Examples of ill-behaved central paths in convex optimization

J. Charles Gilbert *   Clóvis C. Gonzaga †   Elizabeth Karas ‡

April, 2001, revised August, 2002.

Abstract

This paper presents some examples of ill-behaved central paths in convex optimization. Some contain infinitely many fixed length central segments; others manifest oscillations with infinite variation. These central paths can be encountered even for infinitely differentiable data.

1 Introduction

Penalization approaches have been central in the whole evolution of optimization techniques. Constrained problems are often solved by a sequence of resolutions of penalized problems, which depend on a single penalization parameter μ. The optimizers associated with each value of μ define a "path of optimizers", which should be well-behaved and lead to a solution to the original problem.

This approach is developed in the classical book by Fiacco and McCormick [12], who describe external, internal, and mixed penalty methods for nonlinear programming. Internal penalty methods, also known as barrier methods, had an explosive development in the last fifteen years, due to the success of interior point methods for linear programming and for linear complementarity problems (see the monographs [10, 18, 19, 23, 25, 26, 31]; see also the extensions to nonlinear programming in [4, 9, 11, 14, 15, 27]).The first deep study of the path of optimizers, now known as central path, is due to Bayer and Lagarias [3] and to Megiddo [21], who gave a definitive characterization of the primal-dual central path. An introduction to path-following methods is given by Gonzaga [16].

We know that for linear programming the path is well defined, infinitely differentiable and has a bounded length, associated with the complexity of path-following methods. The limiting behavior of the path derivatives is described

*INRIA Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France, Jean-Charles.Gilbert@inria.fr.

†Department of Mathematics, Federal University of Santa Catarina, Florianópolis, Brazil, clovis@mtm.ufsc.br. Research partially supported by CAPES, Brazil.

‡Department of Mathematics, Federal University of Paraná, Curitiba, Brazil, karas@mat.ufpr.br. Research partially supported by CAPES and CNPq, Brazil.

by Adler and Monteiro [1] and Witzgall et al. [30], who show that the path approaches the optimal face with a well-defined inclination. Similar results are obtained for convex problems by McLinden [20]: he shows that under reasonable hypotheses, which include strict complementarity, the central path converges to the "analytic center" of the optimal set for the problem. Vavasis and Ye [28] show that for linear programs in $\mathbb{R}^n$ the central path is composed of a sequence of no more than n^2 alternated curved and "almost straight" sectors.

From these results, it seems reasonable to expect smooth and calm sets of optimizers in convex programming problems. In this paper, however, we show that this is not necessarily true. We give examples of convex problems for which the path associated with any penalty function is undesirably weird.

We shall work with an extremely simple region in $\mathbb{R}^2$ and construct increasingly complex objective functions. First, we exhibit a continuous convex function giving rise to an "antenna-like" central path, containing an infinite number of horizontal segments of constant length. The second example is obtained by slightly perturbing the first one and results in a zig-zagging central path with infinite variation. We proceed by smoothing both functions, and produce the same behaviors as before for a differentiable objective function, and finally for a C^∞ function.

We conclude that convexity, even when added to differentiability of any degree, is not a sufficiently strong property to ensure reasonably attractive central paths.

Finally, we study the effect of these results on the complexity of algorithms. We prove that no penalized function constructed as in this paper can be self-concordant, and thus it is not possible to prove polynomiality of any algorithm using Nesterov and Nemirovskii's theory. Assuming that a nice path following algorithm exists, overcoming each turn of the zig-zags in a fixed number of iterations, then this algorithm will be polynomial for our examples, but a slight change in the problem statement makes the path following algorithm converge in infinite time to a non-optimal set.

Stating the problem in Nesterov-Nemirovskii's format and defining the central path by means of a self-concordant barrier, we show the following results: for the examples in which we obtain the antenna central trajectories, the path will now be a straight line. For the zig-zag examples, the trajectories will be damped zig-zags in some cases, but we conjecture (and explain the reason for our guess) that in the simplest example of zig-zag behavior (Example 3), the self-concordant barrier will still generate a zig-zag with infinite length, for which good path following algorithms will be polynomial, which is amazing.

2 Background

We are interested in nonlinear programming problems with the shape

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & c_i(x) \geq 0, \qquad i = 1, \ldots, m, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $(-c_i) : \mathbb{R}^n \to \mathbb{R}$ are closed convex functions.

A penalty method is based on some penalty function $p : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ and proceeds by solving problems of the form

$$\min\; f(x) + \mu\, p(c(x)), \tag{1}$$

for some penalty parameter $\mu > 0$. If these problems are well-defined, the image of the set-valued map ($\mathbb{R}_{++}$ denotes the set of positive real numbers)

$$\mu \in \mathbb{R}_{++} \mapsto \chi(\mu) := \operatorname{argmin}\{f(x) + \mu\, p(c(x))\}$$

is called the *central path*. There is also a *dual* central path, but this paper only deals with the *primal* central path defined above.

Interior point methods use the logarithmic barrier function, defined for $y \in \mathbb{R}^m_{++}$ (the set of m-vectors with positive components) by

$$p(y) = -\sum_{i=1}^{m} \log\, y_i.$$

In that case $\chi(\mu) = \operatorname{argmin}\{f(x) - \mu \sum_{i=1}^m \log(c_i(x)) \mid c(x) > 0\}$. Our examples do not depend on a specific penalty function, however.

A simple way of visualizing the central path is the following. Assume that $\bar{x} \in \chi(\mu)$ for some $\mu > 0$. Then $\bar{x}$ also solves

$$\begin{array}{ll} \text{minimize} & f(x) + \mu p(c(x)) \\ \text{subject to} & p(c(x)) = p(c(\bar{x})), \end{array}$$

or equivalently

$$\text{minimize}\ \{f(x) \mid p(c(x)) = p(c(\bar{x}))\}.$$

Pictorially, this means that central points are points where f and $(p \circ c)$ have tangent level curves.

Let us particularize this to a very simple 2-dimensional problem. Let $F : z = (x, y) \in \mathbb{R}^2 \to \mathbb{R}$ be a closed convex function. We study problems like

$$\begin{array}{ll} \text{minimize} & F(x, y) \\ \text{subject to} & y \geq 0, \end{array} \tag{2}$$

and assume that its solutions are on the x-axis $\{(x, 0) : x \in \mathbb{R}\}$. For this extremely simple feasible set, most penalty functions (e.g., those studied in [2]) give the same sets of optimizers, those described by

$$\operatorname{argmin}\{F(x, y) \mid y = \text{constant}\}. \tag{3}$$

Depending on whether the penalties are internal or external, y will assume positive or negative values. We consider here the case of interior point methods and parameterize the central path by $y > 0$:

$$y \in \mathbb{R}_{++} \mapsto \chi(y) = \operatorname*{argmin}_{x \in \mathbb{R}} F(x, y). \tag{4}$$

We denote by $\{e^1, e^2\}$ the canonical basis of $\mathbb{R}^2$ and by $\mathbb{N}$ the set of nonnegative integer numbers.

3 Examples with continuous objectives

This section is devoted to the construction of examples of weird central paths associated with continuous convex objectives. After two trivial examples, we construct Example 3, which has a cross-shaped central path. The constructions made for this example are then used as building blocks for Example 4, the most interesting one in the paper: here we generate central paths shaped as a TV antenna with an infinity of branches or as a zig-zag with unbounded variation (see Figure 3).

Let us start with two simple examples of problems with obvious central paths.

Example 1 The function F is defined by $(x,y) \in \mathbb{R}^2 \mapsto F(x,y) = a\,y + b$, where $a > 0$ and $b \in \mathbb{R}$. In this case, $\chi(y) = \mathbb{R}$ for any $y > 0$. □

Example 2 The function F is defined by $(x,y) \in \mathbb{R}^2 \mapsto F(x,y) = a\,y + b + \varepsilon\,x$, where $a > 0$, b and $\varepsilon \in \mathbb{R}$, with $\varepsilon \neq 0$. Now, $\chi(y) = \emptyset$ for any $y > 0$. □

We now perturb the functions in Examples 1 and 2 to obtain the first non-trivial example. Let $y_k > 0$ be a given number (the index $k \in \mathbb{N}$ will be useful in a moment) and define the function g_k^0 by

$$z = (x,y) \in \mathbb{R}^2 \mapsto g_k^0(x,y) = \frac{1}{2}\left(\|z - z_k^-\| + \|z - z_k^+\| - 2\right), \tag{5}$$

where $\|\cdot\|$ denotes the 2-norm, and

$$z_k^- = (-1, y_k) \qquad \text{and} \qquad z_k^+ = (1, y_k).$$

Function g_k^0 is convex.

Its set of minimizers is the segment

$$[z_k^-, z_k^+] = \{(1-t)z_k^- + tz_k^+ : t \in [0,1]\},$$

on which g_k^0 takes the value 0. This function is also differentiable everywhere but at the points z_k^- and z_k^+. Note that, since $\|\cdot\| \leq \|\cdot\|_1$ (the 1-norm), $g_k^0(x,y) \leq \frac{1}{2}(|x-1| + |x+1| + 2|y - y_k| - 2)$, so that

$$g_k^0(x,y) \leq \begin{cases} |y - y_k| & \text{if } |x| \leq 1, \\ |x| + |y - y_k| - 1 & \text{otherwise.} \end{cases} \tag{6}$$

In our constructions, we use convex functions g_k that are smoothings of g_k^0, satisfying

$$0 \leq g_k(z) \leq g_k^0(z), \tag{7}$$

for z in a large region.

We can now give our first non trivial example, the function f_k. It is obtained by adding to the function of Example 2 (with a, b, and ε indexed by k), the convex perturbation function g_k just described, multiplied by a positive constant c_k, and the constant $-a_k y_k$.

Example 3 Let a_k, b_k, c_k, and y_k be positive parameters and $\varepsilon_k \in I\!R$. Then, function f_k is defined by

$$z \in I\!R^2 \mapsto f_k(z) = a_k(y - y_k) + b_k + c_k g_k(z) + \varepsilon_k x. \tag{8}$$

where the convex function $g_k : I\!R^2 \to I\!R$ satisfies (7) with g_k^0 given by (5). $\square$

To get familiar with function f_k, which will be used continually, consider two particular cases with $g_k \equiv g_k^0$.

- If $\varepsilon_k = 0$, we have Example 1 with an added perturbation $c_k g_k^0(\cdot)$. Let c_k be positive, but small enough such that problem (2) has its solutions on the x-axis. We easily see using (4) that for any $y > 0$, $\chi(y) = \{0\}$ if $y \neq y_k$, and $\chi(y_k) = [-1, +1]$. The central path is the cross, depicted on the left in Figure 1.

- Suppose now that ε_k is nonzero, but small in module, $|\varepsilon_k| < c_k$ say. For $y \neq y_k$, $x \mapsto f_k(x, y)$ is strictly convex with compact level sets, so that $\chi(y)$ is a singleton. Suppose now that $y = y_k$. Observe that $f_k'(z_k^-; -e^1) = c_k - \varepsilon_k$, $f_k'(z_k^-; e^1) = \varepsilon_k$, $f_k'(z_k^+; -e^1) = -\varepsilon_k$, and $f_k'(z_k^+; e^1) = c_k + \varepsilon_k$. Therefore, the inequality $|\varepsilon_k| < c_k$ implies that $\chi(y_k) = \{-1\}$ if $\varepsilon_k > 0$ and $\chi(y_k) = \{+1\}$ if $\varepsilon_k < 0$. The central path is perturbed sideways, like on the right in Figure 1: the path deviates to the left or to the right respectively if $\varepsilon_k > 0$ or $\varepsilon_k < 0$.

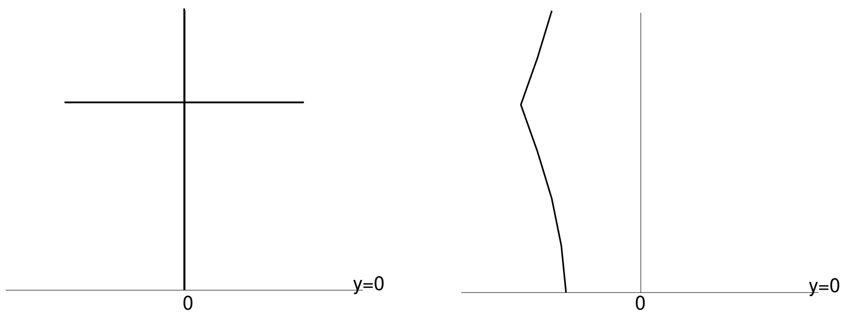

Figure 1: The central path for Example 3 with $g_k = g_k^0$, and $\varepsilon_k = 0$ (left) or $\varepsilon_k > 0$ (right)

The remarkable examples of this paper are constructed with infinitely many copies of the function f_k, defined in Example 3, with various coordinates y_k and coefficients a_k, b_k, c_k, and ε_k. These functions are tangent to a *support function* $(x, y) \in I\!R \times [-0.5, 1] \mapsto \psi(y)$, where $\psi : [-0.5, 1] \to I\!R$ has the following properties:

$$\begin{cases} \psi \text{ is continuously differentiable,} \\ \psi(y) = 0, \text{ for } y \in [-0.5, 0], \\ y \in [0, 1] \mapsto \psi(y) \text{ is strictly convex.} \end{cases} \tag{9}$$

Clearly, these properties imply that $\psi(y) > 0$ and $\psi'(y) > 0$ for all $y \in (0, 1]$. The functions f_k are tangent to the support function ψ in the sense that the coefficients a_k and b_k are chosen using the slope and value of ψ at y_k, as below:

$$y_k = 2^{-k}, \quad a_k = \psi'(y_k), \quad \text{and} \quad b_k = \psi(y_k). \tag{10}$$

From the assumptions (9) made on ψ, the sequences $\{y_k\}$, $\{a_k\}$, and $\{b_k\}$ are positive, decreasing, and converge to zero, when $k \to \infty$.

