Singular Riemannian Barrier Methods and Gradient-Projection Dynamical Systems for Constrained Optimization.

H. ATTOUCH * J. BOLTE * P. REDONT * M. TEBOULLE †

Abstract. This work is devoted to the dynamical system (SRB): $\frac{d}{dt}\partial h(x(t)) + \nabla\Phi(x(t)) \ni 0$, with h a proper lower semicontinuous convex function. Existence and uniqueness of solutions are examined. Systems (SRB) include the class of gradient systems with respect to a Hessian Riemannian metric induced by a convex Legendre function h: $\dot{x}(t) + \nabla^2 h(x(t))^{-1}\nabla\Phi(x(t)) = 0$. Moreover, class (SRB) is closed in a variational sense: links are made with regularized Lotka-Volterra systems and the limit equations obtained by letting the barrier parameter go to 0. Of particular interest is the case $h(x) = \frac{1}{2}|x|^2 + \delta_C(x)$: this way, one obtains a new gradient-projection method. System (SRB) bears a direct relation with the minimization of Φ over the domain of ∂h; the asymptotic behaviour of solutions, as time goes to infinity, is a real issue, which is addressed for a convex Φ and a h of the form $h = k + \delta_C$, with k convex $\mathcal{C}^1$ and C a finite dimensional polyhedron.

Keywords: Dynamical systems, continuous gradient method, asymptotic analysis, barrier methods in constrained optimization, singular Hessian Riemannian metric, convex Legendre functions, Bregman distances.

AMS classification: 34A60, 37Bxx, 37Lxx, 37N40, 49K24.

1 Introduction

This paper proposes to introduce and study a new continuous dynamical system of a gradient-projection type

$$\text{(SRB)} \qquad \frac{d}{dt}\partial h(x(t)) + \nabla\Phi(x(t)) \ni 0$$

where $\partial h(x)$ is the subgradient set at point x of the proper lower semicontinuous convex function h. As it was primarily designed to deal with constraints by means of Riemannian metrics with singularities acting as barriers, it was called SRB for Singular Riemannian Barrier.

System (SRB) is one of the many descendants of the classical continuous steepest descent method (also known as the gradient method)

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

The steepest descent method possesses an extraordinarily wide range of applications extending from mechanics and partial differential equations to economics and optimization where it may be used to minimize the function Φ. But in order to deal with a constrained optimization problem

$$(\mathcal{P}) \qquad \min\{\Phi(x) : x \in C\}$$

the method needs some arrangement.

Penalization methods have long been providing with means for circumventing constraints, too often at the expense of numerical stiffness. Our approach here is in the line of the recent studies on smooth continuous interior descent methods which have been developed in order to study constrained optimization problems. The notion of variable metric lies at the root of these methods. The Euclidean metric is relinquished and replaced by a local Riemannian metric capable of dealing with constraints. A unifying approach indeed has progressively emerged, which is based on a general study of Riemannian steepest descent dynamics associated with Legendre functions: Attouch-Teboulle [3] (link with Lotka-Volterra equations), Bolte-Teboulle [6], Alvarez-Bolte-Brahic [1], Bayer-Lagarias[5], Iusem-Svaiter-Da Cruz [10]...

A general form of such systems is the following

*ACSIOM-CNRS UMR 5149, Département de Mathématiques, case 51, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France.

†School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv 69978, Israel.

$$(\mathrm{D}) \qquad \dot{x}(t)+\nabla^2 h(x(t))^{-1}\nabla\Phi(x(t))=0$$

where $\nabla^2 h(x)^{-1}$ stands for the inverse of the Hessian mapping, $\nabla^2 h(x)$, of a Legendre function h associated with the constraint C. Since $\nabla^2 h(x)^{-1}\nabla\Phi(x)$ is the gradient of Φ calculated with respect to the Riemannian metric given at point x by the scalar product $\langle\nabla^2 h(x).,.\rangle$, system (D) is no more than the steepest descent method expressed in that metric.

In the absence of constraints a celebrated example of such a system is the continuous Newton method where h coincides with Φ

$$\dot{x}(t)+\nabla^2\Phi(x(t))^{-1}\nabla\Phi(x(t))=0.$$

But more typically, h is a smooth function on the interior of C, h is convex with $\nabla^2 h(x)$ symmetric positive definite and ∇h blows up on the boundary of C, which means that $|\nabla h(x^k)|\to+\infty$ for every sequence $(x^k)\subset \operatorname{int} C$ converging to a boundary point of C as $k\to+\infty$. It may be advantageous, and equivalent indeed, to write the system (D) as follows

$$\frac{d}{dt}\nabla h(x(t))+\nabla\Phi(x(t))=0.$$

Usually such a h is obtained by using a barrier term (like the log barrier term) which makes a barrier parameter $\mu>0$ appear in the dynamics. So doing, for each μ, one can exhibit a dynamical system

$$(D_\mu) \qquad \dot{x}(t)+\nabla^2 h_\mu(x(t))^{-1}\nabla\Phi(x(t))=0$$

which, for given Cauchy data, generates a family (x_μ) of smooth continuous interior descent trajectories.

The questions which are addressed here are:

- Which is the dynamical system obtained as a limit when $\mu\to 0$?
- What properties does it enjoy?

The regularized Lotka-Volterra dynamical system [3] will serve as a model example and guideline to help answer these questions. With x subject to the constraint $x\in\mathbb{R}^p_{++}=\{x\in\mathbb{R}^p, x_j>0\}$, the system reads

$$(\mathrm{RLV}_\mu) \qquad \frac{d}{dt}x_{\mu,j}(t)+\frac{x_{\mu,j}(t)}{\mu+\nu x_{\mu,j}(t)}\frac{\partial\Phi}{\partial x_j}(x_\mu(t))=0, \qquad j=1\dots p.$$

The Riemannian metric naturally attached to this system is defined by $h_\mu(x)=\frac{\nu}{2}|x|^2+\mu\sum_{j=1}^p x_j\ln x_j$ which makes sense on $\mathbb{R}^p_{++}$; it is a compromise between the Euclidean metric, dominant far from the boundary of the constraints, and the *log-barrier* metric, effective near the boundary. The (RLV_μ) system is of the form (D_μ). Letting the barrier parameter μ go to zero preserves the viability of the constraint (limit trajectories starting from $x_0\in\operatorname{int}C$ remain in C, but no more necessarily in $\operatorname{int}C$, they may hit the boundary after a finite time interval). The descent property is also preserved. This strongly suggests that the limit dynamics is of gradient-projection type.

Indeed, and this is the main result of the paper, so doing, we obtain a new gradient-projection dynamical system which is of the following type (p_C stands for the projection on C)

$$\begin{cases} x(t)=p_C(y(t)) \\ \dot{y}(t)+\nabla\Phi(p_C(y(t))=0 \end{cases}$$

This system is a smooth differential equation with respect to y, which generates classical $\mathcal{C}^1$ trajectories $t\mapsto y(t)$. By projecting them onto C one obtains the trajectories $t\mapsto x(t)$ of the gradient-projection system under study.

This system can equivalently be described using the (SRB) differential inclusion formulation; it is enough to choose $h(x)=\frac{1}{2}|x|^2+\delta_C(x)$, and then $\partial h(x)=x+N_C(x)$, where δ_C is the indicator function of C and $N_C(x)$ is the normal cone to C at point x ($y\in\partial h(x)\Leftrightarrow x=p_C(y)$), see e.g., [13]. This formulation involves subdifferential blowing terms, which formally suggests a link with phase transition dynamical systems, see [11].

We stress the fact that this gradient-projection dynamical system is in general different from the classical gradient-projection system

$$\dot{x}(t)=p_{T_C(x(t))}[-\nabla\Phi(x(t))]$$

where $T_C(x)$ is the convex tangent cone to C at x. When Φ is convex and lower semicontinuous, the latter system can be given the more general form of a differential inclusion, the *generalized steepest descent method*, see [7]

$$(\mathrm{GSD}) \qquad \dot{x}(t)+\partial\varphi(x(t))\ni 0,$$

with $\varphi=\Phi+\delta_C$. Indeed a striking property is that (GSD) is a lazy system insofar as it selects the least-norm element $\partial\varphi^0(x)$ of the subgradient set $\partial\varphi(x)$ at point x; and one can prove: $\dot{x}(t^+)+\partial\varphi^0(x(t))=0$, for all $t\geq 0$.

