cones), our construction yields completely new polynomial time stationary cutting plane methods.

We analyze the theoretical complexity properties of the wide family of polynomial time methods we have discovered. It turns out that in the worst-case setting, no one of these methods can outperform the Ellipsoid algorithm. At the same time, it is known that the practical behaviour of the ellipsoid method is nearly the same as its theoretical worst-case behaviour. On the other hand, there are clear theoretical reasons to expect that the practical behaviour of other polynomial time cutting plane algorithms will be essentially better than their theoretical worst-case behaviour, so that in principle one could hope that the new algorithms are of certain practical interest. However, a numerical comparison of the existing and the new polynomial time algorithms is beyond the scope of this Thesis.

[1)] Our construction,however, cannot yield the Inscribed Ellipsoid method, since it is *not* stationary.

Chapter 1

Introduction

1.1 The goal: background, outline and motivation

In nonlinear optimization, research aims at the mathematical study of problems and solution algorithms, the implementation of the algorithms and experimentations on computers as well as practical applications. In the three directions, the development is different and is changing from time to time. Parallel with the widespread utilization of nonlinear optimization, the developments of algorithms and software engineering (e.g., expert systems, decision support systems, symbolic computation), and the rapid growth of computing experience, the need for exploring the mathematical nature of the arising problems is increasing, which means a serious help in modelling, designing and developing algorithms, i.e., the solid theoretical background makes the handling and solution of the problems substantially easier. In the theory of nonlinear optimization, remarkable results have been obtained till now.

One of the most important classes of nonlinear optimization is constituted by convex optimization problems. The importance of this problem class is due to the fact that convex programs, in a sense, form a "solvable case" in Continuous Optimization: under mild regularity assumptions, a generic convex optimization program admits an efficient – polynomial time – solution algorithm (for details, see [4], Chapter 5). *The goal of this Thesis is to develop new polynomial time solution algorithms for general-type Convex Programming problems.*

Polynomial Time algorithms in Convex Programming – preliminaries. The known polynomial time Convex Programming algorithms can be split into two groups:

A. *Universal "black-box-oriented methods"*. The characteristic features of these methods are as follows:

   – The methods, essentially, impose no restrictions on the problem, except for convexity;

   – The methods do not require complete a priori knowledge of the problem to be solved. The solution process is based solely upon the possibility to compute the values and subgradients of the objective function and the constraints at a point.

     The latter possibility is offered by a kind of black box routine – "oracle", and this is where the name "black-box-oriented methods" comes from.

B. *Methods for specific "well-structured" convex problems*, like Linear, Quadratic Linearly/Quadratically Constrained and Semidefinite Programming programs. These methods (as a matter of fact, all of them belong to the family of *Polynomial Time Interior Point methods*) can be characterized as follows:

   – The methods are aimed at solving programs of fixed analytical structure, so that the objective function and the constraints of a particular problem instance are fully specified by the values of the coefficients of known in advance analytical expressions;

   – The solution process utilizes the complete a priori knowledge of the problem under consideration.

The development of polynomial time algorithms for Convex Optimization was started in 1976 with the Ellipsoid method (Nemirovski and Yudin [10]; Shor [14]), followed by the Circumscribed Simplex method (Bulatov and Shepot'ko [5]; Yamnitski and Levin [16]) and the Inscribed Ellipsoid method (Khachiyan et al. [8]). All these methods belong to group A, and no other polynomial time methods of this group are known.

The first polynomial time Interior Point method was proposed by N. Karmarkar in 1984 [7] for Linear Programming. The discovery of Karmarkar initiated explosion of activity (sometimes called "Interior Point Revolution") in the area of IP methods first for Linear, and then - for well-structured nonlinear convex problems. As a result of 15-year effort of hundreds of researchers expressed in several thousands of theoretical papers and numerous software projects, the Interior Point revolution , extended dramatically the power and the scope of Convex Optimization techniques.

Black-box-oriented polynomial time methods vs. Interior Point ones. The strongest feature of black box oriented polynomial time methods is their *universality* – as it was already stated, the only "application restrictions" imposed by these methods are the convexity and the "efficient computability" of the objective and the constraints of the problem to be solved. This feature is crucial in theoretical results on *polynomial time solvability* of generic convex programs (see, e.g., [4], Chapter 5); the latter fact is one of the most fundamental discoveries in Convex Optimization and Computer Science. In contrast to this, Interior Point polynomial time methods are oriented at particular families of convex programs and thus are far from being universal. At the same time, universality, which makes black-box-oriented methods that attractive in theoretical studies, implies severe limitations on the computational power of these methods as actual computational tools. Indeed, it is known [11] that *for every* black-box-oriented method $\mathcal{B}$ as applied to programs

$$\min_x \{f(x) : x \in D_n = \{x \in \mathbf{R}^n,\ -1 \le x_i \le 1, 1 \le i \le n\}\} \qquad (B)$$

with convex continuous objective functions f normalized by the requirement $0 \le f(x) \le 1$ for $x \in D_n$ and every $\epsilon < 1/2$, there exists an instance of (B) where building ϵ-solution[1)] by $\mathcal{B}$ takes *at least* $0.8n \ln \frac{1}{\epsilon}$ calls to the oracle computing the values and the derivatives of f. Note that the outlined bound $O(n) \ln \frac{1}{\epsilon}$ on performance of black-box oriented methods is a theoretical *lower bound* expressed in terms of the number of oracle calls; the actual number of oracle calls $N(\epsilon)$ required to build ϵ-solution by a particular polynomial time black-box-oriented method can be significantly larger than this lower bound. Besides this, the answers of the oracle should be somehow processed by the method, which means additional computational effort. The actual complexity characteristics of known polynomial time black-box-oriented methods as applied to solving problems (B) within accuracy ϵ are given in Table 1.1. From these data is clear that

> *For black-box-oriented methods, the "computational cost per accuracy digit"* (the factor at $\ln \frac{1}{\epsilon}$ in the operation count) *blows up with the dimension n of the problem* <u>at least</u> *as* n^4.

It should be stressed that, as a matter of fact, the above bound reflects both the worst case and the *typical* behaviour of black-box-oriented methods. As a result,

---

[1)]i.e., a point $x_\epsilon \in D_n$ such that $f(x_\epsilon) - \min_{D_n} f \le \epsilon$

Method	# of oracle calls	Total # of arithmetic operations to process oracle's answers
Ellipsoid	$O(1)n^2 \ln\left(\frac{n}{\epsilon}\right)$	$O(1)n^4 \ln\left(\frac{n}{\epsilon}\right)$
Circumscribed Simplex	$O(1)n^3 \ln\left(\frac{n}{\epsilon}\right)$	$O(1)n^5 \ln\left(\frac{n}{\epsilon}\right)$
Inscribed Ellipsoid	$O(1)n \ln\left(\frac{n}{\epsilon}\right)$	$O(1)n^{4.5} \ln\left(\frac{n}{\epsilon}\right)$

Table 1.1: Performance characteristics of known polynomial time black-box-oriented methods when solving problems (B) within accuracy ϵ ($f : D_n \to [0, 1]$ is convex and continuous).

> *All known polynomial time black-box-oriented methods become impractical when solving problems with more than few tens of design variables.*

The polynomial time Interior Point methods are free from the outlined limitations; as a result, these methods allow to handle problems with hundreds and thousands, and in some cases (e.g., in Linear Programming with favourable sparsity pattern of the constraint matrix) – hundreds *of* thousands of design variables. This is perhaps the major reason of huge interest in Interior Point methods during last two decades.

The goal of the Thesis is to develop essentially new polynomial time black-box-oriented methods. The motivation behind this goal is as follows.

1. *As far as theoretical aspects are concerned*, the three existing universal polynomial time methods look a kind of "ad hoc" inventions. We are about to demonstrate that *both the Ellipsoid and the Circumscribed Simplex methods are particular cases of a general construction with allows to associate such a method with every symmetric self-scaled cone*. The Ellipsoid and the circumscribed Simplex algorithms are given by this construction as applied to a Lorentz cone and a nonnegative orthant, respectively, and these are just two very particular cases of self-scaled cones; the general case includes the cones of real symmetric, Hermitian and Quaternion positive semidefinite matrices, as well as finite direct products of cones from the just outlined series. We believe that a general theory of *stationary polynomial time black-box-oriented methods* we intend to develop yields a new insight and is of definite theoretical interest.

2. *As far as computational power is concerned*, the polynomial time methods we intend to develop are subject to the same severe limitations, intrinsically related to universality, as the existing black-box-oriented methods. These limitations, however, still leave enough room for the methods in question to be of practical interest. Indeed, when solving convex optimization problems of "small design dimension" n, like 10 – 40, the bounds in Table 1.1 are not prohibitively large, especially when one takes into account several attractive computational features of black-box-oriented polynomial time methods, specifically

   (a) ability to handle in polynomial time (and thus – to very high accuracy) very complicated objective and constraints with no transparent analytical structure;

   (b) perfect numerical stability, much better than the one of Interior Point methods, and thus – high reliability;

   (c) algorithmic and implementation simplicity.

   These attractive properties, known to be possessed by the Ellipsoid and the Circumscribed Simplex methods, in fact come from the specific structure of the algorithms. Since this structure, as we shall see, is shared by the novel methods we are about to develop, we have all reasons to believe that the outlined advantages will be possessed by the new algorithms as well.

   Of course, one could argue that a convex optimization method capable to solve problems with no more than 10 – 40 design variables is of no practical interest whatsoever, independently of how good is it in its "application domain". We strongly believe that this is not the case. Moreover, surprisingly enough, these are exactly extremely large-scale convex programs which give rise to interest in rapid and reliable low-dimensional optimization. Just two examples:

   (a) *Decomposition.* Consider a large-scale convex optimization program of the form

$$\min_{x=(x[1],\ldots,x[N])} \left\{ \sum_{i=1}^{N} f_i(x[i]) : F(x) = \begin{bmatrix} F_0(x) = \sum_{i=1}^{N} F_{0i}(x[i]) \\ F_1(x[1]) \\ \vdots \\ F_N(x[N]) \end{bmatrix} \leq 0 \right\} \tag{D}$$

where $x[i]$ are certain blocks of design variables and $F_0(x)$, $F_i(x[i])$, $i \geq 1$, are vector functions. If there were no "linking constraints" $F_0(x) \leq 0$, (D) would be merely a collection of N independent convex optimization problems; when N is large and $\dim x[i] << \dim x$, $\dim F_i << \dim F$, it is mush easier to solve the N problems from the collection one by one than to solve a single "large" problem with $\dim x$ variables and $\dim F$ constrains. Now consider the case when the linking constrains are present. In this case, by Lagrange Duality, problem (D), under mild regularity assumptions, is equivalent to finding saddle point (max in λ, min in x) of the function

$$L(x,\lambda) = \sum_{i=1}^{N} f_i(x[i]) + \sum_i \lambda^T F_{0i}(x[i])$$

on the set

$$\underbrace{\{\lambda \geq 0\}}_{\Lambda} \times \underbrace{\{x : F_i(x[i]) \leq 0,\, i = 1, ..., N\}}_{X}.$$

The saddle point problem, in turn, can be reduced to the problem of minimizing in $\lambda \in \Lambda$ the function

$$G(\lambda) = -\inf_{x \in X} L(x,\lambda).$$

This function is "easily computable"; indeed, to compute the value and a subgradient of G at a given point is, essentially, the same as to solve one by one N "small" problems

$$\min_{x[i]} \left\{ f(x[i]) + \lambda^T F_{0i}(x[i]) : F_i(x[i]) \leq 0 \right\},\ i = 1, ..., N.$$

Note that the problem of minimizing G over Λ is a typical black-box-represented convex program: even in the case when (D) is perfectly structured (say, is a Linear Programming program), the function $G(\lambda)$, although convex, has no "usable" structure. It follows that *the only way to implement the outlined Decomposition scheme is to solve* (G) *by a black-box-oriented method.* When the number $\dim F_0$ of "linking constraints" is small, say, ≤ 40 (which well may happen in the case when (D) by itself is extremely large-scale), then a natural candidate to the role of the latter method might be a polynomial time black-box-oriented algorithm.

(b) *Extremely large-scale optimization via restricted memory bundle methods.* When solving extremely large-scale convex programs (tens and hundreds of thousands of design variables; problems of this type do arise in many applications, e.g., in 3D Medical Imaging or Structural Design), one usually cannot use Interior Point methods even in the case when the program is perfectly structured. This impossibility comes from the fact that IP methods require solving Newton-type systems of the size equal to the design dimension of the problem, and solving a linear system with, say, 10,000 (not speaking of 100,000) variables is an absolutely impossible task[2)]. The problems in question typically are constrained problems of the form

$$\min_{x \in X} f(x), \qquad (*)$$

where X is a "simple" set (like Euclidean ball, or box, or simplex), and f is a convex objective defined on X. One of the most promising approaches to handling extremely large-scale problems of the form $(*)$ is offered by *restricted memory bundle methods* (see, e.g., [9, 17] and references therein). At a step of such a method, one needs to solve an auxiliary problem of the form

$$\min_{x \in X} \{\omega(x) : Ax \leq b\},$$

where

- $\omega(x)$ is a "simple" function (say, linear, or $\frac{1}{2}x^T x$),
- the column size of the *bundle* $[A, b]$ is under our full control and can be (and is) kept small, like $1 - 30$.

When the dimension of X is really large, the natural way to solve the auxiliary problem is by passing to its Lagrange dual problem

$$\min_{\lambda \geq 0} G(\lambda), \quad G(\lambda) = -\min_X [\omega(x) + \lambda^T [Ax - b]].$$

Here, same as in the case of decomposition, the dual objective $G(\cdot)$ is convex and "easily computable"[3)], although has no "usable structure". When minimizing G over $\lambda \geq 0$, we again are solving a low-dimensional convex program with black-box-represented objective, and again can use a polynomial time black-box-oriented algorithm.

[2)]unless the system possess an appropriate sparsity structure, which is *not* the case in many applications, e.g. those we have mentioned.

[3)]Indeed, to compute the value and a subgradient of G at a given point, one should minimize the *simple* function $\omega(x) + \lambda^T [Ax - b]$ over the *simple* set X; say, when X is Euclidean ball (or the standard simplex, or a box) in $\mathbf{R}^n$ and $\omega(x)$ is either linear, or $\frac{1}{2}x^T x$, it takes just $O(n \ln n)$ arithmetic operations to compute G and G'.

The bottom line is: *Although the main emphasis in this Thesis is on purely theoretical development of new polynomial time black-box-oriented optimization algorithms, associated with self-scaled cones, we do believe that our research may be of practical potential as well.*

1.2 Overview of contents

Our presentation is as follows. In the remaining sections of Introduction, we describe in details the *black-box-based* setting of a convex optimization program along with a general *Cutting Plane Scheme* for solving problems in this setting, thus specifying the framework for our further developments and allowing for detailed formulation of the goals of the research (Section 1.3). In Section 1.4, we outline the results of our research. These results are obtained and discussed in Chapter 2.

1.3 Black-box-represented convex programs and the Cutting Plane scheme

In what follows we focus on convex optimization programs in the form

$$\min_{x \in X} f(x), \tag{P}$$

where

- the *domain* X of the problem is a solid (convex compact set with a nonempty interior) in an n-dimensional Euclidean space $\mathbf{E}$ with inner product $\langle \cdot, \cdot \rangle_{\mathbf{E}}$;
- the *objective* $f : X \mapsto \mathbf{R}$ is a convex continuous function on X.

1.3.1 Black-box-represented convex programs

When speaking about solution methods for (P), we always assume that the problem is *black box represented*, meaning that

- X is represented by a *Separation oracle* – a routine which, given on input a point $x \in \mathbf{R}^n$, reports whether $x \in \operatorname{int} X$, and if it is not the case, reports a *separator* of x and X – a nonzero vector e and a real $a \geq 0$ such that
$$\langle e, x \rangle_{\mathbf{E}} \geq a + \max_{y \in X} \langle e, y \rangle_{\mathbf{E}}$$

(the existence of such a separator is given by the Separation Theorem for convex sets, see, e.g., [2], Theorem ???).

- f is represented by a *First Order oracle* which, given on input a point $x \in \operatorname{int} X$, returns on output the value $f(x)$ and a subgradient $f'(x)$ of f at x.

Note that outlined "working environment" covers many other traditional Convex Optimization settings, for example, problems given in the Nonlinear Programming form

$$\min_x \{f(x) : g_i(x) \leq 0,\, i = 1, ..., m\}. \tag{1.1}$$

Assuming that the problem (1.1) is convex (i.e., $f(x)$ and $g_i(x)$, $i = 1, ..., m$, are finite convex functions on $\mathbf{E}$) with a bounded feasible set $X = \{x : g_i(x) \leq 0,\, i = 1, ..., m\}$ and assuming the Slater condition

$$\exists \bar{x} : \quad g_i(\bar{x}) < 0,\, i = 1, ..., m,$$

we can rewrite (1.1) in the form of (P); from the assumptions we have just made it follows that X indeed is a solid. To equip the resulting problem (P) with Separation and First Order oracles, it suffices to be able to compute the values and subgradients of f and $g_1, ..., g_m$ at any given point. Indeed, in this case we can mimic the required oracles as follows.

• In order to mimic the Separation oracle, we, given $x \in \mathbf{E}$, compute the quantities $g_i(x)$, $1 \leq i \leq m$. If all these quantities are negative, we report that $x \in \operatorname{int} X$, otherwise we report that $x \notin \operatorname{int} X$, identify an index i^* such that $g_{i^*}(x) \geq 0$ and set

$$e = g'_{i^*}(x).$$

From the definition of the subgradient it follows that

$$\forall (y \in X) : 0 \leq g_{i^*}(x) - g_{i^*}(y) \leq \langle g'_{i^*}(x), x - y \rangle_{\mathbf{E}},$$

so that $g'_{i^*}(x)$ is the required separator.

• The First Order oracle for f is readily given by the postulated possibility to compute values and subgradients of f at every point.