We still have some freedom for the determination of each f_k, due to the unspecified parameters $c_k > 0$ and ε_k, and the function g_k. The parameters c_k and ε_k will be fixed to control the overlapping of two consecutive functions f_k and f_{k-1}, in order to get particular properties of the central path; while the function g_k will be an appropriate modification of g_k^0 to get smoothness properties.

The fine control of the objective function in the next examples need only to be done in the closed rectangle of $I\!R^2$ defined by

$$\Omega = \{(x, y) \in I\!R^2 \mid |x| \le 1.5, \ -0.5 \le y \le 1\}. \tag{11}$$

For example, smoothness results will be proved for points in Ω only. These properties can easily be extended to the whole $I\!R^2$, by a procedure shown at the end of the paper (in Section 7). At places, we shall pay attention to particular properties satisfied in the horizontal strips of Ω, defined for $k = 1, 2, \ldots$ by

$$\Omega_k = \{(x, y) \in \Omega \mid y \in [y_k, y_{k-1}]\}. \tag{12}$$

Our main concern now is to define the perturbation parameters c_k, and next ε_k. Let us examine two consecutive functions f_{k-1} and f_k, for some index $k \ge 1$ (see Figure 2).

We want the functions f_{k-1} and f_k to “cross” in the strip Ω_k, with a positive slack r_k at y_k and y_{k-1} between the linearized models (see Figure 2). The slack r_k must satisfy

$$\begin{aligned} r_k &\le b_k - (b_{k-1} + a_{k-1}(y_k - y_{k-1})), \\ r_k &\le b_{k-1} - (b_k + a_k(y_{k-1} - y_k)). \end{aligned}$$

These inequalities are compatible with the positivity of r_k, since the right hand sides are positive by the strict convexity of ψ on $[0, 1]$. Our choice for r_k, $k \ge 1$, is:

$$r_k = \frac{1}{2} \min\{b_k - (b_{k-1} + a_{k-1}(y_k - y_{k-1})), \ b_{k-1} - (b_k + a_k(y_{k-1} - y_k))\}. \tag{13}$$

Therefore, $\{r_k\}$ is positive and tends to zero, when $k \to \infty$. Now, one sets (to start the induction, set $r_0 = +\infty$ and $\varepsilon_{-1} = +\infty$):

$$c_k = \frac{1}{4} \min\{r_k, r_{k+1}\} \text{ for all } k \tag{14}$$

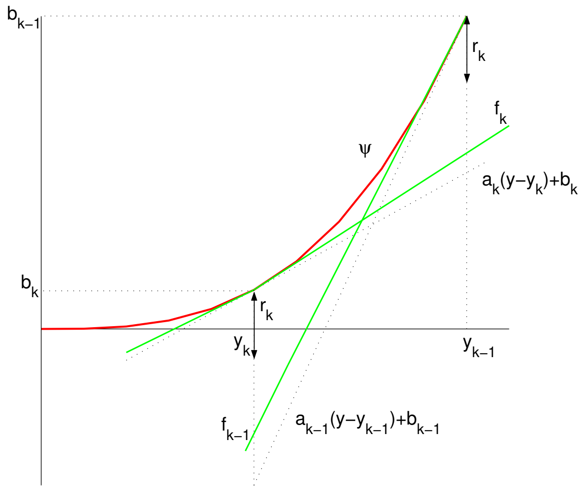

Figure 2: Choice of the coefficients of function f_k

and

$$\begin{array}{c} \text{either } \varepsilon_k = 0 \text{ for all } k \\ \text{or } \varepsilon_k = (-1)^k |\varepsilon_k| \text{ for all } k, \text{ with } 0 < |\varepsilon_k| \leq \min\{c_k/4, |\varepsilon_{k-1}|\}. \end{array} \tag{15}$$

Therefore, either the sequence $\{|\varepsilon_k|\}$ is identically zero or it is positive and non-increasing. Note that $r_k \leq b_{k-1}/2$. Hence $r_{k+1} \leq b_k/2$ and

$$|\varepsilon_k| \leq \frac{c_k}{4} \leq \frac{r_{k+1}}{16} \leq \frac{b_k}{32}. \tag{16}$$

We can now specify our 4th example of objective function.

Example 4 Let $\psi : [-0.5, 1] \to I\!R$ be a function satisfying (9) and let the functions f_k, $k \in I\!N$, be defined as in Example 3, with y_k and the coefficients a_k, b_k, c_k, ε_k satisfying (10) and (13)–(15). The convex functions g_k are unspecified, but must satisfy (7) on Ω. The objective function F is then defined for $z \in \Omega$ by

$$F(z) = \sup_{k \in I\!N} f_k(z). \tag{17}$$

□

In the rest of this section, we highlight some interesting features of the objective function F introduced in Example 4, including the shape of its central path. We start with an overlapping property of two successive functions f_{k-1} and f_k, which implies that these functions cross in the strip Ω_k.

Lemma 1 *Let the functions f_k and the parameters r_k be as in Example 4. Then, for $k \geq 1$ and $(x, y) \in \Omega$:*

$$f_{k-1}(x, y) \leq f_k(x, y) - r_k \qquad \text{if} \quad y \leq y_k \tag{18}$$

and

$$f_k(x,y) \le f_{k-1}(x,y) - r_k \qquad \textit{if} \quad y \ge y_{k-1}. \tag{19}$$

Proof. Let $(x,y) \in \Omega$, with $y \le y_k$. Since $g_{k-1} \le g^0_{k-1}$, $|x| \le 1.5$ and $|y - y_{k-1}| \le 1.5$, (6) gives us $g_{k-1}(z) \le 2$, so that

$$f_{k-1}(x,y) \le a_{k-1}(y - y_{k-1}) + b_{k-1} + 2\,c_{k-1} + 1.5\,|\varepsilon_{k-1}|. \tag{20}$$

On the other hand,

$$\begin{aligned} f_k(x,y) &\ge a_k(y - y_k) + b_k - |\varepsilon_k|\,|x| \\ &\ge a_{k-1}(y - y_k) + b_k - 1.5\,|\varepsilon_k|, \end{aligned} \tag{21}$$

because $a_k \le a_{k-1}$, $y \le y_k$ and $g_k(\cdot) \ge 0$. Grouping (20) and (21):

$$\begin{aligned} &f_k(x,y) - f_{k-1}(x,y) \\ &\ge \; a_{k-1}(y_{k-1} - y_k) + b_k - b_{k-1} - 2c_{k-1} - 1.5\,|\varepsilon_k| - 1.5\,|\varepsilon_{k-1}|. \end{aligned}$$

We have $|\varepsilon_k| \le |\varepsilon_{k-1}| \le \frac{r_k}{16}$ and $c_{k-1} \le \frac{r_k}{4}$. Hence, using (13),

$$f_k(x,y) - f_{k-1}(x,y) \ge 2\,r_k - \frac{1}{2}\,r_k - \frac{3}{16}\,r_k \ge r_k.$$

This completes the proof of (18). The proof of (19) is similar.

□

Observe that Lemma 1 implies that in Ω

$$\begin{array}{ll} f_0 \le f_1 \le f_2 \le \ldots \le f_{k-1} & \text{for } y \le y_{k-1} \\ f_k \ge f_{k+1} \ge f_{k+2} \ge \cdots & \text{for } y \ge y_k. \end{array}$$

Lemma 2 *Let the functions f_k and F be defined as in Example 4, and $z = (x,y) \in \Omega$. Then $\lim_{k\to\infty} f_k(z) = 0$. In addition:*

(i) $F(x,y) = 0$, *if* $y \le 0$;
(ii) $F(z) = \max\{f_{k-1}(z), f_k(z)\}$, *if* $z \in \Omega_k$;
(iii) $F(x,y_k) = f_k(x,y_k)$, *if* $(x,y_k) \in \Omega$.

Proof. Let $z = (x,y) \in \Omega$. Then $\lim_{k\to\infty} f_k(z) = 0$, since g_k satisfies (7) on Ω, g^0_k is bounded by 2 on Ω, and all the coefficients of f_k converge to 0.

(i) If $y \le 0$, from Lemma 1, $f_k(z)$ increases with k. Hence $F(z) = \sup_k f_k(z)$ $= \lim_{k\to\infty} f_k(z) = 0$.

(ii) and (iii) follow immediately from Lemma 1. □

Lemma 3 *The objective function F defined in Example 4 is convex and continuous on Ω. Let $k \ge 1$ and assume that $g_k \equiv g^0_k$. Then*

$$\operatorname*{argmin}_{|x|\le 1.5} F(x,y_k) = \begin{cases} [-1,1] & \textit{if } \varepsilon_k = 0, \\ \{(-1)^{k+1}\} & \textit{otherwise.} \end{cases} \tag{22}$$

Furthermore, for $y \in (y_k, y_{k-1})$, $\operatorname{argmin}_{|x|\le 1.5} F(x,y)$ is reduced to a single point. This point is $x = 0$ when $\varepsilon_k = \varepsilon_{k-1} = 0$.

Proof. Function F is clearly convex, since it is the supremum of the convex functions f_k. On the other hand, the functions f_k are bounded on Ω (by $1.5\,a_0 + b_0 + 2\,\bar{c} + 1.5\,\varepsilon_0$, where $\bar{c}$ is a bound on $\{c_k\}$), hence so is F. As a bounded convex function, F is continuous.

When $g_k \equiv g_k^0$ and $y = y_k$, we can use the properties of the function f_k given after Example 3 (note that $|\varepsilon_k| < c_k$). If $\varepsilon_k = 0$, $\chi(y_k) = [-1, +1]$; if $\varepsilon_k > 0$, $\chi(y_k) = \{-1\}$; and if $\varepsilon_k < 0$, $\chi(y_k) = \{+1\}$. This leads to formula (22).

Consider now the case when $g_k \equiv g_k^0$ and $y \in (y_k, y_{k-1})$. If $\varepsilon_k = \varepsilon_{k-1} = 0$, $f_k(\cdot, y)$ and $f_{k-1}(\cdot, y)$ are both minimized at $x = 0$, hence so is $F(\cdot, y)$. If $\varepsilon_k \neq 0$, since $F(\cdot, y)$ is strictly convex (maximum of two strictly convex functions), it has still a unique minimizer at some $x \in (-1.5, 1.5)$. $\square$

Theorem 1 *Suppose that $g_k \equiv g_k^0$ in Example 4. Then the central path $y \in (0,1] \mapsto \chi(y)$ satisfies the following properties.*

(i) The points $z_k^- = (-1, 2^{-k})$ for k even and $z_k^+ = (1, 2^{-k})$ for k odd belong to the path.

(ii) If $\varepsilon_k = 0$ for all $k \in I\!N$, then the line segments $[z_k^-, z_k^+]$ belong to the path.

(iii) For any $y \in (0,1]$ such that $y \neq 2^{-k}$ for all $k \in I\!N$, $\chi(y)$ is a singleton.

(iv) If $\varepsilon_k \neq 0$ for all $k \in I\!N$, then $\chi(\cdot)$ is a continuous curve.

Proof. Items (*i*) to (*iii*) follow directly from Lemma 3. Let us prove (*iv*).

The map $y \in (0,1] \mapsto \chi(y)$ is outer semicontinuous (in the sense of Rockafellar and Wets [24, Section 5.B]) because of the following: if $y \to \bar{y} > 0$ and some $x_y \in \chi(y) \to \bar{x}$, then $\bar{x} \in \chi(\bar{y})$, as can be seen by taking the limit in the inequality $F(x_y, y) \leq F(x, y)$ for any x s.t. $|x| \leq 1.5$.

Now, if $\varepsilon_k \neq 0$ for all $k \in I\!N$, then Lemma 3 implies that $\chi(y)$ is a singleton. This and the facts that the path is in the bounded set Ω and $\chi(\cdot)$ is outer semicontinuous imply that $y \mapsto \chi(y)$ is a continuous curve, completing the proof. $\square$

From this theorem we can recognize the behavior of the path, as shown in Figure 3. If $\varepsilon_k = 0$, then the path is like a TV antenna with infinitely many branches (left picture). If $\varepsilon_k \neq 0$, then the curve is continuous and visits alternately the points z_k^- for k even, and z_k^+ for k odd, resulting in a zig-zag with infinite variation (right picture).