Both systems (the classical gradient-projection system and the new one we are introducing) enjoy optimization properties, and offer different numerical features. Moreover they model different behaviours of economical or biological agents with respect to constraints.

2 The model example. The regularized Lotka-Volterra system.

The regularized Lotka-Volterra system (RLV$_\mu$) has been introduced in [3], where its well-posedness, viability and minimizing properties have been investigated (see also [1] where the authors provide with a general frame based on Riemannian geometry, Legendre functions and Bregman distances, which can advantageously be applied to studying (RLV$_\mu$)).

The system (RLV$_0$), simply obtained by setting $\mu = 0$ in (RLV$_\mu$), is the plain steepest descent, which does not respect the positiveness constraint. Yet if (x_μ) is ever to converge to some limit as $\mu \to 0$, the latter is expected to be nonnegative, a property that solutions of (RLV$_0$) do not automatically share. Actually the right limit system is the following (which could be dubbed Lotka-Volterra with Vanishing Barrier) with β defined below

$$\text{(LVVB)} \qquad \nu \frac{d}{dt} x_j(t) + \frac{\partial \Phi}{\partial x_j}(x(t)) + \frac{d}{dt}\beta(x_j(t)) \ni 0, \qquad j = 1 \ldots p.$$

As this is not an ordinary differential equation, the notion of solution has to be made explicit: a function x is a solution to (LVVB) if there exists a function η which, along with x, satisfies the following properties

1. $x, \eta \in W^{1,1}_{\text{loc}}([0, +\infty[; \mathbb{R}^p) \cap \mathcal{C}([0, +\infty[; \mathbb{R}^p)$,

2. $\eta_j(t) \in \beta(x_j(t))$, $\quad j = 1...p$, $\forall t \geq 0$,

3. $\nu \frac{d}{dt} x_j(t) + \frac{\partial \Phi}{\partial x_j}(x(t)) + \frac{d}{dt}\eta_j(t) = 0$, $\quad j = 1...p$, $\forall t \geq 0$ a. e.

Let us now make the hypotheses precise:

1. $\Phi : \mathbb{R}^p \mapsto \mathbb{R}$ a $\mathcal{C}^1$ function, bounded below on $\mathbb{R}^p_+$, with gradient $\nabla\Phi$ Lipschitz continuous and bounded on the bounded subsets of $\mathbb{R}^p$,

2. $\beta : \mathbb{R} \rightrightarrows \mathbb{R}$ the maximal monotone operator with graph $u \in \beta(x) \Leftrightarrow u \leq 0$, $x \geq 0$, $ux = 0$,

3. $\nu > 0$ a fixed parameter, and $\mu > 0$ a small parameter.

Theorem 2.1 *Under the hypotheses stated above, for any $x_0 \in \mathbb{R}^p_{++}$ there exists a unique solution of the dynamical system (LVVB) with Cauchy data $x_j(0) = x_{0j}$. Moreover, this solution is obtained as the limit when $\mu \to 0$ of the filtered sequence $(x_\mu)_{\mu>0}$, where x_μ is the solution of the regularized Lotka-Volterra dynamical system (RLV$_\mu$) with Cauchy data $x_{\mu,j}(0) = x_{0j}$. Any trajectory of (LVVB) satisfies the following properties*

1. *$x_j(t) \geq 0$ for all $t \geq 0$, all $j = 1 \ldots p$,*

2. *$\nu\dot{x}(t) + \nabla\Phi(x(t)) = 0$ whenever the trajectory is in $\mathbb{R}^p_{++}$,* i. e. *on $\mathbb{R}^p_{++}$ the dynamics is the steepest descent,*

3. *whenever $x_j(t) = 0$ for $t \in [t_0, t_1]$, with t_0 (resp. t_1) being the minimal (resp. maximal) time for which $x_j(t) = 0$,* i. e. *$x_j(t) > 0$ for $t > t_1$, t sufficiently close to t_1, $x_j(t) > 0$ for $t < t_0$, t sufficiently close to t_0, we have*

$$\int_{t_0}^{t_1} \frac{\partial \Phi}{\partial x_j}(x(s))ds = 0. \tag{1}$$

Except for its last conclusion, theorem 2.1 actually is a corollary of theorems 3.1 and 5.1 below; the proof of point **3.** is omitted for the sake of brevity.

Systems (RLV$_\mu$) and (LVVB) both are of (SRB) type. Indeed, if we set $h_\mu(x) = \frac{\nu}{2}|x|^2 + \mu \sum_{j=1}^p x_j \ln x_j$ and $h(x) = \frac{\nu}{2}|x|^2 + \delta_{\mathbb{R}^p_+}$, then (RLV$_\mu$) and (LVVB) may respectively be written

$$\frac{d}{dt}\nabla h_\mu(x(t)) + \nabla\Phi(x(t)) = 0,$$

$$\frac{d}{dt}\partial h(x(t)) + \nabla\Phi(x(t)) \ni 0.$$

While the former system is relevant to Riemannian steepest descent systems associated to Legendre functions and Bregman distances (see [1]), little is known about the latter where h may appear as a singular Riemannian metric. Theorem 2.1 proves the existence of a solution to (LVVB) by a limit process as $\mu \to 0$. In the next section the existence of a solution to (SRB) will be proved directly *via* a smart change of unknown function. Later (section 5) it will be shown that the class of (SRB) systems is closed with respect to the graph convergence $\partial h_\mu \to \partial h$.

Figure 1 illustrates theorem 2.1 for a quadratic $\Phi : \mathbb{R}^2 \mapsto \mathbb{R}$ with a minimum point M inside the constraint set $\mathbb{R}^2_{++}$; dotted lines are level curves of Φ. Starting from point O and ending at point M, four (RLV$_\mu$) trajectories are displayed with $\mu = 1, 10^{-1}, 10^{-2}, 10^{-4}$. The last one, plotted with larger line width, is indistinguishable indeed from the limit trajectory $OABM$ solution of (LVVB). This graph also examplifies the difference between the systems (LVVB) and (GSD), which reads here $\dot{x}(t) + \nabla\Phi(x(t)) + N_{\mathbb{R}^2_+}(x(t)) \ni 0$. The steepest descent trajectory $OACM$ first coincides with the (LVVB) trajectory inside the constraint set and next on that part of the boundary where $-\nabla\Phi$ points outwards. Afterwards it leaves the boundary smoothly (dashed line CM), while the (LVVB) trajectory remains farther on the boundary till point B, satisfying (1), where it leaves the boundary transversally. Thus the behaviour of an (LVVB) trajectory is much like the gradient-projection behaviour of a (GSD) trajectory, yet it is definitely different as the former remains longer on the boundary of the constraint set.

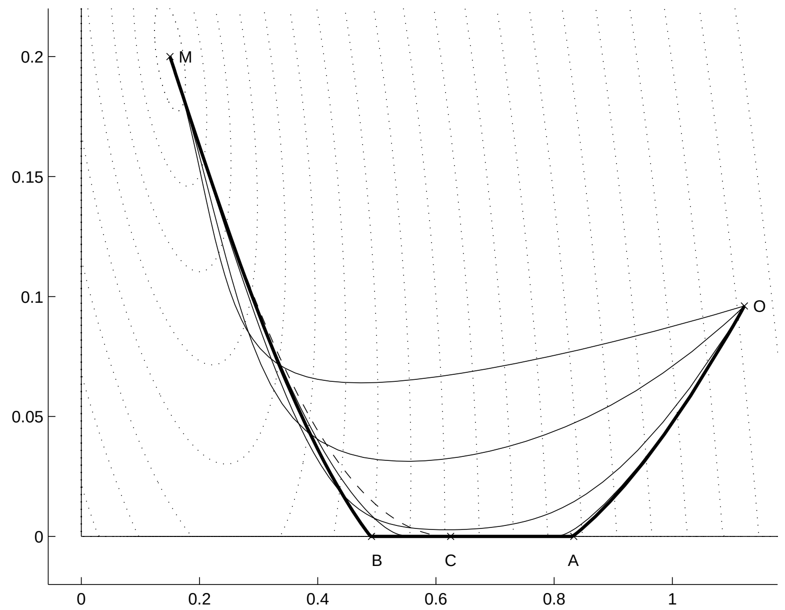

Figure 1: Convergence of the solutions of (LVB_μ) as $\mu \to 0$.

