JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 120, No. 2, pp. 247–273, February 2004 (©2004)

Steepest Descent with Curvature Dynamical System[1,2]

F. Alvarez[3] and A. Cabot[4]

Communicated by G. Di Pillo

Abstract. Let H be a real Hilbert space and let $\langle \cdot,\cdot \rangle$ denote the corresponding scalar product. Given a $\mathscr{C}^2$ function Φ: $H \to \mathbb{R}$ that is bounded from below, we consider the following dynamical system:

$$\text{(SDC)} \quad \dot{x}(t) + \lambda(x(t))\nabla\Phi(x(t)) = 0, \qquad t \geq 0,$$

where $\lambda(x)$ corresponds to a quadratic approximation to a linear search technique in the direction $-\nabla\Phi(x)$. The term $\lambda(x)$ is connected intimately with the normal curvature radius $\rho(x)$ in the direction $\nabla\Phi(x)$. The remarkable property of (SDC) lies in the fact that the gradient norm $|\nabla\Phi(x(t))|$ decreases exponentially to zero when $t \to +\infty$.

When Φ is a convex function which is nonsmooth or lacks strong convexity, we consider a parametric family $\{\Phi_\epsilon, \epsilon > 0\}$ of smooth strongly convex approximations of Φ and we couple this approximation scheme with the (SDC) system. More precisely, we are interested in the following dynamical system:

$$\text{(ASDC)} \quad \dot{x}(t) + \lambda(t, x(t))\nabla_x\Phi(t, x(t)) = 0, \qquad t \geq 0,$$

where $\lambda(t, x)$ is a time-dependent function involving a curvature term. We find conditions on the approximating family and on $\epsilon(\cdot)$ ensuring the asymptotic convergence of the solution trajectories $x(\cdot)$ toward a particular solution of the problem $\min \{\Phi(x), x \in H\}$. Applications to barrier

[1]The authors are grateful to Professors P. Armand, H. Attouch, R. Cominetti, P. Redont, and to the anonymous referees whose comments and suggestions have done much to bring this article to its present form.

[2]This work was partially supported by ECOS C00E05. The research of the first author was supported by grants Fondecyt 1020610 and Fondap from Conicyt-Chile and by Iniciativa Científica Milenio from Mideplan-Chile.

[3]Profesor Asistente, Departamento de Ingeniería Matemática, Centro de Modelamiento y Matemática, Universidad de Chile, Santiago, Chile.

[4]Maître de Conférences, Laboratoire LACO, Faculté des Sciences, Université de Limoges, Limoges, France.

and penalty methods in linear programming and to viscosity methods are given.

Key Words. Gradient-like systems, asymptotic analysis, convex optimization, approximate methods, optimal trajectories.

1. Introduction

Let H be a real Hilbert space with scalar product and corresponding norm denoted by $\langle\cdot,\cdot\rangle$ and $|\cdot|$ respectively. We are interested in the asymptotic behavior at infinity of the trajectories generated by some first-order in time dynamical systems associated with an objective function that we want to minimize. More precisely, given a smooth function $\Phi: H \to \mathbb{R}$ which is bounded from below, we consider the following initial-value problem of the gradient type:

$$\dot{x} + \lambda(t, x)\nabla\Phi(x) = 0, \qquad t \geq 0,$$
$$x(0) = x_0 \in H,$$

where $\lambda(t, x)$ is the stepsize coefficient.

When $\lambda(t, x) \equiv \lambda_0 > 0$ (constant), this is simply the steepest descent method with fixed stepsize. The main drawback of this method lies in the fact that there is no influence of the shape of the function on the stepsize coefficient. A first alternative consists in taking a coefficient $\lambda = \lambda(t)$ depending on only the time. But in practice, the trajectories are not known a priori, so that it becomes apparent that such a strategy will not be very useful.

Let us consider a coefficient of the type $\lambda = \lambda(x)$. More precisely, assuming that Φ is of class $\mathscr{C}^2$ on H and that

$$\langle\nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle > 0,$$
$$\text{for all } x \in H \text{ such that } |\nabla\Phi(x)| \neq 0,$$

we study the following dynamical system:

$$\text{(SDC)} \quad \dot{x} + [|\nabla\Phi(x)|^2/\langle\nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle]\nabla\Phi(x) = 0, \qquad t \geq 0,$$

where SDC means steepest descent with curvature. This choice of $\lambda(x)$ corresponds to a quadratic approximation to a linear search technique in the direction $-\nabla\Phi(x)$. Of course, the trajectories of (SDC) are simply the steepest descent ones, but they are described with different speeds. The remarkable property of (SDC) lies in the fact that the gradient norm $|\nabla\Phi(x(t))|$ decreases exponentially to zero when $t \to +\infty$. A similar situation occurs in the case of the continuous Newton method (Ref. 1). It is interesting to notice that our expression of $\lambda(x)$ is related closely to the normal curvature radius $\rho(x)$ in the direction $\nabla\Phi(x)$. Roughly speaking, the motion on the trajectory is accelerated [resp. decelerated] when the normal curvature radius $\rho(x)$ increases [resp. decreases].

The numerical implementation of (SDC) is very simple because the computation of an approximate of the Hessian term $\langle\nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle$ is reduced to two evaluations of the function Φ. From a numerical point of view, the coefficient $\lambda(x)$ in the (SDC) system is better adapted to the shape of the function Φ than a simple constant coefficient λ: this is due to the influence of the second-order information contained in $\lambda(x)$.

When minimizing a convex function Φ which is nonsmooth (in constrained optimization problems for example) or which is not strongly convex, it is suggested to replace Φ by a better behaved parametric approximation. More precisely, let us consider the family of problems

$$(\mathrm{P}_\epsilon) \quad \min\{\Phi_\epsilon(x), x\in H\},$$

where for any $\epsilon>0$, the function $\Phi_\epsilon: H\rightarrow\mathbb{R}$ is a smooth strongly convex function such that $\Phi_\epsilon\rightarrow\Phi$ as $\epsilon\rightarrow 0$. The coupling of dynamical systems with parametric approximation schemes has been developed by Attouch-Cominetti (Ref. 2) and Cominetti-Courdurier (Ref. 3) for the steepest descent method and by Alvarez-Pérez (Ref. 1) for the Newton method. Our goal here is to couple the (SDC) system with the previous approximation scheme. For that purpose, we introduce the method of approximate steepest descent with curvature,

$$(\mathrm{ASDC}) \quad \dot{x}+[\langle\nabla_x\Phi(t,x), \nabla_x\Phi(t,x)+\nabla^2_{tx}\Phi(t,x)\rangle/ \langle\nabla^2_{xx}\Phi(t,x)\nabla_x\Phi(t,x), \nabla_x\Phi(t,x)\rangle]\nabla_x\Phi(t,x)=0, \qquad t\geq 0,$$

where for simplicity of notation we write $\Phi(t,x)$ instead of $\Phi_{\epsilon(t)}(x)$. The solutions to (ASDC) satisfy

$$|\nabla_x\Phi(t,x(t))| = e^{-t}|\nabla_x\Phi(0,x_0)|.$$

In a relatively general setting, we obtain a weak condition on the function $\epsilon(\cdot)$ ensuring the asymptotic convergence of the solution trajectories $x(\cdot)$ toward a particular solution of the approximated problem. Finally, we apply our results to barrier and penalty methods in linear programming and to viscosity methods.

2. Preliminaries

2.1. Descent Direction and Scaling. We begin by recalling some well-known facts about the classical steepest descent algorithm for the unconstrained minimization of a continuously differentiable function Φ: $H \to \mathbb{R}$, which consists in recursively computing

$$\text{(SDA)} \quad x^{k+1} = x^k + \lambda_k d^k,$$

where $d^k = -\nabla\Phi(x^k)$ is the steepest descent direction at x^k and the steplength $\lambda_k > 0$ is an approximate solution of the following one-dimensional minimization problem:

$$\min_{\lambda > 0} \Phi(x^k + \lambda d^k).$$

In practice, the stepsize $\lambda_k > 0$ is selected by means of an inexact line search rule with the aim of ensuring an appropriate reduction in the objective function Φ. Although (SDA) is globally convergent under some conditions, for a poorly scaled optimization problem (that is, when the corresponding objective function has highly elongated level sets), the steepest descent direction may not provide much reduction in the function, forcing the line search method to choose a very small steplength. This is due to the fact that the gradient vector may be nearly orthogonal to the direction that leads to the minimizer, producing zigzagging iterates that converge very slowly. Moreover, in such situations, even divergence may occur if one does not proceed carefully with the control of the stepsize.