This example shows that a continuous convex problem can give rise to ill-behaved central paths $y \mapsto \chi(y)$. We have constructed a path that is not a curve, with $\chi(y)$ alternating between a point and a line segment an infinite number of times, and another one that is a curve making a zig-zag with unbounded variation.

The optimal solutions in this example do not satisfy strict complementarity, since $F(x, y) = 0$ if $y \leq 0$. The central paths do not change, however, if a term ay, with $a > 0$, is added to F. Then, strict complementarity holds. Note that the shape of the central paths is compatible with McLinden's result [20], according to which, when strict complementarity holds, the limit points of any

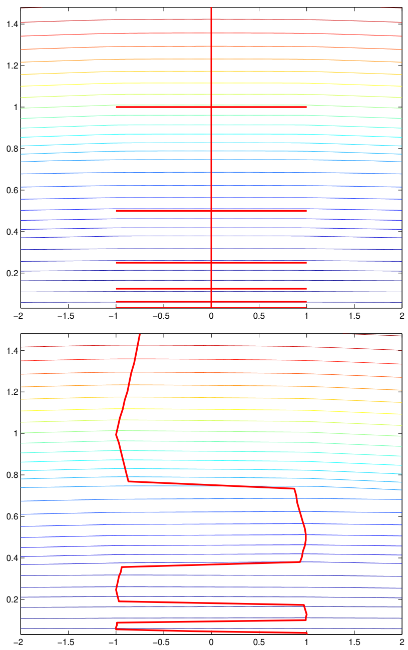

Figure 3: Central paths for a C^0 objective function: the antenna and the zig-zag patterns

convergent selection of central points are in the analytic center of the optimal set. Here, the analytic center is identical to the optimal set since the single constraint is always active at the solutions.

In this example, F is convex but non-smooth and the resulting central paths are not smooth curves. Sections 5 and 6 will be dedicated to obtaining similar paths for smooth problems. Before this, we settle technical tools that will be useful for smooting function F of Example 4.

4 Smoothing the maximum of two functions

This is a self-contained section, in which we show how to smooth the maximum of two real-valued convex functions. This operation will be needed in the subsequent sections to generate examples with smooth objectives similar to the one in Example 4.

Consider two functions $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$, both convex of class C^q, where q is a nonnegative integer or $q = \infty$. The maximum of f_1 and f_2 is denoted by $f_{\max}$:

$$z \in \mathbb{R}^n \mapsto f_{\max}(z) = \max\{f_1(z), f_2(z)\}.$$

This function is in general only continuous but it can be smoothed in a simple way, by rounding it around its kinks, with no changes far from it. Since, for t_1 and $t_2 \in \mathbb{R}$, $\max\{t_1, t_2\} = \frac{1}{2}(t_1 + t_2 + |t_1 - t_2|)$, the max-function can be smoothed by introducing a smooth approximation of the absolute value. This is the track we follow.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function of class C^q, with the following property:

$$\varphi(w) = |w|, \quad \text{for } |w| \geq 1. \tag{23}$$

An example of such a function is given in the left picture in Figure 4 and a concrete construction will be detailed below. Function φ is a C^q version of the absolute value that does not modify it for $|w| \geq 1$. It will be necessary to make this approximation more and more precise. Let $r > 0$ be a scalar measuring the precision of the approximation, which increases when $r \to 0$, and consider the convex function $\varphi_r : \mathbb{R} \to \mathbb{R}$ defined by

$$\varphi_r(w) = r\,\varphi\left(\frac{w}{r}\right), \quad \text{for all } w \in \mathbb{R}. \tag{24}$$

Of course φ_r is convex and $\varphi_r(w) = |w|$ for $|w| \geq r$. It follows from the convexity of φ_r that, for any w and $w' \in \mathbb{R}$:

$$|w| \leq \varphi_r(w) \leq \varphi_r(w') + |w - w'|. \tag{25}$$

Given $r > 0$, we introduce the function $M_r : \mathbb{R}^2 \to \mathbb{R}$

$$M_r(t_1, t_2) = \frac{1}{2}(t_1 + t_2 + \varphi_r(t_1 - t_2)), \tag{26}$$

which is a smooth version of the max-function. It is therefore natural to approximate the function $f_{\max}$ by (see the picture on the right in Figure 4):

$$z \in \mathbb{R}^n \mapsto f(z) = M_r(f_1(z), f_2(z)). \tag{27}$$

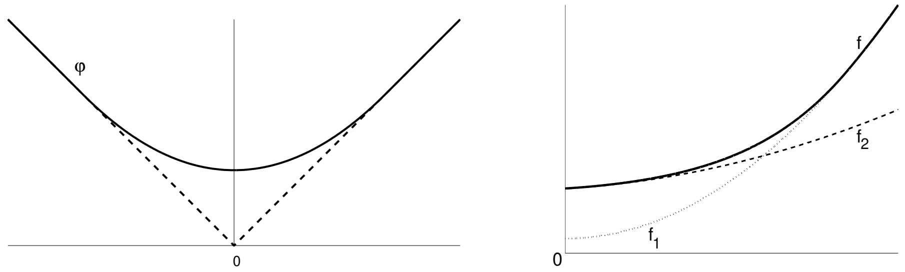

Figure 4: Smoothing of the absolute value and of the maximum of two convex functions

The next lemma claims that M_r is convex and increasing, while Lemma 5 shows in what sense the function f defined by (27) is a convex smoothing of the maximum function.

Lemma 4 *The function M_r defined by (26) is convex and increasing: if $t_1 \leq t_1'$ and $t_2 \leq t_2'$, then $M_r(t_1, t_2) \leq M_r(t_1', t_2')$.*

Proof. The convexity of M_r is a clear consequence of that of φ_r. For the monotonicity, use (25):

$$\begin{aligned} M_r(t_1', t_2') &= \frac{1}{2}(t_1' + t_2' + \varphi_r(t_1' - t_2')) \\ &\geq \frac{1}{2}(t_1' + t_2' + \varphi_r(t_1 - t_2) - |(t_1' - t_2') - (t_1 - t_2)|) \\ &\geq \frac{1}{2}(t_1' + t_2' + \varphi_r(t_1 - t_2) - (t_1' - t_1) - (t_2' - t_2)) \\ &= M_r(t_1, t_2). \end{aligned}$$

$\square$

Lemma 5 *Let f_1 and f_2 be two convex functions of class C^q, with $0 \leq q \leq \infty$, and $f_{\max} = \max\{f_1, f_2\}$. Consider the function f defined by (27), where M_r is constructed as above, with a function φ of class C^q. Then*

(i) f is convex,
(ii) f is of class C^q,
(iii) $z \in \mathbb{R}^n$, $|f_1(z) - f_2(z)| \geq r \Longrightarrow f(z) = f_{\max}(z)$,
(iv) for any $z \in \mathbb{R}^n$, $f_{\max}(z) \leq f(z) \leq f_{\max}(z) + r/2$.

Proof. (i) The convexity of f follows from that of f_1 and f_2 and from the convexity and monotonicity of M_r provided by Lemma 4.

(ii) This follows from the fact that f_1, f_2, φ, φ_r, and M_r are all of class C^q.

(iii) Let $z \in I\!\!R^n$ and set $\Delta f = f_1(z) - f_2(z)$. If $|\Delta f| \geq r$, then from the definition of φ_r, one has $\varphi_r(\Delta f) = |\Delta f|$. Hence

$$f(z) = \frac{1}{2}\left(f_1(z) + f_2(z) + |f_1(z) - f_2(z)|\right) = f_{\max}(z).$$

(iv) From (25) with $w' = 0$ and $w \in I\!\!R$, $|w| \leq \varphi_r(w) \leq |w| + r$. Then, according to (27):

$$\frac{1}{2}\left(f_1(z) + f_2(z) + |\Delta f|\right) \leq f(z) \leq \frac{1}{2}\left(f_1(z) + f_2(z) + |\Delta f| + r\right),$$

which is (iv). $\square$

Examples of C^1 and C^∞ smoothing

We now discuss the construction of a class of smoothing functions by integrating twice a probability density function. This smoothing technique is proposed by Chen and Mangasarian in [6] and [7]. We start with a probability density function $w \in I\!\!R \mapsto \sigma(w) \in I\!\!R$ with compact support:

$$\sigma(w) \geq 0, \qquad \int_{-\infty}^{+\infty} \sigma(w)\,dw = 1, \qquad \text{and} \qquad \sigma(w) = 0, \text{ for } w \notin (0,1).$$

Define $\bar{\varphi}$ by integrating σ twice:

$$w \in I\!\!R \mapsto \bar{\varphi}(w) = \int_{-\infty}^{w} \int_{-\infty}^{t} \sigma(\tau)\,d\tau\,dt.$$

Let $c = 1 - \bar{\varphi}(1)$. Since $\bar{\varphi}'(w) = 0$ for $w \leq 0$, and $\bar{\varphi}'(w) \leq 1$ for all $w \in [0,1]$, then $\bar{\varphi}(1) \leq 1$ and $c \geq 0$. The smoothing function $\varphi : I\!\!R \to I\!\!R$ is the symmetrization of $\bar{\varphi}$, defined by

$$\varphi(w) = \bar{\varphi}(|w|) + c. \tag{28}$$

This function is convex and satisfies (23). Its smoothness depends on that of σ.

We now provide two examples. The first one is based on the pulse function defined below and will be used in Section 5:

$$\sigma(w) = \begin{cases} 1 & \text{if } w \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Integrating twice this pulse function, we obtain a continuously differentiable smoothing function

$$\varphi(w) = \begin{cases} \frac{1}{2}w^2 + \frac{1}{2} & \text{if } w \in [-1,1], \\ |w| & \text{otherwise.} \end{cases} \tag{29}$$

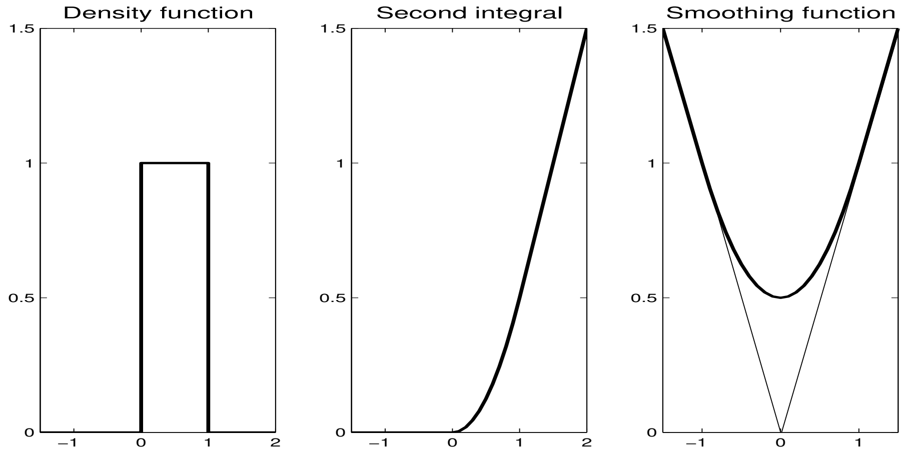

Figure 5: Smoothing function obtained by twice integrating the pulse function

The second example, used in Section 6, is a smoothing function of class C^∞ but not analytic. To obtain this example, consider initially a classical example of a non-analytic function of class C^∞ (see [13, page 51]) defined by

$$w \in I\!\!R \mapsto \theta(w) = \begin{cases} e^{-1/w^2} & \text{if } w > 0, \\ 0 & \text{if } w \leq 0. \end{cases} \tag{30}$$

For $w \neq 0$ and $q \in I\!\!N$, the qth order derivatives $\theta^{(q)}(w)$ can be computed by elementary calculus. For $w < 0$, $\theta^{(q)}(w) = 0$; while for $w > 0$, $\theta^{(q)}(w)$ is a polynomial of degree $3q$ in w^{-1}, times $\theta(w)$. It is known that

$$\lim_{w \to 0} \frac{1}{w^p}\, \theta(w) = 0, \quad \forall\, p \in I\!\!N, \tag{31}$$

so that $\theta^{(q)}(0) = 0$, for any $q \in I\!\!N$. Now, define

$$w \in I\!\!R \mapsto \sigma(w) = \beta\theta(w)\theta(1-w) \tag{32}$$

where $\beta > 0$ is such that

$$\int_{-\infty}^{+\infty} \sigma(w)dw = 1.$$

This is a probability density function, with derivatives of any order that vanish out of $[0,1]$. Finally the non-analytic smoothing function φ of class C^∞ is obtained by twice integrating this density function.