1.3.2 The Cutting Plane scheme

The Cutting Plane scheme is one of the standard schemes for solving black-box-represented convex optimization problems in the form of (P); as a matter of fact, both known universal polynomial time optimization methods and the methods we intend to develop are particular implementations of this general scheme. The scheme is a straightforward multidimensional analogy of Bisection. As applied to (P), a generic cutting plane algorithm works as follows:

1. Initialization. Choose an initial localizer G_0 – a subset of $\mathbf{E}$ which contains X. Set $f^0 = +\infty$ (f^t is the best (the smallest) value of the objective along the feasible solutions found at the first t steps).

2. Step t, $t \geq 1$. Given previous localizer – a set $G_{t-1} \subset \mathbf{E}$ – we act as follows:

   (a) We choose somehow t-th search point $x_t \in \mathbf{E}$ and call the separation oracle to check whether $x_t \in \operatorname{int} X$.

      i. If the separation oracle says that $x_t \notin \operatorname{int} X$ (a *non-productive* step), it returns a separator $(e_t \neq 0, \alpha_t \geq 0)$:

      $$\langle e_t, x_t \rangle_{\mathbf{E}} \geq \alpha_t + \max_{y \in X} \langle e_t, y \rangle_{\mathbf{E}}, \qquad e_t \neq 0.$$

      ii. If the separation oracle says that $x_t \in \operatorname{int} X$ (a *productive* step), we call the first order oracle, x_t being the input, to get $f(x_t)$, $f'(x_t)$, and set

      $$\begin{array}{rcl} f^t & = & \min[f^{t-1}, f(x_t)] \\ e_t & = & f'(x_t) \\ \alpha_t & = & f(x_t) - f^t \quad [\geq 0]. \end{array}$$

      If $e_t = 0$, x_t is the exact solution of (P), and we terminate.

   (b) We set

   $$\widehat{G}_t = \{x \in G_{t-1} \mid \langle e_t, x - x_t \rangle_{\mathbf{E}} + \alpha_t \leq 0\},$$

   we choose somehow a new localizer G_t such that

   $$\widehat{G}_t \subset G_t.$$

   Loop to step $t + 1$.

3. Approximate solution x^t generated in course of the first t steps of the method is well-defined only if among these steps there were productive ones. In the latter case, x^t is the best – with the smallest value of f – of the search point x_τ associated with productive steps $\tau \leq t$. Note that x^t is well-defined if and only if $f^t < \infty$, and in this case $f(x^t) = f^t$.

1.3.3 Rate of convergence of a Cutting Plane algorithm

We have assumed once for ever that the objective f in (P) is convex and continuous on X. For the time being (and *only* for the time being) it makes sense to relax slightly this assumption and to assume that f is convex with $\mathrm{int}\, X \subset \mathrm{Dom} f$ (so that f can be $+\infty$ at certain boundary point of X) and that f is *semi-bounded*:

$$V(f) = \sup_{x \in \mathrm{int}\, X, y \in X} \langle y - x, f'(x) \rangle_{calE} < \infty. \tag{1.2}$$

Here, as always, $f'(x)$ is a subgradient of f at x (it is easily seen that the value of the right hand side in (1.2) is independent of how we choose this subgradient at every point). Note that if f is bounded on X, then f is semi-bounded, and

$$V(f) \leq \sup_{x \in \mathrm{int}\, X} f(x) - \inf_{x \in \mathrm{int}\, X} f(x).$$

Indeed, the right hand side in (1.2) remains unchanged when we restrict y to vary in $\mathrm{int}\, X$, and for $x, y \in \mathrm{int}\, X$ we have $(y - x)^T f'(x) \leq f(y) - f(x)$ (since f is convex and $\mathrm{Dom} f \supset \mathrm{int}\, X$).

The standard analysis of Cutting Plane methods is based on the following simple

Proposition 1.1 [12] *Let all localizers G_τ generated by a cutting plane method as applied to* (P) *be measurable. Assume that t is such that*

(a) in course of the first t steps the method did not terminate with an optimal solution to the problem, and

(b) one has

$$\epsilon(t) \equiv \left(\frac{\mathrm{Vol}(G_t)}{\mathrm{Vol}(X)} \right)^{1/n} < 1$$

(from now on, $\mathrm{Vol}(A)$ *is the* $\dim \mathbf{E}$*-dimensional Lebesque measure of a measurable set* $A \subset \mathbf{E}$*).*

Then the result x^t of the first t steps is well-defined, belongs to $\mathrm{int}\, X$ *and is an $\epsilon(t)$-solution to (P), i.e.,*

$$f(x^t) - \inf_X f \leq \frac{\epsilon(t)}{1 - \epsilon(t)} V(f). \tag{1.3}$$

Proof (see [12]). Let us choose $\epsilon' \in (\epsilon(t), 1)$ and $\delta > 0$, and let us set

$$X_\delta = \left\{x \in \mathrm{int}X \mid f(x) \le \inf_X f + \delta\right\}, \qquad X_\delta^{\epsilon'} = (1-\epsilon')X_\delta + \epsilon' X.$$

By evident reasons, $\mathrm{Vol}(X_\delta^{\epsilon'}) \ge (\epsilon')^n \mathrm{Vol}(X) > \epsilon^n(t)\mathrm{Vol}(X) = \mathrm{Vol}(G_t)$, so that the set $X_\delta^{\epsilon'} \backslash G_t$ is nonempty. Let y be a point of the latter set. Since $X_\delta^{\epsilon'} \subset \mathrm{int}X$, we have $y \in \mathrm{int}X$, whence also $y \in G_0 \supset X$. Thus, $y \in G_0$ and $y \notin G_t$; it follows that there exists $\tau \le t$ such that $y \in G_{\tau-1}$ and $y \notin G_\tau$. Looking at how G_t and $G_{\tau-1}$ are linked, we conclude that $y \in G_{\tau-1} \backslash \widehat{G}_\tau$, i.e., that

$$\langle e_\tau, y - x_\tau \rangle_{\mathbf{E}} + \alpha_\tau > 0. \tag{1.4}$$

We claim that the step τ is productive ($x_\tau \in \mathrm{int}\, X$). Indeed, otherwise (e_τ, α_τ) would separate x_τ and X, i.e., it would hold $\langle e_\tau, x_\tau \rangle_{\mathbf{E}} \ge \alpha_\tau + \langle e_\tau, z \rangle_{\mathbf{E}}$ for all $z \in X$, in particular, for $z = y$ (as we have seen, $y \in X$); but the relation $\langle e_\tau, x_\tau \rangle_{\mathbf{E}} \ge \alpha_\tau + \langle e_\tau, y \rangle_{\mathbf{E}}$ contradicts (1.4).
Thus, the step τ is productive ($x_\tau \in \mathrm{int}\, X$). Since $\tau \le t$, it follows that x^t is well-defined; by construction of x^t, this point, being well-defined, belongs to $\mathrm{int}\, X$ and satisfies the relation

$$f(x^t) = f^t \le f^\tau. \tag{1.5}$$

Furthermore, since τ is a productive step, (1.4) implies that $\langle f'(x_\tau), y - x_\tau \rangle_{\mathbf{E}} + \alpha_\tau > 0$, whence by convexity of f

$$f(y) \ge f(x_\tau) - \alpha_\tau = f^\tau \tag{1.6}$$

(the concluding equality is given by the definition of α_τ for a productive step τ). Now recall that $y \in X_\delta^{\epsilon'}$; by construction of the latter set, it means that there exist $u \in \mathrm{int}\, X$, $v \in X$, with $f(u) \le \inf_X f + \delta$, such that $y = (1-\epsilon')u + \epsilon' v$. We now have

$$\begin{aligned} \inf_X f + \delta \ge f(u) \ge f(y) + \langle f'(y), u - y \rangle_{\mathbf{E}} &= f(y) - \epsilon' \langle f'(y), v - u \rangle_{\mathbf{E}} \\ &= f(y) - \tfrac{\epsilon'}{1-\epsilon'} \langle f'(y), v - y \rangle_{\mathbf{E}}, \end{aligned}$$

whence

$$f(y) \le \inf_X f + \delta + \frac{\epsilon'}{1-\epsilon'} \langle f'(y), v - y \rangle_{\mathbf{E}} \le \inf_X f + \delta + \frac{\epsilon'}{1-\epsilon'} V(f).$$

Combining this relation with (1.6) and (1.5), we come to

$$f(x^t) \le \inf_X f + \delta + \frac{\epsilon'}{1-\epsilon'} V(f).$$

The resulting inequality if valid for all $\delta > 0$ and all $\epsilon' \in (\epsilon(t), 1)$, and (1.3) follows. ∎

Corollary 1.1 *Assume that in a cutting plane method the policies of choosing*

(a) the initial localizer G_0,

(b) the current search point x_t, given G_{t-1}, and

(c) the new localizer G_t, given $\widehat{G}_t$, are such that whatever are nonzero vectors e_t, it is guaranteed that

$$\left(\frac{\mathrm{Vol}(G_t)}{\mathrm{Vol}(G_{t-1})}\right)^{1/n} \leq \omega < 1. \tag{1.7}$$

(recall that $n = \dim \mathbf{E}$).

Then, for every solid X and every semi-bounded on X convex objective f, one has for all $t \geq 1$:

$$\delta(t) \equiv \omega^t \left(\frac{\mathrm{Vol}(G_0)}{\mathrm{Vol}(X)}\right)^{1/n} < 1 \Longrightarrow f(x^t) - \inf_X f \leq \frac{\delta(t)}{1-\delta(t)} V(f).$$

Moreover, when f is bounded, if $\omega = \omega(n)$ approaches 1 as $n \to \infty$ in a polynomial fashion:

$$\omega(n) = 1 - \frac{O(1)}{n^\beta} \qquad [\beta > 0] \tag{1.8}$$

then (1.3) *implies a polynomial-time, in terms of the iteration count, efficiency estimate, specifically, the estimate*

$$\begin{array}{l} \forall(\epsilon \in [0,1)) : \\ t \geq O(1) n^\beta \ln\left(\frac{D_0}{\epsilon}\right) \implies f(x^t) - \min\limits_X f \leq \epsilon \left[\max\limits_X f - \min\limits_X f\right], \\ D_0 = \left(\frac{\mathrm{Vol}(G_0)}{\mathrm{Vol}(X)}\right)^{1/n}. \end{array} \tag{1.9}$$

Due to the latter fact, a Cutting Plane algorithm satisfying (1.7)–(1.8) implies an efficient algorithm for non-smooth convex optimization, provided that one can efficiently implement the Cutting Plane step (i.e., compute x_t, given G_{t-1}, and compute G_t, given G_{t-1}, x_t, e_t).

1.3.4 Known polynomial time implementation of the Cutting Plane scheme

For the time being, just three polynomial time implementations of the Cutting Plane scheme satisfying are known, namely

- The Ellipsoid method (Nemirovski & Yudin [10]; Shor [14]). Here all localizers G_t are ellipsoids, x_t is the center of G_{t-1}, and G_t is the minimum volume ellipsoid containing $\widehat{G}_t$. For this algorithm, $\beta = 2$, and the arithmetic cost of the step is $O(n^2)$.

- The Circumscribed Simplex method (Bulatov & Shepot'ko [5]; Yamnitski & Levin [16]). Here all localizers G_t are simplexes, x_t is the barycenter of G_{t-1}, and G_t is certain approximation to the smallest volume simplex containing $\widehat{G}_t$. For this method, $\beta = 3$ (which is worse than for the Ellipsoid method) and the arithmetic cost of a step is $O(n^2)$ (same as in the Ellipsoid method).

- The Inscribed Ellipsoid method (Khachiyan et al [8]). Here X is assumed to be a polytope, all localizers G_t also are polytopes, x_t is the center of the maximal in volume ellipsoid inscribed into the previous localizer G_{t-1}, and $G_t = \widehat{G}_t$. For this method, $\beta = 1$ (which is better than for the Ellipsoid method) and the arithmetic cost of a step is $O(n^{3.5})$ (which is worse than for the Ellipsoid method).

It should be stressed that the existence of "polynomial time Cutting Plane algorithms" has extremely important theoretical and rather important practical consequences: this fact underlies the fundamental (and a fairly general) theorem on polynomial-time solvability of generic Convex Optimization programs (see [4], Chapter 5); in particular, the Ellipsoid method is the major "working horse" in the famous proof of polynomial time solvability of Linear Programming with rational data (Khachiyan, 1978).

1.3.5 Stationary polynomial time Cutting Plane algorithms

Both the Ellipsoid and the Simplex implementations of the Cutting Plane scheme are "stationary" in the sense that here all the localizers are affine images of a once for ever fixed n-dimensional solid – the n-dimensional Euclidean ball for the first method and the standard n-dimensional simplex for the second method. More generally, a stationary cutting plane algorithm for solving n-dimensional convex programs is a Cutting Plane algorithm where

- all the localizers are "geometrically the same" – all of them are the images of certain "perfect" solid $\mathbf{B} \subset \mathbf{E}$, $0 \in \operatorname{int}\mathbf{B}$, under invertible affine mappings

$$x \mapsto C(x) = Cx + c;$$

- if the current localizer is

$$G_t = C_t\mathbf{B} + c_t,$$

  then the corresponding search point is

$$x_{t+1} = c_t.$$

In fact, all we need in order to build a stationary, in the aforementioned sense, Cutting Plane algorithm satisfying (1.7) is a *perfect* solid $\mathbf{B} \subset \mathbf{E}$, specifically, a solid with the following properties:

P.1. $\mathbf{B}$ contains the centered at the origin unit Euclidean ball and is contained in the concentric ball of radius $\gamma[\mathbf{B}]$;

P.2. Given a nonzero vector $p \in \mathbf{E}$, we can efficiently point out a one-to-one affine transformation

$$x \mapsto A^p(x) = A_p x + a_p$$

such that

i. $\mathbf{B}^+ \equiv A^p(\mathbf{B}) \supset \{x \in \mathbf{B} \mid \langle p, x\rangle_{\mathbf{E}} \leq 0\}$;

ii. $|\mathrm{Det}(A_e)|^{1/n} \leq \omega < 1$ ($n = \dim \mathcal{E}$). Geometrically: a part of $\mathbf{B}$ cut off $\mathbf{B}$ by a hyperplane passing through the origin can be covered by an affine image $\mathbf{B}^+$ of $\mathbf{B}$ under an affine mapping which reduces volumes by factor at least ω^n. The quantity ω will be called the *index* of $\mathbf{B}$.

Assume that when solving (P) we know in advance two reals R, r, $R \geq r > 0$, such that X is contained in the centered at the origin Euclidean ball of radius R and contains (unknown) ball of radius r. Then, given a perfect solid $\mathbf{B}$, we can associate with it a stationary cutting plane method as follows.

SCP Algorithm associated with a perfect solid B:

1. All localizers G_t generated by the method are affine images of $\mathbf{B}$:

$$G_t = C^t(\mathbf{B}),$$

   where $C^t(x) = C_t x + c_t$ and C_t are nonsingular square matrices. Analytically (and algorithmically), the method operates on the data (C_t, c_t) of these affine mappings.

2. The initialization policy (a) is just to set

$$C_0 = RI, \qquad c_0 = 0.$$

With this policy, $G_0 = R\mathbf{B}$ contains the centered at the origin Euclidean ball of radius R (since $\mathbf{B}$ contains the centered at the origin unit ball), and this ball contains X, as required for the initial localizer.

3. The search policy (b) is to set

$$x_t = C^{t-1}(0) = c_{t-1}.$$

4. The embedding policy (c) is as follows:

   i. Given $e_t \neq 0$, we set
   $$p_t = C_{t-1}^* e_t$$
   ($x \mapsto C_t^* x$ is the mapping conjugate to the mapping $x \mapsto C_t x$). Note that since $G_{t-1} = C^{t-1}(\mathbf{B})$ and $x_t = C^{t-1}(0)$, we have
   $$\widehat{G}_t \equiv \{x \in G_{t-1} \mid \langle e_t, x - x_t\rangle_{\mathcal{E}} + \alpha_t \leq 0\} = C^{t-1}(\widehat{B}^t),$$
   $$\widehat{B}^t = \{y \in \mathbf{B} \mid \langle p_t, y\rangle_{\mathbf{E}} + \alpha_t \leq 0\}$$
   and that $p_t \neq 0$ (since $e_t \neq 0$ and C_{t-1} is nonsingular).

   ii. By assumption and since $\alpha_t \geq 0$, the set $\widehat{B}^t \subset \{x \in \mathbf{B} : \langle p_t, x\rangle_{\mathbf{E}} \leq 0\}$ is contained in the image of $\mathbf{B}$ under the one-to-one affine mapping $y \mapsto A^{p_t}(y)$, whence the set $C^{t-1}(\widehat{B}^t)$ (which contains the set $\widehat{G}_t$) is contained in the image of $\mathbf{B}$ under the one-to-one affine mapping $y \mapsto C^{t-1}(A^{p_t}(y))$. Thus, setting
   $$C_t = C_{t-1}A_{p_t}, \qquad c_t = C_{t-1}a_{p_t} + c_{t-1}$$
   (so that $C^t(y) = C^{t-1}(A^{p_t}(y))$ for all y), we get
   $$G_t \equiv C^t(\mathbf{B}) = C^{t-1}(A^{p_t}(\mathbf{B})) \supset C^{t-1}(\widehat{B}_n^t) = \widehat{G}_t,$$
   i.e., $G^t \supset \widehat{G}_t$, as required in the Cutting Plane scheme.