We remark that (SDA) is a discrete version of the first-order differential equation

$$\text{(SD)} \quad \dot{x} + \nabla\Phi(x) = 0, \qquad t > 0.$$

Indeed, the Euler polygonal method for approximating the solution trajectory $x(t)$ consists in applying successively the iterative scheme (SDA), the size of the integration step λ_k being selected in order to keep the approximation error $e_k = |x(t_k) - x^k|$, with $t_k = \sum_{i=0}^{k-1} \lambda_i$, within prescribed bounds over a fixed time interval. Of course, the numerical integration methods of Runge–Kutta type and Adams type can be considered also. Nevertheless, it is important to note that, if the objective function is poorly scaled, then (SD) is a stiff differential system, which requires the integration step to remain small despite slow changes in the state variable; otherwise, numerical instabilities may produce drastic increases in the approximation error (see Refs. 4–5).

Typically, stiffness occurs when there are some components of $x(t)$ that decay exponentially fast with quite different speed. For instance, take a quadratic function,

$$\Phi(x) = (1/2)\langle Ax, x\rangle,$$

defined on $H = \mathbb{R}^n$, where A is a $n \times n$ symmetric and positive-definite real matrix. The corresponding (SD) system is

$$\text{(SD)} \quad \dot{x} + Ax = 0,$$

whose general solution is obtained as a linear combination of fundamental solutions of the type

$$x_i(t) = e^{-\mu_i t} v_i,$$

where $0 < \mu_1 \leq \cdots \leq \mu_n$ are the eigenvalues of A. If $\mu_1 \ll \mu_n$, then the level sets are very elongated, the corresponding fundamental solutions of (SD) have quite different convergence rates, and there appears an inherent numerical instability. This drawback holds for any descent method that is highly sensitive to poor scaling; thus, an appropriate selection of the stepsize is crucial.

For comparison, let us now consider the Newton iteration

$$x^{k+1} = x^k + \lambda_k d^k,$$

where the descent direction is given by

$$d^k = -\nabla^2\Phi(x^k)^{-1}\nabla\Phi(x^k),$$

provided that the Hessian $\nabla^2\Phi(x^k)$ is positive definite. The continuous Newton method is given by the differential equation

$$\text{(N)} \quad \dot{x} + \nabla^2\Phi(x)^{-1}\nabla\Phi(x) = 0, \qquad t > 0.$$

The scale invariance of this method can be illustrated simply by applying it to a quadratic function,

$$\Phi(x) = (1/2)\langle Ax, x\rangle.$$

In fact, in such case, the discrete Newton algorithm (with unit stepsize) converges in a single iteration, while the solutions to (N) are straight line trajectories of the type

$$x(t) = e^{-t} x_0,$$

independently of the eigenvalues of A.

Generally speaking, algorithms that satisfy some kind of scale invariance, such as the Newton method, are expected to behave better than those that do not. We refer the reader to Refs. 6–8 for deeper discussions of descent methods, poor scaling, and line search rules.

2.2. Asymptotic Convergence of SD. Let us recall briefly some asymptotic results concerning the continuous steepest descent method. The following theorem summarizes the main results about the differential system (SD).

Theorem 2.1. Let H be a Hilbert space and let $\Phi: H \to \mathbb{R}$ be a $\mathscr{C}^1$ function which is bounded from below on H. Assume that $\nabla\Phi$ is Lipschitz continuous on the bounded subsets of H. Then, the following properties hold:

(i) For every $x_0 \in H$, there exists a unique maximal solution $x: \mathbb{R}_+ \to H$ of (SD), which is of class $\mathscr{C}^1$ and satisfies the initial condition $x(0) = x_0$.

(ii) If the trajectory x is bounded [i.e., $x \in L^\infty([0, +\infty); H)$], then $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$.

(iii) Assuming that Φ is convex and that Argmin $\Phi \neq \emptyset$, then there exists $x_\infty \in$ Argmin Φ such that $x(t) \rightharpoonup x_\infty$ weakly in H as $t \to +\infty$.

Points (i) and (ii) are classical. Point (iii) uses an argument due to Opial (Ref. 9). We refer the reader to Refs. 10–12 for all details and more general results in this direction.

Let us introduce in (SD) a continuous stepsize coefficient λ: $[0, +\infty) \to (0, +\infty)$ as follows:

$$(\mathrm{SD})_{\lambda(t)} \quad \dot{x} + \lambda(t)\nabla\Phi(x) = 0, \qquad t > 0.$$

Defining

$$s(t) := \int_0^t \lambda(\tau) d\tau,$$

we have

$$\frac{ds}{dt}(t) = \lambda(t) > 0 \quad \text{and} \quad s(0) = 0.$$

Thus, setting

$$s_\infty := \lim_{t\to+\infty} s(t),$$

the function s: $[0, +\infty) \to [0, s_\infty)$ is onto and increasing. It is direct to verify that, if $y(\cdot)$ is the solution of

$$\dot{y}(s) + \nabla\Phi(y(s)) = 0, \qquad s \in [0, s_\infty),$$

with the initial condition $y(0) = x(0)$, then $x(t) := y(s(t))$ is the solution of $(\mathrm{SD})_{\lambda(t)}$. In this sense, the trajectories associated with $(\mathrm{SD})_{\lambda(t)}$ are obtained by reparametrization in time of those generated by the standard (SD). If we assume that

$$\int_0^{+\infty} \lambda(\tau) d\tau = +\infty,$$

then $s_\infty = +\infty$ and the asymptotic results at infinity for (SD) are valid for $(\mathrm{SD})_{\lambda(t)}$.

In the general case, however, since the trajectories are not known a priori, it is clear that a coefficient $\lambda = \lambda(t)$ depending on only t will not be very useful. In this direction, it is more natural to consider a space-dependent coefficient $\lambda = \lambda(x)$. This is the aim of the next section.

3. Steepest Descent with Curvature

3.1. SDC Dynamical System. The introduction of a variable steplength coefficient in (SD) allows one to have a certain control on the speed at which the steepest descent trajectories are described. This is particularly relevant for those minimization problems that are poorly scaled (cf. Section 2.1) in order to prevent the numerical instabilities that are associated with the steepest descent direction. Since in practice one has only some local information about the objective function Φ, it seems natural to consider a space-dependent coefficient $\lambda = \lambda(x)$ that takes into account this kind of data. More precisely, let us consider a differential equation of the type

$$\dot{x} + \lambda(x)\nabla\Phi(x) = 0, \qquad t > 0,$$

where $\Phi: H \to \mathbb{R}$ is a smooth function that is bounded from below. If $x(t)$ is a solution to this equation and Φ is of class $\mathscr{C}^2$, then

$$\begin{aligned}(d/dt)[(1/2)|\nabla\Phi(x)|^2] &= \langle \nabla\Phi(x), \nabla^2\Phi(x)\dot{x}\rangle \\ &= -\lambda(x)\langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle,\end{aligned}$$

where $\nabla^2\Phi(x)$ is the Hessian of Φ at $x \in H$. This suggests the following choice for the steplength

$$\lambda(x) := |\nabla\Phi(x)|^2 / \langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle, \tag{1}$$

whenever $\langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle \neq 0$. Indeed, if $\lambda(x)$ is given by (1), then we get

$$(d/dt)[(1/2)|\nabla\Phi(x)|^2] = -|\nabla\Phi(x)|^2$$

over the whole trajectory, which yields

$$|\nabla\Phi(x(t))| = |\nabla\Phi(x_0)|e^{-t}, \tag{2}$$

where $x(0) = x_0 \in H$ is the initial condition. This is a nice scale-invariant property because the rate of convergence is independent of Φ. In some sense, the use of second-order information about Φ gives a normalized exponential decay under no strong convexity condition.