5 A continuously differentiable example

Our aim now is to smooth the objective function F introduced in Example 4, without modifying the layout of the central paths obtained for $\varepsilon_k = 0$ (antenna) and $\varepsilon_k \neq 0$ (zig-zag). A close examination of the construction of F shows that to reach this goal, it is necessary to smooth the elementary functions f_k introduced in (8), which in turn requires the smoothing of g_k^0, and the "sup" operator appearing in (17).

Smoothing g_0^k is very easy, while smoothing the sup operation will use the results of the previous section. We start by showing how these operations are performed, and end up with Example 5, in which the central path has again the antenna and zig-zag behaviors. The zig-zag central path is still non-smooth, however.

Then we improve the smoothing to obtain a C^q example, with $q > 1$: again the same shapes for the central paths appear, but now the zig-zag path is $(q-1)$ times differentiable.

5.1 An example of class C^1

Let us first consider the smoothing of the elementary functions f_k. For this, we must construct g_k in formula (8) as a smooth approximation of the continuous only function g_k^0 given by (5). The analysis of Section 3 has shown that the lemmas involving g_k remain true if that function takes nonnegative values not exceeding those of g_k^0; see condition (7). In other words, this smoothing operation has to be done from below, by decreasing g_k^0. The reason is that for $\varepsilon_k = 0$ the horizontal segment of the central path will remain central, while for $\varepsilon_k \neq 0$ the central path will do even larger loops. Therefore, the weird pattern of the previous central paths will be preserved.

For this example, we define $g_k : \Omega \to \mathbb{R}$ for $k \in \mathbb{N}$ by

$$z = (x, y) \in \Omega \mapsto g_k(z) = \frac{1}{4} \left(g_k^0(z)\right)^2 . \tag{33}$$

Lemma 6 *The function g_k defined by* (33), *with g_k^0 given in* (5), *is convex, continuously differentiable, and satisfies* (7) *on Ω.*

Proof. The function g_k^0 is convex and takes nonnegative values, and $t \in \mathbb{R}_+ \mapsto t^2/4$ is convex and increasing. Therefore, g_k is convex.

On the other hand, g_k^0 is of class C^∞ near any $z \in \Omega\backslash\{z_k^-, z_k^+\}$. Hence, the same is true for g_k. We now analyze the smoothness of g_k at z_k^+, knowing that the argument is similar at z_k^-. The function g_k is differentiable at z_k^+ if its subdifferential is reduced to a singleton (see [17, Section VI.2.1]). From [17, theorem VI.6.3.1], we know that $\partial g(z_k^+)$ is the convex hull of the set of limits of convergent sequences $\nabla g(z^j)$ for $z^j \to z_k^+$, $z^j \notin \{z_k^-, z_k^+\}$. Consider such a sequence $(z^j) \subset \Omega$. We have $\nabla g_k(z^j) = \frac{1}{2} g_k^0(z^j) \nabla g_k^0(z^j)$. Since $\|\nabla g_k^0(z^j)\|$ is bounded by 1 and $g_k^0(z^j) \to 0$, we deduce that $\lim_{j\to\infty} \nabla g_k(z^j) = 0$. Hence $\partial g_k(z_k^+) = \{0\}$. The continuous differentiability of g_k follows from a well-known result of convex analysis (see [17, remark VI.6.2.6]).

To conclude, note that g_k satisfies (7) on Ω, since by (6), $g_k^0(z) \leq 2$ on that set. $\square$

We now describe our 5th example.

Example 5 Function $F : \Omega \to I\!\!R$ is made from the following ingredients.

- The support function $\psi : [-0.5, 1] \to I\!\!R$ is defined for $y \in [-0.5, 1]$ by

$$\psi(y) = \frac{1}{2} (\max\{0, y\})^2 . \tag{34}$$

  Conditions in (9) are satisfied.

- For $k \in I\!\!N$, g_k is the function defined by (33), where g_k^0 is given by (5). According to Lemma 6, g_k is a convex C^1 function satisfying (7).

- The elementary functions f_k, $k \in I\!\!N$, are then defined by (8) (as in Example 3), with coordinate y_k and coefficients a_k, b_k, c_k and ε_k satisfying (10) and (13)–(15).

- The smoothing function φ is given by (29).

The objective function F is now constructed by a recursive process.

- Set $F_0(z) = f_0(z)$, for all $z \in \Omega$.

- For $k = 1, 2, \ldots$, set

$$F_k(z) = M_{r_k}\left(f_k(z), F_{k-1}(z)\right), \quad \text{for all } z \in \Omega, \tag{35}$$

  where r_k is defined by (13), and M_{r_k} is defined by (24) and (26).

Then F is obtained as the pointwise limit of the functions F_k:

$$F(z) = \lim_{k\to\infty} F_k(z), \quad \forall z \in \Omega. \tag{36}$$

□

The next lemma gives some elementary properties of the functions F_k and F. We denote by $\operatorname{int}\Omega = \{(x, y) \in I\!\!R^2 : |x| < 1.5, \ -0.5 < y < 1\}$ the interior of the set Ω.

Lemma 7 *Let $k \in I\!\!N$. The function F_k defined in Example 5 is convex and continuously differentiable on* $\operatorname{int}\Omega$. *Furthermore:*

(i) $F(z) = F_{k+i}(z)$, *if* $i \in I\!\!N$ *and* $z = (x, y) \in \Omega$ *with* $y \geq y_k$;
(ii) $F_k(z) = f_k(z)$, *if* $z = (x, y) \in \Omega$ *and* $y \leq y_k$;
(iii) $F(z) = 0$, *if* $z = (x, y) \in \Omega$ *and* $y \leq 0$;
(iv) $F(z) = F_k(z) = M_{r_k}(f_k(z), f_{k-1}(z))$, *if* $z \in \Omega_k$.

Proof. Convexity and differentiability are proven by induction. According to (8) and Lemma 6, $F_0 = f_0$ is convex and C^1. Suppose now that $k \geq 1$ and that F_{k-1} is convex and C^1. Since f_k is convex and C^1 (like f_0), then (35), the fact that φ is C^1, and Lemma 5 show that F_k is also convex and C^1.

(i) From Lemma 5(iv) and definition (35): $F_k \geq f_k$. For $z \in \Omega$ with $y \geq y_k$, Lemma 1 gives us

$$F_k(z) \geq f_k(z) \geq f_{k+1}(z) + r_{k+1}.$$

Based on Lemma 5(iii) and definition (35), we have $F_{k+1}(z) = F_k(z)$.

For the same $z \in \Omega$, one has $y \geq y_k \geq y_{k+1}$, so that the result just proven shows that $F_{k+2}(z) = F_{k+1}(z) = F_k(z)$. By induction, $F_{k+i}(z) = F_k(z)$, for any $i \in \mathbb{N}$. The definition (36) of F now leads to the result.

(ii) By induction: by definition $F_0 = f_0$. Assume now that (ii) holds for $k-1$, with $k \geq 1$, and let $z \in \Omega$ be such that $y \leq y_k$. By induction and Lemma 1:

$$F_{k-1}(z) = f_{k-1}(z) \leq f_k(z) - r_k.$$

Hence, according to Lemma 5(iii) and the definition (35), we have $F_k(z) = f_k(z)$.

(iii) For $z = (x, y) \in \Omega$ with $y \leq 0$, $F(z) = \lim F_k(z)$ [by (36)] $= \lim f_k(z)$ [by point (ii)] $= 0$ [by Lemma 2].

(iv) Consider $z \in \Omega_k$; hence $y \in [y_k, y_{k-1}]$. From point (ii) and $y \leq y_{k-1}$, $F_{k-1}(z) = f_{k-1}(z)$, so that the definition (35) leads to

$$F_k(z) = M_{r_k}(f_k(z), f_{k-1}(z)).$$

On the other hand, $F(z) = F_k(z)$ by point (i). □

The lemma above shows that the recursive construction is in fact not needed. The function F can be defined simply by applying (iv) for $k \in \mathbb{N}$. Denote by $\bar{f} = \sup_{k \in \mathbb{N}} f_k$ the objective function of Example 4 with the data of Example 5. Then Lemma 7(iv) and Lemma 5(iv) yield

$$\bar{f}(z) \leq F(z) \leq \bar{f}(z) + r_k/2, \quad \text{for } z \in \Omega_k. \tag{37}$$

This shows that function F is a smoothing of $\bar{f}$. The smoothing process rounds up the kinks within each strip Ω_k, and the smoothing of each kink does not affect the others.

Lemma 8 *The function F constructed in Example 5 is convex and continuously differentiable on* $\operatorname{int} \Omega$.

Proof. We know from Lemma 7(iii) that $F(x, y) = 0$ if $y \leq 0$. Hence F is convex and C^1 on $\Omega_- = \{(x, y) \in \operatorname{int} \Omega \mid y < 0\}$.

Let us now show that F is convex and C^1 on $\Omega_+ = \{(x, y) \in \operatorname{int} \Omega \mid y > 0\}$. Let $k \in \mathbb{N}$. By Lemma 7, $F(z) = F_k(z)$ in the open set $O_k = \{(x, y) \in \operatorname{int} \Omega \mid y > y_k\}$, so that F is convex and smooth on this set. Since k is arbitrary and $\Omega_+ = \cup_{k \in \mathbb{N}} O_k$, F is convex and smooth on Ω_+.

Note that F is continuous on Ω. Indeed by Lemmas 2 and 3, $\bar{f}$ vanishes on Ω_- and is continuous on Ω, so that (37) implies the continuity of F at any point $(x, 0) \in \Omega$. To show the convexity of F, consider two points $z \in \Omega_-$ and $z' \in \Omega_+$, and a scalar $t \in [0, 1]$. Since F vanishes on Ω_-, it is enough to show that $F((1-t)z + tz') \leq tF(z')$. This inequality clearly holds if $(1-t)z + tz' \in \Omega_-$, since then the left hand side vanishes and $F(z') \geq 0$ (use (37) and the non-negativity of $\bar{f}$ on Ω_+). Otherwise, let z'' be the point in $[z, z']$ such that $(z'')^\top e^2 = 0$. Then $(1-t)z + tz' = (1-t')z'' + t'z'$ for some $t' \in [0, t]$. Since F is nonnegative and convex on the closure of Ω_+ [17, proposition IV.1.2.6] and $F(z'') = 0$, one deduces $F((1-t)z + tz') = F((1-t')z'' + t'z') \leq t'F(z') \leq tF(z')$.

We still have to prove that F is C^1 at an arbitrary point $z = (x, 0)$ with $-1.5 < x < 1.5$. It is enough to show that $\partial F(z) = \{0\}$. Let $\delta \in \partial F(z)$. Along any direction $h = (h_1, h_2)$ such that $h_2 \leq 0$ we must have $F'(z, h) = 0$, because F vanishes on the closure of Ω_-. For $h = (0, -1)$, $0 = F'(z, h) \geq \delta^\top h = -\delta_2$. Hence $\delta_2 \geq 0$. For $h = (\pm 1, 0)$, $0 = F'(z, h) \geq \delta^\top h = \pm\delta_1$. Hence $\delta_1 = 0$. We conclude that $\delta = (0, \delta_2)$, with $\delta_2 \geq 0$. By convexity,

$$F(x, y_k) \geq F'((x, 0); (0, y_k)) \geq \delta_2 \, y_k. \tag{38}$$

According to Lemma 7, $F(x, y_k) = f_k(x, y_k) \leq b_k + c_k g_k(x, y_k) + 1.5\,|\varepsilon_k|$. From (16) and the fact that $g_k(x, y_k) \leq 0.5$, it follows immediately that $F(x, y_k) \leq 2b_k = y_k^2$. Then the limit in (38) implies that $\delta = 0$. $\square$

Lemma 8 shows that the function F constructed in Example 5 is indeed convex and smooth. Let us examine its central path.

For $\varepsilon_k = 0$, the antenna pattern of the central path is preserved, like in example 4. This is essentially a consequence of Lemma 7. Indeed, each segment $[z_k^-, z_k^+]$ belongs to the central path because $F(z) = f_k(z)$ when $z = (x, y_k) \in \Omega$. For $y \in (y_k, y_{k-1})$, we know that both $f_k(\cdot, y)$ and $f_{k-1}(\cdot, y)$ are uniquely minimized at $x = 0$. Then, this is also the case for $F(\cdot, y) = M_{r_k}(f_k(\cdot, y), f_{k-1}(\cdot, y))$ by the monotonicity property of M_{r_k} (Lemma 4).

Consider now the case when $\varepsilon_k \neq 0$ is computed like in (15). For $y \in (y_k, y_{k-1})$, $F(\cdot, y)$ has a unique minimizer (for the same reasons as above). For $y = y_k$, $F(\cdot, y_k) = f_k(\cdot, y_k)$, and an easy calculation using (8) and (33) yields

$$\operatorname*{argmin}_{x \in I\!R} F(x, y_k) = \{(-1)^{k+1}(1 + \tau_k)\},$$

where $0 < \tau_k = 2|\varepsilon_k|/c_k \leq 0.5$. Like in Example 4, the central path is a continuous curve (for $y > 0$) forming a zig-zag with even larger loops.