Observe that the resulting policies (a), (b), (c) ensure that

$$\begin{aligned}\left(\frac{\mathrm{Vol}(G_t)}{\mathrm{Vol}(G_{t-1})}\right)^{1/n} &= \left(\frac{\mathrm{Vol}(C^t(\mathbf{B}))}{\mathrm{Vol}(C^{t-1}(\mathbf{B}))}\right)^{1/n} = \left(\frac{|\,\mathrm{Det}\ C_t\,|}{|\,\mathrm{Det}\ C_{t-1}\,|}\right)^{1/n} \\ &= \left(\frac{|\,\mathrm{Det}\ (C_{t-1}A_{p_t})\,|}{|\,\mathrm{Det}\ C_{t-1}\,|}\right)^{1/n} = |\,\mathrm{Det}\ A_{p_t}\,|^{1/n} \\ &\leq \omega,\end{aligned}$$

so that the resulting cutting plane method converges linearly with the rate ω. In particular, if

$$\omega \leq 1 - O(1)n^{-\beta}, \tag{1.10}$$

then the outlined stationary Cutting Plane algorithm is a polynomial time one. Specifically, the *iteration complexity of finding ϵ-solution*, i.e., the number of steps sufficient to build a point $x_\epsilon \in X$ such that

$$f(x_\epsilon) \leq \min_X f + \epsilon[\max_X f - \min_X f],$$

for every $\epsilon \in (0,1)$ does not exceed

$$N(\epsilon) = O(1)n^\beta \ln\left(\frac{\gamma[\mathbf{B}]R}{\epsilon r}\right). \tag{1.11}$$

This result is readily given by (1.9) combined with the fact that in our case

$$D_0 = \left(\frac{\mathrm{Vol}(G_0)}{\mathrm{Vol}(X)}\right)^{1/n} = R\left(\frac{\mathrm{Vol}(\mathbf{B})}{\mathrm{Vol}(\{x : \|x\|_2 \leq r\})}\right)^{1/n} \leq \frac{R\gamma[\mathbf{B}]}{r}$$

(recall that X contains Euclidean ball of radius r, and $\mathbf{B}$ is contained in a Euclidean ball of radius $\gamma[\mathbf{B}]$).

To better visualize aforementioned algorithm, see Fig. 1.1.

1.4 Main Result

For the time being, exactly two "generic solids" were known to possess properties P.1 – P.2, specifically, n-dimensional Euclidean ball and n-dimensional simplex; the associates stationary Cutting Plane algorithms were exactly the Ellipsoid and the Circumscribed Simplex polynomial time methods (note that the third of the existing polynomial time Cutting Plane algorithms – the Inscribed Ellipsoid one – is *not* stationary). The main theoretical result of this Thesis is that

> *Every self-scaled cone gives rise to a perfect solid satisfying* (1.10) *and thus gives rise to a stationary polynomial time Cutting Plane algorithm.*

Here a *self-scaled cone* is defined as a closed pointed cone $\mathbf{K}$ with a nonempty interior in $\mathbf{E}$ such that $\mathbf{K}$ is self-dual:

$$\mathbf{K} = \mathbf{K}_* \equiv \{\xi \in \mathbf{E} : \langle \xi, x\rangle_{\mathbf{E}} \geq 0\, \forall x \in \mathbf{K}\}$$

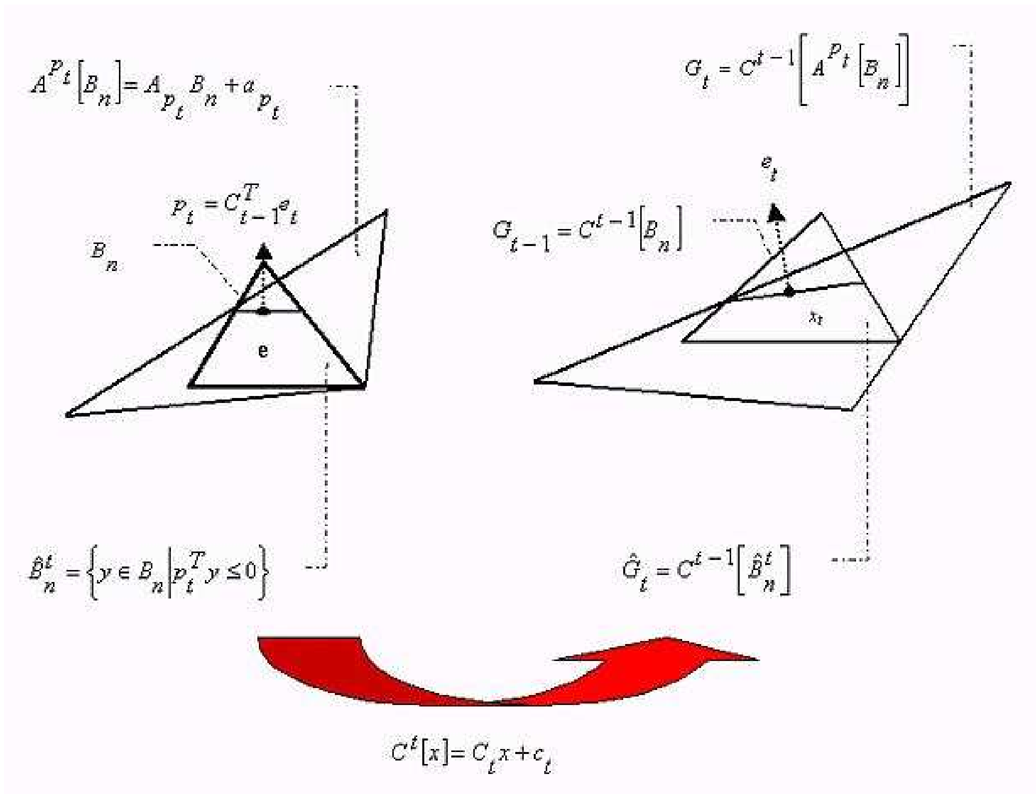

Figure 1.1: Geometry of a step in a stationary Cutting Plane algorithm

and, moreover, $\operatorname{int}\mathbf{K}$ is a *homogeneous space*: for every pair of points $u, v \in \operatorname{int}\mathbf{K}$, there exists a linear invertible mapping on $\mathbf{E}$ which maps $\mathbf{K}$ onto itself and maps u onto v.

It is known [15] that self-scaled cones are exactly cones representable as direct products of *irreducible components* of the following 5 types:

1. Lorentz cone $\mathbf{L}^n_+ = \{x \in \mathbf{R}^n : x_n \geq \sqrt{x_1^2 + ... + x_{n-1}^2}\}, n = 2, 3, ...$;

2. Semidefinite cone $\mathbf{S}^n_+$ (the cone of positive semidefinite real symmetric $n \times n$ matrices), $n = 1, 2, ...$;

3. Hermitian cone $\mathbf{H}^n_+$ (the cone of positive semidefinite Hermitian $n \times n$ matrices with complex entries), $n = 1, 2, ...$;

4. Quaternion cone $\mathbf{Q}^n_+$ (the cone of positive semidefinite Hermitian $n \times n$ matrices with quaternion entries), $n = 1, 2, ...$;

5. Exceptional 27-dimensional *Octonion* cone.

We demonstrate that *if $\mathbf{K}$ is a self-scaled cone with no Octonion irreducible components, then appropriate translation of the set*

$$\mathbf{K}_f = \{x \in \mathbf{K} : \langle f, x \rangle_{\mathbf{E}} = 1\} \qquad [f \in \operatorname{int}\mathbf{K}] \tag{1.12}$$

(treated as a subset of its affine hull) *is, up to a dilatation, a perfect solid with certain explicit value of ω satisfying* (1.10).

In connection with this result, it should be noted that

1. The sets of the form $\mathbf{B}_f$ corresponding to different choices $f \in \operatorname{int}\mathbf{K}$ are images of each other under invertible affine mappings, so that in our context a given self-scaled cone $\mathbf{K}$ produces exactly one stationary polynomial time Cutting Plane method, and this method is capable to solve problems (P) of design dimension by one less than the dimension of $\mathbf{K}$;

2. Our construction covers the two previously known stationary polynomial time Cutting Plane algorithms. Specifically, when speaking about solving n-dimensional problems (P), the Ellipsoid method is given by our construction as applied to the Lorentz cone $\mathbf{L}^{n+1}$. To get the Circumscribed Simplex method, the construction should be applied to the $n+1$-dimensional nonnegative orthant $\mathbf{R}^{n+1}_+$ (which clearly is a self-scaled cone – it is direct product of nonnegative rays $\mathbf{R}_+ = \mathbf{S}^1_+$.

3. Our construction extends dramatically the spectrum of stationary polynomial time Cutting Plane algorithms, adding, e.g., methods associated with *spectahedrons*

$$\mathbf{B} = \left\{ x = [x_{ij}] \in \mathbf{R}^{k\times k} : x = x^T, \mathrm{Tr}(x) = 0, I + x \succeq 0 \right\}$$

(**B** is affine equivalent to the set given by (1.12) as applied to $\mathbf{K} = \mathbf{S}^n_+$ and to the unit matrix in the role of f), or with cross-sections of direct products of Lorentz cones, or with cross-sections of direct products of several Lorentz and several Semidefinite cones, etc.

Chapter 2

The SCP algorithm associated with self-scaled cone

This Chapter contains the main result of the Thesis – the construction of a polynomial time Cutting Plane algorithm associated with a given self-scaled cone $\mathbf{K}$ (with no Octonian irreducible components) in a Euclidean space $\mathbf{E}$. As it was explained in Introduction, to achieve our goal, we should demonstrate that for every $f \in \operatorname{int}\mathbf{K}$ and appropriate $x_f \in \operatorname{int}\mathbf{K}$ the set

$$\mathbf{B}_f = \{x - x_f : x \in \mathbf{K} : \langle f, x - x_f\rangle_{\mathbf{E}} = 0\} \subset \mathbf{F} \equiv \{x \in \mathbf{E} : \langle f, x\rangle_{\mathbf{E}} = 0\}$$

is, up to a dilatation, a *perfect solid* in $\mathbf{F}$, i.e.,

> ($\clubsuit$) $\mathbf{B}_f$ *contains a Euclidean ball centered at the origin, of certain radius* κ, *and is contained in the concentric ball of radius* $\gamma[\mathbf{B}]\kappa$. *Moreover, whenever* $g \in \mathbf{F}$ *is nonzero, the part*
>
> $$\mathbf{B}_f^+[g] = \{x \in \mathbf{B}_f \mid \langle g, x\rangle_{\mathbf{E}} \leq 0\}$$
>
> *of* $\mathbf{B}_f$ *"cut off" the solid* $\mathbf{B}_f$ *by the hyperplane* $\langle g, x\rangle_{\mathbf{E}} = 0$ *passing through the origin can be covered by the image*
>
> $$A^g(\mathbf{B}) + b^g$$
>
> *of* $\mathbf{B}_f$ *under an invertible affine mapping* $x \mapsto A^g x + b^g$ *of* $\mathbf{F}$ *onto itself, and this affine mapping should reduce volumes:*
>
> $$\exists(\omega < 1): \qquad |\mathrm{Det}(A^g)|^{1/\dim \mathbf{F}} \leq \omega \qquad \forall(g \neq 0).$$
>
> *Besides this, given* g, *we should be able to compute efficiently* A^g *and* b^g, *and the quantity* ω *should be at a "polynomial in* $\dim \mathbf{F}$*" distance from 1:*
>
> $$1 - \omega \geq \frac{O(1)}{(\dim \mathbf{F})^\beta}. \qquad (2.1)$$

Note that since the cone $\mathbf{K}$ is self-scaled, all sets of the form $\mathbf{B}$ corresponding to $f, x_f \in \operatorname{int} \mathbf{K}$ are affine images of each other, so that it suffices to prove the outlined "perfectness" statement for just one pair $f, x_f \in \operatorname{int} \mathbf{K}$, no matter which.

Our plan of action is as follows. In Section 2.1, we present an explicit geometric construction which, given on input a (not necessary self-scaled) cone $\mathbf{K}$ in a Euclidean space $\mathbf{E}$, provides a mechanism for covering a set of the form $\mathbf{B}_f^+[g]$ by an affine image of the set $\mathbf{B}_h$ with appropriately chosen h. When $\mathbf{K}$ is self-scaled, $\mathbf{B}_h$ is an affine image of $\mathbf{B}_f$ (since all compact cross-sections of a self-scaled cone by hyperplanes intersecting the interior of the cone are affine equivalent to each other). This, in the self-scaled case the construction from Section 2.1 in fact yields a mechanism of covering $\mathbf{B}_f^+[g]$ by the image of $\mathbf{B}_f$ under certain explicit affine mapping. It turns out that with the parameters of the construction properly chosen, *the corresponding affine mapping reduces volumes and thus provides us with the desired result.* This fact (Theorem 2.1) is established in Section 2.5; its derivation is based upon the intermediate results (Propositions 2.2, 2.3, 2.4) proved in Section 2.4. In our analysis, we heavily exploit the theory of *self-concordant functions* developed in [13] and, of course, the basic facts on self-scaled cones; the related background is outlined in Sections 2.2 and 2.3, respectively. In the concluding Section 2.6 we present and discuss the complexity characteristics of the polynomial time Cutting Plane methods we have developed.

2.1 Geometry of the construction

2.1.1 The construction

Consider a cone $\mathbf{K}$ (convex, closed, pointed and with a nonempty interior) in a Euclidean space $\mathbf{E}$ with inner product $\langle \cdot, \cdot \rangle$, and let $\mathbf{j}$ be a once for ever fixed unit vector in $\mathbf{E}$ which belongs both to the interior of the cone $\mathbf{K}$ and the interior of the dual to $\mathbf{K}$ cone $\mathbf{K}_*$. Let us set

$$\mathbf{F} = \Big\{ x \in \mathbf{E} \mid \langle \mathbf{j}, x \rangle = 0 \Big\}.$$

Then $\mathbf{F}$ is a Euclidean space.
Note that whenever $f \in \mathbf{K}_*$, the cross-section of $\mathbf{K}$ by the hyperplane $\Pi[f] = \Big\{ x \mid f^T x = f^T \mathbf{j} \Big\}$ passing through $\mathbf{j}$ and orthogonal to f, i.e., the set

$$\mathbf{K}[f] = \Big\{ x \in \mathbf{K} \mid \langle x - \mathbf{j}, f \rangle = 0 \Big\}$$

is convex and compact, and $\mathbf{j}$ is in its relative interior. In particular, the set

$$\mathbf{D} = \Big\{h \in \mathbf{F} \mid \mathbf{j} + h \in \mathbf{K}\Big\} = \mathbf{K}[\mathbf{j}] - \mathbf{j}$$

is a convex solid in $\mathbf{F}$. Geometrically: $\mathbf{D}$ is the orthogonal projection of the intersection of the hyperplane $\Pi[\mathbf{j}] = \mathbf{F} + \mathbf{j}$ with the cone $\mathbf{K}$ onto the plane $\mathbf{F}$.

Example 1. Let

$$\begin{aligned} \mathbf{K} &= \mathbf{R}^n_+ \equiv \{x \in \mathbf{R}^n : x \geq 0\}, \\ \mathbf{j} &= n^{-1/2}(1, \ldots, 1)^T. \end{aligned}$$

In this case,

$$\mathbf{K}_* = \mathbf{K} = \mathbf{R}^n_+,$$

$$\mathbf{F} = \{x \in \mathbf{R}^n : \sum_{i=1}^n x_i = 0\},$$

$$\mathbf{D} = \{x \in \mathbf{R}^n : \sum_{i=1}^n x_i = 0, x_i \geq -n^{-1/2}, i = 1, \ldots, n\}.$$

Geometrically: $\mathbf{D}$ is the perfect simplex in the $(n-1)$-dimensional Euclidean space $\mathbf{F}$, see Fig. 2.1.1

Example 2. Let

$$\mathbf{K} = \mathbf{L}^{n+1} \equiv \{(x,t) \in \mathbf{R}^{n+1} = \mathbf{R}^n_x \times \mathbf{R}^t : t \geq \| x \|_2\}$$

(this is called the Lorentz, (or ice-cream, or second order) cone), and let

$$\mathbf{j} = (0, \ldots, 0, 1)^T.$$

In this case

$$\mathbf{K}_* = \mathbf{K} = \mathbf{L}^{n+1},$$

$$\mathbf{F} = \{(x, 0) : x \in \mathbf{R}^n\},$$

$$\mathbf{D} = \{(x, 0) : x \in \mathbf{R}^n, \| x \|_2 \leq 1\}.$$

Geometrically: $\mathbf{D}$ is the unit Euclidean ball in $\mathbf{F} = \mathbf{R}^n$.

We have described certain way of producing solids from cones; note that *every* solid can be obtained in such a way with appropriately chosen cone $\mathbf{K}$ and $\mathbf{j} \in \operatorname{int} \mathbf{K} \cap \operatorname{int} \mathbf{K}_*$.

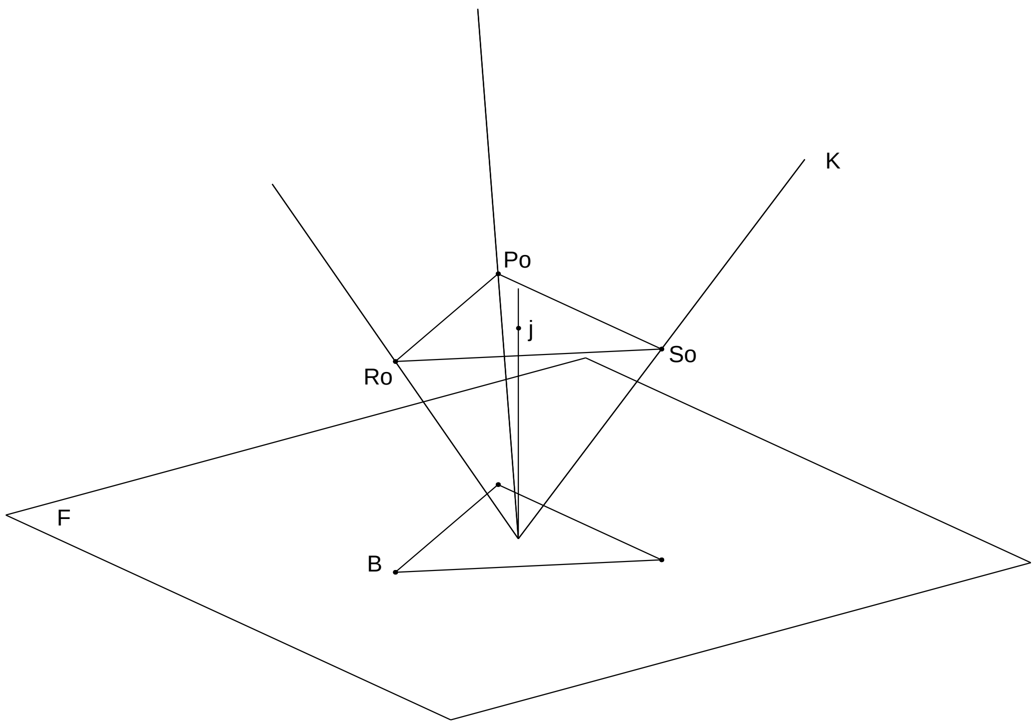

Figure 2.1: Perfect simplex

Now assume that we treat the above solid $\mathbf{D}$ as the solid responsible for a stationary cutting plane algorithm, and set $\mathbf{e} = 0$. How could we ensure ($\clubsuit$)?