It is worth pointing out that the stepsize coefficient defined by (1) corresponds to a line search technique by quadratic approximation. In fact, given $x \in H$, define

$$q(\lambda) := \Phi(x - \lambda \nabla \Phi(x)),$$

so that the optimal steplength along the steepest descent direction is given by the minimum of $q(\lambda)$. But the second-order approximation of $q(\lambda)$ on a neighborhood of 0 is given by the Taylor formula

$$q(\lambda) = q(0) + \lambda q'(0) + (\lambda^2/2) q''(0) + o(\lambda^2),$$

where it is direct to verify that

$$q(0) = \Phi(x),$$

$$q'(0) = -|\nabla \Phi(x)|^2,$$

$$q''(0) = \langle \nabla^2 \Phi(x) \nabla \Phi(x), \nabla \Phi(x) \rangle.$$

If $q''(0) > 0$, then the quadratic approximation of $q(\lambda)$ attains its minimum at

$$\lambda(x) = -q'(0)/q''(0),$$

which coincides with (1).

Another interesting feature of (1) is its connection with the notion of curvature. Indeed, if $M \subset H \times \mathbb{R}$ stands for the submanifold that is described by the equation $y = \Phi(x)$, then the normal curvature of M at x in the direction $h \in H$ is given by

$$\kappa(x, h) = [1/\sqrt{1 + |\nabla \Phi(x)|^2}] \langle \nabla^2 \Phi(x) h, h \rangle / |h|^2.$$

Consequently, (1) may be written equivalently

$$\lambda(x) = \rho(x) / \sqrt{1 + |\nabla \Phi(x)|^2},$$

where

$$\rho(x) = 1/\kappa(x, \nabla \Phi(x))$$

is the radius of normal curvature in the direction $\nabla \Phi(x)$. According to our previous discussion, we are interested in the following steepest descent with curvature dynamical system:

$$(\mathrm{SDC}; x_0) \quad \dot{x} + [|\nabla \Phi(x)|^2 / \langle \nabla^2 \Phi(x) \nabla \Phi(x), \nabla \Phi(x) \rangle] \nabla \Phi(x) = 0, \qquad t > 0,$$

with initial condition

$$x(0) = x_0 \in H.$$

The first-order differential equation in $(\mathrm{SDC}; x_0)$ may be written as

$$\dot{x} = F(x)$$

with

$$F(x) = -\rho(x)\nabla\Phi(x)/\sqrt{1+|\nabla\Phi(x)|^2}.$$

We remark that this vector field combines two effects simultaneously: (i) a nondegenerate pseudonormalization of the steepest descent direction when dividing by $\sqrt{1+|\nabla\Phi(x)|^2}$; (ii) a steplength $\rho(x)$ along the normalized direction that takes into account the local geometry (curvature) of the objective function.

In the following, we study the (SDC) system on a connected component Ω of the open set $H\setminus\nabla\Phi^{-1}(0)$ and we assume that

$$\langle\nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle > 0, \qquad \forall x\in\Omega. \tag{3}$$

Suppose that $\nabla^2\Phi$ is locally Lipschitz continuous on Ω, so the vector field in (SDC) is also locally Lipschitz continuous on Ω. Fix $x_0\in\Omega$. Then, the Cauchy-Lipschitz theorem gives the existence of a unique maximal solution of (SDC; x_0), defined on an interval of the type $[0, T_{\max})$ for some $T_{\max} > 0$. We claim that $T_{\max} = +\infty$. To see this, define the function $s(t)$ by setting

$$s(0) = 0 \quad \text{and} \quad \frac{ds}{dt}(t) = \lambda(x(t)), \qquad \text{for all } t\in[0, T_{\max}).$$

By (3), we get an increasing diffeomorphism,

$$s: [0, T_{\max}) \to [0, \bar{s}), \qquad \text{for some } \bar{s} > 0.$$

On the other hand, given the maximal solution $y: [0, +\infty) \to H$ of the steepest descent equation

$$(\mathrm{SD}; x_0) \quad \dot{y}(s) + \nabla\Phi(y(s)) = 0, \qquad s > 0,$$

with initial condition $y(0) = x_0$, let us define $x(t) := y(s(t))$. It is direct to verify that $x(t)$ is the solution of (SDC). Suppose that $\bar{s} < +\infty$. Consequently,

$$\lim_{t\to T_{\max}} x(t) = \lim_{s\to\bar{s}} y(s) = y(\bar{s}) \in \Omega,$$

so that we can apply to (SDC) the Cauchy-Lipschitz theorem with initial condition $y(\bar{s})$, which allows one to extend the maximal solution to an interval that is strictly larger than $[0, T_{\max})$, contradicting the maximality of $T_{\max}$ if we assume that $T_{\max} < +\infty$. On the other hand, if $\bar{s} = +\infty$, then

$$\lim_{t\to T_{\max}} \nabla\Phi(x(t)) = \lim_{s\to+\infty} \nabla\Phi(y(s)) = 0.$$

As

$$|\nabla\Phi(x(t))| = |\nabla\Phi(x_0)|e^{-t}, \qquad \text{for all } t\in[0, T_{\max}),$$

we obtain $T_{\max} = +\infty$. In any case, we deduce that $T_{\max} = +\infty$ as claimed.

Summarizing, we have the following result.

Theorem 3.1. Let Φ: $H \to \mathbb{R}$ be a $\mathscr{C}^2$ function such that $\nabla\Phi$ and $\nabla^2\Phi$ are locally Lipschitz continuous on H. Suppose that Φ is bounded from below and that $\Omega \neq \emptyset$ is a connected component of $H \setminus \nabla\Phi^{-1}(0)$ satisfying (3). Then, for each $x_0 \in \Omega$, there exists a unique global solution x: $[0, +\infty) \to H$ of (SDC; x_0), which is of class $\mathscr{C}^1$ and satisfies $x(t) \in \Omega$ for all $t \in [0, +\infty)$. Furthermore, the decay property (2) holds.

Remark 3.1. We have proved actually that there exists an increasing diffeomorphism s: $[0, +\infty) \to [0, +\infty)$ such that $x(t) = y(s(t))$, where $y(s)$ is the maximal solution to (SD; x_0). Therefore, the trajectories generated by (SD; x_0) and (SDC; x_0) are the same in the geometric sense, but they are described with different speeds. Consequently, the asymptotic convergence properties of (SD) can be extended immediately to (SDC). In particular, if Φ is convex with Argmin $\Phi \neq \emptyset$, then there exists $x_\infty \in$ Argmin Φ such that

$$x(t) \rightharpoonup x_\infty, \qquad \text{weakly in } H \text{ as } t \to +\infty.$$

Remark 3.2. For one-dimensional problems ($H = \mathbb{R}$), the (SDC; x_0) system coincides with the continuous Newton method,

$$(\mathrm{N}; x_0) \quad \dot{x} + \nabla^2\Phi(x)^{-1}\nabla\Phi(x) = 0, \qquad t > 0,$$

with initial condition $x(0) = x_0$. In higher dimensions, (N) has the advantage of satisfying scale invariance. Furthermore, it is easy to see that the Newton trajectory satisfies

$$\nabla\Phi(x(t)) = \nabla\Phi(x_0)e^{-t},$$

which ensures that all the components of the gradient decay at the same exponential rate. On the other hand, (SDC) gives only (2), which is also a scale invariant property but may not eliminate completely the intrinsic instability associated with the steepest descent direction.

3.2. Some Variants of the SDC System. In this section, we study the asymptotic properties of some variants of the (SDC) system.

SDC$_\alpha$ System. Given any $\alpha > 0$, consider the following more general dynamical system:

$$(\mathrm{SDC})_\alpha \quad \dot{x} + \lambda_\alpha(x)\nabla\Phi(x) = 0, \qquad t > 0,$$

where the stepsize coefficient $\lambda_\alpha(x)$ is given by

$$\lambda_\alpha(x) = |\nabla\Phi(x)|^\alpha / \langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle. \tag{4}$$

Theorem 3.2. Under the hypotheses of Theorem 3.1, for each $x_0 \in \Omega$ there exists a unique maximal solution $x: [0, T_{\max}) \to H$ of $(\mathrm{SDC})_\alpha$, which is of class $\mathscr{C}^1$ and satisfies $x(0) = x_0$ and $x(t) \in \Omega$ for all $t \in [0, T_{\max})$. Moreover, for all $t \in [0, T_{\max})$,

$$\left(\frac{d}{dt}\right)|\nabla\Phi(x(t))|^2 = -2|\nabla\Phi(x(t))|^\alpha \quad \text{and} \quad \lim_{t\to T_{\max}} \nabla\Phi(x(t)) = 0.$$

More precisely, the following estimates hold.