3 The main result. The Singular Riemannian Barrier dynamical system SRB.

In this section the existence of a solution to

$$\text{(SRB)} \qquad \frac{d}{dt}\partial h(x(t)) + \nabla\Phi(x(t)) \ni 0$$

will be proved under the following assumptions

H1. H a real Hilbert space,

H2. $\Phi : H \mapsto \mathbb{R}$ a bounded below $\mathcal{C}^1$ function, with $\nabla\Phi$ Lipschitz continuous on the bounded subsets of H,

H3. $h : H \mapsto \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ a strongly convex function: *i. e.* there exists $\alpha > 0$ such that

$$h(\lambda x + (1-\lambda)y) \le \lambda h(x) + (1-\lambda)h(y) - \frac{\alpha}{2}\lambda(1-\lambda)|y-x|^2$$

for all x, y in the domain of h and all $\lambda \in]0,1[$; or equivalently

$$(u,v) \in \partial h(x) \times \partial h(y) \ \Rightarrow\ \langle v-u, y-x\rangle \ \ge\ \alpha|y-x|^2.$$

Definition. A function $x : [t_0, +\infty[\mapsto H$ is termed a solution of (SRB) with initial data $x_0 \in \operatorname{dom}\partial h$ if and only if

1. $x \in W^{1,1}_{\text{loc}}([t_0,+\infty[; H) \cap \mathcal{C}([t_0,+\infty[; H)$,

2. $x(t_0) = x_0$,

3. there exists a function $u \in \mathcal{C}^1([t_0,+\infty[; H)$ such that

. $u(t) \in \partial h(x(t))$, $\forall t \in [t_0,+\infty[$,

. $\dot{u}(t) + \nabla\Phi(x(t)) = 0$, $\forall t \in [t_0,+\infty[$.

For x to be a solution of (SRB) it is required that $u(t) \in \partial h(x(t))$. The latter is equivalent to $x(t) \in \partial h^*(u(t))$, $h^* : H \mapsto \bar{\mathbb{R}}$ being the convex conjugate of h. It is important to realize that, in view of the strong convexity of h, h^* is everywhere defined on H; moreover h^* is differentiable and its gradient ∇h^* is Lipschitz continuous with $1/\alpha$ as Lipschitz modulus (see e.g., [7],[14, Proposition 12.60]). So x and u are related by $x(t) = \nabla h^*(u(t))$, and u has to satisfy the following auxiliary regular ordinary differential equation with some initial value u_0 arbitrary in $\partial h(x_0)$:

$$\text{(SRBaux)} \qquad \dot{u}(t) + \nabla\Phi(\nabla h^*(u(t))) = 0.$$

Problems (SRB) and (SRBaux) are related by the change of unknown function $x = \nabla h^*(u)$; using this idea can be traced back to [1, 5] and to general duality principles, see Attouch-Théra [4].

Theorem 3.1 *Under the assumptions stated above, and for any $x_0 \in$ dom ∂h*

- **a.** *x is a solution of (SRB) with Cauchy data $x(t_0) = x_0$ if and only if there exist $u_0 \in \partial h(x_0)$ and a solution $u : [t_0, +\infty[\mapsto H$ of (SRBaux) such that $u(t_0) = u_0$ and $x(t) = \nabla h^*(u(t))$ for all $t \in [t_0, +\infty[$;*
- **b.** *(SRB) with initial data $x(t_0) = x_0$ possesses one solution at least;*
- **c.** *if $\partial h(x_0)$ is a singleton, (SRB) with initial data $x(t_0) = x_0$ has exactly one solution.*

Proof. Since ∇h^* is Lipschitz continuous, point **a** is a mere rephrasing of x being a solution of (SRB).

Taking the existence for granted, the uniqueness asserted by **c** follows from the well-posedness of the Cauchy-Lipschitz problem (SRBaux).

There remains to show the existence of a solution $x : [t_0, +\infty[\mapsto H$ to (SRB).

Since ∇h^* is Lipschitz continous, the Cauchy problem (SRBaux) admits a maximal local solution $u : [t_0, t_{\max}[\mapsto H$ which is continuous together with its derivative $\dot{u}$. To show that it is global indeed, it is enough to show that it is bounded if $t_{\max}$ is supposed finite.

Define $x : [t_0, t_{\max}[\mapsto H$ by $x(t) = \nabla h^*(u(t))$. Since ∇h^* is Lipschitz continuous and $u \in \mathcal{C}^1([t_0, t_{\max}[, H)$, x is locally absolutely continuous and hence differentiable almost everywhere in $[t_0, t_{\max}[$; hence

$$\langle \dot{u}(t), \dot{x}(t)\rangle + \langle \nabla\Phi(x(t)), \dot{x}(t)\rangle = 0, \quad \text{a.e. in } [t_0, t_{\max}[$$

At every point $t \in [t_0, t_{\max}[$ where x is differentiable we have

$$\begin{aligned}
\langle \dot{u}(t), \dot{x}(t)\rangle &= \lim_{h\to 0} \langle \frac{u(t+h) - u(t)}{h}, \frac{x(t+h) - x(t)}{h}\rangle \\
&\geq \alpha \lim_{h\to 0} \left| \frac{x(t+h) - x(t)}{h} \right|^2, \text{ after hypothesis } \mathbf{H3} \\
&\geq \alpha |\dot{x}(t)|^2.
\end{aligned}$$

Hence

$$\alpha|\dot{x}(t)|^2 + \langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \leq 0 \quad \text{a.e. in } [t_0, t_{\max}[. \tag{2}$$

But Φ is $\mathcal{C}^1$, and x and $\Phi \circ x$ are locally Lipschitz continuous, hence almost everywhere differentiable; so we have

$$\frac{d}{dt}\Phi(x(t)) = \langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \text{ a.e. in } [t_0, t_{\max}[. \tag{3}$$

Since $\Phi \circ x$ is locally Lipschitz continuous, integrating (2) yields

$$\alpha \int_{t_0}^{t} |\dot{x}(\tau)|^2 d\tau + \Phi(x(t)) - \Phi(x_0) \leq 0, \quad \forall t \in [t_0, t_{\max}[$$

$$\alpha \int_{t_0}^{t} |\dot{x}(\tau)|^2 d\tau \leq \Phi(x_0) - \inf_H \Phi, \quad \forall t \in [t_0, t_{\max}[.$$

If $t_{\max}$ were finite then x would be bounded on $[t_0, t_{\max}[$ in view of

$$|x(t) - x(t_0)| \leq \int_{t_0}^{t} |\dot{x}(\tau)| d\tau \leq \sqrt{t_{\max} - t_0}\sqrt{\int_{t_0}^{t_{\max}} |\dot{x}(\tau)|^2 d\tau}.$$

Hence $\dot{u} = -\nabla\Phi(x)$ would be bounded too; and so would u, which completes the proof. $\square$

Although (SRB) systems are primarily designed to deal with constraints, as is examplified by systems (RLV$_\mu$), (LVVB) and figure 1, they nevertheless allow singularities for the function h inside its domain of definition. This is illustrated by figure 2 for a quadratic $\Phi : \mathbb{R}^2 \mapsto \mathbb{R}$ with minimum point M; dotted lines are level curves of Φ. The function h defining the singular Riemannian metric is the sum of a Euclidean term plus a term with two singularities concentrated on the horizontal x_1-coordinate axis and on the vertical x_2-coordinate axis.; namely $h(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) + |x_1| + 2|x_2|$. Two x trajectories OM and $O'M$ are displayed, showing the delay effect of the singularities on the axes. While a constraint may be compared to an infinite height wall, the singular terms $|x_1|$ and $|x_2|$ may be thought of as two walls of finite height (respectively 1 and 2) that the x trajectories have to climb up and down before proceeding according to the steepest descent rule. Dashed lines are the u trajectories; they are plotted on the same graph as the x trajectories, but actually they lie in the space dual to that of x, and they do not bear a straight relation with the background of the graph. Their $\mathcal{C}^1$ regularity is noteworthy in contrast to the $\mathcal{C}$ regularity of x. From a numerical point of view, the solutions u are computed first, and next the solutions x via $x = \nabla h^*(u)$.

