On Self-Concordant Convex-Concave Functions

Arkadi Nemirovski*

April 19, 1999

Abstract

In this paper, we introduce the notion of a *self-concordant convex-concave* function, establish basic properties of these functions and develop a path-following interior point method for approximating saddle points of "good enough" convex-concave functions – those which admit natural self-concordant convex-concave regularizations. The approach is illustrated by its applications to developing an exterior penalty polynomial time method for Semidefinite Programming and to the problem of inscribing the largest volume ellipsoid into a given polytope.

1 Introduction

Self-concordance-based approach to interior point methods for variational inequalities: state of the art. The self-concordance-based theory of interior point (IP) polynomial methods in Convex Programming [5] is commonly recognized as the standard approach to the design of theoretically efficient IP methods for convex optimization programs. A natural question is whether this approach can be extended to other problems with convex structure, like saddle point problems for convex-concave functions, and, more generally, variational inequalities with monotone operators. The goal of this paper is to make a step in this direction.

The indicated question was already considered in [5], Chapter 7. To explain what and why we intend to do, let us start with outlining the relevant results from [5].

We want to solve a variational inequality

$$(\mathcal{V})\quad \text{find } z^* \in \operatorname{cl} Z : (z - z^*)^T A(z) \geq 0 \quad \forall z \in Z,$$

where Z is an open and, say, bounded convex set in $\mathbf{R}^N$ and $A(\cdot) : Z \to \mathbf{R}^N$ is a monotone operator. The domain Z of the inequality is equipped with a ϑ-self-concordant barrier (s.-c.b.) B ([5], Chapter 2), and two classes of monotone operators on Z are considered:

*Faculty of Industrial Engineering and Management at Technion – the Israel Institute of Technology, Technion City, Haifa 32000, Israel. e-mail: `nemirovs@ie.technion.ac.il`. The research was partially funded by the Fund for the Promotion of Research at Technion, the G.I.F. contract No. I-0455-214.06/95 and the Israel Ministry of Science grant # 9636-1-96

(a) "self-concordant" (s.-c.);
(b) "compatible with B".
Both classes are defined in terms of certain differential inequalities imposed on A.

The variational inequalities which actually can be solved by the technique in question are those with operators of the type (b). In order to solve such an inequality, we associate with the operator of interest A the penalized family

$$\{A_t(z) = tA(z) + B'(z)\}_{t>0}.$$

It turns out that every A_t is s.-c. and possesses a unique zero $z^*(t)$ on Z, and that the path $z^*(t)$ converges to the solution set of $(\mathcal{V})$. In order to solve $(\mathcal{V})$, we trace the indicated path:

> Given current iterate $(t^{i-1}, z^{i-1}) = (t, z)$ with z close, in certain precise sense, to $z^*(t)$, we update (t, z) into a new pair $(t^i, z^i) = (t^+, z^+)$, also close to the path, according to
>
> $$(a) \quad t^+ = (1 + 0.01\vartheta^{-1/2})t, \qquad (b) \quad z^+ = z - [(A_{t^+})'(z)]^{-1}A_{t^+}(z); \tag{1}$$
>
> here $(A_t)'(z)$ is the Jacobian of the mapping $A_t(\cdot)$ at z.

It turns out that the outlined process converges to the set of solutions of $(\mathcal{V})$ linearly (w.r.t. certain meaningful accuracy measure) with the convergence ratio $(1 - O(1)\vartheta^{-1/2})$.

The main ingredient of the complexity analysis of the outlined construction is an affine-invariant *local* theory of the Newton method for approximating zero of a s.-c. monotone operator. The method is responsible for the "centering step" $(1.b)$, and the penalty updating policy $(1.a)$ is given by the desire to keep the previous iterate z in the domain of local quadratic convergence of the Newton method as applied to A_{t^+}.

Note that in the potential case, i.e., when $A(z)$ is the gradient field of a "good" (say, linear or quadratic) convex function f, the outlined scheme becomes the standard short-step IP method for minimizing f over Z, and the indicated complexity result yields the standard, the best known so far theoretical complexity bound for the corresponding convex optimization program, which is a good news. The potential case, however, offers us much more, since here we possess not only a local, but also a *global* affine-invariant convergence theory for the Newton method. As a result, in the potential case we may use instead of the penalty rate $(1.a)$ other penalty updating policies as well, at the cost of replacing a single pure Newton centering step $(1.b)$ with several *damped* Newton steps, until the required closeness to the new target point $z^*(t^+)$ is achieved. The number of required damped Newton steps can be bounded from above, in a universal data-independent fashion, via the residual

$$V(t^+, z) = [t^+ f(z) + B(z)] - \min_{z'}[t^+ f(z') + B(z')]$$

(t, z) being the previous iterate. Thus, in the potential case the Newton complexity (# of Newton steps in z per one updating of the penalty parameter t) of the path-following method in question can be controlled not only for the worst-case oriented penalty updating policy $(1.a)$, but for any other policy capable to control the residuals $V(t^+, z)$. This observation provides us with possibility to trace the path with "mediate" steps (arbitrary absolute constant instead of 0.01 in $(1.a)$, see [5], Section 3.2.6) or even with "long" on-line adjusted steps (see [6, 7]), which is very attractive for practical computations.

In contrast to these favourable features of the potential case, in the case of a non-potential monotone operator A compatible with a s.-c.b. for $\mathrm{cl}\, Z$ all which has been offered to the moment by the self-concordance-based approach is the short-step policy ($1.a$). With current understanding of the subject we are unable to say what happens with the method when the constant 0.01 in ($1.a$) is replaced by, say, 1. This is a definite bad news about the known so far extensions of the self-concordance-based theory to the case of variational inequalities: in order to get a complexity bound, no matter $O(\sqrt{\vartheta})$-one or worse, we should restrict ourselves with the worst-case-oriented and therefore very conservative short-step policy ($1.a$). This drawback of the theory finally comes from the fact that we have no global theory of convergence of the Newton method as applied to a (non-potential) s.-c. monotone operator.

The goal of this paper is to investigate the situation which is in-between the potential case and the general monotone case, namely, the one of monotone mappings associated with convex-concave saddle point problems. We intend to demonstrate that there exists a quite reasonable extension of the notion of a self-concordant convex function – the basic ingredient of the self-concordance-based theory of IP methods – to the case of convex-concave functions. The arising entities – *self-concordant convex-concave* (s.-c.c.-c.) functions – possess a rich theory very similar to the theory developed in [5] for convex s.-c. functions. In particular, we develop a *global* theory of convergence of (a kind of) the Newton method as applied to the problem of approximating a saddle point of a s.-c.c.-c. function. Finally, this global theory allows us to build mediate-step path-following methods for approximating saddle point of a convex-concave function which is "compatible" with self-concordant barrier for its domain.

The contents of the paper is as follows. In Section 2 we introduce our central notion – the one of a self-concordant convex-concave function – and investigate the basic properties of these functions. Section 3 is devoted to the duality theory for s.-c.c.-c. functions; the central result here is that the class in question is closed w.r.t. the Legendre transformation. The latter fact underlies the global theory of (a kind of) the Newton method for approximating saddle point of a s.-c.c.-c. function (Section 5). Our main result here is that the number of Newton steps needed to pass from a point $(\widehat{x}, \widehat{y})$ from the domain of a s.-c.c.-c. function $f(x, y)$ to an ϵ-saddle point (x, y) – a point where

$$\mu(x, y) \equiv \sup_{y'} f(x, y') - \inf_{x'} f(x', y) \leq \epsilon$$

– is bounded from above by a quantity of the type

$$\Theta(\mu(\widehat{x}, \widehat{y})) + O(1) \ln \ln(\epsilon^{-1} + 3),$$

where $\Theta(\cdot)$ is a universal function and $O(1)$ is an absolute constant. This result is very similar to the basic result on the Newton method for minimizing a convex s.-c. function f: the number of Newton steps needed to pass from a point $\widehat{x}$ of the domain of the function to an ϵ-minimizer of f – to a point x where $v(x) \equiv f(x) - \min f \leq \epsilon$ – is bounded from above by $O(1)[v(\widehat{x}) + \ln \ln(\epsilon^{-1} + 3)]$.

Equipped with a global theory of our "working horse" – the Newton method for approximating saddle points of s.-c.c.-c. functions – we get the possibility to develop a general theory of path-following method for approximating saddle points of convex-concave functions "compatible" with self-concordant barriers for their domains. This development is carried out in Section 6. We conclude this Section by constructing a polynomial time *exterior* penalty method for semidefinite programming problems (the construction works for linear and conic quadratic programming as well).

Finally, in Section 7 we apply our general constructions and results to the well-known problem of finding the maximum volume ellipsoid contained in a given polytope. This problem is of interest for Control Theory (see [1]) and especially for Nonsmooth Convex Optimization, where it is the basic auxiliary problem arising in the *Inscribed Ellipsoid method* (Khachiyan, Tarasov, Erlikh [2]); the latter method, as applied to general convex problems, turns out to be optimal in the sense of Information-based Complexity Theory. The best known so far complexity estimates for finding an ϵ-optimal (with volume $\geq (1-\epsilon)$ times the maximum one) ellipsoid inscribed into a n-dimensional polytope, given by $m = O(n)$ linear inequalities, are $O(n^{4.5}\ln(n\epsilon^{-1}R))$ and $O(n^{3.5}\ln(n\epsilon^{-1}R)\ln(n\epsilon^{-1}\ln R))$ arithmetic operations (see [5], Chapter 6, and [3], respectively). Here R is an a priori known ratio of radii of two centered at the origin Euclidean balls, the smaller being contained in, and the larger containing the polytope. These complexity bounds are given by interior point methods as applied to the standard semidefinite reformulation of the problem. The latter reformulation is of a "large" – $O(n^2)$ – design dimension. We demonstrate that the problem admits saddle point reformulation with "small" – $O(n+m)$ – design dimension, and that the resulting saddle point problem can be straightforwardly solved by the path-following method from Section 6; the arithmetic complexity of finding ϵ-optimal ellipsoid in this manner turns out to be $O(n^{3.5}\ln(n\epsilon^{-1}R))$. In contrast to the complexity bounds from [5, 3], the indicated – better – bound arises naturally, without sophisticated ad hoc tricks heavily exploiting problem's structure.

The proofs of the major part of the results to be presented are rather technical. To improve the readability of the paper, all such proofs are placed in Appendices.

2 Self-concordant convex-concave functions

In this section we introduce the main concept to be studied – the one of a *self-concordant convex-concave* (s.-c.c.-c.) *function* – and establish several basic properties of these functions.

2.1 Preliminaries: self-concordant convex functions

The notion of a s.-c.c.-c. function is closely related to the one of a self-concordant (s.-c.) convex function, see [5]. For reader's convenience, we start with the definition of the latter notion.

Definition 2.1 *Let X be an open nonempty convex domain in $\mathbf{R}^n$. A function $f : X \to \mathbf{R}$ is called* self-concordant (s.-c.) *on X, if f is convex,* C^3 *smooth and*

(i) *f is a barrier for X: $f(x_i) \to \infty$ along every sequence of points $x_i \in X$ converging to a boundary point of X.*

(ii) *For every $x \in X$ and $h \in \mathbf{R}^n$ one has*

$$|D^3 f(x)[h,h,h]| \leq 2\left(D^2 f(x)[h,h]\right)^{3/2} \qquad (2)$$

(from now on, $D^k f(x)[h_1,...,h_k]$ denotes k-th differential of a smooth function f taken at a point x along directions $h_1,...,h_k$).

If, in addition to (i), (ii)*, for some $\vartheta \geq 1$ and all $x \in X$, $h \in \mathbf{R}^n$ one has*

$$|Df(x)[h]| \leq \vartheta^{1/2}\sqrt{D^2 f(x)[h,h]},$$

then f is called ϑ-self-concordant barrier (s.-c.b.) *for* $\mathrm{cl}\, X$.

A s.-c. function f is called nondegenerate, if $f''(x)$ is nonsingular at least at one point (and then – at every point, [5], Corollary 2.1.1) of the domain of the function.

For a nondegenerate s.-c. function f, the quantity

$$\lambda(f,x) = \sqrt{[f'(x)]^T[f''(x)]^{-1}f'(x)} \quad [x \in \mathrm{Dom} f] \tag{3}$$

is called the Newton decrement of f at x.

Note that what here is called self-concordance of a function in [5] is called strong self-concordance.

A summary of the basic properties of convex s.-c. functions is as follows (for proofs, see [5, 4]):

Proposition 2.1 *Let f be a convex s.-c. function. Then for every $x \in \mathrm{Dom} f$ one has:*

$$\begin{array}{c} \forall h: \quad f(x) + h^T f'(x) + \rho\left(-\sqrt{h^T f''(x)h}\right) \le f(x+h) \le f(x) + h^T f'(x) + \rho\left(\sqrt{h^T f''(x)h}\right), \\ \rho(s) = -s - \ln(1-s) \end{array} \tag{4}$$

(both f and ρ are $+\infty$ outside their domains); in particular, if $h^T f''(x)h < 1$, then $x+h \in \mathrm{Dom} f$.

Besides this,

$$h^T f''(x)h < 1 \Rightarrow \left\{ \begin{array}{l} f''(x+h) \succeq \left(1 - \sqrt{h^T f''(x)h}\right)^2 f''(x), \\ f''(x+h) \preceq \left(1 - \sqrt{h^T f''(x)h}\right)^{-2} f''(x) \end{array} \right. \tag{5}$$

(from now on, an inequality $A \succeq B$ with symmetric matrices A, B of the same size means that $A - B$ is positive semidefinite).

In the case of a nondegenerate f one has

$$\rho\left(-\lambda(f,x)\right) \le f(x) - \inf f \le \rho\left(\lambda(f,x)\right). \tag{6}$$

2.2 Self-concordant convex-concave functions: definition and local properties

We define the notion of a s.-c.c.-c. function as follows:

Definition 2.2 *Let X, Y be open convex domains in $\mathbf{R}^n$, $\mathbf{R}^m$, respectively, and let*

$$f(x,y) : X \times Y \to \mathbf{R}$$

be C^3 smooth function. We say that the function is self-concordant convex-concave on $X \times Y$, if f is convex in $x \in X$ for every $y \in Y$, concave in $y \in Y$ for every $x \in X$, and

(i) *For every $x \in X$, $[-f(x,\cdot)]$ is a barrier for Y, and for every $y \in Y$, $f(\cdot,y)$ is a barrier for X*

(ii) *For every $z = (x,y) \in X \times Y$ and every $dz = (dx,dy) \in \mathbf{R}^n \times \mathbf{R}^m$ one has*

$$|D^3 f(z)[dz,dz,dz]| \le 2[dz^T \mathrm{S}_f(z) dz]^{3/2}, \qquad \mathrm{S}_f(z) = \begin{pmatrix} f''_{xx}(z) & 0 \\ 0 & -f''_{yy}(z) \end{pmatrix}. \tag{7}$$

A s.-c.c.-c. function f is called nondegenerate, if $\mathrm{S}_f(z)$ is positive definite for some (and then, as we shall see, for all) $z \in Z$.

Remark 2.1 Note that if $f : X \times Y \to \mathbf{R}$ is a s.-c.c.-c. function, then the convex function $f(\cdot,y)$ is s.-c. on X for every $y \in Y$, and the convex function $-f(x,\cdot)$ is s.-c. on Y for every $x \in X$.

Our current goal is to establish several basic properties of s.-c.c.-c. functions.

Proposition 2.2 [Basic differential inequality and recessive subspaces] *Let $f(z)$ be a s.-c.c.-c. function on $Z = X \times Y \subset \mathbf{R}^n \times \mathbf{R}^m$. Then*

(i) *For every $z \in Z$ and every triple dz_1, dz_2, dz_3 of vectors from $\mathbf{R}^n \times \mathbf{R}^m$ one has*

$$|D^3 f(z)[dz_1, dz_2, dz_3]| \leq 2 \prod_{j=1}^{3} \sqrt{dz_j^T \mathrm{S}_f(z) dz_j} \tag{8}$$

(ii) *Let $E(z) = \{dz = (dx, dy) \mid dz^T \mathrm{S}_f(z) dz = 0\}$. Then*

(ii.1) *the subspace $E(z) \equiv E_f$ is independent of $z \in Z$ and is the direct sum of its projections $E_{f,x}$, $E_{f,y}$ on $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. In particular, if S_f is nondegenerate at some point of Z, then S_f is nondegenerate everywhere on Z.*

(ii.2) *$Z = Z + E_f$.*

(iii) [Sufficient condition for nondegeneracy] *If both X and Y do not contain straight lines, then f is nondegenerate.*

The next statement says, roughly speaking, that a s.-c.c.-c. function f is fairly well approximated by its second-order Taylor expansion at a point $z \in \mathrm{Dom} f$ in the respective *Dikin ellipsoid* $\{z' | (z' - z)^T \mathrm{S}_f(z)(z' - z) < 1\}$. This property (similar to the one of convex s.-c. functions) underlies all our further developments.

Proposition 2.3 [Local properties] *Let $f : Z = X \times Y \to \mathbf{R}$, $X \subset \mathbf{R}^n$, $Y \subset \mathbf{R}^m$, be a s.-c.c.-c. function. Then*

(i) *For every $z = (x, y) \in Z$, the set $W_x^f(z) = \{x' \mid (x' - x)^T f''_{xx}(z)(x' - x) < 1\}$ is contained in X, and the set $W_y^f(z) = \{y' \mid (y' - y)^T [-f''_{yy}(z)](y' - y) < 1\}$ is contained in Y.*

(ii) *Let $z \in Z$ and $h = (u, v) \in \mathbf{R}^n \times \mathbf{R}^m$. Then*

$$\begin{array}{lrcl}
r \equiv \sqrt{h^T \mathrm{S}_f(z) h} < 1 \Rightarrow & & & \\
(a) & z + h & \in & Z; \\
(b) & (1-r)^2 \mathrm{S}_f(z) & \preceq & \mathrm{S}_f(z+h) \preceq (1-r)^{-2} \mathrm{S}_f(z); \\
(c) & \forall (h_1, h_2 \in \mathbf{R}^n \times \mathbf{R}^m) : & & \\
 & |h_1^T [f''(x+z) - f''(x)] h_2| & \leq & \left[\frac{1}{(1-r)^2} - 1\right] \sqrt{h_1^T \mathrm{S}_f(z) h_1} \sqrt{h_2^T \mathrm{S}_f(z) h_2}; \\
(d) & \forall h' \in \mathbf{R}^n \times \mathbf{R}^m : & & \\
 & \left| (h')^T [f'(z+h) - f'(z) - f''(z) h] \right| & \leq & \frac{r^2}{1-r} \sqrt{(h')^T \mathrm{S}_f(z) h'}.
\end{array} \tag{9}$$

Relation to self-concordant monotone mappings. We conclude this section by demonstrating that in the case of monotone mappings coming from convex-concave functions the notion of self-concordance of the mapping, as defined in [5], Chapter 7, is equivalent to self-concordance, as defined here, of the underlying convex-concave function.

In [5], Chapter 7, a *strongly self-concordant monotone mapping* is defined as a single-valued C^2 monotone mapping $A(\cdot)$ defined on an open convex domain Z such that

1. For every $z \in Z$ and every triple of vectors h_1, h_2, h_3 one has

$$|h_2^T \nabla_z^2 (h_1^T A(z)) h_3| \leq 2 \prod_{i=1}^{3} \sqrt{h_i^T \widehat{A}(z) h_i}, \quad \widehat{A}(z) = \frac{1}{2} \left[A'(z) + [A'(z)]^T \right]; \tag{10}$$

2. Whenever a sequence $\{z_i \in Z\}$ converges to a boundary point of Z, the sequence of matrices $\{\widehat{A}(z_i)\}$ is unbounded.

Proposition 2.4 *Let $f(x, y) : Z = X \times Y \to \mathbf{R}$ be C^3 convex-concave function, X, Y being open convex sets in $\mathbf{R}^n$, $\mathbf{R}^m$, respectively. The function is s.-c.c.-c. if and only if the monotone mapping*

$$A(x, y) = \begin{pmatrix} f'_x(x, y) \\ -f'_y(x, y) \end{pmatrix} : Z \to \mathbf{R}^{n+m}$$

is strongly self-concordant.

2.3 Saddle points of self-concordant convex-concave functions: existence and uniqueness

Our ultimate goal is to approximate saddle points of s.-c.c.-c. functions, and to this end we should know when saddle points do exist. The simple necessary and sufficient condition to follow is completely similar to the fact that a convex s.-c. function attains its minimum if and only if it is below bounded:

Proposition 2.5 *Let $f(z)$ be a nondegenerate s.-c.c.-c. function on $Z = X \times Y \subset \mathbf{R}^n \times \mathbf{R}^m$. Then f possesses a saddle point on Z if and only if*

> (*) *$f(x_0, \cdot)$ is above bounded on Y for some $x_0 \in X$, and $f(\cdot, y_0)$ is below bounded on X for some $y_0 \in Y$.*

Whenever (*) is the case, the saddle point of f on Z is unique.

3 Duality for self-concordant convex-concave functions

We are about to study the notion heavily exploited in the sequel – the one of the *Legendre transformation* of a nondegenerate s.-c.c.-c. function. The construction goes back to Rockafellar [8] and defines the Legendre transformation of a convex-concave function $f(x, y)$ as

$$f_*(\xi, \eta) = \inf_y \sup_x [\xi^T x + \eta^T y - f(x, y)]$$

(cf. the definition of the Legendre transformation of a convex function f: $f_*(\xi) = \sup_x [\xi^T x - f(x)]$). Our local goal is to describe the domain of the Legendre transformation of a nondegenerate s.-c.c.-c. function and to demonstrate that this transformation also is s.-c.c.-c.

Definition 3.1 *Let $f(z)$ be a s.-c.c.-c. function on $Z = X \times Y \subset \mathbf{R}^n \times \mathbf{R}^m$. We say that a vector $\xi \in \mathbf{R}^n$ is x-appropriate for f, if the function $\xi^T x - f(x, y)$ is above bounded on X for some $y \in Y$. Similarly, we say that a vector $\eta \in \mathbf{R}^m$ is y-appropriate for f, if the function $\eta^T y - f(x, y)$ is below bounded on Y for some $x \in X$. We denote by $X^*(f)$ the set of those ξ which are x-appropriate for f, and denote by $Y^*(f)$ the set of those η which are y-appropriate for f.*

Proposition 3.1 *Let $f(z)$ be a nondegenerate s.-c.c.-c. function on $Z = X \times Y \subset \mathbf{R}^n \times \mathbf{R}^m$. Then*

(i) *The sets $X^*(f)$ and $Y^*(f)$ are open nonempty convex sets in $\mathbf{R}^n$, $\mathbf{R}^m$, respectively;*

(ii) *The set $Z^*(f) = X^*(f) \times Y^*(f)$ is exactly the set of those pairs $(\xi, \eta) \in \mathbf{R}^n \times \mathbf{R}^m$ for which the function*

$$f_{\xi,\eta}(x, y) = f(x, y) - \xi^T x - \eta^T y$$

*possesses a saddle point (*min *in x,* max *in y) on $X \times Y$;*

(iii) *The set $Z^*(f) = X^*(f) \times Y^*(f)$ is exactly the image of the set Z under the mapping*

$$z \mapsto f'(z)$$

and the mapping is a one-to-one twice continuously differentiable mapping of Z onto $Z^*(f)$ with twice continuously differentiable inverse;

(iv) *The function*

$$f_*(\xi, \eta) = \inf_{y\in Y} \sup_{x\in X} [\xi^T x + \eta^T y - f(x, y)]$$

is s.-c.c.-c. on $Z^*(f)$ and is equal to

$$\sup_{x\in X} \inf_{y\in Y} [\xi^T x + \eta^T y - f(x, y)],$$

and the mapping

$$\zeta \mapsto f'_*(\zeta)$$

is inverse to the mapping $z \mapsto f'(z)$.

Definition 3.2 *Let $f : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function. The function*

$$f_* : Z^*(f) = X^*(f) \times Y^*(f) \to \mathbf{R}$$

defined in Proposition 3.1 is called the Legendre transformation of f.

Theorem 3.1 *The Legendre transformation f_* of a nondegenerate s.-c.c.-c. function f is a nondegenerate s.-c.c.-c. function, and f is the Legendre transformation of f_*. Besides this, one has*

$$\{z \in Z, \zeta = f'(z)\} \Leftrightarrow \{\zeta \in Z^*(f), z = f'_*(\zeta)\} \Rightarrow \left\{ \begin{array}{lrcl} (a) & f''(z) & = & [f''_*(\zeta)]^{-1}; \\ (b) & [f''(z)]^{-1}\mathrm{S}_f(z)[f''(z)]^{-1} & = & \mathrm{S}_{f_*}(\zeta); \\ (c) & f(z) + f_*(\zeta) & = & \zeta^T z. \end{array} \right. \tag{11}$$

4 Operations preserving self-concordance of convex-concave functions

In order to apply the machinery we are developing, we need tools to recognize self-concordance of a convex-concave function. These tools are given by (A) a list of "raw materials" − simple s.-c.c.-c. functions, and (B) a list of "combination rules" preserving the property in question. The simplest versions of (A) and (B), following immediately from Definition 2.2, can be described as follows:

Proposition 4.1 (i) *Let $f(x,y)$ be a quadratic function of (x,y) which is convex in $x \in \mathbf{R}^n$ and concave in $y \in \mathbf{R}^m$. Then f is s.-c.c.-c. function on $\mathbf{R}^n \times \mathbf{R}^m$.*

(ii) *Let $\phi(x)$, $\psi(y)$ be s.-c. convex functions on $X \subset \mathbf{R}^n$, $Y \subset \mathbf{R}^m$, respectively. Then $f(x,y) = \phi(x) - \psi(y)$ is a s.-c.c.-c. function on $X \times Y$.*

(iii) *Let $X_i \subset \mathbf{R}^n$, $Y_i \subset \mathbf{R}^m$, let $\alpha_i \geq 1$, and let f_i be s.-c.c.-c. function on $Z_i = X_i \times Y_i$, $i = 1, ..., k$. If the set $Z = \bigcap\limits_{i=1}^{k} Z_i$ is nonempty, then the function $f(z) = \sum\limits_{i=1}^{k} \alpha_i f_i(z)$ is s.-c.c.-c. on Z.*

(iv) *Let $f(z)$ be s.-c.c.-c. function on $Z = X \times Y \subset \mathbf{R}^n \times \mathbf{R}^m$, and let $\Pi(u,v) = \begin{pmatrix} x = Pu + p \\ y = Qv + q \end{pmatrix}$ be affine mapping from $\mathbf{R}^\nu \times \mathbf{R}^\mu$ to $\mathbf{R}^n \times \mathbf{R}^m$ with the image intersecting Z. Then the function $\phi(u,v) = f(\Pi(u,v))$ is s.-c.c.-c. on $\Pi^{-1}(Z)$.*

More "advanced" combination rules (Propositions 4.3, 4.4 below) state, essentially, that minimization/maximization of a s.-c.c.-c. function with respect to (a part of) "convex", respectively, "concave", variables preserves the self-concordance. To arrive at the corresponding results, we start with important by its own right "dual representation" of the extremum values of a s.-c.c.-c. function.

