Journal of Convex Analysis
Volume 2 (1995), No.1/2, 145–152

Asymptotic Convergence of the Steepest Descent Method for the Exponential Penalty in Linear Programming

R. Cominetti[1]
Universidad de Chile,
Casilla 170/3 Correo 3, Santiago, Chile.

Received 30 May 1994
Revised manuscript received 6 February 1995

Dedicated to R. T. Rockafellar on his 60th Birthday

We study the asymptotic behavior of the integral curves of the differential equation

$$\dot{u}(t) = -\nabla_x f(u(t),\ r(t)),\ u(t_0) = u_0$$

where $f(x,r)$ is the exponential penalty function associated with the linear program $\min\{c'x : Ax \leq b\}$, and $r(t)$ decreases to 0 as t goes to ∞. We show that for each initial condition (t_0, u_0) the solution $u(t)$ is defined on the whole interval $[t_0, \infty)$ and, under suitable hypothesis on the rate of decrease of $r(t)$, we establish the convergence of $u(t)$ towards an optimal solution u_∞ of the linear program. In particular we find sufficient conditions for u_∞ to coincide with the limit of the unique minimizer $x(r)$ of $f(\cdot, r)$.

Keywords : linear programming, exponential penalty, steepest descent, asymptotic behavior, nonlinear evolution equations.

1991 Mathematics Subject Classification: 90C05, 90C25, 90C31, 49M30.

1. Introduction

In this paper we study the asymptotic behavior of a non-autonomous nonlinear evolution equation of the form

$$(E) \qquad \begin{cases} \dot{u}(t) = -\nabla_x f(u(t), r(t)) \\ u(t_0) = u_0 \end{cases}$$

where $f(x,r)$ is convex and differentiable with respect to the x variable for every fixed $r > 0$, and $r(t)$ is a positive real differentiable function decreasing to 0 as $t \to \infty$.
Our motivation for studying such type of evolution equations comes from the coupling of the steepest descent method with different approximation and/or regularization schemes for optimization problems. Namely, we interpret the function $f(\cdot, r)$ as a smooth convex

[1] Partially supported by Fondecyt 93/1100.

ISSN 0944-6532 / \$ 2.50

approximation of a certain lower semicontinuous convex function f_0 which we are willing to minimize. The approximation $f(\cdot, r)$ is assumed to converge to f_0 when $r \downarrow 0$, so that the evolution equation (E) corresponds to applying the steepest descent method to the approximating functions while simultaneously controlling the degree of approximation by means of a parameter function $r(t)$.

More precisely, in the present work we consider the specific case of the exponential penalty function

$$f(x, r) := c'x + r \sum_{i=1}^{m} \exp[(A_i x - b_i)/r]$$

which is a smooth approximation for the function

$$f_0(x) = \begin{cases} c'x & \text{if } Ax \leq b \\ +\infty & \text{otherwise} \end{cases}$$

associated with a linear program of the form

$$(P) \quad \min_x \{c'x : Ax \leq b\}.$$

Our goal is to find conditions on the function $r(t)$ to ensure that the integral curves of (E) converge as $t \to \infty$ towards an optimal solution of (P).

Evolution equations, specially in infinite dimensional spaces, have received a great deal of attention since many physical phenomena fall into such framework (see *e.g.* [2], [4], [7], [8]). For autonomous subdifferential inclusions of the form

$$\text{(A)} \quad \begin{cases} \dot{u}(t) \in -\partial f_0(u(t)) \\ u(t_0) = u_0 \end{cases}$$

the asymptotic convergence of its integral curves towards a minimizer of f_0 was established by Brézis [2] and Brück [3]. The asymptotic convergence for so-called *semi-autonomous* differential inclusions may also be found in [2], [7]. Fully non-autonomous evolution equations have been considered recently by Furuya, Miyashiba and Kenmochi [6]. Roughly speaking, they consider the case in which the approximating functions $f(\cdot, r(t))$ approach the limit function f_0 sufficiently fast, so that the asymptotic behavior of the non-autonomous system (E) is close to the one of the autonomous system (A). This approach leads to results requiring a fast convergence of $r(t)$ towards 0.