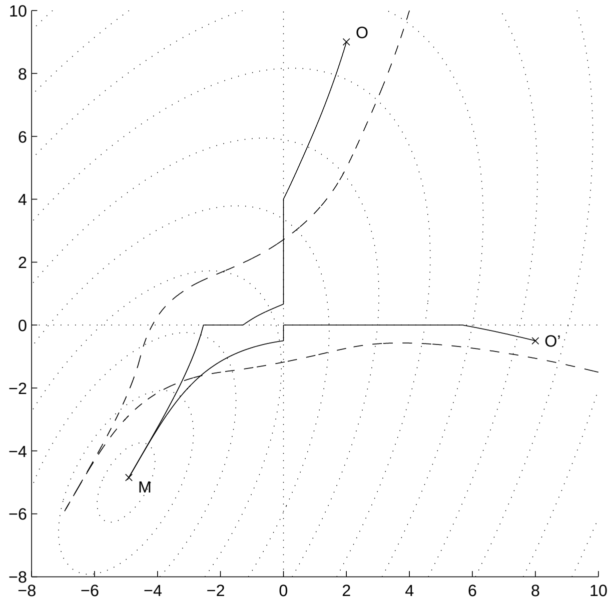

Figure 2: Effects of inner singularities of h.

4 Minimizing properties of the (SRB) system.

In the finite dimensional case, and for a convex Φ, system (SRB) enjoys nice optimization properties. Indeed it is proved in [1] that $\Phi(x(t))$ converges to the infimum value of Φ on dom h as $t \to +\infty$; and further if h is $\mathcal{C}^2$ on its domain and is both of Legendre and Bregman type, then $x(t)$ converges to a minimum point of Φ on dom h (provided there do exist minimum points). The analysis heavily relies on the properties of the so-called Bregman distance $(a, b) \mapsto h(b) - h(a) - \langle \nabla h(a), b - a \rangle$. But for a general convex lower semicontinuous h, the Bregman distance, as it stands, does not even make sense. Yet enforcing the very ideas of [1] allows to prove the convergence of the trajectories of (SRB), at least in the particular case where the set of constraints is convex polyhedral.

Fix x, u two solutions of (SRB) and (SRBaux) related by $u(t) \in \partial h(x(t))$. For $a \in$ dom h, and for $t \geq t_0$ define $D(a, t) = h(a) - h(x(t)) - \langle u(t), a - x(t) \rangle$. The dependence of D on the pair (x, u) is implicit. Observe that D is nonnegative since h is convex and $u(t) \in \partial h(x(t))$. Obviously $D(a, t)$ is meant to act as a Bregman distance between points a and $x(t)$.

The next two lemmas, which establish an infinitesimal property of D, prepare the theorem on the convergence of $\Phi(x(t))$.

Lemma 4.1 *([9]) Let $I \subseteq \mathbb{R}$ be an open interval; let $f : I \mapsto \mathbb{R}$ be such that* $\limsup_{\tau \to 0, \tau \neq 0} \frac{1}{\tau}\{f(t + \tau) - f(t)\} \leq 0$. *Then f is nonincreasing on I.*

Proof. Fix $\lambda > 0$, set $f_\lambda(t) = f(t) - \lambda t$. Then for all $t \in I$ we have $\limsup_{\tau \to 0, \tau \neq 0} \frac{1}{\tau}\{f_\lambda(t + \tau) - f_\lambda(t)\} \leq -\lambda < 0$. Hence for all $t \in I$ there exists $\varepsilon > 0$ such that

$$t \leq s < t + \varepsilon \Rightarrow f_\lambda(s) - f_\lambda(t) < 0$$

$$t - \varepsilon < r \leq t \Rightarrow f_\lambda(r) - f_\lambda(t) > 0.$$

Fix a, b in I with $a < b$. There exists a finite cover of $[a, b]$ by intervals $(]t_i - \varepsilon_i/2, t_i + \varepsilon_i/2[)_{i=1\dots n}$, where the t_i's and ε_i's satisfy the relations above. It is harmless to suppose that the cover is minimal with respect to inclusion, and that the t_i's build an increasing sequence. Then it enjoys the following properties:
- two consecutive intervals $]t_i - \varepsilon_i/2, t_i + \varepsilon_i/2[$ and $]t_{i+1} - \varepsilon_{i+1}/2, t_{i+1} + \varepsilon_{i+1}/2[$ overlap;
- if either of the inclusions $t_i \in]t_{i+1} - \varepsilon_{i+1}, t_{i+1}[$, $t_{i+1} \in]t_i, t_i + \varepsilon_i[$ is false, the other one is true; hence $f_\lambda(t_i) - f_\lambda(t_{i+1}) > 0$;
- $a \in]t_0 - \varepsilon_0/2, t_0]$, hence $f_\lambda(a) - f_\lambda(t_0) > 0$;
- $b \in [t_n, t_n + \varepsilon_n/2[$, hence $f_\lambda(t_n) - f_\lambda(b) > 0$.

Summing the inequalities above yields $f_\lambda(a) - f_\lambda(b) > 0$. Since λ is arbitrary we get $f(a) - f(b) \geq 0$, which completes the proof. □

Lemma 4.2 *Let x and u be solutions of (SRB) and (SRBaux) with $u(t) \in \partial h(x(t))$ for all $t \geq t_0$. Then*

- **a.** $\limsup_{\tau\to 0, \tau\neq 0} \frac{1}{\tau}\{D(a, t+\tau) - D(a, t)\} \leq -\langle \nabla\Phi(x(t)), x(t) - a\rangle$ *for all $t \geq t_0$, and all a in dom h;*
- **b.** *If Φ is convex and $S = \{x, \Phi(x) = \inf_{dom\ h} \Phi\}$ is nonempty, then for all $a \in S$ the function $t \in [t_0, +\infty[\to D(a, t)$ is nonincreasing and $\lim_{t\to+\infty} D(a, t)$ exists and is nonnegative; further the solution x is bounded.*

Proof. a. For r, s in $]t_0, +\infty[$ with $u(r) \in \partial h(x(r))$, $u(s) \in \partial h(x(s))$, invoking the three points identity [8, Lemma 3.1] for $D(\cdot, \cdot)$ associated with the convex lower semicontinuous function h, we have

$$D(a, s) - D(a, r) = \langle u(s) - u(r), x(s) - a\rangle - D(x(r), s)$$

and thus, since $D(\cdot, \cdot) \geq 0$, we get

$$D(a, s) - D(a, r) \leq \langle u(s) - u(r), x(s) - a\rangle.$$

Hence, if $s > r$

$$\frac{1}{s - r}\{D(a, s) - D(a, r)\} \leq \langle \frac{u(s) - u(r)}{s - r}, x(s) - a\rangle.$$

Now choose $s = t + \tau$, $r = t$ if $\tau > 0$, and $s = t$, $r = t + \tau$ if $\tau < 0$ to obtain the desired conclusion.

b. Let a be a point in S. Then, by (a), for all $t \geq t_0$, we have $\limsup_{\tau\to 0, \tau\neq 0} \frac{1}{\tau}\{D(a, t+\tau) - D(a, t)\} \leq -\langle \nabla\Phi(x(t)), x(t) - a\rangle$. In view of the convexity inequality $\Phi(a) \geq \Phi(x(t)) + \langle \nabla\Phi(x(t)), a - x(t)\rangle$ we have further

$$\limsup_{\tau\to 0, \tau\neq 0} \frac{1}{\tau}\{D(a, t+\tau) - D(a, t)\} \leq \Phi(a) - \Phi(x(t)). \tag{4}$$

But $\Phi(x(t)) \geq \Phi(a)$, hence, with lemma (4.1), the function $t \mapsto D(a, t)$ is nonincreasing, and $\lim_{t\to+\infty} D(a, t)$ exists and is nonnegative.