(i) Polynomial Decay: If $\alpha > 2$, then $T_{\max} = +\infty$ and

$$|\nabla\Phi(x(t))| = [C + (\alpha - 2)t]^{-1/(\alpha-2)},$$

with

$$C = |\nabla\Phi(x_0)|^{-(\alpha-2)}.$$

(ii) Exponential Decay: If $\alpha = 2$, then $T_{\max} = +\infty$ and

$$|\nabla\Phi(x(t))| = |\nabla\Phi(x_0)|e^{-t}.$$

(iii) Finite-Time Convergence: If $\alpha < 2$, then

$$T_{\max} = |\nabla\Phi(x_0)|^{(2-\alpha)}/(2-\alpha) < +\infty$$

and

$$|\nabla\Phi(x(t))| = [(2-\alpha)(T_{\max} - t)]^{1/(2-\alpha)}.$$

Proof. We follow the proof of Theorem 3.1, where the case $\alpha = 2$ was treated. The Cauchy-Lipschitz theorem gives the existence of a unique maximal solution $x(\cdot)$ of $(\mathrm{SDC})_\alpha$, defined on $[0, T_{\max})$ for some $T_{\max} > 0$. On the other hand, it is direct to verify that

$$\left(\frac{d}{dt}\right)|\nabla\Phi(x(t))|^2 = -2|\nabla\Phi(x(t))|^\alpha.$$

Integrating this equation, we obtain

$$|\nabla\Phi(x(t))| = |\nabla\Phi(x_0)|e^{-t}$$

in the case where $\alpha = 2$; when $\alpha \neq 2$, we get

$$|\nabla\Phi(x(t))| = [C + (\alpha - 2)t]^{1/(2-\alpha)}, \tag{5}$$

with

$$C = |\nabla\Phi(x_0)|^{2-\alpha}.$$

Let us define the function $s(\cdot)$ by $s(0) = 0$ and, for all $t \in [0, T_{\max})$,

$$\frac{ds}{dt}(t) = |\nabla\Phi(x(t))|^{\alpha} / \langle \nabla^2\Phi(x(t))\nabla\Phi(x(t)), \nabla\Phi(x(t))\rangle.$$

It is clear that

$$s: [0, T_{\max}) \to [0, s(T_{\max}))$$

is an increasing diffeomorphism of class $\mathscr{C}^1$. If $y(\cdot)$ is the solution of

$$\text{(SD)} \quad \dot{y}(s) + \nabla\Phi(y(s)) = 0, \qquad s > 0,$$

with initial condition $y(0) = x_0$, then $x(t) := y(s(t))$ is the solution of $(\text{SDC})_\alpha$. Let us now prove that $s(T_{\max}) = +\infty$. Let us argue by contradiction and assume that $s(T_{\max}) < +\infty$, so that

$$\begin{aligned} \lim_{t \to T_{\max}} x(t) &= \lim_{s \to s(T_{\max})} y(s) \\ &= y(s(T_{\max})) \in \Omega. \end{aligned} \tag{6}$$

We now distinguish the cases where $T_{\max} < +\infty$ and $T_{\max} = +\infty$.

Case $T_{\max} < +\infty$. In view of (6), we can apply to $(\text{SDC})_\alpha$ the Cauchy-Lipschitz theorem with initial condition $(T_{\max}, y(s(T_{\max}))) \in \mathbb{R}_+ \times \Omega$, which allows one to extend the solution $x(\cdot)$ to an interval that is strictly larger than $[0, T_{\max})$, contradicting the maximality of the solution.

Case $T_{\max} = +\infty$. In view of (5), it is clear that $T_{\max} = +\infty$ implies $\alpha \geq 2$ and then we have

$$\lim_{t \to +\infty} \nabla\Phi(x(t)) = 0,$$

which contradicts (6).

Finally, we have

$$\lim_{t \to T_{\max}} |\nabla\Phi(x(t))| = \lim_{s \to +\infty} |\nabla\Phi(y(s))| = 0;$$

therefore, if $\alpha \geq 2$, then $T_{\max} = +\infty$; if $\alpha < 2$, then

$$T_{\max} = C/(2 - \alpha) = |\nabla\Phi(x_0)|^{2-\alpha}/(2 - \alpha). \qquad \square$$

Corollary 3.1. Under the hypotheses of Theorem 3.2 with $\alpha \geq 1$, if there exists $M > 0$ such that, for all $x \in \Omega$,

$$\langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle \geq M|\nabla\Phi(x)|^2, \tag{7}$$

then there exists $x_\infty \in \nabla\Phi^{-1}(0)$ such that

$$\lim_{t\to T_{\max}} x(t) = x_\infty,$$

where $x: [0, T_{\max}) \to \Omega$ is the maximal solution of $(\mathrm{SDC})_\alpha$. Furthermore:

(i) if $\alpha > 2$, then $|\dot{x}(t)| = O(t^{-(\alpha-1)/(\alpha-2)})$ in the neighborhood of $+\infty$;
(ii) if $\alpha = 2$, then $|\dot{x}(t)| = O(e^{-t})$ in the neighborhood of $+\infty$;
(iii) if $\alpha \in]1, 2[$, then $\lim_{t\to T_{\max}} |\dot{x}(t)| = 0$;
(iv) if $\alpha = 1$, then $|\dot{x}|$ is bounded on $[0, T_{\max})$.

Proof. It follows from (7) that we obtain

$$\begin{aligned}|\dot{x}(t)| &= |\nabla\Phi(x(t))|^{\alpha+1} / \langle \nabla^2\Phi(x(t))\nabla\Phi(x(t)), \nabla\Phi(x(t))\rangle \\ &\le (1/M)|\nabla\Phi(x(t))|^{\alpha-1}. \end{aligned} \tag{8}$$

We are going to use (8) by distinguishing the cases $\alpha > 2$, $\alpha = 2$, and $\alpha \in [1, 2[$.

(i) Case $\alpha > 2$. By Theorem 3.2,

$$|\nabla\Phi(x(t))| \sim Ct^{-1/(\alpha-2)}, \qquad t \to +\infty.$$

In view of (8), then we have

$$|\dot{x}(t)| = O(t^{-(\alpha-1)/(\alpha-2)})$$

in the neighborhood of $+\infty$. On the other hand,

$$(\alpha - 1)/(\alpha - 2) > 1 \quad \text{implies} \quad |\dot{x}| \in L^1(0, +\infty).$$

As a consequence $s - \lim_{t\to+\infty} x(t)$ exists. Denoting by x_∞ this limit, we have

$$\nabla\Phi(x_\infty) = \lim_{t\to+\infty} \nabla\Phi(x(t)) = 0,$$

i.e.,

$$x_\infty \in \nabla\Phi^{-1}(0).$$

(ii) Case $\alpha = 2$. By Theorem 3.2,

$$|\nabla\Phi(x(t))| \sim Ce^{-t}.$$

Hence, $|\dot{x}(t)| = O(e^{-t})$ in the neighborhood of $+\infty$. Then, we conclude as in the case $\alpha > 2$.

(iii)–(iv) Case $\alpha \in [1, 2[$. By Theorem 3.2,

$$\lim_{t\to T_{\max}} |\nabla\Phi(x(t))| = 0.$$

Then, in view of (8), it is clear that

$$\lim_{t\to T_{\max}} |\dot{x}(t)| = 0, \qquad \text{if } \alpha > 1,$$

and that $|\dot{x}|$ is bounded if $\alpha = 1$. In both cases, the function $|\dot{x}|$ is bounded on $[0, T_{\max})$, which combined with $T_{\max} < +\infty$ yields immediately the existence of

$$x_\infty = \lim_{t\to T_{\max}} x(t). \qquad \square$$

SDPC System. We now turn to the steepest descent with pure curvature system,

$$\text{(SDPC)} \quad \dot{x} + \rho(x)\nabla\Phi(x) = 0, \qquad t > 0,$$

where

$$\rho(x) = 1/\kappa(x, \nabla\Phi(x))$$

is the radius of normal curvature in the direction $\nabla\Phi(x)$.