Proposition 4.2 *Let $f : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let $f_* : Z^* = X^* \times Y^* \to \mathbf{R}$ be the Legendre transformation of f. Then*

(i) *Whenever $y \in Y$, the function $f(\cdot, y)$ is below bounded on X if and only if $0 \in X^*$ and the function $\eta^T y - f_*(0, \eta)$ is below bounded on Y^*. When it is the case, one has*

$$\inf_{x \in X} f(x,y) = \min_{x \in X} f(x,y) = \min_{\eta \in Y^*} [\eta^T y - f_*(0,\eta)]; \tag{12}$$

(ii) *Whenever $x \in X$, the function $f(x, \cdot)$ is above bounded on Y if and only if $0 \in Y^*$ and the function $\xi^T x - f_*(\xi, 0)$ is above bounded on X^*. When it is the case, one has*

$$\sup_{y \in Y} f(x,y) = \max_{y \in Y} f(x,y) = \max_{\xi \in X^*} [\xi^T x - f_*(\xi, 0)].$$

Proposition 4.3 *Let $f(x,y) : Z \equiv X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function. If the set $X^+ = \{x \in X \mid \sup\limits_{y \in Y} f(x,y) < \infty\}$ is nonempty, then it is an open convex set, and the function*

$$\bar{f}(x) = \sup_{y \in Y} f(x,y) : X^+ \to \mathbf{R}$$

is a s.-c. convex function on X^+.

Similarly, if the set $Y^+ = \{y \in Y \mid \inf\limits_{x \in X} f(x,y) > -\infty\}$ is nonempty, then it is an open convex set, and the negative of the function

$$\underline{f}(y) = \inf_{x \in X} f(x,y) : Y^+ \to \mathbf{R}$$

is a s.-c. convex function on Y^+.

In the case when one of the sets X, Y – say, Y – is bounded, the latter proposition can be strengthened as follows:

Proposition 4.4 *Let $f : Z \equiv X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let the set Y be bounded. Let also $y = (u, v)$ be a partitioning of the variables y, and let U be the projection of Y onto the u-space. Then the function*

$$\phi(x, u) = \max_{v:(u,v)\in Y} f(x, (u, v)) : X \times U \to \mathbf{R} \tag{13}$$

is nondegenerate s.-c.c.-c.

5 The Saddle Newton method

5.1 Outline of the method

The crucial role played by self-concordance in the theory of interior-point polynomial time methods for convex optimization comes from the fact that these functions are well-suited for the Newton minimization. Specifically, if f is a nondegenerate s.-c. convex function, then a (damped) Newton step

$$x \mapsto x^+ = x - \frac{1}{1 + \lambda(f, x)}[f''(x)]^{-1} f'(x) \tag{14}$$

(see (3)) maps Dom$f$ into itself and

(A) reduces f at least by $O(1)$, provided that $\lambda(f, x)$ is not small: $f(x^+) \le f(x) - \rho(-\lambda(f, x))$;

(B) always "nearly squares" the Newton decrement: $\lambda(f, x^+) \le 2\lambda^2(f, x)$.

Property (A) ensures global convergence of the Newton minimization method, provided that f is below bounded: by (A), the number of Newton steps before a point $\bar{x}$ with $\lambda(f, \bar{x}) \le 0.25$ is reached, does not exceed $O(1)(f(\widehat{x}) - \inf f)$, $\widehat{x}$ being the starting point. Property (B) ensures local quadratic convergence of the method in terms of the Newton decrement (and in fact – in terms of the residual $f(x) - \min f$, since this residual can be bounded from above in terms of the Newton decrement solely, provided that the latter is less than 1). As a result, for every $\epsilon < 0.5$, an ϵ-minimizer of f – a point x satisfying $f(x) - \min f \le \epsilon$ – can be found in no more than

$$O(1)\left([f(\widehat{x}) - \min f] + \ln\ln\frac{1}{\epsilon}\right) \tag{15}$$

steps of the damped Newton method (14).

When passing from approximating the minimizer of a nondegenerate s.-c. convex function to approximating the saddle point of a nondegenerate s.-c.c.-c. function f, a natural candidate to the role of the Newton iteration is something like

$$z \mapsto z^+ = z - \gamma(z)[f''(z)]^{-1} f'(z), \tag{16}$$

$\gamma(z)$ being a stepsize. For such a routine, there is no difficulty with extending (B) (see [5], Chapter 7) and thus – with establishing local quadratic convergence. There is, however, a severe difficulty with extending (A), and, consequently, with establishing global convergence of the method, since now it is unclear what could play the role of the Lyapunov function of the process, the role played in the minimization case by the objective itself. To overcome this difficulty, recall that process (16), started at a point $\widehat{z}$, is a discretization of the "continuous time" process

$$\frac{d}{ds} z = -[f''(z)]^{-1} f'(z), \quad z(0) = \widehat{z},$$

and that the trajectory of the latter process is, up to the re-parameterization $s \mapsto t = \exp\{-s\}$, the path given by

$$f'(z(t)) = tf'(\widehat{z}), \ \ 1 \geq t \geq 0, \tag{17}$$

so that $z(t)$ is the saddle point of a s.-c.c.-c. function $f_t(z) = f(z) - tz^T f'(\widehat{z})$. Now, path (17) can be traced not only by (16), but via the path-following scheme

$$(t^k, z^k) \mapsto \left(t^{k+1} < t^k, z^{k+1} = z^k - [f''_{t^{k+1}}(z^k)]^{-1} f'_{t^{k+1}}(z^k)\right), \quad (t^1, z^1) = (1, \widehat{z}). \tag{18}$$

An advantage of (18) as compared to (16) is that with a proper control of the rate at which the homotopy parameter t^k decreases with k, z^k all the time is in the domain of quadratic convergence, as given by a straightforward "saddle point" extension of (B), of the Newton method as applied to $f_{t^{k+1}}$. Thus, one can analyze process (18) on the basis of the results on *local* convergence of the Newton method for approximating saddle points. The resulting complexity bound for (18) (see Theorem 5.1 below) states that the number of steps of (18) required to reach an *ϵ-saddle point* – a point z satisfying

$$\mu(f, z) \equiv \sup_{y'} f(x, y') - \inf_{x'} f(x', y) \leq \epsilon \quad [z = (x, y)]$$

for $\epsilon < 0.5$ is bounded from above by

$$\Theta(\mu(z_{\text{ini}})) + O(1) \ln \ln \frac{1}{\epsilon}, \tag{19}$$

where $\Theta(\cdot)$ is an *universal* (i.e., problem-independent) continuous function on the nonnegative ray. Note that the residual $\mu(f, z)$ is a natural extension to the saddle point case of the usual residual $\phi(x) - \min \phi$ associated with the problem of minimizing a function ϕ (indeed, the latter problem can be though of as the problem of finding a saddle point of the function $f(x, y) \equiv \phi(x)$, and $\mu(f, (x, y)) = \phi(x) - \min \phi$). Thus, the complexity bound (19) is of the same spirit as the bound (15), up to the fact that in the minimization case the universal function $\Theta(\cdot)$ is just linear.

The goal of this section is to implement the outlined scheme and to establish (19).

5.2 Newton decrement and proximities

We start with introducing several quantities relevant to the construction outlined in the previous subsection.

Definition 5.1 *Let $f : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let $z = (x, y) \in Z$. We define*

- *the Newton direction of f at z as the vector $e(f, z) = [f''(z)]^{-1} f'(z)$;*
- *the Newton decrement of f at z as the quantity $\omega(f, z) = \sqrt{e^T(f, z) \mathrm{S}_f(z) e(f, z)}$;*
- *the weak proximity of z as the quantity $\mu(f, z) = \sup_{y' \in Y} f(x, y') - \inf_{x' \in X} f(x', y) \leq +\infty$;*
- *the strong proximity of z as the quantity $\nu(f, z) = \sqrt{[f'(z)]^T [\mathrm{S}_f(z)]^{-1} f'(z)}$.*

The following proposition establishes useful connections between the introduced entities.

Proposition 5.1 *Let $f : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let $f_* : Z^* = X^* \times Y^* \to \mathbf{R}$ be the Legendre transformation of f. Then*

(i) *One has*

$$\forall (z \in Z): \quad \omega(f,z) = \sqrt{[f'(z)]^T \mathrm{S}_{f_*}(f'(z)) f'(z)}. \tag{20}$$

(ii) *The set $K(f) = \{z \in Z \mid \mu(f,z) < \infty\}$ is open, and $\mu(f,z)$ is continuous on $K(f)$.*

(iii) *The following properties are equivalent to each other:*

(iii.1) *$K(f)$ is nonempty;*
(iii.2) *f possesses a saddle point on Z;*
(iii.3) *$(0,0) \in Z^*$;*
(iii.4) *there exists $z \in Z$ with $\omega(f,z) < 1$.*

(iv) *For all $z \in Z$ one has*

$$\begin{array}{llcll} (a) & & \omega(f,z) & \leq & \nu(f,z); \\ (b) & \nu(f,z) < 1 \Rightarrow & \mu(f,z) & \leq & \rho(\nu(f,z)) \left[\text{recall that } \rho(s) = \begin{cases} -s - \ln(1-s), & s < 1 \\ +\infty, & s \geq 1 \end{cases}\right]; \\ (c) & z \in K(f) \Rightarrow & \frac{\nu^2(f,z)}{2(1+\nu(f,z))} & \leq & \mu(f,z). \end{array} \tag{21}$$

(v) *Let $z \in Z$ be such that $\omega(f,z) < 1$. Then the Newton iterate of z – the point $z^+ = z - e(f,z)$ – belongs to Z, and*

$$\nu(f,z^+) \leq \frac{\omega^2(f,z)}{(1-\omega(f,z))^2}. \tag{22}$$

(vi) *Assume that $K(f) \neq \emptyset$, and let z^* be the saddle point of f (it exists by* (iii)*). For every $z \in K(f)$ one has*

$$\sqrt{[z-z^*]^T \mathrm{S}_f(z^*)[z-z^*]} \leq 2\left[\mu(f,z) + \sqrt{\mu(f,z)}\right]. \tag{23}$$

5.3 The Saddle Newton method

Let $f(x,y) : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let $\widehat{z} \in K(f)$. The *Saddle Newton method for approximating the saddle point of f with starting point $\widehat{z}$* is as follows.

Let

$$f_t(z) = f(z) - t z^T f'(\widehat{z}), \ 0 \leq t \leq 1,$$

and let $z^*(t)$ be the unique saddle point of the function f_t, $0 \leq t \leq 1$ [1)]. In the method, we trace the path $z^*(t)$ as $t \to +0$, i.e., generate a sequence of pairs $(t^i, z^i = (x^i, y^i))$ such that

$$\{t^i \geq 0\} \ \& \ \{z^i \in Z\} \ \& \ \{\nu_i \equiv \nu(f_{t^i}, z^i) \leq 0.1\} \tag{P_i}$$

We start the method with the pair $(t^1, z^1) = (1, \widehat{z})$ which clearly satisfies (P_1). At i-th step, given a pair (t^i, z^i) satisfying (P_i), we

---

[1)] The existence of $z^*(t)$ is given by Proposition 5.1.(iii): since $K(f)$ is nonempty, $(0,0) \in Z^*(f)$, and of course $f'(\widehat{z}) \in Z^*(f)$, so that $tf'(\widehat{z}) \in Z^*(f)$ for all $t \in [0,1]$. The uniqueness of $z^*(t)$ follows from the fact that f_t is nondegenerate.

1. *Choose the smallest $t = t^{i+1} \in [0, t^i]$ satisfying the predicate*

$$\omega(f_t, z^i) \leq 0.2$$

2. *Set*

$$z^{i+1} = z^i - [f''(z^i)]^{-1} f'_{t^{i+1}}(z^i). \tag{24}$$

We are about to establish one of our central results – the efficiency bound for the Saddle Newton method.

Theorem 5.1 *Let $f : X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let the starting point $\widehat{z}$ be such that $\mu \equiv \mu(f, \widehat{z}) < \infty$. As applied to $(f, \widehat{z})$, the Saddle Newton method possesses the following properties ($\widehat{\Theta}$ and Θ below are properly chosen universal positive continuous and nondecreasing functions on the nonnegative ray):*

(i) *All iterates (t^i, z^i) are well-defined and satisfy (P_i).*

(ii) *Whenever $t^{i+1} > 0$, one has*

$$t^i - t^{i+1} \geq \widehat{\Theta}^{-1}(\mu).$$

In particular,

$$i > \widehat{\Theta}(\mu) + 1 \Rightarrow t^i = 0.$$

(iii) *If i is such that $t^i = 0$, then*

$$\nu_{i+1} = \nu(f, z^{i+1}) \leq \frac{\nu_i^2}{(1 - \nu_i)^2}. \tag{25}$$

(iv) *Let $\epsilon \in (0, 1)$. The method finds an ϵ-approximate saddle point z_ϵ of f:*

$$\nu(f, z_\epsilon) \leq \epsilon \quad [\Rightarrow \mu(f, z_\epsilon) \leq -\epsilon - \ln(1 - \epsilon)]$$

in no more than

$$\Theta(\mu(f, \widehat{z})) + O(1) \ln \ln \left(\frac{3}{\epsilon}\right)$$

steps, $O(1)$ being an absolute constant.

6 The path-following scheme

We are about to achieve our main target – developing an efficient interior point method for approximating saddle point of a "good enough" convex-concave function, one which admits s.-c.c.-c. regularizations. The method in question is based on the path-following scheme. Namely, let $f : X \times Y \to \mathbf{R}$ be the function in question. We associate with f and "good" (s.-c.) barriers F and G for $\operatorname{cl} X$, $\operatorname{cl} Y$, respectively, the family

$$f_t(x, y) = t f(x, y) + F(x) - G(y),$$

$t > 0$ being penalty parameter; a specific *regularity* assumption we impose on f implies that all functions from the family are nondegenerate s.-c.c.-c. Under minimal additional assumptions (e.g., in the case of bounded X, Y) every function f_t has a unique saddle point $z^*(t) = (x^*(t), y^*(t))$ on $X \times Y$, and the path $z^*(t)$ converges to the saddle set of f in the sense that

$$\mu(f, z^*(t)) \equiv \sup_{y \in Y} f(x^*(t), y) - \inf_{x \in X} f(x, y^*(t)) \to 0, \ t \to \infty.$$

In the method, we trace the path $z^*(t)$ as $t \to \infty$. Namely, given a current iterate (t^i, z^i) with z^i "close", in certain exact sense, to $z^*(t^i)$, we update it into a new pair (t^{i+1}, z^{i+1}) of the same type with $t^{i+1} = (1+\alpha)t^i$, the "penalty rate" $\alpha > 0$ being a parameter of the scheme. In order to compute z^{i+1}, we apply to $f_{t^{i+1}}$ the Saddle Newton method, z^i being the starting point, and run the method until closeness to the new target point $z^*(t^{i+1})$ is restored.

We start our developments with specifying the regularity assumption we need.

6.1 Regular convex-concave functions

For an open convex domain $G \subset \mathbf{R}^k$ and a point $u \in G$ let

$$\pi_u^G(h) = \inf\{t \mid u \pm t^{-1}h \in G\}$$

be the Minkowski function of the symmeterized domain $(G-u) \cap (u-G)$. In what follows, we associate with a positive semidefinite $k \times k$ matirx Q the seminorm

$$\| h \|_Q \equiv \sqrt{h^T Q h}$$

on $\mathbf{R}^k$.

Definition 6.1 *Let $X \subset \mathbf{R}^n$, $Y \subset \mathbf{R}^m$ be open convex sets, let $Z = X \times Y$ and let $f(x,y) : Z \to \mathbf{R}$ be a C^3 function which is convex in $x \in X$ for every $y \in Y$ and is concave in $y \in Y$ for every $x \in X$. Let also $\beta \geq 0$.*

(i) *Let B be a convex s.-c. function on Z. We say that f is β-compatible with B, if for all $z \in Z, h \in \mathbf{R}^n \times \mathbf{R}^m$ we have*

$$|D^3 f(z)[h,h,h]| \leq \beta \| h \|^2_{\mathrm{S}_f(z)} \| h \|_{B''(z)} .$$

(ii) *We say that f is β-regular, if for all $z \in Z, h \in \mathbf{R}^n \times \mathbf{R}^m$ one has*

$$|D^3 f(z)[h,h,h]| \leq \beta \| h \|^2_{\mathrm{S}_f(z)} \pi_z^Z(h).$$

Note that a β-regular c.-c. function $f : Z \equiv X \times Y \to \mathbf{R}$ is β-compatible with any s.-c. convex function on Z, see Proposition 6.1.(i) below.

6.1.1 Examples of regular functions

Let us look at examples of regular functions. We start with the following evident observation:

Example 6.1 *Let $f(x,y)$ be a (perhaps nonhomogeneous) quadratic function convex in $x \in \mathbf{R}^n$ and concave in $y \in \mathbf{R}^m$. Then f is 0-regular on $\mathbf{R}^n \times \mathbf{R}^m$.*

The next two examples are less trivial:

Example 6.2 *Let*

$$S(y) : \mathbf{R}^m \to \mathbf{S}^n$$

be a (perhaps nonhomogeneous) quadratic mapping taking values in the space $\mathbf{S}^n$ *of symmetric* $n \times n$ *matrices, and assume that the mapping is concave w.r.t. the cone* $\mathbf{S}^n_+$ *of positive semidefinite* $n \times n$ *matrices:*

$$\lambda \in [0,1] \Rightarrow S(\lambda y') + S((1-\lambda)y'') \preceq S(\lambda y' + (1-\lambda)y'') \quad \forall y', y'' \in \mathbf{R}^m.$$

[example: $S(y) = A + By + y^T B^T - y^T C y$, where y runs over the space of $m \times n$ matrices, A, B, C are fixed matrices of appropriate sizes, A, C are symmetric and C is positive semidefinite].

Denote

$$Y = \{y \mid \; S(y) \in \mathbf{S}^n_{++}\},$$

$\mathbf{S}^n_{++}$ *being the interior of* $\mathbf{S}^n_+$*, and let*

$$f(x,y) = x^T S(y) x : Z \equiv \mathbf{R}^n \times Y \to \mathbf{R}.$$

The function f *is* 5*-regular.*

Example 6.3 *For* $a = (a_1, ..., a_m)^T \in \mathbf{R}^m$ *and* $u = (u_1, ..., u_m)^T \in \mathbf{R}^m_{++}$, $\mathbf{R}^m_{++}$ *being the interior of* $\mathbf{R}^m_+$*, let* $u^a = (u_1^{a_1}, u_2^{a_2}, ..., u_m^{a_m})^T$*. Now, let* $a, b \in \mathbf{R}^m_+$ *be such that* $0 \le a_i \le 1$, $0 \le b_i$, $i = 1, ..., m$*. The function*

$$f(x,y) = \ln \mathrm{Det}(E^T \mathrm{Diag}(y^a)\mathrm{Diag}(x^{-b})E) : Z \equiv \mathbf{R}^m_{++} \times \mathbf{R}^m_{++} \to \mathbf{R},$$

E *being an* $m \times n$ *matrix of rank* n*, is* $21(1+ \parallel b \parallel_\infty)^2$*-regular.*

The number of examples can be easily extended by applying "combination rules" as follows:

Proposition 6.1 (i) *If* $f : X \times Y \to \mathbf{R}$ *is* β*-regular,* f *is* β*-compatible with any s.-c. function on* Z*.*

(ii) *Let* $f(x,y) : X \times Y \to \mathbf{R}$ *be* β*-regular, and let* $X' \subset X$, $Y' \subset Y$ *be nonempty open convex sets. Then the restriction of* f *on* $X' \times Y'$ *is* β*-regular.*

(iii) *Let* $\alpha_i \ge 0$, $X_i \subset \mathbf{R}^n$, $Y_i \subset \mathbf{R}^m$*, let* $Z_i = X_i \times Y_i$, $i = 1, ..., k$*, and let the set* $Z = \bigcap\limits_{i=1}^{k} Z_i$ *be nonempty. Let also* $f_i : Z_i \to \mathbf{R}$, $i = 1, ..., k$*. If* f_i *are* β_i*-compatible with s.-c. functions* $B_i : Z_i \to \mathbf{R}$, $i = 1, ..., k$*, then the function* $f(z) = \sum\limits_{i=1}^{k} \alpha_i f_i(z) : Z \to \mathbf{R}$ *is* $(\max\limits_i \beta_i)$*-compatible with the s.-c. function* $B = \sum\limits_i B_i : Z \to \mathbf{R}$*. If* f_i *are* β_i*-regular on* Z_i, $i = 1, ..., k$*, then* $f(z)$ *is* $(\max\limits_i \beta_i)$*-regular on* Z*.*

(iv) *Let* $Z = X \times Y \subset \mathbf{R}^n \times \mathbf{R}^m$*, and let* $\Pi(u,v) = \begin{pmatrix} x = Pu + p \\ y = Qv + q \end{pmatrix}$ *be an affine mapping from* $\mathbf{R}^\nu \times \mathbf{R}^\mu$ *to* $\mathbf{R}^n \times \mathbf{R}^m$ *with the image intersecting* Z*. If a function* $f : Z \to \mathbf{R}$ *is* β*-compatible with a s.-c. function* $B : Z \to \mathbf{R}$*, then the superposition*

$$f^+(u,v) = f(\Pi(u,v)) : \Pi^{-1}(Z) \to \mathbf{R}$$

is β*-compatible with the s.-c. function* $B^+(u,v) = B(\Pi(u,v))$*. If a function* $f : Z \to \mathbf{R}$ *is* β*-regular, so is* $f^+(u,v)$*.*

(v) *Let* $f(x,y) : Z = X \times Y \to \mathbf{R}$ *be* β*-regular, and let*

$$U \;=\; \{(\lambda, u) \mid \lambda > 0, \lambda^{-1} u \in X\}, \quad V \;=\; \{(\mu, v) \mid \mu > 0, \mu^{-1} v \in Y\}.$$

Then the function

$$\phi((\lambda,u),(\mu,v)) = \lambda\mu f(\lambda^{-1}u, \mu^{-1}v) : U \times V \to \mathbf{R}$$

is $(4\beta + 9)$*-regular.*

6.2 Path-following method: preliminaries

The following two results are basic for us:

Proposition 6.2 *Let $X \subset \mathbf{R}^n, Y \subset \mathbf{R}^m$ be open nonempty convex domains, let $\vartheta \geq 1$, and let F, G be ϑ-s.-c. barriers for $\operatorname{cl} X$, $\operatorname{cl} Y$, respectively. Assume that a function $f : Z \equiv X \times Y \to \mathbf{R}$ is β-compatible with the s.-c.b. $F(x) + G(y)$ for $\operatorname{cl} Z$, and that*

> **(C)** *the matrices $\nabla^2_{xx}[f(x,y) + F(x)]$, $\nabla^2_{yy}[G(y) - f(x,y)]$ are nondegenerate for some $(x,y) \in Z$.*

Then the family

$$\{f_t(x,y) = \gamma\,[tf(x,y) + F(x) - G(y)]\}_{t>0}\,, \quad \gamma = \left(\frac{\beta+2}{2}\right)^2$$

is comprised of nondegenerate s.-c.c.-c. functions.

Condition **(C)** *is for sure satisfied when X, Y do not contain lines.*

Proposition 6.3 *Under the same assumptions and in the same notation as in Proposition 6.2, assume that a pair $(t > 0, \bar{z} = (\bar{x}, \bar{y}) \in Z)$ is such that*

$$\nu(f_t, \bar{z}) \leq 0.1.$$

Then

$$\mu(f, \bar{z}) = \sup_{y \in Y} f(\bar{x}, y) - \inf_{x \in X} f(x, \bar{y}) \leq \frac{4\vartheta}{t}. \tag{26}$$

Let also $\alpha > 0$ and

$$t^+ = (1+\alpha)t.$$

Then

$$\mu(f_{t^+}, \bar{z}) \leq \mathcal{R}_{\beta,\vartheta}(\alpha) \equiv 0.25\alpha\sqrt{\gamma\vartheta} + 0.02(1+\alpha) + 2\gamma\vartheta[\alpha - \ln(1+\alpha)]. \tag{27}$$

In particular, for every $\chi \geq 0$

$$\alpha = \frac{\chi}{\sqrt{\gamma\vartheta}} \Rightarrow \mu(f_{t^+}, \bar{z}) \leq 0.02 + 0.3\chi + \chi^2.$$

6.3 The Basic path-following method

Now we can present the Basic path-following method for approximating a saddle point of a convex-concave function $f : Z = X \times Y \to \mathbf{R}$. We assume that

A.1. $X \subset \mathbf{R}^n$, $Y \subset \mathbf{R}^m$ are open and convex;

A.2. f is β-compatible with the barrier $F(x) + G(y)$ for $\operatorname{cl} Z$, F, G being given ϑ-s.-c. barriers for $\operatorname{cl} X$, $\operatorname{cl} Y$, respectively, and there exists $(x,y) \in X \times Y$ such that both $\nabla^2_{xx}[f(x,y) + F(x)]$ and $\nabla^2_{yy}[G(y) - f(x,y)]$ are positive definite;

A.3. There exists $\widehat{y} \in Y$ such that $f(\cdot, \widehat{y})$ has bounded level sets $\{x \in X \mid f(x, \widehat{y}) \leq a\}$, $a \in \mathbf{R}$, and there exists $\widehat{x} \in X$ such that $f(\widehat{x}, \cdot)$ has bounded level sets $\{y \in Y \mid f(\widehat{x}, y) \geq a\}$, $a \in \mathbf{R}$.

Let us associate with f, F, G the family

$$\{f_t(x,y) = \gamma[tf(x,y) + F(x) - G(y)]\}_{t\geq 0} \quad \left[\gamma = \left(\frac{\beta+2}{2}\right)^2\right] \tag{28}$$

of convex-concave mappings $Z \to \mathbf{R}$. Note that the family is comprised of nondegenerate s.-c.c.-c. functions (Proposition 6.2).