Assume that we are given a unit vector $g \in \mathbf{F}$ and, as required in ($\clubsuit$), are interested to cover the set

$$\mathbf{D}^+[g] = \left\{ x \in \mathbf{D} \mid \langle g, x \rangle \leq 0 \right\}$$

by an affine image of $\mathbf{D}$. Let us look at the following geometric construction:

1. Given g, we choose a positive real λ and set

$$f = \mathbf{j} + \lambda g.$$

Since $\mathbf{j} \in \operatorname{int} \mathbf{K}_*$, the vector f, for all not large coefficients λ, also belongs to $\operatorname{int} \mathbf{K}_*$. Consequently, the intersection $\mathbf{K}[f]$ of $\mathbf{K}$ and the hyperplane orthogonal to f and passing through $\mathbf{c}$ is a convex compact set.
Geometric intuition says to us that the projection $\widehat{\mathbf{D}}[f]$ of the set $\mathbf{K}[f]$ onto $\mathbf{F}$ all the time contains $\mathbf{D}^+[g]$, see Fig. 2.1.1. When $\lambda = 0$, $f \equiv \lambda g + \mathbf{j} = \mathbf{j}$, and the cross-section $\mathbf{K}[f]$ of the cone $\mathbf{K}$ by the hyperplane $\Pi[f]$ is the domain $P_o Q_o R_o$. the orthogonal projection of this domain onto $\mathbf{F}$ is exactly $\mathbf{D}$. On our picture, $\mathbf{D}^+[g]$ is the dashed part of $\mathbf{D}$.
When we increase λ, starting from zero, the hyperplane $\Pi[f]$ rotates, and eventually its intersection with $\mathbf{K}$ becomes unbounded. This, however, does not happen while λ is not too large, and the projection $\widehat{\mathbf{D}}[f] = prs$ of the set $\mathbf{K}[f] = PRS$ onto $\mathbf{F}$, as is seen on the picture, all the time contains $\mathbf{D}^+[g]$.

2. In fact, the set $\widehat{\mathbf{D}}[f]$ can be shrunk in the direction of g in such a way that the shrunken set, let it be called $\mathbf{D}[f]$, still contains $\mathbf{D}^+[g]$. This is what we see on the above picture – when we shrink (but not too much!) the set prs in the direction of g, the shrunken set still contains the dashed part of $\mathbf{D}$, which is our $\mathbf{D}^+[g]$. This fact is true in the general case as well (see below).

3. It follows that if the cone $\mathbf{K}$ is such that all its compact cross-sections by hyperplanes are affine images of each other, then the above construction provides us with certain mechanism of building affine images of $\mathbf{D}$ which contain the set $\mathbf{D}^+[g]$ – in this case, all compact sets $\mathbf{K}[f] = \mathbf{K} \cap \Pi[f]$ are affine images of $\mathbf{D}$, and consequently so are their projections $\widehat{\mathbf{D}}[f]$ onto $\mathbf{F}$, as well as the shrinkages $\mathbf{D}[f]$ of these projections (since shrinkage is an affine

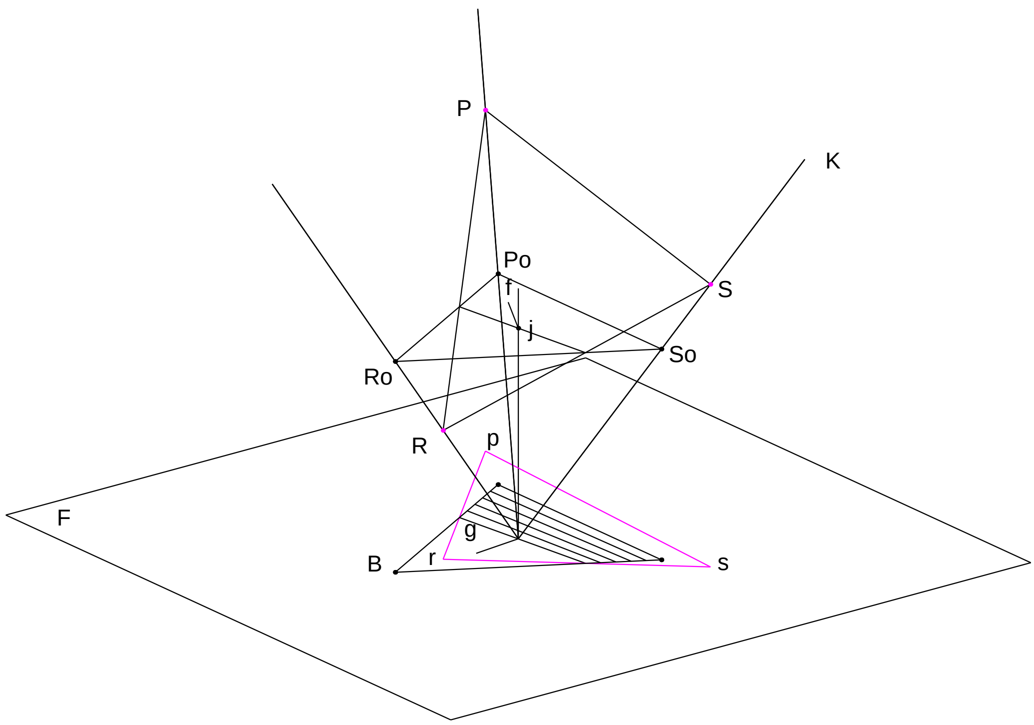

Figure 2.2: The construction.

mapping). As we have mentioned, all sets $\mathbf{D}[f]$ contain $\mathbf{D}^+[g]$.
Now we can play with the parameter λ in order to find the smallest in volume projection, thus getting (hopefully good!) covering of $\mathbf{D}^+[g]$ by an affine image of $\mathbf{D}$.

There are cones where the required property indeed takes place. The first example is nonnegative orthant $-$ one can easily verify that all its compact intersections with hyperplanes are of the form

$$\mathcal{S} = \{x \geq 0, \sum_{i=1}^{n} c_i x_i = c_0\},$$

where $c_0, c_1, \ldots, c_n$ are positive. But all these sets are affinely equivalent to the standard simplex

$$\Delta = \Big\{x \geq 0, \sum_{i=1}^{n} x_i = 1\Big\}.$$

To realize this, look what happens with $\mathcal{S}$ under the linear mapping ("scaling")

$$x \mapsto \left(\frac{c_1}{c_0} x_1, \ldots, \frac{c_n}{c_0} x_n\right)^T.$$

Another example is the Lorentz cone; here all compact intersections of the cone with the hyperplanes are ellipsoids, and they are affinely equivalent to the standard Euclidean ball.

2.1.2 A geometric fact

Now we can prove the geometric fact which has been mentioned in item 2. Here is the precise statement:

Proposition 2.1 *Let*

- $\mathbf{K}$ *be a cone (closed, pointed, convex and with a nonempty interior) in a Euclidean space* $\mathbf{E}$ *with inner product* $\langle \cdot, \cdot \rangle$*;*
- $\mathbf{j}$ *be a unit vector from* $\operatorname{int} \mathbf{K} \cap \operatorname{int} \mathbf{K}_*$*, where* $\mathbf{K}_*$ *is the cone dual to* $\mathbf{K}$*;*
- $\mathbf{F}$ *be the orthogonal complement of* $\mathbf{j}$*:*

$$\mathbf{F} = \Big\{x \in \mathbf{E} \mid \langle \mathbf{j}, x \rangle = 0\Big\};$$

- $\mathbf{D}$ *be the set given by*

$$\mathbf{D} = \Big\{ h \in \mathbf{F} \mid \mathbf{j} + h \in \mathbf{K} \Big\}.$$

Let, further, $g \in \mathbf{F}$ *be a unit vector,* λ *be a positive real such that*

$$f = \mathbf{j} + \lambda g \in \operatorname{int} \mathbf{K}_*,$$

and let $\widehat{x} \in \mathbf{D}$ *be a vector and* $\alpha \geq 0$ *be a real such that*

$$\langle g, \widehat{x} \rangle \leq -\alpha.$$

Let P *be the orthoprojector of* $\mathbf{E}$ *onto* $\mathbf{F}$:

$$P(x) = x - \langle \mathbf{j}, x \rangle \mathbf{j},$$

and $L[x]$ *be the linear transformation of* $\mathbf{F}$ *given by*

$$L[x] = x + \lambda \langle g, x \rangle \widehat{x} + \lambda \alpha \widehat{x}$$

($L(\cdot)$ *"shrinks"* $\mathbf{F}$ *in the direction of* g*). Consider the sets*

$$Y^+ = \Big\{ u \in \mathbf{K} \mid \langle u - (1 - \lambda\alpha)\mathbf{j}, f \rangle = 0 \Big\},$$

$$Y = P(Y^+).$$

Then the image of the set Y *under the affine mapping* $L(\cdot)$ *contains the set*

$$\mathbf{D}^+[g] = \Big\{ x \in \mathbf{D} \mid \langle g, x \rangle \leq -\alpha \Big\}.$$

Proof. In order to show that $\mathbf{D}^+[g] \subset L[P(Y^+)]$, it is enough to prove that if $x \in \mathbf{D}^+[g]$ then $x \in L[P(Y^+)]$. Assume that $x \in \mathbf{D}^+[g]$. Let us first find $y \in \mathbf{F}$ such that $L[y] = x$, i.e., $y \in \mathbf{F}$ which satisfies the equation $y + \lambda \langle g, y \rangle \widehat{x} + \lambda \alpha \widehat{x} = x$. Multiplying both sides of this equation by g, we get

$$\langle g, y \rangle = \frac{\langle g, x \rangle - \lambda \alpha \langle g, \widehat{x} \rangle}{1 + \lambda \langle g, \widehat{x} \rangle}, \tag{2.2}$$

whence

$$y = x - \lambda \frac{\langle g, x \rangle - \lambda \alpha \langle g, \widehat{x} \rangle}{1 + \lambda \langle g, \widehat{x} \rangle} \widehat{x} - \lambda \alpha \widehat{x} = x - \lambda \frac{\langle g, x \rangle + \alpha}{1 + \lambda \langle g, \widehat{x} \rangle} \widehat{x}. \tag{2.3}$$

Now let us find a point $y^+ \in \mathbf{M} \equiv \{u \in \mathbf{E} : \langle u - (1-\lambda\alpha)\mathbf{j}, f\rangle = 0\}$ such that $P(y^+) = y$. We have $y^+ = y + t\mathbf{j}$, where t should ensure the relation

$$\langle y + t\mathbf{j} - (1-\lambda\alpha)\mathbf{j}, f\rangle = 0,$$

which is a linear equation for t. Solving this equation, we get

$$t = 1 - \lambda\frac{\langle g, x\rangle + \alpha}{1 + \lambda\langle g, \widehat{x}\rangle},$$

whence

$$\begin{array}{rcl} y^+ & = & x - \lambda\frac{\langle g,x\rangle+\alpha}{1+\lambda\langle g,\widehat{x}\rangle}\widehat{x} + \left[1 - \lambda\frac{\langle g,x\rangle+\alpha}{1+\lambda\langle g,\widehat{x}\rangle}\right]\mathbf{j} \\ & = & (x+\mathbf{j}) + \frac{-\lambda[\langle g,x\rangle+\alpha]}{1+\lambda\langle g,\widehat{x}\rangle}(\widehat{x}+\mathbf{j}). \end{array} \tag{2.4}$$

By construction, we have $L(P(y^+)) = x$; in order to conclude the proof, it suffices to verify that $y^+ \in Y^+ = \mathbf{M} \cap \mathbf{K}$, i.e., due to $y^+ \in \mathbf{M}$, that $y^+ \in \mathbf{K}$. The verification is as follows. We first claim that $1 + \lambda\langle g, \widehat{x}\rangle > 0$. Indeed, since $\widehat{x} \in \mathbf{D}$, we have $0 \neq \mathbf{j} + \widehat{x} \in \mathbf{K}$, while $\mathbf{j} + \lambda g \in \operatorname{int}\mathbf{K}_*$. Since $\mathbf{K}_*$ is the cone dual to $\mathbf{K}$, we have

$$0 < \langle \mathbf{j} + \widehat{x}, \mathbf{j} + \lambda g\rangle = 1 + \lambda\langle g, \widehat{x}\rangle.$$

Moreover, we have $\mathbf{j} + x \in \mathbf{K}$ due to $\mathbf{D}^+[g] \subset \mathbf{D}$ and $x \in \mathbf{D}^+[g]$.
Since $\langle g, \widehat{x}\rangle \leq -\alpha$ and $1 + \lambda\langle g, \widehat{x}\rangle > 0$, the denominator in the fractions in (2.4) is positive, and the numerator in the concluding fraction is nonnegative. Since $\mathbf{K}$ is a cone and, as we have just verified, y^+ is a combination, with nonnegative coefficients, of the vectors $\mathbf{j} + \widehat{x}$ and $\mathbf{j} + x$ from $\mathbf{K}$, we have $y^+ \in \mathbf{K}$. ■

It turns out that there exists a family of cones (the so called *self-scaled* ones) where the outlined construction ensures the desired property (♣), and, moreover, results in efficient stationary cutting plane algorithms.

2.2 Preliminaries on self-concordant functions

We start by summarizing the properties of self-concordant functions and barriers we will frequently use in the sequel; for the proofs, see [13].

2.2.1 Notation

In what follows letters like $\mathbf{E}$, $\mathbf{F}$, etc., denote Euclidean linear spaces; corresponding inner products are denoted $\langle\cdot,\cdot\rangle_{\mathbf{E}}$, $\langle\cdot,\cdot\rangle_{\mathbf{F}}$. We skip subscripts in

$\langle\cdot,\cdot\rangle$, when it is clear from the context what the Euclidean space in question is.

For a linear operator $x \mapsto Bx : \mathbf{F} \to \mathbf{E}$, B^* stands for the conjugate operator: $\langle y, Bx\rangle_{\mathbf{E}} = \langle B^* y, x\rangle_{\mathbf{F}}$. We write $B \succeq 0$ ($B \succ 0$) to express that B is a symmetric and positive semidefinite (resp., positive definite) operator on E, with evident interpretation of relations like $A \succeq B$ or $B \prec A$.

We associate with an operator $B \succ 0$ on $\mathbf{E}$ a conjugate pair of Euclidean norms on $\mathbf{E}$:

$$\begin{array}{rcl} \|x\|_B & = & \langle x, Bx\rangle^{1/2}, \\ \|x\|_B^* & = & \max\{\langle x, y\rangle : \|y\|_B \le 1\} = \|x\|_{B^{-1}}. \end{array}$$

From now on, we set

$$\rho(t) \quad = \quad t - \ln(1+t) \quad \left[= \tfrac{t^2}{2}(1+o(t)),\, t \to 0\right], \; .$$

For a convex C^2 on its domain and nondegenerate ($f'' \succ 0$) function $f : \mathbf{E} \mapsto \mathbf{R} \cup \{+\infty\}$ and $x \in \mathrm{Dom} f$, we define the *Newton decrement of f at x* as

$$\lambda(f,x) = \|f'(x)\|^*_{f''(x)} = \sqrt{[f'(x)]^T [f''(x)]^{-1} f'(x)}.$$

2.2.2 Self-concordant functions and barriers: definitions

A convex function $f : \mathbf{E} \to \mathbf{R} \cup \{+\infty\}$ is called *self-concordant* (s.c.), if the domain Q of f is open, f is C^3 on Q, satisfies the differential inequality

$$\left| \frac{d^3}{dt^3}\Big|_{t=0} f(x+th) \right| \le 2 \left(\frac{d^2}{dt^2}\Big|_{t=0} f(x+th) \right)^{3/2} \qquad \forall (x \in Q, h \in \mathbf{E}) \tag{2.5}$$

and is a barrier for Q: $f(x_i) \to \infty$ along every sequence $\{x_i\} \subset Q$ converging to a boundary point of Q.

A s.c. function f is called *nondegenerate*, if its Hessian $f''(x)$ is nondegenerate at some (and then automatically at every) point $x \in \mathrm{Dom} f$.

Let $\nu \ge 1$. Function f is called *ν-self-concordant barrier* (ν-s.c.b.) *for* $\mathrm{cl}\mathrm{Dom} f$, if f is self-concordant and

$$\left| \frac{d}{dt}\Big|_{t=0} f(x+th) \right| \le \sqrt{\nu} \left(\frac{d^2}{dt^2}\Big|_{t=0} f(x+th) \right)^{1/2} \qquad \forall (x \in \mathrm{Dom} f, h \in \mathbf{E}). \tag{2.6}$$

A nondegenerate s.c. function f is ν-s.c.b. if and only if $\lambda(f,x) \le \sqrt{\nu}$ for all $x \in \mathrm{Dom} f$.

Let $\mathbf{E}_+$ be a cone in E (closed, pointed, convex and with a nonempty interior). A *ν-logarithmically homogeneous self-concordant barrier* (l.-h.s.-c-b) *for* $\mathbf{E}_+$ is a self-concordant function f with $\mathrm{Dom} f = \mathrm{int}\,\mathbf{E}_+$ such that

$$\forall (x \in \mathrm{int}\, E, t > 0) : f(tx) = f(x) - \nu \ln t.$$

A ν-logarithmically homogeneous self-concordant barrier for $\mathbf{E}_+$ is a self-concordant barrier for $\mathbf{E}$ with the self-concordance parameter equal to ν. Such a barrier satisfies a number of useful identities, specifically:

$$\begin{array}{rcll} \multicolumn{4}{l}{\forall (x \in \mathrm{int}\,\mathbf{E}_+, t > 0) :} \\ f'(tx) & = & t^{-1} f'(x) & (a) \\ f''(tx) & = & t^{-2} f''(x) & (b) \\ f'(x) & = & f''(x)x & (c) \\ \langle f'(x), x\rangle & = & -\langle f''(x)x, x\rangle = -\nu & (d) \end{array} \tag{2.7}$$

2.2.3 Basic properties of self-concordant functions

We summarize these properties in the following list.