The antenna and zig-zag central paths for the C^1 objective function of example 5 are given in Figure 6 (the zig-zag is the dashed curve).

5.2 An example of class C^q with $q > 1$

It is straightforward to modify the ingredients determining the function F in Example 5 to get a smoother objective function. Let $q \geq 2$ be an integer specifying the required degree of smoothness ($q = 1$ in Example 5). The modifications to bring are the following. First, instead of defining g_k by (33), set

$$g_k(z) = \left(\frac{g_k^0(z)}{2}\right)^{q+1}, \quad \text{for } z \in \Omega. \tag{39}$$

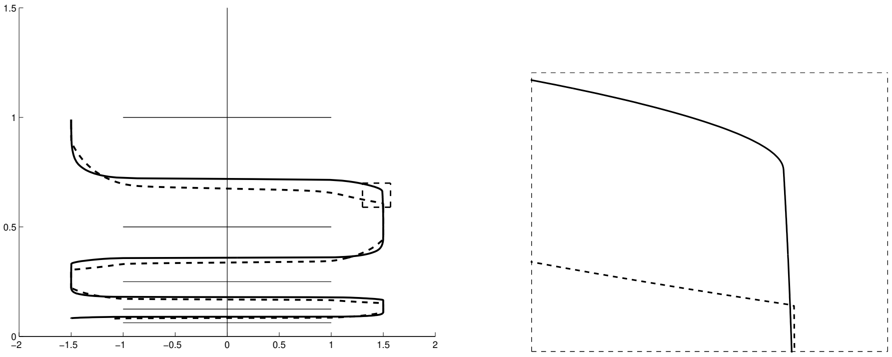

Figure 6: The zig-zag central path for a C^1 and a C^2 objective functions.

The support function $\psi : [-0.5, 1] \to I\!R$ in (34) is now defined by

$$\psi(y) = \frac{1}{q+1} \left(\max\{0, y\}\right)^{q+1}, \quad \text{for } y \in [-0.5, 1], \tag{40}$$

and the smoothing function $\varphi : I\!R \to I\!R_+$ in (29) is now obtained as in Section 4, starting with the probability density function

$$\sigma(w) = \begin{cases} \beta w^{q-2}(1-w)^{q-1} & \text{if } w \in [0,1], \\ 0 & \text{otherwise,} \end{cases} \tag{41}$$

where $\beta > 0$ is set to have $\int \sigma(w)dw = 1$. It can be shown, using arguments similar to those of Sections 5 and 6 that, with these modifications, the objective function in Example 5 is convex and of class C^q.

The next lemma gives conditions ensuring the smoothness of the zig-zag central path. The one corresponding to the C^2 objective function, obtained with $q = 2$ in the data above, is the continuous curve in Figure 6.

Lemma 9 *Consider Example 5, in which $g_k = \gamma \circ g_k^0$, where $\gamma : I\!R \to I\!R$ is twice differentiable and satisfies $\gamma'(0) = 0$, as well as $\gamma'(t) > 0$ and $\gamma''(t) > 0$ when $t > 0$. Suppose also that the smoothing function φ is twice differentiable and that the objective function F is convex and of class C^q, with $q \geq 2$. Then, the zig-zag central path is a function of class C^{q-1} of $y > 0$.*

Proof. The central path is defined by (4), where χ is *a priori* multivalued. Since F is convex and differentiable, a point $z = (x, y)$ is on the central path if and only if it satisfies the optimality condition

$$\nabla_x F(x, y) = 0.$$

We want to show that χ is single valued and of class C^{q-1}. Since ∇F is C^{q-1}, this is a clear consequence of the implicit function theorem, if $\nabla^2_{xx} F(x, y)$ is nonzero along the zig-zag central path. Actually, we are going to show that $\nabla^2_{xx} F(x, y) > 0$, which will conclude the proof.

Let $z = (x, y)$, with $y > 0$, be an arbitrary point on the zig-zag central path. Observe that when $\varepsilon_k > 0$:

$$F'(z_k^-; -e^1) = f_k'(z_k^-; -e^1) = c_k \gamma'(0)(g_k^0)'(z_k^-; -e^1) - \varepsilon_k = -\varepsilon_k < 0.$$

Similarly, when $\varepsilon_k < 0$: $F'(z_k^+; e^1) = \varepsilon_k < 0$. Therefore, $z \notin [z_k^-, z_k^+]$, $g_k^0(z) > 0$, and g_k^0 is C^∞ around z.

Let us now show that $\nabla_{xx}^2 f_k(z) > 0$ or equivalently, since $c_k > 0$, that $\nabla_{xx}^2 g_k(z) > 0$. By the assumptions and the smoothness of g_k^0 around z:

$$\nabla_{xx}^2 g_k(z) = \gamma'(g_k^0(z))\nabla_{xx}^2 g_k^0(z) + \gamma''(g_k^0(z))\left(\nabla_x g_k^0(z)\right)^2.$$

Because $g_k^0(z) > 0$, one has $\gamma'(g_k^0(z)) > 0$ and $\gamma''(g_k^0(z)) > 0$. Also, since $z \notin \{z_k^-, z_k^+\}$:

$$\nabla_x g_k^0(z) = \frac{1}{2}\left(\frac{x+1}{\|z - z_k^-\|} + \frac{x-1}{\|z - z_k^+\|}\right).$$

This derivative vanishes only if $x = 0$. On the other hand, for $x = 0$:

$$\nabla_{xx}^2 g_k^0(z) = \frac{(y - y_k)^2}{\|(1, y - y_k)\|^3}$$

This quantity vanishes only if $y = y_k$. Since $(0, y_k)$ is not on the zig-zag central path, one deduces that $\nabla_{xx}^2 g_k(z) > 0$, hence $\nabla_{xx}^2 f_k(z) > 0$.

If $y = y_k$, $F(z) = f_k(z)$ and $\nabla_{xx}^2 F(x, y) = \nabla_{xx}^2 f_k(z) > 0$.

Suppose now that $y \in (y_k, y_{k-1})$. Then, according to Lemma 7(iv)

$$F(z) = M_{r_k}(f_k(z), f_{k-1}(z)) = \frac{1}{2}\Big(f_k(z) + f_{k-1}(z) + \varphi_{r_k}(f_k(z) - f_{k-1}(z))\Big).$$

Therefore

$$\begin{aligned}\nabla_{xx}^2 F(x, y) \quad = \quad & \frac{1}{2}\Big(\Big[1 + \varphi_{r_k}'(f_k(z) - f_{k-1}(z))\Big]\nabla_{xx}^2 f_k(z) + \\ & \Big[1 - \varphi_{r_k}'(f_k(z) - f_{k-1}(z))\Big]\nabla_{xx}^2 f_{k-1}(z) + \\ & \varphi_{r_k}''(f_k(z) - f_{k-1}(z))\Big[\nabla_x f_k(z) - \nabla_x f_{k-1}(z)\Big]^2\Big).\end{aligned}$$

By construction of φ in (23), $|\varphi_r'(t)| \le 1$ for any $t \in I\!R$. Hence, the three terms in the main parenthesis above are all nonnegative. On the other hand, we have shown that $\nabla_{xx}^2 f_k(z) > 0$ and $\nabla_{xx}^2 f_{k-1}(z) > 0$, and since the factor of each of these two quantities cannot both vanish, one deduces that $\nabla_{xx}^2 F(x, y) > 0$. $\square$

6 An example of class C^∞

This section is quite technical, but solves our final quest: constructing an infinitely smooth example. The construction is very similar to the one in Section 5, but now all functions involved is the construction and smoothing of F must be of class C^∞. Again we obtain the antenna and zig-zag paths, and the zig-zag path is an infinitely differentiable curve.

Example 6 Function $F : \Omega \to I\!R$ is constructed in the same way as in Example 5, with the following modifications.

- The support function $\psi : [-0.5, 1] \to I\!R$ is now defined for $y \in [-0.5, 1]$ by
$$\psi(y) = \theta\left(\frac{y}{2}\right), \tag{42}$$
where θ is the C^∞ function given in (30)

- The perturbation functions $g_k : \Omega \to I\!R$, for $k \in I\!N$, are now obtained by smoothing with θ the functions g_k^0 defined in (5): for $z = (x, y) \in \Omega$,
$$g_k(z) = g_k^\infty(z) := \theta\left(\frac{1}{3} g_k^0(z)\right). \tag{43}$$

- The smoothing function φ is now the C^∞ function given at the end of Section 4, using function θ and the probability density function defined by (32).

$\square$

Let us motivate the choices made in Example 6. It is easy to check (calculating second derivatives), that θ is strictly convex on $(0, \sqrt{2/3})$. Therefore, the support function ψ satisfies the conditions in (9). On the other hand, we know by (6) that, for $z \in \Omega$, $0 \le g_k^0(z) \le 2$. Hence, g_k^∞ is convex in Ω, as a composition of a convex nondecreasing function and a convex function. It is easy to check that for $w \ge 0$, $0 \le \theta(w) \le w$. Hence, for $z \in \Omega$, we have $0 \le g_k^\infty(z) \le g_k^0(z)$ and (7) is satisfied.

Our main work in this section is to prove that F is of class C^∞ (its convexity and the behavior of the central paths will result from arguments similar to those in Section 5). For this, we use the following result quoted by Fleming [13, page 50] as a special case of a theorem of Whitney [29].

Lemma 10 *Let $\Gamma \subset I\!R^p$ be the closure of an open set Γ_0, and assume that Γ is convex. Let ϕ be of class C^q on Γ_0, for some $q \in I\!N$, and continuous on Γ. Moreover, assume that for each $m \ge 0$ and $n \ge 0$ with $m + n = q$, there is a function $\bar{\phi}_{m,n}$ continuous on Γ such that*
$$\bar{\phi}_{m,n}(z) = \frac{\partial^q \phi}{\partial x^m \partial y^n}(z), \quad \text{for all } z \in \Gamma_0.$$
Then there exists a function $\bar{\phi}$ of class C^q on $I\!R^p$ such that $\bar{\phi}(z) = \phi(z)$ for every $z \in \Gamma$. In particular, ϕ is of class C^q on Γ.

In plain words, this lemma ensures that if any partial derivative of order q of ϕ on Γ_0 can be continuously extended to Γ, then ϕ is of class C^q on Γ.

According to Lemma 10, we have to look at the partial derivatives of the functions involved in the definition of F, and to control their behavior near possible singular points. We do so in sequence for g_k^0, θ, g_k^∞, f_k, φ, and finally F.

Derivatives of g_k^0. The next lemma establishes a bound for these derivatives.

Lemma 11 *For any $q \in I\!N$, there exists a positive constant N_q such that, for all k, m, $n \in I\!N$ with $m+n=q$ and all $z \in \Omega$ with $g_k^0(z) > 0$, there holds*

$$\left| \frac{\partial^q g_k^0}{\partial x^m \partial y^n}(z) \right| \le \frac{N_q}{(g_k^0(z))^{q-1}}. \tag{44}$$

Proof. When $q=0$, the result is true with $N_0 = 1$. Now let us fix k, m, $n \in I\!N$ with $m+n=q>0$. For $z \in \Omega$ with $g_k^0(z) > 0$,

$$\frac{\partial^q g_k^0}{\partial x^m \partial y^n}(z) = \frac{1}{2} \left(\frac{\partial^q}{\partial x^m \partial y^n} \|z - z_k^-\| + \frac{\partial^q}{\partial x^m \partial y^n} \|z - z_k^+\| \right).$$

Calculating these derivatives, we obtain

$$\frac{\partial^q g_k^0}{\partial x^m \partial y^n}(z) = \frac{Q_{q,k}^-(z)}{\|z - z_k^-\|^{2q-1}} + \frac{Q_{q,k}^+(z)}{\|z - z_k^+\|^{2q-1}},$$

where $Q_{q,k}^{\mp}$ (in this sentence, the order of the superscripts + and − matters) is a sum of products of a constant (depending on m and n) times q factors chosen among $(x \pm 1)$ and $(y - y_k)$. Because $|x \pm 1|$ and $|y - y_k|$ are bounded by $\|z - z_k^{\mp}\|$, there is a constant $N_{m,n}$ independent of k such that

$$\left| \frac{\partial^q g_k^0}{\partial x^m \partial y^n}(z) \right| \le N_{m,n} \left(\frac{1}{\|z - z_k^-\|^{q-1}} + \frac{1}{\|z - z_k^+\|^{q-1}} \right).$$

Using the triangle inequality and $\|z_k^- - z_k^+\| = 2$, we have

$$g_k^0(z) = \frac{1}{2} \left(\|z - z_k^-\| + \|z - z_k^- + z_k^- - z_k^+\| - 2 \right) \le \|z - z_k^-\|.$$

Similarly, $g_k^0(z) \le \|z - z_k^+\|$. It follows that

$$\left| \frac{\partial^q g_k^0}{\partial x^m \partial y^n}(z) \right| \le \frac{2N_{m,n}}{(g_k^0(z))^{q-1}}.$$

The result follows with $N_q := \max\{2N_{m,n} \mid m+n=q,\ m,\ n \in I\!N\}$. □

Derivatives of a composite function. Consider the composite function $\phi = \gamma \circ \eta$, where $\gamma : I\!R \to I\!R$ and $\eta : I\!R^2 \to I\!R$ are functions of class C^q. Let m, $n \in I\!N$ be such that $q := m+n > 0$. The partial derivative of ϕ can be written as a finite sum of the form

$$\frac{\partial^q \phi}{\partial x^m \partial y^n}(z) = \sum_j \alpha_j \, \gamma^{(l_j)}(\eta(z)) \, \mathcal{P}_j, \tag{45}$$

where α_j are integers, $l_j \in \{1, 2, \ldots, q\}$, and $\mathcal{P}_j$ are products of l_j partial derivatives of η at z, with sum of orders equal to q. *Derivatives of θ.* The derivative of order $l \in I\!N$ of θ defined in (30) at w is given by

$$\theta^{(l)}(w) = \begin{cases} \theta(w) P_{3l}\left(\frac{1}{w}\right) & \text{if } w > 0 \\ 0 & \text{if } w \le 0, \end{cases} \tag{46}$$

where $P_{3l} : I\!R \to I\!R$ is a polynomial of degree $3l$.