Proposition 3.1. Under the hypotheses of Theorem 3.1, for each $x_0 \in \Omega$ there exists a unique solution $x\colon [0, +\infty) \to \Omega$ of (SDPC) satisfying $x(0) = x_0$; moreover,

$$|\nabla\Phi(x(t))| = 1/\sinh(t + C), \tag{9}$$

with

$$C = \operatorname{argsinh}(1/|\nabla\Phi(x_0)|).$$

Proof. We give only the proof of (9). Define the Lyapunov function

$$L(t) = \sqrt{1 + |\nabla\Phi(x(t))|^2},$$

which satisfies

$$\dot{L}(t) = -|\nabla\Phi(x(t))|^2;$$

since

$$|\nabla\Phi(x(t))|^2 = L(t)^2 - 1,$$

$L(t)$ solves the differential equation

$$\dot{L}(t) = 1 - L(t)^2.$$

But

$$L(t) \geq 1, \qquad \text{for all } t \geq 0,$$

so we have classically

$$L(t) = \coth(t + C), \qquad \text{for some } C \in \mathbb{R}.$$

Therefore,

$$\begin{aligned}|\nabla\Phi(x(t))| &= \sqrt{L(t)^2 - 1}\\ &= 1/\sinh(t + C).\end{aligned}$$

By taking $t = 0$ in the previous formula, we find that

$$C = \operatorname{argsinh}(1/|\nabla\Phi(x_0)|). \qquad \square$$

4. Coupling SDC with Parametric Approximation in Convex Optimization

When the objective function Φ is nonsmooth or has lack of strong convexity, a standard technique consists in replacing Φ with a family of better behaved parametric approximations $(\Phi_\epsilon)_{\epsilon>0}$ such that $\Phi_\epsilon \to \Phi$ in a suitable variational sense.

Let us consider an abstract family of problems,

$$(\mathrm{P}_\epsilon) \quad \min\{\Phi_\epsilon(x): x \in H\},$$

where for each $\epsilon > 0$, $\Phi_\epsilon: H \to \mathbb{R}$ is a smooth convex function such that $\Phi_\epsilon \to \Phi$ as $\epsilon \to 0^+$. Moreover, we suppose that Φ_ϵ is $\beta(\epsilon)$-strongly convex with $\beta(\epsilon) > 0$; that is to say,

$$\langle \nabla\Phi_\epsilon(x) - \nabla\Phi_\epsilon(y), x - y\rangle \geq \beta(\epsilon)|x - y|^2, \qquad \forall x, y \in H. \tag{10}$$

We assume also that there exists an optimal path of the optimal solutions x_ϵ of (P_ϵ) that satisfies the following statement:

$$\text{there exists } x^* \in \operatorname{Argmin} \Phi \text{ such that } \lim_{\epsilon \to 0+} x_\epsilon = x^*. \tag{11}$$

Observe that x_ϵ is characterized as the unique solution to the stationary condition $\nabla\Phi_\epsilon(x_\epsilon) = 0$, which together with the strong convexity condition (10) yields

$$|x - x_\epsilon| \leq |\nabla\Phi_\epsilon(x)|/\beta(\epsilon), \qquad \forall x \in H. \tag{12}$$

4.1. Coupling SD and the Newton Method with Approximation. In the sequel, we consider a parametrization $\epsilon: [0, +\infty) \to \mathbb{R}_+$, which is supposed to be strictly positive and decreasing to 0; for simplicity of notation, we write $\Phi(t, x)$ instead of $\Phi_{\epsilon(t)}(x)$. Concerning the coupling of (SD) with parametric approximation schemes, the so-called descent and approximation dynamical asymptotical method

$$\text{(DADA)} \quad \dot{x} + \nabla_x \Phi(t, x) = 0, \qquad t > 0,$$

was introduced in Ref. 2, where it is proved that, if $\epsilon(t)$ converges to 0 sufficiently slow, then the solution $x(t)$ of (DADA) approaches asymptotically the optimal path $x_{\epsilon(t)}$ and is attracted toward $x^* = \lim_{\epsilon \to 0} x_\epsilon$. The speed of convergence of $\epsilon(t)$ is measured in terms of the strong convexity parameter $\beta(\epsilon)$. See also Ref. 13. In the framework of convex programming, Cominetti and Courdurier (Ref. 3) couple a general penalty scheme with the steepest descent method and they prove that, under very mild conditions on ϵ [ϵ measurable with $\epsilon(t) \to 0$ as $t \to +\infty$], the trajectory $x(t)$ converges toward some element of the optimal set, which may be different to x^*, the limit of the optimal path.

When the approximate function Φ_ϵ is of class $\mathscr{C}^2$, it is natural to consider second-order descent methods such as Newton's. In Ref. 1, the following approximate continuous Newton method is studied:

$$\text{(ACN)} \quad \dot{x} + \nabla^2_{xx} \Phi(t, x)^{-1} [\nabla_x \Phi(t, x) + \nabla^2_{tx} \Phi(t, x)] = 0, \qquad t > 0.$$

Here, the vector field combines a Newton correction with an extrapolation direction that takes into account the changes in the objective function $\Phi(t, \cdot)$ as t increases. It is direct to verify that, if $x(t)$ is a solution of (ACN), then

$$\left(\frac{d}{dt}\right) \nabla_x \Phi(t, x) = -\nabla_x \Phi(t, x);$$

consequently,

$$\nabla_x \Phi(t, x(t)) = e^{-t} \nabla_x \Phi(0, x_0) = e^{-t} \nabla \Phi_{\epsilon 0}(x_0), \tag{13}$$

and by (12), we get

$$|x(t) - x_{\epsilon(t)}| \leq |\nabla \Phi_{\epsilon 0}(x_0)| e^{-t} / \beta(\epsilon(t)).$$

Therefore, under the condition

$$\lim_{t \to +\infty} e^{-t} / \beta(\epsilon(t)) = 0, \tag{14}$$

we have that, thanks to (11),

$$\lim_{t \to +\infty} x(t) = \lim_{\epsilon \to 0+} x_\epsilon = x^* \in \text{Argmin } \Phi. \tag{15}$$

4.2. Approximate Steepest Descent with Curvature. We begin by observing that there is not a standard extension of the (SDC) system to the parametric approximation setting. The most straightforward alternative seems to be the nonautonomous differential equation

$$\dot{x} + [|\nabla_x \Phi(t, x)|^2 / \langle \nabla^2_{xx} \Phi(t, x) \nabla_x \Phi(t, x), \nabla_x \Phi(t, x) \rangle] \nabla_x \Phi(t, x) = 0,$$

where $\Phi(t, x)$ stands for $\Phi_{\epsilon(t)}(x)$ and the parametrization $\epsilon(t)$ is as before. However, there is no reason for this equation to be the most efficient way of combining (SDC) with an approximation scheme. Instead, the dynamical system that we propose is the following:

$$\text{(ASDC)} \quad \dot{x} + \lambda(t, x) \nabla_x \Phi(t, x) = 0, \qquad t > 0,$$

where

$$\lambda(t, x) = \langle \nabla_x \Phi(t, x), \nabla_x \Phi(t, x) + \nabla^2_{tx} \Phi(t, x) \rangle / \langle \nabla^2_{xx} \Phi(t, x) \nabla_x \Phi(t, x), \nabla_x \Phi(t, x) \rangle. \tag{16}$$

We refer to (ASDC) as the approximate steepest descent with curvature method. From now on, when dealing with (ASDC), we assume that Φ is of class $\mathscr{C}^2$, which imposes a differentiability condition on ϵ. The remarkable property of this system is that, if $x(t)$ denotes its solution, then

$$|\nabla_x \Phi(t, x(t))| = |\nabla_x \Phi(0, x_0)| e^{-t}. \tag{17}$$

Indeed, setting

$$E(t) = (1/2) |\nabla_x \Phi(t, x(t))|^2,$$

we obtain

$$\begin{aligned} \dot{E}(t) &= \langle \nabla_x \Phi(t, x), \nabla^2_{xx} \Phi(t, x) \dot{x} + \nabla^2_{tx} \Phi(t, x) \rangle \\ &= -\lambda(t, x) \langle \nabla_x \Phi(t, x), \nabla^2_{xx} \Phi(t, x) \nabla_x \Phi(t, x) \rangle + \langle \nabla_x \Phi(t, x), \nabla^2_{tx} \Phi(t, x) \rangle \\ &= -|\nabla_x \Phi(t, x)|^2 - \langle \nabla_x \Phi(t, x), \nabla^2_{tx} \Phi(t, x) \rangle + \langle \nabla_x \Phi(t, x), \nabla^2_{tx} \Phi(t, x) \rangle \\ &= -2E(t), \end{aligned}$$

from which (17) follows immediately. Therefore, under the condition (14), we have (15) as well as with (ACN), but for a simpler evolution problem from the computational point of view.