In this work we consider the specific setting of exponential penalties in linear programming [5], and we present two results on the asymptotic convergence of (E), which go in some sense in opposite directions in terms of assumptions. The first one is concerned with the case of a function $r(t)$ converging to zero sufficiently *fast*, being close in spirit to the results in [6]. Nevertheless, the conditions we obtain are more explicit and seem to be weaker than the ones imposed by the abstract results in [6]. The second result considers the opposite situation: we show that asymptotic convergence will also be attained if the function $r(t)$ goes to zero sufficiently *slow*. An advantage of this second approach is that the limit point of the integral curve may be identified and characterized in variational terms.

The paper is organized as follows. In section 2 we review the basic results concerning the exponential penalty trajectory in linear programming needed for the asymptotic convergence analysis of (E). The existence of global solutions for (E) as well as the asymptotic convergence towards an optimal solution of (P) are then presented in section 3.

2. Preliminaries

Let us consider the linear program

$$(P) \quad \min_x \{c'x : Ax \leq b\}$$

where $c \in \mathbb{R}^n$, A is an $m \times n$ matrix of full rank $n \leq m$, and $b \in \mathbb{R}^m$. We assume throughout that the optimal solution set of (P), which we denote by S, is *nonempty and bounded*. Equivalently, this amounts to suppose that the dual problem

$$(D) \quad \min_\lambda \{b'\lambda : A'\lambda + c = 0,\ \lambda \geq 0\},$$

has a strictly positive feasible solution.
Following [5], for each $r > 0$ we consider the unconstrained penalized problem

$$(P_r) \quad \min_x \ c'x + r \sum_{i \in I} \exp[(A_i x - b_i)/r]$$

where A_i denote the rows of A for $i \in I := \{1, \ldots, m\}$. The dual problem of (P_r) is

$$(D_r) \quad \min_\lambda \left\{ b'\lambda + r \sum_{i \in I} \lambda_i(\ln \lambda_i - 1) : A'\lambda + c = 0,\ \lambda \geq 0 \right\},$$

which can be interpreted as a penalty method which introduces the positivity constraints of (D) into the objective function through the barrier term "$r \sum \lambda_i(\ln \lambda_i - 1)$".
From [5, Prop. 2.1] we know that both problems (P_r) and (D_r) have unique optimal solutions $x(r)$ and $\lambda(r)$ respectively, which are related as

$$\lambda_i(r) = \exp[(A_i x(r) - b_i)/r],$$

and are of class $\mathcal{C}^\infty$ on the open interval $(0, \infty)$.
Moreover, according to [5, Thm. 5.8], when r tends to 0 the primal trajectory $x(r)$ approaches (quite rapidly indeed) a straight line directed towards a center of the optimal face of (P). More precisely we have the asymptotic expansion

$$x(r) = x^* + rd^* + \eta(r)$$

where x^* is an optimal solution of (P) which is called *centroid*, the directional derivative d^* is completely characterized in variational terms, and the error $\eta(r)$ goes to zero exponentially fast, that is, at the speed of $\exp(-\mu/r)$ for some $\mu > 0$.
For the dual trajectory it turns out [5, Thm. 5.8] that $\lambda(r)$ behaves asymptotically as a constant, that is

$$\lambda(r) = \lambda^* + \nu(r),$$

Figure 1. An example of primal (a) and dual (b) trajectories.

where λ^* is a center of the optimal face of (D) and the error $\nu(r)$ goes to zero exponentially fast as r goes to 0.

We recall a couple of additional properties which we shall use in section 3 (these may also be found in [5]). Firstly we remark that the centroid x^* is a (relative) interior point of the optimal face S. Namely, denoting

$$I_0 = \{i \in I : A_i x = b_i \text{ for all } x \in S\}$$

the set of constraints which are binding at every optimal solution of (P), we have

$$A_i x^* < b_i \quad \text{for all } i \notin I_0. \tag{2.1}$$

Secondly, the derivative of $x(r)$ stays bounded as r goes to zero. In fact we have

$$\frac{dx}{dr}(r) = d^* + \rho(r)$$

with $\rho(r)$ converging exponentially fast towards 0.