Derivatives of g_k^∞. We now apply the results above to obtain a bound on the derivatives of g_k^∞ defined in (43). The next lemma also shows that g_k^∞ is of class C^∞, which extends Lemma 6 to Example 6.

Lemma 12 *The functions g_k^∞, $k \geq 0$, are convex and of class C^∞ on Ω. Furthermore, for any $q \in I\!N$, there exists a positive constant K_q such that, for all k, m, $n \in I\!N$ with $m + n = q$ and for all $z \in \Omega$, there holds*

$$\left| \frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z) \right| \leq K_q. \tag{47}$$

Proof. Let k, m, $n \in I\!N$ be fixed, $q = n + m$. For $q = 0$, the fact that g_k^0 is bounded by 2 on Ω readily implies (47) with $K_0 = \theta(2)$. On the other hand g_k^∞ is clearly continuous on Ω. Therefore, we can assume $q > 0$. To simplify the notation we set $g(\cdot) \equiv g_k^0(\cdot)/3$; hence $g_k^\infty = \theta \circ g$.

We know that g_k^0 is C^∞, except at the points z_k^- and z_k^+, where it is non-differentiable. Therefore, g_k^∞ is C^∞ at any point $z \in \Omega \backslash \{z_k^-, z_k^+\}$. To show its smoothness at $z_k^\pm$, we prove that any partial derivative of order q of g^∞ converges to zero when $z \to z_k^\pm$ and apply Lemma 10.

Let us compute the derivatives of g_k^∞ at a point $z \in \Omega \backslash \{z_k^-, z_k^+\}$. As above:

$$\frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z) = \sum_j \alpha_j \, \theta^{(l_j)}\left(g(z)\right) \, \mathcal{P}_j, \tag{48}$$

where α_j are integers, $l_j \in \{1, 2, \cdots, q\}$, and $\mathcal{P}_j$ are products of l_j partial derivatives of g, with sum of orders equal to q.

We know from (46) that, if $g_k^0(z) = 0$, then $\theta^{(l_j)}\left(g(z)\right) = 0$ for all j, so that $\frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z) = 0$. Consider now the case when $g_k^0(z) \neq 0$. Then from (46) again

$$\theta^{(l_j)}\left(g(z)\right) = \theta\left(g(z)\right) \, P_{3l_j}\left(\frac{1}{g(z)}\right), \tag{49}$$

where $P_{3l_j} : I\!R \to I\!R$ is a polynomial of degree $3l_j$. On the other hand, each derivative of g satisfies (44); hence

$$|\mathcal{P}_j| \leq \prod_{l=1}^{l_j} \frac{N_{q_l}}{(g(z))^{q_l - 1}} \leq \frac{\bar{N}_q^q}{(g(z))^{r_j}},$$

where $\bar{N}_q := \max\{1, N_1, \ldots, N_q\}$ and $r_j \leq q$. Combining this with (49) and (48) yields

$$\left| \frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z) \right| \leq \theta\left(g(z)\right) \sum_j \alpha_j \, P_{3l_j}\left(\frac{1}{g(z)}\right) \frac{\bar{N}_q^q}{(g(z))^{r_j}} = \theta\left(g(z)\right) \tilde{P}_{4q}\left(\frac{1}{g(z)}\right),$$

where $\tilde{P}_{4q} : I\!R \to I\!R$ is a polynomial of degree $\leq 4q$. We have shown that for $z \in \Omega\backslash\{z_k^-, z_k^+\}$:

$$\begin{cases} \dfrac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z) = 0 & \text{if } g_k^0(z) = 0, \\ \left|\dfrac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z)\right| \leq \theta\left(\frac{g_k^0(z)}{3}\right)\ \tilde{P}_{4q}\left(\frac{3}{g_k^0(z)}\right) & \text{otherwise.} \end{cases} \tag{50}$$

Let us now prove that g_k^∞ is C^∞ at z_k^+ (the proof is similar at z_k^-). By Lemma 10, it is enough to show that

$$\lim_{z \to z_k^+} \frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z) = 0.$$

Consider a sequence $\{z_i\}$ in $\Omega\backslash\{z_k^-, z_k^+\}$ such that $z_i \to z_k^+$. Then $g_k^0(z_i) \to 0$ and the limit above follows from (50) and (31).

We still have to prove (47). From (31), the function

$$w \in I\!R \mapsto \begin{cases} \theta(w)\tilde{P}_{4q}(1/w) & \text{if } w > 0 \\ 0 & \text{if } w \leq 0 \end{cases}$$

is continuous and g_k^0 takes its values in the compact set $[0, 2]$. Therefore, from (50), there exists a positive constant K_q dependent only on q, such that for any $z \in \Omega$:

$$\left|\frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z)\right| \leq K_q.$$

□

Recall that f_k is defined by (8), in which $g_k = g_k^\infty$ is now given by (43). Since g_k^∞ is convex and of class C^∞ on Ω (Lemma 12), so is f_k. Also, the smoothing function φ is convex and C^∞. It is then possible to extend Lemma 7.

Lemma 13 *For any $k \in I\!N$, the function F_k of Example 6 is convex and of class C^∞ and all the properties (i)-(iv) in Lemma 7 hold.*

Proof. The proof is similar to that of Lemma 7. □

To proceed, we need a technical result.

Lemma 14 *For k sufficiently large,*

$$2 \leq \frac{\psi(y_{k-1})}{r_k} \leq 3 \qquad \text{and} \qquad \frac{2}{y_k^3} \leq \frac{a_{k-1}}{r_k} \leq \frac{3}{y_k^3}.$$

Proof. We simplify the notation by setting

$$\begin{aligned} \alpha_k &:= \psi(y_k) - [\psi(y_{k-1}) + \psi'(y_{k-1})(y_k - y_{k-1})] \\ \beta_k &:= \psi(y_{k-1}) - [\psi(y_k) + \psi'(y_k)(y_{k-1} - y_k)]. \end{aligned} \tag{51}$$

By (10) and (13), $r_k = \frac{1}{2}\min\{\alpha_k,\ \beta_k\}$.

Let us first prove that $\beta_k \le \alpha_k$, for $k \ge 1$. Since ψ is positive and increasing on $(0,1)$, we have from (42):

$$\beta_k \le \psi(y_{k-1}) = \theta(y_k).$$

Now consider (51) for $k \ge 1$, and use the following facts: $\psi(y_k) \ge 0$, $y_k - y_{k-1} = -y_k$, $\psi'(y_{k-1}) = \psi(y_{k-1})/y_k^3$, $\psi(y_{k-1}) = \theta(y_k)$, and $y_k \le 1/2$. We get

$$\alpha_k \ge \psi'(y_{k-1})y_k - \psi(y_{k-1}) = \left(\frac{1}{y_k^2} - 1\right)\theta(y_k) \ge \theta(y_k).$$

Hence $\beta_k \le \alpha_k$.

As a result $r_k = \beta_k/2$. Then, using again $y_k - y_{k-1} = -y_k$ and $\psi(y_k) = e^{-3/y_k^2}\psi(y_{k-1})$, we obtain

$$\begin{aligned} r_k &= \tfrac{1}{2}\left(\psi(y_{k-1}) - \psi(y_k) - \psi'(y_k)y_k\right) \\ &= \tfrac{1}{2}\left(\psi(y_{k-1}) - \left(1 + \tfrac{8}{y_k^2}\right)\psi(y_k)\right) \\ &= \tfrac{1}{2}\left(1 - e^{-3/y_k^2}\left(1 + \tfrac{8}{y_k^2}\right)\right)\psi(y_{k-1}). \end{aligned}$$

Since $e^{-3/y_k^2}(1 + 8/y_k^2)$ is positive and tends to zero (by (31)), we have for k large enough

$$2 \le \frac{\psi(y_{k-1})}{r_k} = \frac{2}{1 - e^{-3/y_k^2}\left(1 + \frac{8}{y_k^2}\right)} \le 3,$$

proving the first bracketing. For the second one, just observe that, by (10) and (42), $a_{k-1} = \psi'(y_{k-1}) = \psi(y_{k-1})/y_k^3$. $\square$

Derivatives of f_k. The first derivatives at $z \in \Omega$ are

$$\frac{\partial f_k}{\partial x}(z) = c_k \frac{\partial g_k^\infty}{\partial x}(z) + \varepsilon_k \qquad \text{and} \qquad \frac{\partial f_k}{\partial y}(z) = a_k + c_k \frac{\partial g_k^\infty}{\partial y}(z). \tag{52}$$

Its higher order derivatives for $m,\ n,\ q \in \mathbb{N}$ such that $m + n = q \ge 2$ are

$$\frac{\partial^q f_k}{\partial x^m \partial y^n}(z) = c_k\ \frac{\partial^q g_k^\infty}{\partial x^m \partial y^n}(z). \tag{53}$$

Lemma 15 *Let $m,\ n \in \mathbb{N}$, such that $q := m + n \ge 1$. For any $k \ge 0$:*

$$\lim_{k\to\infty} \frac{\partial^q f_k}{\partial x^m \partial y^n}(z) = 0, \quad \textit{uniformly for } z \in \Omega. \tag{54}$$

Furthermore, there is a positive constant R_q, such that for $z \in \Omega$ and k sufficiently large:

$$\left|\frac{\partial^q f_k}{\partial x^m \partial y^n}(z) - \frac{\partial^q f_{k-1}}{\partial x^m \partial y^n}(z)\right| \le R_q\ \frac{r_k}{y_k^3}. \tag{55}$$

Proof. The first claim of the lemma follows immediately from the formulas of f_k and its derivatives (see (52) and (53)), from the bound (47) obtained in Lemma 12, and the fact that a_k, b_k, c_k, and ε_k tend to zero.

The second result is proven by examining three cases, depending on the values of q, m, and n.

- Case 1: $q \geq 2$. Using (53), the bound (47), and the fact that both c_{k-1} and c_k do not exceed $r_k/4$ (see (14)), we obtain for any $k \geq 1$:

$$\left|\frac{\partial^q f_k}{\partial x^m \partial y^n}(z) - \frac{\partial^q f_{k-1}}{\partial x^m \partial y^n}(z)\right| \leq (c_k + c_{k-1})\, K_q \leq \frac{r_k}{2}\, K_q.$$

  Hence, (55) follows with $R_q = K_q/2$, because $y_k \leq 1$.

- Case 2: $q = m = 1$. Using (52), the bound (47), the fact that both c_{k-1} and c_k do not exceed $r_k/4$, and $|\varepsilon_k| \leq |\varepsilon_{k-1}| \leq r_k/16$ (see (16)), we have for any $k \geq 1$:

$$\left|\frac{\partial f_k}{\partial x}(z) - \frac{\partial f_{k-1}}{\partial x}(z)\right| \leq \frac{r_k}{2}K_1 + 2|\varepsilon_{k-1}| \leq \frac{r_k}{2}\left(K_1 + \frac{1}{4}\right).$$

  Inequality (55) follows as in case 1.

- Case 3: $q = n = 1$. Using (52), the bound (47), the fact that both c_{k-1} and c_k do not exceed $r_k/4$, and $a_k \leq a_{k-1} \leq 3r_k/y_k^3$ (see Lemma 14), we obtain for k large enough:

$$\left|\frac{\partial f_k}{\partial y}(z) - \frac{\partial f_{k-1}}{\partial y}(z)\right| \leq 2a_{k-1} + \frac{r_k}{2}K_1 \leq \frac{6r_k}{y_k^3} + \frac{r_k}{2}K_1 \leq \frac{r_k}{y_k^3}\left(6 + \frac{K_1}{2}\right),$$

  since $y_k \leq 1$. Inequality (55) follows.