4.3. Examples. In the sequel, we take a differentiable parametrization $\epsilon(\cdot) \in C^1([0, +\infty); \mathbb{R}_+)$ that satisfies $\epsilon(t) \to 0^+$ as $t \to +\infty$.

Example 4.1. Tikhonov Approximation. The Tikhonov approximation of a closed, proper, and convex function $\Phi: H \to \mathbb{R} \cup \{+\infty\}$ is given by

$$\Phi_\epsilon(x) = \Phi(x) + (\epsilon/2) |x|^2,$$

which regularizes Φ by adding a strongly convex term. If Argmin $\Phi \neq \emptyset$, then the unique minimizer x_ϵ of Φ_ϵ converges as $\epsilon \to 0^+$ to the element of minimal norm in Argmin Φ (see Ref. 14). This approximation scheme is known also as the viscosity method (see Ref. 15). Assume that $\Phi \in C^2(H; \mathbb{R})$ and set

$$\Phi(t, x) = \Phi(x) + [\epsilon(t)/2]|x|^2.$$

In this case,

$$\nabla_x \Phi(t, x) = \nabla \Phi(x) + \epsilon(t) x,$$
$$\nabla^2_{xx} \Phi(t, x) = \nabla^2 \Phi(x) + \epsilon(t) I,$$
$$\nabla^2_{tx} \Phi(t, x) = \dot{\epsilon}(t) x,$$

and (ASDC) corresponds to

$$\dot{x} + \lambda(t, x)[\nabla \Phi(x) + \epsilon(t) x] = 0, \qquad t > 0,$$
$$x(0) = x_0,$$

with

$$\lambda(t, x) = \langle \nabla \Phi(x) + \epsilon(t) x, \nabla \Phi(x) + \epsilon(t) x + \dot{\epsilon}(t) x \rangle / \langle (\nabla^2 \Phi(x) + \epsilon(t) I)(\nabla \Phi(x) + \epsilon(t) x), \nabla \Phi(x) + \epsilon(t) x \rangle.$$

Of course, the strong convexity condition (10) holds in this situation with $\beta(\epsilon) = \epsilon$. Therefore, the convergence condition (14) is

$$\lim_{t \to +\infty} e^{-t}/\epsilon(t) = 0. \tag{18}$$

Take for instance

$$\epsilon(t) = e^{-\alpha t}, \quad \text{with } 0 < \alpha < 1,$$

or

$$\epsilon(t) = (1 + t)^{-r}, \quad \text{with } r > 0.$$

Example 4.2. Log-Barrier in Linear Programming. Let us consider the linear program

$$\text{(LP)} \quad \min_{x \in \mathbb{R}^n} \{c^T x : Ax \leq b\},$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ is a full-rank matrix, and $b \in \mathbb{R}^m$. Assume that the open set

$$\Omega = \{x \in \mathbb{R}^n : Ax < b\}$$

is nonempty and bounded. The log-barrier associated with (LP) is given by

$$\Phi_\epsilon(x) = c^T x - \epsilon \sum_{i=1}^{m} \log(b_i - A_i^T x),$$

where $A_i \in \mathbb{R}^n$ denotes the ith row of A. It is well-known that the approximate problem $\min_{x\in\mathbb{R}^n}\Phi_\epsilon(x)$ has a unique solution x_ϵ which converges toward the so-called analytic center of the optimal set S(LP), a particular solution of (LP) that is characterized as the unique solution of

$$\max\left\{\sum_{i\notin I_0} \log(b_i - A_i^T x) : x \in \mathrm{S(LP)}\right\},$$

where

$$I_0 = \{i : A_i^T x = b_i, \qquad \text{for all } x \in \mathrm{S(LP)}\}. \tag{19}$$

Defining

$$\Phi(t, x) = \Phi_{\epsilon(t)}(x),$$

we have that $\Phi(\cdot,\cdot) \in C^\infty((0,+\infty)\times\Omega;\mathbb{R})$. A straightforward computation yields that, for every $x \in \Omega$,

$$\nabla_x \Phi(t, x) = c + \epsilon(t) A^T d(x),$$
$$\nabla_{tx}^2 \Phi(t, x) = \dot{\epsilon}(t) A^T d(x),$$
$$\nabla_{xx}^2 \Phi(t, x) = \epsilon(t) A^T D(x)^2 A,$$

where the vector-valued function $d : \Omega \to \mathbb{R}^m$ is defined by

$$d(x) = (1/(b_i - A_i^T x) : i = 1, \ldots, m)$$

and $D(x)$ is the diagonal matrix

$$D(x) = \mathrm{diag}(d(x)).$$

Thus,

$$\lambda(t, x) = \langle c + \epsilon(t)A^T d(x), c + (\epsilon(t) + \dot{\epsilon}(t))A^T d(x)\rangle / \epsilon(t)\langle A^T D(x)^2 A(c + \epsilon(t)A^T d(x)), c + \epsilon(t)A^T d(x)\rangle.$$

Furthermore, it is simple to verify that the strong convexity hypothesis holds with $\beta(\epsilon) = \beta_0\epsilon$ for a suitable constant $\beta_0 > 0$ so that the convergence condition (14) is also given by (18). Figure 1 shows the (ASDC) trajectory and the optimal path in the case of a two-dimensional problem with the following data:

$$c = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{20a}$$

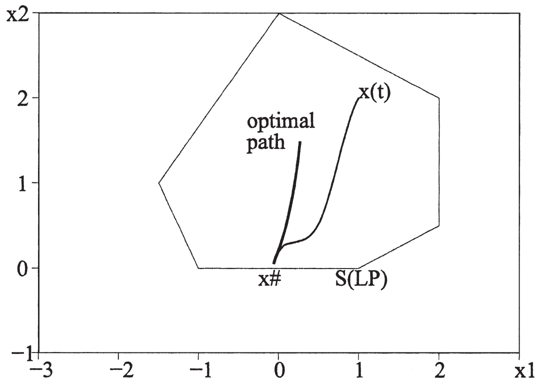

Fig. 1. Log-barrier in linear programming: optimal path and trajectory of the (ASDC) system. Convergence toward the analytic center $x^{\#}$ of S(LP).

$$A^T = \begin{bmatrix} 0 & 1 & 1 & 1/2 & -4/3 & -2 \\ -1 & -2 & 0 & 1 & 1 & -1 \end{bmatrix}, \tag{20b}$$

$$b^T = (0, 1, 2, 3, 3, 2). \tag{20c}$$

The control parameter is taken equal to

$$\epsilon(t) = 1/\log(t+2)$$

and the initial condition satisfies $x_0^T = (1, 2)$.

Example 4.3. Exponential Penalty and Dual Convergence. The exponential penalty for the linear program (LP) is given by

$$\Phi_\epsilon(x) = c^T x + \epsilon \sum_{i=1}^{m} \exp[(A_i^T x - b_i)/\epsilon].$$

If we assume that S(LP) is nonempty and bounded, then there exists a unique optimal path x_ϵ of the minimizers of Φ_ϵ which converges to a particular point $x^* \in$ S(LP) called the centroid (see Ref. 16). Defining

$$\Phi(t, x) = \Phi_{\epsilon(t)}(x),$$

we have

$$\nabla_x \Phi(t, x) = c + A^T \mu(t, x),$$
$$\nabla_{tx}^2 \Phi(t, x) = -(\dot{\epsilon}(t)/\epsilon(t)) A^T S(\mu(t, x)),$$
$$\nabla_{xx}^2 \Phi(t, x) = (1/\epsilon(t)) A^T D(t, x) A,$$

where $\mu(\cdot,\cdot): [0,+\infty)\times\mathbb{R}^n\to\mathbb{R}^m$ is defined by

$$\mu(t,x) := (\exp[(A_i^T x - b_i)/\epsilon(t)]: i = 1,\dots,m),$$

the diagonal matrix by

$$D(t,x) := \mathrm{diag}(\mu(t,x)),$$

and $S: \mathbb{R}^m_+\to\mathbb{R}^m$ by

$$S(\mu) := (\mu_i \log \mu_i: i = 1,\dots,m),$$

with the convention that $0\log 0 = 0$. Thus,

$$\lambda(t,x) = \langle c + A^T\mu(t,x), \epsilon(t)c + \epsilon(t)A^T\mu(t,x) - \dot{\epsilon}(t)A^T S(\mu(t,x))\rangle / \langle A^T D(t,x)A(c + A^T\mu(t,x)), c + A^T\mu(t,x)\rangle.$$