Lemma 6.1 *Under assumptions* **A.1** – **A.3** *every function f_t, $t > 0$, has a unique saddle point $z^*(t)$ on Z, and the path $x^*(t)$ is continuously differentiable.*

Now we can present the Basic path-following method associated with f, F, G:

Basic path-following method:

- *Initialization:* Find starting pair $(t^0, z^0) \in \mathbf{R}_{++} \times Z$ such that

$$\nu(f_{t^0}, z^0) \leq 0.1 \tag{$\mathcal{P}_0$}$$

- *Step i, $i \geq 1$:* Given previous iterate $(t^{i-1}, z^{i-1}) \in \mathbf{R}_{++} \times Z$ satisfying

$$\nu(f_{t^{i-1}}, z^{i-1}) \leq 0.1, \tag{$\mathcal{P}_{i-1}$}$$

  1. Set $t^i = (1+\alpha)t^{i-1}$, $\alpha > 0$ being the parameter of the method
  2. Apply to $g(z) \equiv f_{t^i}(z)$ the Saddle Newton method from Section 5.3, $\widehat{z} \equiv z^{i-1}$ being the starting point. Run the method until a point satisfying the relation $\nu(g, \cdot) \leq 0.1$ is generated, and take this point as z^i. Step i is completed.

The efficiency of the Basic path-following method is given by the following

Theorem 6.1 *Under assumptions* **A.1** – **A.3** *the Basic path-following method is well-defined (i.e., for every i rule 2 yields z^i in finitely many steps). The approximations $z^i = (x^i, y^i)$ generated by the method satisfy the accuracy bound*

$$\sup_{y\in Y} f(x^i, y) - \inf_{x\in X} f(x, y^i) \leq (1+\alpha)^{-i}\frac{4\vartheta}{t_0}.$$

The Newton complexity (# of Newton steps required by rule 2) of every iteration of the method can be bounded from above as

$$\Theta_*(\mathcal{R}_{\beta,\vartheta}(\alpha)).$$

Here $\Theta_*(\cdot)$ is a universal continuous function on the nonnegative ray and $\mathcal{R}_{\beta,\theta}$ is given by (27).

In particular, with the setup

$$\alpha = \frac{2\chi}{(\beta+2)\sqrt{\vartheta}} \quad [\chi > 0] \tag{29}$$

the Newton complexity of every iteration of the Basic path-following method does not exceed a universal function of χ.

The result of the Theorem is an immediate consequence of Propositions 6.2 and 6.3.

Remark 6.1 In the case of bounded X, Y, in order to initialize the Basic path-following method one can use the same scheme as in the optimization case, namely, find a tight approximation z^0 to the minimizer of the barrier $B(x,y) = \gamma(F(x) + G(y))$ for $\mathrm{cl}\, Z$, say, one with $\lambda(B, z^0) \leq 0.05$. After such a point is found, one may choose as t^0 the largest t such that $\nu(f_t, z^0) \leq 0.1$ (such a t exists, since $\nu(f_t, z^0)$ is continuous in t and $\nu(f_0, z^0) = \lambda(B, z^0)$).

6.4 Self-concordant families associated with regular convex-concave functions

In some important cases family (28) is so simple that one can optimize analytically the functions $f_t(\cdot,\cdot)$ in either x or y. Whenever it is the case, the Basic path-following method from the previous section can be simplified, and the complexity bound can be slightly improved. Namely, let us add to **A.1** – **A.3** the assumption

A.4. The functions

$$\Phi(t,x) = \sup_{y\in Y} f_t(x,y)$$

are well-defined on X and "are available" (i.e., one can compute their values, gradients and Hessians at every $x \in X$).

Note that under the assumptions **A.1** – **A.4** the family $\{\Phi(t,\cdot)\}_{t>0}$ is comprised of nondegenerate s.-c. convex functions (Propositions 6.2 and 4.3), and by Lemma 6.1 the functions $\Phi(t,\cdot)$ are below bounded on X, so that the path

$$x^*(t) = \operatorname*{argmin}_x \Phi(t,x)$$

is well-defined; it is the x-component of the path $z^*(t)$ of the saddle points of the functions f_t. Moreover, by **A.4** for every $\xi \in X$ and $\tau > 0$ the function $-f_\tau(\xi,\cdot)$ is a convex nondegenerate (see Proposition 6.2) s.-c. and below bounded function on Y; consequently, there exists a unique $\widehat{y} \equiv \widehat{y}(\tau,\xi) \in Y$ such that

$$\Phi(\tau,\xi) = f_\tau(\xi,\widehat{y}(\tau,\xi)).$$

Consider the standard scheme of tracing the path $x^*(t)$:

(S) *Given an iterate $(\bar{t},\bar{x})$ "close to the path" – satisfying the predicate*

$$\lambda(\Phi(t,\cdot),x) \le \kappa \quad [<0.1], \tag{$\mathcal{C}(t,x)$}$$

we update it into a new iterate (t^+,x^+) with the same property as follows:

• *first, we compute the "improved" iterate*

$$\widetilde{x} = \bar{x} - [\nabla^2_{xx}\Phi(\bar{t},\bar{x})]^{-1}\nabla_x\Phi(\bar{t},\bar{x}); \tag{30}$$

• *second, we choose a $t^+ > \bar{t}$ and apply to the function $\Phi(t^+,\cdot)$ the Damped Newton method*

$$x \mapsto x - \frac{1}{1+\lambda(\Phi(t^+,\cdot),x)}[\nabla^2_{xx}\Phi(t^+,x)]^{-1}\nabla_x\Phi(t^+,x),$$

starting the method with $x = \widetilde{x}$, until a point x^+ such that $\lambda(\Phi(t^+,\cdot),x^+) \le \kappa$ is generated.

Theorem 6.2 *Under assumptions* **A.1** – **A.4** *for every pair $(\bar{t},\bar{x})$ satisfying $(\mathcal{C}(\cdot,\cdot))$ one has*

$$\sup_{y\in Y} f(\widetilde{x},y) - \inf_{x\in X}\sup_{y\in Y} f(x,y) \le \frac{4\vartheta}{\bar{t}}. \tag{31}$$

Moreover, for every $t^+ > \bar{t}$ the number of damped Newton steps required by the updating **(S)** *does not exceed the quantity*

$$O(1)\left[\rho(\kappa) + \frac{3}{2}\left(1+\sqrt{\gamma\vartheta}\right)\frac{t^+-\bar{t}}{\bar{t}} + \gamma\vartheta\left[\frac{t^+-\bar{t}}{\bar{t}} - \ln\frac{t^+}{\bar{t}}\right] + \ln\ln\frac{1}{\kappa}\right], \tag{32}$$

$O(1)$ being an absolute constant.

Theorem 6.2 says, e.g., that under assumptions **A.1** – **A.4** we can trace the path of minimizers $x^*(t) = \operatorname*{argmin}_{x \in X} \Phi(t,x)$ of the functions $\Phi(t,x) = \max_{y \in Y} f_t(x,y)$, increasing the penalty parameter t linearly at the rate $(1 + O(1)(\gamma\vartheta)^{-1/2})$ and accompanying every updating of t by an absolute constant Newton-type "corrector" steps in x. The outlined possibility to trace efficiently the path $x^*(\cdot)$ correlates to the fact that we can use the Basic path-following method to trace the path $\{(x^*(t), y^*(t))\}$ of saddle points of the underlying family $\{f_t(x,y)\}$ of s.-c.c.-c. functions. Both processes have the same theoretical complexity characteristics.

6.4.1 Application: Exterior penalty method for Semidefinite Programming

Consider a semidefinite program with convex quadratic objective:

$$\phi(x) \equiv \frac{1}{2}x^T Ax + b^T x \to \min \mid \mathcal{A}(x) \succeq 0 \quad [x \in \mathbf{R}^n], \tag{SDP}$$

$\mathcal{A}(x)$ being an affine mapping taking values in the space $\mathbf{S} = \mathbf{S}^\vartheta$ of symmetric $\vartheta \times \vartheta$ matrices of a given block-diagonal structure. For the sake of simplicity, we assume A to be positive definite.

Let

$$f(x,y) = \phi(x) - \mathrm{Tr}(y\mathcal{A}(x)) : \mathbf{R}^n \times \mathbf{S} \to \mathbf{R};$$

this function clearly is convex-concave and therefore 0-regular (it is quadratic!). Given large positive T, let us set

$$X = \mathbf{R}^n, \; Y_T = \{y \in \mathbf{S} \mid 0 \preceq y \preceq 2TI\},$$

I being the unit matrix of the size ϑ. We have

$$\max_{y \in Y_T} f(x,y) = \phi(x) + 2T\rho_-(x),$$

where $\rho_-(x)$ is the sum of modulae of the negative eigenvalues of $\mathcal{A}(x)$; thus, for large T the saddle point of f on $X \times Y_T$ is a good approximate solution to (SDP). Moreover, if (SDP) satisfies the Slater condition – there exists x with positive definite $\mathcal{A}(x)$ – then, for all large enough T, the x-component of a saddle point of f on $X \times Y_T$ is an exact optimal solution to (SDP). In order to approximate this saddle point, we can use the Basic path-following method, choosing as $F(x)$ the trivial barrier – identically zero – for $X = \mathbf{R}^n$ and choosing as G the (2ϑ)-s.-c. barrier

$$G(y) = -\ln \mathrm{Det}(y) - \ln \mathrm{Det}(2TI - y),$$

thus coming to the family

$$\{f_t(x,y) = t[\phi(x) - \mathrm{Tr}(y\mathcal{A}(x))] + \ln \mathrm{Det}(y) + \ln \mathrm{Det}(2TI - y)\}_{t>0}.$$

Since A is positive definite, our data satisfy the assumptions from Proposition 6.2, and we can apply the Basic path-following method to trace the path $(x^*(t), y^*(t))$ of saddle points of f_t as $t \to \infty$, thus approximating the solution to (SDP). Note that in the case in question the assumption **A.4** also is satisfied: a straightforward computation yields

$$\begin{array}{rcl}\Phi(t,x) & \equiv & \max_{y \in Y_T} f_t(x,y) = t\phi(x) + \Phi_T(t\mathcal{A}(x)), \\ \Phi_T(y) & = & \mathrm{Tr}\left[T^2y^2\left(I + (I + T^2y^2)^{1/2}\right)^{-1} - Ty\right] - \ln \mathrm{Det}\left(I + (I + T^2y^2)^{1/2}\right) + 2\ln T.\end{array}$$

Note that in both processes – tracing the saddle point path $(x^*(t), y^*(t))$ by the Basic path-following method and tracing the path $x^*(t)$ by iterating updating **(S)** – no problem of finding initial feasible solution to (SDP) occurs, so that the methods in question can be viewed as infeasible-start methods for (SDP) with $\sqrt{\vartheta}$-rate of convergence. Note also that the family $\{\Phi(t, \cdot)\}$ is in fact an *exterior penalty* family: as $t \to \infty$, the functions $\frac{1}{t}\Phi(t, x)$ converge to the function $\phi(x) + 2T\rho_-(x)$, which, under the Slater condition and for large enough value of T, is an exact exterior penalty function for (SDP).

Note also that we could replace in the outlined scheme the set Y_T with another "bounded approximation" of the cone $\mathbf{S}_+^{\vartheta}$, like $Y_T = \{0 \preceq y, \mathrm{Tr}(y) \leq T\}$ or $Y_T = \{0 \preceq y, \mathrm{Tr}(y^2) \leq T^2\}$, using as G the standard s.-c. barriers for the resulting sets. All indicated sets Y_T are simple enough to allow for explicit computation of the associated functions $\Phi(t, \cdot)$ and lead therefore to polynomial time exterior penalty schemes. Our scheme works also for Linear and Conic Quadratic Programming problems (cf. [5], Section 3.4).

7 Application: Inscribing maximal volume ellipsoid into a polytope

7.1 The problem

Consider a (bounded) polytope represented as

$$\Pi = \{\xi \in \mathbf{R}^n \mid e_i^T \xi \leq 1, \ i = 1, ..., m\}.$$

We are interested to approximate the ellipsoid of the largest volume among those contained in the polytope. This problem is of interest for Control Theory (see [1]) and especially for Nonsmooth Convex Optimization, where it is the basic auxiliary problem arising in the *Inscribed Ellipsoid method* (Khachiyan, Tarasov, Erlikh [2]); the latter method, as applied to general convex problems, turns out to be optimal in the sense of Information-based Complexity Theory.

The best known so far complexity estimates for finding an ϵ-solution to the problem, i.e., for identifying an inscribed ellipsoid with the volume $\geq (1 - \epsilon)$ times the maximal one, in the case of $m = O(n)$ are $O(n^{4.5} \ln(n\epsilon^{-1}R))$ and $O(n^{3.5} \ln(n\epsilon^{-1}R) \ln(n\epsilon^{-1} \ln R))$ arithmetic operations, see [5], Chapter 6, and [3], respectively; here R is an a priori known ratio of radii of two centered at the origin Euclidean balls, the smaller being contained in Π and the larger containing the polytope. These complexity bounds are given by interior point methods as applied to the standard setting of the problem:

$$-\ln \mathrm{Det} A \to \min \ \text{s.t.} \ \sqrt{e_i^T A^2 e_i} \leq x_i(\xi) \equiv 1 - e_i^T \xi, \ i = 1, ..., m, \ A \in \mathbf{S}_{++}^n, \qquad (P_{\rm ini})$$

where $\mathbf{S}_{++}^n$ is the interior of the cone $\mathbf{S}_+^n$ of positive semidefinite $n \times n$ matrices. The design variables in the problem are a symmetric $n \times n$ matrix A and a vector $\xi \in \mathbf{R}^n$ which together define the ellipsoid $\{u = \xi + Av \mid v^T v \leq 1\}$, and the design dimension of the problem is $O(n^2)$. We shall demonstrate that the problem admits saddle point reformulation with "small" – $O(m)$ – design dimension; moreover, the arising convex-concave function is $O(1)$-regular, $O(1)$ being an absolute constant, so that one can solve the resulting saddle point problem by the path-following method from Section 6. As a result, we come to the complexity bound $O(n^{3.5} \ln(n\epsilon^{-1}R))$. In contrast to the constructions from [5, 3], this bound arises naturally, without sophisticated ad hoc tricks exploiting the specific structure of $(P_{\rm ini})$.

The rest of the section is organized as follows. We start with the saddle point reformulation of $(P_{\rm ini})$ and demonstrate that the arising convex-concave function indeed is $O(1)$-regular (Section 7.2). Section 7.3 presents the Basic path-following algorithm as applied to the particular saddle point problem we are interested in, including initialization scheme, issues related to recovering nearly-optimal ellipsoids from approximate solutions to the saddle point problem and the overall complexity analysis of the algorithm.

7.2 Saddle point reformulation of $(P_{\rm ini})$

Note that for every feasible solution (ξ, A) to the problem one has $\xi \in \mathrm{int}\,\Pi$, i.e.,

$$x(\xi) \equiv (x_1(\xi), ..., x_m(\xi))^T > 0.$$

In other words, $(P_{\rm ini})$ is the optimization problem

$$\mathcal{V}(\xi) \equiv \inf\{-\ln \mathrm{Det} A \mid A \in \mathrm{int}\, \mathbf{S}^n_+,\ e_i^T A^2 e_i \le x_i^2(\xi),\ i = 1, ..., m\} \to \min \mid x(\xi) > 0. \tag{33}$$

Setting $B = A^2$, we can rewrite the definition of $\mathcal{V}$ as

$$\mathcal{V}(\xi) = \inf\{-\frac{1}{2} \ln \mathrm{Det} B \mid B \in \mathbf{S}^n_{++},\ e_i^T B e_i - x_i^2(\xi) \le 0,\ i = 1, ..., m\} \quad [\xi \in \mathrm{int}\,\Pi].$$

The optimization problem in the right hand side of the latter relation clearly is convex in B and satisfies the Slater condition; therefore

$$\mathcal{V}(\xi) = \sup_{z \in \mathbf{R}^m_+} \inf_{B \in \mathbf{S}^n_{++}} \left[-\frac{1}{2} \ln \mathrm{Det} B + \sum_{i=1}^m z_i (e_i^T B e_i - x_i^2(\xi)) \right]. \tag{34}$$

It is easily seen that in the latter formula we can replace $\sup\limits_{z \in \mathbf{R}^m_+}$ with $\sup\limits_{z \in \mathbf{R}^m_{++}}$. For $z \in \mathbf{R}^m_{++}$, optimization with respect to B in (34) can be carried out explicitly: the corresponding problem is

$$\inf_{B \in \mathbf{S}^m_{++}} \phi(B), \quad \phi(B) = -\frac{1}{2} \ln \mathrm{Det} B + \sum_{i=1}^m z_i (\mathrm{Tr}(B e_i e_i^T) - x_i^2(\xi));$$

for a positive definite symmetric B, we have

$$D\phi(B)[H] = \mathrm{Tr}\left(\left[-\frac{1}{2} B^{-1} + \sum_{i=1}^m z_i e_i e_i^T \right] H \right).$$

Setting

$$Z = \mathrm{Diag}(z),\ E = [e_1; ...; e_m],\ B^* = \left[2 \sum_{i=1}^m z_i e_i e_i^T \right]^{-1} = \left[2E^T Z E \right]^{-1},$$

(B^* is well defined, since $z > 0$ and E is of rank n – otherwise Π would be unbounded), we conclude that $\phi'(B^*) = 0$, and therefore B^* is the desired minimizer of the convex function ϕ. We now have

$$\begin{aligned}
\phi(B^*) &= -\tfrac{1}{2} \ln \mathrm{Det}([2E^T Z E]^{-1}) + \sum_{i=1}^m z_i (\mathrm{Tr}(B^* e_i e_i^T) - x_i^2(\xi)) \\
&= \tfrac{n \ln 2}{2} + \tfrac{1}{2} \ln \mathrm{Det}(E^T Z E) + \mathrm{Tr}(B^* E^T Z E) - \sum_{i=1}^m z_i x_i^2(\xi) \\
&= \tfrac{n \ln 2 + n}{2} + \tfrac{1}{2} \ln \mathrm{Det}(E^T Z E) - \sum_{i=1}^m z_i x_i^2(\xi).
\end{aligned}$$

Thus, (34) becomes

$$\begin{array}{rcl}\mathcal{V}(\xi) &=& \frac{n\ln 2+n}{2} + \sup\limits_{z\in\mathbf{R}^m_{++}}\left[\frac{1}{2}\ln\mathrm{Det}(E^TZE) - \sum\limits_{i=1}^m z_i x_i^2(\xi)\right]\\ &=& \frac{n\ln 2+n}{2} + \sup\limits_{y\in\mathbf{R}^m_{++}}\left[\frac{1}{2}\ln\mathrm{Det}(E^TYX^{-1}(\xi)E) - y^Tx(\xi)\right],\\ Y=\mathrm{Diag}(y), && X(\xi)=\mathrm{Diag}(x(\xi)) \qquad [\text{substitution } z_i = y_i x_i^{-1}(\xi)].\end{array}$$

We have proved the following

Proposition 7.1 *Whenever* $\xi\in\mathrm{int}\,\Pi$, *one has*

$$\mathcal{V}(\xi) = \frac{n\ln 2+n}{2} + \sup_{y\in\mathbf{R}^m_{++}}\left[\frac{1}{2}\ln\mathrm{Det}(E^TYX^{-1}(\xi)E) - y^Tx(\xi)\right],\quad Y=\mathrm{Diag}(y),\ X(\xi)=\mathrm{Diag}(x(\xi)).$$

Consequently, as far as the ξ*-component of the solution is concerned, to solve* ($P_{\rm ini}$) *is the same as to solve the saddle point problem*

$$\min_{\xi\in\mathrm{int}\,\Pi}\sup_{y\in\mathbf{R}^m_{++}} f(\xi,y),\quad f(\xi,y)=\ln\mathrm{Det}(E^TYX^{-1}(\xi)E) - 2y^Tx(\xi). \qquad (P_{\rm s})$$

Note that the equivalence between the problems ($P_{\rm ini}$) and ($P_{\rm s}$) is only partial: the component A of the solution to ($P_{\rm ini}$) "is not seen explicitly" in ($P_{\rm s}$). However, we shall see in Section 7.3 that there is a straightforward computationally cheap way to update an "ϵ-approximate saddle point of f" into an ϵ-solution to ($P_{\rm ini}$). Note also that in some applications, e.g., in the Inscribed Ellipsoid method, we are not interested in the A-part of a nearly optimal ellipsoid; all used by the method is the center of such an ellipsoid, and this center is readily given by a good approximate saddle point of ($P_{\rm s}$).

The following fact is crucial for us:

Proposition 7.2 *The function* $f(\xi,y)$ *defined in* ($P_{\rm s}$) *is 84-regular on its domain* $Z=\mathrm{int}\,\Pi\times\mathbf{R}^m_{++}$.

Proof. Indeed, from Example 6.3 we know that the function

$$g(x,y)=\ln\mathrm{Det}(E^T\mathrm{Diag}(y)\mathrm{Diag}^{-1}(x)E)$$

is 84-regular on $\mathbf{R}^m_{++}\times\mathbf{R}^m_{++}$. The function $f(\xi,y)$ is obtained from g by affine substitution of argument $\xi\mapsto x(\xi)$ with subsequent adding bilinear function $-2y^Tx(\xi)$, and these operations preserve regularity in view of Proposition 6.1. ∎

7.3 Basic path-following method as applied to ($P_{\rm s}$)

In view of Proposition 7.2, we can solve the saddle point problem ($P_{\rm s}$) by the Basic path-following method from Section 6. To this end we should first of all specify the underlying s.-c. barriers $F(\xi)$ for Π and $G(y)$ for $\mathbf{R}^m_+$. Our choice is evident:

$$F(\xi) = -\sum_{i=1}^m \ln x_i(\xi);\quad G(y) = -\sum_{i=1}^m \ln y_i,$$

which results in the parameter of self-concordance $\vartheta = m$.

The next step is to check validity of the assumptions **A.1** – **A.3**, which is immediate. Indeed, **A.1** is evident; **A.2** is readily given by Propositions 7.2 and 6.1.(i). To verify **A.3**, it suffices to set $\widehat{\xi} = 0$, $\widehat{y} = e$, where $e = (1, ..., 1)^T \in \mathbf{R}^m$. Indeed, since in the case in question $X = \operatorname{int}\Pi$ is bounded, all we need to prove is that the function $g(y) = \ln\operatorname{Det}(E^T \operatorname{Diag}(y) E) - 2e^T y$ on $Y = \mathbf{R}^m_{++}$ has bounded level sets $\{y \in Y \mid g(y) \geq a\}$. This is evident, since

$$g(y) \leq O(1 + \ln \| y \|_\infty) - 2e^T y.$$

Thus, we indeed are able to solve (P_s) by the Basic path-following method as applied to the family

$$\left\{ f_t(\xi, y) = \gamma \left[t \ln\operatorname{Det}\left(E^T \operatorname{Diag}(y) \operatorname{Diag}^{-1}(x(\xi)) E \right) - 2ty^T x(\xi) - \sum_{i=1}^m \ln x_i(\xi) + \sum_{i=1}^m \ln y_i \right] \right\}_{t>0}, \quad (35)$$

γ being an appropriate absolute constant.

For the sake of definiteness, let us speak about the Basic path-following method with the penalty updating rate of type (29):

$$\alpha = \frac{\chi}{\sqrt{\vartheta}}, \quad (36)$$

$\chi > 0$ being the parameter of the method.

To complete the description of the method, we should resolve the following issues:

- How to initialize the method, i.e., to find a pair (t^0, z^0) with $\nu(f_{t^0}, z^0) \leq 0.1$;
- How to convert a “nearly saddle point” of f to a “nearly maximal inscribed ellipsoid”.

These are the issues we are about to consider.

7.3.1 Initialization

Initialization of the Basic path-following method can be implemented as follows.

We start with the *Initialization Phase* – approximating the analytic center of Π. Namely, we use the standard interior point techniques to find “tight” approximation of the analytic center

$$\xi^* = \operatorname*{argmin}_{\xi \in \operatorname{int}\Pi} F(\xi) \quad [F(\xi) = -\sum_{i=1}^m \ln x_i(\xi)]$$

of the polytope Π. We terminate the Initialization Phase when a point $\xi^0 \in \operatorname{int}\Pi$ such that

$$\sqrt{[F'(\xi^0)]^T [F''(\xi^0)]^{-1} F'(\xi^0)} \leq \frac{0.05}{2\sqrt{\gamma}} \quad (37)$$

is generated and set

$$t^0 = \tfrac{0.05}{\sqrt{2m\gamma}}, \quad y^0 = [2t^0 X(\xi^0)]^{-1} e, \; e = (1, ..., 1)^T \in \mathbf{R}^m, \quad z^0 = (\xi^0, y^0).$$

We claim that the pair (t^0, z^0) can be used as the initial iterate in the path-following scheme:

Lemma 7.1 *One has* $\nu(f_{t^0}, z^0) \leq 0.1$.

7.3.2 Accuracy of approximate solutions

We start with the following

Proposition 7.3 *Assume that a pair* $(t > 0, z = (\xi, y))$ *is such that* $\nu(f_t, z) \le 0.1$, $f_t(\cdot)$ *being function from family* (35). *Then*

$$\mathcal{V}(\xi) - \inf_{\xi' \in \operatorname{int}\Pi} \mathcal{V}(\xi') \le \frac{2m}{t},$$

$\mathcal{V}$ *being given by* (33).