SC.I. [Stability w.r.t. linear operations]

1) *Let f_i, $i = 1, .., m$, be s.c. functions on* $\mathbf{E}$, *and let $\lambda_i \geq 1$, and let the function $f(x) = \sum\limits_i \lambda_i f_i(x)$ possess a nonempty domain. Then the function f is s.c. If every f_i is ν_i-s.c.b., then f is $(\sum\limits_i \lambda_i \nu_i)$-s.c.b. If every f_i is a ν_i-l.-h.s.-c.b., then f is $(\sum\limits_i \lambda_i \nu_i)$-l.-h.s.c.b.*

2) *Let f be s.c. on* $\mathbf{E}$, *and let $y \mapsto Ay + b$ be an affine embedding from Euclidean space* $\mathbf{F}$ *to* $\mathbf{E}$ *with image intersecting* $\mathrm{Dom} f$. *Then the function $g(y) = f(Ay + b)$ is s.c. If f is ν-s.c.b., then so is g. If $b = 0$ and f is ν-l.-h.s.c.b., then so is g.*

SC.II. [Local behaviour and damped Newton step] *Let f be a nondegenerate s.c. function with $Q = \mathrm{Dom} f$. Then*

1) *For every $x \in Q$, the ellipsoid $\{y : \|y - x\|_{f''(x)} < 1\}$ is contained in Q. Besides this,*

$$\begin{array}{rcll} r \equiv \|y - x\|_{f''(x)} < 1 & \Rightarrow & (1-r)^2 f''(x) \preceq f''(y) \preceq (1-r)^{-2} f''(x) & (a) \\ r \equiv \|y - x\|_{f''(x)} < 1 & \Rightarrow & f(y) \leq f(x) + \langle f'(x), y - x\rangle + \rho(-r) & (b.1) \\ y \in Q, r \equiv \|y - x\|_{f''(x)} & \Rightarrow & f(y) \geq f(x) + \langle f'(x), y - x\rangle + \rho(r). & (b.2) \end{array} \tag{2.8}$$

2) *For $x \in Q$, we define the damped Newton iterate of x as*

$$x_+ = x - \frac{1}{1 + \lambda(f, x)} [f''(x)]^{-1} f'(x).$$

For every $x \in Q$ we have

$$\begin{array}{rcll} x_+ & \in & Q & (a) \\ f(x_+) & \leq & f(x) - \rho(\lambda(f,x)) & (b) \\ \lambda(f,x_+) & \leq & 2\lambda^2(f,x). & (c) \end{array} \qquad (2.9)$$

SC.III. [Additional properties of s.c.b.'s] *Let f be a nondegenerate ν-s.c.b., and let $Q = \mathrm{Dom} f$. Then*

1) *one has*

$$\begin{array}{rl} \forall(x,y \in Q): \quad \langle y-x, f'(x)\rangle \leq \nu & (a) \\ \forall(x,y \in Q): \langle y-x, f'(x)\rangle \geq 0 \Rightarrow \|y-x\|_{f''(x)} \leq \nu + 2\sqrt{\nu} & (b) \end{array} \qquad (2.10)$$

2.3 Self-scaled cones

A self-scaled cone is a closed convex pointed cone $\mathbf{K}$ in an Euclidean space $\mathbf{E}$ with inner product $\langle\cdot,\cdot\rangle_{\mathbf{E}}$ such that

1. $\mathbf{K}$ is *self-dual*: the cone dual to $\mathbf{K}$, i.e., the cone

$$\mathbf{K}_* = \{\xi \in \mathbf{E} : \langle \xi, x\rangle_{\mathbf{E}} \geq 0 \,\forall x \in \mathbf{K}\}$$

   is $\mathbf{K}$ itself;

2. $\mathrm{int}\,\mathbf{K}$ is a "homogeneous space": for every pair of points $u, v \in \mathrm{int}\,\mathbf{K}$ there exists an invertible linear mapping $y = Ax$ which maps the cone onto itself and maps u to v: $A\mathbf{K} = \mathbf{K}$, $Au = v$.

Given two self-scaled cones, one can form their direct product, which clearly is a self-scaled cone in the direct product of the respective Euclidean spaces. A well-known fact [15, 1] is that every self-scaled cone can be obtained, by taking direct products, from "irreducible" self-scaled cones (those which cannot be represented as direct products of self-scaled components), and that the irreducible self-scaled cones admit complete description. Specifically, every irreducible self-scaled cone, up to isomorphism, is

1. either a Lorentz cone $\mathbf{L}^n_+$;

2. or the cone $\mathbf{S}^n_+$ of positive semidefinite matrices in the space $\mathbf{S}^n$ of symmetric $n \times n$ matrices (the latter space is equipped with the Frobenius inner product $\langle A, B\rangle = \mathrm{Tr}(AB^T)$);

3. or the cone $\mathbf{H}^n_+$ of positive semidefinite Hermitian matrices in the space $\mathbf{H}^n$ of $n \times n$.

   A Hermitian $n \times n$ matrix can be thought of as a real symmetric $2n \times 2n$ matrix x with n^2 2×2 blocks of the form $\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}$. $\mathbf{H}^n$ is the linear space of all Hermitian $n \times n$ matrices, equipped with the inner product $\langle x, y\rangle_{\mathbf{H}_n} = \frac{1}{2}\mathrm{Tr}(xy)$. The cone $\mathbf{H}^n_+$ is comprised of all positive semidefinite Hermitian $n \times n$ matrices.

4. or the cone of positive semidefinite quaternion matrices in the space $\mathbf{Q}^n$ of $n \times n$ Hermitian quaternion matrices

   An $n \times n$ Hermitian quaternion matrix can be thought of as a real symmetric $4n \times 4n$ matrix with n^2 4×4 blocks of the form $\begin{bmatrix} \alpha & -\beta & \gamma & -\delta \\ \beta & \alpha & \delta & \gamma \\ -\gamma & \delta & \alpha & -\beta \\ \delta & -\gamma & \beta & \alpha \end{bmatrix}$. $\mathbf{Q}^n$ is the linear space of all Hermitian quaternion $n \times n$ matrices, equipped with the inner product $\langle x, y\rangle_{\mathbf{Q}_n} = \frac{1}{4}\mathrm{Tr}(xy)$. The cone $\mathbf{Q}^n_+$ is comprised of all positive semidefinite Hermitian quaternion $n \times n$ matrices.

5. or the exceptional 27-dimensional "Octonian" cone.

Note that the specific feature of the Hermitian/Hermitian quaternion matrices which makes the corresponding cones self-scaled comes from the following well-known algebraic fact:

Lemma 2.1 *Let $s = s_1^{p_1} s_2^{p_2} \ldots s_m^{p_m}$, where s_j are symmetric matrices of the same size, and p_j are integers. If all the matrices s_j are Hermitian (Hermitian quaternion) and s is well-defined and symmetric, then s is Hermitian (respectively, Hermitian quaternion).*

It is easily seen that if $\mathbf{K}$ is a self-scaled, then all sets of the form

$$\mathbf{K}_f = \{x \in \mathbf{K} : \langle f, x\rangle = 1\}$$

given by $f \in \mathrm{int}\,\mathbf{K}$ (in other words, all *compact* cross-sections of $\mathbf{K}$ by hyperplanes not passing through the origin) are solids in their affine hulls, and all these solids are affinely equivalent to each other. The main fact discovered in our research is that

> (!) *If all irreducible components of $\mathbf{K}$ are distinct from the Octonian cone*[1)]*, then the aforementioned solids do satisfy (♣) and thus give rise to polynomial time SCP algorithms.*

[1)] This is the only case we have considered

It should be mentioned that the two known polynomial time SCP algorithms – the Ellipsoid and the Simplex ones – correspond to a very particular cases of (!): the Ellipsoid method – to the case when $\mathbf{K} = \mathbf{L}^n_+$, and the Simplex method – to the case when $\mathbf{K}$ is the nonnegative orthant, i.e., direct product of the simplest – one-dimensional – semidefinite cones $\mathbf{S}^1_+ = \mathbf{R}_+$.

Our current goal is to establish (!) for the case when $\mathbf{K}$ is a self-scaled cone with Lorentz and Semidefinite irreducible components.

2.4 Equipped spaces

Let us call an *equipped* space a quadruple $(\mathbf{E}, \mathbf{E}_+, E, \mathbf{e})$, where

- $\mathbf{E}$ is a Euclidean space with inner product $\langle \cdot, \cdot \rangle_{\mathbf{E}}$, and associated Euclidean norm $\| \cdot \|_{\mathbf{E}}$;

- $\mathbf{E}_+$ is a closed convex pointed cone in $\mathbf{E}$ with a nonempty interior which is self-dual:
$$\mathbf{E}_+ = \{x \mid \langle x, y \rangle_{\mathbf{E}} \geq 0 \quad \forall y \in \mathbf{E}_+\};$$

- E is a logarithmically homogeneous self-concordant barrier for the cone $\mathbf{E}_+$ (see Section 2.2.2); the parameter of self-concordance of this barrier will be denoted by $\nu(E)$;

- $\mathbf{e}$ is a vector from $\operatorname{int} \mathbf{E}_+$ such that $E''(\mathbf{e})$ is the unit matrix $I_{\dim \mathbf{E}}$.

2.4.1 Simple equipped spaces

We shall be interested in the following four series of equipped spaces, which we call *simple*:

1. $\mathcal{L}^k = (\mathbf{L}^k, \mathbf{L}^k_+, L^k, \mathbf{l}^k), \quad k \geq 0:$

- $\mathbf{L}^k$ is the standard Euclidean coordinate space $\mathbf{R}^{k+1}$;
- $\mathbf{L}^k_+$ is the Lorentz cone in $\mathbf{L}^k$:
$$\mathbf{L}^k_+ = \{(x_0, x_1, \ldots, x_k) \in \mathbf{R}^{k+1} : x_0 \geq \sqrt{x_1^2 + x_2^2 + \ldots + x_k^2}\};$$
- $L^k(x) = -\ln(x_0^2 - x_1^2 - x_2^2 - \ldots - x_k^2) \quad [\nu(L^k) = 2];$

- $\mathbf{l}^k = \begin{pmatrix} \sqrt{2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

2. $\mathcal{S}^k = (\mathbf{S}^k, \mathbf{S}^k_+, S^k, \mathbf{s}^k), \quad k \geq 1:$

- $\mathbf{S}^k$ is the space of $k \times k$ symmetric matrices with real entries equipped with the standard Frobenius inner product $\langle x, y\rangle_{\mathbf{S}^k} = \mathrm{Tr}(xy)$;
- $\mathbf{S}^k_+$ is the cone of positive semidefinite matrices from $\mathbf{S}^k$;
- $S^k = -\ln \mathrm{Det}(x) \quad [\nu(S^k) = k]$;
- $\mathbf{s}^k$ is the $k \times k$ unit matrix.

3. $\mathcal{H}^k = (\mathbf{H}^k, \mathbf{H}^k_+, H^k, \mathbf{h}^k)$, $k \geq 1$:

- $\mathbf{H}^k$ is the space of $k \times k$ Hermitian matrices (recall that the elements of this space are treated as $2k \times 2k$ real symmetric matrices, and that $\langle x, y\rangle_{\mathbf{H}^k} = \frac{1}{2}\mathrm{Tr}(xy)$);
- $\mathbf{H}^k_+$ is the cone of positive semidefinite matrices from $\mathbf{H}^k$;
- $H^k(x) = -\frac{1}{2}\ln \mathrm{Det}(x) = -\frac{1}{2}\ln \mathrm{Det}(R[x])$ $[\nu(H^k) = k]$;
- $\mathbf{h}^k$ is the unit Hermitian $k \times k$ matrix (i.e., $2k \times 2k$ real unit matrix).

4. $\mathcal{Q}^k = (\mathbf{Q}^k, \mathbf{Q}^k_+, Q^k, \mathbf{q}^k)$, $k \geq 1$:

- $\mathbf{Q}^k$ is the space of $k \times k$ Hermitian quaternion matrices (recall that the elements of this space are treated as $4k \times 4k$ real symmetric matrices, and that $\langle x, y\rangle_{\mathbf{Q}^k} = \frac{1}{4}\mathrm{Tr}(xy)$);
- $\mathbf{Q}^k_+$ is the cone of positive semidefinite matrices from $\mathbf{Q}^k$;
- $Q^k(x) = -\frac{1}{4}\ln \mathrm{Det}(x)$ $[\nu(Q^k) = k]$;
- $\mathbf{q}^k$ is the $k \times k$ unit matrix with quaternion entries, i.e., $4k \times 4k$ unit real matrix.

The fact that the aforementioned barriers indeed are self-concordant logarithmically homogeneous barriers with the indicated values of the parameters was established in [13] for the case of $\mathcal{L}^k$, $\mathcal{S}^k$ and in [6] for the remaining cases.

The following two facts are crucial for us:

Proposition 2.2 *Let $\mathcal{E} = (\mathbf{E}, \mathbf{E}_+, E, \mathbf{e})$ be an equipped space from one of the above four series. Then for every $s \in \operatorname{int} \mathbf{E}_+$ the mapping $E''(s)$ is a nondegenerate linear mapping which maps $\mathbf{E}_+$ onto itself.*

Proposition 2.3 *Let $\mathcal{E} = (\mathbf{E}, \mathbf{E}_+, E, \mathbf{e})$ be an equipped space from one of the above two series, and let $d \in \operatorname{int} \mathbf{E}_+$. Then*

(A). *There exists a unique $s \in \operatorname{int} \mathbf{E}_+$ such that*

$$E''(s)\mathbf{e} = d, \tag{2.11}$$

(B). *For the above s it holds*

$$\ln \operatorname{Det}(E''(s)) = \Theta[\mathcal{E}][E(d) - E(\mathbf{e})],$$

where

$$\begin{array}{rcl} \Theta[\mathcal{L}^k] & = & \frac{k+1}{2} \\ \Theta[\mathcal{S}^k] & = & \frac{k+1}{2} \\ \Theta[\mathcal{H}^k] & = & k \\ \Theta[\mathcal{Q}^k] & = & 2k-1 \end{array} \tag{2.12}$$

Proof of Proposition 2.2. Consider 4 cases.

Case 1: $\mathcal{E} = \mathcal{L}^k$. Let $s = \begin{pmatrix} t \\ x \end{pmatrix} \in \operatorname{int} \mathbf{E}_+$, $ds = \begin{pmatrix} dt \\ dx \end{pmatrix} \in \mathbf{E}$, and let

$$E(\tau, \xi) = -\ln(\tau^2 - \xi^T \xi)$$

be the canonical barrier for the Lorentz cone $\mathbf{E}_+$. We have

$$DE(s)[ds] = -\frac{2t}{t^2 - x^T x}\, dt + \frac{2x^T}{t^2 - x^T x}\, dx, \tag{2.13}$$

$$D^2E(s)[ds, ds] = \frac{-2dt^2 + 2(dx)^T dx}{t^2 - x^T x} + \frac{(2tdt - 2x^T dx)^2}{(t^2 - x^T x)^2}. \tag{2.14}$$

Setting

$$J = \begin{pmatrix} -1 & 0 \\ 0 & I_k \end{pmatrix},\ d = t^2 - x^T x, \tag{2.15}$$

we get from (2.14), (2.15):

$$\begin{aligned}
D^2E(s)[ds,ds] &= 2\frac{(ds)^T Jds}{d} + 4\frac{(s^T Jds)^2}{d^2} \\
&= 2\frac{(ds)^T Jds}{d} + 4\frac{s^T Jdss^T Jds}{d^2} \\
&= 2\frac{(ds)^T Jds}{d} + 4\frac{\left[(J^T s)^T ds\right]^T s^T Jds}{d^2} \\
&= 2\frac{(ds)^T Jds}{d} + 4\frac{(ds)^T \left[(J^T s)^T\right]^T s^T Jds}{d^2} \\
&= 2\frac{(ds)^T Jds}{d} + 4\frac{(ds)^T Jss^T Jds}{d^2} \\
&= (ds)^T \left[\frac{2}{d}J + \frac{4}{d^2}Jss^T J\right] ds.
\end{aligned}$$

Consequently,

$$E''(s) = \frac{2}{d}J + \frac{4}{d^2}Jss^T J. \tag{2.16}$$

Let us prove that the mapping $\alpha \mapsto E''(s)\alpha$ maps the cone $\mathbf{E}_+$ into itself. Let $E''(s)\alpha = \beta$, where $\alpha = \begin{pmatrix} r \\ y \end{pmatrix} \in \mathbf{E}_+$, $\beta = \begin{pmatrix} q \\ z \end{pmatrix}$, with $r, q \in \mathbf{R}$ and $y, z \in \mathbf{R}^k$. Taking into account (2.16), we have

$$E''(s)\begin{pmatrix} r \\ y \end{pmatrix} = -\frac{2}{d}\begin{pmatrix} r \\ -y \end{pmatrix} - \frac{4}{d^2}(x^T y - tr)\begin{pmatrix} t \\ -x \end{pmatrix},$$

whence

$$q = -\frac{2}{d}r - \frac{4}{d^2}t(x^T y - tr), \quad z = \frac{2}{d}y + \frac{4}{d^2}x(x^T y - tr). \tag{2.17}$$

In order to verify that $\beta \in \mathbf{E}_+$, we should check that $q \geq 0$ and $q^2 \geq z^T z$. First of all, let us show that q is nonnegative. Let $\| x \|_2 = a$. From (2.15), (2.17) it follows that

$$q = -\frac{2}{d}r - \frac{4}{d^2}t(x^T y - tr) \geq -\frac{2}{d}r - \frac{4}{d^2}t(ar - tr) = \frac{2r}{(t+a)^2} \geq 0,$$

where the first $\geq$ is given by the Cauchy inequality, and the concluding inequality follows from $r \geq 0$ due to $\alpha \in \mathbf{E}_+$. Now let us demonstrate that

$q^2 \geq \| z \|_2^2$:

$$\begin{array}{rcl}
z^T z & = & \left[\frac{2}{d}y + \frac{4}{d^2}x(x^T y - tr)\right]^T \left[\frac{2}{d}y + \frac{4}{d^2}x(x^T y - tr)\right] \\
& = & \frac{4}{d^2}y^T y + \frac{16}{d^3}x^T y(x^T y - tr) + \frac{16}{d^4}x^T x(x^T y - tr)^2 \\
& \leq & \frac{4}{d^2}r^2 + \frac{16}{d^3}tr(x^T y - tr) + \frac{16}{d^4}t^2(x^T y - tr)^2 \\
& & \text{[by Cauchy's inequality]} \\
& = & \left[-\frac{2}{d}r - \frac{4}{d^2}t(x^T y - tr)\right]^2 = q^2.
\end{array}$$

Thus $\beta \in \mathbf{E}_+$, so that the multiplication by $E''(s)$ maps $\mathbf{E}_+$ into itself.