Now for $t \geq t_0$ one has by definition:

$$D(a, t) = h(a) - h(x(t)) - \langle u(t), a - x(t)\rangle.$$

Since, h is strongly convex, it follows that

$$D(a, t) \geq \frac{\alpha}{2}|a - x(t)|^2.$$

Therefore, since $t \to D(a, t)$ is nonincreasing, one obtains for all $t \geq t_0$:

$$|a - x(t)|^2 \leq \frac{2}{\alpha} D(a, t) \leq \frac{2}{\alpha} D(a, t_0) < \infty,$$

showing that $\sup_{t\in[t_0, +\infty[} |x(t)| < \infty$, i.e., $x(t)$ is bounded.□

Theorem 4.1 *Let us assume the hypotheses* **H1-3**. *Then*

- **a.** *The function $t \in [t_0, +\infty[\mapsto \Phi(x(t)) \in \mathbb{R}$ is nonincreasing along all solution x of (SRB).*
- **b.** *If Φ is convex, then $\Phi(x(t))$ converges to the infimum value of Φ on dom h as $t \to +\infty$.*

Proof. a. Equations (2) and (3) in theorem 3.1 entail

$$\frac{d}{dt}\Phi(x(t)) \leq -\alpha|\dot{x}(t)|^2$$

almost everywhere on $[t_0, +\infty[$. Since $\Phi \circ x$ is locally Lipschitz continuous it is nonincreasing.

b. Let a be any point in dom h. From (4), for all $t \geq t_0$, we have

$$\limsup_{\tau\to 0, \tau\neq 0} \frac{1}{\tau}\{D(a, t+\tau) - D(a, t)\} \leq \Phi(a) - \Phi(x(t)).$$

Fix some $T > t_0$. Since $\Phi \circ x$ is nonincreasing, we thus have for all $t \in [t_0, T]$

$$\limsup_{\tau \to 0, \tau \neq 0} \frac{1}{\tau}\{D(a, t+\tau) - D(a,t)\} \leq \Phi(a) - \Phi(x(T)).$$

In view of lemma 4.1, the function $t \in [t_0, T] \mapsto D(a,t) - [\Phi(a) - \Phi(x(T))]t$ is nonincreasing, and we have

$$D(a,T) - T[\phi(a) - \Phi(x(T))] \leq D(a,t_0) - t_0[\phi(a) - \Phi(x(T))],$$

$$\Phi(x(T)) \leq \Phi(a) + \frac{1}{T - t_0}\{D(a,t_0) - D(a,T)\},$$

$$\Phi(x(T)) \leq \Phi(a) + \frac{1}{T - t_0} D(a, t_0),$$

since $D(a,T) \geq 0$. Hence $\lim_{t \to +\infty} \Phi(x(t)) \leq \Phi(a)$ for all $a \in \text{dom } h$. On the other hand, since $x(t) \in \text{dom } h$ entails $\Phi(x(t)) \geq \inf_{\text{dom } h} \Phi$, we obtain $\lim_{t \to +\infty} \Phi(x(t)) = \inf_{\text{dom } h} \Phi$. $\square$

We now turn to the asymptotic behaviour of a solution x to (SRB) in the case of a convex polyhedral constraint set included in a finite dimensional space. Recall (see [13, p. 162]) that a face of a convex set C is a convex subset F of C such that every (closed) line segment in C with a relative interior point in F has both endpoints in F.

Lemma 4.3 *Let C be a closed polyhedral convex subset of a finite dimensional Hilbert space.*

- **a.** *Let F be a face of C; let a, α be two points of F with $\alpha \in ri(F)$; let ν belong to $N_C(\alpha)$. Then $\langle \nu, \alpha - a \rangle = 0$.*
- **b.** *Let a, b two points in C, along with two sequences $a_n \in C$ and $\nu_n \in N_C(a_n)$ satisfying: $a_n \to a$ and $\lim_{n \to \infty} \langle \nu_n, a - b \rangle$ exists. Then $\lim_{n \to \infty} \langle \nu_n, a - b \rangle \geq 0$.*

Proof. **a.** For any real t in some neighbourhood of 0, the point $(1-t)\alpha + ta$ belongs to face F; hence $0 \leq \langle \nu, \alpha - [(1-t)\alpha + ta] \rangle = t\langle \nu, \alpha - a \rangle$.
b. As a polyhedral convex set, C is the finite union of the relative interiors of its faces ([13, th. 18.2, th. 19.1]). Hence there must exist a face F along with a subsequence $a_{\sigma(n)}$ of a_n such that $a_{\sigma(n)} \in \text{ri}(F)$, $a \in F$. Part **a** then entails $\langle \nu_{\sigma(n)}, a_{\sigma(n)} - a \rangle = 0$. Hence $0 \leq \langle \nu_{\sigma(n)}, a_{\sigma(n)} - b \rangle = \langle \nu_{\sigma(n)}, a - b \rangle$; whence $\lim_{n \to +\infty} \langle \nu_{\sigma(n)}, a - b \rangle = \lim_{n \to +\infty} \langle \nu_n, a - b \rangle \geq 0$. $\square$

Theorem 4.2 *Assume hypotheses* **H1-3** *with H finite dimensional. Assume further*
. Φ is convex and the set $S = \{x \in C, \Phi(x) = \inf_C \Phi\}$ is nonempty;
. $h = k + \delta_C$, where C is a closed convex polyhedron and $k : H \mapsto \mathbb{R}$ is convex continuously differentiable.
Then $x(t)$ converges to a minimum point of Φ on C as $t \to +\infty$.

Proof. With lemma 4.2, and since Φ is convex and S is nonempty, x is bounded and admits cluster points in dom $h = C$. Inspired by a lemma of Opial [12], next we assume that x admits two cluster points a and b in dom h. In view of theorem 4.1 and of the continuity of $\Phi \circ x$, a and b belong to S. Then the quantity

$$D(b,t) - D(a,t) = h(b) - h(a) - \langle u(t), b - a \rangle \tag{5}$$

has a limit as $t \to +\infty$. Let α_n, β_n be two sequences with $\alpha_n \to +\infty$, $\beta_n \to +\infty$, $x(\alpha_n) \to a$, $x(\beta_n) \to b$ as $t \to +\infty$. Then specializing (5) to the sequences α_n and β_n, and taking the limit we get

$$\lim_{n \to +\infty} \langle u(\beta_n), b - a \rangle = \lim_{n \to +\infty} \langle u(\alpha_n), b - a \rangle. \tag{6}$$

In view of $h = k + \delta_C$, we have $\partial h = \nabla k + N_C$; let us write $u(t) = \nabla k(x(t)) + \nu(t)$ with $\nu(t) \in N_C(x(t))$. The equality (6) reads

$$\lim_{n \to +\infty} \langle \nu(\alpha_n), a - b \rangle + \lim_{n \to +\infty} \langle \nu(\beta_n), b - a \rangle + \langle \nabla k(b) - \nabla k(a), b - a \rangle = 0, \tag{7}$$

and further since a, b lie in C, together with the strong convexity of h it follows that,

$$\lim_{n \to +\infty} \langle \nu(\alpha_n), a - b \rangle + \lim_{n \to +\infty} \langle \nu(\beta_n), b - a \rangle + \alpha |b - a|^2 \leq 0. \tag{8}$$

Owing to lemma 4.3 the first two terms in (8) are nonnegative; hence $|b - a| = 0$ and thus $x(t)$ converges to one point which is a minimum point of Φ in C. $\square$

5 The SRB dynamics as a singular perturbation limit of interior gradient systems

In this section, we turn to one of the questions raised in the introduction: what is the dynamical system obtained as a limit of (D_μ) when $\mu \to 0$? Section 2 has already shown that the dynamical systems (RLV_μ), with $h_\mu(x) = \frac{\nu}{2}|x|^2 + \mu \sum_{j=1}^p x_j \ln x_j$, admit (LVVB), with $h(x) = \frac{\nu}{2}|x|^2 + \delta_{\mathbb{R}_+^p}(x)$, as a limit when $\mu \to 0$. Thus pointwise convergence of the h_μ family, which would yield $x \mapsto \frac{\nu}{2}|x|^2$ as a limit, is not the right notion of convergence in order to obtain the limit dynamical system (LVVB). Instead Mosco convergence for functionals ensures $h_\mu \to h, \mu \to 0$ and proves to be the right notion.