Proof. By Proposition 7.1, for every $\xi' \in \operatorname{int}\Pi$ one has

$$\mathcal{V}(\xi') = c(n) + \frac{1}{2} \sup_{y' \in \mathbf{R}^m_{++}} f(\xi', y'), \tag{38}$$

while by Proposition 6.3 we have (note that in the case in question $\vartheta = m$)

$$\sup_{y' \in \mathbf{R}^m_{++}} f(\xi, y') - \inf_{\xi' \in \operatorname{int}\Pi} f(\xi', y) \le \frac{4m}{t}. \tag{39}$$

It remains to note that

$$\begin{aligned}
\mathcal{V}(\xi) \quad & \underbrace{=}_{(a)} c(n) + \frac{1}{2} \sup_{y' \in \mathbf{R}^m_{++}} f(\xi, y') \underbrace{\le}_{(b)} \frac{2m}{t} + c(n) + \frac{1}{2} \inf_{\xi' \in \operatorname{int}\Pi} f(\xi', y) \\
& \le \frac{2m}{t} + c(n) + \frac{1}{2} \inf_{\xi' \in \operatorname{int}\Pi} \sup_{y' \in \mathbf{R}^m_{++}} f(\xi', y') \underbrace{=}_{(c)} \frac{2m}{t} + \inf_{\xi' \in \operatorname{int}\Pi} \mathcal{V}(\xi'),
\end{aligned}$$

with (a), (c) given by (38) and (b) given by (39). ■

We are about to prove the following

Proposition 7.4 *Let* $0 < \delta \le 0.01$, *and let* $(t, z = (\xi, y))$ *be such that* $\nu(f_t, z) \le \delta$. *Denote*

$$\begin{array}{c}
Y = \operatorname{Diag}(y); \quad X = \operatorname{Diag}(x(\xi)); \quad B = (E^T Y X^{-1}(\xi) E)^{-1}; \\
A = 2^{-1/2} B^{1/2}; \quad \widehat{A} = (1 + 10\delta)^{-1/2} A; \quad \epsilon = \frac{5m}{2t} + 15 n \delta.
\end{array}$$

and consider the ellipsoid

$$W = \{\xi + \widehat{A} u \mid u^T u \le 1\}$$

The ellipsoid W *is contained in* Π *and is* ϵ*-optimal: for any ellipsoid* $W' \subset \Pi$ *one has*

$$\ln \operatorname{Vol}(W') \le \ln \operatorname{Vol}(W) + \epsilon, \tag{40}$$

Vol *being the* n*-dimensional volume.*

7.3.3 Algorithm and complexity analysis

The entire path-following algorithm for solving the saddle point problem (P_s) is as follows:

Input: a matrix E specifying the polytope Π according to $\Pi = \{\xi \mid E\xi \le e\}$; a tolerance $\epsilon \in (0,1)$.

Initialization: apply the Initialization Phase from Section 7.3.1 to get a starting pair

$$\left(t^0 = \frac{0.05}{\sqrt{2m\gamma}}, z^0\right) \tag{41}$$

satisfying $\nu(f_{t^0}, z^0) \le 0.1$, $\{f_t\}_{t>0}$ being given by (35).

Main Phase: starting with (t^0, z^0), apply the Basic path-following method with penalty updating rate (36) to trace the path of saddle points of the family $\{f_t\}$. Terminate the process when for the first time an iterate $(\bar{t}, \bar{z} = (\bar{y}, \bar{\xi}))$ with $\bar{t} > 5m\epsilon^{-1}$ is generated.

Recovering of a nearly optimal ellipsoid: starting with $z = \bar{z}$, apply to the s.-c. function $f_{\bar{t}}(\cdot)$ the Saddle Newton method (Section 5.3) until an iterate z with

$$\nu(f_{\bar{t}}, z) < \delta \equiv \frac{\epsilon}{30n}$$

is generated. After it happens, use the pair $(\bar{t}, z)$ to build the resulting ellipsoid W as explained in Proposition 7.4.

The complexity of the presented algorithm is given by the following

Theorem 7.1 *Assume that for some $R \ge 1$ the polytope Π contains the centered at 0 Euclidean ball of a radius r and is contained in the concentric ball of the radius Rr. Assume also that the Initialization Phase is carried out by the standard path-following method for approximating the analytic center of a polytope, the method being started at the origin. Then for every given tolerance $\epsilon \in (0,1)$*

(i) *The algorithm terminates with an ϵ-optimal inscribed ellipsoid W, i.e., $W \subset \Pi$ and*

$$\ln \mathrm{Vol}(W) \ge \ln \mathrm{Vol}(W') - \epsilon$$

for every ellipsoid $W' \subset \Pi$.

(ii) *The Newton complexity of the method – the total # of Newton steps in course of running the algorithm – does not exceed*

$$N_{\mathrm{Nwt}} \le \Theta_+(\chi)\sqrt{m}\ln\left(\frac{2mR}{\epsilon}\right), \tag{42}$$

where $\Theta_+(\cdot)$ is a universal continuous function on the axis and $\chi > 0$ is the parameter of the penalty updating rule from (36).

(iii) *The arithmetic complexity of every Newton step does not exceed $O(1)m^3$, $O(1)$ being an absolute constant.*

Proof. Below $O(1)$'s denote appropriate positive absolute constants.

(i) is readily given by Proposition 7.4. Let us prove (ii). It is well-known (see, e.g., [5], Section 3.2.3) that under the premise of the theorem the Newton complexity of the Initialization Phase does not exceed $N^{(1)} = O(1)\sqrt{m}\ln(2mR)$. In view of Theorem 6.1, (36) and (41) the Newton complexity of the

Main Phase does not exceed $N^{(2)} = \Theta_+(\chi)\sqrt{m}\ln(2m/\epsilon)$. The Newton complexity of the Recovering Phase, in view of the fact that

$$\nu(f_{\bar{t}}, \bar{z}) \leq 0.1 \Rightarrow \mu(f_{\bar{t}}, \bar{z}) \leq O(1)$$

(see (21.a)) and by virtue of Theorem 5.1, does not exceed $N^{(3)} = O(1)\ln(2m/\epsilon)$. (ii) is proved.

It remains to prove (iii). The arithmetic cost of a step of the Initialization Phase, i.e., in the path-following approximation of the analytic center of Π, is known to be $\leq O(1)m^3$, so that all we need is to prove a similar bound for the arithmetic cost of a step at the two subsequent phases of the algorithm. These latter steps are of the same arithmetic cost, and for the sake of definiteness we can focus on a step of the Main Phase. The amount of computations at such a step is dominated by the necessity (a) to compute, given $(t, z = (\xi, y))$, the gradient $g = (f_t)'(z)$ and the Hessian $H = (f_t)''(z)$, and (b) to compute $H^{-1}g$. The arithmetic cost of (b) clearly is $O(1)m^3$, so that all we need is to get a similar bound for the arithmetic cost of (a).

It is easily seen that the amount of computations in (a) is dominated by the necessity to compute at a given point the gradient and the Hessian of the function

$$f(\xi, y) = \ln \mathrm{Det}(E^T \mathrm{Diag}(y)\mathrm{Diag}^{-1}(x(\xi))E).$$

The function in question is obtained from the function

$$\phi(x, y) = \ln \mathrm{Det}(E^T \mathrm{Diag}(y)\mathrm{Diag}^{-1}(x)E)$$

by affine substitution of argument $(\xi, y) \mapsto (x(\xi), y)$, so that computation of the gradient and the Hessian of f is equivalent to similar computations for ϕ plus $O(1)mn^2 \leq O(1)m^3$ additional computations needed to "translate" the arguments/results of the latter computation to those of the former one. Thus, all we need is to demonstrate that it is possible to compute the gradient g and the Hessian H of ϕ at a given point $(x, y) \in \mathbf{R}^m_{++} \times \mathbf{R}^m_{++}$ at the arithmetic cost $O(1)m^3$.

Denoting $h = \begin{pmatrix} s \\ r \end{pmatrix} \in \mathbf{R}^m \times \mathbf{R}^m$, we have

$$\begin{array}{c} h^T g = \mathrm{Tr}(A^{-1}E^T R X^{-1} E) - \mathrm{Tr}(A^{-1}E^T Y X^{-2} S E), \\ X = \mathrm{Diag}(x),\ Y = \mathrm{Diag}(y),\ A = E^T Y X^{-1} E,\ R = \mathrm{Diag}(r),\ S = \mathrm{Diag}(s). \end{array}$$

In other words, the y- and the x-components of g are formed by the diagonal entries of the matrices $(X^{-1}EA^{-1}E^T)$, $(-EA^{-1}E^T Y X^{-2})$, respectively; straightforward computation of these two matrices clearly costs $O(1)m^3$ arithmetic operations, and this is the cost of computing g.

Now let $h' = \begin{pmatrix} s' \\ r' \end{pmatrix} \in \mathbf{R}^m \times \mathbf{R}^m$. We have

$$\begin{array}{rcl} h^T H h' & = & -\mathrm{Tr}(A^{-1}E^T R' X^{-1} E A^{-1} E^T R X^{-1} E) \\ & & +\mathrm{Tr}(A^{-1}E^T Y S' X^{-2} E A^{-1} E^T R X^{-1} E) - \mathrm{Tr}(A^{-1}E^T R S' X^{-2} E) \\ & & +\mathrm{Tr}(A^{-1}E^T R' X^{-1} E A^{-1} E^T Y X^{-2} S E) \\ & & -\mathrm{Tr}(A^{-1}E^T Y X^{-2} S' E A^{-1} E^T Y X^{-2} S E) \\ & & -\mathrm{Tr}(A^{-1}E^T R' X^{-2} S E) + 2\mathrm{Tr}(A^{-1}E^T Y X^{-3} S' S E), \\ & & \qquad R' = \mathrm{Diag}(r'), \quad S' = \mathrm{Diag}(s'). \end{array}$$

It is easily seen that

$$h^T H h' \quad = \quad \sum_{i=1}^{7} \mathrm{Tr}(A_i \mathrm{Diag}(h') B_i^T \mathrm{Diag}(h)),$$

where A_i, B_i are $(2m)\times(2m)$ matrices independent of h, h' and computable at the cost $O(1)m^3$. Thus, H is the sum of the 7 matrices H_i, $i = 1, ..., 7$, of bilinear forms $\mathrm{Tr}(A_i \mathrm{Diag}(h') B_i^T \mathrm{Diag}(h))$ of h, h'. Clearly, j-th column of H_i is the diagonal of the matrix $A_i f_j f_j^T B_i^T$, where f_j are the standard basic orths of $\mathbf{R}^{2m}$. Given A_i, B_i, one can compute the diagonal of the matrix $A_i f_j f_j^T B_i^T = (A_i f_j)(B_i f_j)^T$ in $O(1)m$ operations; thus, after A_i, B_i are computed, the computation of a column in H costs only $O(1)m$ operations, and the computation of the entire matrix H costs $O(1)m^2$ operations. ■

References

[1] Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory*, SIAM Studies in Applied Mathematics, SIAM, Philadelphia, 1995

[2] Khachiyan, L.G., Tarasov, S.P., and Erlikh A.I., "The inscribed ellipsoid method", Dokl. Akad. Nauk SSSR, v. 298, 1988 (Engl. trans. Soviet Math. Dokl.)

[3] Khachiyan. L.G., and Todd, M.J., "On the complexity of approximating the maximal inscribed ellipsoid for a polytope", *Mathematical Programming* v. 61 (1993), pp. 137-159.

[4] Nemirovski, A. *Interior point polynomial methods in Convex Programming*, Lecture Notes, Spring Semester 1996, Faculty of Industrial Engineering and Management at Technion, Technion City, Haifa 32000, Israel.

[5] Nesterov, Yu., and Nemirovski, A. *Interior point polynomial algorithms in Convex Programming*, SIAM Publications, Philadelphia, USA, 1994.

[6] Nesterov, Yu. "Long-step strategies in interior point path-following methods" – Preprint, Departement d'Economie Commerciale et Industrielle, Universite de Geneve, 102 Bd. Carl Vogt, CH-1211 Geneve 4, Switzerland, September 1993.

[7] Nesterov, Yu., and Nemirovski, A. "Multi-parameter surfaces of analytic centers and long-step surface-following interior point methods", *Mathematics of Operations Research* v. 23 (1998), pp. 1-38.

[8] Rockafellar, R.T. *Convex Analysis*, Princeton University Press, 1970.

8 Appendices: Proofs

8.1 Proof of Proposition 2.2

(i): This statement immediately follows from the fact that the form $D^3 f(z)[dz_1, dz_2, dz_3]$ of dz_1, dz_2, dz_3 is 3-linear and symmetric, see [5], Appendix 1.

(ii): Let $dz = (dx, dy) \in E(z)$, $dz_x = (dx, 0)$, $dz_y = (0, dy)$, let $z' \in Z$, and let $h = z' - z$. Since $f''_{xx}(z)$ and $-f''_{yy}(z)$ are positive semidefinite, we have $dz_x, dz_y \in E(z)$; vice versa, the latter inclusions imply that $dz \in E(z)$. Thus, $E(z)$ is the direct sum of its projections $E_{f,x}(z)$ and $E_{f,y}(z)$ on $\mathbf{R}^n$, $\mathbf{R}^m$, respectively. Setting

$$\phi(t) = dz_x^T f''_{xx}(z + th) dz_x = D^2 f(z + th)[dz_x, dz_x], \; 0 \le t \le 1,$$

we get a continuously differentiable function on $[0,1]$. We have

$$|\phi'(t)| = |D^3 f(z+th)[dz_x, dz_x, h]| \leq 2\phi(t)\sqrt{h^T \mathrm{S}_f(z+th)h}$$

(we have used (8)), and since $\phi(0) = 0$, we have $\phi(t) = 0$, $0 \leq t \leq 1$. Thus, $dz_x \in E_{f,x}(z')$. "Symmetric" reasoning implies that $dz_y \in E_{f,y}(z')$, and since, as we just have seen, $E(z') = E_{f,x}(z') + E_{f,y}(z')$, we end up with $dz \in E(z')$. (ii.1) is proved.

By Remark 2.1, $f(\cdot, y)$ is s.-c. on X for every $y \in Y$, and therefore, due to [5], Theorem 2.1.1.(ii), $X = X + E_{f,x}$. By similar reasons $Y = Y + E_{f,y}$, whence $Z = Z + E_f$.

(iii) is an immediate consequence of (ii.2). ■

8.2 Proof of Proposition 2.3

(i): This is an immediate consequence of Remark 2.1 and Proposition 2.1.

(ii): (ii.a) immediately follows from (i).

(ii.b): For $0 \leq t \leq 1$, denote

$$\phi(t) = h^T \mathrm{S}_f(z+th)h = \phi_x(t) + \phi_y(t), \ \left\{ \begin{array}{lcl} \phi_x(t) & = & D^2 f(z+th)[(u,0),(u,0)], \\ \phi_y(t) & = & -D^2 f(z+th)[(0,v),(0,v)]. \end{array} \right.$$

We have (see (8))

$$\left. \begin{array}{lcl} |\phi_x'(t)| & = & |D^3(z+th)[(u,0),(u,0),h]| \leq 2\phi_x(t)\sqrt{\phi(t)} \\ |\phi_y'(t)| & = & |D^3(z+th)[(0,v),(0,v),h]| \leq 2\phi_y(t)\sqrt{\phi(t)} \end{array} \right\} \Rightarrow |\phi'(t)| \leq 2\phi^{3/2}(t), \ 0 \leq t \leq 1.$$

From the resulting differential inequality it immediately follows (note that $r = \phi^{1/2}(0)$)

$$z+h \in Z \Rightarrow \frac{r}{1+rt} \leq \sqrt{\phi(t)}, \ 0 \leq t \leq 1; \quad r < 1 \Rightarrow \sqrt{\phi(t)} \leq \frac{r}{1-rt}, \ 0 \leq t \leq 1. \tag{43}$$

Now let $dz = (dx, dy) \in \mathbf{R}^n \times \mathbf{R}^m$, and let

$$\psi(t) = dz^T \mathrm{S}_f(z+th)dz = \psi_x(t) + \psi_y(t), \ \left\{ \begin{array}{lcl} \psi_x(t) & = & D^2 f(z+th)[(dx,0),(dx,0)], \\ \psi_y(t) & = & -D^2 f(z+th)[(0,dy),(0,dy)], \end{array} \right.$$

$t \in [0,1]$. We have (see (8))

$$\left. \begin{array}{lcl} |\psi_x'(t)| & = & |D^3(z+th)[(dx,0),(dx,0),h]| \leq 2\psi_x(t)\sqrt{\phi(t)} \\ |\psi_y'(t)| & = & |D^3(z+th)[(0,dy),(0,dy),h]| \leq 2\psi_y(t)\sqrt{\phi(t)} \end{array} \right\} \Rightarrow |\psi'(t)| \leq 2\psi(t)\sqrt{\phi(t)}, \ 0 \leq t \leq 1.$$

From the resulting differential inequality and (43) it immediately follows that if $r < 1$, then

$$\psi(t) \geq (1-rt)^2 \psi(0), \ 0 \leq t \leq 1; \quad \psi(t) \leq (1-rt)^{-2}\psi(0), \ 0 \leq t \leq 1. \tag{44}$$

Since the resulting inequalities are valid for every dz, (ii.b) follows.

(ii.c): Let $r < 1$, let $h_1, h_2 \in \mathbf{R}^n \times \mathbf{R}^m$, and let

$$\psi_i(t) = h_i^T \mathrm{S}_f(z+th)h_i, \ i = 1,2; \quad \theta(t) = D^2 f(z+th)[h_1, h_2].$$

By (8), (43), (44) we have for $0 \leq t \leq 1$:

$$|\theta'(t)| = |D^3 f(z+th)[h_1,h_2,h]| \leq 2\sqrt{\psi_1(t)}\sqrt{\psi_2(t)}\sqrt{\phi(t)} \leq 2r(1-rt)^{-3}\sqrt{\psi_1(0)\psi_2(0)},$$

whence for $r < 1$ one has

$$\forall (h_1, h_2): \quad |h_1^T[f''(z+h) - f''(z)]h_2| \le \left[\frac{1}{(1-r)^2} - 1\right] \sqrt{h_1^T \mathrm{S}_f(z) h_1} \sqrt{h_2^T \mathrm{S}_f(z) h_2}, \tag{45}$$

as claimed in (ii.c).

(ii.d): Let $r < 1$, and let $h' \in \mathbf{R}^n \times \mathbf{R}^m$. We have

$$\begin{array}{rcl} |(h')^T[f'(z+h) - f'(z) - f''(z)h]| & = & |\int\limits_0^1 (h')^T[f''(z+th) - f''(z)]h dt| \\ & \underbrace{\le}_{(*)} & \int\limits_0^1 \left[\frac{1}{(1-tr)^2} - 1\right] r\sqrt{(h')^T \mathrm{S}_f(z) h'} dt = \frac{r^2}{1-r} \sqrt{(h')^T \mathrm{S}_f(z) h'} \end{array}$$

(($*$) is given by (45)), and (ii.d) follows. ■

8.3 Proof of Proposition 2.4

"If" part: in the case in question the left hand side expression in (10) is

$$|D^3 f(x,y)[Jh_1, h_2, h_3]|, \quad J = \begin{pmatrix} I_n & 0 \\ 0 & -I_m \end{pmatrix},$$

(from now on, I_k is the $k \times k$ unit matrix), while $\widehat{A}(z)$ is exactly $\mathrm{S}_f(z)$. Thus, (10) implies that

$$|D^3 f(x,y)[h,h,h]| = |D^3 f(x,y)[J(Jh), h, h]| \le 2[h^T \mathrm{S}_f(z) h][(Jh)^T \mathrm{S}_f(z) Jh]^{1/2} = 2(h^T \mathrm{S}_f(z) h)^{3/2}$$

(note that $J^T \mathrm{S}_f(z) J = \mathrm{S}_f(z)$), as required in (7). It remains to verify Definition 2.2.(i). Let $y \in Y$ and let $\{x_i \in X\}$ converge to a boundary point of X, so that the points $z_i = (x_i, y)$ converge to a boundary point of Z. According to [5], Proposition 7.2.1, the ellipsoids $W_i = \{z \mid (z - z_i)^T \mathrm{S}_f(z_i)(z - z_i) < 1\}$ (recall that $\mathrm{S}_f(\cdot) = \widehat{A}(\cdot)$) are contained in Z. Now let $x_0 \in X$ be a once for ever fixed point, and let $e_i = x_i - x_0$. We claim that the quantities $\delta_i = \sqrt{e_i^T f''_{xx}(z_i) e_i}$ tend to $+\infty$. Indeed, since

$$(\widehat{x}_i, y) = (x_i + (1+\delta_i)^{-1} e_i, y) \in W_i \subset Z,$$

in the case of bounded $\{\delta_i\}$ the limit of x_i, i.e., a boundary point of X, would be a convex combination, with positive weights, of a limiting point of $\widehat{x}_i \in X$ and x_0, which is impossible.

We now have

$$f(x_i, y) = f(x_0, y) + e_i^T f'_x(x_0, y) + \int\limits_0^1 t g_i(t) dt, \;\; g_i(t) = e_i^T f''_{xx}(x_0 + (1-t)e_i, y) e_i. \tag{46}$$

Now note that in view of self-concordance of $A(\cdot)$ we clearly have

$$g_i'(t) \ge -2 g_i^{3/2}(t), \;\; 0 \le t \le 1,$$

whence for $0 \le t \le 1$

$$g_i^{-1/2}(t) \le g_i^{-1/2}(0) + t \Rightarrow g_i(t) \ge \frac{g_i(0)}{(1 + t\sqrt{g_i(0)})^2} = \frac{\delta_i^2}{(1 + t\delta_i)^2}.$$

Consequently, the integral in (46) can be bounded from below by

$$\int\limits_0^1 \frac{t\delta_i^2}{(1+t\delta_i)^2} dt = \int\limits_0^{\delta_i} \frac{s}{(1+s)^2} ds,$$

and since $\delta_i \to \infty$ as $i \to \infty$, we see that the integral in (46) tends to $+\infty$ as i grows. Since the remaining terms in the expression for $f(x_i, y)$ in (46) are bounded uniformly in i, we conclude that $f(x_i, y) \to \infty$, as required in Definition 2.2.(i). By symmetric reasons, $f(x, y_i) \to -\infty$ whenever $x \in X$ and a sequence $\{y_i \in Y\}$ converges to a boundary point of Y. The "if" part is proved.

"Only if" part: Assuming f s.-c.c.-c. and taking into account Proposition 2.2.(i), we get

$$|h_2^T \nabla_z^2 (h_1^T A(z)) h_3| = |D^3 f(z)[Jh_1, h_2, h_3]| \leq 2(h_1^T J^T \mathrm{S}_f(z) J h_1)^{1/2} (h_2^T \mathrm{S}_f(z) h_2)^{1/2} (h_3^T \mathrm{S}_f(z) h_3)^{1/2},$$

as required in (10) (recall that $\widehat{A}(z) = \mathrm{S}_f(z)$ and $J^T \mathrm{S}_f(z) J = \mathrm{S}_f(z)$). Now, if a sequence $\{z_i = (x_i, y_i) \in Z\}$ converges to a boundary point of Z, then the sequence of matrices $\widehat{A}(z_i) = \mathrm{S}_f(z_i)$ is unbounded due to Proposition 2.3.(i). Thus, $A(\cdot)$ is a strongly self-concordant monotone operator. ■

8.4 Proof of Proposition 2.5

If f possesses a saddle point on Z, then, of course, (*) takes place. In the case in question the saddle point is unique in view of the nondegeneracy of f.

It remains to verify that if f satisfies (*), then f possesses a saddle point on Z. Assume that f satisfies (*), and let x_0, y_0 be the corresponding points. Setting

$$\phi(x) = \sup_{y \in Y} f(x, y),$$

we get a lower semicontinuous convex function on X (taking values in $\mathbf{R} \cup \{+\infty\}$) which is finite at x_0. We claim that the level set

$$X^-(a) = \{x \in X \mid \phi(x) \leq a\}$$

is compact for every $a \in \mathbf{R}$. Indeed, the set clearly is contained in the set

$$\widehat{X}(a) = \{x \in X \mid f(x, y_0) \leq a\},$$

and since $f(\cdot, y_0)$ is a s.-c. nondegenerate below bounded function on X, $\widehat{X}$ is a compact set (this is an immediate corollary of [5], Proposition 2.2.3). Since ϕ is lower semicontinuous, $X^-(a)$ is a closed subset of $\widehat{X}(a)$ and is therefore compact.

By "symmetric reasons", the function

$$\psi(y) = \inf_{x \in X} f(x, y)$$

is an upper semicontinuous function on Y which is finite at y_0 and has compact level sets

$$Y^+(a) = \{y \in Y \mid \psi(y) \geq a\}.$$

Since ϕ has compact level sets and is finite at least at one point, ϕ attains its minimum on X at a convex compact set X^*, and by similar reasons ψ attains its maximum on Y at a convex compact set Y^*. In order to prove that f possesses a saddle point on Z, it suffices to demonstrate that the inequality in the following chain

$$a_* \equiv \max_{y \in Y} \psi(y) \leq \min_{x \in X} \phi(x) \equiv a^*$$

is in fact equality. Assume, on contrary, that $a_* < a^*$, and let $a \in (a_*, a^*)$. Denoting $X(y) = \{x \in X \mid f(x, y) \leq a\}$, we conclude that $\bigcap_{y \in Y} X(y) = \emptyset$. Since $f(\cdot, y)$ is a nondegenerate s.-c. function on X for every $y \in Y$, the sets $X(y)$ are closed; as we just have seen, $X(y_0)$ is compact. Consequently, $\bigcap_{y \in Y} X(y) = \emptyset$ implies that $\bigcap_{y \in Y'} X(y) = \emptyset$ for some finite subset $Y' \in Y$. In other words, $\max_{y \in Y'} f(x, y) \geq a$ for every $x \in X$, and therefore a convex combination $\sum_{y \in Y'} \lambda_y f(x, y)$ is $\geq a$ everywhere on X. But $\sum_{y \in Y'} \lambda_y f(x, y) \leq f(x, y^*)$, $y^* = \sum_{y \in Y'} \lambda_y y$, and we see that $\inf_{x \in X} f(x, y^*) \geq a > a_*$, which is a contradiction. ■

8.5 Proof of Proposition 3.1

(i): Since $f(\cdot, y)$ is a nondegenerate s.-c. function on X for every $y \in Y$, by ([5], Theorem 2.4.1) the set

$$X^*(f, y) = \{f'_x(x, y) \mid x \in X\}$$

is open, nonempty and convex and is exactly the set of those ξ for which the function $\xi^T x - f(x, y)$ is above bounded on X. From these observations it immediately follows that $X^*(f) = \bigcup_{y \in Y} X^*(f, y)$; in particular, the set $X^*(f)$ is open. Let us prove that this set is convex. Indeed, assume that $\xi_1, \xi_2 \in X^*(f)$, so that for some $y_1, y_2 \in Y$ the functions $\xi_i^T x - f(x, y_i)$ are above bounded on X, $i = 1, 2$. Whenever $\lambda \in [0, 1]$, we have

$$\lambda[\xi_1^T x - f(x, y_1)] + (1 - \lambda)[\xi_2^T x - f(x, y_2)] \geq [\lambda\xi_1 + (1 - \lambda)\xi_2]^T x - f(x, \lambda y_1 + (1 - \lambda)y_2),$$

so that $\lambda\xi_1 + (1-\lambda)\xi_2 \in X^*(x, \lambda y_1 + (1-\lambda)y_2) \subset X^*(f)$, and consequently $X^*(f)$ is convex. Similar arguments demonstrate that $Y^*(f)$ also is open and convex.