□

Derivatives of φ: We do not need an explicit expression for the derivatives of the smoothing function φ, but just bounds. From (23), if $|w| \geq 1$, then $\varphi'(w) = \pm 1$ and $\varphi^{(l)}(w) = 0$ for $l \geq 2$. Therefore, since $\varphi \in C^\infty$, for any integer $l \geq 1$, there exists a positive constant S_l, such that

$$|\varphi^{(l)}(w)| \leq S_l, \quad \text{for any } w \in I\!R. \tag{56}$$

The main result of this section is the following.

Lemma 16 *The cost function F of Example 6 is convex and of class C^∞ on Ω.*

Proof. Using the same arguments as in the beginning of the proof of Lemma 8, we have that F vanishes on $\Omega_- = \{(x, y) \in \operatorname{int}\Omega \mid y < 0\}$ and is C^∞ on $\Omega_+ = \{(x, y) \in \operatorname{int}\Omega \mid y > 0\}$. It is also convex and continuous on Ω.

We still have to show that F is C^∞ on Ω, knowing that it has this smoothness on $\Omega_0 = \Omega_- \cup \Omega_+$. For this, we apply Lemma 10: it is sufficient to show that any partial derivative of F has a continuous extension from Ω_0 to Ω. This extension at a point $(x,y) \in \Omega \backslash \Omega_0$, with $y \neq 0$, is straightforward; therefore, we concentrate on those points of the form $\tilde{z} = (x,0)$, with $x \in [-1.5, 1.5]$.

Let us fix m, $n \in I\!N$, such that $q := m+n > 0$. It is sufficient to show that $\frac{\partial^q F}{\partial x^m \partial y^n}(z_i)$ converges to zero when $z_i = (x_i, y_i) \in \Omega_0$ converges to $\tilde{z}$. We can assume that $y_i > 0$, since otherwise the partial derivative vanishes (an easy case).

For any $i \in I\!N$, there exists $k_i \in I\!N$ such that z_i belongs to the strip Ω_{k_i}. Let us simplify the notation by setting, for $z \in \Omega$

$$h_{k_i}(z) = \frac{f_{k_i}(z) - f_{k_i-1}(z)}{r_{k_i}}.$$

From Lemma 13,

$$F(z_i) = M_{r_{k_i}}(f_{k_i}(z_i), f_{k_i-1}(z_i)) = \frac{1}{2}\left(f_{k_i}(z_i) + f_{k_i-1}(z_i) + r_{k_i}\varphi\left(h_{k_i}(z_i)\right)\right).$$

Therefore

$$\frac{\partial^q F}{\partial x^m \partial y^n}(z_i) = \frac{1}{2}\left(\frac{\partial^q f_{k_i}}{\partial x^m \partial y^n}(z_i) + \frac{\partial^q f_{k_i-1}}{\partial x^m \partial y^n}(z_i) + r_{k_i}\, \frac{\partial^q(\varphi \circ h_{k_i})}{\partial x^m \partial y^n}(z_i)\right). \quad (57)$$

When $z_i \to \tilde{z}$, $k_i \to \infty$, and thus Lemma 15 implies that the first two terms in (57) tend to zero.

Using (45), the last term in (57) can be written

$$r_{k_i}\frac{\partial^q(\varphi \circ h_{k_i})}{\partial x^m \partial y^n}(z_i) = r_{k_i}\sum_j \alpha_j\, \varphi^{(l_j)}(h_{k_i}(z_i))\, \mathcal{P}_j, \quad (58)$$

where α_j are integers, $l_j \in \{1, 2, \ldots, q\}$, and $\mathcal{P}_j$ are products of l_j partial derivatives of h_{k_i} at z_i, with sum of orders equal to q. By (56), for all $l_j \in \{1, 2, \cdots, q\}$

$$|\varphi^{(l_j)}(h_{k_i}(z_i))| \leq \bar{S}_q, \quad (59)$$

where $\bar{S}_q := \max\{1, S_1, \ldots, S_q\}$ is independent of i. Let us now examine the product $\mathcal{P}_j$:

$$|\mathcal{P}_j| = \left|\prod_{l=1}^{l_j} \frac{\partial^{q_l} h_{k_i}}{\partial x^{m_l}\, \partial y^{n_l}}(z_i)\right| = \frac{1}{r_{k_i}^{l_j}} \prod_{l=1}^{l_j} \left|\frac{\partial^{q_l} f_{k_i}}{\partial x^{m_l}\, \partial y^{n_l}}(z_i) - \frac{\partial^{q_l} f_{k_i-1}}{\partial x^{m_l}\, \partial y^{n_l}}(z_i)\right|.$$

By Lemma 15, for i (hence k_i) sufficiently large

$$|\mathcal{P}_j| \leq \frac{1}{r_{k_i}^{l_j}} \prod_{l=1}^{l_j}\left(R_{q_l}\, \frac{r_{k_i}}{y_{k_i}^3}\right) \leq \left(\frac{\bar{R}_q}{y_{k_i}^3}\right)^q,$$

where $\bar{R}_q := \max\{1, R_1, \ldots, R_q\}$ is independent of i. Combining this estimate with (58) and (59) yields

$$r_{k_i}\left|\frac{\partial^q(\varphi \circ h_{k_i})}{\partial x^m \partial y^n}(z_i)\right| \leq r_{k_i}\, \bar{S}_q\, \frac{\bar{R}_q^q}{y_{k_i}^{3q}} \sum_j \alpha_j \leq 2K\, \frac{r_{k_i}}{y_{k_i}^{3q}},$$

where K is a constant independent of i. From Lemma 14, $r_{k_i} \le \psi(y_{k_i-1})/2 = \theta(y_{k_i})/2$. Hence

$$r_{k_i}\left|\frac{\partial^q(\varphi\circ h_{k_i})}{\partial x^m \partial y^n}(z_i)\right| \le K\,\frac{\theta(y_{k_i})}{y_{k_i}^{3q}}.$$

By (31) and $y_{k_i} \to 0$, the right hand side of this inequality converges to zero. □

To conclude, let us look at the central path (4), associated with function F in Example 6. It depends on the values of ε_k satisfying (15).

If $\varepsilon_k = 0$ for all $k \in I\!N$, the central path shows an antenna pattern, as before. Indeed, for $y \in (y_k, y_{k-1})$, $f_k(\cdot, y)$ and $f_{k-1}(\cdot, y)$ are both uniquely minimized at $x = 0$ ($g_k^0(\cdot, y)$ is positive and strictly convex and θ is strictly convex on $(0, 2/3)$), hence so is $F(\cdot, y)$ (see Lemmas 13-7). For $y = y_k$, $F(\cdot, y_k) = f_k(\cdot, y_k)$, which is minimized for $x \in [-1, +1]$.

If $\varepsilon_k \ne 0$ and $|\varepsilon_k|$ is sufficiently small, for all $k \in I\!N$, the central path shows a zig-zag pattern with infinite variation. Indeed, $F(\cdot, y_k)$ is minimized at a single point $x \notin [-1, +1]$ characterized by

$$\frac{\partial F}{\partial x}(x, y_k) = \frac{\partial f_k}{\partial x}(x, y_k) = c_k \frac{\partial g_k^\infty}{\partial x}(x, y_k) + \varepsilon_k = 0.$$

We have

$$\begin{aligned}\frac{\partial g_k^\infty}{\partial x}(x, y_k) &= \frac{18}{(g_k^0(x, y_k))^3}\frac{\partial g_k^0}{\partial x}(x, y_k)\theta\left(\frac{1}{3}g_k^0(x, y_k)\right)\\ &= \begin{cases}\frac{18}{(x+1)^3}\,\theta(|x+1|/3) & \text{if } x < -1\\ \frac{18}{(x-1)^3}\,\theta(|x-1|/3) & \text{if } x > +1.\end{cases}\end{aligned}$$

Then an easy calculation shows that, if $\varepsilon_k > 0$ is small enough, the minimizer $x \in [-1.5, -1)$, and if $-\varepsilon_k > 0$ is small enough, the minimizer $x \in (1, 1.5]$. On the other hand, due to the convexity of g_k and Lemmas 13-7, the central path is entirely in Ω and, by Lemma 9, it is a C^∞ function of $y > 0$.

Additional notes

In the last examples we proved smoothness results for F in $\Omega \subset I\!R^2$. It is easy to modify these examples so that the central path does not change much and F becomes smooth or C^∞ in the whole space. Define the following smooth convex function:

$$z = (x, y) \in I\!R^2 \mapsto q(z) = \alpha\left(x^2 - (1.4)^2\right) + \psi(y),$$

where $\alpha > 0$ is large and ψ is the support function used to form F, and set $\bar{F}(\cdot) = \max\{F(\cdot), q(\cdot)\}$. This function coincides with F in the region of interest (for $|x| \le 1.4$ and $y \in [-0.5, 1]$). Smoothing $\bar{F}$ by the method of Section 4 makes it as smooth as F, and the central path stays almost the same.

As we did in the end of Section 3, strict complementarity will be satisfied by all our examples by adding a linear term ay to F.

It is not possible to construct an antenna-like or a zig-zag central path with an objective function that is analytic. Monteiro and Zhou [22] ensure indeed that if the functions involved in the definition of the problem are analytic, together with some other mild assumptions, then the primal central path is a curve that converges to a single point in the optimal set (related results are given by Cominetti [8] and Champion [5]). It is therefore remarkable that C^∞ smoothness still allowed us to construct central paths with weird layout.

7 Complexity results

We now discuss the consequences of these examples on the complexity of path-following algorithms. Here the problem dimension is fixed, and we seek complexity results in terms of the precision of a solution. We shall establish polynomiality for some algorithms in the sense that for any $\epsilon > 0$ given, the algorithm reaches points (x^j, y^j) such that $y^j \leq y^* + \epsilon$ in $h = O(|\ln \epsilon|)$ iterations, where y^* is the optimal value of y.

We shall establish the following facts.

- (i) No penalized function in the format used here can be self-concordant, and hence polynomiality of path-following algorithms cannot be proved.
- (ii) If we assume that a predictor-corrector scheme needs only a fixed number of steps to overcome each turn of the zig-zag, then the algorithm will be polynomial.
- (iii) A slight change in the problem statement destroys the property above, and worse, makes the predictor-corrector method converge to a non-optimal set in infinite time.
- (iv) Using the Nesterov-Nemirovskii formulation of the problem with a self-concordant barrier, the trajectory will be a straight line for $\varepsilon_k = 0$ and a zig-zag for the case $\varepsilon_k \neq 0$. This zig-zag will be damped with the changes introduced in (iii), but we explain why we believe that it may be of unbounded variation.

Below, we consider any function F of class C^3, constructed as in Section 5 or 6, and any penalty function $(x, y) \mapsto p(y)$.

Outcome (i): lack of self-concordence of the penalized functions

To prove (i), we use the following result quoted from Nesterov and Nemirovskii [23, Corollary 2.1.1].

Lemma 17 *Let Φ be self-concordant on $Q \subseteq \mathbb{R}^n$. Then the subspace*

$$\{h \in \mathbb{R}^n \mid h^\top \nabla^2 \Phi(x) h = 0\}$$

does not depend on $x \in Q$.

Item (i) above follows from the following lemma.

Lemma 18 *The function $z = (x, y) \in \Omega \mapsto \Phi(z) = F(z) + p(y)$ is not self-concordant.*

Proof. Let $z = (0, 2^{-k})$ for some $k \in I\!N$. Then by construction F is affine along $z + \lambda h$ for $h = (1, 0)$, $\lambda \in [-1, 1]$. Hence $h^\top \nabla^2 F(z) h = 0$ and also $h^\top \nabla^2 \Phi(z) h = 0$ because p does not depend on x. For any $z = (x, y)$ such that $y \neq 2^{-k}$, $k \in I\!N$, $h^\top \nabla^2 \Phi(z) h \neq 0$ (the proof is similar to the one in Lemma 9). The result follows from Lemma 17, completing the proof. $\square$

Outcome (ii): polynomiality of an ideal predictor-corrector algorithm

Imagine that a predictor-corrector method (possibly using an oracle to perform perfect corrector steps) follows the central path for some of our zig-zag examples so that each turn of the zig-zag is overcome in less than P steps, for some fixed $P \in I\!N$. Then starting from $z = (0, 1)$, a point $z = (x, y)$ such that $y \leq \epsilon = 2^{-L}$, $L > 0$, would be achieved in no more than PL steps, because of our definition $y_k = 2^{-k}$. Hence such a predictor-corrector algorithm would be polynomial.

Outcome (iii): examples without convergence

The examples become more interesting with the following modification. Let $a > 0$ and $0 < \delta < 1$ be given constants, and F be any of our functions with a zig-zag central path. Consider the problem

$$\begin{array}{rl} \text{minimize} & F(z) + ay \\ \text{subject to} & y \geq -\delta. \end{array}$$

The function F is flat in a large region $\Upsilon \subset \{z \in \Omega \mid y \in [-\delta, 0]\}$. The central points for $y > 0$ are the same zig-zag as for our examples. Although we are not interested in the central points for $y \in [-\delta, 0]$, it is easy to see that all points in Υ are central.