To the previous assumptions **H1-3** we add the following extra hypotheses:

H4. H is a finite dimensional Hilbert space;

H5. for $\mu > 0$, $h_\mu : H \mapsto \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ is a family of uniformly strongly convex functions: i.e.,

$$h_\mu(\lambda x + (1-\lambda)y) \leq \lambda h_\mu(x) + (1-\lambda)h_\mu(y) - \frac{\alpha}{2}\lambda(1-\lambda)|y-x|^2$$

for all x, y in the domain of h_μ and all $\lambda \in]0,1[$; or equivalently

$$(u,v) \in \partial h_\mu(x) \times \partial h_\mu(y) \ \Rightarrow\ \langle v-u, y-x\rangle \ \geq\ \alpha|y-x|^2;$$

H6. the family h_μ converges to h in the sense of Mosco, as $\mu \to 0$;

H7. $x_{0,\mu}$ is a family in H satisfying $x_{0,\mu} \in \mathrm{dom}\ \partial h_\mu$ for each $\mu > 0$, and converging to $x_0 \in \mathrm{dom}\ \partial h$ as $\mu \to 0$;

H8. the set $\bigcup_{\mu>0} \partial h_\mu(x_{0,\mu})$ is bounded in H.

Mosco convergence is a variational notion of convergence; its formulation, in the finite dimensional Hilbert space H, is the following (see [2, proposition 3.19]):

- for all $x \in H$, there exists a family x_μ converging to x such that $h_\mu(x_\mu)$ converges to $h(x)$;
- for all $x \in H$, for all family x_μ converging to x, the inequality $h(x) \leq \liminf_{\mu>0,\mu\to 0} h_\mu(x_\mu)$ holds.

The key property of the Mosco convergence $h_\mu \to h$ is its entailing the graph convergence of the subdifferential operators $\partial h_\mu \to \partial h$ (see [2, theorem 3.66]), which means that, for every (x,u) with $u \in \partial h(x)$, there exists a family (x_μ, u_μ) with $u_\mu \in \partial h_\mu(x_\mu)$, $x_\mu \to x$, $u_\mu \to u$, as $\mu \to 0$.

The graph convergence $\partial h_\mu \to \partial h$ in turn implies that, for any family (x_μ, u_μ) with $u_\mu \in \partial h_\mu(x_\mu)$, $x_\mu \to x$, $u_\mu \to u$, then $u \in \partial h(x)$ (see [2, proposition 3.59]).

In accordance with theorem 3.1, for each $\mu > 0$ the problem

$$(\mathrm{SRB}_\mu) \qquad \frac{d}{dt}\partial h_\mu(x(t)) + \nabla\Phi(x(t)) \ni 0,$$

with initial data $x(t_0) = x_{0,\mu}$ admits at least a solution x_μ in $W^{1,1}_{\mathrm{loc}}([t_0,+\infty[;H) \bigcap \mathcal{C}([t_0,+\infty[,H)$.

Theorem 5.1 *Under the hypotheses* **H1-8**, *every family of solutions x_μ of (SRB_μ) with initial data $x_\mu(t_0) = x_{0,\mu}$ admits a sequence $x_{\mu(n)}$, with $\mu(n) \to 0$, $n \to +\infty$, converging to a solution x of (SRB) in the following sense:*

- $x_{\mu(n)} \to x$, $n \to +\infty$ *in* $\mathcal{C}([t_0,T];H)$ *for all* $T > t_0$,
- $\dot{x}_{\mu(n)} \rightharpoonup \dot{x}$, $n \to +\infty$ *weakly in* $\mathcal{L}^2([t_0,+\infty[;H)$.

If (SRB) has a unique solution x, then the whole family x_μ converges to x in the above-mentioned sense as $\mu \to 0$.

Proof. For every $\mu > 0$, there exists a pair of functions (x_μ, u_μ) in $W^{1,1}_{\mathrm{loc}}([t_0,+\infty[;H)$ which satisfy

$$\dot{u}_\mu(t) + \nabla\Phi(x_\mu(t)) = 0, \quad \forall t \geq t_0, \quad , x_\mu(0) = x_{0,\mu} \tag{9}$$

$$u_\mu(t) \in \partial h_\mu(x_\mu(t)), \quad \forall t \geq t_0; \tag{10}$$

Set $u_{0,\mu} = u_\mu(0)$. Since the family $u_{0,\mu}$ lies in $\bigcup_{\mu>0} \partial h_\mu(x_{0,\mu})$, which is bounded, we may extract some sequence $u_{0,\mu(n)}$, with $\mu(n) \to 0$, $n \to +\infty$, converging to some u_0 in H.

As in the proof of theorem 3.1 we may show that the ordinary differential equation

$$\dot{u}(t) + \nabla\Phi(\nabla h^*(u(t))) = 0, \quad u(t_0) = u_0 \tag{11}$$

has a unique solution u in $W^{1,1}_{\text{loc}}([t_0,+\infty[;H)$. Define the function x in $W^{1,1}_{\text{loc}}([t_0,+\infty[;H)$ by $x(t)=\nabla h^*(u(t))$. In view of the graph-convergence of ∂h_μ towards ∂h, we have $u_0\in\partial h(x_0)$ and, so, x is a solution of (SRB).

Coming back to (9) and (10), as in the proof of theorem 3.1 we can prove

$$\alpha\int_{t_0}^t |\dot{x}_\mu(\tau)|^2 d\tau \le \Phi(x_{0,\mu})-\inf_H \Phi,\quad \forall t\ge t_0.$$

Since the sequence $x_{0,\mu(n)}$ is convergent and since Φ is continuous, the quantities $\Phi(x_{0,\mu(n)})$ are uniformly bounded above. Hence the sequence $\dot{x}_{\mu(n)}$ is bounded in $\mathcal{L}^2([t_0,+\infty[;H)$. For any fixed $T>t_0$, the sequence $x_{\mu(n)}$ is thus uniformly equicontinuous on $[0,T]$ in view of

$$|x_{\mu(n)}(t)-x_{\mu(n)}(s)|\le\int_s^t|\dot{x}_{\mu(n)}(\tau)|d\tau\le\sqrt{t-s}\sqrt{\int_{t_0}^T|\dot{x}_{\mu(n)}(\tau)|^2d\tau}$$

for $t_0\le s\le t\le T$; further the ranges of the functions $x_{\mu(n)}$ lie in a bounded subset of H. Ascoli's theorem then asserts the existence of a subsequence, still denoted with $x_{\mu(n)}$, converging to some $\bar{x}$ in $\mathcal{C}([t_0,T];H)$ as $n\to+\infty$.

Equation (9) then shows that $\dot{u}_{\mu(n)}$ admits a limit in $\mathcal{C}([t_0,T];H)$ as $n\to+\infty$. Since $u_{0,\mu}$ converges to u_0, $u_{\mu(n)}$ admits a limit, say $\bar{u}$, in $\mathcal{C}([t_0,T];H)$ which satisfies

$$\dot{\bar{u}}(t)+\nabla\Phi(\bar{x}(t))=0,\quad\forall t\in[t_0,T],\quad \bar{u}(0)=u_0.$$

But in view of the graph-convergence of ∂h_μ to ∂h equation (10) yields in the limit

$$\bar{u}(t)\in\partial h(\bar{x}(t)),\quad\forall t\in[t_0,T].$$

Now the well-posedness of (11) shows that $\bar{u}$ is the restriction of u to $[t_0,T]$; likewise $\bar{x}$ is the restriction of x to $[t_0,T]$. This settles the first convergence result.

The weak convergence $\dot{x}_{\mu(n)}\rightharpoonup\dot{x}$, $n\to+\infty$ in $\mathcal{L}^2([t_0,+\infty[;H)$ (up to another subsequence) is a consequence of the boundedness of $\dot{x}_{\mu(n)}$ in $\mathcal{L}^2([t_0,+\infty[;H)$.