(ii): Whenever $(\xi, \eta) \in Z^*(f)$, the function $f_{\xi,\eta}(x, y)$ (which is s.-c.c.-c. on Z by Proposition 4.1.(i) and is nondegenerate together with f) possesses property (*) from Proposition 2.5, and by this proposition it possesses a unique saddle point on Z. Vice versa, if (ξ, η) is such that $f_{\xi,\eta}(z)$ possesses a saddle point (x^*, y^*) on Z, the function $\xi^T x - f(x, y^*)$ is above bounded on X, and the function $\eta^T y - f(x^*, y)$ is below bounded on Y, so that $(\xi, \eta) \in Z^*(f)$.

(iii): If $z_0 \in Z$, then z_0 clearly is the saddle point of the function $f_{f'(z_0)}(\cdot, \cdot)$ on Z, so that $f'(z_0) \in Z^*(f)$ by (ii). Vice versa, if $(\xi, \eta) \in Z^*(f)$, then the function $f_{\xi,\eta}(z)$ possesses a saddle point z_0 on Z by (ii); we clearly have $(\xi, \eta) = f'(z_0)$. Thus, $z \mapsto f'(z)$ maps Z onto $Z^*(f)$. This mapping is a one-to-one mapping, since the inverse image of a point $(\xi, \eta) \in Z^*(f)$ is exactly the saddle set of the function $f_{\xi,\eta}(z)$ on Z, and the latter set, being nonempty, is a singleton by Proposition 2.5. It remains to prove that the mapping and its inverse are twice continuously differentiable. To this end it suffices to verify that $f''(z)$ is nonsingular for every $z \in Z$. The latter fact is evident: since f is convex-concave and nondegenerate, we have

$$f''(z) = \begin{pmatrix} A & Q \\ Q^T & -B \end{pmatrix}$$

with positive definite symmetric A, B, and a matrix of this type always is nonsingular. Indeed, assuming

$$Au + Qv = 0, \quad Q^T u - Bv = 0,$$

multiplying the first equation by u^T, the second by $-v^T$ and adding the results, we get $u^T Au + v^T Bv = 0$, whence $u = 0$ and $v = 0$; consequently, $\mathrm{Ker} f''(z) = \{0\}$.

(iv): First let us verify that f_* is convex in ξ and concave in η on $Z^*(f)$. Indeed,

$$f_*(\xi, \eta) = \inf_{y \in Y} \sup_{x \in X} [\xi^T x + \eta^T y - f(x, y)] = \inf_{y \in Y} [\eta^T y + \sup_{x \in X} [\xi^T x - f(x, y)]],$$

so that $f_*(\xi, \eta)$ is the lower bound of a family of affine functions of η and therefore it is concave in η. Convexity in ξ follows, via similar arguments, from the representation

$$f_*(\xi, \eta) = \sup_{x \in X} \inf_{y \in Y} [\xi^T x + \eta^T y - f(x, y)]$$

coming from the fact that $f_{\xi,\eta}(x, y)$ possesses a saddle point on Z when $(\xi, \eta) \in Z^*(f)$, see (ii).

Now let us prove that f_* is differentiable and that the mapping $\zeta \to f'_*(\zeta)$ is inverse to $f'(z)$. Indeed, let $\zeta = (\xi, \eta) \in Z^*(f)$, so that the function $f_\zeta(x, y)$ possesses a unique saddle point $z_\zeta = (x_\zeta, y_\zeta)$ on Z ((ii) and Proposition 2.5). Note that by evident reasons $\zeta = f'(z_\zeta)$. We claim that x_ζ is a subgradient of $f_*(\cdot, \eta)$ at the point ξ, and y_ζ is the super-gradient of $f_*(\xi,\cdot)$ at the point η. By symmetry, it suffices to prove the first claim, which is evident:

$$\begin{array}{rcl}
f_*(\xi',\eta) & = & \sup\limits_{x\in X}\inf\limits_{y\in Y}[(\xi')^Tx+\eta^Ty-f(x,y)]\\
& \geq & \inf\limits_{y\in Y}[(\xi')^Tx_\zeta+\eta^Ty-f(x_\zeta,y)]=(\xi')^Tx_\zeta+\inf\limits_{y\in Y}[\eta^Ty-f(x_\zeta,y)]\\
& = & (\xi'-\xi)^Tx_\zeta+\inf\limits_{y\in Y}[\xi^Tx_\zeta+\eta^Ty-f(x_\zeta,y)]\\
& = & (\xi'-\xi)^Tx_\zeta+[\xi^Tx_\zeta+\eta^Ty_\zeta-f(x_\zeta,y_\zeta)] \qquad \text{[since } (x_\zeta,y_\zeta) \text{ is a saddle point of } f_\zeta(\cdot,\cdot)]\\
& = & (\xi'-\xi)^Tx_\zeta+\inf\limits_{y\in Y}\sup\limits_{x\in X}[\xi^Tx+\eta^Ty-f(x,y)] \qquad \text{[by the same reasons]}\\
& = & (\xi'-\xi)^Tx_\zeta+f_*(\xi,\eta).
\end{array}$$

Since, on one hand, the mapping $\zeta\to z_\zeta$ is inverse to the mapping $z\to f'(z)$ and is therefore twice continuously differentiable by (iii), and, on the other hand, the components of this mapping are partial sub- and supergradients of f_*, the function f_* is C^3 smooth on $Z^*(f)$, and its gradient mapping is inverse to the one of f. In particular,

$$\{\zeta=f'(z)\}\Leftrightarrow\{z=f_*'(\zeta)\}\Rightarrow\{f_*''(\zeta)=[f''(z)]^{-1}\}. \tag{47}$$

It remains to prove that f_* is s.-c.c.-c. on $Z^*(f)$. Let us first prove the corresponding differential inequality. We have

$$\begin{array}{rcl}
\multicolumn{3}{l}{d\zeta^Tf_*''(\zeta)d\zeta=d\zeta^T[f''(f_*'(\zeta))]^{-1}d\zeta \quad \Rightarrow}\\
D^3f_*(\zeta)[d\zeta,d\zeta,d\zeta] & = & -D^3f(f_*'(\zeta))[f_*''(\zeta)d\zeta,f_*''(\zeta)d\zeta,f_*''(\zeta)d\zeta]=-D^3f(z)[dz,dz,dz],\\
& & [z=f_*'(\zeta),\ dz=f_*''(\zeta)d\zeta=[f''(z)]^{-1}d\zeta.]
\end{array}$$

It follows that

$$|D^3f_*(\zeta)[d\zeta,d\zeta,d\zeta]|=|D^3f(z)[dz,dz,dz]|\leq 2(d\zeta^T[f''(z)]^{-1}\mathrm{S}_f(z)[f''(z)]^{-1}d\zeta)^{3/2}. \tag{48}$$

Now let $J=\begin{pmatrix}I_n & 0\\ 0 & -I_m\end{pmatrix}$. We have

$$\mathrm{S}_f(z)=\frac{1}{2}\left([f''(z)]J+J[f''(z)]\right),$$

whence

$$\begin{array}{rcl}
[f''(z)]^{-1}\mathrm{S}_f(z)[f''(z)]^{-1} & = & \frac{1}{2}[f''(z)]^{-1}\left([f''(z)]J+J[f''(z)]\right)[f''(z)]^{-1}=\frac{1}{2}\left(J[f''(z)]^{-1}+[f''(z)]^{-1}J\right)\\
& = & \frac{1}{2}\left(J[f_*''(\zeta)]+[f_*''(\zeta)]J\right) \qquad \text{[we have used (47)]}\\
& = & \mathrm{S}_{f_*}(\zeta).
\end{array} \tag{49}$$

Consequently, (48) becomes

$$|D^3f_*(\zeta)[d\zeta,d\zeta,d\zeta]| \quad\leq\quad 2(d\zeta^T\mathrm{S}_{f_*}(\zeta)d\zeta)^{3/2},$$

which is exactly the differential inequality required in Definition 2.2.

It remains to prove that $f_*(\cdot,\eta)$ is a barrier for $X^*(f)$ for every $\eta\in Y^*(f)$, and that $-f_*(\xi,\cdot)$ is a barrier for $Y^*(f)$ for every $\xi\in X^*(f)$. By symmetry, it suffices to prove the first of these statements. Let us fix $\eta\in Y^*(f)$, and let a sequence $\{\xi_i\in X^*(f)\}$ converge to a point ξ and be such that $f_*(\xi_i,\eta)\leq a$ for some $a\in\mathbf{R}$ and all i. We should prove that under these assumptions $\xi\in X^*(f)$. Indeed, we have

$$\begin{array}{rcl}
-a\leq -f_*(\xi_i,\eta) & = & \inf\limits_{x\in X}\sup\limits_{y\in Y}[f(x,y)-\xi_i^Tx-\eta^Ty]\underbrace{=}_{(a)}=\sup\limits_{y\in Y}\inf\limits_{x\in X}[f(x,y)-\xi_i^Tx-\eta^Ty]\\
& \underbrace{=}_{(b)} & \min\limits_{x\in X}[f(x,y_i)-\xi_i^Tx-\eta^Ty_i]\ [y_i\in Y].
\end{array} \tag{50}$$

with (a), (b) given by Proposition 3.1.(iv,ii), respectively. Since $\eta \in Y^*(f)$, there exists $x_0 \in X$ such that the function

$$g(y) = f(x_0, y) - \eta^T y - \xi^T x_0$$

is above bounded on Y. Since $(-g)$ is s.-c. convex and nondegenerate on Y, the level set

$$Y^+ = \{y \in Y \mid f(x_0, y) - \eta^T y - \xi^T x_0 \geq -a - 1\}$$

is compact (we already have mentioned this fact). We claim that all points y_i, starting from some i_0, belong to Y^+. Indeed, whenever $y_i \notin Y^+$, we have by (50)

$$\begin{array}{rcl} -a & \leq & \min\limits_{x \in X}[f(x, y_i) - \xi_i^T x - \eta^T y_i] \leq f(x_0, y_i) - \xi_i^T x_0 - \eta^T y_i \\ & = & [f(x_0, y_i) - \xi^T x_0 - \eta^T y_i] + (\xi - \xi_i)^T x_0 < -a - 1 + (\xi - \xi_i)^T x_0, \end{array}$$

and the resulting inequality, due to $\xi_i \to \xi$, can be valid for finitely many i's only.

Since Y^+ is compact and contains all y_i except finitely many of them, we can, passing to a subsequence, assume that $y_i \in Y^+$ and y_i converge to $y_0 \in Y^+ \subset Y$. We claim that the function $f(x, y_0) - \xi^T x$ is below bounded on X, which yields the desired contradiction. Indeed, by (50) we have for every $x \in X$ and all i

$$-a \leq f(x, y_i) - \xi_i^T x - \eta^T y_i,$$

and passing to limit, we get

$$-a \leq f(x, y_0) - \xi^T x - \eta^T y_0,$$

so that $f(x, y_0) - \xi^T x \geq -a + \eta^T y_0$. Thus, $\xi \in X^*(f)$. ■

8.6 Proof of Theorem 3.1

The fact that f_* is s.-c.c.-c. is given by Proposition 3.1.(iv). The equivalence in the premise of (11) is stated by (47). Under this premise, the validity of (a), (b) is given by (47), (49), respectively. Now, under the premise of (11) z clearly is the saddle point of the function $f(z') - \zeta^T z'$ of $z' \in Z$, whence

$$f_*(\zeta) = -[f(z) - \zeta^T z],$$

as required in (c). Nondegeneracy of f_* follows from (b). It remains to prove that the Legendre transformation of f_* is exactly f. From Proposition 3.1.(iii,iv) it follows that the domain of the Legendre transformation of f_* is exactly Z, so that all we need is to prove that

$$(f_*)_*(x_0, y_0) = f(x_0, y_0)$$

for every $(x_0, y_0) \in Z$. This is immediate: setting $\xi_0 = f'_x(x_0, y_0)$, $\zeta_0 = f'_y(x_0, y_0)$, we clearly have

$$f_*(\xi_0, \eta_0) = \xi_0^T x_0 + \eta_0^T y_0 - f(x_0, y_0),$$

and by Proposition 3.1.(iv)

$$(f_*)'_\xi(\xi_0, \eta_0) = x_0, (f_*)'_\eta(\xi_0, \eta_0) = y_0,$$

whence

$$(f_*)_*(x_0, y_0) = x_0^T \xi_0 + y_0^T \eta_0 - f_*(\xi_0, \eta_0) = f(x_0, y_0).$$ ■.

8.7 Proof of Proposition 4.2

By symmetry, it suffices to prove (i).

"If" part and derivation of (12): Assume that $0 \in X^*$ and the function $\eta^T y - f_*(0,\eta)$ is below bounded on Y^*. Since the latter function is s.-c. on Y^* (Theorem 3.1 and Remark 2.1), it attains its minimum on Y^* at some point η_*. Setting $x^* = (f_*)'_\xi(0,\eta_*)$, we have $f'_*(0,\eta_*) = (x^*, y)$, whence by Theorem 3.1

$$f(x^*, y) = 0^T x^* + \eta_*^T y - f_*(0,\eta_*),$$

while by Theorem 3.1 and Proposition 3.1 for every $x \in X$ one has

$$f(x,y) = \sup_{\xi \in X^*} \inf_{\eta \in Y^*} [\xi^T x + \eta^T y - f_*(\xi,\eta)] \geq \inf_{\eta \in Y^*} [\eta^T y - f_*(0,\eta)] = \eta_*^T y - f_*(0,\eta_*),$$

and (12) follows.

"Only if" part: assume that $f(\cdot, y)$ is below bounded on X. Then $0 \in X^*$ by the definition of X^*, and since $f(\cdot, y)$ is s.-c. on X (Remark 2.1), the function attains its minimum over X at some point x^*. Setting $\eta_* = f'_y(x^*, y)$, we get $f'(x^*, y) = (0, \eta_*)$, whence by Theorem 3.1 $(0, \eta_*)$ is the saddle point of the function

$$f_*(\xi,\eta) - \xi^T x^* - \eta^T y,$$

so that the function $f_*(0,\eta) - \eta^T y$ is above bounded in $\eta \in Y^*$. Thus, the function $\eta^T y - f_*(0,\eta)$ is below bounded. ∎

8.8 Proof of Proposition 4.3

By symmetry, it suffices to prove the first statement. Convexity of X^+ is evident. To prove openness, note that if $x \in X^+$, then $-f(x,\cdot)$ is a nondegenerate s.-c. below bounded convex function on Y, so that $f(x,\cdot)$ attains its maximum on Y at a unique point $y(x)$, and the Newton decrement of $-f(x,\cdot)$ at $y(x)$ is zero. Consequently, there exists a neighbourhood U of x such that the Newton decrements, taken at $y(x)$, of the s.-c. on Y functions $-f(x', \cdot)$, $x' \in U$, are < 1, and therefore the indicated functions are below bounded on Y ([5], Theorem 2.2.2.(i)). Thus, $U \subset X^+$, whence X^+ is open.

As we have seen, for $x \in X^+$ the function $f(x,\cdot)$ attains its maximum on Y at a unique point $y(x)$ given by

$$f'_y(x, y(x)) = 0. \tag{51}$$

Since f''_{yy} is nondegenerate, by the Implicit Function Theorem (51) defines a C^2 function $y(\cdot)$ on X^+; consequently, the function

$$\phi(x) = f(x, y(x)) \equiv \max_{y \in Y} f(x,y) : X^+ \to \mathbf{R}$$

is C^2 smooth (and clearly convex). Since by evident reasons

$$D\phi(x)[dx] = dx^T f'_x(x, y(x)),$$

we see that in fact ϕ is C^3-smooth.

Let us prove that ϕ is s.-c. on X^+. The barrier property is evident: if $x_i \in X^+$ and $x_i \to x \notin X^+$, then either $x \notin X$ – and then $\phi(x_i) \geq f(x_i, y) \to \infty$, $y \in Y$ being fixed, – or $x \in X$. In the latter case the sequence of functions $\{f(x_i, \cdot)\}$ does not contain a uniformly above bounded on Y subsequence – otherwise $f(x,\cdot)$ were above bounded on Y, which is not the case – and therefore $\phi(x_i) \to \infty$ as $i \to \infty$.

It remains to verify the differential inequality (2) responsible for self-concordance. Let us fix $x \in X^+$ and a direction dx in the x-space, and let

$$\begin{array}{rclrclrcl} y &=& y(x), & dy &=& Dy(x)[dx], & d^2y &=& D^2y(x)[dx,dx], \\ z &=& (x,y), & dz &=& (dx,dy), & f''(z) &=& \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}. \end{array}$$

We have for every $h \in \mathbf{R}^m$ and $u \in X^+$:

$$
\begin{array}{lrcl}
 & Df(u,y(u))[(0,h)] & = & 0 \\
\Rightarrow (a) & D^2 f(u,y(u))[(dx, Dy(u)[dx]),(0,h)] & = & 0 \\
\Rightarrow (b) & D^3 f(u,y(u))[(dx, Dy(u)[dx]),(dx, Dy(u)[dx]),(0,h)] & & \\
 & +D^2 f(u,y(u))[(0, D^2 y(u)[dx,dx]),(0,h)] & = & 0.
\end{array}
\tag{52}
$$

Note that from (52.a) we get

$$
dy = C^{-1} B dx, \tag{53}
$$

while from (52.b) we get

$$
\forall h \in \mathbf{R}^m : \quad h^T C d^2 y = D^3 f(z)[dz, dz, (0,h)]. \tag{54}
$$

Now,

$$
\begin{array}{rrcl}
 & D\phi(u)[dx] & = & Df(u,y(u))[(dx, Dy(u)[dx])] = Df(u,y(u))[(dx,0)] \\
 & & & \text{[since } Df(u,y(u))[(0,h)] \equiv 0] \\
\Rightarrow & D^2\phi(u)[dx,dx] & = & D^2 f(u,y(u))[(dx, Dy(u)[dx]),(dx,0)] \\
\Rightarrow & D^3\phi(u)[dx,dx,dx] & = & D^3 f(u,y(u))[(dx, Dy(u)[dx]),(dx, Dy(u)[dx]),(dx,0)] \\
 & & & +D^2 f(u,y(u))[(0, D^2 y(u)[dx,dx]),(dx,0)],
\end{array}
$$

so that

$$
\begin{array}{rcl}
D^2\phi(x)[dx,dx] & = & D^2 f(z)[(dx,dy),(dx,0)] = dx^T A dx + dx^T B^T dy \\
 & \underbrace{=}_{(a)} & dx^T A dx + dy^T C dy = dz^T \mathrm{S}_f(z) dz; \\
D^3\phi(x)[dx,dx,dx] & = & D^3 f(z)[dz,dz,(dx,0)] + D^2 f(z)[(0,d^2y),(dx,0)] \\
 & = & D^3 f(z)[dz,dz,(dx,0)] + dx^T B^T d^2 y \\
 & = & D^3 f(z)[dz,dz,(dx,0)] + (C^{-1}Bdx)^T C d^2 y \\
 & \underbrace{=}_{(b)} & D^3 f(z)[dz,dz,(dx,0)] + D^3 f(z)[dz,dz,(0,C^{-1}Bdx)] \\
 & \underbrace{=}_{(c)} & D^3 f(z)[dz,dz,(dx,0)] + D^3 f(z)[dz,dz,(0,dy)] \\
 & = & D^3 f(z)[dz,dz,dz]
\end{array}
\tag{55}
$$

((a), (c) are given by (53), (b) is given by (54)), and since f is s.-c.c-c., we end up with

$$
|D^3\phi(x)[dx,dx,dx]| \leq 2\left(D^2\phi(x)[dx,dx]\right)^{3/2}. \qquad \blacksquare
$$

8.9 Proof of Proposition 4.4

First of all, the function ϕ is well-defined: since Y is bounded, $f(x,\cdot)$ is above bounded on Y whenever $x \in X$, and consequently the function $-f(x,(u,\cdot))$ is below bounded on its domain whenever $x \in X, u \in U$. Since the latter function is s.-c. ([5], Proposition 2.1.1.(i)), it attains its minimum.

Now let us prove that ϕ is s.-c.c.-c. Convexity-concavity of ϕ is evident. As we have seen, the maximum in the right hand side of (13) is achieved, and the maximizer $v(x,u)$ is unique, since f''_{vv} is negative definite. By the latter reason and the Implicit Function Theorem the function $v(x,u)$ is twice continuously differentiable on $X \times U$, and since

$$
D\phi(x,u)[(dx,du)] = Df(x,(u,v(x,u)))[(dx,(du,0))]
$$

due to $f'_v(x,(u,v(x,u))) = 0$, ϕ is C^3 smooth.

Let us verify the barrier properties required by Definition 2.2.(i). Let $u \in U$, and let $\{x_i \in X\}$ be a sequence converging to a boundary point of X. Then $\phi(x_i,u) \geq f(x_i,(u,v))$, where v is such that $(u,v) \in Y$, and consequently $\phi(x_i,u) \to \infty$, $i \to \infty$. Now let $x \in X$, and let $\{u_i \in U\}$ be a sequence converging to a boundary point of U. Passing to a subsequence we may suppose that $v_i = v(x,u_i)$ has a limit (Y is bounded!), and of course the limit of the sequence (u_i, v_i) is a boundary point of Y. Consequently, $\phi(x, u_i) = f(x, (u_i, v_i)) \to -\infty$, $i \to \infty$, as required.

In order to verify the differential inequality required in Definition 2.2(2), let us first note that the computations which led to (55) do not use the s.-c.c.-c. property of the underlying function and in fact establish the following

Lemma 8.1 *Let $g(p, q)$ be a C^3 function defined in a neighbourhood of a point $(\widehat{p}, \widehat{q})$ such that $g'_q(\widehat{p}, \widehat{q}) = 0$, and let $g''(\widehat{p}, \widehat{q}) = \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}$ with nonsingular $C = -g''_{qq}(\widehat{p}, \widehat{q})$. Then the equation $g'_q(p, q) = 0$ in a neighbourhood of $\widehat{p}$ has a C^2-smooth solution $q(p)$, $q(\widehat{p}) = \widehat{q}$, the function $h(p) = g(p, q(p))$ is C^3-smooth in a neighbourhood of $\widehat{p}$, and for every vector dp in the p-space we have*

$$D^2 h(\widehat{p})[dp, dp] = dr^T S dr, \quad D^3 h(\widehat{p})[dp, dp, dp] = D^3 g(\widehat{p}, \widehat{q})[dr, dr, dr],$$

where

$$dr \ = \ (dp, dq), \quad dq \ = \ C^{-1} B dp, \quad S \ = \ \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}.$$

In particular,

$$h''(\widehat{p}) = A + B^T C^{-1} B. \tag{56}$$

Let $\widehat{w} \equiv (\widehat{x}, \widehat{u}) \in X \times U$, and let

$$g((x, u), v) = f(x, (u, v)).$$

Let also $\widehat{v} = v(\widehat{x}, \widehat{u})$, $\widehat{z} = (\widehat{x}, (\widehat{u}, \widehat{v}))$, and let

$$f''(\widehat{z}) = \left(\begin{array}{c|c|c} Q & B_u^T & B_v^T \\ \hline B_u & -P & D^T \\ \hline B_v & D & -R \end{array}\right), \quad Q = f''_{xx}(\widehat{z}), -P = f''_{uu}(\widehat{z}), -R = f''_{vv}(\widehat{z}).$$

Applying Lemma 8.1 to g $(p = (x, u), q = v)$ with $\widehat{p} = (\widehat{x}, \widehat{u})$, $\widehat{q} = v(\widehat{x}, \widehat{u})$ and $dp = (dx, du)$ being a direction in the (x, u)-space, we get

$$\begin{array}{rcl} D^2 \phi(\widehat{w})[dp, dp] & = & dz^T \begin{pmatrix} Q & B_u^T & 0 \\ B_u & -P & 0 \\ 0 & 0 & R \end{pmatrix} dz, \quad \left[dz = \begin{pmatrix} dx \\ du \\ dv \end{pmatrix}, dv = R^{-1}[B_v dx + D du] \right]; \\ D^3 \phi(\widehat{w})[dp, dp, dp] & = & D^3 f(\widehat{z})[dz, dz, dz]. \end{array} \tag{57}$$

From (57) we get

$$\begin{array}{rcl} \phi''_{xx}(\widehat{w})[dx, dx] & = & dx^T Q dx + [R^{-1} B_v dx]^T R [R^{-1} B_v dx] = dx^T Q dx + dx^T B_v^T R^{-1} B_v dx; \\ -\phi''_{uu}(\widehat{w})[du, du] & = & du^T P du - du^T [R^{-1} D du]^T R [R^{-1} D du] = du^T P du - du^T D^T R^{-1} D du, \end{array}$$

whence

$$dp^T \mathrm{S}_\phi(\widehat{w}) dp = dx^T Q dx + dx^T B_v^T R^{-1} B_v dx + du^T P du - du^T D^T R^{-1} D du. \tag{58}$$

We have (see (57), (58))

$$\begin{array}{rcl} dz^T \mathrm{S}_f(\widehat{z}) dz & = & dx^T Q dx + du^T P du + dv^T R dv - 2 du^T D^T dv \\ & = & dx^T Q dx + du^T P du + [R^{-1}(B_v dx + D du)]^T R [R^{-1}(B_v dx + D du)] \\ & & -2 du^T D^T [R^{-1}(B_v dx + D du)] \\ & = & dx^T Q dx + du^T P du + dx^T B_v^T R^{-1} B_v dx + 2 dx^T B_v^T R^{-1} D du \\ & & + du^T D^T R^{-1} D du - 2 du^T D^T R^{-1} B_v dx - 2 du^T D^T R^{-1} D du \\ & = & dx^T Q dx + dx^T B_v^T R^{-1} B_v dx + du^T P du - du^T D^T R^{-1} D du = dp^T \mathrm{S}_\phi(\widehat{w}) dp. \end{array} \tag{59}$$

Thus,

$$\begin{array}{rcll} |D^3\phi(\widehat{w})[dp,dp,dp]| & = & |D^3 f(\widehat{z})[dz,dz,dz]| & \\ & \leq & 2(dz^T \mathrm{S}_f(\widehat{z})dz)^{3/2} & \text{[since } f \text{ is s.-c.c.-c.]} \\ & = & 2(dp^T \mathrm{S}_\phi(\widehat{w})dp)^{3/2} & \text{[see (59)]}, \end{array}$$

as required in (7). The fact that ϕ is nondegenerate is readily given by (59). ■

8.10 Proof of Proposition 5.1

(i), (ii) are immediate consequnces of (11.b) and Proposition 4.3, respectively. Let us prove (iii).