Now let us find the inverse to the matrix $E''(s)$. By the well-known formula,

$$\left[I + fg^T\right]^{-1} = I - \frac{fg^T}{1 + g^T f},$$

whence

$$\begin{array}{rcl}
[E''(s)]^{-1} & = & \left[\frac{2}{d}J + \frac{4}{d^2}Jss^T J\right]^{-1} = \left[\frac{2}{d}J(I + \frac{2}{d}ss^T J)\right]^{-1} \\
& = & \frac{d}{2}\left[I + \frac{2}{d}ss^T J\right]^{-1} J = \frac{d}{2}\left[I + \frac{2}{d}ss^T J\right]J \\
& = & \frac{d}{2}J\left[I + \frac{2}{d}(\underbrace{-Js}_{\eta})(-s)^T\right] = \frac{d}{2}J\left[I + \frac{2}{d}\eta(J\eta)^T\right] \\
& = & \frac{d}{2}J\left[I + \frac{2}{d}\eta\eta^T J\right] = \frac{d^2}{4}E''(\eta) = \frac{d^2}{4}E''(-Js).
\end{array}$$

We see that the $[E''(s)]^{-1}$ is proportional, with positive coefficient, to $E''(\eta)$, where $\eta = -Js$ belongs to $\mathbf{E}_+$ along with s, so that the multiplication by $[E''(s)]^{-1}$, same as the multiplication by $E''(s)$, maps the cone $\mathbf{E}_+$ into itself. Thus, the multiplication by $E''(s)$ induces a one-to-one mapping of $\mathbf{E}_+$ onto itself, as claimed.

Case 2: $\mathcal{E} = \mathcal{S}^k$. Let $s \in \operatorname{int} \mathbf{E}_+$, and

$$E(x) = -\ln \operatorname{Det}(x)$$

is k-self-concordant barrier for the cone $\mathbf{E}_+$ of positive semidefinite symmetric $k \times k$ matrices, and $F(x) = \operatorname{Det}(x)$.

Let us compute the derivatives of $E(x)$:

$$
\begin{aligned}
DE(s)[h] &= -D\ln(F(s))\Big[DF(s)[h]\Big] = -F^{-1}(s)\Big[DF(s)[h]\Big] \\
&= -\mathrm{Det}^{-1}(s)\lim_{t\to+0} t^{-1}\Big[\mathrm{Det}(s+t\,h)-\mathrm{Det}(s)\Big] \\
&= -\mathrm{Det}^{-1}(s)\lim_{t\to+0} t^{-1}\Big[\mathrm{Det}(s(I+t\,s^{-1}h))-\mathrm{Det}(s)\Big] \\
&= -\mathrm{Det}^{-1}(s)\lim_{t\to+0} t^{-1}\Big[\mathrm{Det}(s)(\mathrm{Det}(I+t\,s^{-1}h)-1)\Big] \\
&= -\lim_{t\to+0} t^{-1}\Big[\mathrm{Det}(I+t\,s^{-1}h)-1\Big] = -\mathrm{Tr}(s^{-1}h) \\
&= -\sum_{i,j}[s^{-1}]_{ji}h_{ij}
\end{aligned}
$$

where the concluding equality

$$
\lim_{t\to+0} t^{-1}\Big[\mathrm{Det}(I+tA)-1\Big] = \mathrm{Tr}(A) \equiv \sum_i A_{i\,i} \tag{2.18}
$$

is immediately given by recalling what is the determinant of $I+tA$: this is a polynomial of t which is the sum of products, taken along all diagonals of a $k\times k$ matrix and assigned certain signs, of the entries of $I+tA$. At every one of these diagonals, except for the main one, there are at least two cells with the entries proportional to t, so that the corresponding products do not contribute to the constant and the linear in t terms in $\mathrm{Det}(I+tA)$ and thus do not affect the limit in (2.18). The only product which does contribute to the linear and the constant terms in $\mathrm{Det}(I+tA)$ is the product $(I+tA_{11})(I+tA_{22})\dots(I+tA_{kk})$ coming from the main diagonal; it is clear that in this product the constant term is 1,and the linear in t term is $t(A_{11}+\ldots+A_{kk})$, and (2.18) follows.
As we have seen,

$$
DE(s)[h] = -\mathrm{Tr}(s^{-1}h). \tag{2.19}
$$

To differentiate the right hand side (2.19) in s, let us first find the derivative of the mapping $G(x) = x^{-1}$. We have

$$\begin{array}{rcl} DG(x)[\Delta x] & = & \lim\limits_{t\to+0} t^{-1}\Big[(x+t\Delta x)^{-1} - x^{-1}\Big] \\ & = & \lim\limits_{t\to+0} t^{-1}\Big[(x(I+t\,x^{-1}\Delta x))^{-1} - x^{-1}\Big] \\ & = & \lim\limits_{t\to+0} t^{-1}\Big[(I+t\,x^{-1}\Delta x)^{-1}x^{-1} - x^{-1}\Big] \\ & = & \Big[\lim\limits_{t\to+0} t^{-1}[(I+t\,x^{-1}\Delta x)^{-1} - I]\Big]x^{-1} \\ & = & \Big[\lim\limits_{t\to+0} t^{-1}[I-(I+t\,x^{-1}\Delta x)](I+t\,x^{-1}\Delta x)^{-1}\Big]x^{-1} \\ & = & \Big[\lim\limits_{t\to+0}[-x^{-1}\Delta x(I+t\,x^{-1}\Delta x)^{-1}]\Big]x^{-1} = -x^{-1}\Delta x x^{-1} \end{array}$$

and we arrive at the important by its own right relation

$$D(x^{-1})[\Delta x] = -x^{-1}\Delta x x^{-1},$$

which is the "matrix extension" of the standard relation $(x^{-1})' = -x^{-2}$, $x \in \mathbf{R}$.
Now we are ready to compute the second derivative of the log-det barrier:

$$\begin{array}{rcl} D^2E(s)[h,h] & = & -\lim\limits_{t\to+0} t^{-1}\Big[\mathrm{Tr}((s+t\,h)^{-1}h) - \mathrm{Tr}(s^{-1}h)\Big] \\ & = & -\lim\limits_{t\to+0} \mathrm{Tr}\Big(t^{-1}\Big[(s+t\,h)^{-1}h - s^{-1}h\Big]\Big) \\ & = & -\lim\limits_{t\to+0} \mathrm{Tr}\Big(t^{-1}\Big[(s+t\,h)^{-1} - s^{-1}\Big]h\Big) \\ & = & -\mathrm{Tr}(-s^{-1}hs^{-1}h) \end{array}$$

and we arrive at the formula

$$D^2E(s)[h,h] = \mathrm{Tr}(s^{-1}hs^{-1}h).$$

Since $\mathbf{E}$ is the space equipped with the standard Frobenius inner product $\langle x,y\rangle_{\mathbf{E}} = \mathrm{Tr}(xy)$, we have

$$\langle E''(s)\,h, h\rangle = \mathrm{Tr}(s^{-1}hs^{-1}h),$$

whence

$$E''(s)\,h = s^{-1}hs^{-1}. \tag{2.20}$$

Taking into account that $s = s^T \succ 0$, we conclude that both the mappings $h \mapsto E''(s)h = s^{-1}hs^{-1}$ and $h \mapsto [E''(s)]^{-1}h = shs$ map the cone $\mathbf{E}_+$ onto itself.

Case 3, 4: $\mathcal{E} = \mathcal{H}^k/\mathcal{Q}^k$**.** These cases are completely similar to Case 2, since here, as it is immediately seen,

$$E''(s)h = s^{-1}hs^{-1},$$

(cf. (2.20)), and it remains to apply Lemma 2.1. ■

Proof of Proposition 2.3. We again consider 4 cases.

Case 1: $\mathcal{E} = \mathcal{L}^k$**.** Let $s = \begin{pmatrix} \sigma \\ x \end{pmatrix} \in \mathbf{E}_+$, $d = \begin{pmatrix} \tau \\ y \end{pmatrix} \in \mathbf{E}_+$, with $\sigma, \tau \in \mathbf{R}$ and $x, y \in \mathbf{R}^k$. Then after a direct computation

$$E''(s) = \frac{2}{(\sigma^2 - x^T x)^2} \begin{pmatrix} \sigma^2 + x^T x & -2\sigma x^T \\ -2\sigma x & 2xx^T + (\sigma^2 - x^T x)I_k \end{pmatrix},$$

and $\mathbf{e} = \begin{pmatrix} \sqrt{2} \\ 0_k \end{pmatrix} \in \mathbf{E}_+$. Thus,

$$E''(s)\mathbf{e} = \frac{2\sqrt{2}}{(\sigma^2 - x^T x)^2} \begin{pmatrix} \sigma^2 + x^T x \\ -2\sigma x \end{pmatrix}. \tag{2.21}$$

Writing down $y = \eta f$ with $\eta = \sqrt{y^T y}$ and $\|f\|_2 = 1$, the right hand side in (2.21) equals d if and only if $x = -\xi f$, where $\sigma, \xi, \sigma \geq |\xi|$, solve the system of equations

$$\begin{aligned} \frac{\sigma^2 + \xi^2}{(\sigma^2 - \xi^2)^2} &= \frac{\tau}{2\sqrt{2}}, \\ \frac{2\sigma\xi}{(\sigma^2 - \xi^2)^2} &= -\frac{\eta}{2\sqrt{2}}. \end{aligned}$$

Adding and subtracting these equations, we see that the system is equivalent to

$$\begin{aligned} \frac{1}{(\sigma + \xi)^2} &= \frac{\tau - \eta}{2\sqrt{2}}, \\ \frac{1}{(\sigma - \xi)^2} &= \frac{\tau + \eta}{2\sqrt{2}}. \end{aligned}$$

Since $d \in \operatorname{int} \mathbf{E}_+$, we have $\tau > \eta$. Thus, the right side quantities in the latter system are positive, whence a solution to it does exist and is unique, provided that $\sigma >|\ \xi\ |$. This unique solution is given by

$$\begin{aligned} \sigma + \xi &= \frac{2^{3/4}}{\sqrt{\tau - \eta}}, \\ \sigma - \xi &= \frac{2^{3/4}}{\sqrt{\tau + \eta}}. \end{aligned} \qquad (2.22)$$

(A) is proved. To prove (B), note that from the explicit formula for $E''(s)$ it follows that

$$\mathrm{Det}(E''(s)) = \left(\frac{2}{\sigma^2 - x^T x}\right)^{k+1},$$

whence

$$\begin{aligned} \ln \mathrm{Det}(E''(s)) &= \tfrac{k+1}{2}[\ln 2 - \ln(\tau^2 - y^T y)] \\ &= \tfrac{k+1}{2}[E(d) - E(\mathbf{e})] = \tfrac{k+1}{2}[E(d) - E(\mathbf{e})], \end{aligned}$$

as required in (B).

Case 2: $\mathcal{E} = \mathcal{S}^k$. Let $s, d \in \mathrm{int}\,\mathbf{E}_+$, so that s, d are positive semidefinite real $k \times k$ matrices, and $\mathbf{e}$ is the unit real $k \times k$ matrix. We have

$$E''(s)\mathbf{e} = s^{-2},$$

so that (A) is trivially true, and

$$s = d^{-1/2}.$$

Let f_i, $i = 1, ..., k$, be an orthonormal eigenbasis of the symmetric matrix $d \succ 0$, and $\lambda_1, ..., \lambda_k > 0$ be the corresponding eigenvalues. Consider the system of $K = \frac{k(k+1)}{2}$ symmetric matrices F_{ij}, $1 \le i \le j \le k$, given by

$$F_{ij} = \begin{cases} f_i f_i^T, & i = j \\ \frac{1}{\sqrt{2}}[f_i f_j^T + f_j f_i^T], & i < j \end{cases}.$$

It is immediately seen that $\{F_{ij}\}$ is an orthonormal basis in $\mathbf{E}$ and that this is an eigenbasis of the mapping $h \mapsto E''(s)h$:

$$E''(s)F_{ij} = d^{1/2} F_{ij} d^{1/2} = \sqrt{\lambda_i \lambda_j} F_{ij}.$$

It follows that

$$\mathrm{Det} E''(s) = \prod_{1 \le i \le j \le k} (\lambda_i \lambda_j)^{1/2} = \left(\prod_{i=1}^{k} \lambda_i\right)^{\frac{k+1}{2}},$$

whence

$$\begin{aligned} \ln \mathrm{Det}(E''(s)) &= \tfrac{k+1}{2} \ln \mathrm{Det}(d) \\ &= \tfrac{k+1}{2} E(d) = \tfrac{k+1}{2}[E(d) - E(\mathbf{e})], \end{aligned}$$

as required in (B).

Case 3,4: $\mathcal{E} = \mathcal{H}^k/\mathcal{Q}^k$. These cases are completely similar to Case 2. ■

2.4.2 Products of simple equipped spaces

Assume we are given m simple equipped spaces

$$\mathcal{E}^\ell = (\mathbf{E}^\ell, \mathbf{E}^\ell_+, E^\ell, \mathbf{g}^\ell), \ \ell = 1, \ldots, m$$

from the above four series (to avoid confusion, here we denote the "centers" of the cones $\mathbf{E}^\ell_+$ by $\mathbf{g}^\ell$. Then we can form the *product* of these spaces

$$\mathbf{E} = \prod_{\ell=1}^m \mathbf{E}^\ell$$

which, by construction, is the equipped space as follows. Let

$$\begin{array}{rcl} \Theta & = & \min\limits_{1\le\ell\le m} \Theta[\mathcal{E}^\ell], \\ \chi_\ell & = & \frac{\Theta[\mathcal{E}^\ell]}{\Theta} \ge 1, \ \ell = 1, \ldots., m. \end{array} \tag{2.23}$$

We set

$$\mathbf{E} = (\mathbf{E}, \mathbf{E}_+, E(\cdot), \mathbf{e}),$$

where

1. The Euclidean space $\mathbf{E}$ is the direct product of the Euclidean spaces $\mathbf{E}^\ell$, $\ell = 1, ..., m$;

2. The cone $\mathbf{E}_+$ is the direct product of the cones $\mathbf{E}^\ell_+$, $\ell = 1, ..., m$;

3. The logarithmically homogeneous self-concordant barrier E for $\mathbf{E}_+$ is

$$E(x^1, ..., x^m) = \sum_{\ell=1}^m \chi_\ell E^\ell(x^\ell), \tag{2.24}$$

   where x^ℓ is the $\mathbf{E}^\ell$-component of $x \in \mathbf{E}$. Note that E indeed is a self-concordant logarithmically homogeneous barrier for $\mathbf{E}_+$ and the parameter of self-concordance of $E(\cdot)$ is

$$\nu(E) = \sum_{\ell=1}^m \chi_\ell \nu(E^\ell)$$

   (see item SC.I.1) in Section 2.2.3);

4. The center of $\mathbf{E}_+$ is the point

$$\mathbf{e} = (\sqrt{\chi_1}\mathbf{g}^1, ..., \sqrt{\chi_m}\mathbf{g}^m). \tag{2.25}$$

It is immediately seen that $\mathbf{e}$ indeed is the center of $\mathbf{E}_+$. Indeed,

$$\begin{array}{rcl} E''(\mathbf{e}) & = & \mathrm{Diag}\left\{\{\nabla^2(\chi_\ell E^\ell)(\sqrt{\chi_\ell}\mathbf{g}^\ell)\}_{\ell=1}^m\right\} \\ & = & \mathrm{Diag}\left\{\{\chi_\ell^{-1}\nabla^2(\chi_\ell E^\ell)(\mathbf{g}^\ell)\}_{\ell=1}^m\right\} \quad \text{[see (2.7.}b\text{)]} \\ & = & \mathrm{Diag}\left\{\{I_{\dim E^\ell}\}_{\ell=1}^m\right\} \end{array}$$

is the unit matrix.

We are about to demonstrate that a product of simple equipped spaces satisfies straightforward extensions of Propositions 2.2, 2.3:

Proposition 2.4 *Let* $\mathbf{E} = (\mathbf{E}, \mathbf{E}_+, E(\cdot), \mathbf{e})$ *be product of simple equipped spaces* $\mathcal{E}^\ell = (\mathbf{E}^\ell, \mathbf{E}^\ell_+, E^\ell, \mathbf{g}^\ell)$, $\ell = 1, \dots, m$. *Then*

(i) *For every* $s \in \mathrm{int}\,\mathbf{E}_+$, *the mapping* $h \mapsto E''(s)h$ *is a one-to-one mapping of the cone* $\mathbf{E}_+$ *onto itself.*

(ii) *For every* $d = (d^1, ..., d^m) \in \mathrm{int}\,\mathbf{E}_+$, *there exists a unique* $s \in \mathrm{int}\,\mathbf{E}_+$ *such that*

$$d = E''(s)\mathbf{e}, \tag{2.26}$$

and for this mapping one has

$$\ln \mathrm{Det} E''(s) = \Theta(E(d) - E(\mathbf{e})),$$

with Θ *defined in* (2.23).

Proof. For $s = (s^1, ..., s^m) \in \mathrm{int}\,\mathbf{E}_+$ we have

$$E''(s) = \mathrm{Diag}\left\{\{\chi_\ell \nabla^2 E^\ell(s^\ell)\}_{\ell=1}^m\right\}; \tag{2.27}$$

since $\chi_\ell > 0$ and multiplication by $\nabla^2 E^\ell(s^\ell)$ is a one-to-one mapping of $\mathbf{E}^\ell_+$ onto itself by Proposition 2.2, (i) follows.