3. The steepest descent method

In the sequel we denote by $f(x, r)$ the *exponential penalty function*

$$f(x, r) = c'x + r \sum_{i \in I} \exp[(A_i x - b_i)/r].$$

In [5] it was suggested that the asymptotic *straight line* character of the primal trajectory $x(r)$, should make it easy for a predictor-corrector method to trace this path approximately. However, instead of being forced to follow and stay close to a single trajectory, the task would be alleviated if we had a flow on $\mathbb{R}^n$ whose integral curves converged towards the optimal set S. This leads us to consider the continuous steepest descent equation

$$(E) \quad \begin{cases} \dot{u}(t) = -\nabla_x f(u(t), r(t)) \\ u(t_0) = u_0 \end{cases}$$

where $r : [t_0, \infty) \longrightarrow \mathbb{R}_+$ is a given continuously differentiable decreasing function such that

$$\lim_{t\to\infty} r(t) = 0.$$

Theorem 3.1. *For each initial condition u_0, problem (E) has a unique solution $u(t)$ which is defined on all of $[t_0, \infty)$ and stays bounded as $t \to \infty$.*

Proof. Since the mapping $(x,t) \longrightarrow \nabla_x f(x, r(t))$ is locally lipschitz we have local existence and uniqueness for (E), so that all we need is to establish an *a priori* bound for $u(t)$. To this end let us consider the function

$$\varphi(t) = \tfrac{1}{2}\|u(t) - x(r(t))\|^2.$$

The optimality of $x(r)$, the convexity of $f(\cdot, r)$ and equation (E) imply that

$$0 \;\le\; f(u(t), r(t)) - f(x(r(t)), r(t)) \;\le\; \langle -\dot{u}(t), u(t) - x(r(t))\rangle. \tag{3.1}$$

Denoting $\Delta(t) := \langle -\dot{u}(t), u(t) - x(r(t))\rangle$, taking a constant M such that

$$\|\frac{dx}{dr}(r)\| \le \sqrt{2}M \quad \text{for all } r < r(t_0),$$

and observing that $\dot{r}(t) \le 0$, we obtain

$$0 \le \Delta(t) = -\dot{\varphi}(t) - \dot{r}(t)\langle \frac{dx}{dr}(r(t)), u(t) - x(r(t))\rangle \le -\dot{\varphi}(t) - 2M\sqrt{\varphi(t)}\dot{r}(t). \tag{3.2}$$

It follows that $\dot{\varphi}(t) \le -2M\sqrt{\varphi(t)}\dot{r}(t)$ from which one easily gets

$$\sqrt{\varphi(t)} \le \sqrt{\varphi(t_0)} + M[r(t_0) - r(t)] \le \sqrt{\varphi(t_0)} + Mr(t_0). \tag{3.3}$$

Since $x(r(t))$ stays bounded as $t \to \infty$, we deduce that $u(t)$ stays bounded as well. □

We intend to prove that the solution $u(t)$ of (E) converges, when t goes to infinity, towards an optimal solution u_∞ of the linear program (P). To this end we need some assumptions on the rate of decrease of $r(t)$ towards 0. We shall distinguish two cases:

$$\text{Fast Decay:} \quad \int_{t_0}^{\infty} r(t)\exp(-\alpha/r(t))dt < \infty \text{ for all } \alpha > 0,$$

$$\text{Slow Decay:} \quad \int_{t_0}^{\infty} \frac{1}{r(t)}\exp(-K/r(t))dt = \infty \text{ for all } K > 0.$$

Fast Decay holds for instance when $r(t) = 1/t^a$ with $a > 0$, while Slow Decay holds in the case $r(t) = 1/\ln(\ln t)$. For $r(t) = 1/\ln t$ neither Fast nor Slow Decay is satisfied.

Theorem 3.2. *If $r(t)$ satisfies the assumption of Fast Decay, then the solution $u(t)$ of (E) converges towards a particular solution $u_\infty \in S$.*

Proof. We split the proof into 4 steps.