(iii.1)⇔(iii.2): If $z \in K(f)$, then $0 \in X^*$ and $0 \in Y^*$, so that $(0,0) \in Z^*$, and f possesses a saddle point on Z in view of Proposition 3.1. Vice versa, if f possesses a saddle point $z^* = (x^*, y^*)$ on Z, then clearly $z^* \in K(f)$ and therefore $K(f)$ is nonempty.
(iii.2)⇔(iii.3): This equivalence is an immediate consequence of Proposition 3.1.
(iii.2)⇒(iii.4): This is evident, since $\omega(f, z^*) = 0$ at a saddle point z^* of f.
(iii.4)⇒(iii.3): Given $z \in Z$ with $\omega(f,z) < 1$, consider the point $\zeta \equiv f'(z) = (\xi, \eta)$. By (20), $0 \in W_\xi^{f_*}(\zeta)$, so that $0 \in X^*$ by Proposition 2.3.(i) as applied to f_* (the latter function is s.-c.c.-c. and nondegenerate by Theorem 3.1). By symmetric reasons, $0 \in Y^*$, so that $(0,0) \in Z^* = X^* \times Y^*$.

(iv): Let us start with the following simple fact:

Lemma 8.2 *Let $A = \begin{pmatrix} P & R^T \\ R & -Q \end{pmatrix}$ be a symmetric matrix with positive definite P, Q, and let $S = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}$. Then A is nonsingular, and one has*

$$A^{-1} \preceq S^{-1} \quad \& \quad A^{-1} S A^{-1} \preceq S^{-1}.$$

In particular, if $f : Z = X \times Y \to \mathbf{R}$ is a nondegenerate s.-c.c.-c. function, then for every $z \in Z$ one has

$$[f''(z)]^{-1} \mathrm{S}_f(z) [f''(z)]^{-1} \preceq \mathrm{S}_f^{-1}(z).$$

Proof of Lemma: Nonsingularity of A was established in the proof of Proposition 3.1. Since A is symmetric and nonsingular, to prove that $A^{-1} \preceq S^{-1}$ is the same as to prove that $A \preceq AS^{-1}A$, which is immediate:

$$AS^{-1}A = \begin{pmatrix} P & R^T \\ R & -Q \end{pmatrix} \begin{pmatrix} P^{-1} & \\ & Q^{-1} \end{pmatrix} \begin{pmatrix} P & R^T \\ R & -Q \end{pmatrix} = \begin{pmatrix} P + R^T Q^{-1} R & \\ & RP^{-1}R^T + Q \end{pmatrix} = A + D,$$

where

$$D = \begin{pmatrix} R^T Q^{-1} R & -R^T \\ -R & RP^{-1}R^T + 2Q \end{pmatrix} \succeq \begin{pmatrix} R^T Q^{-1} R & -R^T \\ -R & Q \end{pmatrix} \succeq 0.$$

Similarly, in order to prove that $A^{-1}SA^{-1} \preceq S^{-1}$, it suffices to prove that $AS^{-1}A - S \succeq 0$; from the above computation, the latter difference is $\begin{pmatrix} R^T Q^{-1} R & \\ & RP^{-1}R^T \end{pmatrix}$, i.e., it indeed is positive semidefinite. ■

By Lemma 8.2,

$$\omega^2(f,z) = (f'(z))^T [f''(z)]^{-1} \mathrm{S}_f(z) [f''(z)]^{-1} f'(z) \leq (f'(z))^T [\mathrm{S}_f(z)]^{-1} f'(z) = \nu^2(f,z),$$

as required in (21.a).

To verify (21.b,c), denote

$$\lambda_x = \sqrt{(f'_x(z))^T [f''_{xx}(z)]^{-1} f'_x(z)}, \quad \lambda_y = \sqrt{(f'_y(z))^T [-f''_{yy}(z)]^{-1} f'_y(z)},$$

so that

$$\nu^2(f,z) = \lambda_x^2 + \lambda_y^2. \tag{60}$$

Setting $z = (x,y)$ and taking into account Remark 2.1 and (6), we get

$$\rho(-\lambda_x) \le [f(z) - \min_{x'\in X} f(x',y)] \le \rho(\lambda_x),\ \ \rho(-\lambda_y) \le [\max_{y'\in Y} f(x,y') - f(z)] \le \rho(\lambda_y). \tag{61}$$

We see that

$$\nu(f,z) < 1 \Rightarrow \mu(f,z) \le \max\{\rho(u)+\rho(v) \mid 0 \le u,v, u^2+v^2 = \nu^2(f,z)\},$$

and since the function $\rho(s^{1/2})$ clearly is convex on $[0,1)$, we come to $(21.b)$. Now, it is easily seen that

$$s \ge 0 \Rightarrow \rho(-s) \ge \frac{s^2}{2(1+s)},$$

whence in view of (61)

$$\frac{\lambda_x^2}{2(1+\lambda_x)} + \frac{\lambda_y^2}{2(1+\lambda_y)} \le \mu(f,z),$$

so that (60) yields

$$\frac{\nu^2(f,z)}{2(1+\nu(f,z))} \le \mu(f,z),$$

as required in $(21.c)$.

(v): The inclusion $z^+ \in Z$ is readily given by $(9.a)$. To prove (22), note that by $(9.d)$ for every $h \in \mathbf{R}^n \times \mathbf{R}^m$ one has

$$\frac{\omega^2(f,z)}{1-\omega(f,z)}\sqrt{h^T \mathrm{S}_f(z) h} \ge |h^T(f'(z^+) - f'(z) + f''(z)e(f,z))| = |h^T f'(z^+)|,$$

while by $(9.b)$ one has

$$\sqrt{h^T \mathrm{S}_f(z) h} \le (1-\omega(f,z))^{-1}\sqrt{h^T \mathrm{S}_f(z^+) h}.$$

Consequently,

$$|h^T f'(z^+)| \le \frac{\omega^2(f,z)}{(1-\omega(f,z))^2}\sqrt{h^T \mathrm{S}_f(z^+) h} \qquad \forall h,$$

which is nothing but (v).

(vi): Let

$$z^* = (x_*,y_*), z = (x,y), d_x = \sqrt{(x-x_*)^T f''_{xx}(z^*)(x-x_*)}, d_y = \sqrt{-(y-y_*)^T f''_{yy}(z^*)(y-y_*)}.$$

Then

$$\begin{aligned}\mu(f,z) &= \sup_{y'\in Y} f(x,y') - \inf_{x'\in X} f(x',y) \ge f(x,y_*) - f(x_*,y) \\ &\ge [f(x_*,y_*) + \rho(-d_x)] - [f(x_*,y_*) - \rho(-d_y)] = \rho(-d_x) + \rho(-d_y)\end{aligned}$$

(we have used (4) and the fact that $f'(z^*) = 0$). From the resulting inequality, as in the proof of $(21.c)$, we get

$$\frac{d_x^2}{2(1+d_x)} + \frac{d_y^2}{2(1+d_y)} \le \mu(f,z),$$

whence, after a simple computation, $d_x + d_y \le 2[\mu(f,z) + \sqrt{\mu(f,z)}]$, and (23) follows. ■

8.11 Proof of Theorem 5.1

In the proof to follow, Θ_i denote properly chosen universal positive continuous nondecreasing functions on the nonnegative ray.

Main Lemma. We start with the following technical result:

Lemma 8.3 *For properly chosen $\Theta_1(\cdot)$, the following holds. Let $f : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let $f_* : Z^* = X^* \times Y^* \to \mathbf{R}$ be the Legendre transformation of f. Let also $z_1 \equiv (x_1, y_1) \in K(f)$, $\zeta_1 \equiv (\xi_1, \eta_1) = f'(z_1)$, and let*

$$z_t = f'_*(t\zeta_1),\ 0 \le t \le 1.$$

Then for $0 \le t \le 1$ one has

$$\begin{array}{lrcl} (a) & \zeta_1^T \mathrm{S}_{f_*}(t\zeta_1)\zeta_1 & \le & \Theta_1(\mu(f, z_1)); \\ (b) & \Theta_1^{-1}(\mu(f, z_1))\mathrm{S}_{f_*}(\zeta_1) & \preceq & \mathrm{S}_{f_*}(t\zeta_1) \preceq \Theta_1(\mu(f, z_1))\mathrm{S}_{f_*}(\zeta_1). \end{array} \tag{62}$$

Proof. Since $K(f)$ is nonempty (it contains z_1), f possesses a saddle point (Proposition 5.1.(iii)). Thus, $(0, 0) \in Z^*$, whence $\zeta_t \in Z^*$, $0 \le t \le 1$.

Let us denote

$$\left.\begin{array}{rcl} \mu_x & = & f(x_1, y_1) - \min\limits_{x \in X} f(x, y_1) \quad [\ge 0] \\ \mu_y & = & \max\limits_{y \in Y} f(x_1, y) - f(x_1, y_1) \quad [\ge 0] \end{array}\right\} \Rightarrow \mu \equiv \mu(f, (x_1, y_1)) = \mu_x + \mu_y;$$
$$g_*(\xi, \eta) = f_*(\xi, \eta) - \xi^T x_1 - \eta^T y_1.$$

1^0. Note that g_* differs from f_* by a linear function and is therefore a nondegenerate s.-c.c.-c. function on Z^*. Besides this, by construction ζ_1 is a saddle point of g_* on Z^*.

2^0. By (12) we have

$$\min_{x \in X} f(x, y_1) = \min_{\eta \in Y^*} [\eta^T y_1 - f_*(0, \eta)] = \min_{\eta \in Y^*} [-g_*(0, \eta)] = -\max_{\eta \in Y^*} g_*(0, \eta),$$

while by Theorem 3.1

$$f(z_1) = \xi_1^T x_1 + \eta_1^T y_1 - f_*(\xi_1, \eta_1) = -g_*(\xi_1, \eta_1).$$

Thus,

$$\max_{\eta \in Y^*} g_*(0, \eta) - g_*(\xi_1, \eta_1) = \mu_x. \tag{63}$$

By symmetric reasoning,

$$g_*(\xi_1, \eta_1) - \min_{\xi \in X^*} g_*(\xi, 0) = \mu_y.$$

3^0. By 1^0, $\max\limits_{\eta \in Y^*} g_*(\xi_1, \eta) = g_*(\zeta_1)$, and since $g_*(\cdot, \eta)$ is convex on X^*, we get from (63)

$$\max_{\eta \in Y^*} g_*(t\xi_1, \eta) - g_*(\zeta_1) \le \mu_x,\ 0 \le t \le 1. \tag{64}$$

By symmetric reasoning,

$$g_*(\zeta_1) - \min_{\xi \in X^*} g_*(\xi, t\eta_1) \le \mu_y,\ 0 \le t \le 1. \tag{65}$$

4^0. Let $\xi_t = t\xi_1$, $\eta_t = t\eta_1$, $0 \le t \le 1$. Since ζ_1 is a saddle point of g_*, the s.-c. on X^* function

$$\phi(\xi) = g_*(\xi, \eta_1)$$

attains its minimum on X^* at the point ξ_1, and by (64) we have

$$\phi(\xi_t) \le \phi(\xi_1) + \mu_x,\ 0 \le t \le 1.$$

The latter inequality, due to the standard properties of s.-c. functions ([5], Section 2.2.4) combined with the fact that $\mu_x \le \mu$, implies that for $t \in [0, 1]$ one has

$$\Theta_2^{-1}(\mu)\phi''(\xi_1) \preceq \phi''(\xi_t) \preceq \Theta_2(\mu)\phi''(\xi_1), \qquad \xi_1^T \phi''(\xi_1)\xi_1 \le \Theta_3(\mu). \tag{66}$$

5^0. Let $h \in \mathbf{R}^m$, and let

$$\gamma(t) = -h^T (g_*)''_{\eta\eta}(\xi_t, \eta_1) h.$$

We have

$$|\gamma'(t)| = |D^3 g_*(\xi_t, \eta_1)[(0,h),(0,h),(-\xi_1,0)]| \underbrace{\leq}_{(a)} 2\gamma(t)\sqrt{\xi_1^T \phi''(\xi_t)\xi_1} \underbrace{\leq}_{(b)} \Theta_4(\mu)\gamma(t)$$

((a) is given by (8) as applied to g_*, (b) is given by (66)), so that

$$\Theta_5^{-1}(\mu)\gamma(1) \leq \gamma(t) \leq \Theta_5(\mu)\gamma(1),$$

for all $t \in [0,1]$, whence

$$\Theta_5^{-1}(\mu)[-(g_*)''_{\eta\eta}(\zeta_1)] \preceq [-(g_*)''_{\eta\eta}(\xi_t, \eta_1)] \preceq \Theta_5(\mu)[-(g_*)''_{\eta\eta}(\zeta_1)], \ 0 \leq t \leq 1. \tag{67}$$

6^0. Now let us fix $t \in [0,1]$ and set

$$\psi(\eta) = -g_*(\xi_t, \eta);$$

note that ψ is a s.-c. function on Y^*.

By (64) we have

$$\min_{\eta \in Y^*} \psi(\eta) \geq -g_*(\zeta_1) - \mu_x, \tag{68}$$

while for $s \in [0,1]$ it holds

$$\psi(\eta_s) = -g_*(\xi_t, \eta_s) \leq -\min_{\xi \in X^*} g_*(\xi, \eta_s) \leq -g_*(\zeta_1) + \mu_y \quad \text{[we have used (65)]}.$$

Combining these relations, we see that

$$\psi(\eta_s) \leq \min_{\eta \in Y^*} \psi(\eta) + \mu, \ 0 \leq s \leq 1. \tag{69}$$

7^0. From (69) and the same standard properties of s.-c. functions as in 4^0 we get for $0 \leq s \leq 1$:

$$\Theta_6^{-1}(\mu)\psi''(\eta_1) \preceq \psi''(\eta_s) \preceq \Theta_6(\mu)\psi''(\eta_1), \quad \eta_1^T \psi''(\eta_1)\eta_1 \leq \Theta_7(\mu);$$

combining this result with (67), we see that for $s, t \in [0,1]$ one has

$$\Theta_8^{-1}(\mu)[-(g_*)''_{\eta\eta}(\zeta_1)] \preceq [-(g_*)''_{\eta\eta}(\xi_t, \eta_s)] \preceq \Theta_8(\mu)[-(g_*)''_{\eta\eta}(\zeta_1)], \ \eta_1^T[-(g_*)''_{\eta\eta}(\zeta_1)]\eta_1 \leq \Theta_9(\mu). \tag{70}$$

By symmetric reasoning,

$$\Theta_8^{-1}(\mu)(g_*)''_{\xi\xi}(\zeta_1) \preceq (g_*)''_{\xi\xi}(\xi_t, \eta_s) \preceq \Theta_8(\mu)(g_*)''_{\xi\xi}(\zeta_1), \ \xi_1^T (g_*)''_{\xi\xi}(\zeta_1)\xi_1 \leq \Theta_9(\mu). \tag{71}$$

Relations (70) and (71) in view of $f''_*(\zeta) \equiv g''_*(\zeta)$ imply (62.a,b). ∎

From Main Lemma to Theorem 5.1. (i): Since (t^1, z^1) clearly satisfies (P_1), all we should prove is the implication $(P_i) \Rightarrow (P_{i+1})$.

1^0. Assume that i is such that (P_i) is valid. Setting

$$e_i = [f''(z^i)]^{-1}(f_{t^i})'(z^i), \ Q_i = \mathrm{S}_f(z^i),$$

we get by (21.a)

$$\omega_i \equiv \omega(f_{t^i}, z^i) = \| e_i \|_{Q_i} \leq \nu_i \equiv \nu(f_{t^i}, z^i) \leq 0.1, \tag{72}$$

the concluding inequality being given by (P_i). It follows that

$$0 \le t \le t^i \;\Rightarrow\; \begin{array}{rcl}\omega(f_t, z^i) & = & \| [f''(z^i)]^{-1} f'_t(z^i) \|_{Q_i} = \| [f''(z^i)]^{-1}[f'_{t^i}(z^i) + (t^i - t)f'(\widehat{z})] \|_{Q_i} \\ & \le & \omega_i + (t^i - t)\gamma_i, \quad \gamma_i \equiv \| [f''(z^i)]^{-1} f'(\widehat{z}) \|_{Q_i} = \| f'(\widehat{z}) \|_{\mathrm{S}_{f_*}(f'(z^i))}\end{array} \tag{73}$$

(see (11.b)).

2^0. Since $\omega_i \le \nu_i \le 0.1$ by (72), (73) says that there indeed exists $t \in [0, t^i]$ such that $\omega(f_t, z^i) \le 0.2$, so that t^{i+1} is well-defined. Note also that by (73)

$$t^{i+1} > 0 \Rightarrow \omega(f_{t^{i+1}}, z^i) = 0.2 \Rightarrow t^i - t^{i+1} \ge 0.1\gamma_i^{-1}. \tag{74}$$

3^0. For the sake of brevity, denote temporarily

$$g(z) = f_{t^{i+1}}(z),\; d = g'(z^i),\; e = [f''(z^i)]^{-1} d = [g''(z^i)]^{-1} d.$$

By definition of t^{i+1} we have

$$\sigma_i \equiv \omega(g, z^i) = \| e \|_{Q_i} \le 0.2,$$

whence (see (24) and (9.a))

$$z^{i+1} = z^i - e \in Z.$$

Moreover, by (9.d) and in view of $\| e \|_{Q_i} = \sigma_i \le 0.2$ we have

$$\forall (h \in \mathbf{R}^{n+m}) : \qquad |h^T g'(z^{i+1})| = |h^T [g'(z^{i+1}) - g'(z^i) + g''(z^i) e]| \le \frac{\sigma_i^2}{1 - \sigma_i} \| h \|_{Q_i},$$

whence

$$\| g'(z^{i+1}) \|_{Q_i^{-1}} \le \frac{\sigma_i^2}{1 - \sigma_i} \le \frac{0.04}{0.8} = 0.05. \tag{75}$$

Now note that by (9.b)

$$\mathrm{S}_f(z^{i+1}) \succeq (1 - \| e \|_{Q_i})^2 \mathrm{S}_f(z^i),$$

so that by (75)

$$\begin{array}{rcl}\nu(f_{t^{i+1}}, z^{i+1}) & \equiv & \nu(g, z^{i+1}) = \| g'(z^{i+1}) \|_{[\mathrm{S}_f(z^{i+1})]^{-1}} \le (1 - \| e \|_{Q_i})^{-1} \| g'(z^{i+1}) \|_{[\mathrm{S}_f(z^i)]^{-1}} \\ & = & (1 - \| e \|_{Q_i})^{-1} \| g'(z^{i+1}) \|_{Q_i^{-1}} \le \frac{\sigma_i^2}{(1-\sigma_i)^2} \le \frac{0.05}{0.8} < 0.1.\end{array} \tag{76}$$

Thus, (P_{i+1}) indeed is valid. (i) is proved.

(ii): Assume that i is such that $t^{i+1} > 0$. By (74)

$$t^i - t^{i+1} \ge 0.1\gamma_i^{-1}, \qquad \gamma_i = \| f'(\widehat{z}) \|_{\mathrm{S}_{f_*}(f'(z^i))}. \tag{77}$$

Now let

$$g(z) = f_{t^i}(z), \tag{78}$$

so that in view of (P_i) we have

$$[\omega(g, z^i) \le] \quad \nu(g, z^i) \le 0.1 \tag{79}$$

(the left inequality is given by (21.a)).

Lemma 8.4 *Let $g : Z = X \times Y \to \mathbf{R}$ be a nondegenerate s.-c.c.-c. function, and let $z_0 = (x_0, y_0) \in Z$ be such that $\omega \equiv \omega(g, z) \le 0.2$. Then g possesses a saddle point z^* on Z, and*

$$\| z^* - z_0 \|_{\mathrm{S}_f(z_0)} \le \omega + 2.7\omega^2. \tag{80}$$

Proof. Consider the Newton iterate

$$z_1 \equiv z_0 - [g''(z_0)]^{-1} g'(z_0) = (x_1, y_1)$$

of z_0. Same as in 3^0, $z_1 \in Z$ and

$$\| z_0 - z_1 \|_{\mathrm{S}_f(z_0)} = \omega, \quad \nu(g, z_1) \equiv \nu \leq \frac{\omega^2}{(1-\omega)^2} \qquad \text{[see (76)]}. \tag{81}$$

Let $z^* = (x^*, y^*)$ be the saddle point of g (it exists by Proposition 5.1.(iii)). Let us set

$$Q \equiv \mathrm{S}_f(z^*) = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, \quad r_x = \| x_1 - x^* \|_A, \quad r_y = \| y_1 - y^* \|_B \ .$$

By (4) we have

$$\max_y g(x_1, y) \geq g(x_1, y^*) \geq g(z^*) + r_x - \ln(1 + r_x), \quad \min_x g(x, y_1) \leq g(x^*, y_1) \leq g(z^*) - r_y + \ln(1 + r_y),$$

whence

$$\mu(g, z_1) \geq r_x + r_y - \ln(1 + r_x) - \ln(1 + r_y),$$

and at the same time

$$\mu(g, z_1) \leq -\nu - \ln(1 - \nu) \qquad \left[\nu \leq \frac{\omega^2}{(1-\omega)^2}\right]$$

(we have used (21.b) and (81)). Thus,

$$-\nu - \ln(1 - \nu) \geq r_x + r_y - \ln(1 + r_x) - \ln(1 + r_y) \tag{82}$$

Since $\omega \leq 0.2$, we have $\nu \leq \frac{\omega^2}{0.64} \leq 0.07$, and therefore $-\nu - \ln(1-\nu) \leq 0.53\nu^2 \leq 0.0026$. Consequently, from (82) it immediately follows that $\max[r_x, r_y] \leq 0.25$, whence $r_x + r_y - \ln(1 + r_x) - \ln(1 + r_y) \geq 0.4(r_x^2 + r_y^2)$. Thus, (82) results in

$$0.53\nu^2 \geq 0.4(r_x^2 + r_y^2),$$

whence

$$\| z_1 - z^* \|_Q \leq 1.2\nu \leq 1.875\omega^2 \leq 0.08.$$

The latter inequality, by (9.b), implies that

$$\| z_1 - z^* \|_{\mathrm{S}_g(z_1)} \leq (1 - 0.08)^{-1} \| z_1 - z^* \|_Q \leq 2.1\omega^2.$$

By the same (9.b), from (81) it follows that $\mathrm{S}_g(z_1) \succeq (1-\omega)^2 \mathrm{S}_g(z_0)$, whence

$$\| z_1 - z^* \|_{\mathrm{S}_g(z_0)} \leq (1 - \omega)^{-1} \| z_1 - z^* \|_{\mathrm{S}_g(z_1)} \leq 2.7\omega^2$$

(recall that $\omega \leq 0.2$). Combining this relation with (81), we come to (80). ∎

Now we are ready to complete the proof of (ii). Applying Lemma 8.4 to the function g given by (78) and taking into account (79), we see that

$$\| z^i - z^*(t^i) \|_{\mathrm{S}_f(z^i)} \leq 0.1 + 2.7 \times 0.1^2 = 0.127. \tag{83}$$

Now let

$$h(\zeta) = f_*(\zeta) - (z^*(t^i))^T \zeta : Z^*(f) \to \mathbf{R}, \quad \zeta_i = f'(z^i), \quad \zeta_i^* = f'(z^*(t^i)) = t^i f'(\widehat{z}). \tag{84}$$

We have

$$h'(\zeta_i) = z^i - z^*(t^i), \quad h'(\zeta_i^*) = 0, \quad h_*(z) = f(z + z^*(t^i)),$$

so that

$$\begin{array}{rrcll} & \omega(h,\zeta_i) & = & \sqrt{[z^i - z^*(t^i)]^T \mathrm{S}_f(z^i)[z^i - z^*(t^i)]} & \text{[we have used (20)]} \\ & & \le & 0.127 & \text{[see (83)]} \\ \Rightarrow & \| \zeta_i - \zeta_i^* \|_{\mathrm{S}_{f_*}(\zeta_i)} & \le & 0.127 + 2.7 \times (0.127)^2 & \text{[we have used Lemma 8.4]} \\ & & \le & 0.171 & \\ \Rightarrow & \mathrm{S}_{f_*}(\zeta_i) & \le & (0.82)^{-2}\mathrm{S}_{f_*}(\zeta_i^*) & \text{[we have used (9.b)]} \\ & & = & (0.82)^{-2}\mathrm{S}_{f_*}(t^i f'(\widehat{z})) & \text{[see (84)].} \end{array}$$

Consequently,

$$\gamma_i = \| f'(\widehat{z}) \|_{\mathrm{S}_{f_*}(f'(z^i))} \le 1.22 \| f'(\widehat{z}) \|_{\mathrm{S}_{f_*}(t^i f'(\widehat{z}))} \le \Theta_2(\mu(f,\widehat{z})),$$

(the concluding inequality is given by (62.a)). The resulting upper bound on γ_i combined with (77) implies (ii).

(iii): In the case of $t^i = 0$ one has (we use the same notation as in 3^0)

$$\sigma_i \equiv \omega(g, z^i) = \omega(f, z^i) \le \nu(f, z^i),$$

the latter inequality being given by (21.a). This inequality combined with (76) implies (25).

(iv): By (ii) it takes no more than $\Theta_{10}(\mu(f,\widehat{z}))$ steps to make t^i equal to zero, and by (25) it takes at most $O(1)\ln\ln(3/\epsilon)$ more steps to make $\nu(f, z^i)$ less than ϵ. ∎

8.12 Proof of Proposition 6.1

(i) is an immediate consequence of the fact that whenever B is s.-c. on Z, the seminorm $\| \cdot \|_{B''(z)}$ majorates the seminorm $\pi_z^Z(\cdot)$, $z \in Z$ (see [5], Theorem 2.1.1.(ii)).

The remaining statements, except (vi), are immediate consequences of definitions. To prove (vi), it clearly suffices to demonstrate that the function

$$\psi(x,(\mu,v)) = \mu f(x, \mu^{-1}v) : W = X \times V \to \mathbf{R}$$

is $(2\beta+3)$-regular.