The approximate Φ_ϵ is strongly convex with

$$\beta(\epsilon) = (\alpha/\epsilon)e^{-M/\epsilon}$$

for suitable positive constants α and M. If we have

$$\lim_{t\to+\infty} \epsilon(t)e^{-t+M/\epsilon(t)} = 0, \qquad \text{for all } M > 0,$$

then $x(t)$ converges toward the centroid x^* of S(LP); moreover, there exists a constant $C>0$ such that

$$|x(t) - x_{\epsilon(t)}| \le C\epsilon(t)e^{-t+M/\epsilon(t)}. \tag{21}$$

Figure 2 illustrates the convergence of the (ASDC) trajectory toward the centroid x^* of S(LP). The data are the same as in (20), the control parameter

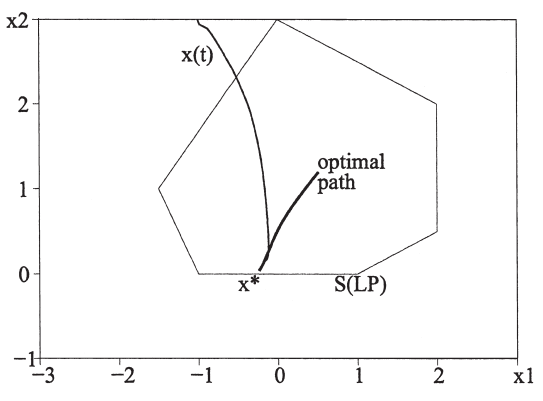

Fig. 2. Exponential penalty in linear programming: optimal path and trajectory of the (ASDC) system. Convergence toward the centroid x^* of S(LP).

equals

$$\epsilon(t) = 1 \log(t+2)$$

and the initial condition is given by $x_0^T = (-1, 3)$.

On the other hand, the Fenchel duality theory applied to this situation permits to associate with (P_ϵ) the following dual problem:

$$(D_\epsilon) \quad \min\{b^T\mu + \epsilon \sum_{i=1}^{m} \mu_i(\log \mu_i - 1): A^T\mu = -c, \mu \geq 0\},$$

which is a penalized version of the classical dual problem of (LP), namely,

$$(D) \quad \min\{b^T\mu: A^T\mu = -c, \mu \geq 0\}.$$

This dual parametric scheme was studied in Ref. 16, where it is shown that, if x_ϵ denotes the optimal solution of (P_ϵ), then the dual optimal path,

$$\mu_\epsilon = (\exp[(A_i^T x_\epsilon - b_i)/\epsilon]: i = 1, \ldots, m),$$

gives the unique optimal solution of (D_ϵ). Moreover, μ_ϵ converges toward a particular solution μ^* of the dual problem (D), characterized as the unique solution of

$$(D_0) \quad \min\left\{\sum_{i \in I_0} \mu_i(\log \mu_i - 1): A'\mu = -c, \mu_i = 0 i \notin I_0, \mu_i \geq 0 i \in I_0\right\},$$

with I_0 given by (19). Define

$$\mu(t) = (\exp[(A_i^T x(t) - b_i)/\epsilon(t)]: i = 1, \ldots, m),$$

with $x(\cdot)$ being solution of (ASDC). Since

$$\mu_i(t) = \exp[A_i^T(x(t) - x_{\epsilon(t)})/\epsilon(t)](\mu_{\epsilon(t)})_i, \qquad \text{for all } i = 1, \ldots, m,$$

if we have

$$\lim_{t \to +\infty} |x(t) - x_{\epsilon(t)}|/\epsilon(t) = 0, \tag{22}$$

then

$$\lim_{t \to +\infty} \mu(t) = \lim_{\epsilon \to 0+} \mu_\epsilon = \mu^*.$$

But

$$\begin{aligned} |x(t) - x_{\epsilon(t)}|/\epsilon(t) &\leq |\nabla_x \Phi_{\epsilon 0}(x_0)| e^{-t}/(\beta(\epsilon(t))\epsilon(t)) \\ &= Ce^{-t + M/\epsilon(t)}, \end{aligned}$$

so that, in order to have (22), it suffices to verify that

$$\lim_{t \to +\infty} e^{-t + M/\epsilon(t)} = 0,$$

or equivalently,

$$\lim_{t\to+\infty} t - M/\epsilon(t) = +\infty, \tag{23}$$

for all $M>0$. For example,

$$\epsilon(t) = (1+t)^{-r}, \qquad \text{with } 0<r<1.$$

Finally, since (22) implies

$$\lim_{t\to+\infty} |x(t) - x_{\epsilon(t)}| = 0,$$

we get

$$(x(t), \mu(t)) \to (x^*, \mu^*), \quad \text{as } t\to+\infty.$$

In this case, the dual convergence for (ASDC) is a consequence of a sharp estimate for the distance between $x(t)$ and the optimal trajectory $x_{\epsilon(t)}$. A similar analysis for the corresponding (DADA) trajectories is more involved; in fact, it is proved in Ref. 13 that, for the convergence of the (DADA) dual trajectories in the exponential penalty framework, it is sufficient to have

$$\lim_{t\to+\infty} e^{\alpha/\epsilon(t)}\dot{\epsilon}(t) = 0, \qquad \text{for all } \alpha>0,$$

for instance,

$$\epsilon(t) = 1/\log(\log(t)),$$

which is much restrictive than (23). Less stringent conditions for dual convergence are given in Ref. 3 by means of a completely different approach.

5. Numerical Experiments in the Unconstrained Case

Given a $\mathscr{C}^2$ function $\Phi: H\to\mathbb{R}$, with $\nabla^2\Phi$ being locally Lipschitz continuous and such that, for all $x\in H\setminus\nabla\Phi^{-1}(0)$,

$$\langle\nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle \neq 0,$$

consider the following discrete version of the (SDC) system:

$$\text{(DSDC)} \quad x^{k+1} = x^k - \lambda_k\nabla\Phi(x^k),$$

where

$$\lambda_k = |\nabla\Phi(x^k)|^2 / |\langle\nabla^2\Phi(x^k)\nabla\Phi(x^k), \nabla\Phi(x^k)\rangle|.$$

The absolute value $|\langle\nabla^2\Phi(x^k)\nabla\Phi(x^k), \nabla\Phi(x^k)\rangle|$ yields a strictly positive stepsize $\lambda_k>0$ whenever $\nabla\Phi(x^k)\neq0$. This permits us to deal with objective functions that are either convex or concave in a neighborhood of x^k.

From the numerical point of view, the main drawback of (DSDC) seems to be the computation of the stepsize λ_k, because of the Hessian term. In fact, it is possible to approximate $\langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle$. by a simple finite-difference technique. Like in Section 3.1, we fix $x \in H$ and define

$$q(\lambda) = \Phi(x - \lambda\nabla\Phi(x)), \qquad \text{for } \lambda \in \mathbb{R},$$

so that

$$q''(0) = \langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle.$$

As $\nabla^2\Phi$ is Lipschitz continuous on a neighborhood of x, so is the function q on a neighborhood of 0; under these conditions, it is well-known that

$$q''(0) = [q(\epsilon) - 2q(0) + q(-\epsilon)]/\epsilon^2 + O(\epsilon),$$

and consequently,

$$\langle \nabla^2\Phi(x)\nabla\Phi(x), \nabla\Phi(x)\rangle \approx [\Phi(x - \epsilon\nabla\Phi(x)) - 2\Phi(x) + \Phi(x + \epsilon\nabla\Phi(x))]/\epsilon^2,$$

for ϵ sufficiently small. Since the value $\Phi(x)$ is supposed to be already known, we just have to compute $\Phi(x - \epsilon\nabla\Phi(x))$ and $\Phi(x + \epsilon\nabla\Phi(x))$, which represents only two extra evaluations. As a consequence, the (DSDC) algorithm can be implemented easily without any excessive computational cost.

We are now going to test the (DSDC) algorithm and compare it with the fixed stepsize(DSD),

$$\text{(DSD)} \quad x^{k+1} = x^k - \lambda\nabla\Phi(x^k),$$

where λ is a fixed constant. We use two objective functions that are well-known in optimization theory: the Beale function (Ref. 17) and the Shekel function (Ref. 18).