Now let us examine the behavior of a predictor-corrector algorithm. It will be the same as for the case (ii). The sequence (x^i, y^i) generated by the method will satisfy $y^i \to 0$, and hence no accumulation point of the sequence can be optimal: the sequence converges to the non-optimal set $\{z \in I\!R^2 \mid y = 0\}$.

Outcome (iv): self-concordant barrier

We now discuss the central paths for our problems using a self-concordant barrier. This discussion will not be rigorous, because this is difficult material, beyond the scope of this paper.

Let us state the problem according to Nesterov and Nemirovskii:

$$\begin{array}{rl} \text{minimize} & t \\ \text{subject to} & G(x, y) - t \leq 0 \\ & y \geq -\delta, \end{array}$$

where $G(x,y) = F(x,y) + ay$, $\delta \geq 0$, $a \geq 0$ and F is the objective function in any one of our examples. Let $z^* = (x^*, y^*)$ be an optimal solution and define $G^* = G(x^*, y^*)$. We shall parameterize the central path as a set of centers: for each $T \in (G^*, 1]$, consider the set

$$\begin{array}{rcl} \Lambda_T & = & \{(x,y,t) \in I\!R^3 \mid G(x,y) - t \leq 0,\ y \geq -\delta,\ t \leq T\} \\ \Lambda_T^o & = & \operatorname{int} \Lambda_T. \end{array}$$

For $T \in (G^*, 1]$, let $p_T(x,y,t) : \Lambda_T^o \to I\!R$ be a self-concordant barrier family: here we assume that p_T is the universal barrier defined by Nesterov and Nemirovskii [23, section 2.5] for the set Λ_T. Then the central points will be

$$w(T) = (x(T), y(T), t(T)) = \operatorname*{argmin}_{w \in \Lambda_T^o} p_T(w).$$

Let us comment on the behavior of the central path in some cases. Note that $T \mapsto w(T)$ is a continuous curve.

(a) $\varepsilon_k = 0$: in this case, $p_T(x,y,t)$ is symmetrical in relation to x for any given y, t. This means that $x(T) = 0$ for all $T \in (G^*, 1]$, and the central path for the original problem is the straight line $\{(0,y) \mid y \in (-\delta, 1]\}$. The antenna branches vanish.

(b) $\varepsilon_k \neq 0$, $\delta \geq 0$, $a \geq 0$. Given T, let $w(T)$ be the central point and let $y^+(T)$

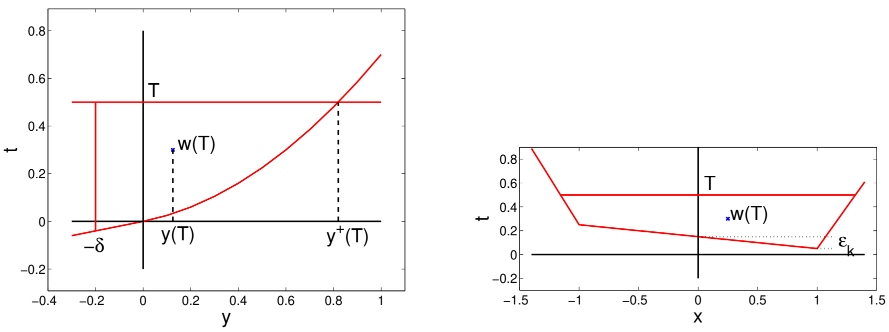

Figure 7: Two bi-dimensional cuts of Λ_T.

be such that $G(0, y^+(T)) = T$, as in Figure 7. From the definition of $w(T)$, we obviously have: assuming that $x(T)$ is known, $(y(T), t(T))$ is the minimizer of the restriction of p_T to the two-dimensional set defined by $x = x(T)$, represented in Figure 7 left. Similarly, given $y(T)$, $(x(T), t(T))$ is the minimizer of p_T in the constant y slice of Λ_T, represented in Figure 7 right.

To reason with analytic centers, let us use the following intuitive assumption: the centers of both two-dimensional sets above are well approximated by the analytic centers of these sets considered as two-dimensional, i.e., using the universal barrier defined for these two-dimensional sets. Then we can state some reasonable hints on the centers.

Assume that $y(T) = 2^{-k}$ for some $k \in I\!N$.

From Figure 7 right, we see that if $\varepsilon_k < 0$, then $x(T) > 0$, if $\varepsilon_k > 0$, then $x(T) < 0$, and the trajectory makes a zig-zag.

The value $|x(T)|$ depends on the relation between $|\varepsilon_k|$ (the tilt in the basis of the set) and the height $T - G(0, 2^{-k})$.

Let us assume that $\delta = 0$. The case $\delta > 0$ is simpler and we comment it below.

As T approaches zero (remember that $T \in (0, 1]$), the relation $y(T)/y^+(T)$ must tend to a constant. Due to the simple shape of the set in Figure 7 left, this constant should be positive. By the same reasoning, as $T \to 0$, the following relation will also converge to a constant:

$$\frac{G(0, y(T))}{T - G(0, y(T))} \to c \in (0, 1).$$

Let us examine the sequence $T^k \to 0$ chosen so that $y(T^k) = 2^{-k}$ (the turns of the zig-zag). Define $z_k = (0, y(T^k))$. We have $G(z_k) = \psi(2^{-k}) + 2^{-k}a$, where ψ is the support function defined by (9). Hence

$$\frac{\psi(2^{-k}) + 2^{-k}a}{T - G(z_k)} \to c. \tag{60}$$

We have two cases:

$\mathbf{a > 0}$. In all our examples, $\lim_{y \to 0^+} \psi(y)/y = 0$. Hence $\psi(2^{-k})/2^{-k} \to 0$ and we conclude from (60) that

$$\frac{\psi(2^{-k})}{T - G(z_k)} \to 0.$$

Since $\varepsilon_k < \psi(2^{-k})$ by construction,

$$\frac{\varepsilon_k}{T - G(z_k)} \to 0.$$

This means that the zig-zag is damped.

If $\delta > 0$ and $a > 0$, as $T \to -\delta$, the zig-zag will also be damped by a slight adaptation of the same reasoning. The last part of the trajectory, for $T \leq 0$, will be a straight line.

There is another way of proving that the trajectory will be a damped zig-zag: a predictor-corrector algorithm developed in Nesterov and Nemirovskii [23] (see Section 3.5 - primal parallel trajectories method) has polynomial complexity. We have seen that if the zig-zag is not damped, the predictor-corrector algorithm will converge to a wrong point in infinite time.

$\mathbf{a = 0}$. This is the most interesting case, for which we only have a guess at the present state of our research. Assume that we are using the simplest objective function, as in Example 3. Then ε_k is of the order of $\psi(2^{-k})$ and we conclude from (60) that

$$\frac{\varepsilon_k}{T - G(z_k)} \to \bar{c} > 0,$$

where $\bar{c}$ is a constant. We arrive to the following amazing conclusion (if our hints are correct): in this case the trajectory will still be a zig-zag with infinite variation. But the predictor-corrector algorithm will be polynomial, as we saw in (ii), and is proved by Nesterov and Nemirovskii.

References

[1] Adler, I., Monteiro, R. D. C. (1991): Limiting behavior of the affine scaling continuous trajectories for linear programming problems. Mathematical Programming **50**, 29–51

[2] Auslender, A., Cominetti, R., Haddou, M. (1997): Asymptotic analysis for penalty and barrier methods in convex and linear programming. Mathematics of Operations Research **22**, 43–62

[3] Bayer, D. A., Lagarias, J. C. (1989): The nonlinear geometry of linear programming, Part I : Affine and projective scaling trajectories. Transactions of the American Mathematical Society **314**, 499–526

[4] Byrd, R. H., Gilbert, J. C., Nocedal, J. (2000): A trust region method based on interior point techniques for nonlinear programming. Mathematical Programming **89**, 149–185

[5] Champion, T. (2001): Asymptotic Convergence of Penalty Trajectories in Convex Programming with Multiple Solutions. Ph.D. thesis, Département de Mathématique, Université de Montpellier II, France

[6] Chen, C., Mangasarian, O. L. (1995): Smoothing methods for convex inequalities and linear complementary problems. Mathematical Programming **71**, 1–112

[7] Chen, C., Mangasarian, O. L. (1996): A class of smoothing functions for non-linear and mixed complementary problems. Computational Optimization and Applications **5**, 97–138

[8] Cominetti, R. (1999): Nonlinear averages and convergence of penalty trajectories in convex programming. in: Théra, M., Tichatschke, R., eds., Ill-posed variational problems and regularization techniques, Springer Verlag, Berlin, vol. 477 of *Lecture Notes in Economics and Mathematical System*, 65–78

[9] Conn, A. R., Gould, N. I. M., Orban, D., Toint, P. L. (1999): A primal-dual trust region algorithm for minimizing a non-convex function subject to general inequality and linear equality constraints. Tech. Rep. RAL-TR-1999-054, Computation Science and Engeneering Department, Atlas Centre, Rutherford Appleton Laboratory, Oxon, OX11 OQX

[10] den Hertog, D. (1992): Interior Point Approach to Linear, Quadratic and Convex Programming. Mathematics and its Applications 277, Kluwer Academic Publishers, Dordrecht

[11] Dennis, J. E., Heinkenschloss, M., Vicente, L. N. (1994): Trust-region interior-point SQP algorithms for a class of nonlinear programming problems. Tech. Rep. 94-45, Department of Computational and Applied Mathematics, Rice University

[12] Fiacco, A. V., McCormick, G. P. (1968): Nonlinear Programming : Sequential Unconstrained Minimization Techniques. John Wiley & Sons, New York, reprint : Volume 4 of *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, PA 19104–2688, USA, 1990

[13] Fleming, W. H. (1965): Functions of Several Variables. Addison-Wesley Publishing Company, United States

[14] Forsgren, A., Gill, P. E. (1998): Primal-dual interior methods for nonconvex nonlinear programming. SIAM Journal on Optimization **8**, 1132–1152

[15] Gay, D. M., Overton, M. L., Wright, M. H. (1998): A primal-dual interior method for nonconvex nonlinear programming. in: Yuan, Y.-X., ed., Advances in Nonlinear Programming, Kluwer Academic Publishers, Dordretch, 31–56

[16] Gonzaga, C. C. (1992): Path following methods for linear programming. SIAM Review **34**, 167–227

[17] Hiriart-Urruty, J.-B., Lemaréchal, C. (1993):
Convex Analysis and Minimization Algorithms I. Springer-Verlag, New York

[18] Jansen, B. (1997): Interior Point Techniques in Optimization. Applied Optimization 6, Kluwer Academic Publishers, Dordrecht

[19] Kojima, M., Megiddo, N., Noma, T., Yoshise, A. (1991): A unified approach to interior point algorithms for linear complementarity problems. in: Lecture Notes in Computer Science, Springer Verlag, Berlin, vol. 538

[20] McLinden, L. (1980): An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem. Pacific Journal of Mathematics **88**, 101–161

[21] Megiddo, N. (1989): Pathways to the optimal set in linear programming. in: Megiddo, N., ed., Progress in Mathematical Programming : Interior Point and Related Methods, Springer Verlag, New York, 131–158, identical version in : *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, pages 1–35, 1986

[22] Monteiro, R. D. C., Zhou, F. (1998): On the existence and convergence of the central path for convex programming and some duality results. Computational Optimization and Applications **10**, 51–77

[23] Nesterov, Y. E., Nemirovskii, A. S. (1994): Interior-Point Polynomial Algorithms in Convex Programming, vol. 13 of *SIAM Studies in Applied Mathematics 13*. SIAM, Philadelphia

[24] Rockafellar, R. T., Wets, R. (1998): Variational Analysis. Springer

[25] Roos, C., Terlaky, T., Vial, J. P. (1997): Theory and Algorithms for Linear Optimization: An Interior Point Approach. John Wiley & Sons, Chichester

[26] Terlaky, T., ed. (1996): Interior Point Methods of Mathematical Programming. Kluwer Academic Press

[27] Vanderbei, R. J., Shanno, D. F. (1999): An interior-point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications **13**, 231–252

[28] Vavasis, S. A., Ye, Y. (1996): A primal-dual interior point method whose running time depends only on the constraint matrix. Mathematical Programming **74**, 79–120

[29] Whitney, H. (1934): Functions differentiable on the boundaries of regions. Ann. of Math. **35**, 482–485

[30] Witzgall, C., Boggs, P. T., Domich, P. D. (1990): On the convergence behavior of trajectories for linear programming. in: Lagarias, J. C., Todd, M. J., eds., Mathematical Developments Arising from Linear Programming, American Mathematical Society, Providence, Rhode Island, USA, vol. 114 of *Contemporary Mathematics*, 161–188

[31] Wright, S. J. (1997): Primal-Dual Interior-Point Methods. SIAM Publication, Philadelphia