Let us fix a point $w = (x, s = (\mu, v)) \in W$ and a direction $dw = (dx, ds = (d\mu, dv))$ in the $(x,(\mu,v))$-space, and let

$$z = (x, \mu^{-1}v), \quad \delta y = \mu^{-1}dv - \mu^{-2}d\mu v, \quad dz = (dx, \delta y).$$

A straightforward computation yields (the derivatives of ψ are taken at w, the derivatives of f are taken at z):

$$\begin{array}{rrcl} & D\psi[dw] & = & \mu Df[dz] + d\mu f; \\ & D^2\psi[dw,dw] & = & \mu D^2 f[dz,dz] + \mu Df[(0, -2\mu^{-1}d\mu\delta y)] + 2d\mu Df[dz] \\ & & = & \mu D^2 f[dz,dz] + 2d\mu Df[(dx,0)] \\ \Rightarrow & D^2\psi[(dx,0),(dx,0)] & = & \mu D^2 f[(dx,0),(dx,0)] \ge 0, \\ & D^2\psi[(0,du),(0,du)] & = & \mu D^2 f[(0,\delta y),(0,\delta y)] \le 0, \\ \Rightarrow \quad (a) & dw^T \mathrm{S}_\psi(w) dw & = & \mu dz^T \mathrm{S}_f(z) dz; \\ & D^3\psi[dw,dw,dw] & = & \mu D^3 f[dz,dz,dz] - 4\mu D^2 f[(dx,\delta y),(0,\mu^{-1}d\mu\delta y)] \\ & & & +2d\mu D^2 f[(dx,0),(dx,\delta y)] + d\mu D^2 f[(dx,\delta y),(dx,\delta y)] \\ & & = & \mu D^3 f[dz,dz,dz] - 4d\mu D^2 f[(dx,0),(0,\delta y)] \\ & & & -4d\mu D^2 f[(0,\delta y),(0,\delta y)] + 2d\mu D^2 f[(dx,0),(dx,0)] \\ & & & +2d\mu D^2 f[(dx,0),(0,\delta y)] + d\mu D^2 f[(dx,0),(dx,0)] \\ & & & +2d\mu D^2 f[(dx,0),(0,\delta y)] + d\mu D^2 f[(0,\delta y),(0,\delta y)] \\ & & = & \mu D^3 f[dz,dz,dz] + 3d\mu D^2 f[(dx,0),(dx,0)] \\ & & & -3d\mu D^2 f[(0,\delta y),(0,\delta y)] \\ \Rightarrow \quad (b) & D^3\psi(w)[dw,dw,dw] & = & \mu D^3 f(z)[dz,dz,dz] + 3\frac{d\mu}{\mu} dw^T \mathrm{S}_\psi(w) dw. \end{array}$$

It remains to note that

$$\max\{2\mu^{-1}d\mu, \pi_z^Z(dz)\} \leq 2\pi_w^W(dw). \tag{85}$$

Indeed, assuming (85), we immediately get from (a), (b) and from regularity of f that

$$\begin{array}{rcl}|D^3\psi(w)[dw,dw,dw]| & \leq & \mu\beta(dz^T \mathrm{S}_f(z)dz)\pi_z^Z(dz) + 3|\mu^{-1}d\mu|(dw^T\mathrm{S}_\psi(w)dw) \\ & \leq & (2\beta+3)\pi_w^W(dw)(dw^T\mathrm{S}_\psi(w)dw),\end{array}$$

as claimed. To prove (85), note that the relation $|\mu^{-1}d\mu| \leq \pi_w^W(dw)$ is evident, and all we need to prove is that if $w \pm dw \in X \times V$, then $z \pm 0.5dz \in X \times Y$, i.e., that if $|d\mu| < \mu$ and $(\mu \pm d\mu)^{-1}(v \pm dv) \in Y$, then

$$y^{\pm} = \mu^{-1}v \pm 0.5(\mu^{-1}dv - \mu^{-2}d\mu v) \in Y.$$

This is immediate: setting $\theta = \mu^{-1}d\mu$, $r = \mu^{-1}v$, $dr = \mu^{-1}dv$, we have $|\theta| < 1$,

$$(1+\theta)^{-1}(r+dr) \in Y, (1-\theta)^{-1}(r-dr) \in Y$$

and therefore

$$\begin{array}{rcl}y^+ & \equiv & r + 0.5[dr - \theta r] \\ & = & [(1+\theta)(0.75-0.25\theta)](1+\theta)^{-1}(r+dr) + [(1-\theta)(0.25-0.25\theta)](1-\theta)^{-1}(r-dr) \\ & \in & Y, \\ y^- & \equiv & r - 0.5[dr - \theta r] \\ & = & [(1+\theta)(0.25+0.25\theta)](1+\theta)^{-1}(r+dr) + [(1-\theta)(0.75+0.25\theta)](1-\theta)^{-1}(r-dr) \\ & \in & Y\end{array}$$

(note that $[(1+\theta)(0.75-0.25\theta)]+[(1-\theta)(0.25-0.25\theta)] = 1$, $[(1+\theta)(0.25+0.25\theta)]+[(1-\theta)(0.75+0.25\theta)] = 1$, and that all four weights in question are nonnegative). ■

8.13 Justification of Examples 6.2 and 6.3

Example 6.2. The convexity of Y is evident. Now, setting

$$z = (x,y) \in Z, \quad dz = (dx,dy) \in \mathbf{R}^n \times \mathbf{R}^m, \quad s = S(y), \quad ds = DS(y)[dy], \quad d^2s = D^2S(y)[dy,dy],$$

we have

$$\begin{array}{lrcl} & Df(z)[dz] & = & 2dx^T sx + x^T ds\, x; \\ & D^2f(z)[dz,dz] & = & 2dx^T s\, dx + 4dx^T ds\, x + x^T d^2s\, x; \\ (a) & dz^T\mathrm{S}_f(z)dz & = & 2dx^T s\, dx - x^T d^2s\, x = a^2+b^2,\ a = \sqrt{2dx^T s\, dx},\ b = \sqrt{x^T[-d^2s]x}; \\ (b) & D^3f(z)[dz,dz,dz] & = & 6dx^T ds\, dx + 6dx^T d^2s\, x.\end{array} \tag{86}$$

Denoting $\pi = \pi_z^Z(dz)$, we have

$$\pi' > \pi \Rightarrow y \pm (\pi')^{-1}dy \in Y \Rightarrow S(y \pm (\pi')^{-1}dy) > 0,$$

and since $S(\cdot)$ is quadratic,

$$S(y \pm (\pi')^{-1}dy) = s \pm (\pi')^{-1}ds + \frac{1}{2}(\pi')^{-2}d^2s,$$

whence for all $\pi' > \pi$ it holds

$$s \pm (\pi')^{-1}ds + \frac{1}{2}(\pi')^{-2}d^2s > 0. \tag{87}$$

Since $d^2s \leq 0$, we conclude, first, that $\pi' > \pi \Rightarrow s \pm (\pi')^{-1}ds > 0$, whence

$$-\pi s \leq ds \leq \pi s; \tag{88}$$

second, taking the arithmetic mean of the two inequalities in (87), we get $\pi' > \pi \Rightarrow \frac{1}{2}(\pi')^{-2}[-d^2s] < s$, whence

$$0 \le -d^2s \le 2\pi^2 s. \tag{89}$$

Finally we have

$$\begin{array}{rcll}
|D^3f(z)[dz,dz,dz]| & \le & 6|dx^T ds\, dx| + 6|dx^T[-d^2s]x| & \text{[see (86.}b\text{)]}\\
& \le & 6\pi dx^T s dx + 6\sqrt{dx^T[-d^2s]dx}\sqrt{x^T[-d^2s]x} & \text{[by (88) and since } -d^2s \ge 0]\\
& \le & 6\pi dx^T s\, dx + 6\pi\sqrt{2}\sqrt{dx^T s\, dx}\sqrt{x^T[-d^2s]x} & \text{[see (89)]}\\
& = & 3\pi(a^2+2ab) & \text{[see (86.}a\text{)]}\\
& \le & 5\pi(a^2+b^2) = 5\pi dz^T \mathrm{S}_f(z)dz. &
\end{array} \quad \blacksquare$$

Example 6.3. Let $x, y \in \mathbf{R}^m_{++}$ and let $h = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbf{R}^m \times \mathbf{R}^m$. In the below computation, lowercase letters like a, b, x, y, u, v denote m-dimensional vectors, and their uppercase counterparts stand for the corresponding diagonal matrices: $A = \mathrm{Diag}(a), X = \mathrm{Diag}(x), Y = \mathrm{Diag}(y), \ldots$ We write Z^d instead of $\mathrm{Diag}(z^d)$ ($z \in \mathbf{R}^m_{++}, d \in \mathbf{R}^m$); e stands for the m-dimensional vector of ones.

Let

$$Q = Q(x,y) = E^T X^{-b} Y^a E;$$

then

$$\begin{array}{rcl}
Df(x,y)[h] & = & \mathrm{Tr}(\,Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E);\\
D^2f(x,y)[h,h] & = & -\mathrm{Tr}(\,Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E\\
& & \quad \times Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E)\\
& & +\mathrm{Tr}(\,Q^{-1}E^T[B(B+I_m)X^{-b-2e}U^2Y^a - 2ABX^{-b-e}UVY^{a-e}\\
& & \quad +A(A-I_m)X^{-b}Y^{a-2e}V^2]E);\\
D^3f(x,y)[h,h,h] & = & 2\mathrm{Tr}(\,Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E\\
& & \quad \times Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E\\
& & \quad \times Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E)\\
& & -3\mathrm{Tr}(\,Q^{-1}E^T[B(B+I_m)X^{-b-2e}U^2Y^a - 2ABX^{-b-e}UVY^{a-e}\\
& & \quad +A(A-I_m)X^{-b}Y^{a-2e}V^2]E\\
& & \quad \times Q^{-1}E^T[-BX^{-b-e}UY^a + X^{-b}AVY^{a-e}]E)\\
& & +\mathrm{Tr}(\,Q^{-1}E^T[-B(B+I_m)(B+2I_m)X^{-b-3e}U^3Y^a\\
& & \quad +3B(B+I_m)AX^{-b-2e}Y^{a-e}U^2V - 3BA(A-I_m)X^{-b-e}Y^{a-2e}UV^2\\
& & \quad +A(A-I_m)(A-2I_m)X^{-b}Y^{a-3e}V^3]E).
\end{array} \tag{90}$$

Now let us set

$$\delta U = UX^{-1}, \quad \delta V = VY^{-1}, \quad P = X^{-b/2}Y^{a/2}EQ^{-1}E^TX^{-b/2}Y^{a/2}.$$

Note that P is an orthoprojector. We clearly have

$$\begin{array}{rcl}
D^2f(x,y)[h,h] & = & -\mathrm{Tr}(P[A\delta V - B\delta U]P[A\delta V - B\delta U]P)\\
& & +\mathrm{Tr}(P[B(B+I_m)\delta U^2 - 2AB\delta U\delta V + A(A-I_m)\delta V^2]P)\\
\Rightarrow D^2f(x,y)[(u,0),(u,0)] & = & -\mathrm{Tr}(PB\delta UPB\delta UP) + \mathrm{Tr}(PB(B+I_m)\delta U^2P)\\
& = & \mathrm{Tr}(PB\delta U^2P) + \mathrm{Tr}(PB\delta U(I_m - P)\delta UBP)\\
& \ge & \mathrm{Tr}(PB\delta U^2P) \quad \text{[since } P \text{ is an orthoprojector]}\\
& \ge & 0;\\
-D^2f(x,y)[(0,v),(0,v)] & = & \mathrm{Tr}(PA\delta VPA\delta VP) + \mathrm{Tr}(PA(I_m - A)\delta V^2P) \ge 0.
\end{array}$$

We see that f is convex in x and concave in y and that

$$\begin{array}{rl}
h^T\mathrm{S}_f(x,y)h & \ge \ \mathrm{Tr}(PB\delta U^2P) + \mathrm{Tr}(PA\delta VP\delta VAP) + \mathrm{Tr}(PA(I_m - A)\delta V^2P) \equiv \omega^2 = p^2+q^2+r^2,\\
& p = \parallel PB^{1/2}\delta U \parallel_2, \quad q = \parallel PA\delta VP \parallel_2, \quad r = \parallel PA^{1/2}(I_m - A)^{1/2}\delta V \parallel_2;
\end{array} \tag{91}$$

from now on, $\parallel S \parallel_2 = \sqrt{\mathrm{Tr}(S^TS)}$ stands for the Frobenius, and $\parallel S \parallel_\infty$ – for the operator norm (maximum singular value) of a matrix S.

Now, (90) can be rewritten as

$$\begin{array}{rcl}D^3f(x,y)[h,h,h] & = & 2\mathrm{Tr}([P[A\delta V - B\delta U]P]^3)\\ & & -3\mathrm{Tr}(P[B(B+I_m)\delta U^2 - 2AB\delta U\delta V + A(A-I_m)\delta V^2]P[A\delta V - B\delta U]P)\\ & & +\mathrm{Tr}(P[-B(B+I_m)(B+2I_m)\delta U^3 + 3B(B+I_m)A\delta U^2\delta V\\ & & -3BA(A-I_m)\delta U\delta V^2 + A(A-I_m)(A-2I_m)\delta V^3]P).\end{array}$$

We have

$$\|\delta U\|_\infty \le \pi \equiv \pi^{\mathbf{R}^m_+\times\mathbf{R}^m_+}_{(x,y)}(h);\quad \|\delta V\|_\infty \le \pi. \tag{92}$$

Therefore

$$\begin{array}{rcl}2\left|\mathrm{Tr}\left([P[A\delta V - B\delta U]P]^3\right)\right| & \le & 2\|P[A\delta V - B\delta U]P\|_\infty \|P[A\delta V - B\delta U]P\|_2^2\\ & \le & 2(1+\|b\|_\infty)\pi[\|PB\delta UP\|_2 + \|PA\delta VP\|_2]^2\\ & & \text{[we have used (92)]}\\ & \le & 2(1+\|b\|_\infty)\pi\,[q+\|PB\delta UP\|_2]^2\\ & & \text{[we have used (91)]}\\ & \le & 2(1+\|b\|_\infty)\pi\,[q+\|b\|_\infty^{1/2}\,p]^2\\ & & [\text{since } \|PB\delta UP\|_2 \le \|PB^{1/2}\delta U\|_2\|B^{1/2}P\|_\infty \le p\|b\|_\infty^{1/2}]\\ & \le & 2(1+\|b\|_\infty)(1+\|b\|_\infty^{1/2})^2\pi\omega^2 \le 4(1+\|b\|_\infty)^2\pi\omega^2,\end{array} \tag{93}$$

the first inequality being given by the following well-known fact:

> *Whenever the product of three matrices Q_1, Q_2, Q_3 makes sense and is a square matrix, the trace of the product is in absolute value $\le$ the product of the operator norm of one of the factors (no matter which one) and the Frobenius norms of the two remaining factors.*

Note that a byproduct of our reasoning is the inequality

$$\|P[A\delta V - B\delta U]P\|_2 \le (1+\|b\|_\infty^{1/2})\omega. \tag{94}$$

By similar reasons,

$$\begin{array}{rl} & 3\left|\mathrm{Tr}\left(P\left[B(B+I_m)\delta U^2 - 2AB\delta U\delta V + A(A-I_m)\delta V^2\right]P\left[A\delta V - B\delta U\right]P\right)\right|\\ \le & 3\|PB^{1/2}\delta U\|_2 \times \|B^{1/2}(B+I_m)\delta U - 2AB^{1/2}\delta V\|_\infty \|P[A\delta V - B\delta U]P\|_2\\ & +3\|PA^{1/2}(I_m-A)^{1/2}\delta V\|_2\|A^{1/2}(I_m-A)^{1/2}\delta V\|_\infty\|P[A\delta V - B\delta U]P\|_2\\ \le & 3p\,[\|b\|_\infty^{1/2}(3+\|b\|_\infty)]\pi\,(1+\|b\|_\infty^{1/2})\omega + 3r\,(\pi/2)\,(1+\|b\|_\infty^{1/2})\omega\\ \le & 10(1+\|b\|_\infty)^2\pi\omega^2 \qquad \text{[we have used (91), (92), (94)];}\end{array} \tag{95}$$

$$\begin{array}{rl} & \left|\mathrm{Tr}\left(P\left[-B(B+I_m)(B+2I_m)\delta U^3 + 3B(B+I_m)A\delta U^2\delta V - 3BA(A-I_m)\delta U\delta V^2 + A(A-I_m)(A-2I_m)\delta V^3\right]P\right)\right|\\ \le & \|PB^{1/2}\delta U\|_2^2\|(B+I_m)(B+2I_m)\delta U\|_\infty + 3\|PB^{1/2}\delta U\|_2^2\|A(B+I_m)\delta V\|_\infty\\ & +3\|PA^{1/2}(I_m-A)^{1/2}\delta V\|_2^2\|B\delta U\|_\infty + \|PA^{1/2}(I_m-A)^{1/2}\delta V\|_2^2\|(A-2I_m)\delta V\|_\infty\\ \le & \left[(1+\|b\|_\infty)(2+\|b\|_\infty) + 3(1+\|b\|_\infty) + 3\|b\|_\infty + 2\right]\pi\omega^2\\ \le & 7(1+\|b\|_\infty)^2\pi\omega^2 \qquad \text{[we have used (91), (92)].}\end{array} \tag{96}$$

Combining (91), (93), (95), (96), we come to

$$|D^3f(x,y)[h,h,h]| \le 21(1+\|b\|_\infty)^2\pi\omega^2 = 21(1+\|b\|_\infty)^2\pi^{\mathbf{R}^m_+\times\mathbf{R}^m_+}_{(x,y)}(h)(h^T\mathrm{S}_f(x,y)h).$$ ■

8.14 Proof of Proposition 6.2

Let $t > 0$. The function f_t clearly is convex in $x \in X$ for every $y \in Y$, is concave in $y \in Y$ for every $x \in X$ and satisfies the requirements from Definition 2.2.(i). In order to verify Definition 2.2.(ii), let us fix

$z=(x,y)\in Z\equiv X\times Y$ and $h=(u,v)\in\mathbf{R}^n\times\mathbf{R}^m$. We have

$$\begin{array}{rcll}
|D^3 f_t(z)[h,h,h]| & \leq & \gamma\left[|tD^3 f(z)[h,h,h]|+|D^3F(x)[u,u,u]|+|D^3G(y)[v,v,v]|\right] & \\
& \leq & \gamma\left[\beta t h^T \mathrm{S}_f(z)h\sqrt{\| u\|^2_{F''(x)}+\| v\|^2_{G''(y)}}\right. & \\
& & \left.+2\| u\|^3_{F''(x)}+2\| v\|^3_{G''(v)}\right] & \text{[since } F,\, G \text{ are s.-c.]}\\
& \leq & \gamma\left[\frac{2\beta}{3}\left(th^T\mathrm{S}_f(z)h\right)^{3/2}+\frac{\beta}{3}\left(\| u\|^2_{F''(x)}+\| v\|^2_{G''(y)}\right)^{3/2}\right. & \\
& & \left.+2\| u\|^3_{F''(x)}+2\| v\|^3_{G''(y)}\right] & \text{[the Hölder inequality]}\\
& \leq & \gamma(\beta+2)\left[th^T\mathrm{S}_f(z)h+u^TF''(x)u+v^TG''(y)v\right]^{3/2} & \\
& = & 2\left[h^T\mathrm{S}_{f_t}(z)h\right]^{3/2} & \text{[since } \gamma\frac{\beta+2}{2}=\gamma^{3/2}\text{]}
\end{array}$$

Finally, the nondegeneracy of f_t is readily given by **(C)**, see Proposition 2.2.(ii). ■

8.15 Proof of Proposition 6.3

Let us set

$$\phi(x) = \gamma f(x,\bar{y}),\quad \Phi(x) = \gamma F(x)-\gamma G(\bar{y}),\quad \psi(x) = t\phi(x)+\Phi(x),\quad \theta = \gamma\vartheta,$$

so that

$$f_\tau(x,\bar{y})=\tau\phi(x)+\Phi(x),\ x\in X,\tau\geq 0,\quad f_t(x,\bar{y})=\psi(x).$$

1^0. We have

$$\lambda(\psi,\bar{x})\leq\nu(f_t,\bar{z})\leq 0.1, \tag{97}$$

$\lambda(\psi,x)=\sqrt{(\psi'(x))^T[\psi''(x)]^{-1}\psi'(x)}$ being the Newton decrement of the convex s.-c. function ψ at a point x. By [4], (2.2.23) - (2.2.24), relation (97) implies existence of $x^*\in X$ such that $\psi'(x^*)=0$, and that

$$\begin{array}{lrcl}
(a) & \|\bar{x}-x^*\| & \leq & \frac{\lambda(\psi,\bar{x})}{1-\lambda(\psi,\bar{x})}\leq 0.12,\quad \| u\|\equiv\sqrt{u^T\psi''(x^*)u};\\
(b) & \psi(\bar{x})-\psi(x^*) & \leq & -\lambda(\psi,\bar{x})-\ln(1-\lambda(\psi,\bar{x}))\leq 0.006.
\end{array} \tag{98}$$

2^0. By (98.b) we have

$$t\phi(\bar{x})+\Phi(\bar{x})\leq t\phi(x^*)+\Phi(x^*)+0.006\Rightarrow\phi(\bar{x})-\phi(x^*)\leq\frac{1}{t}[\Phi(x^*)-\Phi(\bar{x})+0.006]. \tag{99}$$

3^0. Let $\| u\|_\Phi=\sqrt{u^T\Phi''(x^*)u}$; note that $\| u\|_\Phi\leq\| u\|$, so that by(98.a) we have

$$r\equiv\| x^*-\bar{x}\|_\Phi\leq 0.12.$$

Since Φ is convex, it follows that

$$\begin{array}{rcll}
\Phi(\bar{x}) & \geq & \Phi(x^*)+(\bar{x}-x^*)^T\Phi'(x^*) & \\
& = & \Phi(x^*)-\|\bar{x}-x^*\|_\Phi\sqrt{\theta} & \text{[since } \Phi \text{ is a } \theta\text{-s.-c.b. for } \mathrm{cl}\, X\text{]}\\
& \geq & \Phi(x^*)-0.12\sqrt{\theta}, &
\end{array}$$

and we come to

$$\Phi(x^*)-\Phi(\bar{x})\leq 0.12\sqrt{\theta}. \tag{100}$$

Combining this result with (99), we obtain

$$\phi(\bar{x})-\phi(x^*)\leq\frac{1}{t}[0.12\sqrt{\theta}+0.006]\leq\frac{1}{5t}\sqrt{\theta}. \tag{101}$$

Now let $x \in X$. We have

$$\begin{array}{rcll}
\phi(x) & \geq & \phi(x^*) + (x-x^*)^T\phi'(x^*) & \\
& = & \phi(x^*) - \frac{1}{t}(x-x^*)^T\Phi'(x^*) & [\text{since } t\phi'(x^*)+\Phi'(x^*)=0] \\
& \geq & \phi(x^*) - \frac{\theta}{t} & [\text{since } \Phi \text{ is } \theta\text{-s.-c.b. for } \mathrm{cl}\, X, \text{ see [5], (2.3.2)}].
\end{array}$$

From this relation and (101) it follows that

$$\phi(\bar{x}) - \inf_X \phi \leq \frac{1}{5t}\sqrt{\theta} + \frac{1}{t}\theta \leq \frac{2\theta}{t},$$

whence

$$f(\bar{x},\bar{y}) - \inf_{x\in X} f(x,\bar{y}) \leq \frac{2\theta}{t\gamma} = \frac{2\vartheta}{t}.$$

Symmetric reasoning yields $\sup\limits_{y\in Y} f(\bar{x},y) - f(\bar{x},\bar{y}) \leq \frac{2\vartheta}{t}$, and (26) follows.

4^0. By (99) we have

$$\begin{array}{rcll}
[t^+\phi(\bar{x})+\Phi(\bar{x})] - [t^+\phi(x^*)+\Phi(x^*)] & \leq & \left[\frac{t^+}{t}-1\right][\Phi(x^*)-\Phi(\bar{x})] + 0.006\frac{t^+}{t} & \\
& = & \alpha[\Phi(x^*)-\Phi(\bar{x})] + 0.006(1+\alpha) & \\
& \leq & 0.12\alpha\sqrt{\theta} + 0.006(1+\alpha) & [\text{see (100)}].
\end{array} \tag{102}$$

5^0. Now let $x \in X$. We have

$$\begin{array}{rcl}
t^+\phi(x)+\Phi(x) & \geq & t^+\phi(x^*) + t^+(x-x^*)^T\phi'(x^*) + \Phi(x) \\
& = & t^+\phi(x^*) - (1+\alpha)(x-x^*)^T\Phi'(x^*) + \Phi(x) = t^+\phi(x^*) + \Phi(x^*) + \Psi_\alpha(x), \\
\Psi_\alpha(x) & = & \Phi(x) - \Phi(x^*) - (1+\alpha)(x-x^*)^T\Phi'(x^*).
\end{array} \tag{103}$$

The remaining reasoning reproduces the one of [5], Section 3.2.6. First, we claim that the function $\Psi_\alpha(x)$, for every $\alpha \geq 0$ (in fact – even for $\alpha > -1$) attains its minimum on X at a point $x(\alpha)$. Indeed, Φ is a θ-s.-c.b. for $\mathrm{cl}\, X$; consequently, the image of X under the mapping $x \mapsto \Phi'(x)$ is the relative interior of a convex cone ([5], Proposition 2.3.2 and Theorem 2.4.2). Since this cone clearly contains the point $\Phi'(x^*)$, it also contains the points $(1+\alpha)\Phi'(x^*)$, so that there exists $x(\alpha) \in X$ with $\Phi'(x(\alpha)) = (1+\alpha)\Phi'(x^*)$, or, which is the same, with $(\Psi_\alpha)'(x(\alpha)) = 0$, as claimed.