Step 1: *Existence of* $\lim \|u(t) - x(r(t))\| = \lim \|u(t) - x^*\|$.
Letting $K = 2M[\sqrt{\varphi(t_0)} + Mr(t_0)]$ and using (3.2) and (3.3) we have $\dot{\varphi}(t) + K\dot{r}(t) \leq 0$. Then the function $\varphi(t) + Kr(t)$ is positive and decreasing so that it has a finite limit when $t \to \infty$. The convergence of $\varphi(t)$ follows at once and, since $x(r(t))$ converges to x^*, we conclude the existence of

$$\lim_{t\to\infty} \|u(t) - x(r(t))\| = \lim_{t\to\infty} \|u(t) - x^*\|.$$

Step 2: $\Delta(t)$ *belongs to* $L^1(t_0, \infty)$.
This follows from (3.2) taking into account that $2M\sqrt{\varphi(t)} \leq K$ and

$$\int_{t_0}^{\infty} [-\dot{\varphi}(\tau) - K\dot{r}(\tau)]d\tau \leq \varphi(t_0) + Kr(t_0).$$

Step 3: *Existence of* $\lim \|u(t) - \bar{u}\|$ *for all* $\bar{u} \in S$.
Let $\bar{u} \in S$. Since

$$\|u(t) - \bar{u}\|^2 = \|u(t) - x^*\|^2 + \|x^* - \bar{u}\|^2 - 2\langle x^*, x^* - \bar{u}\rangle + 2\langle u(t), x^* - \bar{u}\rangle,$$

the existence of $\lim \|u(t) - \bar{u}\|$ is equivalent to the existence of

$$\lim_{t\to\infty} \langle u(t), x^* - \bar{u}\rangle.$$

To prove this fact we shall establish that $\theta(t) := \langle \dot{u}(t), x^* - \bar{u}\rangle$ belongs to $L^1(t_0, \infty)$. Indeed, for $\epsilon \in \mathbb{R}$ let us set $x_\epsilon(t) := x(r(t)) + \epsilon(\bar{u} - x^*)$. The convexity of $f(\cdot, r)$ and (E) imply that

$$\langle -\dot{u}(t), x_\epsilon(t) - u(t)\rangle \leq f(x_\epsilon(t), r(t)) - f(u(t), r(t)),$$

so that using the optimality of $x(r(t))$ and the definitions of $\theta(t)$ and $\Delta(t)$ we deduce

$$\epsilon\theta(t) = \epsilon\langle \dot{u}(t), x^* - \bar{u}\rangle \leq \Delta(t) + f(x_\epsilon(t), r(t)) - f(x(r(t)), r(t)).$$

Now, since x^* and $\bar{u}$ are optimal solutions for (P) we have

$$\begin{aligned} c'x(r(t)) &= c'x_\epsilon(t) \\ A_i x(r(t)) &= A_i x_\epsilon(t), \quad i \in I_0 \end{aligned}$$

and using the expression of $f(\cdot, r)$, we obtain the bound

$$\epsilon\theta(t) \leq \Delta(t) + r(t) \sum_{i \notin I_0} \exp[(A_i x_\epsilon(t) - b_i)/r(t)].$$

We may then use (2.1) to find $\alpha > 0$ such that for all ϵ close enough to 0 (either positive or negative) and all t sufficiently large

$$A_i x_\epsilon(t) - b_i \leq -\alpha.$$

Then

$$\epsilon\theta(t) \leq \Delta(t) + mr(t)e^{-\alpha/r(t)}$$

so that the assumption of Fast Decay together with Step 2 imply that $\theta(t)$ belongs to $L^1(t_0, \infty)$ as claimed.

Step 4: *Conclusion.*

By Step 2, the function $\Delta(t)$ belongs to $L^1(t_0, \infty)$ so we may find a sequence $t_n \to \infty$ such that $\Delta(t_n) \to 0$. By extracting a subsequence we may further assume that $u(t_n)$ converges towards a certain u_∞. Using (3.1) and the results quoted from [5] in section 2, one may easily deduce that u_∞ is a solution of (P). But then, using Step 3 we have the existence of

$$\lim_{t\to\infty} \|u(t) - u_\infty\| = \lim_{n\to\infty} \|u(t_n) - u_\infty\| = 0,$$

achieving the proof of our theorem. □

Theorem 3.3. *If $r(t)$ satisfies the assumption of Slow Decay, then the solution $u(t)$ of (E) converges to the centroid x^* as t goes to infinity.*

Proof. Since $x(r)$ tends to x^* as $r \to 0$, it suffices to show that $\varphi(t)$ tends to 0 as t goes to ∞. By contradiction assume that the limit of $\varphi(t)$ is not 0 (see Step 1 in the previous proof).