(ii): Taking into account (2.24), (2.25), we have for $s = (s^1, ..., s^m) \in \mathrm{int}\,\mathbf{E}_+$:

$$E''(s)\mathbf{e} = \{\chi_\ell^{3/2}\nabla^2 E^\ell(s^\ell)\mathbf{g}^\ell\}_{\ell=1}^m, \tag{2.28}$$

so that (2.26) is equivalent to the relations

$$(\nabla^2 E^\ell)(s^\ell)\mathbf{g}^\ell = \chi_\ell^{-3/2} d^\ell,\ \ell = 1, ..., m.$$

Applying Proposition 2.3, we conclude that the latter relations are satisfied by a unique collection $s^\ell \in \operatorname{int} \mathbf{E}^\ell_+$, $\ell = 1, ..., m$. Further, by Proposition 2.3 we have

$$\begin{array}{rcl}
\ln \mathrm{Det}(\nabla^2 E^\ell(s^\ell)) & = & \Theta[\mathcal{E}^\ell]\left[E^\ell(\chi_\ell^{-3/2} d^\ell) - E^\ell(\mathbf{g}^\ell)\right] \\
& = & \frac{\Theta[\mathcal{E}^\ell]}{\chi_\ell}\left[(\chi_\ell E^\ell)(\chi_\ell^{-3/2} d^\ell) - (\chi_\ell E^\ell)(\chi_\ell^{-1/2}\mathbf{e}^\ell)\right] \\
& = & \Theta\Big[(\chi_\ell E^\ell)(d^\ell) + \chi_\ell \nu(E^\ell)\ln(\chi_\ell^{3/2}) - (\chi_\ell E^\ell)(\mathbf{e}^\ell) \\
& & -\chi_\ell \nu(E^\ell)\ln(\chi_\ell^{1/2})\Big] \\
& & \qquad [\text{since } E^\ell(ty) = E^\ell(y) - \nu(E^\ell)\ln t] \\
& = & \Theta\left[(\chi_\ell E^\ell)(d^\ell) - (\chi_\ell E^\ell)(\mathbf{e}^\ell)\right] + \gamma_\ell.
\end{array}$$

Taking into account (2.27), we arrive at

$$\begin{array}{rcl}
\ln \mathrm{Det} E''(s) & = & \sum\limits_{\ell=1}^{m}\left[\dim(\mathbf{E}^\ell)\ln\chi_\ell + \ln \mathrm{Det}(\nabla^2 E^\ell(s^\ell))\right] \\
& = & \sum\limits_{\ell=1}^{m}\left[\dim(\mathbf{E}^\ell)\ln\chi_\ell + \Theta\left[(\chi_\ell E^\ell)(d^\ell) - (\chi_\ell E^\ell)(\mathbf{e}^\ell)\right] + \gamma_\ell\right] \\
& = & \Theta\left[E(d) - E(\mathbf{e})\right] + \gamma
\end{array} \tag{2.29}$$

with γ independent of d. In fact $\gamma = 0$, since $E''(\mathbf{e}) = I$, i.e., $s = \mathbf{e}$ when $d = \mathbf{e}$, so that (2.29) is valid when $d = s = \mathbf{e}$, whence

$$0 = \ln \mathrm{Det} I = \ln \mathrm{Det} E''(\mathbf{e}) = \Theta\left[E(\mathbf{e}) - E(\mathbf{e})\right] + \gamma = \gamma.$$

Since $\gamma = 0$, (2.29) is exactly the relation required in (ii). ∎

2.5 Main result

Let $\mathcal{E}^\ell = (\mathbf{E}^\ell, \mathbf{E}^\ell_+, E^\ell, \mathbf{e}_\ell)$, $\ell = 1, \ldots, m$, be simple equipped spaces, and let $\mathcal{E} = (\mathbf{E}, \mathbf{E}_+, E, \mathbf{e})$ be a product of these spaces, see Section 2.4.2. To save notation, we denote the inner products in $\mathbf{E}^\ell$, $\mathbf{E}$ by $\langle\cdot,\cdot\rangle_\ell$, $\langle\cdot,\cdot\rangle$, respectively, so that

$$\langle x, y\rangle = \sum_{\ell=1}^{m}\langle x^\ell, y^\ell\rangle_\ell.$$

Let us set

$$\mathbf{F} = \Big\{x \in \mathbf{E} \mid \langle \mathbf{e}, x\rangle = 0\Big\},$$

$$\vartheta = \sqrt{\langle \mathbf{e}, \mathbf{e}\rangle},$$

$$\mathbf{j} = \vartheta^{-1}\mathbf{e},$$

$$\mathbf{D} = \Big\{ x \in \mathbf{F} \mid x + \mathbf{j} \in \mathbf{E}_+ \Big\}.$$

Let, further, $g \in \mathbf{F}$ be a unit vector, $\alpha \geq 0$ be a real, and let

$$\mathbf{D}^+[g] = \Big\{ x \in \mathbf{D} \mid \langle g, x \rangle \leq -\alpha \Big\}.$$

Our main result is as follows:

Theorem 2.1 *There exists (and can be explicitly written down) an affine mapping*

$$x \mapsto A[x] : \ \mathbf{F} \to \mathbf{F}$$

such that

$$\mathbf{D}^+[g] \subset A[\mathbf{D}] \tag{2.30}$$

and

$$\mathrm{Vol}(A[\mathbf{D}]) \leq \exp\{-\Xi\}\,\mathrm{Vol}(\mathbf{D}), \tag{2.31}$$

where Vol *stands for the* $\dim(\mathbf{F})$*-dimensional volume and*

$$\Xi = \left[\frac{1}{\nu(E)\sqrt{\Theta}} - \ln\left(1 + \frac{1}{\nu(E)\sqrt{\Theta}}\right) \right], \quad \nu(E) = \sum_\ell \chi_\ell \nu(E^\ell). \tag{2.32}$$

Proof. $\mathbf{1}^0$. Let $\widehat{x} \in \mathbf{D}$ be such that

$$\langle g, \widehat{x} \rangle \leq -\alpha, \tag{2.33}$$

and let $\lambda \geq 0$ be a real such that

$$d_\lambda \equiv \mathbf{e} + \vartheta \lambda g \in \mathrm{int}\, \mathbf{E}_+. \tag{2.34}$$

Note that from (2.34) and $\widehat{x} \in \mathbf{D}$ it follows that

$$1 - \lambda\alpha > 0;$$

indeed, we have $\widehat{x} + \mathbf{j} \in \mathbf{E}_+ \backslash \{0\}$ and $\langle \widehat{x}, \mathbf{e} \rangle = 0$, $\langle g, \mathbf{j} \rangle = 0$, $\langle \mathbf{j}, \mathbf{e} \rangle = \vartheta$, whence by (2.34) one has

$$0 < \langle \widehat{x} + \mathbf{j}, \mathbf{e} + \vartheta\lambda g \rangle = \vartheta\lambda \langle \widehat{x}, g \rangle + \vartheta = \vartheta(1 - \lambda\alpha).$$

Finally, let P be the orthoprojector of $\mathbf{E}$ onto $\mathbf{F}$:

$$P(x) = x - \langle x, \mathbf{j} \rangle \mathbf{j},$$

and let L be the affine operator on $\mathbf{F}$ given by

$$L[x] = x + \lambda\langle g, x\rangle\widehat{x} + \lambda\alpha\widehat{x}, \quad x \in \mathbf{F}.$$

Let us set

$$\begin{aligned} f &= \mathbf{j} + \lambda g, \\ Y^+ &= \Big\{y \in \mathbf{E}_+ \mid \langle y - (1-\lambda\alpha)\mathbf{j}, f\rangle = 0\Big\}, \\ Y &= P(Y^+). \end{aligned}$$

2^0. By Proposition 2.4 and (2.35), there exists $s_\lambda \in \operatorname{int}\mathbf{E}_+$ such that for

$$B_\lambda = E''(s_\lambda) \tag{2.35}$$

it holds

$$B_\lambda \mathbf{e} = d_\lambda \equiv \mathbf{e} + \vartheta\lambda g, \tag{2.36}$$

the mapping $x \mapsto B_\lambda x$ is nonsingular and maps $\mathbf{E}_+$ onto itself, and

$$\ln\operatorname{Det}(B_\lambda) = \Theta[E(d_\lambda) - E(\mathbf{e})] \tag{2.37}$$

Lemma 2.2 *The mapping $C_\lambda = B_\lambda^{-1}$ maps the set $\widetilde{\mathbf{D}} \equiv \mathbf{D} + \mathbf{j}$ onto the intersection $Y^\sharp$ of the cone $\mathbf{E}_+$ with the affine hyperplane*

$$H = \Big\{y \in \mathbf{E} \mid \langle y - \mathbf{j}, \mathbf{j} + \lambda g\rangle = 0\Big\}.$$

Proof. Indeed,

$$\begin{aligned} \Big\{y \in \mathbf{E} \mid \langle y - \mathbf{j}, \mathbf{j} + \lambda g\rangle = 0\Big\} &= \Big\{y \in \mathbf{E} \mid \langle y - \mathbf{j}, \mathbf{e} + \vartheta\lambda g\rangle = 0\Big\} \\ &= \Big\{y \in \mathbf{E} \mid \langle y - \mathbf{j}, B_\lambda \mathbf{e}\rangle = 0\Big\} \\ &= \Big\{y \in \mathbf{E} \mid \langle B_\lambda y, \mathbf{e}\rangle = \langle \mathbf{j}, B_\lambda \mathbf{e}\rangle\Big\} \\ &= \Big\{y \in \mathbf{E} \mid \langle B_\lambda y, \mathbf{e}\rangle = \langle \mathbf{j}, \mathbf{e} + \vartheta\lambda g\rangle\Big\} \\ &= \Big\{y \in \mathbf{E} \mid \langle B_\lambda y, \mathbf{e}\rangle = \vartheta\Big\} \\ &\qquad \Big[\text{recall that } \langle \mathbf{e}, g\rangle = 0,\ \langle \mathbf{j}, \mathbf{e}\rangle = \vartheta\Big] \\ &= \Big\{y \in \mathbf{E} \mid \langle B_\lambda y, \mathbf{j}\rangle = 1\Big\} \\ &= \Big\{y \in \mathbf{E} \mid \langle B_\lambda y - \mathbf{j}, \mathbf{j}\rangle = 0\Big\}, \end{aligned}$$

i.e.,

$$B_\lambda(H) = \mathbf{j} + \mathbf{F} \Leftrightarrow H = B_\lambda^{-1}[\mathbf{j} + \mathbf{F}],$$

and since B_λ is a one-to-one transformation of $\mathbf{E}$ which maps $\mathbf{E}_+$ onto itself, $B_\lambda y \in \mathbf{E}_+$ if and only if $y \in \mathbf{E}_+$. Thus, $B_\lambda \underbrace{[H \cap \mathbf{E}_+]}_{Y^\sharp} = \underbrace{(\mathbf{F} + \mathbf{j}) \cap \mathbf{E}_+}_{\widetilde{\mathbf{D}}}$. ∎

3^0. Now consider the following five affine mappings:
(a) $R: \ \mathbf{F} \to \mathbf{E}: \ R(x) = x + \mathbf{j}$;
(b) $C_\lambda: \ \mathbf{E} \to \mathbf{E}$, which was already defined;
(c) $S: \ \mathbf{E} \to \mathbf{E}: \ S(y) = (1 - \lambda\alpha)y$;
(d) $P: \ \mathbf{E} \to \mathbf{F}$, which was already defined as the orthoprojector of $\mathbf{E}$ onto $\mathbf{F}$;
(e) $L: \ \mathbf{F} \to \mathbf{F}: \ L[z] = z + \lambda\langle g, z\rangle\widehat{x} + \lambda\alpha\widehat{x}$
along with their composition

$$A = L \circ P \circ S \circ C_\lambda \circ R: \ \mathbf{F} \to \mathbf{F}.$$

Our first claim is that

Lemma 2.3 *One has $\mathbf{D}^+[g] \subset A[\mathbf{D}]$.*

Proof. Indeed,

- By Lemma 2.2, the set $C_\lambda(R(B))$ is exactly $Y^\sharp$;
- By evident reasons, $S(Y^\sharp)$ is exactly Y^+;
- Applying Proposition 2.1, we conclude that the set $L[P(Y^+)]$ contains $\mathbf{D}^+[g]$. ∎

4^0. Now let us compute the absolute value κ of the determinant of the (homogeneous part of the) mapping A. In other words, we should understand by which factor κ the $\dim(\mathbf{F})$-dimensional volume of a set $Q \subset \mathbf{F}$ is multiplied under the action of A. Since R is just a shift and thus does not vary $\dim(\mathbf{F})$-dimensional volumes, κ is the product of the following four quantities:

- the factor μ by which the linear transformation C_λ of $\mathbf{E}$ multiplies $\dim(\mathbf{F})$-dimensional volumes of sets contained in $\mathbf{F}$;
- the factor $(1-\lambda\alpha)^{\dim(\mathbf{F})}$ by which the dilatation S multiplies $\dim(\mathbf{F})$-dimensional volumes of sets contained in the hyperplane $\mathbf{G} = C_\lambda\mathbf{F}$;

- the factor ν by which the orthoprojection P multiplies the $\dim(\mathbf{F})$ - dimensional volumes of the sets belonging to the hyperplane $\mathbf{G}$.
  Note that by Lemma 2.2 $\mathbf{G}$ is the hyperplane in $\mathbf{E}$ orthogonal to the vector $\mathbf{j} + \lambda g$, so that ν is the cosine of the angle between the normal $\mathbf{j} + \lambda g$ to $\mathbf{G}$ and the normal $\mathbf{j}$ to $\mathbf{F}$:

$$\nu = \frac{\langle \mathbf{j}, \mathbf{j} + \lambda g \rangle}{\sqrt{\langle \mathbf{j}, \mathbf{j} \rangle}\sqrt{\langle \mathbf{j} + \lambda g, \mathbf{j} + \lambda g \rangle}} = \frac{1}{\sqrt{1+\lambda^2}};$$

- the factor θ by which the linear transformation L of $\mathbf{F}$ multiplies the $\dim(\mathbf{F})$-dimensional volumes of sets in $\mathbf{F}$.
  From the expression for L it is immediately seen that

$$\theta = 1 + \lambda \langle g, \widehat{x} \rangle.$$

Thus,

$$\kappa = \frac{1 + \lambda \langle g, \widehat{x} \rangle}{\sqrt{1+\lambda^2}} (1 - \lambda\alpha)^{\dim(\mathbf{F})} \mu. \tag{2.38}$$

5^0. It remains to find μ. Note that $C_\lambda = B_\lambda^{-1}$, so that the mapping C_λ multiplies $\dim(\mathbf{E})$-dimensional volumes by the factor

$$\mu^+ = \mathrm{Det}(C_\lambda) = \frac{1}{\mathrm{Det}(B_\lambda)}. \tag{2.39}$$

On the other hand, μ^+ is exactly the product of μ and the distance σ from the image of $\mathbf{j}$ under the mapping C_λ to the image of $\mathbf{F}$ under the same mapping. Thus,

$$\sigma^2 = \min_{h \in \mathbf{F}} \left\{ \langle C_\lambda[\mathbf{j} - h], C_\lambda[\mathbf{j} - h] \rangle \mid \langle \mathbf{j}, h \rangle = 0 \right\}.$$

The solution h^* to the right hand side optimization problem clearly is $h^* = \mathbf{j} - \phi^* C_\lambda^{-2}\mathbf{j}$ where $\phi^* = \frac{1}{\langle C_\lambda^{-2}\mathbf{j}, \mathbf{j} \rangle}$. A direct computation implies that

$$\begin{aligned} \sigma^2 &= \frac{\langle C_\lambda^{-1}\mathbf{j}, C_\lambda^{-1}\mathbf{j} \rangle}{\langle C_\lambda^{-2}\mathbf{j}, \mathbf{j} \rangle^2} = \frac{1}{\langle C_\lambda^{-2}\mathbf{j}, \mathbf{j} \rangle} \\ &= \frac{1}{\langle B_\lambda \mathbf{j}, B_\lambda \mathbf{j} \rangle} = \frac{\vartheta^2}{\langle B_\lambda \mathbf{e}, B_\lambda \mathbf{e} \rangle} \\ &= \frac{\vartheta^2}{\langle \mathbf{e} + \vartheta\lambda g, \mathbf{e} + \vartheta\lambda g \rangle} = \frac{1}{1+\lambda^2}. \end{aligned}$$

We now get

$$\mu = \frac{\mu^+}{\sigma} = \mu^+\sqrt{1+\lambda^2},$$

whence

$$\begin{array}{rcl}\kappa & = & \mu\nu\theta(1-\lambda\alpha)^{\dim(\mathbf{F})} = \mu^+\sqrt{1+\lambda^2}\frac{1}{\sqrt{1+\lambda^2}}(1+\lambda\langle g,\widehat{x}\rangle)(1-\lambda\alpha)^{\dim(\mathbf{F})} \\ & = & (1+\lambda\langle g,\widehat{x}\rangle)\mu^+(1-\lambda\alpha)^{\dim(\mathbf{F})} = (1+\lambda\langle g,\widehat{x}\rangle)\frac{1}{\mathrm{Det}(B_\lambda)}(1-\lambda\alpha)^{\dim(\mathbf{F})}.\end{array}$$

Applying Proposition 2.4. (ii), we finally get

$$\begin{array}{rcl}\ln(\kappa) & = & \ln(1+\lambda\langle g,\widehat{x}\rangle) + \dim(\mathbf{F})\ln(1-\lambda\alpha) + \Theta\left[E(\mathbf{e}+\vartheta\lambda g) - E(\mathbf{e})\right] \\ & \leq & \lambda[\langle g,\widehat{x}\rangle - \alpha\dim(\mathbf{F})] + \Theta\left[E(\mathbf{e}+\vartheta\lambda g) - E(\mathbf{e})\right] \\ & \leq & \lambda\langle g,\widehat{x}\rangle + \Theta\left[E(\mathbf{e}+\vartheta\lambda g) - E(\mathbf{e})\right] \\ & \equiv & \Psi(\lambda)\end{array} \tag{2.40}$$

6^0. To prove Theorem 2.1, it remains to optimize $\ln(\kappa)$ in the "free parameters" of our construction, i.e., $\widehat{x}$ and λ. In fact, of course, we intend to optimize $\Psi(\cdot)$.