From [5], Proposition 2.3.2 and Theorem 2.4.2 one can easily derive that $x(\alpha)$ can be chosen to satisfy $x(0) = x^*$ and to be differentiable in α. The point $x(\alpha)$ solves the equation

$$\Phi'(x) = (1+\alpha)\Phi'(x^*), \tag{104}$$

and therefore

$$\Phi''(x(\alpha))x'(\alpha) = \Phi'(x^*). \tag{105}$$

Now let

$$g(\alpha) = \Psi_\alpha(x(\alpha)).$$

We have

$$\begin{array}{rcl}
g'(\alpha) & = & [x'(\alpha)]^T[\Phi'(x(\alpha)) - (1+\alpha)\Phi'(x^*)] - (x(\alpha)-x^*)^T\Phi'(x^*) \\
& = & -(x(\alpha)-x^*)^T\Phi'(x^*) \quad [\text{since } x = x(\alpha) \text{ is a solution to (104)}],
\end{array} \tag{106}$$

whence

$$\begin{array}{rcll}
g''(\alpha) & = & -[x'(\alpha)]^T\Phi'(x^*) & \\
& = & \min\limits_h [2h^T\Phi'(x^*) + h^T\Phi''(x(\alpha))h] & [\text{see (105)}] \\
& = & (1+\alpha)^{-2}\min\limits_h [2h^T\Phi'(x(\alpha)) + h^T\Phi''(x(\alpha))h] & [\text{see (105)}] \\
& \geq & -(1+\alpha)^{-2}\theta & [\text{since } \Phi \text{ is a } \theta\text{-s.-c.b.}].
\end{array} \tag{107}$$

Since $x(0) = x^*$, we get from (106), (107)

$$g(0) = 0; \quad g'(0) = 0; \quad g''(\alpha) \geq -\theta(1+\alpha)^{-2},$$

whence

$$\alpha \geq 0 \Rightarrow g(\alpha) \geq -\theta[\alpha - \ln(1+\alpha)].$$

Consequently, (103) ensures that

$$\inf_{x \in X} [t^+\phi(x) + \Phi(x)] \geq [t^+\phi(x^*) + \Phi(x^*)] - \theta[\alpha - \ln(1+\alpha)].$$

Combining this result with (102), we come to

$$f_{t^+}(\bar{z}) - \inf_{x \in X} f_{t^+}(x, \bar{y}) \leq 0.12\alpha\sqrt{\theta} + 0.006(1+\alpha) + \theta[\alpha - \ln(1+\alpha)].$$

Symmetric reasoning implies that

$$\sup_{y \in Y} f_{t^+}(\bar{x}, y) - f_{t^+}(\bar{z}) \leq 0.12\alpha\sqrt{\theta} + 0.006(1+\alpha) + \theta[\alpha - \ln(1+\alpha)].$$

and (27) follows. ∎

8.16 Proof of Lemma 6.1

Let us first prove the existence and the uniqueness of the saddle point $z^*(t)$. Since f_t is a nondegenerate s.-c.c.-c. function, it suffices to verify that $f_t(\cdot, \widehat{y})$ is below bounded on X, and $f(\widehat{x}, \cdot)$ is above bounded on Y (Proposition 2.5). By symmetry, we may prove the first statement only. Since $g(x) = f(x, \widehat{y})$ has bounded level sets and is convex on X, it admits lower bound of the type $g(x) \geq a + b \parallel x \parallel_2$, $x \in X$, with $b > 0$, while for a ϑ-s.-c.b. F for $\mathrm{cl}\, X$ we have (see [5], (2.3.3))

$$x \in X \Rightarrow F(x) \geq F(x_0) + \vartheta \ln(1 - \pi_x(x_0)),$$

where $x_0 \in X$ is arbitrary and

$$\pi_x(x_0) = \inf\{t > 0 \mid x + t^{-1}(x_0 - x) \in X\}.$$

For a once for ever fixed $x_0 \in X$, the quantity $1 - \pi_x(x_0)$ is of course bounded from below by a function of the type $c/(1+ \parallel x \parallel_2)$. Thus,

$$f_t(x, \widehat{y}) \geq \gamma[tb \parallel x \parallel_2 - \vartheta \ln(1+ \parallel x \parallel_2) + \mathrm{const}(t)],$$

and the right hand side in this inequality is below bounded for every $t > 0$.

The fact that the path $z^*(t)$ is continuously differentiable is readily given by the Implicit Function Theorem (recall that the Hessian of a nondegenerate s.-c.c.-c. function is nondegenerate, see the proof of Proposition 3.1.(iii)). ∎

8.17 Proof of Theorem 6.2

1^0. To prove (31), we start with the following observation:

Lemma 8.5 *Let $\bar{y} = \widehat{y}(\bar{t}, \bar{x})$. Then the Newton decrement $\omega(f_{\bar{t}}, (\bar{x}, \bar{y}))$ of the s.-c.c.-c. function f_t at the point $\bar{z} = (\bar{x}, \bar{y})$ is equal to the Newton decrement $\lambda(\Phi(\bar{t}, \cdot), \bar{x})$ of the convex s.-c. function $\Phi(\bar{t}, \cdot)$ at the point $\bar{x}$. Moreover, the Newton iterate $\widetilde{x}$ of $\bar{x}$ given by (30) is exactly the x-component of the pair*

$$z^+ = \bar{z} - [\nabla^2 f_{\bar{t}}(\bar{z})]^{-1} \nabla f_{\bar{t}}(\bar{z}).$$

Proof. Denoting $\Phi(t,\cdot)=\Psi(\cdot)$, $f_{\bar t}\equiv\phi$, let $\phi''(\bar z)=\begin{pmatrix} A & B^T\\ B & -C\end{pmatrix}$ and $g=\phi'_x(\bar z)$. Since $\phi'_y(\bar z)=0$, the Newton direction $e=e(\phi,\bar z)=\begin{pmatrix} p\\ q\end{pmatrix}$ of ϕ at $\bar z$ is the solution to the system of equations

$$Ap+B^Tq=g,\quad Bp-Cq=0,$$

whence

$$p=(A+B^TC^{-1}B)^{-1}g,\quad q=C^{-1}B(A+B^TC^{-1}B)^{-1}g.$$

Since $g=\Psi'(\bar x)$ and $\Psi''(\bar x)=A+B^TC^{-1}B$ (see (56)), we get

$$\widetilde{x}=\bar x-p,\ \omega^2(f,\bar z)=p^TAp+q^TCq=g^T(A+B^TC^{-1}B)^{-1}g=[\Phi'(\bar x)]^T[\Phi''(\widehat{x})]^{-1}\Phi'(x).$$ ■

Lemma combined with (22) implies that

$$\nu(f_{\bar t},z^+)\le\frac{\lambda^2(\Phi(\bar t,\cdot),\bar x)}{(1-\lambda(\Phi(\bar t,\cdot),\bar x))^2}\le\frac{(0.1)^2}{(0.9)^2}\le 0.1. \tag{108}$$

By Proposition 6.3 it follows that

$$\sup_{y\in Y} f(\widetilde{x},y)\le\inf_{x\in X} f(x,y^+)+\frac{4\vartheta}{\bar t}, \tag{109}$$

where y^+ is the y-component of z^+. Since $\inf\limits_{x\in X} f(x,y^+)\le\inf\limits_{x\in X}\sup\limits_{y\in Y} f(x,y)$, (31) follows.

2^0. Let us prove the bound (32). Since $\Phi(t,\cdot)$ is a s.-c. convex below bounded function on X, the number of damped Newton steps in the updating $(\bar t,\bar x)\mapsto(t^+,x^+)$ is bounded from above by $O(1)\left[\Delta(t^+)+\ln\ln\frac{1}{\kappa}\right]$, where

$$\Delta(t)=\Phi(t,\widetilde{x})-\min_{x'\in X}\Phi(t,x')$$

and $O(1)$ is an absolute constant (see [5], Theorem 2.2.3). Thus, all we need is to verify that

$$t\ge\bar t\Rightarrow\Delta(t)\le\rho(\kappa)+\frac{3}{2}\left(1+\sqrt{\gamma\vartheta}\right)\frac{t-\bar t}{\bar t}++3\gamma\vartheta\left[\frac{t-\bar t}{\bar t}-\ln\frac{t}{\bar t}\right]. \tag{110}$$

Let $z^*(t)=(x_*(t),y_*(t))$ be the trajectory of saddle points of the functions f_t; note that $z^*(\cdot)$ is continuously differentiable by Lemma 6.1. As we remember,

$$\Phi(t,x)=\gamma\left[tf(x,\widehat{y}(t,x))+F(x)-G(\widehat{y}(t,x))\right],$$

with continuously differentiable function $\widehat{y}(t,x)$ satisfying the relation

$$y=\widehat{y}(t,x)\Rightarrow tf'_y(x,y)-G'(y)=0.$$

From these relations we get (values of all functions and their derivatives are taken at the point $(t,x,\widehat{y}(t,x))$):

$$\begin{array}{rcl}\widehat{y}'_t &=& [G''-tf''_{yy}]^{-1}f'_y=[G''-tf''_{yy}]^{-1}t^{-1}G',\\ \Phi'_t &=& \gamma f,\\ \Phi''_{tt} &=& \gamma[f'_y]^T\widehat{y}'_t=\gamma t^{-1}[G']^T\widehat{y}'_t=\gamma t^{-2}[G']^T[G''-tf''_{yy}]^{-1}G'.\end{array} \tag{111}$$

Taking into account that G is a ϑ-s.c-c.b. and that f is concave in y, we come to

$$0\le\Phi''_{tt}(t,x)\le\gamma\vartheta t^{-2}. \tag{112}$$

Now, setting $\Psi(x,y)=F(x)-G(y)$ we have (values of all functions and their derivatives are taken at the point $(t,z^*(t))$):

$$t\nabla_z f+\nabla_z\Psi=0\Rightarrow\frac{d}{dt}z^*(t)=-[t\nabla_z^2 f+\nabla_z^2\Psi]^{-1}\nabla_z f\Rightarrow\frac{d}{dt}z^*(t)=t^{-1}[t\nabla_z^2 f+\nabla_z^2\Psi]^{-1}\nabla_z\Psi,$$

whence

$$
\begin{array}{rcll}
\frac{d}{dt}\Phi(t,x_*(t)) & = & \gamma f, & \\
-\frac{d^2}{dt^2}\Phi(t,x_*(t)) & = & -\gamma[\nabla_z f]^T\frac{d}{dt}z^*(t)=\gamma t^{-2}[\nabla_z\Psi]^T[t\nabla_z^2 f+\nabla_z^2\Psi]^{-1}\nabla_z\Psi & \\
& \le & t^{-2}\gamma^2[\nabla_z\Psi]^T[\mathrm{S}_{f_t}(z^*(t))]^{-1}\nabla_z\Psi \qquad \text{[see Lemma 8.2]} & \\
& \le & 2t^{-2}\gamma\vartheta, &
\end{array}
\tag{113}
$$

the concluding inequality being given by $\mathrm{S}_{f_t}\ \ge\ \gamma\begin{pmatrix}F'' & \\ & G''\end{pmatrix}$, $\nabla_z\Psi\ =\ \begin{pmatrix}F'\\ -G'\end{pmatrix}$ and the fact that F,G are ϑ-s.-c.b.'s.

Since $\lambda(\Phi(\bar{t},\cdot),\bar{x})\le\kappa\le 0.1$, we have

$$\Phi(\bar{t},\widetilde{x})\le\Phi(\bar{t},\bar{x})\le\min_{x\in X}\Phi(\bar{t},x)+\rho(\kappa)$$

(see [5], Theorem 2.2.1, and (6)). Consequently,

$$\Delta(\bar{t})\le\rho(\kappa). \tag{114}$$

Besides this,

$$\Delta'(t)=\Phi'_t(t,\widetilde{x})-\frac{d}{dt}\Phi(t,x_*(t)),$$

whence in view of (111), (112), (113)

$$\Delta'(t)=\gamma\left[f(\widetilde{x},\widehat{y}(t,\widetilde{x}))-f(z^*(t))\right] \tag{115}$$

and

$$\Delta''(t)=\Phi''_{tt}(t,\widetilde{x})-\frac{d^2}{dt^2}\Phi(t,x_*(t))\le 3\gamma t^{-2}\vartheta. \tag{116}$$

3^0. In view of the inequalities (114), (116) all we need to bound $\Delta(t^+)$ from above is to bound from above the quantity

$$\Delta'(\bar{t})=\gamma\left[f(\widetilde{x},\widetilde{y})-f(z^*(\bar{t}))\right],\quad \widetilde{y}=\widehat{y}(\bar{t},\widetilde{x}). \tag{117}$$

(see (115)).

3^0.1). For the sake of brevity, let $g(x,y)=f_{\bar{t}}(x,y)$, $z^*\equiv(x_*,y_*)=z^*(\bar{t})$, and let

$$d_x=\sqrt{[\widetilde{x}-x_*]^T g''_{xx}(z^*)[\widetilde{x}-x_*]},\quad d_y=\sqrt{-[y^+-y_*]g''_{yy}(z^*)[y^+-y_*]}.$$

By (108), $\nu(g,z^+)<0.1$; applying (21.b), we get $\mu(g,z^+)\le\rho(0.1)\le 0.01$, whence (see (23))

$$\sqrt{d_x^2+d_y^2}\le 2\left[\mu(g,z^+)+\sqrt{\mu(g,z^+)}\right]\le 0.5. \tag{118}$$

Since $\mu(g,z^+)\equiv\sup\limits_y g(\widetilde{x},y)-\inf\limits_x g(x,y^+)\le 0.01$, the quantity $g(z^+)$ differs from the saddle value $g(z^*)$ of g by at most 0.01:

$$|g(z^+)-g(z^*)|\equiv\gamma\left|\bar{t}[f(z^+)-f(z^*)]+F(\widetilde{x})-F(x_*)+G(y_*)-G(y^+)\right|\le 0.01. \tag{119}$$

We have $g''_{xx}(z^*)\succeq\gamma F''(x_*)$, so that $(\widetilde{x}-x_*)^T[\gamma F''(x_*)](\widetilde{x}-x_*)\le d_x^2$. Therefore

$$
\begin{array}{rcll}
\gamma|F(\widetilde{x})-F(x_*)| & \le & |(\widetilde{x}-x_*)^T(\gamma F'(x_*))|+\rho(d_x) & \text{[(4) as applied to }\gamma F\text{]}\\
& \le & \sqrt{\gamma\vartheta}d_x+\rho(d_x) & \text{[since }\gamma F\text{ is }(\gamma\vartheta)\text{-s.-c.b.]}\\
& \le & 0.5\sqrt{\gamma\vartheta}+0.45 & \text{[see (118)]}
\end{array}
$$

and similarly $\gamma|G(y^+)-G(y_*)|\le 0.45\sqrt{\gamma\vartheta}+0.5$. Thus, (119) implies that

$$\gamma|f(z^*)-f(z^+)|\le\frac{1+\sqrt{\gamma\vartheta}}{\bar{t}}. \tag{120}$$

$3^0.2$). Now consider the self-concordant function $\phi(y) = -\gamma\,[\bar{t}f(\widetilde{x}, y) + F(\widetilde{x}) - G(y)]$. This function attains its minimum on Y at $y = \widetilde{y}$, and from $\mu(g, z^+) \leq 0.01$ it follows that $\phi(y^+) - \min\limits_y \phi(y) \leq 0.01$. Denoting $\delta = \sqrt{[y^+ - \widetilde{y}]^T \phi''(\widetilde{y})[y^+ - \widetilde{y}]}$ and applying (4), we get $\rho(-\delta) \leq 0.01$, whence $\delta \leq 0.5$. We have

$$0.01 \geq \phi(y^+) - \phi(\widetilde{y}) = \gamma \left|\bar{t}[f(\widetilde{x}, y^+) - f(\widetilde{x}, \widetilde{y})] - [G(y^+) - G(\widetilde{y})]\right|,$$

and from $\delta \leq 0.5$, same as in $3^0.1$), it follows that $\gamma|G(y^+) - G(\widetilde{y})| \leq 0.5\sqrt{\gamma\vartheta} + 0.45$. Combining these observations, we get

$$\gamma|f(z^+) - f(\widetilde{x}, \widetilde{y})| \leq \frac{0.5 + 0.5\sqrt{\gamma\vartheta}}{\bar{t}},$$

which together with (120) and (117) implies that

$$\Delta'(\bar{t}) \leq \frac{3(1 + \sqrt{\gamma\vartheta})}{2\bar{t}}. \tag{121}$$

Relations (114), (116) and (121) imply (110). ∎

8.18 Proof of Lemma 7.1

Let $h \in \mathbf{R}^m, r \in \mathbf{R}^n$ be two arbitrary vectors, and let

$$\begin{array}{ll} Y_0 = \mathrm{Diag}(y^0), \quad H = Y_0^{-1}\mathrm{Diag}(h), \quad X_0 = X(\xi_0), \\ R = X_0^{-1}\mathrm{Diag}(Er), \quad P = Y_0^{1/2}X_0^{-1/2}E(E^TY_0X_0^{-1}E)^{-1}E^TX_0^{-1/2}Y_0^{1/2}. \end{array}$$

Note that P is an orthoprojector.

We have

$$\begin{array}{rcl}
h^Tf'_y(z^0) & = & \mathrm{Tr}\left((E^TY_0X_0^{-1}E)^{-1}E^T\mathrm{Diag}(h)X_0^{-1}E\right) - 2h^Tx(\xi^0) \\
& = & \mathrm{Tr}\left((E^TY_0X_0^{-1}E)^{-1}E^TY_0X_0^{-1}HE\right) - 2\mathrm{Tr}(X_0Y_0H) \\
& = & \mathrm{Tr}(PH) - 2\mathrm{Tr}(X_0Y_0H); \\
r^Tf'_\xi(z^0) & = & \mathrm{Tr}\left((E^TY_0X_0^{-1}E)^{-1}E^TY_0X_0^{-1}\mathrm{Diag}(Er)X_0^{-1}E\right) + 2(y^0)^TEr \\
& = & \mathrm{Tr}\left((E^TY_0X_0^{-1}E)^{-1}E^TY_0X_0^{-1}RE\right) + 2\mathrm{Tr}(Y_0X_0R) \\
& = & \mathrm{Tr}(PR) + 2\mathrm{Tr}(X_0Y_0R); \\
h^TG''(y^0)h + r^TF''(\xi^0)r & = & h^TY_0^{-2}h + r^TE^TX_0^{-2}Er \\
& = & \mathrm{Tr}(H^2 + R^2); \\
h^TG'(y^0) & = & -\mathrm{Tr}(H); \\
r^TF'(\xi^0) & = & \mathrm{Tr}(R).
\end{array}$$

Consequently,

$$\begin{array}{rcl}
\gamma^{-1}(r^T, h^T)(f_{t^0})'(z^0) & = & t^0\mathrm{Tr}(PH) + t^0\mathrm{Tr}(PR) - 2t^0\mathrm{Tr}(Y_0X_0H) + 2t^0\mathrm{Tr}(Y_0X_0R) + \mathrm{Tr}(H) + \mathrm{Tr}(R) \\
& = & t^0\mathrm{Tr}(PH) + t^0\mathrm{Tr}(PR) + 2\mathrm{Tr}(R)
\end{array} \tag{122}$$

(note that by construction $2t^0Y_0X_0 = I$, I being the unit $m \times m$ matrix). In view of (122) and due to the fact that P is an orthoprojector one has

$$\left|(r^T, h^T)(f_{t^0})'(z^0)\right| \leq \gamma t^0\sqrt{2m\mathrm{Tr}(H^2 + R^2)} + 2\gamma|\mathrm{Tr}(R)|,$$

while in view of (37)

$$|\mathrm{Tr}(R)| = |r^TF'(\xi^0)| \leq \frac{0.05}{2\sqrt{\gamma}}\sqrt{r^TF''(\xi^0)r} = \frac{0.05}{2\sqrt{\gamma}}\sqrt{\mathrm{Tr}(R^2)};$$

thus,

$$\left|(r^T, h^T)(f_{t^0})'(z^0)\right| \leq \gamma t^0\sqrt{2m\mathrm{Tr}(H^2 + R^2)} + 0.05\sqrt{\gamma}\sqrt{\mathrm{Tr}(R^2)} \leq 0.1\sqrt{\gamma}\sqrt{\mathrm{Tr}(H^2 + R^2)}$$

(see the definition of t^0). On the other hand,

$$(r^T, h^T)\mathrm{S}_{f_{t^0}}(z^0)\begin{pmatrix} r \\ h \end{pmatrix} \geq \gamma[h^T G''(y^0)h + r^T F''(\xi^0)r] = \gamma \mathrm{Tr}(H^2 + R^2),$$

and we conclude that for all $h \in \mathbf{R}^m, r \in \mathbf{R}^n$ one has

$$\left|(r^T, h^T)(f_{t^0})'(z^0)\right| \leq 0.1\sqrt{(r^T, h^T)\mathrm{S}_{f_{t^0}}(z^0)\begin{pmatrix} r \\ h \end{pmatrix}},$$

whence $\nu(f_{t^0}, z^0) \leq 0.1$, as claimed. ∎

8.19 Proof of Proposition 7.4

Let vol be the normalization of Vol which makes the normalized volume of the unit Euclidean ball equal to 1; of course, it suffices to prove (40) for the normalized volume vol instead of the standard volume Vol.

Let us set

$$\Phi(\lambda) \quad = \quad \gamma t f(\xi, \lambda), \quad \phi(\lambda) \quad = \quad \gamma \sum_{i=1}^{m} \ln \lambda_i, \quad H(\lambda) \quad = \quad -[\Phi(\lambda) + \phi(\lambda)].$$

Note that since $-f_t(\xi, \cdot)$ differs from $H(\cdot)$ by a constant, the function H is s.-c. on $\mathbf{R}^m_{++}$, and its Newton decrement at y (see Definition 2.1) is majorated by $\nu(f_t, z)$. Thus, $\lambda(H, y) \leq \delta \leq 0.1$, whence by [5], Theorem 2.2.2, there exists $y^* \in \mathbf{R}^m_{++}$ such that

$$\| y^* - y \|_{H''(y)} \leq 10\delta, \quad H'(y^*) = 0. \tag{123}$$

Let us set

$$\begin{array}{l} Y_* = \mathrm{Diag}(y^*), \quad B_* = (E^T Y_* X^{-1} E)^{-1}, \quad A_* = 2^{-1/2} B_*^{1/2}, \\ d_i = y_i^*/x_i(\xi), \ i = 1, ..., m, \quad D = \mathrm{Diag}(d_1, d_2, ..., d_m). \end{array}$$

The second relation in (123), after straightforward computation, implies that

$$e_i^T B_* e_i = 2x_i^2(\xi) - t^{-1} d_i^{-1}, \ i = 1, ..., m, \tag{124}$$

whence

$$e_i^T A_*^2 e_i = x_i^2(\xi) - (2t)^{-1} d_i^{-1}, \ i = 1, ..., m. \tag{125}$$

Since $d_i > 0$ for all i, we conclude that the ellipsoid

$$W_* = \{\xi + A_* u \mid u^T u \leq 1\}$$

is contained in Π.

In view of (124) we have

$$\begin{array}{rcl} n & = & \mathrm{Tr}(B_* B_*^{-1}) = \mathrm{Tr}(B_* E^T D E) = \mathrm{Tr}([E B_* E^T] D) \\ & = & \sum\limits_{i=1}^{m} e_i^T B_* e_i d_i = \sum\limits_{i=1}^{m} (2x_i^2(\xi) - t^{-1} d_i^{-1}) d_i = \sum\limits_{i=1}^{m} 2 y_i^* x_i(\xi) - m t^{-1}, \end{array}$$

whence

$$f(\xi, y^*) = -\ln \mathrm{Det} B_* - 2(y^*)^T x(\xi) = -\ln \mathrm{Det} B_* - n - m t^{-1}.$$

On the other hand, due to Proposition 7.1 and to the second relation in (123) we have

$$\mathcal{V}(\xi) = \frac{n \ln 2 + n}{2} + \frac{1}{2} f(\xi, y^*).$$

Combining our observations, we get

$$\mathcal{V}(\xi) = \frac{n\ln 2 + n}{2} - \frac{1}{2}\ln \mathrm{Det} B_* - \frac{n}{2} - \frac{m}{2t} = -\ln \mathrm{Det} A_* - \frac{m}{2t}. \tag{126}$$

Let $\mathcal{V}^*$ be the infimum of $\mathcal{V}(\cdot)$ over $\mathrm{int}\,\Pi$. Proposition 7.3 states that

$$\mathcal{V}(\xi) \leq \mathcal{V}^* + \frac{2m}{t}.$$

From this inequality and (126) we conclude that

$$\ln\ \mathrm{vol}(W_*) = \ln \mathrm{Det} A_* \geq -\mathcal{V}^* - \frac{5m}{2t}. \tag{127}$$

Now, from the first relation in (123) and from the fact that $\Phi''(y) \leq 0$ it follows that

$$\sqrt{\gamma \sum_{i=1}^{m} (y_i^* - y_i)^2 y_i^{-2}} = \parallel y^* - y \parallel_{-\phi''(y)} \leq 10\delta,$$

whence

$$(1 - 10\delta) y_i \leq y_i^* \leq (1 + 10\delta) y_i,\ \ i = 1, ..., m,$$

so that

$$(1 - 10\delta)^{-1} B \succeq B_* \succeq (1 + 10\delta)^{-1} B, \tag{128}$$

i.e.,

$$\parallel A_* u \parallel_2 \geq \parallel \widehat{A} u \parallel_2 \quad \forall u.$$

From this inequality and (125) it follows that $W \subset \Pi$. Finally, we have

$$\begin{array}{rcl}
\ln \mathrm{Det}\widehat{A} & = & -\frac{n}{2}\ln(1+10\delta) + \ln \mathrm{Det} A = -\frac{n}{2}\ln(1+10\delta) - \frac{n\ln 2}{2} + \frac{1}{2}\ln \mathrm{Det} B \\
& \geq & -\frac{n}{2}\ln(1+10\delta) - \frac{n\ln 2}{2} + \frac{1}{2}\ln \mathrm{Det} B_* + \frac{n}{2}\ln(1-10\delta) \\
& & \qquad\qquad\qquad\qquad [\text{we have used the first inequality in (128)}] \\
& = & -\frac{n}{2}\ln\left(\frac{1+10\delta}{1-10\delta}\right) + \ln \mathrm{Det} A_* \\
\Rightarrow \quad \ln\ \mathrm{vol}(W) & = & \ln \mathrm{Det}\widehat{A} \\
& \geq & \ln\ \mathrm{vol}(W_*) - \frac{n}{2}\ln\left(\frac{1+10\delta}{1-10\delta}\right) \geq -\mathcal{V}^* - \frac{5m}{2t} - \frac{n}{2}\ln\left(\frac{1+10\delta}{1-10\delta}\right) \qquad [\text{see (127)}] \\
& \geq & -\mathcal{V}^* - \epsilon.
\end{array}$$

It remains to recall that $(-\mathcal{V}^*)$ is the logarithm of the normalized volume of the maximal ellipsoid contained in Π. ■