Using the expression of $f(\cdot, r)$ it is easy to see that for every $R > 0$, there exist constants K and L such that for all $x, y \in B(0, R)$ and all $r > 0$ one has

$$\langle \nabla f(y,r) - \nabla f(x,r), y - x\rangle \ \geq \ \frac{L}{r}\exp(-K/r)\|y - x\|^2.$$

Since $u(t)$ and $x(r(t))$ stay bounded, we may find appropiate constants so that the previous inequality holds with $x = x(r(t))$, $y = u(t)$ and $r = r(t)$, to deduce

$$\Delta(t) \ \geq \ \frac{2L}{r(t)}\exp(-K/r(t))\varphi(t).$$

Denoting $\beta(t) = \frac{L}{r(t)}\exp(-K/r(t))$ we obtain

$$\begin{aligned} \dot\varphi(t) \ &= \ -\Delta(t) - \dot r(t)\langle \frac{dx}{dr}(r(t)), u(t) - x(r(t))\rangle \\ &\leq \ -2\beta(t)\varphi(t) - 2M\dot r(t)\sqrt{\varphi(t)}. \end{aligned}$$

Since $\varphi(t) \neq 0$ for t large, we may divide by $2\sqrt{\varphi(t)}$ to deduce

$$\frac{1}{2\sqrt{\varphi(t)}}\dot\varphi(t) + \beta(t)\sqrt{\varphi(t)} \ \leq \ -M\dot r(t).$$

Let $B(t) = \int_{t_0}^t \beta(\tau)d\tau$. Multiplying this inequality by $\exp(B(t))$ we get

$$\frac{d}{dt}[\exp(B(t))\sqrt{\varphi(t)}] \ \leq \ -M\exp(B(t))\dot r(t)$$

which integrated between s and t gives

$$\sqrt{\varphi(t)} \leq \exp(B(s) - B(t))\sqrt{\varphi(s)} - M \int_s^t \exp(B(\tau) - B(t))\dot{r}(\tau)d\tau.$$

We recall that $\dot{r}(t) \leq 0$, so that using the fact that $B(\cdot)$ is increasing we obtain the bound

$$\begin{aligned} \sqrt{\varphi(t)} &\leq \exp(B(s) - B(t))\sqrt{\varphi(s)} - M \int_s^t \dot{r}(\tau)d\tau \\ &= \exp(B(s) - B(t))\sqrt{\varphi(s)} + M[r(s) - r(t)]. \end{aligned}$$

The assumption of Slow Decay implies that $B(t)$ tends to infinity as $t \to \infty$. Thus, passing to the limit we deduce

$$\lim_{t\to\infty} \sqrt{\varphi(t)} \leq Mr(s).$$

It suffices now to let s approach infinity to conclude that $\varphi(t)$ tends to 0, contradicting our assumption and proving the result. □

Remark 3.4. The convergence analysis for the case of Slow Decay may be adapted to more general non-autonomous evolution equations. This will be the subject of the forthcoming paper [1].

References

[1] H. Attouch, R. Cominetti: A dynamical approach to convex minimization coupling approximation with the steepest descent method, submitted 1994.

[2] H. Brézis: Opérateurs maximaux monotones et semigroups de contractions dans les espaces de Hilbert, Mathematical Studies 5, North Holland 1973.

[3] R. Brück: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, Journal of Functional Analysis 18 1975, 15–26.

[4] T. Cazenave, A. Haraux: Introduction aux problèmes d'évolution semi-linéaires, Mathématiques & Applications 1, Ellipses 1993.

[5] R. Cominetti, J. San Martín: Asymptotic behavior of the exponential penalty trajectory in linear programming, Mathematical Programming 67 1994, pp. 169–187.

[6] H. Furuya, K. Miyashiba, N. Kenmochi: Asymptotic behavior of solutions to a class of nonlinear evolution equations, Journal of Differential Equations 62 1986, 73–94.

[7] A. Haraux: Nonlinear Evolution Equations: Global Behavior of Solutions, Lecture Notes in Mathematics 841, Springer-Verlag 1981.

[8] A. Pazy: Semigroups of linear operators and applications to partial differential equations, Springer-Verlag, New York 1983.