6^0.1) We start with maximizing in $\widehat{x}$. The only restriction on this vector is that $\widehat{x} \in \mathbf{D}$ and $\langle g,\widehat{x}\rangle \leq -\alpha$; and of course we are interested to make $\langle g,\widehat{x}\rangle$ as negative as possible. Thus, it makes sense to define $\widehat{x}$ as an optimal solution to the problem

$$\min_{\widehat{x}}\left\{\langle g,\widehat{x}\rangle : \widehat{x}\in\mathbf{D} = \left\{x \mid \langle\mathbf{e},x\rangle = 0, \mathbf{e}+\vartheta x\in\mathbf{E}_+\right\}\right\}. \tag{2.41}$$

It is easily seen that the problem can be solved explicitly. What is important for our subsequent analysis, is the optimal value of this problem, let us denote it by $(-\vartheta^{-1}\sigma)$. Then σ is the optimal value in the optimization problem

$$\max_y\left\{\langle -g,y\rangle : \langle\mathbf{e},y\rangle = 0, \mathbf{e}+y\in\mathbf{E}_+\right\}. \tag{$*$}$$

We claim that

$$1 \leq \sigma \leq \nu(E) + 2\sqrt{\nu(E)}. \tag{2.42}$$

Indeed, E is self-concordant barrier for the cone $\mathbf{E}_+$ with the self-concordance parameter $\nu(E)$, and $E''(\mathbf{e})$ is the unit matrix. Thus, the unit ball U, centered at $\mathbf{e}$, is contained in $\mathbf{E}_+$ (see item SC.II.1) in Section 2.2.3). Since g is a unit vector orthogonal to $\mathbf{e}$, this observation implies the lower bound on σ in (2.42). Since $E''(\mathbf{e})$ is the unit matrix and E is a logarithmically homogeneous self-concordant barrier, $E'(\mathbf{e})$ is equal to $\mathbf{e}$ (see (2.7.c)); it follows that the $(\nu(E)+2\sqrt{\nu(E)})$ times larger than U concentric ball contains the intersection of $\mathbf{E}_+$ and the plane $\mathbf{e}+\mathbf{F}$ (see (2.10.b)), i.e., it contains the feasible set of $(*)$, which implies the upper bound on σ in (2.42).

6^0.2) From now on, we specify $\widehat{x}$ as explained in 6^0.1). With this choice of $\widehat{x}$, the function $\Psi(\lambda)$ from (2.42) becomes

$$\Psi(\lambda) = -\vartheta^{-1}\sigma\lambda + \Theta[E(\mathbf{e}+\vartheta\lambda g) - E(\mathbf{e})].$$

Now, the restrictions imposed on λ by our construction are $\lambda \geq 0$ and (2.34); the latter restriction is equivalent to $\lambda \in \mathrm{Dom}\Psi$. The function Ψ is self-concordant by item SC.I.1), Section 2.2.3, and

$$\begin{array}{rcl}
\Psi(0) & = & 0; \\
\Psi'(0) & = & -\vartheta^{-1}\sigma + \Theta\vartheta\langle E'(\mathbf{e}), g\rangle \\
 & = & -\vartheta^{-1}\sigma \\
 & & [\text{since } E''(\mathbf{e}) = I, \text{ whence } E'(\mathbf{e}) = \mathbf{e} \text{ by (2.7.}c\text{), and } \langle g, \mathbf{e}\rangle = 0] \\
\Psi''(0) & = & \Theta\vartheta^2\langle E''(\mathbf{e})g, g\rangle \\
 & = & \Theta\vartheta^2 \\
 & & [\text{recall that } E''(\mathbf{e}) \text{ is the unit matrix and } g \text{ is a unit vector}]
\end{array}$$

Consequently, the Newton decrement (Section 2.2.2)

$$\lambda(\Psi, 0) = \frac{|\Psi'(0)|}{\sqrt{\Psi''(0)}}$$

is

$$\psi = \frac{\sigma}{\vartheta^2\sqrt{\Theta}};$$

since E is logarithmically homogeneous barrier with the parameter $\nu(E) = \sum\limits_{\ell} \nu(E^\ell)$ and $E''(\mathbf{e})$ is the unit matrix, we have by (2.7.d)

$$\vartheta^2 = \langle \mathbf{e}, \mathbf{e}\rangle = \langle E''(\mathbf{e})\mathbf{e}, \mathbf{e}\rangle = \nu(\mathcal{E}),$$

so that

$$\psi = \frac{\sigma}{\nu(\mathcal{E})\sqrt{\Theta}}.$$

In view Applying item SC.II.2) of Section 2.2.3 to $x = 0$ and $f = \Psi$, we see that with

$$\lambda_* = -\frac{1}{1+\psi}\frac{\Psi'(0)}{\Psi''(0)} > 0$$

one has (cf. (2.9))

$$\begin{array}{rcl}\Psi(\lambda_*) & \leq & \Psi(0) - [\psi - \ln(1+\psi)] \\ & = & -[\psi - \ln(1+\psi)] \\ & \leq & -\left[\frac{1}{\nu(E)\sqrt{\Theta}} - \ln\left(1 + \frac{1}{\nu(E)\sqrt{\Theta}}\right)\right]\end{array}$$

(we have used the fact that $\sigma \geq 1$ by (2.42)). Thus, the choice $\lambda = \lambda_*$ results in

$$\ln \kappa \leq -\left[\frac{1}{\nu(E)\sqrt{\Theta}} - \ln\left(1 + \frac{1}{\nu(E)\sqrt{\Theta}}\right)\right],$$

and (2.32) follows. ∎

2.6 Discussion

Theorem 2.1 combines with the constructions and discussions of Section 1.3.5 to yield the following conclusion:

Theorem 2.2 *Every self-scaled cone with no Octonian irreducible components produces a perfect solid* (Section 1.3.5)*. Specifically, if equipped space* $\mathcal{E} = v$ *is the direct product of simple equipped spaces* $\mathcal{E}^\ell = (\mathbf{E}^\ell, \mathbf{E}^\ell_+, E^\ell, \mathbf{g}^\ell)$, $\ell = 1, ..., m$ (Section 2.4.2), *then the set*

$$\mathbf{B} = \{x : \langle \mathbf{e}, x\rangle_{\mathcal{E}} = 0, x + \mathbf{e} \in \mathbf{E}_+\}$$

is a perfect solid in the space

$$\mathbf{F} = \{x \in \mathbf{E} : \langle \mathbf{e}, x\rangle_{\mathcal{E}}\},$$

and this solid is contained in the Euclidean ball of the radius $\gamma[\mathbf{B}] = \nu(E) + 2\sqrt{\nu(E)}$. *The corresponding index* ω *is given by*

$$\ln(\omega) \leq -\frac{1}{n}\left[\frac{1}{\nu(E)\sqrt{\Theta}} - \ln\left(1 + \frac{1}{\nu(E)\sqrt{\Theta}}\right)\right], \qquad (2.43)$$

where

- $n = \sum\limits_\ell n_\ell - 1$ *is the dimension of* $\mathbf{F}$ *(*n_ℓ *are the dimensions of* $\mathbf{E}_\ell$*);*
- $\nu(E) = \sum\limits_\ell \xi_\ell \nu(E^\ell)$, *where* $\xi_\ell = \frac{\Theta[\mathcal{E}^\ell]}{\Theta}$ *and* $\Theta = \min\limits_\ell \Theta[\mathcal{E}^\ell]$.

The characteristics n_ℓ, $\nu(E^\ell)$, $\Theta[calE^\ell]$ of underlying simple equipped spaces are as follows:

Type of $\mathcal{E}^\ell$	n_ℓ	$\nu(E^\ell)$	$\Theta[\mathcal{E}^\ell]$
$\mathcal{L}^k$	k	2	$\frac{k+1}{2}$
$\mathcal{S}^k$	$\frac{k(k+1)}{2}$	k	$\frac{k+1}{2}$
$\mathcal{H}^k$	k^2	k	k
$\mathcal{Q}^k$	$2k^2-k$	k	$2k-1$

(2.44)

Proof. Note that the set $\mathbf{B}$ is nothing but the dilatation, by factor $\vartheta = \|\mathbf{e}\|_{\mathbf{E}}$, of the set $\mathbf{D}$ considered in Theorem 2.1, so that for every nonzero vector $g \in \mathbf{F}$ one has (cf. (2.30))

$$\{x \in \mathbf{B} : \langle g, x\rangle_{\mathcal{E}} \leq 0\} \subset A[\mathbf{B}],$$

where $A[z] = Az + b$ is the invertible affine transformation of $\mathbf{F}$ given by the proof of Theorem 2.1 in the case of $\alpha = 0$, and $\omega \equiv (\mathrm{Det}(A))^{1/n}$ satisfies (2.43) by (2.31) – (2.32). Besides this, $\mathbf{E}_+$ contains the centered at $\mathbf{e}$ unit Euclidean ball (by SC.II.1), since $E''(\mathbf{E})$ is the unit matrix), and the intersection of $\mathbf{E}_+$ with the set $\mathbf{e} + \mathbf{F}$ is contained in the concentric ball of the radius $\nu(E) + 2\sqrt{\nu(E)}$ (we have seen this when proving (2.42)). It follows that $\mathbf{B}$ is "in-between" the unit ball and the ball of the radius $\nu(E) + 2\sqrt{\nu(E)}$. Thus, $\mathbf{B}$ indeed is a perfect solid with $\gamma[\mathbf{B}]$ and the index ω indicated in Theorem. ∎

Via the construction presented in Section 1.3.5, perfect solids described in Theorem 2.2 give rise to stationary cutting plane methods, and all these methods are polynomial time ones, since the indices ω of the underlying perfect solids satisfy the relation

$$1 - \omega \geq \frac{O(1)}{n^\beta} \tag{2.45}$$

(see (2.43) and the discussion in Section 1.3.5). We are about to investigate the complexity characteristics of these polynomial time SCP methods.

2.6.1 Arithmetic cost of an operation

A straightforward analysis, based on the explicit description of the affine transformation $A[\cdot]$ as given in Theorem 2.1, demonstrates that *for all SCP methods given by our construction, the arithmetic cost of an iteration* (i.e., the number of arithmetic operations per iteration, excluding the computations carried out by the Separation and the First Order oracles) *is proportional to n^2*, where n is the dimension of the problem. Thus, as far as complexity of a step is concerned, *all our polynomial SCP methods are "the same"*.

2.6.2 Iteration count

As it was explained in Introduction (see Section 1.3.5), the iteration complexity of finding ϵ-solution for a stationary Cutting Plane method $\mathcal{B}$ associated with a perfect n-dimensional solid $\mathbf{B}$ with index ω satisfying (2.45) does not exceed

$$N(\epsilon) \leq O(1)n^{\beta} \ln\left(\frac{\gamma[\mathbf{B}]R}{\epsilon r}\right), \tag{2.46}$$

where $R \geq r > 0$ are such that the domain X of the problem contains a ball of radius r and is contained in the centered at the origin ball of radius R. The "most important" contribution of the underlying solid to this complexity bound is the parameter β – the smaller is β, the better. By (2.43), we have

$$1 - \omega \geq O(1)\frac{1}{n\nu^2(E)\Theta} \tag{2.47}$$

(indeed, since $\nu(E) \geq 1$ and $\Theta \geq 1$, one has $\left[\frac{1}{\nu(E)\sqrt{\Theta}} - \ln\left(1 + \frac{1}{\nu(E)\sqrt{\Theta}}\right)\right] = O(1)\frac{1}{\nu^2(E)\Theta}$).

In the case when $\mathcal{E} = \mathcal{L}^{n+1}$ (so that the associated method $\mathcal{B}$ is the Ellipsoid method), we have $\nu(E) = 2$, $\Theta = O(n)$, and (2.47) implies that $1 - \omega \geq O(n^{-2})$, so that here $\beta = 2$. The question is, whether an appropriate implementation of our construction yields a smaller β. The answer, unfortunately, is negative:

Proposition 2.5 *In the family of stationary Cutting Plane methods associated with self-scaled cones, the best guaranteed complexity bound, as given by (2.46) – (2.47), is the bound with $\beta = 2$ associated with the Ellipsoid method (and thus – with the perfect solid given by the Lorentz cone).*

Proof. The fraction in the right hand side of (2.47) is

$$\frac{1}{n\nu^2(E)\Theta} = \frac{\Theta}{n\left(\sum\limits_{\ell}\Theta[\mathcal{E}^\ell]\nu(E^\ell)\right)^2}$$

(see Theorem 2.2). From (2.44) it is seen that $\Theta[\mathcal{E}^\ell]\nu(E^\ell) \geq O(1)n_\ell$, and we arrive at

$$\frac{1}{n\nu^2(E)\Theta} \leq O(1)\frac{\Theta}{n\left(\sum\limits_{\ell} n_\ell\right)^2} = \frac{\Theta}{n^3}.$$

From the same (2.44), $\Theta = \min\limits_{\ell}\Theta[\mathcal{E}^\ell] \leq O(1)n$, and we end up with

$$\frac{1}{n\nu^2(E)\Theta} \leq O(1)\frac{1}{n^2}.$$

It follows that the right hand side in (2.45) is always $\leq O(1)n^{-2}$, i.e., we never can get $\beta < 2$. ∎

To get more insight, here are the values of β for several possible choices of the self-scaled cones:

$\mathcal{E}$	β
$\mathcal{L}^n$ [Ellipsoid]	2
$\mathcal{S}^n$	2.5
$\mathcal{H}^n$	2.5
$\mathcal{Q}^n$	2.5
$\mathcal{S}^1 \times ... \times \mathcal{S}^1$ [Circumscribed Simplex]	3

Note that although the Ellipsoid method is the best in terms of the guaranteed worst-case complexity, it does not necessarily mean that all other methods are of no practical interest. Indeed, our complexity analysis is worst-case-oriented, which, in our context, essentially means that we deal with the worst – the smallest – value of σ as given by (2.42), i.e., the value 1. Now, in the case of $\mathcal{E} = \mathcal{L}^n$ (i.e., in the case of the ellipsoid method) this "worst case" is in fact the *only* case (since here $\mathbf{B}$ is just the unit Euclidean ball). In contrast to this, for other self-scaled cones (e.g., for the semidefinite cone or the nonnegative orthant), the value of σ depends on the direction of g and is equal to 1 in very special (and "hardly probable") cases only; in these cases, a "typical" value of σ – the one corresponding to randomly oriented g uniformly distributed on the unit sphere – is much larger than 1 (it is as large as $\sqrt{k}$ in the cases of $\mathbf{E}_+ = \mathbf{R}^k_+$ and of $\mathbf{E}_+ = \mathbf{S}^k_+$). It follows that *with some luck, the actual behaviour of a "non-Lorentz" SCP algorithm may be much better than it is demonstrated by our worst-case complexity analysis.* This hope is seemingly supported by the numerical experience with the Circumscribed Ellipsoid algorithm. Consequently, there is a hope that the new SCP methods we have developed in practice may outperform the two known SCP algorithms. Experimental verification of this hope is beyond the scope of this Thesis.

Bibliography

[1] Alizadeh, F., Schmieta, S., *Symmetric cones, potential reduction methods, and word-by-ord extensions* – in: R. Saigal, H. Wolkowitcz, L. Vandenberghe, Eds. Handbook on Semidefinite Programming, Kluwer Academic Publishers (2000) 195–233.

[2] Ben-Tal, A., Nemirovski, A. *Optimization I: Convex Analysis, Nonlinear Programming Theory, Nonlinear Programming Algorithms* – Lecture Notes, Faculty of IE & M, Technion.

[3] Nemirovski, A., *Optimization II: Numerical Methods for Nonlinear Continuous Optimization* – Lecture Notes, Faculty of IE & M, Technion.

[4] Ben-Tal, A., Nemirovski, A., *Lectures on Modern Convex Optimization*, SIAM, Philadelphia (2001).

[5] Bulatov, V., Shepot'ko, I. *Method of centers of orthogonal simplexes for solving convex programming problems* (in Russian), In: Methods of Optimization and their Application, Novosibirsk, Nauka (1982), 79-84.

[6] Güler, O., Tuncel, L., *Characterization of the barrier parameter of homogeneous convex cones*, Mathematical Programming **81** (1998), 55-76.

[7] Karmarkar, N. *A new polynomial-time algorithm for linear programming*, Combinatorica **4** (1984), 373-395.

[8] Khachiyan, L., Tarasov, S., Erlikh, I., "Inscribed Ellipsoid method" – *Doklady Akademii Nauk SSSR* (*Soviet MAth. Doklady*), 1987.

[9] Kiwiel, K., *Proximal level bundle method for convex nondifferentable optimization, saddle point problems and variational inequalities*, Mathematical Programming Series B **69** (1995), 89-109.

[10] Nemirovski,A., Yudin, D., *Information-based complexity and efficient methods of convex optimization* (in Russian) – Ekonomika i Matematicheskie Metody, **12:2** (1976) (the journal is translated into English as *Matekon*).

[11] Nemirovski,A., Yudin, D. *Problem complexity and method efficiency in Optimization* – J. Wiley & Sons (1983).

[12] Nemirovski, A., *Efficient Methods in Convex Programming* – Lecture Notes, Faculty of IE & M, Technion.

[13] Nesterov, Yu., Nemirovski, A., *Interior-Point Polynomial Algorithms in Convex Programming* – SIAM, Philadelphia (1994).

[14] Shor, N. *Cutting plane method with space dilation for the solution of convex programming problems* (in Russian), Kibernetika (1977), 94-95.

[15] Vinberg, E., *The theory of homogeneous cones*, Trans. Moscow Math. **12** (1965), 340-403.

[16] Yamnitsky, B., Levin, L. *The 1965 linear programming algorithm runs in polynomial time*, In: Proceedings of the 3rd Annual IEEE Symposium on Foundation of Computer Science 1982 – IEEE Computer Society Press, Los Angeles (1982), 327-328.

[17] Schramm, H., Zowe, J., *A version of bundle idea for minimizing a non-smooth function: conceptual idea, convergence analysis, numerical results*, SIAM Journal on Optimization **2** (1992), 121-152.