Finally the last assertion is an obvious consequence of the uniqueness of the solution x to (SRB).□

Theorem 5.1 is particularly illustrative when applied to Riemannian gradient systems, such as (RLV$_\mu$), with a metric h_μ involving a Legendre function acting as a barrier.

Corollary 5.1 *Assume* **H1-4**. *Let h_μ be of the form $h_\mu(x)=\frac{\nu}{2}|x|^2+\mu\theta(x)$, where θ is a nonnegative Legendre function. Set $C=\overline{\text{dom}\,\theta}$, and let x_0 belong to int C. Then (SRB$_\mu$) with initial data $x_\mu(t_0)=x_0$ and (SRB) with initial data $x(t_0)=x_0$ admit unique solutions x_μ and x respectively. Further the whole family x_μ converges to x as $\mu\to 0$ in the following sense*

- $x_\mu\to x$, *in $\mathcal{C}([t_0,T];H)$ for all $T>t_0$,*
- $\dot{x}_\mu\rightharpoonup\dot{x}$, *weakly in $\mathcal{L}^2([t_0,+\infty[;H)$.*

Proof. Observe first that **H5** is true with $\alpha=\nu$ and that **H7-8** are true with $x_{0,\mu}\equiv x_0$. Now **H6**, that is the Mosco convergence of h_μ to $h=\frac{\nu}{2}|.|^2+\delta_C$, is a consequence of the pointwise monotone decreasing convergence of h_μ to $\frac{\nu}{2}|.|^2+\delta_{\text{dom }h}$ (see [2, theorem 3.20]). The uniqueness of the solutions x_μ and x to (SRB$_\mu$) and (SRB) is then a consequence of $\partial h_\mu(x_0)=\{\nu x_0\}$ and $\partial h(x_0)=\{\nu x_0\}$ being singletons (see theorem 3.1). The convergence asserted in the corollary follows immediately from theorem 5.1 and from the uniqueness of the solution of (SRB).□

6 Extensions of the (SRB) system.

A closer look at the proof of theorem 3.1 suggests that its assumptions may be relaxed: the equation may be more complex and, to some extent, the subgradient ∂h may be replaced by a monotone operator.

6.1. In addition to the hypotheses **H1-3** we assume the following

H9. $F:H\mapsto H$ a globally Lipschitz continuous operator with $L\ge 0$ as Lipschitz modulus;

H10. $\beta\ge 0$ a real constant.

We now turn to the following dynamical system

$$\text{(XSRB)}\qquad \frac{d}{dt}\partial h(x(t))+\beta\partial h(x(t))+\nabla\Phi(x(t))+F(x(t))\ni 0.$$

Definition. A function $x : [t_0, +\infty[\mapsto H$ will be termed a solution of (XSRB) with initial data $x(t_0) = x_0 \in$ dom ∂h if and only if

1. $x \in W^{1,1}_{\text{loc}}([t_0, +\infty[; H) \cap \mathcal{C}([t_0, +\infty[; H)$,
2. $x(t_0) = x_0$,
3. there exists a function $u \in \mathcal{C}^1([t_0, +\infty[; H)$ such that
   . $u(t) \in \partial h(x(t))$, $\forall t \in [t_0, +\infty[$,
   . $\dot u(t) + \beta u(t) + \nabla\Phi(x(t)) + F(x(t)) = 0$, $\forall t \in [t_0, +\infty[$.

As in section 3 we are led to consider an auxiliary regular ordinary differential equation for the function u, with some initial value u_0 arbitrary in $\partial h(x_0)$

$$\text{(XSRBaux)} \qquad \dot u(t) + \beta u(t) + \nabla\Phi(\nabla h^*(u(t))) + F(\nabla h^*(u(t))) = 0.$$

Theorem 6.1 *Under the assumptions* **H1-3**, **H9-10**, *and for any $x_0 \in$ dom ∂h*

a. *x is a solution of (XSRB) with Cauchy data $x(t_0) = x_0$ if and only if there exist $u_0 \in \partial h(x_0)$ and a solution $u : [t_0, +\infty[\mapsto H$ of (XSRBaux) such that $u(t_0) = u_0$ and $x(t) = \nabla h^*(u(t))$ for all $t \in [t_0, +\infty[$;*

b. *(XSRB) with initial data $x(t_0) = x_0$ possesses one solution at least;*

c. *if $\partial h(x_0)$ is a singleton, (XSRB) with initial data $x(t_0) = x_0$ has exactly one solution.*

Proof. Only assertion **b** deserves a proof. In view of the local Lipschitz continuity of the operators $\nabla\Phi \circ \nabla h^*$ and $F \circ \nabla h^*$, the Cauchy problem (XSRBaux) admits a maximal local solution $u : [t_0, t_{\max}[\mapsto H$, which is continuous together with its derivative $\dot u$. To show that it is global indeed, it is enough to show that it is bounded on a left neighbourhood of $t_{\max}$, if $t_{\max}$ is supposed finite.

Define $x : [t_0, t_{\max}[\mapsto H$ by $x(t) = \nabla h^*(u(t))$, so that u and x satisfy

$$\dot u(t) + \beta u(t) + \nabla\Phi(x(t)) + F(x(t)) = 0. \tag{12}$$

Let t_1, t_2, with $t_1 < t_2$, be two arbitrary points in $[t_0, t_{\max}[$ to be specified later. Let t vary in the interval $[t_1, t_2]$. For simplicity, from now on, we do not make the variable t explicit, and we write u_1, u_2, x_1, x_2 in lieu of $u(t_1)$, $u(t_2)$, $x(t_1)$, $x(t_2)$. Observe that x is differentiable almost everywhere, and at every differentiability point of x we have

$$\langle \dot u, \dot x\rangle + \beta\langle u, \dot x\rangle + \langle \nabla\Phi(x), \dot x\rangle + \langle F(x), \dot x\rangle = 0. \tag{13}$$

Our aim is to derive a bound for $\int_{t_1}^{t_2} |\dot x|$. Let us examine, and perform some calculation on, each term involved in this equality.

First, we have $\langle \dot u, \dot x\rangle \ge \alpha|\dot x|^2$, and integrating this inequality on $[t_1, t_2]$ yields

$$\int_{t_1}^{t_2} \langle \dot u, \dot x\rangle \ge \alpha \int_{t_1}^{t_2} |\dot x|^2 \ge \frac{\alpha}{t_2 - t_1}\left(\int_{t_1}^{t_2} |\dot x|\right)^2 \ge \frac{\alpha}{t_{\max} - t_1}\left(\int_{t_1}^{t_2} |\dot x|\right)^2. \tag{14}$$

Secondly, we have $\langle u, \dot x\rangle = \frac{d}{dt}\langle u, x\rangle - \frac{d}{dt}h^*(u)$, and integrating this equality on $[t_1, t_2]$ yields

$$\int_{t_1}^{t_2} \langle u, \dot x\rangle \;=\; \langle u_2, x_2\rangle - \langle u_1, x_1\rangle - h^*(u_2) + h^*(u_1).$$

In view of the convexity inequality $h^*(u_1) \ge h^*(u_2) + \langle u_1 - u_2, x_2\rangle$ we further obtain

$$\int_{t_1}^{t_2} \langle u, \dot x\rangle \;\ge\; \langle u_1, x_2 - x_1\rangle. \tag{15}$$

Thirdly, integrating $\langle \nabla\Phi(x), \dot x\rangle$ on $[t_1, t_2]$ yields

$$\int_{t_1}^{t_2} \langle \nabla\Phi(x), \dot x\rangle = \Phi(x_2) - \Phi(x_1) \ge \inf_H \Phi - \Phi(x_1). \tag{16}$$

Fourthly, we have

$$\begin{aligned} \langle F(x), \dot x\rangle &= \langle F(x) - F(x_1), \dot x\rangle + \langle F(x_1), \dot x\rangle \\ &\ge -L|x - x_1|\,|\dot x| + \langle F(x_1), \dot x\rangle \\ &\ge -L\left(\int_{t_1}^{t_2} |\dot x|\right)|\dot x| + \langle F(x_1), \dot x\rangle. \end{aligned}$$