Example 5.1. Two-Dimensional Example. Let us start with the Beale function,

$$\Phi(x_1, x_2) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2.$$

This function admits a unique global minimum point $x_{\min} = (3, 0.5)$ on the subset $[-5, 5]\times[-5, 5]$ and its minimum value is $\Phi(x_{\min}) = 0$. It is very rough and steep, especially at the point $(-5, -5)$, leading (DSD) to fail; see Fig. 3 where we have represented the isocontours of the Beale function. Generally speaking, if one chooses a small stepsize $\lambda \ll 1$, the discrete trajectory associated with (DSD) is described very slowly and does not converge toward $x_{\min}$ because the function is very flat in a neighborhood of this point. On the other hand, the choice of a large stepsize in (DSD) gives trajectories that are very unstable and strongly oscillating. On the contrary, when using the (DSDC) method, the trajectory converges to $x_{\min}$ very fast (see Figs. 3 and 4).

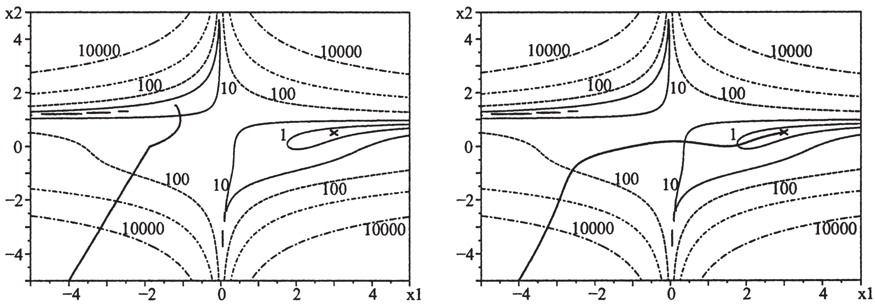

Fig. 3. Minimization of the Beale function. Starting point $x_0 = (-4, -5)$. Minimum point $x_{\min} = (3, 0.5)$. Left: (DSD) trajectory. Right: (DSDC) trajectory.

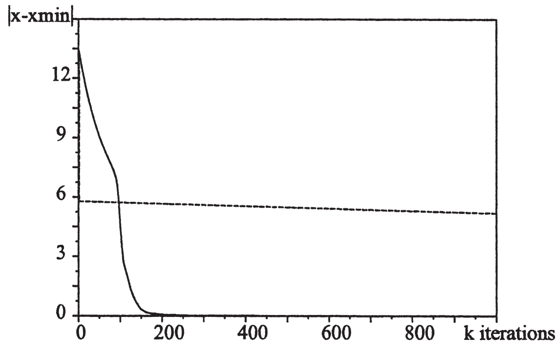

Fig. 4. Convergence history for $|x - x_{\min}|$ during the optimization process. Dashed line: (DSD) algorithm with $\lambda = 2.10^{-5}$; continuous line: (DSDC) algorithm.

Example 5.2. Four-Dimensional Example. We consider now the Shekel function,

$$\Phi(x) = -\sum_{i=1}^{5} 1/(|x - d_i|^2 + c_i),$$

where $x \in [0, 10]^4$. The coefficients c_i and d_i are respectively given by

$$(c_1, c_2, c_3, c_4, c_5) = (0.1, 0.2, 0.2, 0.4, 0.4),$$
$$d_1 = (4, 4, 4, 4), \qquad d_2 = (1, 1, 1, 1),$$
$$d_3 = (8, 8, 8, 8), \qquad d_4 = (6, 6, 6, 6), \qquad d_5 = (3, 7, 3, 7).$$

The Shekel function has four local minima and one global minimum $x_{\min} = (4, 4, 4, 4)$; the minimum value is $\Phi(x_{\min}) \approx -10.15$. This function is very flat, except in the neighborhoods of the points d_1, d_2, d_3, d_4, d_5 where it

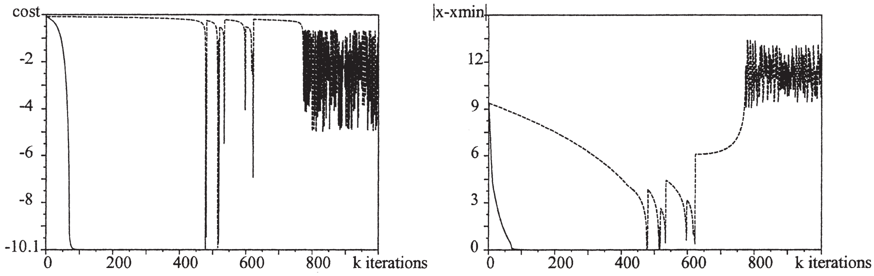

Fig. 5. Minimization of the Shekel function with the (DSD) algorithm (dashed line) and the (DSDC) algorithm (continuous line). Starting point $x_0 = (5, 0, 5, 0)$. Left: Evolution of the cost function during the optimization process. Right: Convergence history for $|x - x_{\min}|$.

is very steep. We have tested the (DSD) and (DSDC) methods and we have represented on Fig. 5 the evolution of the function values (left) and $|x - x_{\min}|$ (right). As in the previous example, (DSD) does not converge, while (DSDC) converges very fast.

References

1. Alvarez, F., and Perez, J. M., *A Dynamical System Associated with Newton's Method for Parametric Approximations of Convex Minimization Problems*, Applied Mathematics and Optimization, Vol. 38, pp. 193–217, 1998.
2. Attouch, H., and Cominetti, R., *A Dynamical Approach to Convex Minimization Coupling Approximation with the Steepest Descent Method*, Journal of Differential Equations, Vol. 128, pp. 519–540, 1996.
3. Cominetti, R., and Courdurier, M., *Coupling General Penalty Schemes for Convex Programming with the Steepest Descent and the Proximal Point Algorithm*, SIAM Journal on Optimization, Vol. 13, pp. 745–765, 2002.
4. Hairer, E., *Solving Ordinary Differential Equations, II: Stiff and Differential-Algebraic Problems*, Springer Series in Computational Mathematics, Springer Verlag, Berlin, Germany, Vol. 14, 1996.
5. Miranker, W. L., *Numerical Methods for Stiff Equations and Singular Perturbation Problems*, Reidel, New York, NY, 1980.
6. Bertsekas, D. P., *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.
7. Dennis, J. E., and Schnabel, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
8. Nocedal, J., and Wright, S. J., *Numerical Optimization*, Springer Series in Operations Research, Springer Verlag, New York, NY, 1999.

9. Opial, Z., *Weak Convergence of the Sequence of Successive Approximations for Nonexpansive Mappings*, Bulletin of the American Mathematical Society, Vol. 73, pp. 591–597, 1967.
10. Bruck, R. E., *Asymptotic Convergence of Nonlinear Contraction Semigroups in Hilbert Space*, Journal of Functional Analysis, Vol. 18, pp. 15–26, 1975.
11. Baillon, J. B., *Un Exemple Concernant le Comportement Asymptotique de la Solution du Problème* $du/dt + \partial\Phi(u) = 0$, Journal of Functional Analysis, Vol. 28, pp. 369–376, 1978.
12. Brezis, H., *Opérateurs Maximaux Monotones*, Mathematical Studies, North-Holland, Amsterdam, Holland, Vol. 5, 1973.
13. Cominetti, R., and Alemany, O., *Steepest Descent Evolution Equations: Asymptotic Behavior of Solutions and Rate of Convergence*, Transactions of the American Mathematical Society, Vol. 341, pp. 4847–4860, 1999.
14. Tikhonov, A., and Arsenine, V., *Méthodes de Résolution de Problèmes Mal Posés*, MIR, Moscow, Russia, 1974.
15. Attouch, H., *Viscosity Solutions of Minimization Problems*, SIAM Journal on Optimization, Vol. 6, pp. 769–806, 1996.
16. Cominetti, R., and San Martin, J., *Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming*, Mathematical Programming, Vol. 67, pp. 169–187, 1994.
17. Walster, G. W., Hansen, E. R., and Sengupta, S., *Test Results for a Global Optimization Algorithm*, Numerical Optimization, SIAM, Philadelphia, Pennsylvania, pp. 272–287, 1984.
18. Pinter, J., *Extended Univariate Algorithms for n-Dimensional Global Optimization*, Computing, Vol. 36, pp. 91–103, 1986.