Integrating $\langle F(x), \dot{x}\rangle$ on $[t_1, t_2]$ yields

$$\int_{t_1}^{t_2} \langle F(x), \dot{x}\rangle \geq -L\left(\int_{t_1}^{t_2} |\dot{x}|\right)^2 + \langle F(x_1), x_2 - x_1\rangle. \tag{17}$$

Now, integrating (13) on $[t_1, t_2]$, and taking the inequalities (14), (15), (16), (17) into account, yields

$$\left(\frac{\alpha}{t_{\max} - t_1} - L\right)\left(\int_{t_1}^{t_2} |\dot{x}|\right)^2 + \langle \beta u_1 + F(x_1), x_2 - x_1\rangle + \inf_H \Phi - \Phi(x_1) \leq 0.$$

But in view of $\langle \beta u_1 + F(x_1), x_2 - x_1\rangle \geq -|\beta u_1 + F(x_1)|\,|x_2 - x_1| \geq -|\beta u_1 + F(x_1)| \int_{t_1}^{t_2} |\dot{x}|$ we finally obtain

$$\left(\frac{\alpha}{t_{\max} - t_1} - L\right)\left(\int_{t_1}^{t_2} |\dot{x}|\right)^2 - |\beta u_1 + F(x_1)| \int_{t_1}^{t_2} |\dot{x}| + \inf_H \Phi - \Phi(x_1) \leq 0.$$

If we choose t_1 so close to $t_{\max}$ that $\alpha/(t_{\max} - t_1) - L > 0$, then the above inequality shows that $\int_{t_1}^{t_2} |\dot{x}|$ is uniformly bounded with respect to t_2 in $[t_1, t_{\max}[$. Hence x is bounded on $[t_1, t_{\max}[$. With (12) u is bounded too on $[t_1, t_{\max}[$.□

One of the interests of systems of (XSRB) form is that they provide with a first breakthrough into second order Riemannian barrier systems. Indeed, consider

$$\frac{d^2}{dt^2}\partial h(x(t)) + \nabla\Phi(x(t)) \ni 0, \tag{18}$$

where h complies with **H3** and Φ, in addition to complying with **H2**, is globally Lipschitz continuous. Define $k : H \times H \mapsto \bar{\mathbb{R}}$ and $F : H \times H \mapsto H \times H$ by $k(x, y) = h(x) + \frac{1}{2}|y|^2$ and $F(x, y) = (y, -\nabla\Phi(x))$. Then

$$\frac{d}{dt}\partial k(x(t), y(t)) + F(x(t), y(t)) \ni 0$$

is an (XSRB) system in $H \times H$, which explicitly reads

$$\begin{aligned} &\frac{d}{dt}\partial h(x(t)) + y(t) \ni 0 \\ &\dot{y}(t) - \nabla\Phi(x(t)) = 0. \end{aligned}$$

This way, (18) may be given a sense, and is relevant to theorem 3.1.

6.2. (XSRB) systems may still be further extended to encompass monotone operators in place of subdifferential operators (but with $\beta = 0$).

In addition to the hypotheses **H1-3**, **H9** we assume the following

H11. $T : H \rightrightarrows H$ is a maximal monotone operator subject to the strong monotonicity condition

$$(u, v) \in Tx \times Ty \Rightarrow \langle v - u, y - x\rangle \geq \alpha|y - x|^2.$$

We now turn to the following dynamical system

$$\text{(XXSRB)} \qquad \frac{d}{dt}Tx(t) + \nabla\Phi(x(t)) + F(x(t)) \ni 0.$$

Definition. A function $x : [t_0, +\infty[\mapsto H$ will be termed a solution of (XXSRB) with initial data $x(t_0) = x_0 \in \operatorname{dom} T$ if and only if

1. $x \in W^{1,1}_{\text{loc}}([t_0, +\infty[; H) \cap \mathcal{C}([t_0, +\infty[; H)$,

2. $x(t_0) = x_0$,

3. there exists a function $u \in \mathcal{C}^1([t_0, +\infty[; H)$ such that

. $u(t) \in Tx(t)$, $\forall t \in [t_0, +\infty[$,

. $\dot{u}(t) + \nabla\Phi(x(t)) + F(x(t)) = 0$, $\forall t \in [t_0, +\infty[$.

In view of property **H11**, the maximal monotone operator T is invertible with T^{-1} everywhere defined and Lipschitz continuous with $1/\alpha$ as Lipschitz modulus (see [7, proposition 2.2]). Thus, as in section 3, we are led to consider an auxiliary regular ordinary differential equation for fonction u, with some initial value u_0 arbitrary in Tx_0

$$\text{(XXSRBaux)} \qquad \dot{u}(t) + \nabla\Phi(T^{-1}u(t)) + F(T^{-1}u(t)) = 0.$$

The following theorem can be proved as theorem 6.1

Theorem 6.2 *Under the assumptions* **H1-3**, **H9**, **H11**, *and for any* $x_0 \in dom\, \partial h$

- **a.** *x is a solution of (XXSRB) with Cauchy data $x(t_0) = x_0$ if and only if there exist $u_0 \in Tx_0$ and a solution $u : [t_0, +\infty[\mapsto H$ of (XXSRBaux) such that $u(t_0) = u_0$ and $x(t) = T^{-1}u(t)$ for all $t \in [t_0, +\infty[$;*
- **b.** *(XXSRB) with initial data $x(t_0) = x_0$ possesses one solution at least;*
- **c.** *if Tx_0 is a singleton, (XXSRB) with initial data $x(t_0) = x_0$ has exactly one solution.*

In view of this theorem which guarantees the existence of a solution to (XXSRB), theorem 5.1 can be extended to encompass maximal monotone operators. Indeed, the main ingredients in the proof of theorem 5.1 are the graph convergence of the subdifferential operators ∂h_μ to ∂h and the uniform strong monotonicity condition for ∂h_μ and ∂h as expressed by **H5**. With an immediate adaptation of its hypotheses, theorem 5.1 can be thus extended to maximal monotone operators.

References

[1] F. Alvarez, J. Bolte, O. Brahic (2004) Hessian riemannian gradient flows in convex programming, *SIAM Journal on Control and Optimization*, (to appear).

[2] H. Attouch (1984) Variational Convergence for Functions and Operators, Pitman Advanced Publishing Program.

[3] H. Attouch, M. Teboulle (2004) A regularized Lotka-Volterra dynamical system as a continuous proximal-like method in Optimization, *JOTA*, (to appear).

[4] H. Attouch, M. Théra (1996) A general duality principle for the sum of two operators, *J. Convex Anal.*, **3**, n° 1, 1–24.

[5] D. A. Bayer, J. C. Lagarias (1989) The nonlinear geometry of linear programming: Part I, Affine and projective scaling trajectories; Part II, Legendre Transform coordinates and central trajectories, *Trans. of AMS*, 314, 499-526 and 527-581.

[6] J. Bolte, M. Teboulle (2003) Barrier operators and associated gradient-like dynamical systems for constrained optimization problems, *SIAM J. Control Optim.*, **42**, 1266–1292.

[7] H. Brézis (1973) Opérateurs maximaux monotones, *Mathematics Studies 5*, North-Holland-American Elsevier.

[8] G. Chen, M. Teboulle (1993) Convergence analysis of a proximal-like algorithm using Bregman functions, *SIAM J. Optimization*, **3**, 538-543.

[9] X. Goudou, oral communication.

[10] A. N. Iusem, B. F. Svaiter, J. F. Da Cruz (1999) Centrals paths, generalized proximal point methods, and Cauchy trajectories in Riemannian manifolds, *SIAM J. Control Optim.*, Vol. 37, n° 2, 566-588.

[11] N. Kenmochi, I. Pawlow (1986) A class of doubly nonlinear elliptic-parabolic equations with time dependent constraints, *Nonlinear Analysis*, Vol. 10, n° 11, 1181-1202.

[12] Z. Opial (1967) Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. of the AMS*, 73, 591-597.

[13] R. T. Rockafellar (1970) Convex Analysis, Princeton University Press.

[14] R.T. Rockafellar, R. J. B Wets (1998) Variational Analysis, Springer Verlag, New York.